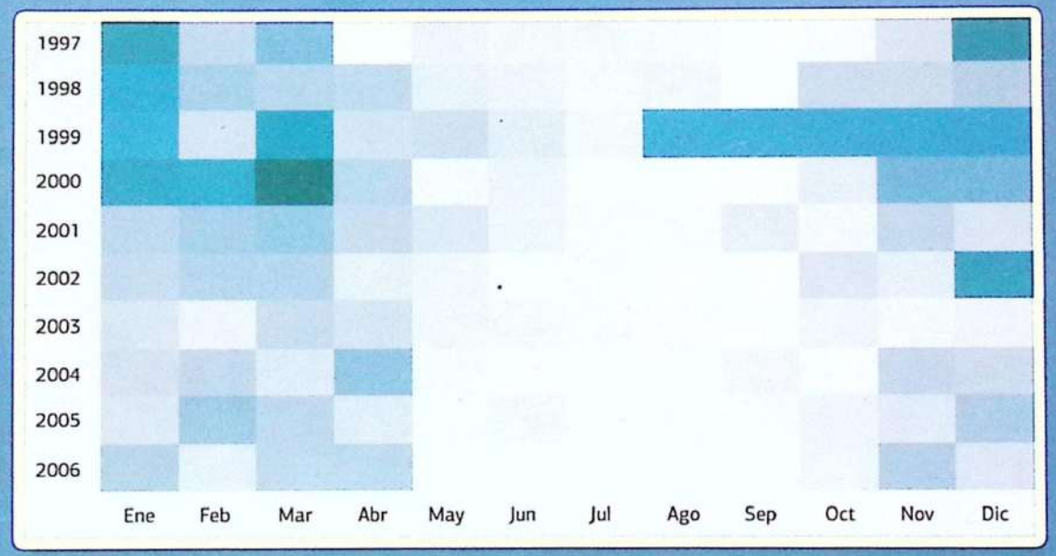
Publicaciones

Raúl Pedro Mentz



UNIVERSIDAD NACIONAL DE TUCUMÁN | AÑO 2008

Este libro realiza un justo homenaje al **Profesor Raúl Pedro Mentz,** Profesor Emérito de la Universidad Nacional de Tucumán e Investigador principal del CONICET. Él es Profesor de la Facultad de Ciencias Económicas de la UNT, que aún jubilado, prefiere continuar trabajando en su Universidad, donde desempeña sus funciones con el mismo entusiasmo de antes. Es también Director del Instituto de Investigaciones Estadísticas, órgano apoyado por el CONICET.

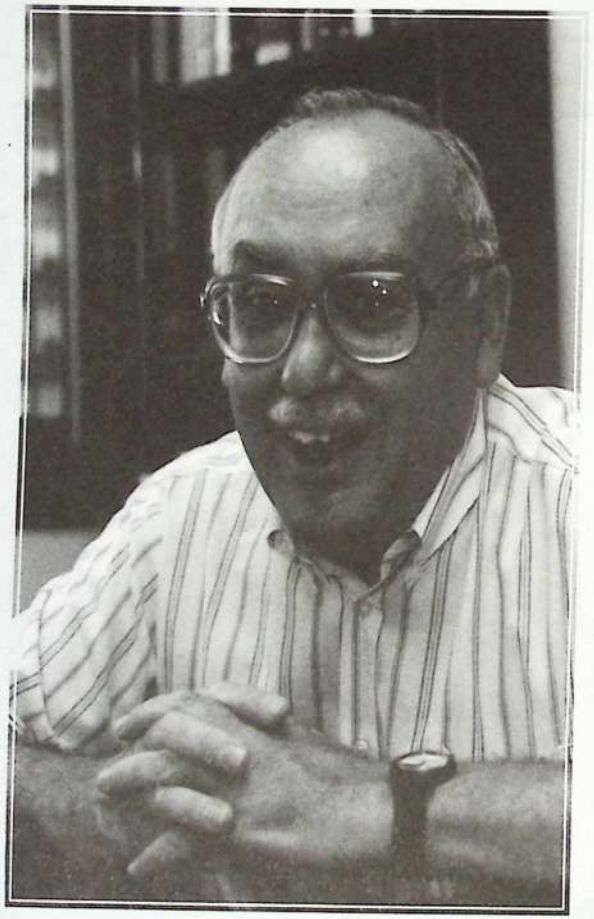
UNIVERSIDAD NACIONAL DE TUCUMÁN | AÑO 2008

Dr. Raúl Pedro Mentz
Profesor Emérito de la Universidad Nacional de Tucumán. Profesor de Estadística de la Facultad de Ciencias Económicas de la UNT. Investigador Principal del Consejo Nacional de Investigaciones Científicas y Técnicas. Director del Instituto de Investigaciones Estadísticas y del "Magíster en Estadística Aplicada" y co-director del Doctorado en Estadística de la Facultad de Ciencias Económicas de la Universidad Nacional de Tucumán.

Publicaciones

Raúl Pedro Mentz

UNIVERSIDAD NACIONAL DE TUCUMÁN | AÑO 2008



Raúl Pedro Mentz

"Dedico esta obra a mi esposa
y a la memoria de mis padres"

ÍNDICE

Índice comentado ____ 6

Prólogo *de Pedro Alberto Morettin* ____ 10

Prólogo *de J. Ricardo Acosta* ____ 12

Prólogo *de Raúl Pedro Mentz* ____ 15

Publicaciones

1. La Estadística Bayesiana: Una descripción de sus ideas básicas. (1967). *Estadística*, pp. 61-80 ____ 17

2. Regression Models for Seasonal Analysis. (1971) con V.J.E. *Estadística*, pp. 476-485. ____ 40

3. Ejemplos sobre Propiedades Asintóticas de los Estimadores. (1971), *Económica*, pp. 367-370. ____ 51

4. Inverting a Covariance Matriz of Toeplitz Type by the Method of Cofactors. (1975), *Revista de la Unión Matemática Argentina*, 27, pp. 106-108. ____ 57

5. A Note on Inverting a Covariance Matrix of Toeplitz Type. ____ 71

6. On the Inverse of Some Covariance Matrices of Toeplitz Type. (1976), SIAM *Journal of Applied Mathematics*, 31(3), pp- 426-437. ____ 77

7. Estimation in the First Order Moving Average Model Based on Sample Autocorelations. (1977), *The Annals of Statistics*, 5(6), pp. 1250-1257. ____ 91

8. Estimation in the First-Order Moving Average Model Through the Finite Autoregressive Approximation. (1977), *Journal of Econometrics*, 6, pp. 225-236. ____ 101

9. The Generalized Variance of a Stationary Autoregressive Process. (1977) con T.W. Anderson. *Journal of Multivariate Analysis*, 7(4), pp. 584-588. ____ 115

10. Modelos de Regresión para Estimaciones Estacionales. (1979) con V.J. Elías. *"Métodos Cuantitativos en las Ciencias Sociales, Ensayos en memoria de Dr. Jos, Barral Souto"*, Ediciones Macchi, pp. 77-92. ____ 121

11. Analysis of Lead-Lag Relations by Cross-Spectral Methods. (1979), Manila, Filipinas, Actas, pp. 341-344. ____ 139

12. On the Structure of the Likelihood Function of Autoregressive and Moving Average Models. (1980) con T.W. Anderson. *Journal of Time Series Analysis*, 1(2), pp. 83-94 ____ 145

13. Encuesta a Centros Estadísticos Latinoamericanos. (1981) con N.M. Jarma. Buenos Aires, Actas de Sesiones invitadas, Tomo I, pp. 378-389. ____ 159

14. Notes on the Estimation of Parameters in Vector Autoregressive Models. (1983) con T.W. Anderson, en P.J. Bickel et al (editores). *"A Festchrift for Erich L. Legmann"*, Wadsworth. ____ 181

15. A Note on the use of the Sample Autocorrelogram. (1988), *Revista Brasileira de Probabilidades y Estadística*, 2, pp. 7-22. ____ 195

16. Estimación en los modelos autorregresivos y de promedios móviles. (1988), *Estadística Española*, 116, pp. 87-106. ____ 211

17. Teaching of Statistics in Latin America. (1989) con P.A. Morettin. *Estadística*, 41(137), pp. 95-103. ____ 233

18. The Inter-American Statistical Institute: Work Program for Regional Change. (1990) con M.E. González y P.A. Morettin. *Estadística*, 42(138), pp. 17-26. ____ 243

19. Evaluation of Quadratic Forms and Traces for Iterative Estimation in First-Order Moving Average Models. (1993) con T.W. Anderson. *Communications in Statistics, Theory and Methods*, 22(4), pp. 931-963. ____ 253

20. Simulations of Iterative Procedures for Maximum Likelihood Estimation in MA(1) Models. (1996) con T.W. Anderson, N.M. Jarma y C.I. Martínez. *Communications in Statistics. Simulation & Computation*, 25(4), pp. 851-865. ____ 287

21. Residual Variance Estimation in Moving Average Models.(1997) con P.A. Morettin y C.M.C. Toloi. *Communication in Statistics. Theory and Methods*, 26(8), pp. 1905-1923. ____ 305

22. On Least-Squares Estimation of the Residual Variance in the First-Order Moving Average Model. (1999) con P.A. Morettin y C.M.C. Toloi. *Computational Statistics & Data Analysis*, 29, pp. 485-499. ____ 327

23. On Residual Variance Estimation in Autoregressive Models. (1998) con P.A. Morettin y C.M.C. Toloi. *Journal of Time Series Analysis*, 19(2), pp. 187-208. ____ 347

24. Cyclical Components of Local Rainfall Data. (2000) con M.A. D'urso, N.M. Jarma y G.B. Mentz. *Internacional Journal of Climatology, Int. J. Climatology*, 20(211), pp. 211-228. ____ 375

25. Likelihood Ratio Test of Equality of Several Variances in the Intraclass Correlation Model. (2001), *Communication in Statistics. Theory and Methods*, 30(1), pp. 75-91. ____ 395

26. Bias Correction for Estimators of the Residual Variance in the ARMA(1,1) Model. (2001) con P.A. Morettin y C.M.C. Toloi. *Estadística*, 53(160, 161), pp. 1-40. ____ 415

27. Robust Estimation in Time Series. (2002) con C.I. Martínez. *Sociedad de Estadística e Investigación Operativa*, 11(2), pp. 385-404 ____ 457

28. On Confidence Bands for Time Series Problems in the Time and Frequency Domains. (2003) con A.J.V. y C.I.M. *Brazilian Journal of Probability and Statistics*, 17, pp. 179-211. ____ 479

29. Correlación, Simetría y Variabilidad. (2003), *Revista de Educación Matemática*, pp. 20-42. ____ 513

30. Nota sobre la Asignación de Fechas a los Picos y Valles de una Serie de Lluvias Mensuales. (2004) con N.M. Jarma. *Estadística*, 56(166 y 167), pp. 21-34. ____ 539

31. Bootstrap Estimation of Standard Errors in Seasonal Analysis.(2005) con N.M. Jarma y C.I. Martínez. *Revista Brasileña de Estadística*, 66(225), pp. 63-108. ____ 553

32. Note on the Autoregressive Spectral Estimator. (2006) con C.I. Martínez. *Brazilian Journal of Probability and Statistics*, 20, pp. 49-66. ____ 601

Índice comentado

En este libro se reúnen copias de 32 trabajos publicados. El criterio para formar este conjunto es que hubiese separatas disponibles en el momento de recopilarlas. Con posterioridad ser descubrieron algunas omisiones importantes que serán comentadas mas abajo.

Para facilitar la lectura a continuación se incluyen breves explicaciones.

Un primer grupo lo forman los trabajos con números 1, 2, 3 y 10. Estos son trabajos publicados alrededor de los años 1965 a 1968, antes de la tesis doctoral. En estos años otros resultados obtenidos no fueron enviados a revistas, sino que fueron difundidos en forma de informes técnicos. Tal es el caso de las publicaciones provenientes de un contrato con la Universidad Nacional de Tucumán para estudiar temas relacionados con el desempeño de alumnos de grado en diversas carreras.

El trabajo de tesis realizó en la Universidad de Stanford, bajo la dirección del muy prestigioso profesor T.W. Anderson, y los temas fueron del área de Series Cronológicas. Un primer grupo lo constituyen las publicaciones 4, 5 y 6. El tema es la búsqueda de una forma explícita para los componentes de la matriz inversa de una matriz de Toeplitz: esta tiene estructura simétrica, con cinco diagonales centrales distintas de cero. El principal trabajo publicado es el número 6, y (como corresponde) apareció en una revista de matemática aplicada.

La forma explícita de los componentes de la matriz inversa citada, se utilizó, junto con otras técnicas, sobre todo de análisis asintótico, para resolver el problema tratados en las publicaciones 7 y 8. Se buscaban formas explícitas para el estimador por máxima verosimilitud del principal parámetro de un modelo de promedios móviles de primer orden. Se trataron las propuestas de los profesores James Durban y Morris Walter. El Primero (publicación 8) se basa en aproximar el promedio móvil por una autorregresión del orden "suficientemente grande" mientras que el segundo (publicación 7) hace el mismo tipo de aproximación utilizando autocorrelaciones. Estos trabajos fueron resumidos en la exposición número 16.

El próximo bloque de publicaciones corresponde a trabajos postdoctorales realizados en forma con junto con el Profesor Anderson. Son las publicaciones 9, 12, 14, 19 y 20. El trabajo número 9 es un resultado de tipo matemático y consiste en obtener una expresión explícita para el determinante de la matriz de varianzas y covarianzas de un modelo autorregresivo; el resultado era conocido en la literatura matemática, pero encontramos procedimientos simples para obtener los resultados deseados.

La publicación 12 trata el importante asunto de las condiciones bajo las cuales existen los estimadores máximo verosímiles (bajo normalidad). Resultó que para el modelo de promedios móviles no es necesario recurrir a condiciones especiales, mientras que para el modelo autorregresivo se requieren condiciones a cumplir con los datos. El trabajo 14 trata algunas extensiones del tema modelos vectoriales.

La publicación 19 trata ya concretamente de los resultados requeridos para poder computar los estimadores máximo verosímiles. Por lo tanto cierra un conjunto de trabajos: 7 y 8 tratan la obtención de estimadores explícitos usando aproximaciones, 12 trata el tema de la existencia de soluciones, 19 los procedimientos para realizar los cómputos y finalmente 20 contiene simulaciones para verificar los resultados citados.

Las publicaciones 13, 17 y 18 tratan aspectos de la enseñanza de estadística en Argentina y en otros países de América. La publicación 18 en particular, describe una propuesta para reorganizar las actividades del IASI.

Con los profesores Pedro Morettin y Clelia Toloi de la Universidad de San Pablo en Brasil encaramos el tema de la estimación de la varianza del término de error ("varianza residual") en modelos ARMA (autorregresivos con residuos de promedios móviles). Los resultados constituyen las publicaciones 21, 22, 23 y 26. Se estudiaron matemáticamente y se ilustró con simulaciones, las estimaciones por los métodos de los momentos, cuadrados mínimos y máxima verosimilitud, en modelos autorregresivos, de promedios móviles y mixtos.

Las publicaciones 29 y 15 tratan respectivamente, de las correlaciones y de las autocorrelaciones. El primero es una nota de tipo didáctico.

Las publicaciones 24 y 30 corresponden a trabajos de tipo aplicado, y están basadas en el procesamiento de datos mensuales sobre precipitaciones.

Las publicaciones 28 y 32 versan sobre análisis espectral, la primera cuando la estimación se realiza usando estimadores con "ventanas", y la segunda cuando se usa el enfoque del estimador autorregresivo. De tipo aplicado es la publicación 11.

Las publicaciones 25, 27 y 31 tratan temas distintos: ña primera incursiona en el importante tema de la "correlación intraclase" y fue motivado por el trabajo número 20, la segunda explora el efecto sobre la estimación de los parámetros, de la presencia de errores con distribución de probabilidad que es una mezcla de distribuciones normales, y finalmente en 30 se utiliza la técnica "bootstrap" para estimar los errores estándares de ciertos estimadores proporcionados en el programa "X-11 ARIMA".

Omisiones importantes advertidas. Al concluir el trabajo descripto, se advirtió la omisión involuntaria de algunos trabajos a los que asigno considerable importancia. Entre ellos los siguientes:

33. Estudio del componente estacional de las series cronológicas mediante el análisis espectral, (1968) Estadística, 518-546.

34. Carlos Eugenio Dieulefalt, 1901-1982, (1980) *Journal of the Royal statistical Society*, Series A, 537-538.

35. Statistics in Latin America (1985) por P.A. Morettin, C.A. Pérez, S.C. Narula y R.P. Mentz, *The American Statistician*, 39, 4 Part 1, 274-278.

36. Modelling and forecasting linear combinations of time series (1987) *Review of the International Statistical Institute*, 349-360.

37. Sobre la historia de la estadística oficial Argentina, (1991) *Estadística Española*, 33, 128, 501-532.

38. Sobre la historia de la enseñanza de la estadística en las universidades Argentinas, (1991) por R.P. Mentz y V.J. Yohai, *Estadística Española*, 33, 128, 533-538.

39. Graduate statistical training in Argentina and Brazil, (1991) por *R.P. Mentz y P.A. Morettin Estadística*, 43, 115-127.

Libros publicados

40. La estadística Bayesiana. Elementos para un Curso introductorio con Aplicaciones a los Negocios y a la Economía, (1967), Instituto Torcuato Di Tella, Buenos Aires.

41. Métodos Estadísticos para el Análisis Cíclico y Estacional (1989), conjunto con E. de Alba, A. Espasa y P.A. Morettin, IASI, Instituto Interamericano de Estadística, Panamá.

42. Estacionalidad: Introducción a los Modelos y Métodos Estadísticos, (2002) Ediciones del Rectorado, Universidad Nacional de Tucumán.

Citaciones

El trabajo 6 fue citado varias veces. Por ejemplo en F.A. Braybill, Matrices with Applications in Statistics, Second Edition (1969), Wadsworth and Brooks/Cole, página 284.

Los trabajos 7 y 8 también fueron citados, ver, por ejemplo, M.B. Priestley, Spectral Analysis and Time Series, Volumen 1, 1981, Academic Press, página 359.

La publicación 12 dio lugar a trabajos destinados a extender el caso vectorial, los resultados obtenidos sobre existencia de estimadores máximo verosímiles en modelos autorregresivos.

Prólogo *de Pedro Alberto Morettin*

Este libro realiza un justo homenaje al Profesor Raúl Pedro Mentz, Profesor Emérito de la Universidad Nacional de Tucumán e Investigador principal del CONICET. Él es Profesor de la Facultad de Ciencias Económicas de la UNT, que aún jubilado, prefiere continuar trabajando en su Universidad, donde desempeña sus funciones con el mismo entusiasmo de antes. Es también Director del Instituto de Investigaciones Estadísticas, órgano apoyado por el CONICET.

Conocí a Raúl hace más de 30 años, durante un viaje de estadísticos latinoamericanos a Estados Unidos, por invitación de la ASA –American Statistical Asociation. Desde entonces percibí su entusiasmo y anhelo de ayudar a desarrollar la Estadística en Latinoamérica. Y esto se materializó en diversos trabajos realizados para el IASI (Interamerican Statistical Institute), ya sea como vicepresidente o presidente, o como editor de su revista Estadística, a la cual ayudó a reestructurar. En el ámbito de su país, estableció los programas de Magíster en Estadística Aplicada y Doctorado en Estadística en la UNT.

Durante todos estos años, Raúl cultivó una gran diversidad de intereses en el área de la Estadística, todos ellos reflejados en los trabajos reunidos en este libro. Encontramos tres publicaciones relacionadas con el trabajo realizado para el IASI: En la primera se efectuó un relevamiento de los centros de la región que poseían algún programa de grado o posgrado en Estadística, cuyos resultados fueron presentados en la Sección del ISI realizada en Buenos Aires, en 1981. La segunda está relacionada con la enseñanza de Estadística en Latinoamérica y la tercera con la actuación del IASI en la región.

Las otras publicaciones que encontramos en este libro reflejan otros intereses. Hay publicaciones relacionadas con la estacionalidad en series de tiempo; un asunto de particular interés para Raúl por muchos años, y que le llevó a publicar un excelente libro sobre métodos estacionales. Otra área de interés, que motivó la publicación de varios trabajos, es aquella de estimación de modelos autoregresivos y de promedios móviles, univariados y multivariados, en algunos de los cuales tuve el inmenso placer en participar. El interés de Raúl en esta última área se remonta a su tesis de doctorado en Stanford, bajo la orientación de T.W. Anderson, sobre las propiedades asintóticas de los estimadores en modelos de promedios móviles.

En el inicio de su carrera encontramos importantes trabajos, publicados en revistas internacionales de renombre, sobre la inversión de matrices de tipo Toeplitz, también reproducidos en este volumen. Completan la lista de publicaciones, trabajos sobre Análisis Espectral y Correlación y trabajos aplicados a las áreas de Economía y Climatología. Hay, todavía una curiosidad: Un artículo publicado en 1967, en la revista Estadística, sobre el Análisis Bayesiano.

Para mi es motivo de orgullo el haber sido invitado a escribir este prólogo sobre Raúl y haber compartido su amistad y la de su familia por más de tres décadas.

Pedro Alberto Morettin
São Paulo, Julho de 2008.

Prólogo *de J. Ricardo Acosta*

"...I've lived a life that's full
I traveled each and every highway
And more, much more than this,
I did it my way..."
(Paul Anka, "My Way")

Hay hombres que pueden enorgullecerse de haber hecho siempre, a lo largo de su vida, lo correcto en el momento correcto. Y luego ser reconocidos por ello.

Casado desde 1956 con Guillermina E. Fierro (Coca), padre de tres hijos (María Isabel, Graciela y Raúl) y abuelo de once nietos, puede afirmarse que la vida de Raúl Mentz -mi abuelo- ha sido y es fructífera en todas sus múltiples facetas.

Nació el 16 de Diciembre de 1932, en el seno de una familia con muchas aspiraciones, que le inculcó desde pequeño la avidez por el conocimiento y la superación personal. Así lo indican sus anécdotas sobre las clases de piano –que no le gustaban demasiado, según lo reconoce- y de inglés, a las cuales asistió a instancias de sus padres, en un tiempo en que el estudio de este idioma no estaba tan difundido como en el presente.

Mi abuelo fue siempre un buen estudiante. Se destacó tanto en su paso por la Escuela Julio A. Roca, como en su posterior tránsito por la Escuela de Comercio y la –por entonces joven- Facultad de Ciencias Económicas de la Universidad Nacional de Tucumán.

Esto mientras alternaba con los pasatiempos habituales de una época que no conocía la TV ni las computadoras. Así, gustaba de departir con amigos, jugaba al billar o a las cartas, iba al cine, y no se perdía de ningún baile, según cuenta mi abuela (con cierto tono de reproche, pero aun así entre risas).

Capítulo aparte merece su sana obsesión por la lectura, hábito que más tarde sabría fomentar en los suyos. Pocas personas conocieron tanto a tan corta edad sobre la vida y obra de Borges –por ejemplo-, como quienes compartimos desde pequeños los almuerzos y cenas de mi abuelo.

Otra de sus sanas obsesiones es la matemática. Su afecto por esta ciencia y su vocación incansable de maestro convertían nuestras visitas de la infancia en una sucesión de pasatiempos interesantes, como aprender que los números primos son aquellos divisibles únicamente por uno y por sí mismos, y que la denominación no alude a ninguna

relación de parentesco. Esto, en la temprana niñez, representaba todo un hallazgo (que además era evaluado al fin de semana siguiente).

Su idilio con la estadística comenzó, lógicamente, de la mano de su afición por la matemática. Stanford, el INIE y lo que sigue es historia conocida. Si bien no puedo precisar la magnitud de lo que hizo mi abuelo por y para la comunidad estadística, confío en que semejantes precisiones serán materia de otros prólogos y se me excusará de brindarlas en el presente. De cualquier modo, el homenaje que hoy se le rinde resulta muestra elocuente de la trascendencia de sus actividades, y es materia de orgullo para su descendencia.

Como cuando junto a mis hermanos y primos lucíamos altivos una remera con la leyenda *"I'm statistically significant"* (comprada en el "gifts-shop" de algún Congreso), e intentábamos reproducir frente a nuestros amigos la explicación del abuelo de que un individuo no es significativo en estadística y de allí la genialidad de la frase. Hay que reconocer que pocos –nosotros incluidos- terminaban de comprenderlo.

Lo cierto es que a lo largo de su vida, Raúl Mentz ha alcanzado el difícil mérito de ser un hombre grande y completo, como docente e investigador, maestro, estadístico y cabeza ejemplar de una familia. Por ello su agasajo, lejos de venir impuesto por el protocolo, representa un reconocimiento justo por parte de sus colegas, discípulos y allegados.

Porque no muchos hombres pueden, como mi abuelo, mirar hacia atrás y afirmar que han vivido una vida plena, que recorrieron cada uno de los caminos, y que lo hicieron a su manera, que ha sido, no caben dudas, la manera correcta.

J. Ricardo Acosta
Agosto de 2008

Prólogo *de Raúl Pedro Mentz*

Como parte del homenaje que generosamente organizó el IASI (Instituto Interamericano de Estadística), se decidió publicar una colección de separatas de trabajos que realicé durante muchos años, desde 1967.

Los tópicos cubiertos son variados, pero hay una clara orientación hacia la teoría y las aplicaciones de los métodos de análisis de series cronológicas. Este fue el tema de mi tesis doctoral (Ph. D. en Stanford en 1975).

Tuve la fortuna de compartir una buena porción de estos proyectos, con estadísticos proveniente de distintos tipos de extracciones. Uno fue el director de mi tesis, Profesor T. W. Anderson, otro el colega Profesor Pedro A. Morettin. En épocas recientes he realizado trabajos que enfatizan el uso del computador, para lo que conté con el aporte de Nora Jarma y de Carlos Martínez.

Estoy sumamente agradecido al IASI y a sus autoridades actuales, en particular a la Presidenta, Licenciada Clyde Trabuchi y al Secretario Técnico, Profesor Evelio Fabbroni.

También deseo agradecer al Profesor Pedro A. Morettin por su apoyo en este proyecto. Esto se une a la labor de muchos de los miembros del Instituto de Investigaciones Estadísticas (INIE) de la Facultad de Ciencias Económicas de la Universidad Nacional de Tucumán.

Raúl Pedro Mentz
Agosto de 2008

La estadística bayesiana: una descripción de sus ideas básicas

Raúl Pedro Mentz, Argentina

Summary

Over a period of roughly the last ten years, various statisticians have originated and developed a current of thinking known as the "Bayesian approach" to decision statistics. In essence, this new trend combines numerous elements of analysis in a fairly original manner known as "Bayesian." Strictly speaking, however, this designation is not altogether correct. The well-known Bayes' Theorem (of Thomas Bayes, published posthumously in 1763) is utilized more frequently in this theory than in other formulations, but this results, among other things, from an attitude that differs from the traditional view regarding what should be interpreted as probabilities. Nevertheless, the name "Bayesian" is accepted and serves to indicate the position with which we are concerned here.

In their modern version, these techniques have spread rapidly, particularly in some European countries and in the United States. They were at first developed as methods for analyzing situations peculiar to business, but a variety of applications were rapidly found and suggested in other fields. Relatively few works have appeared on this topic in Spanish, however, which leads us to the belief that even now, there is still justification for presenting this paper which, without adding new elements to the body of theory concerned, sets forth its principal characteristics in a compact but orderly way and shows the relationship between it and other more traditional methods.

Accordingly, this paper presents (1) a brief review of the traditional concepts relating to tests of statistical hypotheses, including the use of characteristic error curves for the selection of the best test; (2) the "decisionist" way of selecting this best test—along the lines of Wald, and more appropriately, according to the Bayesian field of thought; (3) the "subjectivist" way of choosing the best test; and (4) the Bayesian way of deciding a given problem statistically.

Although the aim of this work is to arrive at topics (3) and (4), we believe that the order in which the topics are presented will facilitate comparison of the theory with which we are concerned and its predecessors.

1. Introducción

Durante los diez últimos años aproximadamente, diversos estadísticos han iniciado y desarrollado una corriente de pensamiento a la que se individualiza como el "enfoque bayesiano" de la estadística de las decisiones. En esencia la nueva corriente de pensamiento *combina* de manera relativamente original numerosos elementos de análisis y es a esa manera de combinarlos que se refiere el nombre de "bayesiano."

---

* Trabajo presentado en el V Coloquio Argentino de Estadística, Rosario, Santa Fe, octubre 1965. Constituye el Cuaderno No. 65-4 del Instituto de Investigaciones Económicas, Facultad de Ciencias Económicas, Universidad Nacional de Tucumán.

** Facultad de Ciencias Económicas, Universidad Nacional de Tucumán.

En rigor, la designación es bastante impropria, pues si bien es cierto que en esta teoría se utiliza con más frecuencia que en otras formulaciones el celebrado "Teorema de Bayes" (original del Reverendo Thomas Bayes y publicado póstumamente en 1763), ello es una consecuencia, entre otras cosas, de una actitud diferente a la tradicional con respecto a lo que debe interpretarse como probabilidades. Sin embargo, el nombre de "bayesiano" está aceptado y sirve para individualizar a la posición que nos ocupa.

Estas técnicas alcanzaron rápida difusión en su versión moderna, especialmente en algunos países de Europa y en Estados Unidos. En principio se desarrollaron como métodos para el análisis de situaciones propias de los negocios, pero rápidamente se han encontrado y se han sugerido aplicaciones variadas fuera de ese campo.

A pesar de esta amplia difusión que comentamos, han sido relativamente pocos los trabajos en español que se han dado a conocer sobre el tema, de manera que creemos que todavía se justifica en este momento un trabajo como el presente, que sin agregar nuevos elementos al cuerpo de la teoría que nos ocupa, exponga sus características principales en forma compacta pero razonada y muestre la relación que dicha teoría guarda con los métodos más tradicionales.

Para cumplir con estos propósitos, presentamos sucesivamente: (1) una rápida revisión de los conceptos tradicionales de los "test de hipótesis estadísticas," incluyendo el uso de curvas características de error para la selección del "mejor test"; (2) la manera "decisionista"—en la línea de Wald—de seleccionar ese "mejor test"; y más propiamente en el campo de las ideas "bayesianas"; (3) la manera "bayesiana" de decidir estadísticamente un problema dado.

En rigor, el objeto de este trabajo es llegar a los temas (3) y (4), pero creemos que el orden que hemos elegido para la presentación de los tópicos facilita la comparación de la teoría que nos ocupa con sus predecesoras.

En el curso de nuestra exposición no damos referencia específica sobre la fuente de cada aseveración; en su reemplazo, presentamos una bibliografía comentada que creemos que llena con ventaja esa función.

2. El diseño de un test de hipótesis estadísticas

1. Nuestro razonamiento arranca desde una de las primeras formas de inferencia probabilística: el intervalo de confianza.

Para fijar mejor las ideas, consideremos el caso particular de una variable aleatoria (estadígrafo muestral) Z, con esperanza $E(Z)$ y desviación estándar $\sigma(Z)$. Por consideraciones teóricas, quizás vinculadas al tamaño de las muestras, o por experiencia que se tiene sobre la naturaleza de la situación física analizada o meramente como una aproximación conveniente, se acepta que Z se distribuye normalmente.

Sea $k_{0.05}$ el número real tal que si X es una variable aleatoria normal con esperanza nula y variancia unitaria, se cumple que

$$Pr\ (X \geqslant k_{0.05}) = 0.05 \tag{1}$$

de manera que $-k_{0.05} = k_{0.05}$. Vemos que

$$Pr\left[-k_{0.05} \leq \frac{Z - E(Z)}{\sigma(Z)} \leq k_{0.05}\right] = 0.90 \tag{2}$$

y también que

$$Pr\left[\frac{Z - E(Z)}{\sigma(Z)} \geq -k_{0.05}\right] = 095\,. \tag{3}$$

De estas probabilidades se pueden derivar formalmente, transponiendo términos dentro de los paréntesis, un *intervalo de confianza del 90 por ciento* en el primer caso, o un *límite superior de confianza del 95 por ciento* en el caso de la Fórmula (3). Por ejemplo, si $\sigma(Z)$ fuera conocida o estimada a satisfacción del investigador, tal que no exista incertidumbre con respecto a su verdadero valor, el intervalo

$$\left[Z - k_{0.05} \cdot \sigma(Z),\ Z + k_{0.05} \cdot \sigma(Z)\right] \tag{4}$$

es el intervalo de confianza del 90 por ciento para $E(Z)$ y

$$\left[-\infty,\ Z + k_{0.05}\,\sigma(Z)\right] \tag{5}$$

el límite superior de confianza del 95 por ciento para $E(Z)$.[1]

Para hacer resaltar la naturaleza de este modelo, no escribimos el intervalo (4), por ejemplo, en la forma de una probabilidad, pues $E(Z)$ no es una variable aleatoria sino un parámetro y en consecuencia carece de distribución de probabilidad.

Una forma habitual de aclarar la naturaleza de (4) y (5) consiste en decir que bajo una interpretación probabilística de pruebas repetidas, intervalos de confianza de este tipo, computados para muchos valores $z_1, z_2, \ldots$ del estadígrafo Z, cubrirán el verdadero (y habitualmente desconocido) valor de $E(Z)$ en el 90 y 95 por ciento de los casos respectivamente.

2. Claramente esta misma línea de argumento puede utilizarse para probar hipótesis (estadísticas) con respecto a $E(Z)$. Así, dada una muestra y un valor z_o computado en ella, "con el 90 por ciento de confianza aceptamos la hipótesis de que $E(Z) = E_o(Z)$" si, y solamente si, $E_o(Z)$ cae dentro del intervalo

$$\left[z_o - k_{0.05} \cdot \sigma(Z),\ z_o + k_{0.05} \cdot \sigma(Z)\right], \tag{6}$$

para el caso de un test "de dos colas" y simétrico en probabilidad.

[1] La distinción entre "intervalo" y "límite" no es esencial y a menudo utilizaremos la primera palabra en forma genérica.

3. En el caso particular de los test de hipótesis, es muy importante advertir los efectos de un diseño descuidado de la investigación. Supongamos que en una muestra determinada z_o haya resultado igual a 25 y el intervalo de confianza (6) sea 25 ± 9 ó (16.34). Si bien en la mayoría de los casos el investigador se sentirá (personalmente) muy confiado al aceptar que $E_o(Z)$ es 25 ó 24 ó 26, su actitud (personal) una vez que se conoce el resultado muestral puede ser mucho menos segura si debe aceptar que E_o (Z) es 33.9 o rechazar que es 34.1, a pesar que el modelo le indica clara e inequívocamente el juicio que debe emitir.

En resumen, sería muy ventajoso que en la etapa de diseño se pudieran considerar, de una manera práctica, los *riesgos* de cada decisión en un test estadístico y así adecuar el tamaño de la muestra y la ubicación del punto crítico, por ejemplo, a los riesgos tolerables para cada valor del parámetro. La manera de evaluar estos riesgos en un plano *probabilístico* que ha adquirido mayor difusión, es la de construir las *curvas características operantes* (o las *curvas de potencia*) para un test determinado. Estas curvas (o funciones) permiten tomar en consideración no sólo el valor del parámetro que corresponde a la hipótesis de nulidad ($E_o(Z)$ en nuestro caso), sino todo un intervalo de valores posibles del parámetro y frecuentemente todos sus valores posibles.

En general, la curva característica operante (CCO) en un test dado, es la representación gráfica de la probabilidad de aceptar la hipótesis de nulidad, como función del parámetro que se investiga y la curva de potencia la complementaria a la unidad, o sea la que muestra la probabilidad de rechazar.

Para un caso como el que corresponde a (4), o sea un test simétrico de dos colas en la distribución normal, la CCO es:

$$f(E(Z)) = Pr\left[E(Z) - k_{0.05} \cdot \sigma(Z) \leq Z \leq E(Z) + k_{0.05} \cdot \sigma(Z)\right] \quad (7)$$

Para un test de una sola cola como el que se deriva de (5), del tipo "acéptese la hipótesis de nulidad si $z \leqslant z_c$ y recháces si $z > z_c$" la CCO es

$$f(E(Z)) = Pr\left[\frac{Z - E(Z)}{\sigma(Z)} \leq k\,\alpha\right] = Pr\left[Z \leq E(Z) + k\,\alpha \cdot \sigma(Z)\right], \quad (8)$$

donde claramente se verifica que $z_c = E(Z) + k\alpha \,.\, \sigma\,(Z)$ y α es el "nivel de significación" del test (0.10 en (7)).

Por ejemplo, para probar la hipótesis de nulidad de que $E(Z) = E_o(Z)$ contra la hipótesis alternativa de que $E(Z) = E_1(Z)$ donde es $E_o(Z) < E_1(Z)$, cuando $\sigma(Z)$ es conocido, el tipo de test que corresponde a (5) es apropiado y las representaciones gráficas de la CCO (8) y de la función de potencia tienen la forma esquemática de los Gráficos 1 y 2, respectivamente.

Una formulación más realista es la siguiente. Por consideraciones propias de la situación que se está investigando, se ha llegado a establecer que el valor $E_o(Z)$ del parámetro que se quiere probar es el *valor de indiferencia (breakeven point)*, de manera que resulta propiado que la hipótesis de nulidad sea, por ejemplo, que $E(Z) \leqslant E_o(Z)$ y la

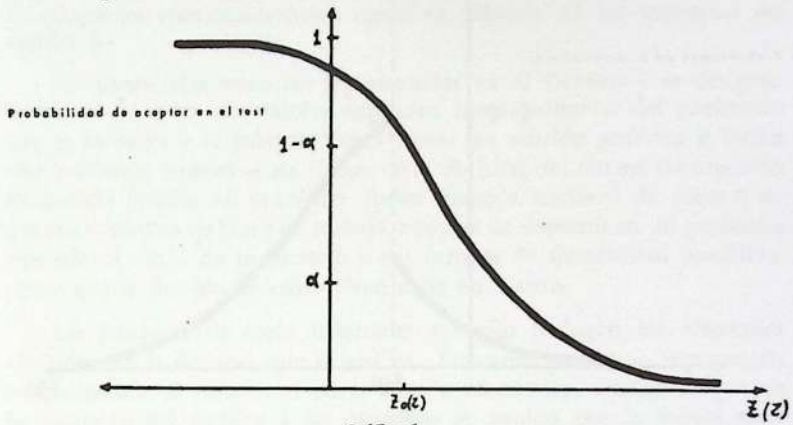


Gráfico 1

hipótesis alternativa que $E(Z) > E_o(Z)$. Para este tipo de test, a menudo resulta más ilustrativo reemplazar la CCO o la función de potencia por la *curva característica de error* que muestra las porciones de las otras dos curvas en las partes que corresponde a una decisión errónea del test. El esquema del Gráfico 3 es típico de la forma de estas curvas.

4. Por lo menos en un plano teórico, el investigador podría plantear bajo la forma del Gráfico 3, una gran cantidad de las combinaciones posibles de test que podrían intentarse, cambiando la ubicación de z_c, modifi-

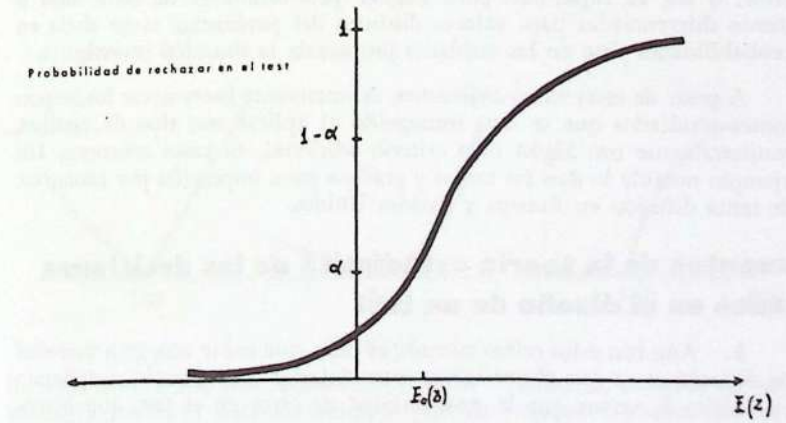


Gráfico 2

cando el tamaño de la muestra, o haciendo ambas cosas a la vez. [2] Con todo ese material a su disposición, conjuntamente con el costo de cada alternativa, estaría entonces en condiciones de decidir sobre el test óptimo.

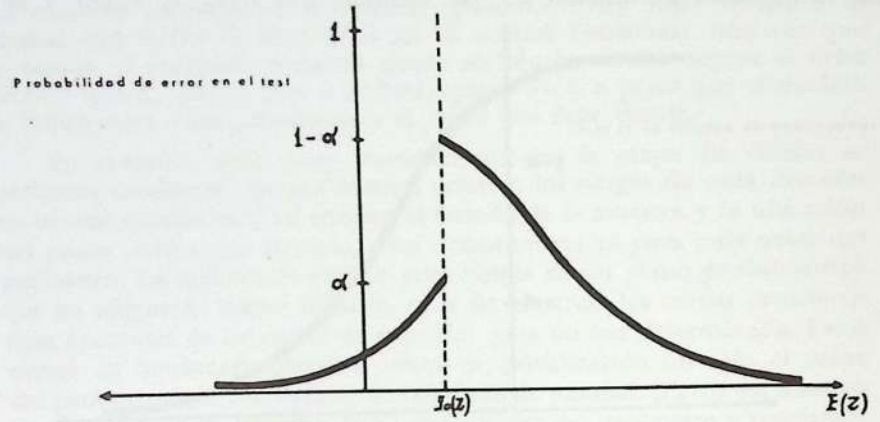


Gráfico 3

Obsérvese que la variedad de curvas características de error que existen para un caso como el que comentamos es infinita. Sin embargo, aún disponiendo de una experiencia limitada, se podría llegar por eliminación a una cantidad manejable.

Considérese además que si bien los costos de realización de los distintos test pueden determinarse bastante adecuadamente en dinero, el *poder discriminador de un test,* tal como aparece en la curva característica de error, o sea su capacidad para asignar probabilidades de error más o menos diferenciadas para valores distintos del parámetro, viene dado en probabilidades y no en las unidades propias de la situación investigada.

A pesar de estas serias objeciones, es interesante hacer notar los importantes resultados que se han conseguido al aplicar este tipo de análisis, conjuntamente con algún otro criterio adicional, en casos concretos. Un ejemplo notable lo dan las tablas y gráficos para inspección por muestreo, de tanta difusión en Europa y Estados Unidos.

3. Elementos de la teoría estadística de las decisiones utilizados en el diseño de un test

5. Aún con estos refinamientos, es claro que existe una gran variedad de situaciones en que el análisis muestra defectos. Por ejemplo, estudiando el Gráfico 3, vemos que la *probabilidad de error* en el test, que hemos

[2] También podría considerarse un "test estadístico secuencial," etc.

elegido como medida del riesgo en su decisión, es mayor en la proximidad del valor de parámetro que corresponde al punto de indiferencia, que para valores muchos más grandes o mucho más chicos que él. Sin embargo, por la propia naturaleza del valor de equilibrio, esperamos que los *riesgos medidos en las unidades originarias del problema* (costos, utilidad, etc.) se comporten aproximadamente como en algunos de los esquemas del Gráfico 4.

Funciones tales como las representadas en el Gráfico 4 se designan funciones de costo. Los valores dependen funcionalmente del parámetro que se investiga y la palabra "costo" tiene un sentido genérico e indica que la función representa los riesgos de la decisión del test en las unidades de medida propias del problema (pesos moneda nacional de costo o de ganancia, número de horas de trabajo, número de alumnos en un problema educacional, etc.). En lo sucesivo y por razones de simplicidad consideraremos que la función de costo viene dada en dinero.

Las funciones de costo habitualmente sólo incluyen los elementos afectados por la decisión que se analiza. En consecuencia no representan necesariamente el costo total en el sentido económico. Obsérvese que en los esquemas del Gráfico 4 las funciones se anulan por lo menos para $E_0(Z)$, precisamente para hacer resaltar esa situación.

Estas funciones de costo no son fáciles de obtener en la práctica y aún puede aseverarse que en aplicaciones científicas tradicionales, como en investigaciones biológicas, médicas, de comportamiento humano no económico, etc., es un verdadero problema separado discutir la naturaleza y aún la existencia de tales funciones. Sin embargo, por lo menos en muchos casos de la investigación económica, de la ingeniería y de los negocios, pueden proponerse funciones de costo si se efectúa un análisis detenido del problema. En tal caso, el hecho de que se deba preparar más

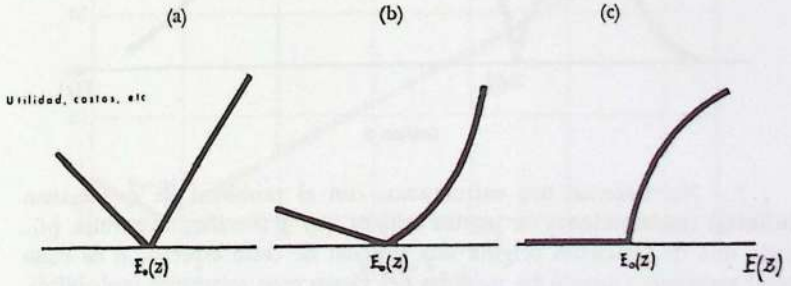


Gráfico 4

información original para aplicar las técnicas estadísticas, redunda en un beneficio posterior, pues los modelos pueden responder adecuadamente muchas preguntas.

6. Aceptada la existencia y admitida la ventaja del uso de la función de costo en un problema dado, nuestro próximo paso es emplearla junto con la curva característica de error en el diseño del test, para lograr una evaluación más apropiada del riesgo asociado con cada decisión. En lo sucesivo y conforme con la tradición en el tema, nos referiremos a "decisiones" y no a "decisiones del test," con la implicación de que, al decidir sobre la base de la estructura económica propia del problema, estamos decidiendo esencialmente éste y no su versión estadística.

La manera propuesta para establecer la vinculación entre costos y probabilidades de error, y posiblemente la que más apela al buen sentido, es ponderar los valores de la función de costo por las probabilidades de la curva característica de error. La función resultante, que depende también del parámetro estudiado, es designada la *función de costo esperado* y adopta aproximadamente la forma esquemática del Gráfico 5 si la curva característica de error es como la del Gráfico 3 y la función de costo como la del Gráfico 4 (a).

Si la función de costo esperado estuviera expresada en dinero, podríamos agregarle el costo de la información muestral que corresponde al test respectivo, el *costo del muestreo,* con lo que tendríamos nuevas funciones de forma similar a la del Gráfico 5. Estas son las *funciones de costo total esperado.*

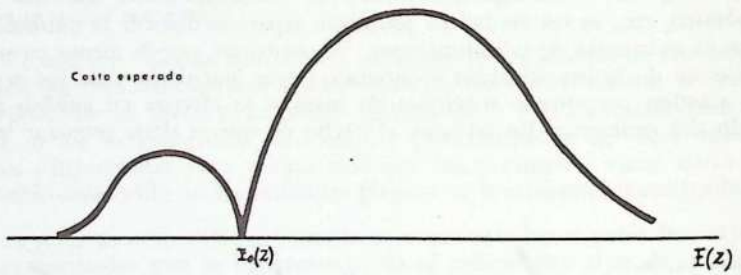


Gráfico 5

7. Nuevamente nos enfrentamos con el problema de que existen infinitas combinaciones de puntos críticos (z_c) y tamaños muestrales (n), cada una de las cuales origina una función de costo esperado o de costo total esperado. Cuando las medidas del riesgo eran solamente probabilidades confiábamos que el investigador pudiera usar su conocimiento del problema para seleccionar el "mejor" test considerando curvas como la del Gráfico 3. Como ahora las curvas de riesgo representan unidades propias del problema y no probabilidades como en los párrafos 3-4, podemos esperar que el juicio informal del investigador sea reemplazado por un *criterio* explícito para definir y seleccionar el "mejor test" de un conjunto dado de test equivalentes.

Una manera de hacer esta selección objetivamente, o sea en forma independiente de la opinión del investigador, es usando el criterio "minimax." Este criterio propone tener en cuenta el resultado más desfavorable de cada test posible (esto es, el máximo de la respectiva curva de costo o costo total esperado) y seleccionar como más conveniente el que tenga el *menor* (mínimo) de esos resultados *más* desfavorables (máximo); de ahí la expresión "minimax."

Por ejemplo, consideremos el siguiente caso (tomado de Robert Schlaifer, *Introduction to Statistics for Business Decisions,* McGraw-Hill, 1961). Se investiga el parámetro p, la verdadera proporción en la población; el valor de indiferencia de p es $p_o = 0.04$; se consideran test basados en una muestra al azar simple de tamaño $n = 20$; el estadígrafo muestral es r, la proporción en la muestra, que se supone sigue una distribución binomial; los test que se comparan son todos del tipo "acéptese que $p \leqslant 0.04$ si y solamente si $r < r_c$." Bajo estas condiciones, cuando la función de costo es como la del Gráfico 6, las curvas de costos esperados para los test donde r_c es 1, 2 y 3 respectivamente, aparecen en el Gráfico 7.

Es fácil advertir que el criterio "minimax" indicará que el test con $r_c = 1$ es superior a los otros dos.

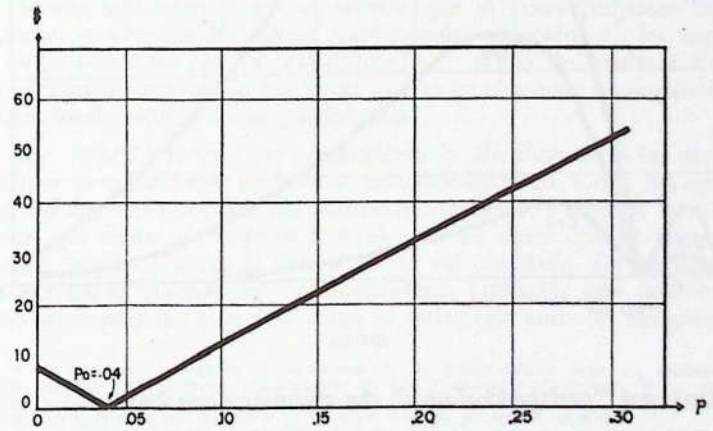


Gráfico 6

8. A pesar de que el criterio "minimax" posee una serie de propiedades muy atractivas y aún desde un punto de vista intuitivo es muy razonable, un análisis detenido permite advertir que presenta algunas dificultades importantes.

En particular, a menudo en este tipo de esquema, resulta imposible combinar explícitamente y como parte del modelo la función de costo esperado con otra información adicional que pudiera tenerse con respecto al parámetro estudiado. Esto representa un verdadero inconveniente pues, por ejemplo, si se tuviese información de mucha credibilidad, pero no necesariamente cierta, de que una situación especial puede representarse para algunos valores particulares del parámetro, sería deseable poder utilizar este conocimiento, efectivo aunque no cierto, para seleccionar el mejor test.

Consideraciones de esta naturaleza, llevan a un enfoque bastante diferente de todo el problema y éste es el tema que analizaremos en el resto de este trabajo.

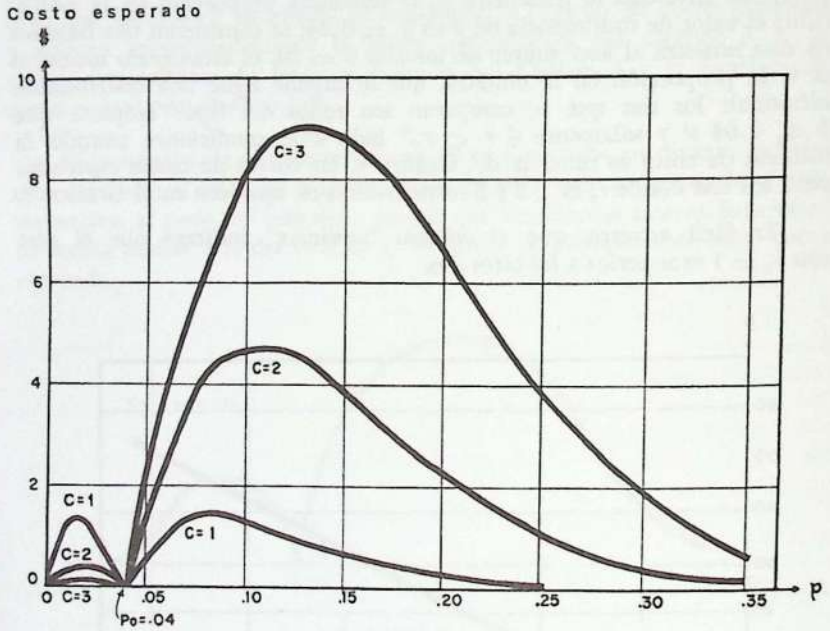


Gráfico 7

4. La manera "subjetivista" de diseñar un test

9. En todo nuestro análisis hasta este punto, hemos utilizado las probabilidades bajo una interpretación tradicional y objetiva de pruebas repetidas, vale decir, como límites de frecuencias relativas. En ningún momento hemos necesitado suponer que ellas representan medidas del grado de credibilidad que se tiene respecto de un suceso y que tienen la interpretación cotidiana que se da a la palabra "probabilidades."

Resulta que esta cuestión teórica pero básica debe ser dilucidada previamente, para poder aceptar un nuevo modo de analizar un problema como el planteado en el párrafo 7. Específicamente, la cuestión se centra en la aceptación de la interpretación subjetiva de las probabilidades.[3]

Si estamos dispuestos a dar ese paso y reconocemos a las probabilidades como medidas (posiblemente subjetivas) del grado de credibilidad de un suceso cualquiera, repetitivo o no, proponemos utilizar el siguiente razonamiento:

Por conocimiento que se tiene sobre la naturaleza del problema, se puede asignar una distribución de probabilidad a $E(Z)$. Vale decir, el parámetro que se investiga, que revestía ese carácter paramétrico bien definido en los modelos descriptos en los párrafos 1-7, es considerado ahora también como una variable aleatoria, la *variable aleatoria básica* del problema. A ella se le asigna, de alguna manera, una distribución de probabilidad, aceptada por el investigador y aún propuesta subjetivamente por él; para distinguir a esta nueva distribución de la distribución muestral del estadígrafo que se considere en el muestreo posterior, la llamamos la *distribución "a priori"* de $E(Z)$. Los elementos que derivamos en los párrafos 1-7 para llegar a la curva de costo esperado, siguen formando parte del modelo, con la única variante que como $E(Z)$ es ahora una variable aleatoria, hablaremos de *curvas* (o *funciones*) *características condicionales de error*, de *curvas condicionales de costo* y de *curvas condicionales de costo esperado;* la palabra "condicional" indica que los valores dependen de la variable aleatoria básica del problema.

Resulta fácil advertir que el criterio que se desea proponer es el siguiente: ponderemos los valores condicionales esperados de las curvas de los Gráficos 5 ó 7 por las probabilidades "a priori" asignadas a $E(Z)$ y seleccionemos como mejor test aquél que tenga el menor *costo esperado* incondicional obtenido de esa ponderación.

10. Dejando de lado por un momento la discusión de si las probabilidades "a priori" son verdaderas probabilidades en todos los casos, obsérvese que hemos logrado una manera razonable de combinar la información que existe con respecto a $E(Z)$ con los datos que se conocen, también "a priori," sobre la naturaleza de un test dado. En particular, esta alternativa "subjetivista," se muestra más razonable que utilizar el juicio del investigador para seleccionar el mejor test mirando únicamente

[3] La interpretación subjetiva o personalista de las probabilidades tiene un moderno y excelente expositor en L. J. Savage (especialmente en *The Foundations of Statistics*, Wiley, 1954). El aclara la naturaleza de la posición (p. 3) diciendo que una probabilidad mide la confianza que un individuo dado tiene con respecto a la veracidad de una proposición específica.

En este orden de ideas, un "personalista" está dispuesto a aceptar como probabilidades juicios sobre la posibilidad de que llueva en 1966, de que Rómulo haya fundado a Roma, etc., todos los cuales son acontecimientos de naturaleza no repetitiva.

La idea central, está en que esas probabilidades personales satisfacen ciertos axiomas, que son los mismos sobre los que se ha demostrado que se puede construir todo el aparato del cálculo de probabilidades como lo conocemos.

No debe suponerse sin embargo que la interpretación que comentamos presupone que las probabilidades serán siempre *propuestas* personalmente por el investigador. Con la adopción del enfoque subjetivista debe asegurarse que el investigador *acepte* las probabilidades, pero éstas pueden originarse en información de fuente histórica sobre el problema, en su propio juicio personal o en otras formas distintas o combinadas.

a curvas características (condicionales) de error; estas últimas vienen expresadas en probabilidades y no puede esperarse en todos los casos que un investigador no estadístico domine su naturaleza. En cambio, puede confiarse que dicho investigador estará en buenas condiciones para evaluar, de la mejor forma posible, su propio conocimiento con respecto a la variable básica del problema.

11. Nuestra presentación ha sido hasta ahora muy informal y sólo ha procurado motivar los elementos de los distintos modelos considerados. Para proponer formalmente el problema de diseño de un test que nos ocupa, utilizaremos esencialmente la notación propuesta por Wilks, [4] tenemos la siguiente formulación:

Sea Z una variable aleatoria con espacio muestral R. Sea θ un punto del espacio paramétrico Ω, donde Ω es, por ejemplo, un conjunto finito, $\Omega = \{\theta_1, \theta_2, \ldots, \theta_s\}$. Sea $p(Z/\theta)$ la distribución de probabilidad de Z cuando θ es el verdadero valor del parámetro, de tal manera que existen s distribuciones discretas de probabilidad $p(Z/\theta_1), \ldots, p(Z/\theta_s)$, una sola de las cuales se supone cierta cuando se realiza una observación de Z.

Sea A el conjunto o espacio de las decisiones, $A = \{a_1, \ldots, a_t\}$ que se tomarán con respecto a θ sobre la base de la observación de Z; A será finito en todas nuestras aplicaciones. Por ejemplo en el problema del párrafo 7, $A = \{a_1, a_2\}$ donde a_1 es la decisión de que $p \leqslant 0.04$ y a_2 de que $p > 0.04$.

Una *función decisoria*, d, que pertenece a una clase de funciones decisorias D, asocia con cada valor de Z una decisión en A. Por ejemplo, en nuestro caso particular tenemos que

$$\begin{aligned} d(r) &= a_1 && \text{cuando } r \leq r_c \\ &= a_2 && \text{cuando } r > r_c . \end{aligned}$$

Asociada con la decisión a de A, existe una función de costo $C(\theta, a)$ que representa el costo de la decisión a cuando θ es el valor del parámetro. Esta función está definida en el espacio $A \times \Omega$ y asigna a cada punto de él un número real.

La *función de costo esperado* es entonces

$$r(\theta, d) = E\left[C(\theta, d(x))\right] = \sum_{x \in R} C(\theta, d(x))\ p(x/\theta), \tag{9}$$

también llamada frecuentemente la *función de riesgo*.

El *criterio minimax* aconseja seleccionar la decisión d^o que satisface:

$$r(\theta, d^o) = \min_{d \in D}\left\{\max_{\theta \in \Omega} r(\theta, d)\right\}. \tag{10}$$

El criterio subjetivista parte de una distribución "a priori" $q(\theta)$ y llega al *costo esperado incondicional*

$$\bar{r}(\theta, d) = \sum_{i=1}^{n} \sum_{x \in R} C(\theta_i, d(x))\ p(x/\theta_i)\ q(\theta_i) . \tag{11}$$

[4] *Mathematical Statistics*, por S. S. Wilks, Wiley 1962, p. 502.

La decisión d^o es propuesta como "mejor" si

$$d^o = \min_{d \in D} \bar{r}(\theta, d) . \tag{12}$$

5. El uso de probabilidades "a priori" y del teorema de Bayes en un problema de decisión

12. A pesar de la nueva interpretación de las probabilidades y del uso de la distribución "a priori," nuestro análisis ha seguido los mismos lineamientos tradicionales, vale decir, ha procurado diseñar un test de una hipótesis estadística seleccionando un punto crítico, etc. Sin embargo, la existencia y aceptación de la distribución "a priori" permite combinar los elementos del problema de una manera diferente. Vamos a presentar a continuación la manera "bayesiana" de analizar un problema de decisión, en el caso de que existan dos acciones alternativas posibles.

Supongamos que el investigador ha logrado proponer una distribución "a priori" para la variable aleatoria básica, a la que llamaremos μ y que se conozca la función condicional de costo, a la que designamos por simplicidad $C(\mu)$. Si no se dispone de otra información, es razonable proponer como criterio de decisión, que se tome aquella acción que minimiza el costo esperado $E'[C(\mu)]$ (o maximiza la utilidad esperada), donde hemos escrito E' para indicar que la esperanza es con respecto a la distribución "a priori" $q(\mu)$. Suponemos que μ sólo toma un número finito de valores, de manera que $q(\mu)$ es discontinua.

Supongamos a continuación que se ha decidido extraer una muestra al azar simple de tamaño n y en ella estimar a μ mediante el estadígrafo Z. Sea $p(Z/\mu)$ la distribución muestral de Z cuando μ es el valor verdadero de la variable básica del problema. Por el teorema de Bayes, tenemos que la distribución "a posteriori" de μ cuando se ha observado Z es

$$p(\mu/Z) = \frac{p(Z/\mu)\ q(\mu)}{\sum_{\mu} p(Z/\mu)\ q(\mu)} \tag{13}$$

Se propone como criterio de decisión en este caso en que existe simultáneamente información "a priori" e información muestral, seleccionar la acción que minimiza el costo esperado, $E''[C(\mu)]$, donde la esperanza E'' es con referencia a la distribución "a posteriori" $p(\mu/Z)$.

13. Para mostrar cómo se calculan estas probabilidades "a posteriori," consideremos el siguiente caso: la distribución "a priori" de π, la proporción favorable a un cierto suceso, es

π	0.2	0.3	0.4	0.5
$q(\pi)$	0.4	0.3	0.2	0.1

Supongamos que se ha seleccionado una muestra al azar simple de tamaño 20 y el estadígrafo r, la proporción muestral, ha resultado ser 0.2; supongamos además que se dan las condiciones para que r pueda considerarse distribuido como una binomial.

Utilizando la nomenclatura de distribuciones conjuntas, marginales y condicionales, tenemos:

Variable aleatoria básica π_i	Distribución "a priori" (Marginal) $q(\pi_i)$	Probabilidad muestral (Condicional) $p(r=0.2/\pi_i)$	Probabilidad conjunta $q(\pi_i)\ p(r/\pi_i)$	Distribución "a posteriori" (Marginal) $p(\pi_i/r=0.2)$
0.2	0.4	0.2182	0.08728	0.6520
0.3	0.3	0.1304	0.03912	0.2923
0.4	0.2	0.0350	0.00700	0.0523
0.5	0.1	0.0046	0.00046	0.0034
	1.0		0.13386	1.0000

Obsérvese cómo la distribución "a posteriori" refleja el hecho de que el resultado muestral ($r = 0.2$) ha contribuido a acrescentar la confianza que se tenía en $\pi = 0.2$.

14. Con el propósito de hacer el problema más realista, puede suponerse que la variable aleatoria básica es continua. En este caso la densidad "a posteriori" es

$$f(\mu/Z) = \frac{p(Z/\mu)\ f(\mu)}{\int p(Z/\mu)\ (f\mu)\ d\mu} \tag{14}$$

En el ejemplo del párrafo 13, supongamos que se acepta que π es uniforme en el intervalo [0.1]. Entonces

$$f(\pi/r) = \frac{\binom{n}{r}\pi^r(1-\pi)^{n-r}}{\binom{n}{r}\int_0^1 \pi^r\ (1-\pi)^{n-r}\ d\pi} \tag{15}$$

Utilizando la solución de la "función beta"

$$\int_0^1 \pi^r(1-\pi)^{n-r}\ d\pi = \frac{r!\,(n-r)!}{(n+1)!} \tag{16}$$

y la identidad

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} \tag{17}$$

resulta que la densidad "a posteriori" de π dado r es:

$$f(\pi/r) = (n+1)\binom{n}{r}\pi^r(1-\pi)^{n-r} \qquad 0 < \pi < 1 \qquad r = 1,2,\ldots,n. \tag{18}$$

15. Un caso interesante ocurre cuando μ es "a priori" normal con media $E'(\mu)$ y desviación estándar $\sigma'(\mu)$ y la distribución muestral del estadígrafo Z es normal $(\mu, \sigma(Z))$ cuando μ es el verdadero valor esperado. En este caso, si $\sigma(Z)$ es conocida, resulta que la distribución "a posteriori" de μ dado Z también es normal con esperanza $E''(\mu)$ y desviación estándar $\sigma''(\mu)$, donde

$$E''(\mu) = \frac{\dfrac{E'(\mu)}{\sigma'^2(\mu)} + \dfrac{Z}{\sigma^2(Z)}}{\dfrac{1}{\sigma'^2(\mu)} + \dfrac{1}{\sigma^2(Z)}} \qquad \sigma''^2(\mu) = \frac{1}{\dfrac{1}{\sigma'^2(\mu)} + \dfrac{1}{\sigma^2(Z)}} \tag{19}$$

Si la variable aleatoria básica, μ, es la esperanza de una distribución normal y el estadígrafo muestral la media $\bar{x}$ (se muestrea al azar simple de una población con varianza σ^2 conocida), entonces la distribución "a posteriori" es normal y los parámetros (19) adoptan la forma particular

$$E''(\mu) = \frac{\dfrac{E'(\mu)}{\sigma'^2(\mu)} + \dfrac{n\bar{x}}{\sigma^2}}{\dfrac{1}{\sigma'^2(\mu)} + \dfrac{n}{\sigma^2}} \qquad \sigma''^2(\mu) = \frac{1}{\dfrac{1}{\sigma'^2(\mu)} + \dfrac{n}{\sigma^2}}. \tag{20}$$

16. En este párrafo consideramos dos aspectos estadísticos de mucha importancia.

En primer lugar, debemos establecer propiedades de los estadígrafos muestrales que permiten seleccionar el más apropiado para utilizar en (13) o (14). Resulta que el criterio deseado es el de *suficiencia* y esto es fácil de aceptar si advertimos que en la formulación "bayesiana" se define a un estadígrafo muestral suficiente como aquél que permite obtener la misma distribución "a posteriori" que una descripción completa de la muestra, y esto para cualquier distribución "a priori" que se proponga. Sin embargo puede probarse que esta definición es equivalente a la formulación más tradicional en término de la factorabilidad de la verosimilitud conjunta de la muestra.

La segunda cuestión se relaciona con la selección de la forma de la distribución "a priori." De la discusión precedente se deduce que no existe ninguna restricción esencial sobre la forma de dicha distribución; sin embargo, si la expresión analítica de la distribución "a priori" se elige convenientemente, el problema matemático se simplifica notablemente, y en particular, la distribución "a posteriori" entonces pertenece a una familia conocida de distribuciones de probabilidad.

Más concretamente, puede probarse que para muchos casos corrientes de distribuciones muestrales, se puede obtener una familia de *distribuciones "a priori" conjugadas,* intercambiando el papel de los parámetros y las variables en la expresión de la distribución muestral, y entonces la distribución "a posteriori" pertenecerá a la misma familia que la distribución "a priori." Ya vimos que si la distribución muestral y la distribución "a priori" son normales, la "a posteriori" también será normal. Igualmente, si la distribución muestral es binomial y la distribución "a priori" es "beta" o si la distribución muestral es Poisson y la "a priori" es "gamma," las distribuciones "a posteriori" serán de la familia "beta" o "gamma" respectivamente.

17. Con la discusión hasta el párrafo 16, hemos reunido un conjunto de elementos interesantísimos para el análisis estadístico de un problema de decisión, pues disponemos de una estructura económica (función de costo), de una distribución "a priori" de la variable aleatoria básica del problema y de un estadígrafo muestral. Además, como se dijo en la última parte del párrafo 12, hemos aceptado una manera precisa de combinar esos elementos mediante una regla decisoria basada en el costo esperado incondicional "a priori" o "a posteriori," según exista o no muestreo.

Para terminar esta presentación, nos referiremos a dos problemas muy importantes: (1) la evaluación del "riesgo terminal," que existe al adoptar una decisión óptima a la luz de la evidencia disponible, antes o después del muestreo; (2) la selección del tamaño muestral.

No es de extrañar que la información económica se use en este modelo de una manera original para lograr una solución satisfactoria para esos problemas. En particular, se ha desarrollado el concepto de *valor esperado de la información perfecta* (o *costo de la incertidumbre*), que es una medida del riesgo asociado con una decisión óptima que se toma en un momento dado del análisis, evaluada en las unidades propias del problema.

Para ilustrar cómo operan estas ideas, consideramos el caso en que se cumplen las siguientes condiciones: (1) la estructura probabilística es como en el párrafo 15; (2) la función de costo es lineal como en el Gráfico 4 (a) y las dos ramas tienen pendiente de igual valor absoluto k, partiendo del valor de indiferencia μ_o. Se prueba que el valor esperado de la información perfecta (VEIP) es

$$VEIP = k \cdot \sigma(\mu) \cdot G(D) \tag{21}$$

donde $\sigma(\mu)$ es la desviación estándar de la distribución de μ ("a priori" o "a posteriori"),

$$D = \frac{|\mu_o - E(\mu)|}{\sigma(\mu)} \tag{22}$$

($E(\mu)$ corresponde a las distribuciones "a priori" o "a posteriori") y la función G definida por

$$G(a) = \int_a^{\infty} (t - a) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \tag{23}$$

se encuentra extensamente tabulada. [5]

Tanto cuando se trabaja con la distribución "a priori" como con la distribución "a posteriori," el VEIP mide el riesgo asociado con la decisión óptima en el momento respectivo y en el caso en que las unidades económicas del problema fueran dinero, entonces de la comparación entre el costo del muestreo y el respectivo VEIP surgirá claramente la ventaja o desventaja de obtener *nueva* información muestral.

En realidad, se han desarrollado también métodos para determinar el tamaño de muestra óptimo con una situación de decisión dada. Sin embargo, no entraremos en ese detalle en este trabajo. [6]

6. Bibliografía y comentarios

18. Las ideas centrales de los párrafos 1 a 4 aparecen prácticamente en cualquier libro de estadística basada en el cálculo de probabilidades. Para quien se interese en la trayectoria histórica de las ideas, recomendamos dos obras del más importante precursor de la estadística moderna:

> *Métodos Estadísticos para Investigadores,* por R. A. Fisher, traducción de J. R. Magán y J. J. Ruiz Rubio, Aguilar, Madrid, 1949. Especialmente el párrafo 5 del capítulo 1, p. 19;
>
> *Statistical Methods and Scientific Inference,* por R. A. Fisher, 2a. edición, Hafner, Nueva York, 1959.

La nomenclatura moderna de los intervalos de confianza y de los test de hipótesis estadísticas fue introducida por J. Neyman y E. S. Pearson. Sus ideas pueden leerse en el libro:

> *First Course in Probability and Statistics,* por J. Neyman, Holt, Nueva York, 1950.

Se atribuye al grupo que trabajaba (en 1925) en técnicas de inspección por muestras en la empresa Bell de Estados Unidos, y en particular a Dodge y Romig, las primeras aplicaciones de las curvas características operantes. Ambos autores publicaron posteriormente una monografía:

> *Sampling Inspection Tables,* por H. F Dodge y H. G. Romig, 2a. edición, Wiley, Nueva York, 1959.

[5] Por ejemplo en los apéndices de los libros de Schlaifer o el de Raiffa y Schlaifer. Para referencias completas consúltese la Parte VI.

[6] Véase Schlaifer, 1959, Cap. 33-38, o Raiffa y Schlaifer, Parte II, *op. cit.* en la Sección VI.

Las ideas básicas de la teoría estadística de las decisiones fueron discutidas inicialmente por Abraham Wald, sobre todo en su libro:

Statistical Decision Functions, por A. Wald, Wiley, Nueva York, 1950.

Las presentaciones de Wald presuponen una gran madurez matemática por parte del lector y sólo adquirieron difusión en el campo académico; su incorporación a la estadística práctica fue muy lenta. Sin embargo en el año 1959 aparecieron dos textos:

Probability and Statistics for Business Decisions, por R. Schlaifer, McGraw-Hill, 1959;

Elementary Decision Theory, por H. Chernoff y L. E. Moses, Wiley, Nueva York, 1957,[7]

que procuraban hacer accesible, a un mayor número de lectores, los principios de la teoría estadística de las decisiones. Ambos tienen niveles matemáticos mucho menos exigentes que las obras de Wald y el libro de Schlaifer no presupone el conocimiento del cálculo infinitesimal.

El libro de Chernoff y Moses presenta un enfoque de las decisiones desde el punto de vista de la teoría de los juegos y no está explícitamente orientado hacia la forma "bayesiana" de decisión que hemos comentado en la última parte de nuestro trabajo. El libro de Schlaifer ha sido preparado para un curso que se ofrece a los alumnos de la Escuela de Administración de Negocios de la Universidad de Harvard y es decididamente "bayesiano" y por supuesto subjetivista.

Es nuestra opinión que una persona que se interese seriamente por la investigación empírica en la economía y los negocios (por lo menos) debería adquirir una apreciación de lo que ofrecen ambos textos. En particular, para quien tiene un interés inmediato en las aplicaciones, sugeriríamos el libro de Schlaifer.

En este texto de 732 páginas, el autor muchas veces se aparta radicalmente de la manera tradicional de encarar los problemas de inferencia probabilística y la nomenclatura y la terminología son entonces diferentes de las de otros textos. Además, la comparación entre los métodos propuestos y la técnicas tradicionales (él las llama "clásicas"), recién se hace en los últimos cuatro capítulos y el libro tiene 421.

Para corregir estos inconvenientes, el propio autor ha presentado el texto (de 382 páginas) :

Introduction to Statistics for Business Decisions, por R. Schlaifer, Mc-Graw-Hill, Nueva York, 1961.

[7] Traducido también al español por L. Dignowity, Editorial Continental, México, 1962, bajo el título *Teoría y Cálculo Elemental de las Decisiones*. Sin embargo, la traducción es excesivamente descuidada como se comprobará con sólo leer el índice, p. 11 y 19.

que contiene nuevamente sus ideas básicas pero comparándolas en todo momento con las "clásicas" o tradicionales. Quizás el lector interesado prefiera comenzar con este texto y referirse al otro de Schlaifer para ampliar el número de técnicas y métodos.

Ambos libros se caracterizan por la claridad con que se presentan los conceptos, por la calidad y minuciosidad de los ejemplos y ejercicios propuestos y, como dijimos, por el escaso uso explícito que se hace de técnicas matemáticas avanzadas. Las pruebas de la mayoría de las proposiciones de estos textos aparecen en el libro

Applied Statistical Decision Theory, por H. Raiffa y R. Schlaifer, División de Investigaciones, Escuela de Administración de Negocios, Universidad de Harvard, Boston, 1964.

Esta obra tiene un papel central en la formulación "bayesiana" que nos ha ocupado. Además de las pruebas formales, contiene presentaciones muy completas de las proposiciones básicas y una exposición detallada de los elementos de la teoría de las probabilidades (Parte III, *Distribution Theory*) necesaria para el análisis. Sin lugar a dudas un estadístico matemático que se interese en el tema deberá consultar extensamente este importante trabajo.

Además del libro de Raiffa y Schlaifer se ha publicado un texto complementario, que trata el problema particular que se plantea cuando se trabaja con la distribución normal y la desviación estándar es desconocida. Esta nueva monografía,

Tables for Normal Sampling with Unknown Variance, por J. Bracken y A. Schleifer, Jr., División de Investigaciones, Escuela de Administración de Negocios, Universidad de Harvard, Boston, 1964,

presenta una breve discusión de ese problema y extensas tablas que se utilizan en su solución.

Esta es la bibliografía básica sobre el tema de la "Estadística Bayesiana." Varios libros recientes sobre estadística matemática y aplicada en general, han comenzado a incluir entre sus temas un capítulo sobre la teoría estadística de las decisiones. Este movimiento es bastante general y quizás los dos mejores ejemplos lo den, a dos niveles diferentes,

Statistical Theory, por B. W. Lindgren, Macmillan, Nueva York, 1962 (Capítulo 5) ;

Mathematical Statistics, por S. Wilks, Wiley, Nueva York, 1962 (Capítulo 16).

En otro orden de ideas, han aparecido numerosos artículos y comentarios sobre el enfoque "bayesiano," que sin lugar a dudas puede ayudar al lector interesado a lograr una buena perspectiva, sobre todo con relación a otras posiciones. En particular, merecen citarse entre los comentarios:

"Review of *Probability and Statistics for Business Decisions*, by R. Schlaiter," por F. J. Anscombe, *Journal of the American Statistical Association*, Vol. 54, No. 288, diciembre 1959, p. 813;

"Review of *Applied Statistical Decision Theory*, by H. Raiffa and R. Schlaifer," por H. V. Roberts, *Ibid.*, Vol. 57, No. 297, marzo 1962, p. 199;
"Review of *Probability and Statistics for Business Decisions*, by R. Schlaifer," por S. Vajda, *Journal of the Royal Statistical Society*, Serie A. Vol. 122, Parte 4, 1959, p. 549.

Este último comentarista no coincide con los otros dos en su apreciación de la labor de los "bayesianos."

En cuanto a artículos mencionaremos solamente tres que nos parece pueden ilustrar, con una lectura breve, las ideas centrales de los modelos que se proponen:

"The New Business Statistics," por H. V. Roberts, *The Journal of Business*, Vol. 32, enero 1960, p. 21;
"The Bayesian Approach to Statistical Decision: An Exposition," por J. Hirschleifer, *Ibid.*, Vol. 34, octubre 1961, p. 471;
"Bayesian Statistics," por F. J. Anscombe, *The American Statistician*, Vol. 15, febrero, 1961, p. 21.

Para cerrar esta reseña bibliográfica, mencionaremos dos trabajos en los que se aplican técnicas "bayesianas" a problemas concretos:

Decisions Under Uncertainty: Drilling Decisions by Oil and Gas Operators, C. Jackson Grayson, Jr., División de Investigaciones, Escuela de Administración de los Negocios, Universidad de Harvard, Boston, 1960,

es una tesis doctoral que presenta el caso de estudiar decisiones propias de la búsqueda de petróleo; este trabajo incluye una estimación de las funciones personales de utilidad de los empresarios respectivos, las que son luego utilizadas en los cálculos estadísticos. El segundo trabajo que mencionaremos es:

Inference and Disputed Authorship, por F. Mosteller y D. L. Wallace, Addison-Wesley, Massachusetts, 1964.

Esta importante monografía analiza la paternidad de los escritos "federalistas" e incluye la derivación de algunos resultados teóricos propios de la estadística "bayesiana."

* * *

Regression models for seasonal analysis

Raúl Pedro Mentz and Víctor Jorge Elías

Regression models for seasonal analysis

Raúl Pedro Mentz and Victor Jorge Elias

In this work we consider a regression approach to estimate simultaneously the trend and seasonal components of a time series, on the basis of intra-annual observations. We analyze several proposals presented in the literature and prove their equivalence; as a consequence the use of one of them is recommended on grounds that its interpretation is simpler and more direct. The analysis is later related to the traditional method of "difference from moving average," which we propose to interpret as a nonparametric variant of the regression scheme. We also consider from a practical point of view the relation between the regression approach with a polynomial trend component, and the method of difference from moving average, and related proposals of empirical extraction.

The regression approach to deal with seasonality is frequent in econometrics, when the main purpose is to estimate the parameters of economic relations on the basis of monthly, quarterly, or similar observations. In that case the approach to the estimation problems is usually conditioned by the desire to obtain "good" estimators of the main parameters, namely those of the economic relation. Different demands are bound to arise when the purpose is to estimate trend (that in our presentation includes all "smooth" movements in the series, relative to the period of observation, and will be taken to include all cyclical components) and seasonality, for multiple uses not necessarily related to a definite economic model. For example, it may then be desired to obtain separate estimates of the trend and seasonal components, and that each one be easy to interpret. This may be important in the preparation of economic indicators responding to an additive or multiplicative model, or to compare estimates coming from different methods.

In this work we study the regression models from the second point of view only, and try to investigate theoretically and practically their performance, "per se" and in relation to well-known approaches of empirical origin. It should be pointed out that it has been questioned whether the regression approach will be effective for the purposes we have in mind here; see, for example, Shiskin, Young and Musgrave (1967). However, works such as that of Jorgenson (1964) tend to increase the confidence of statisticians in regression models, since it is proved there that under suitable hypotheses their statistical treatment can be made to satisfy important optimality criteria.

1. The regression model

Given an observed time series $[y_t: t = 1,2,...,T]$ with monthly observations, a proposal to estimate simultaneously the trend and seasonal components is expressed by the multiple linear regression model

*University of Tucumán, Argentina.

$$y_t = \alpha_0 + \alpha_1 t + ... + \alpha_m t^m + \beta_1 D_{1t} + \beta_2 D_{2t} + ... + \beta_{12} D_{12,t} + u_t,$$
$$t = 1,2, ..., T, \qquad (1)$$

where the polynomial in t of degree m ($m \geq 1$) represents the trend component, $\alpha_0, ..., \alpha_m, \beta_1, ..., \beta_{12}$ are the unknown m + 13 parameters of interest, $[u_t]$ is a set of random variables satisfying for all subindices $\zeta u_t = O, \zeta u_t^2 = \sigma^2, \zeta u_t u_s = O$, when $t \neq s$. D_{st} for s = 1, ..., 12 is a seasonal dummy variable that takes the value 1 when t corresponds to month s, and 0 otherwise. We regard m as known. Model (1) is a particular case of the general linear models considered by Jorgenson (1964) for this purpose, and one of the most widely used.

To simplify matters slightly we take T = 12N. From a theoretical and computational standpoint the general case merely makes the form of some expressions more complex. We consider monthly data for the sake of definiteness, but clearly the same approach is valid for any other similar subdivision of the year.

Expression (1) is interpreted as the desire to fit to the observations 13 "parallel" polynomials in t, each of degree m, one being the average (whose constant term is α_0) and the remaining 12, one for each set of N observations for month s, having constants $\alpha_0 + \beta_s$, s = 1, ..., 12. We note that this is a case of constant seasonality, characterized by the fact that β_s does not depend on t.

In (1) it is easily verified that

$$I = D_{1t} + D_{2t} + ... + D_{12,t}, \qquad t = 1,2, ..., T, \qquad (2)$$

and that the matrix of regressors has rank m + 12 instead of the full rank m + 13. Several proposals to deal with this problem will now be studied. Noted that a solution of the normal equations using some definition of a generalized inverse matrix is also feasible.

1.1 *The solution as in the analysis of variance.* Jorgenson (1964) proposed a solution parallel to that frequently used in the analysis of variance, namely to impose the additional restriction

$$\sum_{s=1}^{12} \beta_s = 0. \qquad (3)$$

Then (1) can be written as

$$y_t = A_0 + A_1 t + ... + A_m t^m + B_1 D_{1t}^* + ... + B_{11} D_{11,t}^* + u_t^*, \quad t = 1,2, ..., T, \quad (4)$$

where D_{st}^* equals 1 when period t corresponds to month s, −1 when it corresponds to s = 12 (December), and 0 otherwise. The summands $B_s D_{st}^*$ are interpreted as deviations from the trend, and Jorgenson (p. 694) notes that under (3) "the annual sums of seasonally adjusted and seasonally unadjusted data must be the same."

1.2. *Omitting the constant.* Sometimes it is proposed to eliminate the constant in (1) and use the model

$$y_t = A_1^* t + ... + A_m^* t^m + B_1^* D_{1t} + ... + B_{12}^* D_{12,t} + v_t, \quad t = 1,2, ..., T, \quad (5)$$

See, for example, Johnston (1972).

This model can be interpreted as proposing 12 regressions, where for month s the equation has regressors t, ..., t^m and constant B_s^*. It is not possible to estimate

the constant of the average polynomial trend, and one can take it arbitrarily equal to 0. Another possibility is to declare its value as $A_0^* = y_1$, or $A_0^* = (1/12)\,\Sigma_{s=1}^{12}\ B_s^*$; any of these would then replace (3) to estimate the 13 + m parameters in (1).

1.3. *Omitting one month*. Instead of (5) sometimes the parameters of

$$y_t = A_0^{**} + A_1^{**}\,t + \ldots + A_m^{**}t^m + B_1^{**}D_{1t} + \ldots + B_{11}^{**}D_{11,t} + w_t, \quad t = 1,2,\ldots,T, \tag{6}$$

are estimated, where for the sake of definiteness we omitted the twelfth month. For example (6) is preferred to (5) to be used with a computer program that forces the use of a constant term. See for example Rao and Miller [(1971), Section 4.7].

2. Comparison of the alternatives

2.1. *Algebraic relations*. Let us consider (4) and (5). In case the parameters are estimated by ordinary least squares (OLS), Goldberger [(1964), Chapter 5] proposed a method to compare the estimators resulting from the normal equations. The first 12 rows of the matrix of regressors corresponding to (4), set in the order $t, \ldots, t^m, 1, D_{1t}^*, \ldots, D_{11,t}^*$, form the matrix

$$(\underline{Q}\,|\,\underline{J}) = \left(\begin{array}{cccc|cccc} 1 & 1 & \ldots & 1 & 1 & 1 & 0 & \ldots & 0 \\ 2 & 2^2 & \ldots & 2m & 1 & 0 & 1 & \ldots & 0 \\ 3 & 3^2 & \ldots & 3m & 1 & 0 & 0 & \ldots & 0 \\ . & . & & . & . & . & . & & . \\ . & . & & . & . & . & . & & . \\ . & . & & . & . & . & . & & . \\ 11 & 11^2 & \ldots & 11m & 1 & 0 & 0 & \ldots & 1 \\ 12 & 12^2 & \ldots & 12m & 1 & -1 & -1 & \ldots & -1 \end{array}\right), \tag{7}$$

while for (5) the corresponding matrix is $(\underline{Q}\,|\,\underline{I})$. Hence

$$(\underline{Q}\,|\,\underline{J}) = (\underline{Q}\,|\,\underline{I}) \left(\begin{array}{c|c} \underline{I} & \underline{Q} \\ \hline \underline{Q} & \underline{J} \end{array}\right) \tag{8}$$

$$\left(\begin{array}{c} \hat{A}_1^* \\ \vdots \\ \hat{A}_m^* \\ \hat{B}_1^* \\ \vdots \\ \hat{B}_{12}^* \end{array}\right) = \left(\begin{array}{c|c} \underline{I} & \underline{Q} \\ \hline \underline{Q} & \underline{J} \end{array}\right) \left(\begin{array}{c} \hat{A}_1 \\ \vdots \\ \hat{A}_m \\ \hat{A}_0 \\ \hat{B}_1 \\ \vdots \\ \hat{B}_{11} \end{array}\right) = \left(\begin{array}{c} \left(\begin{array}{c} \hat{A}_1 \\ \vdots \\ \hat{A}_m \end{array}\right) \\ \hline \underline{J} \left(\begin{array}{c} \hat{A}_0 \\ \hat{B}_1 \\ \vdots \\ \hat{B}_{11} \end{array}\right) \end{array}\right) \tag{9}$$

In terms of components (9) says that the OLS estimators of the coefficients of $t, \ldots, t^m$ are identical in (4) and (5), while the other estimators are related by

$$\begin{aligned} \hat{B}_s^* &= \hat{A}_0 + \hat{B}_s, \quad s = 1,2,\ldots,11, \\ \hat{B}_{12} &= \hat{A}_0 - \hat{B}_1 - \ldots - \hat{B}_{11} = \hat{A}_0 + \hat{B}_{12}, \end{aligned} \tag{10}$$

the last line because the sum of the $\hat{B}_s$ satisfies (3).

Summing over the equations in (10) we find $\hat{A}_0$, so that

$$\begin{aligned} \hat{A}_0 &= (1/12)\,(\hat{B}_1^* + \ldots + \hat{B}_{12}^*), \\ \hat{B}_1 &= \hat{B}_1^* - \hat{A}_0 = (1/12)\,(11\hat{B}_1^* - \hat{B}_2^* - \ldots - \hat{B}_{12}^*), \\ &\vdots \qquad\quad \vdots \qquad\qquad \vdots \\ \hat{B}_{11} &= \hat{B}_{11}^* - \hat{A}_0 = (1/12)\,(-\hat{B}_1^* - \hat{B}_2^* - \ldots - \hat{B}_{10}^* + 11\hat{B}_{11}^* - \hat{B}_{12}^*). \end{aligned} \tag{11}$$

This shows that

$$\underline{\Gamma}^1 = (1/12)\left(\begin{array}{ccccc} 1 & 1 & 1 & \ldots & 1 \\ 11 & -1 & -1 & \ldots & -1 \\ -1 & 11 & -1 & \ldots & -1 \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ -1 & -1 & -1 & \ldots & -1 \end{array}\right) \tag{12}$$

The first equality in (11) says that if to estimate the constant in the average polynomial trend in (5), we use the average of the estimated coefficients of the seasonal variables, as was suggested in 1.2, we show the equivalence with model (4) for the trend component. The equivalence becomes complete if we further adjust the seasonal estimators according to (11). Another way to express the relation between the trend components, is that (4) derived under the condition (3), is equivalent to (5) with

$$\hat{A}_0 = (1/12)\sum_{s=1}^{12} \hat{B}_s^*.$$

Instead of deriving the relation among the OLS estimators, an interesting and in this case equivalent method, is to relate the corresponding parameters. We note that (2) holds and that

$$D_{st}^* = D_{st} - D_{12,t}, \tag{13}$$

for all values of s and t. Substituting in (4) we obtain

$$\begin{aligned} y_t &= A_0(D_{1t} + \ldots + D_{12,t}) + A_1\,t + \ldots + A_m t^m + B_1(D_{1t} - D_{12,t}) + \\ &\quad + \ldots + B_{11}(D_{11,t} - D_{12,t}) + u_t^* \\ &= A_1\,t + \ldots + A_m t^m + (B_1 + A_0)D_{1t} + \ldots + (B_{11} + A_0)D_{11,t} + \\ &\quad + (A_0 - B_1 - \ldots - B_{11})D_{12,t} + u_t^*. \end{aligned} \tag{14}$$

If we set this model equal to (5), the relations in (10) are read for the corresponding parameters. Reciprocally, since $D_{12,t} = (1/12)\,(1 - D_{1t}^* - \ldots - D_{11,t}^*)$, we have that

$$D_{st} = D^*_{st} + D_{12,t} = D^*_{st} + (1/12)\,(1 - D^*_{1t} - \ldots - D^*_{11,t}); \qquad (15)$$

substituting in (5) and setting the resulting expression equal to (4), we read the relations in (11) among the corresponding parameters.

Note that finding (15) is equivalent to inverting $\underset{\sim}{J}$, and that finding (13) is equivalent to inverting $\underset{\sim}{J}^{-1}$

An interesting by-product of this analysis is that the matrix $\underset{\sim}{Q}$ of regressors which are non-zero powers of t, does not enter into the computations, and hence the results hold for any set of m linearly independent regressors, which are linearly independent of the constant and the included seasonal variables. Clearly this observation is valid for any one of the proposals studied in this section.

To compare (5) and (6) we proceed in the same way. Substitution of (2) in (6) gives

$$y_t = A^{**}_1 t + \ldots + A^{**}_m t^m + (B^{**}_1 + A^{**}_0)D_{1t} + \ldots + (B^{**}_{11} + A^{**}_{11})D_{11,t} + \\ + A^{**}_0 D_{12,t} + w_t, \qquad (16)$$

which is of the form (5). On the contrary, if we deduce $D_{12,t}$ from (2) and replace the resulting value in (5), we obtain

$$y_t = B^*_{12} + A^*_1 t + \ldots + A^*_m t^m + (B^*_1 - B^*_{12})D_{1t} + \ldots + \\ + (B^*_{11} - B^*_{12})\,D_{11,t} + v_t, \qquad (17)$$

which is of the form (6). In terms of matrices we proved that

$$\begin{pmatrix} 1 & 1 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 0 & 1 & \ldots & 0 \\ . & . & . & . & & . \\ . & . & . & . & & . \\ . & . & . & . & & . \\ 1 & 0 & 0 & 0 & \ldots & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & 0 & \ldots & 0 & -1 \\ 0 & 1 & 0 & \ldots & 0 & -1 \\ . & . & . & & . & . \\ . & . & . & & . & . \\ . & . & . & & . & . \\ 0 & 0 & 0 & \ldots & 1 & -1 \end{pmatrix} \qquad (18)$$

The relations between the OLS estimators in (5) and (6) can then be written as follows, where we repeat a portion of (10) and another of (11) to facilitate the comparisons:

$$\hat{B}^*_s = \hat{A}^{**}_0 + \hat{B}^{**}_s = \hat{A}_0 + \hat{B}_s, \quad s = 1, \ldots, 11, \qquad (19)$$
$$\hat{B}^*_{12} = \hat{A}^{**}_0 = \hat{A}_0 + \hat{B}_{12},$$

and

$$\hat{A}^{**}_0 = \hat{B}^*_{12} = \hat{A}_0 + \hat{B}_{12},$$
$$\hat{B}^{**}_s = \hat{B}^*_s - \hat{B}^*_{12} = \hat{B}_s - \hat{B}_{12}, \quad s = 1, \ldots, 11. \qquad (20)$$

In (20) we interpret that in incorporating A^{**}_0 in (6), we can only estimate the differences between each specific monthly effect and that of the omitted month, which serves as a base. We see that the relation still holds when we incorporate a constant and impose (3). Similar interpretations can be made for (19).

2.2. *Statistical properties*. The analysis of the proposals in paragraphs 1.1, 1.2 and 1.3 shows that the parameters in models (4), (5) and (6) stand among

themselves in one-to-one linear relations like (10), (11), (19) or (20). Since the three models have full rank (equal to m + 12) the OLS estimators in each case are the unique, best linear unbiased estimators of the corresponding parameters, by the Gauss-Markoy theorem, and the estimators of any set of parameters, obtained from those of any of the other models, have the same optimal properties [of. Scheffé (1959), page 14, Theorem 2].

From the point of view of prediction of values of y, relations like (14), (16) or (17) show that any of the models, with parameters estimated by OLS, will lead to identical predictions.

Finally, if we assume normal distributions for the error terms, the OLS estimators coincide with those by maximum likelihood, and are unique, best unbiased for each model. By the invariance property of maximum likelihood estimation, the properties extend to the estimators of any of the sets of parameters, no matter what model is used to do the initial calculations.

The inferencial procedures for the normal case are given, for example, in Jorgenson (1964).

2.3. *Connection with the analysis of variance*. A situation similar to that studied here, arises in the analysis of variance. Note that (1) is a model of "analysis of covariance."

A classical approach to the analysis of variance is to impose to models like (1) the restriction (3); this is for example the form most frequently used in Scheffé (1959). However, to exploit the close relation among linear models in the analysis of variance with those of regression (which are usually of full rank), the approach as in (5) has been often used [see for example Rao (1965), Section 4.d], and also the approach as in (6) [see for example Mendenhall (1968)].

Graybill [(1961), Section 11.2.3] proves explicitly that every linear model can be reparametrized so that the resulting model has full rank, and that for any such reparametrization the unbiased estimators of the various linear combinations of the parameters are identical. From this point of view we found the explicit relation among several full-rank reparametrizations of (1).

2.4 *Conclusion*. The foregoing discussion shows that the choice of model to perform the estimations is statistically irrelevant. Model (4) provides a complete estimation of trend without further elaborations. Moreover, its seasonal estimates are directly comparable to those obtained by methods such as that of "difference from moving average," to be described below, when in the latter the sum of the seasonal components is set equal to zero. Hence there is a clear advantage in using this model.

Note that in (6), when one of the months is omitted, $\hat{A}^{**}_0$ is not the estimated trend constant. Nor is recommended to use the $\hat{B}^*_s$ of model (5) directly as seasonal components, since $\Sigma_{s=1}^{12}\hat{B}^* = 12\hat{A}_0$, which in general will not equal zero, and will make different the sums of the given observations and of them after seasonal adjustment.

3. Estimation in two steps and the method of difference from moving average

The procedures studied in Sections 1 and 2 will now be considered as operating in one step, against alternative procedures in two steps. In several

respects the contents of this section specializes the general treatment in Jorgenson (1964) to the case of polynomial trends.

3.1. *Least-squares in two steps.* To estimate the 13 + m parameters in (1) we can regress

$$y_t = A'_0 + A'_1 t + \ldots + A'_m t^m + e_t, \quad t = 1,2, \ldots, T, \tag{21}$$

and then fit to the residuals of the OLS estimation, the seasonal parameters

$$y_t - \hat{A}'_0 - \hat{A}'_1 t - \ldots - \hat{A}'_m t^m = B'_1 D_{1t} + \ldots + B'_{12} D_{12,t} + e'_t, \quad t = 1,2, \ldots, T. \tag{22}$$

In (22) the regressors are orthogonal ($\sum_{t=1}^{T} D_{it} D_{jt} = N \delta_{ij}$, where δ_{ij} is Kronecker's delta function) and hence each B'_s is estimated by OLS as if we had posed twelve regressions through the origin

$$y_t - \hat{A}'_0 - \hat{A}'_1 t - \ldots - \hat{A}'_m t^m = B'_s D_{st} + e''_{st}, \quad s = 1,2, \ldots, 12, \tag{23}$$

each one fitting the observations for the corresponding month. Then

$$\hat{B}'_s = \frac{1}{N} \Sigma_{(s)} (y_t - A'_0 - A'_1 t - \ldots - A'_m t^m), \quad s = 1,2, \ldots, 12, \tag{24}$$

the simple arithmetic average of the residuals for the N observations corresponding to month s.

Since in (4) the D^*_{st} are not orthogonal to the remaining regressors, (24) will in general differ from the OLS estimator in (4). Note that $\Sigma_{s=1}^{12} \hat{B}'_s = 0$ when $A'_0 \neq 0$, since the sum of (24) over t = 1,2, ..., T is the normal equation obtained by taking the derivative of the least squares criterion $\Sigma_{t=1}^{T} (y_t - A'_0 - A'_1 t - \ldots - A'_m t^m - B'_s D_{st})^2$ with respect to A'_0 and the resulting equation is set equal to zero in the least-squares procedure.

Under this approach the use of (1) as a simultaneous regression problem for trend and seasonality, is in some respect only apparent. However Jorgenson (op. cit., Section 3) showed that if the trend removal operation in the first step is totally effective, and accomplised by a matrix procedure ("linear filter"), the corresponding procedure in two steps, similar to the one we described above, is equivalent to that of paragraph 1.1, and hence to the other proposals in one step that we studied.

Hence, if model (1) holds, if we want a simultaneous formulation for trend and seasonality, and if the statistical properties described in 2.2 are judged relevant, the procedure in one step considered in 1.1 (or its equivalents) should be preferred to the proposal in two steps of this paragraph.

3.2. *The method of difference from moving average.* The approach in the preceeding paragraph can be used to clarify and reinterpret the traditional method of difference from moving average. In this scheme we start with the model

$$y_t = T_t + \beta'_1 D_{1t} + \ldots + \beta'_{12} D_{12,t} + z_t, \quad t = 1,2, \ldots, T, \tag{25}$$

where T_t is the series trend component, not necessarily parametrized as in (1). It is proposed to estimate T_t by smoothing the available observations using moving

averages (which are also non-parametric procedures), usually a twelve-month centered moving average, which is defined by

$$\hat{T}_t = (1/24) (y_{t-6} + 2 \sum_{j=-5}^{5} y_{t+j} + y_{t+6}), \quad t = 7,8, \ldots, T-6. \tag{26}$$

Then the β'_s are estimated through a formula equivalent to (24), namely

$$\hat{\beta}'_s = \frac{1}{N-1} \Sigma'_{(s)} (y_t - \hat{T}_t), \quad s = 1,2, \ldots, 12, \tag{27}$$

where we omit from the sums the first observation for months corresponding to t = 1, ..., 6, and the last one for t = T–5, ..., T. (In case the period of observation consists of complete years, these are simply the first and last six months, respectively.)

Note that (3) will not hold in general for the $\hat{\beta}'_s$, so that an additional adjustment is usually proposed.

The estimation procedure (26) for the trend component, has shown empirically to be quite useful, and it is expected that the competing possibility (21) will lead to high values of m. For example, Johnston [(1972), Section 6.2] reports that m = 6 was needed in one case. For a critical analysis of moving averages in this context see Durbin (1963).

In relation to the result of Jorgenson mentioned in 3.1, we can interpret the present method as an *approximation* to the requirement that the linear "filtering" of the trend component be totally effective. It should be pointed out that Jorgenson proved that one can always find a linear trend to match *exactly* any moving average filter, but the resulting expression may be very different from a polynomial in t.

3.3. *Regression models and the method of difference from moving average in practice.* The argument in 3.2 can be taken as a pedagogical effort to relate the method of difference from moving average, often presented without much justification, with the well-known regression approach with polynomial trend. However the discussion left unexplained the popularity of the former, and in particular of some refinements of it, in relation to the regression approach with polynomial trend component and least-squares estimation, which looks sounder on a statistical ground. One key aspect that we want to emphasize is the empirical treatment of m, the degree of the polynomial used in (1) (we regard as less important the restriction imposed by the polynomial structure, in so far as one is free to choose m, and has enough observations to perform the analysis). From a practical point of view, m is not part of the data but a parameter to be estimated from the available observations. If we consider the problem of fitting (1) to a set of time series, the choice of m must be solved for each series separately, and a wrong choice may produce important specification errors.

One possibility is to use procedure such as that described in Anderson [(1971), Section 3.2.] to decide upon the value of m. In that case, if we consider the procedure consisting of deciding upon the value of m, and then estimating the 12 + m parameters in (4), it is not clear if the resulting estimators will have the statistical properties of paragraph 2.2.

A practical alternative is then to estimate non-parametrically the trend component and hope that for many observed time series the resulting procedure is endowed with statistical properties at least as good as those of the combined procedure described above, and is highly efficient when the series is truly generated by (1).

This argument may contribute to explain the success obtained in practical problems by methods such as that of difference from moving average and other variants and refinements of it (notably the Bureau of the Census Methods), most of which are clearly of empirical origin and lack a rigorous statistical justification. Further the argument contributes to anticipate that it is not imperative that in applied work these methods of empirical extraction be neatly superseded by those based on the regression approach with polynomial trend component.

3.4. *A proposal.* The preceeding analysis shows that it is worthwhile to explore alternative estimation procedures for trend and seasonality, whose known statistical properties are only of the large-sample type.

Model (4) can be replaced by the autoregressive model:

$$y_t = \beta'_0 + \beta'_1 y_{t-1} + \ldots + \beta'_p y_{t-p} + B'_1 D^*_{1t} + \ldots + B'_{11} D^*_{11,t} + u'_t. \qquad (28)$$

Here $y_t, y_{t-1}, \ldots, y_{t-p}$ satisfy a linear stochastic difference equation of order p, or we say that the model is autoregressive of order p.

An autoregression may model the trend component in a more flexible way than a polynomial in t. Model (28) can be taken as a parametric version of (25) which is less restrictive than (1). In this sense it may be expected that it will be a stronger competitor to the empirical-based methods mentioned above.

The 12 + p parameters in (28) can be estimated by least squares. These are the maximum likelihood estimators when the Σ_t are assumed normal. Asymptotically the estimators are consistent and normally distributed. The asymptotic properties are parallel to those of the OLS estimators in the finite-sample case, when the regressors are nonstochastic. See, for example, Anderson [(1971), Chapter 5].

The choice of p, the order of the autoregression, is a statistical problem parallel to that of choosing m in (1), and can also be tackled satisfactorily. See Anderson [(1971), Section 6.4].

The asymptotic results mentioned in the preceeding arguments require that the β'_j be such that the roots of the equation $z^p - \beta'_1 z^{p-1} - \ldots - \beta'_p = 0$ be less than one in absolute value. This assumption makes stationary the corresponding stochastic process. However, in time series with a strong trend component, the estimated values may contradict this assumption. The nonstationary case has also been considered in the literature, but no reference to this treatment will be given here.

The use of an autoregression as trend component, and its estimation by least squares, were suggested by Parzen in a situation closely related to that studied here, as mentioned by Nerlove [(1964), page 258, note 34].

References

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*. Wiley.

Durbin, J. (1963). "Trend elimination for the purpose of estimating seasonal and periodic components of time series," in *Proceedings of the Symposium on Time Series Analysis*, M. Rosemblat, editor, Wiley, p. 3.

Goldberger, A. S. (1964). *Econometric Theory*. Wiley.

Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*, Volume I. McGraw-Hill.

Johnston, J. (1972). *Econometric Methods*, 2nd. Edition, McGraw-Hill.

Jorgenson, D. W. (1964). "Minimum linear, unbiased seasonal adjustment of economic time series," *Journal of the American Statistical Association*. Vol. 59, p. 681.

Mendenhall, W. (1968). *Introduction to Linear Models and the Design and Analysis of Experiments*. Duxbury Press.

Nerlove, M. (1964). "Spectral analysis of seasonal adjustment procedures." *Econometrica*, Vol. 32, p. 241.

Rao, C. R. (1965), *Linear Statistical Inference and Its Applications*. Wiley.

Rao, P. and R. L. Miller (1971). *Applied Econometrics*, Wadsworth.

Scheffé, H. (1959), *The Analysis of Variance*. Wiley.

Shiskin, J., A. H. Young and C. Musgrave (1967). "The XII Variant of the Census Method II Seasonal Adjustment Program." Bureau of the Census Technical Paper No. 15.

Raúl P. Mentz

Ejemplos sobre propiedades asintóticas de los estimadores

Separata de Económica, La Plata, República Argentina
Nº 3, Septiembre - Diciembre, 1971

Ejemplos sobre propiedades asintóticas de los estimadores

Raúl Pedro Mentz

Al enseñar técnicas de estimación y en la investigación aplicada de muchas áreas, a menudo es necesario considerar conceptos asintóticos o "de muestras grandes", pues para muchos problemas prácticos importantes no se conocen estimadores con las propiedades comunes (esto es, las "de muestras finitas"). Una rápida revisión de algunos textos elementales o intermedios que tratan estas cuestiones hace pensar que la lista de ejemplos (más propiamente, de contraejemplos) que se presenta será útil.

Definiciones

Sea $\hat{\Theta}_n = \hat{\Theta}_n (X_1, \ldots, X_n)$ un estimador de Θ que es función de n variables aleatorias observables. Entonces:

(a) $\hat{\Theta}_n$ es *consistente* si la sucesión $\{\hat{\Theta}_n : n = 1, 2, \ldots\}$ satisface $\operatorname{plim}_{n\to\infty} \hat{\Theta}_n = \Theta$.

(b) $\hat{\Theta}_n$ es *asintóticamente insesgado* si las esperanzas matemáticas $E\hat{\Theta}_n$ existen para todo n y su sucesión satisface $\lim_{n\to\infty} E\hat{\Theta}_n = \Theta$.

(c) $\hat{\Theta}_n$ es *insesgado* si $E\hat{\Theta}_n = \Theta$ para cada n.

Queremos mostrar por vía de ejemplo, que ninguna de estas propiedades implica a las otras dos, excepto que (c) implica a (b) trivialmente.

1. Ejemplos sencillos

Sean $X_1, \ldots, X_n$ variables aleatorias normales e independientes con esperanza matemática μ y varianza σ^2 $(0 < \sigma^2 < \infty)$. En realidad para varios de los casos sólo necesitamos hipótesis menos restrictivas.

* Profesor de la Universidad Nacional de Tucumán.

1) $\hat{\mu}_n = (1\text{-}1/n)\,\overline{X}$ y $\hat{\sigma}_n^2 = (1/n)\sum_{i=1}^{n} (X_i - \overline{X}_n)^2$ son asintóticamente insesgados pero sesgados para cada n (fijo).

2) $\hat{\mu}_n = (1/k)\sum_{i=1}^{k} X_i = \overline{X}_k$ y $\hat{\sigma}_n^2 = (1/(k\text{-}1))\sum_{i=1}^{k} (X_i - \overline{X}_k)^2$ donde k es fijo (no cambia con n) y $n = k + 5, k + 6, \ldots$, son insesgados pero inconsistentes. Lo que ocurre aquí es que las varianzas de las distribuciones muestrales son constantes que no cambian con n y en consecuencia aún para n muy grande la variabilidad muestral de los estimadores no tiende a desaparecer

3) $\hat{\mu}_n = (1-1/n)\,\overline{X}_k$ y $\hat{\sigma}_n^2 = (1-1/n)\,(1/(k-1))\sum_{i=1}^{k} (X_i - \overline{X}_k)^2$, donde k es como en el ejemplo 2), son sesgados, asintóticamente insesgados pero inconsistentes.

4) $\hat{\mu}_n = (1/n)\sum_{i=1}^{n-1} X_i$ y $\hat{\sigma}_n^2 = (1/(n+1))\sum_{i=1}^{n} (X_i - X_n)^2$ son consistentes pero sesgados para cada n. En el ejemplo 6) más abajo se considera un caso donde puede utilizarse un estimador como $\hat{\mu}_n$. El estimador σ^2 es "óptimo" en el sentido de que minimiza el error medio cuadrático $E(\tilde{\sigma}_n^2 - \sigma^2)^2$ entre todos los estimadores $\tilde{\sigma}_n^2$ que son funciones de las n variables aleatorias dadas.

5) Supongamos que $\hat{\Theta}_n$ es un estimador consistente e insesgado de Θ.

Definamos un nuevo estimador ("aleatorizado") $\overset{*}{\Theta}_n$ de la siguiente manera: $P(\overset{*}{\Theta}_n = \hat{\Theta}_n) = 1 - 1/n$ y $P(\overset{*}{\Theta}_n = n^a) = 1/n$ para $n=1,2,\ldots$. Entonces $\operatorname{plim}_{n\to\infty} \overset{*}{\Theta}_n = \Theta$ pero

$$E\,\overset{*}{\Theta}_n = (1\text{-}1/n)\,\Theta + n^{a-1} \text{ y si } a > 1$$

la sucesión de esperanzas no converge. En consecuencia la consistencia no implica la carencia de sesgo asintótico. El problema en este ejemplo radica en que mientras el grueso de la *probabilidad* se concentra alrededor de Θ, cantidades pequeñas pero positivas están asociadas con el *valor* n^a que crece indefinidamente cuando

n crece, si a > 1. Como consecuencia a veces se dice que $E\,\overset{*}{\Theta}_n$, $\mathrm{var}(\overset{*}{\Theta}_n)$, etc., "no existen asintóticamente".

2. Ejemplos del área de series cronológicas

Los ejemplos de la sección I tienen como objeto mostrar la naturaleza de las dificultades, pero algunos son comparativamente triviales pues nadie consideraría a los estimadores propuestos como razonables. Ejemplos más realistas provienen de los problemas de estimación en series cronológicas, donde la teoría asintótica juega un papel central.

Sea $\{ y_t \}$ una serie cronológica estacionaria con $E\, y_t = 0$ para todo t, función de densidad espectral $f(\lambda)$ continua en $-\pi \leq \lambda \leq \pi$, y covarianzas $\{ \sigma(h) : h=1,2,\ldots \}$; aquí $\sigma(h) = \sigma(-h)$. Suponga que observamos $y_1, y_2, \ldots, y_T$.

6) $I_T(\lambda) = (1/2\pi) \sum_{r=-(T-1)}^{T-1} \overset{\wedge}{\sigma}_T(r) \cos \lambda r$, es el *periodograma de Schuster*. Los estimadores de las covarianzas son $\overset{\wedge}{\sigma}_T(-h) = (1/T\text{-}h)) \sum_{t=1}^{T-h} y_t y_{t+h}$, para h=0,1,...,T-1 . Es conocido que este estimador posee las siguientes propiedades:

(i) $E\, I_T(\lambda) \neq f(\lambda)$ en general

(ii) $\lim_{T \to \infty} E\, I_T(\lambda) = f(\lambda)$

(iii) Aún bajo supuestos fuertes con respecto al proceso $\{ y_t \}$ y a su densidad espectral $f(\lambda)$, $I_T(\lambda)$ no es un estimador consistente de $f(\lambda)$; en realidad el estimador (multiplicado por una constante) tiene como distribución asintótica una distribución "ji al cuadrado".

En consecuencia el periodograma, como estimador de $f(\lambda)$ y aún bajo hipótesis comparativamente fuertes, es sesgado, asintóticamente insesgado e inconsistente. *

* Algunos autores utilizan $\sigma_T^*(h) = ((T\text{-}h)/T)\, \overset{\wedge}{\sigma}_T(h)$ como estimador de $\sigma(h)$ para formar estimadores de $f(\lambda)$. Este es sesgado (mientras que $\overset{\wedge}{\sigma}_T(h)$ no lo es), asintóticamente insesgado y consistente, bajo algunas hipótesis.

7) Supongamos que además de las hipótesis generales de esta sección, la serie cronológica dada es un proceso autoregresivo del primer orden, esto es, que satisface la ecuación diferencial (estocástica) del primer orden $y_t = -\beta\, y_{t-1} + u_t$, $t=\ldots,-1,0,1,\ldots$ donde $|\beta| < 1$ y $\{ u_t \}$ es una sucesión de variables aleatorias no correlacionadas con $E\, u_t = 0$ y $E\, u_t^2 = 1$. El estimador del tipo de mínimos cuadrados.

$$\overset{\wedge}{\beta}_T = - \frac{(1/(T\text{-}1)) \sum_{t=1}^{T-1} y_t y_{t+1}}{(1/T) \sum_{t=1}^{T-1} y_t^2}$$

tiene las propiedades:

(i) $E\overset{\wedge}{\beta}_T \neq \beta$

(ii) $\operatorname{plim}_{T \to \infty} \overset{\wedge}{\beta}_T = \beta$

En consecuencia $\overset{\wedge}{\beta}_T$ es sesgado pero consistente.

Inverting a covariance matrix of toeplitz type by the method of cofactors

Raúl Pedro Mentz

SUMMARY. A matrix is said to be of Toeplitz type if it has equal elements along diagonals. These matrices, with the additional property of symmetry, arise frequently in statistical work, as covariance matrices of wide-sense stationary stochastic processes, in nonparametric theory, etc. The inverse is often of interest, and a method is used to find its components when the TxT matrix has only 5 nonvanishing central diagonals. The method is to express the cofactors of the components of the given matrix in terms of some determinants, that are shown to satisfy certain linear difference equations, and to solve these explicitly. The complexity of the resulting expressions for the components of the inverse is comparable to those known in the literature.

1. Introduction

A matrix $\underset{\sim}{A} = (a_{ij})$ is called a *Toeplitz matrix* if $a_{ij} = a_{i-j}$.

In mathematical statistics Toeplitz matrices arise in several contexts; see, for example, Grenader and Szegö [3]. They appear as covariance matrices of wide-sense stationary stochastic processes, in which case they are symmetric and positive semidefinite. In the present paper we assume throughout that they are positive definite and symmetric (i.e. are covariance matrices in nonsingular cases), even when some of our results hold for nonsingular symmetric matrices in general.

If $\underset{\sim}{\Sigma} = (\sigma_{ij}) = (\sigma_{|i-j|})$ is the given matrix, the underlying assumptions often imply that the components vanish if $|i-j| > m$, where $m \geq 0$ is an integer. We may call the corresponding processes finitely correlated of order m, and m=0 is the case of lack of correlation. This occurs, for example, in the moving average model of order m, $x_t = v_t + a_1 v_{t-1} + \ldots + a_m v_{t-m}$, the v_t being uncorrelated random variables with common means and (finite) variances.

There exists wide interest in finding either exact or approximate forms for the components of the inverses of these Toeplitz matrices, since that knowledge can be used to derive the statistical theory of procedures defined in terms of them. For example, the author's interest in the inverse matrix studied here stems from the analysis of Walker's [12] estimation procedure for the moving average time-series model. See Mentz [6].

For the case of m=1, explicit forms for the components of $\underset{\sim}{\Sigma}_T^{-1}$ obtained by means of the method of cofactors are well known; see, for example, Shaman [8]. It has been conjectured that it would be feasible to extend the procedure to m > 1. In the present work we deal with the case of m=2 in detail.

In the case of m=1 several other approaches also solved the problem of finding the components of $\underset{\sim}{\Sigma}_T^{-1}$. For higher values of m the problem proved difficult. Shaman [9] exhibited a close-form expression for the components of the inverse matrix when m=2, and for that case also Mentz [5] has an expression. For general Toeplitz matrices there is a paper by Calderon, Spitzer and Widom [1], but it appears as a hard problem to deduce from their solution an useful explicit form for the components of $\underset{\sim}{\Sigma}_T^{-1}$ for finite T.

A useful notation is $\underset{\sim}{\Sigma} = \sum_{k=0}^{m} \sigma_k \underset{\sim}{G}_k$, where $\underset{\sim}{G}_k$ has components $g_{ij}^{(k)}=1$ for $|i-j| = k$, and otherwise equal to 0. For a solution of a similar problem with different $\underset{\sim}{G}_j$ see Mustafi [7]. In this notation $\underset{\sim}{\Sigma} = \sigma_0 \sum_{k=0}^{m} \rho_k \underset{\sim}{G}_k$, where $\rho_j = \sigma_j/\sigma_0$ ($\sigma_0 > 0$ because $\underset{\sim}{\Sigma}$ is positive definite), and we see that there is no loss of generality in taking the coefficient of $\underset{\sim}{G}_0$ to be equal to 1, as will be done below.

In Section 2 we show that the cofactors of the components of $\underset{\sim}{\Sigma}_T^{-1}$ can be written in terms of some determinants that in turn satisfy linear difference equations that we propose to solve explicitly. Then in Section 3 we use the analysis of Section 2 to derive some close-form and recursive expressions that are comparatively easy to apply analitically and computationally. However, we do not study the computational merits of the proposals as compared with computer rutines prepared for certain Toeplitz matrices; see, for example, Trench [10].

2. The inverse of $\underset{\sim}{I} + \rho_1\underset{\sim}{G}_1 + \rho_2\underset{\sim}{G}_2$ by evaluation of cofactors

Let $\underset{\sim}{\Sigma}_T = (\sigma_{ij}) = \underset{\sim}{I} + \rho_1\underset{\sim}{G}_1 + \rho_2\underset{\sim}{G}_2$, $\rho_2 \neq 0$, and $\underset{\sim}{\Sigma}_T^{-1} = \underset{\sim}{W}_T = (w_{ij}^{(T)})$.

The components of $\underset{\sim}{W}_T$ can be computed from

$$(2.1) \qquad w_{ij}^{(T)} = \frac{\text{cofactor of } \sigma_{ij}}{|\underset{\sim}{\Sigma}_T|}$$

In this section we use the following notation, where a subscript denotes the order of the corresponding matrix or determinant, and we omit the superscripts in the components to simplify the writing. We also use the notation of partitioned matrices:

$$(2.2) \qquad \Sigma_s = |\underset{\sim}{\Sigma}_s|$$

$$\underset{\sim}{L}_s = \begin{bmatrix} \rho_1 & 1 & \rho_1 & \rho_2 & 0 & \cdots & 0 & 0 & 0 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 & \cdots & 0 & 0 & 0 \\ 0 & \rho_2 & \rho_1 & 1 & \rho_1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & \rho_2 & \rho_1 & 1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \rho_2 & \rho_1 \end{bmatrix} = \left[\begin{array}{c|ccc} \rho_1 & & & \\ \rho_2 & & & \\ 0 & & \underset{\sim}{\Sigma}_{s-1} & \\ \vdots & & & \\ 0 & & & \\ \hline 0 & 0 & \cdots 0 & \rho_2\ \rho_1 \end{array}\right]$$

$$(2.3) \qquad L_s = |\underset{\sim}{L}_s|$$

$$\underset{\sim}{K}_s = \begin{bmatrix} \rho_1 & \rho_2 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \rho_1 & 1 & \rho_1 & \rho_2 & 0 & \cdots & 0 & 0 & 0 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \rho_2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & \rho_1 & 1 & \rho_1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & \rho_2 & \rho_1 & 1 \end{bmatrix} = \left[\begin{array}{c|ccc} \rho_1 & \rho_2 & 0 & \cdots\ 0 \\ \hline \rho_1 & & & \\ \rho_2 & & & \\ 0 & & \underset{\sim}{\Sigma}_{s-1} & \\ \vdots & & & \\ 0 & & & \end{array}\right]$$

$$(2.4) \qquad K_s = |\underset{\sim}{K}_s|$$

By expanding Σ_T in terms of the components in its first row, Durbin ([2], p. 315) found that the determinants satisfy the linear, homogeneous, fifth-order difference equation

$$(2.5) \quad -\Sigma_n + (1-\rho_2)\Sigma_{n-1} + (\rho_2-\rho_1^2)\Sigma_{n-2} + \rho_2(\rho_1^2-\rho_2)\Sigma_{n-3} + \rho_2^3(\rho_2-1)\Sigma_{n-4} + \rho_2^5\Sigma_{n-5} = 0$$

The associated polynomial equation

$$(2.6) \quad -z^5 + (1-\rho_2)z^4 + (\rho_2-\rho_1^2)z^3 + \rho_2(\rho_1^2-\rho_2)z^2 + \rho_2^3(\rho_2-1)z + \rho_2^5 = 0$$

can be written in a symmetric way using the substitution $-\rho_2 x = z$;

after division by ρ_2^5 we obtain

$$(2.7)\quad x^5 + \frac{1-\rho_2}{\rho_2}x^4 + \frac{\rho_1^2-\rho_2}{\rho_2^2}x^3 + \frac{\rho_1^2-\rho_2}{\rho_2^2}x^2 + \frac{1-\rho_2}{\rho_2}x + 1 = 0$$

Zero is not a root of (2.7), and if x^* is a root so is $1/x^*$. Since there must be five roots, +1 or -1 must be one of them. (They are the only "self inverses".) By inspection we see it is -1. Then

$$(2.8)\quad z_1 = \rho_2\ ,\quad z_2 = \frac{-\rho_2}{2}(d_1 + \sqrt{d_1^2-4})\ ,\quad z_3 = \frac{-2\rho_2}{d_1+\sqrt{d_1^2-4}}$$

$$z_4 = \frac{-\rho_2}{2}(d_2+\sqrt{d_2^2-4})\ ,\quad z_5 = \frac{-2\rho_2}{d_2+\sqrt{d_2^2-4}}$$

where

$$(2.9)\quad d_1, d_2 = (2\rho_2)^{-1}[2\rho_2 - 1 \pm \sqrt{(2\rho_2+1)^2 - 4\rho_1^2}]$$

In general the roots (2.8) can be real or complex, and some or all can be identical. Hence the solution of (2.5) will take different forms depending on this fact. As an example, which will be also used as illustration in subsequent derivations, if all roots are distinct then (2.5) has solution

$$(2.10)\quad \Sigma_n = \sum_{i=1}^{5} C_i z_i^n$$

where the z_i are the roots given in (2.8).

Since Σ_n is defined only for $n \geq 1$, (2.5) holds for $n \geq 6$, and the sequence satisfying the difference equation and for which (2.10) is the general solution is $\Sigma_1, \Sigma_2, \ldots$. The boundary conditions to determine C_i, $i = 1,\ldots,5$, can be taken to be (2.10) for $n = 1,\ldots,5$, with the left-hand sides evaluated explicitly as

$$(2.11)\quad \begin{aligned} \Sigma_1 &= 1 \\ \Sigma_2 &= 1-\rho_1^2 \\ \Sigma_3 &= (1-\rho_2)(1+\rho_2-2\rho_1^2) \\ \Sigma_4 &= \Sigma_3 - (\rho_1^2+\rho_2^2) + (\rho_1^4+\rho_2^4) + 2\rho_1^2\rho_2 - 2\rho_1^2\rho_2^2 \\ \Sigma_5 &= \Sigma_4 - \rho_1^2\Sigma_3 + 2\rho_1^2\rho_2(1-\rho_1^2-\rho_2+\rho_2^2) - \rho_2^2(1-\rho_1^2-\rho_2^2) \end{aligned}$$

Following the same approach we expand L_T in terms of the components

in its first row and find that

$$(2.12)\quad L_n - \rho_1 L_{n-1} + \rho_2 L_{n-2} - \rho_1\rho_2^2 L_{n-3} + \rho_2^4 L_{n-4} = 0$$

The polynomial equation is

$$(2.13)\quad y^4 - \rho_1 y^3 + \rho_2 y^2 - \rho_1\rho_2^2 y + \rho_2^4 = 0$$

and after replacing $\rho_2 x = y$ it becomes

$$(2.14)\quad \rho_2^4 x^4 - \rho_1\rho_2^3 x^3 + \rho_2^3 x^2 - \rho_1\rho_2^3 x + \rho_2^4 = 0$$

which has symmetric coefficients and can be studied in the same way as equation (2.7). The roots y_s, $s = 1,2,3,4$, of (2.13) are obtained from

$$d_1 = \frac{1}{2\rho_2}(\rho_1+\sqrt{\rho_1^2-4\rho_2+8\rho_2^2})\ ,\quad d_2 = \frac{1}{2\rho_2}(\rho_1-\sqrt{\rho_1^2-4\rho_2+8\rho_2^2})$$

$$(2.15)\quad y_1 = \frac{\rho_2}{2}(d_1+\sqrt{d_1^2-4})\ ,\quad y_2 = \frac{2\rho_2}{d_1+\sqrt{d_1^2-4}}\ ,$$

$$y_3 = \frac{\rho_2}{2}(d_2+\sqrt{d_2^2-4})\ ,\quad y_4 = \frac{2\rho_2}{d_2+\sqrt{d_2^2-4}}$$

The particular case of all roots distinct leads to solving (2.12) by the sequence

$$(2.16)\quad L_n = \sum_{i=1}^{4} C_i^* y_i^n \qquad n = 1,2,\ldots\ .$$

The four boundary conditions needed to determine the C_i^*'s can be taken to be (2.16) for $n = 1,2,3,4$ with the left-hand sides evaluated explicitly as

$$(2.17)\quad \begin{aligned} L_1 &= \rho_1 \\ L_2 &= \rho_1^2-\rho_2 \\ L_3 &= \rho_1^3+\rho_2^2\rho_1-2\rho_1\rho_2 \\ L_4 &= \rho_1 L_3 - \rho_2[(\rho_1^2-\rho_2) - \rho_2(\rho_1^2-\rho_2^2)] \end{aligned}$$

Expanding K_T by the components in its first row we have

$$(2.18)\quad K_n = \rho_1\Sigma_{n-1} - \rho_2 K_{n-1}\ ,\quad n = 2,3,\ldots\ .$$

In the special case that Σ_n is given by (2.10),

$$(2.19) \qquad K_n + \rho_2 K_{n-1} = \rho_1 \sum_{i=1}^{5} C_i z_i^{n-1}$$

which is a first-order, inhomogeneous, linear difference equation. Provided that only one root (for Σ_T) equals ρ_2,

$$(2.20) \qquad K_n = \tilde{C}(-\rho_2)^n + \frac{1}{2}\rho_1 C_1 \frac{1}{\rho_2}\rho_2^n + \rho_1 \sum_{j=2}^{5} C_j \frac{1}{z_j+\rho_2} z_j^n$$

The second summand corresponds to the root $z_1 = \rho_2$; no other z_j can be equal to ρ_2 in (2.20); if more than one root equals ρ_2, instead of the factor $1/(z_j+\rho_2)$ we have to use $1/2\rho_2$.

The new constant $\tilde{C}$ in (2.20) will be evaluated from (2.20) for n=2, with $K_1 = \rho_1$. Note that

$$(2.21) \qquad \begin{aligned} K_2 &= \rho_1(1-\rho_2) \\ K_3 &= \rho_1(1-\rho_1^2) - \rho_1\rho_2(1-\rho_2) \\ K_4 &= \rho_1\Sigma_3 - \rho_2 K_3 = \rho_1(1-\rho_2)(1+\rho_2-2\rho_1^2) - \rho_2 K_3 \end{aligned}$$

With this background we now find expressions for the components w_{ij} of $\underset{\sim}{W}_T = \underset{\sim}{\Sigma}_T^{-1}$. Since $\underset{\sim}{W}_T$ is symmetric we restrict attention to the components on and above the main diagonal.

1st case: i=j. Then $w_{ii} = B_{ii}/\Sigma_T$, where B_{ii} is the cofactor of σ_{ii}. In terms of submatrices

$$(2.22) \qquad B_{ii} = \begin{bmatrix} \underset{\sim}{\Sigma}_{i-1} & \rho_2 \underset{\sim}{E}^{*\prime} \\ \rho_2 \underset{\sim}{E}^{*} & \underset{\sim}{\Sigma}_{T-i} \end{bmatrix}$$

where $\underset{\sim}{E}^*$ has its upper right-hand element equal to 1 and all other elements equal to zero. We use Laplace's expansion in terms of minors of the first i-1 columns; then

$$(2.23) \qquad B_{ii} = \Sigma_{i-1}\Sigma_{T-i} - \rho_2^2 \Sigma_{i-2}\Sigma_{T-i-1}$$

To make (2.23) valid for all i, we define $\Sigma_0 = 1$, $\Sigma_{-1} = 0$.

2nd case: $i < j$.

$$(2.24) \qquad (-1)^{i+j} B_{ij} = \begin{bmatrix} \underset{\sim}{\Sigma}_{i-1} & \underset{\sim}{F} & \underset{\sim}{0} \\ \rho_2 \underset{\sim}{E}^{*} & \underset{\sim}{L}_{j-i} & \underset{\sim}{F} \\ \underset{\sim}{0} & \rho_2 \underset{\sim}{E}^{*} & \underset{\sim}{\Sigma}_{T-j} \end{bmatrix}$$

where $\underset{\sim}{F}$ has its lower left-hand element equal to ρ_1, the two adjacent elements equal to ρ_2, and all other elements equal to 0. We expand (2.24) by Laplace's formula in terms of minors of the first i-1 columns. In these columns there are three non-vanishing minors with non-zero complementary minors, namely

$$(2.25) \qquad |\underset{\sim}{\Sigma}_{i-1}| = \Sigma_{i-1}$$

$$(2.26) \qquad \left[\begin{array}{ccc|c} & \underset{\sim}{\Sigma}_{i-2} & & \begin{matrix} 0 \\ \vdots \\ \rho_2 \\ \rho_1 \end{matrix} \\ \hline 0 & \cdots\cdots & 0 & \rho_2 \end{array}\right] = \rho_2 \Sigma_{i-2}$$

and

$$(2.27) \qquad \left[\begin{array}{ccc|c} & \underset{\sim}{K}^*_{i-2} & & \begin{matrix} 0 \\ \vdots \\ \rho_2 \\ 1 \end{matrix} \\ \hline 0 & \cdots\cdots & 0 & \rho_2 \end{array}\right] = \rho_2 K_{i-2}$$

where $\underset{\sim}{K}^*_s$ is $\underset{\sim}{K}_s$ flipped about its secondary diagonal, so that $|\underset{\sim}{K}^*_s| = K_s$. If we denote by $A_s(i,j)$, s = 1,2,3, the corresponding cofactors, then

$$(2.28) \qquad (-1)^{i+j} B_{ij} = \Sigma_{i-1}A_1(i,j) - \rho_2\Sigma_{i-2}A_2(i,j) + \rho_2 K_{i-2}A_3(i,j)$$

The $A_s(i,j)$'s are computed using Laplace's expansion in terms of the last T-j columns. Then

$$A_1(i,j) = \Sigma_{T-j}L_{j-i} - \rho_2 K_{T-j}L_{j-i-1} + \rho_2^3 \Sigma_{T-j-1}L_{j-i-2}$$

$$(2.29) \qquad \begin{aligned} A_2(i,j) = {} & \Sigma_{T-j}(\rho_1 L_{j-i-1} - \rho_2^2 L_{j-i-2}) - \rho_2 K_{T-j}(\rho_1 L_{j-i-2} - \rho_2^2 L_{j-i-3}) + \\ & + \rho_2^3 \Sigma_{T-j-1}(\rho_1 L_{j-i-3} - \rho_2^2 L_{j-i-4}) \end{aligned}$$

$$A_3(i,j) = \rho_2 \Sigma_{T-j}L_{j-i-1} - \rho_2^2 K_{T-j}L_{j-i-2} + \rho_2^4 \Sigma_{T-j-1}L_{j-i-3}$$

For $j = [T/2]+1,\ldots,T$, say, these formulas are valid for all $i < j$, provided we define $\Sigma_0 = L_0 = 1$, $K_0 = 0$, $\Sigma_{-s} = L_{-s} = K_{-s} = 0$ for $s > 0$. For $j < [T/2]+1$ similar arrangements could be made. In fact, due to the structure of $\underset{\sim}{W}_T$ we only need to compute those components of the last $[(T+1)/2]$ columns on and between the principal and sec-

ondary diagonals, and then deduce the remaining components using the symmetry of $\underset{\sim}{W}_T$ and its *persymmetry* (symmetry with respect to its secondary diagonal). They lead for example to

$$(2.30) \qquad w_{ij} = w_{T-i+1,T-j+1} \quad , \quad i=j,\ldots,T-j; \quad j=1,2,\ldots,[T/2]$$

We summarize these results as follows:

PROPOSITION 2.1. *Let* $\underset{\sim}{\Sigma}_T = \underset{\sim}{I}+\rho_1\underset{\sim}{G}_1+\rho_2\underset{\sim}{G}_2$, *with* $\rho_2 \neq 0$, *and* $\underset{\sim}{\Sigma}_T^{-1} = \underset{\sim}{W}_T = (w_{ij}^{(T)})$. *Then*

$$(2.31) \qquad w_{ij}^{(T)} = (-1)^{i+j}\frac{B_{ij}}{\Sigma_T} \quad , \quad i=j,\ldots,T-j+1; \quad j=[T/2]+1,\ldots,T$$

where the B_{ij} *are given in* (2.23) *when* i=j *and in* (2.28) *and* (2.29) *when* $i<j$, *in terms of the determinants* Σ_s, L_s, K_s, *which are defined in* (2.2)-(2.4) *and satisfy the difference equations* (2.5), (2.12) *and* (2.18), *respectively. The remaining elements of* $\underset{\sim}{W}_T$ *are obtained using* $w_{ij}^{(T)} = w_{ji}^{(T)}$ *and* (2.30).

If $\rho_2 = 0$ but $\rho_1 \neq 0$, then $L_s = \rho_1^s$, $K_s = \rho_1\Sigma_{s-1}$, $(-1)^{i+j}B_{ij} = \Sigma_{i-1}A_1(i,j)$, $A_1(i,j) = \rho_1^{j-i}\Sigma_{T-j}$, and the solution reduces to

$$(2.32) \qquad w_{ij}^{(T)} = (-\rho_1)^{j-i}\frac{\Sigma_{i-1}\Sigma_{T-j}}{\Sigma_T}$$

Here Σ_n satisfies the corresponding version of (2.5), namely

$$(2.33) \qquad \Sigma_n - \Sigma_{n-1} + \rho_1^2\,\Sigma_{n-2} = 0$$

with boundary conditions $\Sigma_1 = 1$, $\Sigma_2 = 1 - \rho_1^2$ instead of (2.11).

3. Alternative forms of the components of the inverse matrix

In this section we want to use the analysis of Section 2 to obtain other forms for the $w_{ij}^{(T)}$ that can be of greater use. We shall find useful the following result:

DEFINITION. *A matrix* $\underset{\sim}{A} = (a_{ij})$ *is said to be "diagonal of type* r" *if* $a_{ij} = 0$ *whenever* $|i-j| > r$.

PROPOSITION 3.1. *Let* $\underset{\sim}{A} = (a_{ij})$ *be a* TxT *symmetric and positive definite matrix. A necessary and sufficient condition that* $\underset{\sim}{A}^{-1}$ *be diagonal of type* r *is that there exist constants* b_{ts} *such that for* t = 1,2,...,T-r+1

$$(3.1) \qquad a_{tt'} + b_{t1}a_{t+1,t'} + \ldots + b_{t,r-1}a_{t+r-1,t'} = 0, \quad t'=t+1,\ldots,T$$

This result was apparently originated with Guttman [4] and Ukita [11], and a detailed proof is given in Mentz [5].

Condition (3.1) states the existence of a linear relation between succesive sets of r adjacent rows of $\underset{\sim}{A}$; an equivalente formulation (to be used beIow) is to relate the first r-1 rows to each of the remaining ones.

We now proceed to derive a closed-form expression for the $w_{ij}^{(T)}$. From (2.28) and (2.29) we derive the components in rows 1 and 2 (or columns T and T-1,respectively) of $\underset{\sim}{W}_T = \underset{\sim}{\Sigma}_T^{-1}$, as follows:

$$(3.2) \qquad w_{1j}=w_{T-j+1,T} = \frac{(-1)^{j+1}}{\Sigma_T}(\Sigma_{T-j}L_{j-1}-\rho_2K_{T-j}L_{j-2}+\rho_2^3\Sigma_{T-j-1}L_{j-3}), \quad j=2,\ldots,T$$

$$w_{2j}=w_{T-j+1,T-1} = \frac{(-1)^{j+2}}{\Sigma_T}[\Sigma_{T-j}L_{j-2}-\rho_2(\rho_1\Sigma_{T-j}+K_{T-j})L_{j-3} +$$

$$(3.3) \qquad +\rho_2^2(\rho_2\Sigma_{T-j}+\rho_2\Sigma_{T-j-1}+\rho_1K_{T-j})L_{j-4}-\rho_2^4(\rho_1\Sigma_{T-j-1}+$$

$$+K_{T-j})L_{j-5}+\rho_2^6\Sigma_{T-j-1}L_{j-6}] \quad , \quad j=3,\ldots,T$$

Since $\underset{\sim}{\Sigma}_T$ is "diagonal of type 3", it follows that there exist constants θ_{1i} and θ_{2i} such that

$$(3.4) \qquad w_{ij} = \theta_{1i}w_{1j} + \theta_{2i}w_{2j} \quad , \quad i=3,\ldots,T; \quad j \geq i$$

Using these relations for j = T-1, T we form the systems

$$(3.5) \qquad \begin{aligned} w_{i,T-1} &= \theta_{1i}w_{1,T-1} + \theta_{2i}w_{2,T-1} \\ w_{i,T} &= \theta_{1i}w_{1T} + \theta_{2i}w_{2T} \quad , \quad i = 3,\ldots,T \end{aligned}$$

that can be solved for the θ_{1i} and θ_{2i}; replacing the resulting values in (3.4) we obtain the following:

PROPOSITION 3.2. *Under the conditions of Proposition 2.1,*

$$w_{ij}^{(T)} = \frac{(w_{2,T-1}w_{iT} - w_{2T}w_{i,T-1})w_{T-j+1,T} - (w_{1,T-1}w_{iT} - w_{1T}w_{i,T-1})w_{T-j+1,T-1}}{w_{1T}w_{2,T-1} - w_{1,T-1}w_{2T}}$$

(3.6)

$$= \frac{(w_{2,T-1}w_{1,T-i+1} - w_{1,T-1}w_{2,T-i+1})w_{1j} - (w_{2T}w_{1,T-i+1} - w_{1T}w_{2,T-i+1})w_{2j}}{w_{1T}w_{2,T-1} - w_{2T}w_{1,T-1}} \qquad i \leq j$$

where the necessary components are given in (3.2) *and* (3.3).

It is easily checked that (3.6) holds for all $i \leq j$. As in the case of (2.31), it suffices to compute w_{ij} for $i = j,\ldots,T-j+1$, $j = [T/2]+1,\ldots,T$.

Expression (3.6) exhibits the components of the inverse matrix as functions of the components in columns T and T-1 (or rows 1 and 2), and in turn, using (3.2) and (3.3), as functions of the roots of (2.6), (2.13) and $-\rho_2$, the latter being the root of the polynomial equation associated with the sequence of K_n's.

While (3.6) may be useful for analytic purposes, the following recursive approach may be simpler, since the $w_{ij}^{(T)}$ will be given as functions of the determinants Σ_s^{ij} and K_s only, not of the L_s.

The components along the diagonals near the main diagonal give rise to some interesting simplifications. In effect from (2.28) we have that

$$(-1)B_{i,i+1} = \Sigma_{i-1}(\Sigma_{T-i-1}L_1 - \rho_2 K_{T-i-1}L_0) - \rho_2\Sigma_{i-2}\Sigma_{T-i-1}\rho_1 L_0 +$$

(3,7)

$$+ \rho_2 K_{i-2}\rho_2\Sigma_{T-i-1} = \Sigma_{i-1}(\rho_1\Sigma_{T-i-1} - \rho_2 K_{T-i-1}) -$$

$$- \rho_1\rho_2\Sigma_{i-2}\Sigma_{T-i-1} + \rho_2^2 K_{i-2}\Sigma_{T-i-1}$$

using the explicit values for some L_s's given in (2.17). Similarly

$$(-1)^2 B_{i,i+2} = \Sigma_{i-1}[(\rho_1^2 - \rho_2)\Sigma_{T-i-2} - \rho_1\rho_2 K_{T-i-2} + \rho_2^3\Sigma_{T-i-3}] - \rho_2\Sigma_{i-2}$$

(3.8)

$$[(\rho_1^2 - \rho_2^2)\Sigma_{T-i-2} - \rho_1\rho_2 K_{T-i-2}] + \rho_2 K_{i-2}(\rho_1\rho_2\Sigma_{T-i-2} - \rho_2^2 K_{T-i-2})$$

$$(-1)B_{i,i+3} = \Sigma_{i-1}[(\rho_1^3 + \rho_1\rho_2^2 - 2\rho_1\rho_2)\Sigma_{T-i-3} - \rho_2(\rho_1^2 - \rho_2)K_{T-i-3} +$$

(3.9)

$$+ \rho_1\rho_2^3\Sigma_{T-i-4}] - \rho_2\Sigma_{i-2}[(\rho_1^2 - \rho_2 - \rho_1\rho_2^2)\Sigma_{T-i-3} - \rho_2(\rho_1^2 - \rho_2^2)K_{T-i-3} -$$

$$- \rho_1\rho_2^3\Sigma_{T-i-4}] + \rho_2 K_{i-2}[\rho_2(\rho_1^2 - \rho_2)\Sigma_{T-i-3} -$$

$$- \rho_1\rho_2^2 K_{T-i-3} + \rho_2^4\Sigma_{T-i-4}]$$

For the present case of r=3 expression (3.1) reads

(3.10) $$w_{ij} = -b_{i1}w_{i+1,j} - b_{i2}w_{i+2,j} \quad , \quad i = 1,\ldots,T-2; \quad j > i$$

using this expression for columns i+2 and i+3 we form the systems

(3.11)

$$w_{i,i+2} = -b_{i1}w_{i+1,i+2} - b_{i2}w_{i+2,i+2}$$

$$w_{i,i+3} = -b_{i1}w_{i+1,i+3} - b_{i2}w_{i+2,i+3} \quad , \quad i = 1,\ldots,T-3$$

From these systems we derive b_{i1} and b_{i2}, substitute them back in (3.10) and obtain

(3.12)

$$w_{ij} = \frac{(w_{i+2,i+3}w_{i,i+2} - w_{i+2,i+2}w_{i,i+3})w_{i+1,j}}{w_{i+1,i+2}w_{i+2,i+3} - w_{i+1,i+3}w_{i+2,i+2}} -$$

$$- \frac{(w_{i+1,i+3}w_{i,i+2} - w_{i+1,i+2}w_{i,i+3})w_{i+2,j}}{w_{i+1,i+2}w_{i+2,i+3} - w_{i+1,i+3}w_{i+2,i+2}}$$

The components of the inverse matrix are then computed as follows:

PROPOSITION 3.3. *Under the conditions of Proposition 2.1, the components* $w_{ij} = w_{ij}^{(T)}$ *of the inverse matrix* $\underset{\sim}{W}_T = \underset{\sim}{\Sigma}_T^{-1}$ *are determined as follows:*

(a) *Determine* $w_{i,i+s}$ *for* s = 0,1,2,3 *according to*

(3.13) $$w_{ii} = \frac{1}{\Sigma_T}(\Sigma_{i-1}\Sigma_{T-i} - \rho_2^2\Sigma_{i-2}\Sigma_{T-i-1}) \quad , \quad i = 1,\ldots,T$$

(3.14) $$w_{i,i+1} = \frac{(-1)}{\Sigma_T}[\Sigma_{T-i-1}(\rho_1\Sigma_{i-1} - \rho_1\rho_2\Sigma_{i-2} + \rho_2^2 K_{i-2}) - \rho_2 K_{T-i-1}\Sigma_{i-1}], \quad i = 2,\ldots,T$$

(3.15)

$$w_{i,i+2} = \frac{1}{\Sigma_T}[\Sigma_{T-i-2}\{(\rho_1^2 - \rho_2)\Sigma_{i-1} - \rho_2(\rho_1^2 - \rho_2^2)\Sigma_{i-2} + \rho_1\rho_2^2 K_{i-2}\} +$$

$$+ \rho_2^3\Sigma_{T-i-3}\Sigma_{i-1} + K_{T-i-2}\{-\rho_1\rho_2\Sigma_{i-1} + \rho_1\rho_2^2\Sigma_{i-2} - \rho_2^3 K_{i-2}\}], \quad i = 3,\ldots,T$$

$$w_{i,i+3} = \frac{(-1)}{\Sigma_T}[\Sigma_{T-i-3}\{(\rho_1^3 + \rho_1\rho_2^2 - 2\rho_1\rho_2)\Sigma_{i-1} - \rho_2(\rho_1^2 - \rho_2 - \rho_1\rho_2^2)\Sigma_{i-2} +$$

(3.16)

$$+ \rho_2^2(\rho_1^2 - \rho_2)K_{i-2}\} + \Sigma_{T-i-4}(\rho_1\rho_2^3\Sigma_{i-1} + \rho_1\rho_2^4\Sigma_{i-2} + \rho_2^5 K_{i-2}) +$$

$$+ K_{T-i-3}\{-\rho_2(\rho_1^2-\rho_2)\Sigma_{i-1}+\rho_2^2(\rho_1^2-\rho_2^2)\Sigma_{i-2}-\rho_1\rho_2^3 K_{i-2}\}] ,$$

$$i = 4,\ldots,T$$

(b) *For column* j, $[T/2]+1 \leq j \leq T$, *find in succession* $w_{j-s,j}$ *for* $s = 4,\ldots,2j-T-1$, *using* (3.12).

(c) *Determine the remaining* w_{ij} *using the symmetry and persymmetry of* $\underset{\sim}{W}_T$.

Expression (3.13) is derived from (2.23), and expressions (3.14)-(3.16) from (3.7)-(3.9), respectively.

As was remarked above, Proposition 3.3 solves the problem of specifying the w_{ij} as functions of the determinants Σ_n and K_n of different orders. Also note [for example, see equation (2.19)] that once the Σ_n are available the K_n are easily determined.

ACKNOWLEDGMENTS. The main part of this study is part of the author's Ph.D. dissertation in the Department of Statistics, Stanford University. Appreciation is expressed to Professor Theodore W. Anderson for generous guidance and help while directing this work and to Professor Paul Shaman for corrections to a first draft. Work at Stanford was supported by a Research Contract with the Office of Naval Research (T.W. Anderson, Project Director).

References

[1] CALDERON,A., SPITZER,F. and WIDOM,H.(1959), *Inversion of Toeplitz Matrices*, Illinois J. of Mathematics, 3,490-498.

[2] DURBIN,J. (1959), *Efficient estimation of parameters in moving average models*, Biometrika, 46, 306-316.

[3] GRENANDER,U. and G.ZSEGÖ (1958), *Toeplitz Forms and Their Applications*, U. of California Press.

[4] GUTTMAN, L. (1955), *A generalized simplex for factor analysis*, Psychometrika, 20, 173-192.

[5] MENTZ,R.P. (1972), *On the inverse of some covariance matrices of Toeplitz type*, Tech. Report N°8, Department of Statistics, Stanford University.

[6] MENTZ,R.P. (1975), *Asymptotic properties of some estimators in moving average models*, Forthcoming as Tech. Report, Dept. of Statistics, Stanford University.

[7] MUSTAFI,Ch.K. (1967), *The inverse of a certain matrix, with an application*, Annals of Math. Statistics, 38, 1289-1292.

[8] SHAMAN,P. (1969), *On the inverse of the covariance matrix of a first order moving average*, Biometrika, 56, 595-600.

[9] SHAMAN,P. (1973), *On the inverse of the covariance matrix for an autoregressive-moving average process*, Biometrika, 60, 193-196.

[10] TRENCH,W.F. (1964), *An algorithm for the inversion of finite Toeplitz matrices*, J. Soc. Indust. Appl. Math., 12, 515-522.

[11] UKITA,Y. (1955), *Characterization of 2-type diagonal matrices with an application to order statistics*, Journal of Hokkaido College of Art and Literature, 6, 66-75.

[12] WALKER,A.M. (1961), *Large sample estimation of parameters for moving average models*, Biometrika, 48, 343-357.

Instituto de Investigaciones Estadísticas

Universidad Nacional de Tucumán, Tucumán, Argentina

Recibido en Octubre de 1974

Versión final Julio de 1975

A note on inverting a covariance of toeplitz type

Raúl Pedro Mentz, University of Tucumán, Argentina

Bulletin of the International Statistical Institute
Volume XLVI, Book 4, 1975, pp.151-155

1. Introduction. A TxT matrix $S_T = (s_{ij})$ is called a Toeplitz matrix if $s_{ij}=s_{i-j}$. A particularly useful case for mathematical statistics is when S_T is symmetric and $s_{ij}=0$ for $|i-j|>m$. For m=1 the components s^{ij} of S_T^{-1} are derived by the method of cofactors, for example in Shaman (1969). In the present work we extend these results to m=2. For m=2 Shaman (1973) and Mentz (1972) have considered the problem, both using approaches different from the present one.

By factoring out the component of S in the main diagonal, we see that without loss of generality we have to deal with $S=I+r_1S_1+r_2S_2$, say, where $S_n=(s_{ij}^{(n)})$, $s_{ij}^{(n)}=1$ if $|i-j|=n$, and equals 0 otherwise, so that $S_0 = I$, the identity matrix.

No details will be given in this paper; these are presented in Mentz (1972).

2. The Inverse of $S=I+r_1S_1+r_2S_2$. Let S_n be of order n and S_n its determinant. Let

$$(2.1) \qquad L_n = \begin{pmatrix} E & S_{n-1} \\ 0 & p \end{pmatrix} \qquad K_n = \begin{pmatrix} r_1 & q \\ E & S_{n-1} \end{pmatrix},$$

where $p=(r_1,r_2,0,\ldots,0)'$, $E=(0,\ldots,0,r_2,r_1)$, $q=(r_2,0,\ldots,0)$, and let L_n and K_n be the corresponding determinants. By expanding S_n in terms of the components in its first row, Durbin (1959) found that the determinants satisfy the linear, homogeneous, fifth-order difference equation

$$(2.2) \quad -S_n+(1-r_2)S_{n-1}+(r_2-r_1^2)S_{n-2}+r_2(r_1^2-r_2)S_{n-3}+r_2^2(r_2-1)S_{n-4}+r_2^3S_{n-5}=0.$$

Hence S_n can be written as a linear combination of the roots of the polynomial equation associated with (2.2), the coefficients being determined from boundary conditions coming from the equations for

n=1,...,5, with the corresponding S_n evaluated explicitly. In fact by means of a simple transformation the polynomial equation can be made symmetric and practically reduced to second order. The same approach can be used with the L_n, because

$$L_n - r_1 L_{n-1} + r_2 L_{n-2} - r_1 r_2^2 L_{n-3} + r_2^4 L_{n-4} = 0, \tag{2.3}$$

while the K_n satisfy the simpler equation

$$K_n + r_2 K_{n-1} = r_1 S_{n-1}. \tag{2.4}$$

Proposition 1. Let $S_T = I + r_1 G_1 + r_2 G_2$ with $r_2 \neq 0$, and $S_T^{-1} = W_T = (w_{ij})$. Then

$$w_{ij} = (-1)^{i+j} B_{ij} / S_T, \qquad i = j, \ldots, T-j+1; \quad j = [T/2]+1, \ldots, T, \tag{2.5}$$

where

$$B_{ii} = S_{i-1} S_{T-i} - r_2^2 S_{i-2} S_{T-i-1}, \tag{2.6}$$

$$(-1)^{i+1} B_{ij} = S_{i-1} A_1(i,j) - r_2 S_{i-2} A_2(i,j) + r_2 K_{i-2} A_3(i,j), \tag{2.7}$$

and

$$A_1(i,j) = S_{T-j} L_{j-i} - r_2 K_{T-j} L_{j-i-1} + r_2^3 S_{T-j-1} L_{j-i-2}, \tag{2.8}$$

$$A_2(i,j) = S_{T-j}(r_1 L_{j-i-1} - r_2^2 L_{j-i-2}) - r_2 K_{T-j}(r_1 L_{j-i-2} - r_2^2 L_{j-i-3}) + r_2^3 S_{T-j-1}(r_1 L_{j-i-3} - r_2^2 L_{j-i-4}), \tag{2.9}$$

$$A_3(i,j) = r_2 S_{T-j} L_{j-i-1} - r_2^2 K_{T-j} L_{j-i-2} + r_2^4 S_{T-j-1} L_{j-i-3}. \tag{2.10}$$

The S_n, L_n and K_n satisfy (2.2), (2.3) and (2.4) respectively. The remaining elements of W_T are obtained by using $w_{ji} = w_{ij}$ (symmetry) and $w_{T-j+1,T-i+1} = w_{ij}$ (persymmetry).

The proof of this proposition consists basically in expanding the cofactors of the s_{ij} by Laplace's rule in terms of minors of the first i-1 columns. The given expressions are valid for all i and j provided we define $S_0 = L_0 = 1$, $K_0 = 0$, $S_{-n} = L_{-n} = K_{-n} = 0$ for $n > 0$.

3. Relation With the Case m=1. If $r_2 = 0$ but $r_1 \neq 0$, then $L_n = r_1^n$, $K_n = r_1 S_{n-1}$, $(-1)^{i+j} B_{ij} = S_{i-1} A_1(i,j)$, $A_1(i,j) = r_1^{j-i} S_{T-j}$, and the solution reduces to

$$w_{ij} = (-r_1)^{j-i} S_{i-1} S_{T-j} / S_T. \tag{3.1}$$

Here S_n satisfies the corresponding version of (2.2), namely $S_n - S_{n-1} + r_1^2 S_{n-2} = 0$, with boundary conditions $S_0 = 1$, $S_{-1} = 1 - r_1^2$. In this

form it is given, for example, by Shaman (1969).

4. Alternative Forms. We now present some forms of the w_{ij} that may be more useful than that in Proposition 1. Note however that (3.5) will eventually be written as function of the roots of the polynomial equations associated with (2.2), (2.3) and (2.4), and that this may be useful for analytic purposes.

Proposition 2. Let $A = (a_{ij})$ be TxT, symmetric and positive definite. A necessary and sufficient conditions that $A^{-1} = (a^{ij})$ satisfies $a^{ij} = 0$ for $|i-j| > r$ is that there exist constants b_{ts} such that for $t = 1, \ldots, T-r+1$,

$$a_{tt'} + b_{t1} a_{t+1,t'} + \ldots + b_{t,r-1} a_{t+r-1,t'} = 0, \qquad t' = t+1, \ldots, T. \tag{4.1}$$

This result apparently originated with Guttman (1955) and Ukita (1955), and a detailed proof is given in Mentz (1972). Condition (4.1) states the existence of a linear relationship between successive sets of r adjacent rows of A; an equivalent formulation (to be used below) is to relate the first r-1 rows to each of the remaining ones.

Proposition 3. Under the conditions of Proposition 1,

$$w_{ij} = \frac{(w_{2,T-1} w_{1T} - w_{2T} w_{i,T-1}) w_{T-j+1,T} - (w_{1,T-1} w_{1T} - w_{1T} w_{i,T-1}) w_{T-j+1,T-1}}{w_{1T} w_{2,T-1} - w_{1,T-1} w_{2T}}$$

$$= \frac{(w_{2,T-1} w_{1,T-i+1} - w_{1,T-1} w_{2,T-i+1}) w_{1j} - (w_{2T} w_{1,T-i+1} - w_{1T} w_{2,T-i+1}) w_{2j}}{w_{1T} w_{2,T-1} - w_{2T} w_{1,T-1}}, \tag{4.2}$$

$$i \leqslant j.$$

Proof. Since $a_{ij} = 0$ for $|i-j| > 3$, by Proposition 2 there exist constants t_{1i} and t_{2i}, such that $w_{ij} = t_{1i} w_{1j} + t_{2i} w_{2j}$, $i = 3, \ldots, T$; $j \geqslant i$. Using these relations for $j = T-1, T$, solving the resulting systems for the t_{1i} and t_{2i}, and replacing back the values, one obtains (4.2). Q.E.D.

Proposition 4. Under the conditions of Proposition 1,

$$w_{ij} = \frac{(w_{i+2,i+3} w_{i,i+2} - w_{i+2,i+2} w_{i,i+3}) w_{i+1,j} - (w_{i+1,i+3} w_{i,i+2} - w_{i+1,i+2} w_{i,i+3}) w_{i+2,j}}{w_{i+1,i+2} w_{i+2,i+3} - w_{i+1,i+3} w_{i+2,i+2}}. \tag{4.3}$$

Proof. The proof is similar to that of Proposition 3, only that we

now use $w_{ij} - b_{i1}w_{i+1,j} - b_{i2}w_{i+2,j}$, which is (4.1). Q.E.D.

Expression (4.2) exhibits the w_{ij} as functions of the w_{1r} and $w_{i,r-1}$ (or w_{1j} and w_{2j}). These in turn can be evaluated directly from Proposition 1.

Expression (4.3) may be computationally simpler than (4.2). We first compute $w_{i,i+c}$ for c=0,1,2,3, directly from Proposition 1, taking advantage that for these cases the L_j determinants that are needed are of low order, and can be evaluated very easily. Then the $w_{i,i+c}$ are functions only of the S_n and H_n determinants, and the latter are determined once the S_n are available. With these components along the central diagonal of W_n, (4.3) can be used recursively to find the other w_{ij}.

5. Acknowledgements. The main part of this study is part of the author's Ph.D. dissertation in the Department of Statistics, Stanford University. Appreciation is expressed to Professor T.W. Anderson for generous guidance and help while directing this work, and to Professor Paul Shaman for corrections to a first draft. Work at Stanford was supported by a Research Contract with the Office of Naval Research (T.W. Anderson, Project Director).

6. References

Durbin, J. (1959). "Efficient estimation of parameters in moving average models". Biometrika, 46, 306-316.

Guttman, L. (1955). "A generalized simplex for factor analysis". Psychometrika, 20, 173-192.

Mentz, R.P. (1972). "On the inverse of some covariance matrices of Toeplitz type". Tech.Rep. N° 8, Dep. of Stat., Stanford Univ.

Shaman, P. (1969). "On the inverse of the covariance matrix of a first order moving average". Biometrika, 56, 595-600.

Shaman, P. (1973). "On the inverse of the covariance matrix for an autoregressive-moving average process". Biometrika, 60, 193-196.

White, J. (1955). "Characterization of 2-type diagonal matrices with an application to order statistics". Journal of Hokkaido College of Art and Literature, 6, 66-75.

7. Summary. A matrix is said to be of Toeplitz type if it has equal elements along diagonals. These matrices, with the additional property of symmetry, arise frequently in statistical work, as covariance matrices of stationary stochastic processes, in nonparametric theory, etc. The inverse is often of interest, and a method is used to find its components when the T x T matrix has only 5 nonvanishing central diagonals. The method is to express the cofactors of the components of the given matrix in terms of some determinants, that are shown to satisfy certain linear difference equations, and to solve these explicitly. The complexity of the resulting expressions for the components of the inverse is comparable to those known in the literature.

8. Résumé. Une matrice est citée comme appartenant au type Toeplitz quand elle a des éléments égaux le long des diagonales. Ces matrices, qui possèdent aussi la propriété spéciale de la symétrie, se trouvent souvent dans les oeuvres statistiques comme des covariances de processus stochastiques stationnaires, dans la théorie non paramétrique, etc. L'inverse est fréquemment intéressante et on utilise une méthode pour en trouver ses composants lorsque la matrice T x T a seulement 5 diagonales centrales non nulles. La méthode consiste à énoncer les mineurs des composants de la matrice donnée en fonction de quelques déterminants. On démontre de ceux-ci qu'ils satisfont certaines équations linéaires aux différences finies et on les résout formellement. La complexité des expressions résultantes pour les composants de l'inverse peut-être comparée à la qu'on connaît déjà des oeuvres techniques.

On the inverse of some covariance matrices of toeplitz type

Raúl Pedro Mentz

Abstract. A matrix is said to be of Toeplitz type if it has equal elements along diagonals. These matrices, with the additional property of symmetry, arise frequently in statistical work, as covariance matrices of wide-sense stationary stochastic processes, in nonparametric theory, etc. The inverse is often of interest, and a method is developed to find its components in close form when the $T \times T$ matrix has only $2m + 1$ nonvanishing (central) diagonals ($1 \leqq m < T$). The method consists in posing difference equations for the components of the inverse, and solving them explicitly. The resulting procedures reproduce known results when $m = 1$, provide an expression for the inverse when $m = 2$, provide approximations in these two important cases and an interpretation for another approximation for general m, and are shown to be particularly suited for two methods of estimation in moving average models of time series. The inverse of a related matrix is also studied.

1. Introduction. A $T \times T$ matrix $\Sigma = (\sigma_{ij})$ is called a *Toeplitz matrix* if $\sigma_{ij} = \sigma_{i-j}$. A particular case is when $\sigma_{ij} = \sigma_{|i-j|}$. A useful notation for this second case is

$$\Sigma = \sum_{k=0}^{T-1} \sigma_k \mathbf{G}_k = \sigma_0 \sum_{k=0}^{T-1} \rho_k \mathbf{G}_k, \tag{1.1}$$

where $\sigma_0 \neq 0$, $\rho_k = \sigma_k/\sigma_0$, and $\mathbf{G}_k$ has components $g_{ij}^{(k)}$ equal to one when $|i - j| = k$, and equal to zero otherwise. Without loss of generality we take $\rho_0 = 1$. A particularly useful case for statistics is when further $\rho_{|i-j|} = 0$ for $|i - j| > m$, $1 \leqq m < T$; the covariance matrices with these components correspond to stochastic processes "finitely correlated of order m" in the theory of wide-sense stationary stochastic processes, and to several other important cases. Note that (1.1) exhibits Σ as a linear combination of known matrices, the coefficients being the (parameters) σ_k or ρ_k. For a solution of a similar problem with different $\mathbf{G}_j$, see Mustafi (1967).

The inversion of Toeplitz matrices has been approached in the literature as a computational problem, where the purpose is to find numerical procedures that operate faster than general inversion procedures, or as a mathematical problem of finding the components of the inverse in explicit form. Of course a solution to the second problem is also a solution to the first one. This paper is devoted to the second objective, and no attempt will be made to evaluate the computational merits of the proposals.

* Received by the editors August 16, 1974.

† Institute of Statistics, University of Tucumán, Tucumán, Argentina. This work is part of the author's doctoral dissertation in the Department of Statistics, Stanford University. This work was supported in part by the Office of Naval Research Contract N000-4-67-A-0112-0030 (NR-042-34) at Stanford University, by the Ford Foundation, and by the University of Tucumán.

The inverses are given in terms of the roots of some polynomial equations involving the ρ_k. This is often convenient for theoretical work, because in many important cases it is possible to prove helpful properties of the roots. One example is when they are less than one in absolute value.

2. The inverse by solving difference equations. In case $\rho_k = 0$ for $|k| > m$, (1.1) becomes $\Sigma = \sum_{k=0}^{m} \rho_k \mathbf{G}_k$, with $\rho_0 = 1$, $\rho_m \neq 0$, $1 \leqq m < T$. If we denote the inverse matrix by $\mathbf{W} = (w_{ij})$, the defining equation for the inverse, $\Sigma\mathbf{W} = \mathbf{I}$, written in terms of components is

$$
\begin{aligned}
\delta_{1j} &= w_{1j} + \rho_1 w_{2j} + \cdots + \rho_m w_{m+1,j} \qquad (i = 1),\\
&\vdots\\
\delta_{mj} &= \rho_{m-1} w_{1j} + \cdots + \rho_1 w_{m-1,j} + w_{mj} + \rho_1 w_{m+1,j}\\
&\quad + \cdots + \rho_m w_{2m,j} \qquad (i = m),\\
\delta_{ij} &= \rho_m w_{i-m,j} + \cdots + \rho_1 w_{i-1,j} + w_{ij} + \rho_1 w_{i+1,j}\\
&\quad + \cdots + \rho_m w_{i+m,j}, \qquad i = m+1, \cdots, T-m,\\
\delta_{T-m+1,j} &= \rho_m w_{T-2m+1,j} + \cdots + \rho_1 w_{T-m,j} + w_{T-m+1,j} + \rho_1 w_{T-m+2,j}\\
&\quad + \cdots + \rho_{m-1} w_{Tj} \qquad i = T-m+1,\\
&\vdots\\
\delta_{Tj} &= \rho_m w_{T-m,j} + \cdots + \rho_1 w_{T-1,j} + w_{Tj} \qquad (i = T),
\end{aligned}
\tag{2.1}
$$

where δ_{ij} is Kronecker's delta function.

This system of linear difference equations will be solved for each j (fixed column of $\mathbf{W}$) and $i \leqq j$, that is, for components above and on the main diagonal.

The linear difference equation of order $2m$ corresponding to rows $m+1, \cdots, T-m$, in the homogeneous case, has the associated characteristic equation

$$
\begin{aligned}
0 &= \rho_m z^{2m} + \rho_{m-1} z^{2m-1} + \cdots + \rho_m\\
&= \rho_m(z^{2m} + 1) + \cdots + \rho_1(z^{m+1} + z^{m-1}) + z^m,
\end{aligned}
\tag{2.2}
$$

which is symmetric in its coefficients. Dividing through by ρ_m, it appears that the roots occur in pairs, z_j and z_j^{-1} ($z_j \neq 0$), and hence (2.2) is equivalent to

$$
0 = \prod_{j=1}^{2m} (z - z_j) = \prod_{j=1}^{m} (z - z_j)(z - z_j^{-1}) = \prod_{j=1}^{m} (z^2 - z\,d_j + 1),
\tag{2.3}
$$

where $d_j = z_j + z_j^{-1}$. Equating coefficients one is led to a system of nonlinear equations for the d_j. When $m = 2$ they are

$$
-\rho_1\rho_2^{-1} = d_1 + d_2, \qquad \rho_2^{-1} = 2 + d_1 d_2,
\tag{2.4}
$$

and when $m = 3$,

$$
\begin{aligned}
-\rho_2\rho_3^{-1} &= d_1 + d_2 + d_3, \quad \rho_1\rho_3^{-1} = 3 + d_1 d_2 + d_1 d_3 + d_2 d_3,\\
-\rho_3^{-1} &= 2d_1 + 2d_2 + 2d_3 + d_1 d_2 d_3.
\end{aligned}
\tag{2.5}
$$

These systems of m equations, and $z_j = \frac{1}{2}(d_j \pm \sqrt{d_j^2 - 4})$ provide the $2m$ roots of (2.2). In fact the reduction to m equations can be achieved preserving the polynomial form. The procedure has been given in detail in Anderson (1971b), and consists in writing (2.2) as

$$
\begin{aligned}
0 &= \rho_m^{-1} z^m + \sum_{h=1}^{m} \rho_h \rho_m^{-1} (z^{m+h} + z^{m-h})\\
&= \rho_m^{-1} z^m + z^m \sum_{h=1}^{m} \rho_h \rho_m^{-1} z^{-h}(z^{2h} + 1),
\end{aligned}
\tag{2.6}
$$

and then using the substitution $zs = z^2 + 1$.

When the roots $z_1, \cdots, z_m$ are distinct, then

$$
w_{ij} = \sum_{h=1}^{m} \{C_h(j) z_h^i + C_h'(j) z_h^{-i}\}, \qquad i \leqq j.
\tag{2.7}
$$

The form will have to be altered if some of the roots coincide. The $2m$ constants in (2.7) can be evaluated from boundary conditions extracted from (2.1). In particular, for columns $1, 2, \cdots, m$ and $T-m+1, \cdots, T$, the boundary conditions are just the first m and last m equations in (2.1), and one of the resulting linear systems determines the constants for the m columns.

To determine the rest of the components we use the symmetry of $\mathbf{W}$ with respect to its two main diagonals (inherited from that of Σ), that leads to

$$
w_{ij} = w_{ji} = w_{T-i+1,T-j+1}, \qquad i, j = 1, 2, \cdots, T,
\tag{2.8}
$$

and one of the following procedures:

PROPOSITION 2.1. *Let Σ be of order $T \times T$, and given by*

$$
\Sigma = \sum_{k=0}^{m} \rho_k \mathbf{G}_k,
\tag{2.9}
$$

with $\rho_0 = 1$, $\rho_m \neq 0$, $1 \leqq m < T$, and let $\Sigma^{-1} = \mathbf{W} = (w_{ij})$. Then the following four steps determine the w_{ij}:

Step 1. Find columns $T-m+1, \cdots, T$ of $\mathbf{W}$ from (2.7) or a similar expression, depending upon the nature of the roots of (2.2), where the $2m$ constants are evaluated from boundary conditions provided by the first m and last m equations in (2.1);

Step 2. Using (2.8) deduce rows $1, 2, \cdots, m$ from columns $T-m+1, \cdots, T$;

Step 3. Determine the proportionality constants θ_{st} in

$$
\theta_{1i} w_{1j} + \theta_{2i} w_{2j} + \cdots + \theta_{mi} w_{mj} = w_{ij}, \qquad i = m+1, \cdots, T, \quad j \geqq i,
\tag{2.10}
$$

(for columns $j = [T/2] + 1, \cdots, T-m$) from columns $T-m+1, \cdots, T$;

Step 4. Using repeatedly (2.10) find those components w_{ij} satisfying $i = j, \cdots, T-j+1$; $j = [T/2]+1, \cdots, T-m$, and all others using (2.8).

Alternatively one can proceed column by column, as follows:

PROPOSITION 2.2. *Under the same hypotheses of Proposition 2.1, the w_{ij} for $i \leqq j$ are given by (2.7) or a similar expression, depending upon the nature of the roots of (2.2). The $2m$ constants are evaluated from boundary conditions extracted from the system (2.1) as follows: For columns $T, T-1, \cdots, T-m+1$, the*

conditions are the first m and last m equations of (2.1). *For any column* $j = T - m, \cdots, [T/2] + 1$, *we assume that columns* $j + 1, \cdots, j + m$ *are already available, and then the conditions are the first m equations of* (2.1) *plus equations for rows* $j - m + 1, \cdots, j$ *with the substitutions*

$$\begin{aligned} w_{j+1,j} &= w_{j,j+1} \equiv w^*_{1j}, \\ w_{j+2,j} &= w_{j,j+2} \equiv w^*_{2j}, \\ &\vdots \\ w_{j+m,j} &= w_{j,j+m} \equiv w^*_{mj}; \end{aligned} \tag{2.11}$$

the remaining columns can be obtained using (2.8).

To prove Proposition 2.1 one only notes that (2.10) is equivalent to the condition of Proposition A.1 in Appendix A. To prove Proposition 2.2, we note that for column j, $[T/2] + 1 \leqq j \leqq T - m$, the complete system of equations is composed of the first m in (2.1), plus the homogeneous equations

$$w_{ij} + \sum_{h=1}^{m} \rho_h (w_{i-h,j} + w_{i+h,j}) = 0, \qquad i = m+1, m+2, \cdots, j-m, \tag{2.12}$$

plus equations

$$\begin{aligned} 0 &= \rho_m w_{j-2m+1,j} + \cdots + \rho_1 w_{j-m,j} + w_{j-m+1,j} + \rho_1 w_{j-m+2,j} \\ &\quad + \cdots + \rho_m w_{j+1,j} \qquad (i = j - m + 1), \\ &\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ 0 &= \rho_m w_{j-m-1,j} + \cdots + \rho_1 w_{j-2,j} + w_{j-1,j} + \rho_1 w_{jj} \\ &\quad + \cdots + \rho_m w_{j+m-1,j} \qquad (i = j - 1), \\ 1 &= \rho_m w_{j-m,j} + \cdots + \rho_1 w_{j-1,j} + w_{jj} + \rho_1 w_{j+1,j} \\ &\quad + \cdots + \rho_m w_{j+m,j} \qquad (i = j), \end{aligned} \tag{2.13}$$

plus other homogeneous equations for rows $i > j$. The sequence satisfying (2.12) is w_{ij}, $i = 1, 2, \cdots, j$, so that the components in the left-hand sides of (2.11) have to be obtained from the columns already determined.

Note 1. To avoid vacuous statements, T is taken to be large compared with m.

Note 2. In some instances, the procedures described in Propositions 2.1 or 2.2, or both, will disclose an explicit general expression for all w_{ij}, as will be exemplified below.

Note 3. Consider the case where $\rho_0 = \rho_0^* \neq 1$ for rows $i = m + 1, \cdots, T$. Then (2.1) will be different only in the (complete) equations corresponding to those rows, and the roots entering in the expression for w_{ij} will be evaluated from (2.2) with the coefficient of z^m replaced by ρ_0^*. Otherwise the procedure is still valid, and so are the forms of the solution, as is made clear by the procedure of Proposition 2.1.

3. Applications. A detailed discussion of applications (a), (b) and (c) below, appears in Mentz (1972).

(a) *The case* $m = 1$, *and some approximations.* In case $m = 1$, either procedure

of section 2 can be used to reproduce some known results. Here x_1 can be taken to be $\frac{1}{2}\rho^{-1}(-1 + \sqrt{1 - 4\rho^2})$. For positive definite matrices we require that $|\rho| < \frac{1}{2}$, and hence $|x_1| < 1$. After deriving columns T, $T - 1$, etc. in succession, the form of column $T - s$ becomes evident, and can be verified by induction, all using Proposition 2.2. This leads to the form

$$w_{ij} = \frac{(1 - x_1^{2T-2j+2})(x_1^{j+1+i} - x_1^{j+1-i})}{\rho(1 - x_1^2)(1 - x_1^{2T+2})}, \qquad i \leqq j, \tag{3.1}$$

which is a new version of the result given by Shaman (1968). This form suggests the following approximations, when T is large and $i \leqq j$:

$$w_{ij} \cong \frac{(1 - x_1^{2T-2j+2})(x_1^{j+1+i} - x_1^{j+1-i})}{\rho(1 - x_1^2)}, \qquad \text{any } j; \tag{3.2}$$

$$w_{ij} \cong \frac{x_1^{j+1+i} - x_1^{j+1-i}}{\rho(1 - x_1^2)}, \qquad j \text{ small}; \tag{3.3}$$

$$w_{ij} \cong -\frac{(1 - x_1^{2T-2j+2})x_1^{j+1-i}}{\rho(1 - x_1^2)}, \qquad j \text{ large}. \tag{3.4}$$

These approximations are based on the fact that $|x_1| < 1$. In particular, under the same hypotheses, $w_{i1} \cong -x_1^i/\rho$, $w_{iT} \cong -x_1^{T+1-i}/\rho$. Shaman gave the approximation

$$w_{ij} \cong \frac{x_1^{j-1-i}}{x_1(1 + 2\rho x_1)}, \qquad \text{any } j, \tag{3.5}$$

based on a different argument.

The form (3.1) (or slight variations of it) can also be found directly by application of Proposition 2.1.

(b) *The case* $m = 2$. The procedure of Proposition 2.1 can be used in this case. Let w_{iT} and $w_{i,T-1}$ be determined for all i. Relation (2.10) now reads $\theta_{1i} w_{1j} + \theta_{2i} w_{2j} = w_{ij}$, for $j = T - 1, T$, and $i = 3, 4, \cdots, T$; solving for θ_{1i} and θ_{2i}, and using the relation for $j = 3, 4, \cdots, T - 2$ and $i \leqq j$, one obtains

$$w_{ij} = \frac{(w_{2,T-1}w_{iT} - w_{2T}w_{i,T-1})w_{T-j+1,T} - (w_{1,T-1}w_{iT} - w_{1T}w_{i,T-1})w_{T-j+1,T-1}}{w_{1T}w_{2,T-1} - w_{1,T-1}w_{2T}}, \qquad i \leqq j. \tag{3.6}$$

This expression is also valid for $i = 1, 2$ and $j = 1, 2, T - 1, T$.

The particular form of w_{iT} and $w_{i,T-1}$ depends on the nature of the roots of the polynomial equation $\rho_1 x^4 + \rho_2 x^3 + x^2 + \rho_1 x + \rho_2 = 0$. If the roots are x_1 and x_1^{-1} each with multiplicity two, then

$$w_{is} = \{C_1(s) + iC_2(s)\}x_1^i + \{C_3(s) + iC_4(s)\}x_1^{-i}, \qquad \begin{matrix} s = T-1, T, \\ i = 1, 2, \cdots, T; \end{matrix} \tag{3.7}$$

if the roots are $x_1, x_1^{-1}, x_2, x_2^{-1}$ and $x_1 \neq x_2, x_1 \neq x_2^{-1}$, then

$$w_{is} = C_1^*(s)x_1^i + C_2^*(s)x_1^{-i} + C_3^*(s)x_2^i + C_4^*(s)x_2^{-i}, \qquad \begin{matrix} s = T-1, T, \\ i = 1, 2, \cdots, T. \end{matrix} \tag{3.8}$$

For the case of (3.7), in Appendix B it is shown that

$$w_{is} = m_s(i)x_1^{T+i} + n_s(i)x_1^{T-i}; \tag{3.9}$$

an approximation is also discussed. The case of (3.8) can be treated in a parallel form, and we omit those details here.

(c) *Estimation in the moving average model.* For the moving average model of time series analysis,

$$y_t = \alpha_0\varepsilon_t + \alpha_1\varepsilon_{t-1} + \cdots + \alpha_q\varepsilon_{t-q}, \tag{3.10}$$

where the random variables ε_t satisfy $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = 1$, $E(\varepsilon_t\varepsilon_s) = 0$ if $t \neq s$, there exists interest in estimating the autocorrelations $\rho(s) = \sigma_{MA}(s)/\sigma_{MA}(0)$, where

$$\sigma_{MA}(s) = \sigma_{MA}(-s) = \begin{cases} \sum_{j=0}^{q-|s|} \alpha_j\alpha_{j+|s|}, & |s| \leqq q, \\ 0, & |s| > q. \end{cases} \tag{3.11}$$

Among other things, this gives a way to estimate the α_j's. On the basis of a sample $y_1, \cdots, y_T$ from (3.10) Walker (1961) proposed an estimator for $\rho' = (\rho(1), \cdots, \rho(q))$ in terms of k $(k > q)$ sample autocorrelations defined by $r_s = C_s/C_0$, $C_s = \sum_{t=1}^{T-s} y_t y_{t-s}/(T-s)$, $s = 0, 1, \cdots, k$. The estimator can be written

$$\hat{\rho} = \mathbf{r}^{(1)} - \mathbf{W}_{12}(\mathbf{r}^{(1)})\mathbf{W}_{22}^{-1}(\mathbf{r}^{(1)})\mathbf{r}^{(2)}, \tag{3.12}$$

where $\mathbf{r}' = (r_1, \cdots, r_k) = (\mathbf{r}^{(1)}, \mathbf{r}^{(2)})$, and $\mathbf{r}^{(1)}$ has q components. $\mathbf{W}(\rho)$ is the covariance matrix of the limiting normal distribution of $\sqrt{T}(\mathbf{r}^{(1)} - \boldsymbol{\rho})$, and is partitioned to conform with $\mathbf{r}$; see also Anderson (1971a, § 5.7.2). Then (3.12) involves the first $2q$ rows, or the last $2q$ columns, of $\mathbf{W}_{22}^{-1}(\rho)$ evaluated at $\rho = \mathbf{r}^{(1)}$. $\mathbf{W}_{22}$ has equal elements along its diagonals, and they vanish if their indices differ by more than $2q$.

Hence, the procedures of § 2 are suited to evaluate the components of the last $2q$ columns of $\mathbf{W}_{22}^{-1}(\mathbf{r}^{(1)})$ explicitly. This was done in detail for the case $q = 1$ in Mentz (1975), to study the asymptotic properties of $\hat{\rho}$ when $k = k(T)$, a function of T such that $\lim_{T\to\infty} k(T) = \infty$. The form (3.9) arose in this context, with $\rho_1 = 2\rho(1 + 2\rho^2)^{-1}$, $\rho_2 = \rho^2(1 + 2\rho^2)^{-1}$.

A proposal to estimate the α_j in (3.10) by Durbin (1959), is based on approximating (3.10) by a finite autoregression of order k larger than q. Then the vector $\boldsymbol{\beta}$ of coefficients of this autoregression is estimated from the least squares equation $\mathbf{C}\boldsymbol{\beta} = \mathbf{c}$, where $\mathbf{C} = (C_{|i-j|})$, $\mathbf{c} = (C_1, C_2, \cdots, C_k)'$, and the C_j are the sample autocovariances defined above. $\mathbf{C}$ is a $k \times k$ Toeplitz matrix and in principle can be evaluated explicitly by the methods of § 2.

A modification suggested by Anderson (1971b) is to take the components of $\mathbf{C}$ equal to 0 when their indices differ by more than q; the resulting matrix estimates

$(\sigma_{MA}(i-j))$ consistently, as does $\mathbf{C}$, because $\sigma_{MA}(i-j) = 0$ for $|i-j| > q$. For this procedure, the methods of § 2 become more appropriate, and for example the explicit forms (3.1) and (3.6) and their approximations can be used when $q = 1$ and 2, respectively. Some details for $q = 1$ are worked out in Mentz (1975).

(d) *Interpreting a well-known approximation.* If $\Sigma = (\sigma_{|i-j|})$, $\sigma_{|i-j|} = 0$ for $|i-j| > q$, and Σ is positive definite for every T, there exist coefficients $\alpha_0, \cdots, \alpha_q$ such that $\Sigma = \Sigma_{MA}$, the covariance matrix of a vector $(y_1, \cdots, y_T)$ generated by the moving average model (3.10); see for example Anderson (1971a, § 5.7.1). This way of looking at Σ gives rise to a different approximation for Σ^{-1} that is frequently used, at least for small values of m, and that we now discuss in the general case. It consists in that $\Sigma_{MA}^{-1} \cong \Sigma_{AR}$, where Σ_{AR} is the covariance matrix of $(x_1, \cdots, x_T)$ generated by the autoregression

$$\alpha_0 x_t + \alpha_1 x_{t-1} + \cdots + \alpha_q x_{t-q} = \varepsilon_t, \qquad t = \cdots, -1, 0, 1, \cdots. \tag{3.13}$$

The merits of this approximation are usually judged by comparing Σ_{MA} with Σ_{AR}^{-1}. These matrices differ in the components in two $q \times q$ submatrices located at both endpoints of the main diagonal. See e.g. Wise (1955).

The form (2.7) in case all roots are different, permits the direct comparison of the components of Σ_{MA}^{-1} with those of Σ_{AR}. In effect, the components of Σ_{AR} are

$$\sigma_{AR}(i, j) = \sum_{s=1}^{q} K_s^* x_s^{j-i} = \sum_{s=1}^{q} K_s(j)x_s^{-i}, \qquad j \geqq i, \tag{3.14}$$

where the x_s are the (distinct) roots of $M(x) = \alpha_0 x^q + \cdots + \alpha_q = 0$ (see for example Anderson (1971a, § 5.2.2)), while under the model of this section (2.7) becomes

$$w_{ij} = \frac{1}{\sigma_{MA}(0)} \sum_{s=1}^{q} \{C_s(j)z_s^i + C_s(j)z_s^{-i}\}, \qquad j \geqq i. \tag{3.15}$$

If in (3.14) we choose all roots x_s to be less than one in absolute value, which without loss of generality we can always do when dealing with second order moments, it then follows that the approximation (3.14) consists of omitting from the exact expression (3.15) the part involving positive powers of those roots. To prove this, note that (2.2) can be written $0 = z^m \sum_{h=-m}^{m} \sigma_{MA}(h)z^h$, because $\sigma_{MA}(h) = \sigma_{MA}(-h)$. But $\sum_{h=-m}^{m} \sigma_{MA}(h)z^h = M(z)M(z^{-1})$, by using (3.11).

The constants in (3.14) and (3.15) are related in a similar fashion: the m equations defining the $\sigma_{AR}(h)$, the so-called Yule–Walker equations, bear to those in (2.13) to be used for columns not near the endpoints of the range of j values, the same type of relation that $M(z)$ bears to $M(z)M(z^{-1})$: further the m boundary conditions provided by rows $1, 2, \cdots, m$ of (2.1), constitute an "end effect" that is neglected as part of the approximating process.

As an illustration, when $q = 1$, $\alpha_0 = 1$, and $\alpha_1 = \alpha$, we have that $\rho = \alpha/(1 + \alpha^2)$, and x_1 of application (a) becomes $x_1 = -\alpha$. Then we can write (3.1) as

$$w_{ij} = -\frac{x_1^j}{1 - x_1^2} \frac{1 - x_1^{2T+2-2j}}{1 - x_1^{2T+2}}(x_1^i - x_1^{-i}), \qquad j \geqq i. \tag{3.16}$$

For the first order autoregression $y_t + \alpha y_{t-1} = \varepsilon_t$, the covariance sequence is $\sigma_{AR}(h) = (-\alpha)^h/(1-\alpha^2)$, so that

$$w_{ij} = \frac{(-\alpha)^{j-i}}{1-\alpha^2} \cong \frac{x_1^j}{1-x_1^2} x_1^{-i}, \qquad j \geqq i. \tag{3.17}$$

Hence in this case it is verified that the term containing x_1^j is neglected, and the "end effect" consists in taking $(1 - x_1^{2T+2-2j})/(1 - x_1^{2T+2})$ as approximately equal to one in the coefficient of x_1^{-i}.

4. A brief review of the literature. Several authors gave forms of the w_{ij} in close form for the case $m = 1$. Shaman (1968), (1969) provided several alternatives, one of which is (3.1). For all values of ρ that make the matrix invertible Lovass-Nagy and Powers (1969) approached the problem through the determination of the characteristic roots and vectors of Σ, and found that

$$w_{ij} = \frac{2}{T+1} \sum_{k=1}^{T} \frac{[\sin(ik\pi/(T+1))][\sin(jk\pi/(T+1))]}{1 + \rho\cos(\pi k/(T+1))}, \qquad \text{all } i \text{ and } j. \tag{4.1}$$

For $|\rho| \geqq \frac{1}{2}$ Kershaw (1969) found that

$$w_{ij} = -\frac{1}{\rho} \frac{U_{i-1}(-1/(2\rho)) U_{T-j}(-1/(2\rho))}{U_T(-1/(2\rho))}, \qquad j \geqq i, \tag{4.2}$$

where $U_r(t)$ is a Chebyshev polynomial of the second kind in t. The method of proof he used is somewhat similar to the one in our § 2.

For the case $m = 2$, Shaman (1971) found expressions for the components of the inverse matrix, based on the association with a moving average process. His answer is in terms of ρ_1 and ρ_2 in our notation, and the level of (algebraic) complexity is somewhat comparable to that involved in (3.6).

Among computing algorithms (for general m), see the papers by Trench (1964), (1967). In the mathematical literature there exists interest in inverting infinite Toeplitz matrices; see for example the work by Calderón, Spitzer and Widom (1959).

Appendix A. On a necessary and sufficient condition for a covariance matrix to have an inverse of diagonal type. We now state and prove in detail a result that appears to have originated with Guttman (1955) and Ukita (1955). The conditions were used by Greenberg and Sarhan (1959), who call them sufficient conditions.

A matrix (a_{ij}) is said to be "diagonal of type r" if $a_{ij} = 0$ whenever $|i-j| \geq r$. From § 1 it follows that a stochastic process which is finitely correlated of order m, gives rise to covariance matrices which are diagonal of type $r = m + 1$.

PROPOSITION A.1. *Let $\Sigma = (\sigma_{ij})$ be a $T \times T$ symmetric and positive definite matrix. A necessary and sufficient condition that $\Sigma^{-1} = (w_{ij})$ be diagonal of type r is that there exist constants b_{ts} such that for $t = 1, 2, \cdots, T - r + 1$,*

$$\sigma_{tt'} + b_{t1}\sigma_{t+1,t'} + \cdots + b_{t,r-1}\sigma_{t+r-1,t'} = 0, \qquad t' = t+1, \cdots, T. \tag{A.1}$$

Proof. For the necessity of the condition, suppose that $\mathbf{y} = (y_1, \cdots, y_T)'$ is a vector of jointly distributed random variables with $E\mathbf{y} = \mathbf{0}$ and covariance matrix Σ such that Σ^{-1} is diagonal of type r. There exists an upper triangular matrix $\mathbf{B} = (b_{ij})$ with $b_{ii} = 1$, and a diagonal matrix $\mathbf{D} = (d_{ij})$ with $d_{ii} > 0$, such that $\Sigma^{-1} = \mathbf{B}'\mathbf{D}\mathbf{B}$. It is readily verified that $\mathbf{B}$ is "upper triangular of type r", that is, $b_{ij} = 0$ if $j - i \geqq r$. Let us introduce the new random vector $\mathbf{u} = (u_1, \cdots, u_T)'$ defined by $\mathbf{u} = \mathbf{B}\mathbf{y}$. Then $E\mathbf{u} = \mathbf{0}$, $E\mathbf{u}\mathbf{u}' = \mathbf{B}\Sigma\mathbf{B}' = \mathbf{B}\mathbf{B}^{-1}\mathbf{D}\mathbf{B}'^{-1}\mathbf{B}' = \mathbf{D}^{-1}$, and the u_t's are uncorrelated. Note that

$$u_t = y_t + b_{t1}y_{t+1} + \cdots + b_{t,r-1}y_{t+r-1}, \qquad t = 1, 2, \cdots, T - r + 1. \tag{A.2}$$

Solving $\mathbf{u} = \mathbf{B}\mathbf{y}$ for $\mathbf{y}$, we note that $\mathbf{B}^{-1} = (b^{ij})$ is also upper triangular, and hence

$$y_t = \sum_{s=t}^{T} b^{ts} u_s, \qquad t = 1, 2, \cdots, T. \tag{A.3}$$

Multiplying (A.2) by (A.3) with t replaced by t', $t' = t+1, \cdots, T$, and taking expected values we obtain (A.1).

Sufficiency. We have to show that if (A.1) holds, $w_{ij} = 0$ for $|i-j| \geqq r$. Consider $i > j$, that is, w_{ij} is below the main diagonal. The $(T-1) \times (T-1)$ cofactor of σ_{ij} will be evaluated by Laplace's expansion using minors formed by its last $T - r - j$ columns. For any one of these minors, if its complementary minor does not include row 1 of Σ, then the first r columns of the latter are linearly dependent. If row 1 of Σ is included, by using relation (A.1) successively and at most $j - 1$ times the complementary minor can be brought into the equivalent form

$$\begin{vmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{0} & \mathbf{M}_{22} \end{vmatrix}, \tag{A.4}$$

where $\mathbf{M}_{11}$ is upper triangular and the first r columns of $\mathbf{M}_{22}$ are linearly dependent. In either case the complementary minor is 0, and hence $w_{ij} = 0$.

Appendix B. Explicit forms of w_{1T} and $w_{1,T-1}$ when $m = 2$. We consider the details of the determination of $C_k(s)$, $k = 1, 2, 3, 4$, $s = T-1, T$, in (3.7). For $s = T$ the boundary conditions are

$$\begin{aligned}
0 &= w_{1T} + \rho_1 w_{2T} + \rho_2 w_{3T}, \\
0 &= \rho_1 w_{1T} + w_{2T} + \rho_1 w_{3T} + \rho_2 w_{4T}, \\
0 &= \rho_2 w_{T-3,T} + \rho_1 w_{T-2,T} + w_{T-1,T} + \rho_1 w_{TT}, \\
1 &= \rho_2 w_{T-2,T} + \rho_1 w_{T-1,T} + w_{TT}.
\end{aligned} \tag{B.1}$$

Replacing the w_{iT} by their form (3.7), one is led to the linear system

$$\text{(B.2)}\qquad \mathbf{C}(T) \equiv \begin{bmatrix} C_1(T) \\ C_2(T) \\ C_3(T) \\ C_4(T) \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} & \begin{bmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{bmatrix} \\ x_1^T \begin{bmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} & x_1^{-T} \begin{bmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{bmatrix} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$
$$\equiv \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ x_1^T \mathbf{A}_{21} & x_1^{-T} \mathbf{A}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{u} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{B}_{12}\mathbf{u} \\ \mathbf{B}_{22}\mathbf{u} \end{bmatrix},$$

where $\mathbf{u} = (0, 1)'$ and the introduction of the $\mathbf{A}_{ij}$ and $\mathbf{B}_{ij}$ is self-explanatory. Here

$$\text{(B.3)}\qquad \begin{aligned}
a_{11} &= x_1 + \rho_1 x_1^2 + \rho_2 x_1^3, \\
a_{12} &= x_1 + 2\rho_1 x_1^2 + 3\rho_2 x_1^3, \\
a_{13} &= x_1^{-1} + \rho_1 x_1^{-2} + \rho_2 x_1^{-3}, \\
a_{14} &= x_1^{-1} + 2\rho_1 x_1^{-2} + 3\rho_2 x_1^{-3}, \\
a_{21} &= \rho_1 x_1 + x_1^2 + \rho_1 x_1^3 + \rho_2 x_1^4, \\
a_{22} &= \rho_1 x_1 + 2x_1^2 + 3\rho_1 x_1^3 + 4\rho_2 x_1^4, \\
a_{23} &= \rho_1 x_1^{-1} + x_1^{-2} + \rho_1 x_1^{-3} + \rho_2 x_1^{-4}, \\
a_{24} &= \rho_1 x_1^{-1} + 2x_1^{-2} + 3\rho_1 x_1^{-3} + 4\rho_2 x_1^{-4}, \\
a_{31} &= \rho_2 x_1^{-3} + \rho_1 x_1^{-2} + x_1^{-1} + \rho_1 && = x_1^{-1} a_{23}, \\
a_{32} &= (T-3)\rho_2 x_1^{-3} + (T-2)\rho_1 x_1^{-2} \\
&\quad + (T-1)x_1^{-1} + T\rho_1 && = Ta_{31} - a_{14}, \\
a_{33} &= \rho_2 x_1^3 + \rho_1 x_1^2 + x_1 + \rho_1 && = x_1 a_{21}, \\
a_{34} &= (T-3)\rho_2 x_1^3 + (T-2)\rho_1 x_1^2 \\
&\quad + (T-1)x_1 + T\rho_1 && = Ta_{33} - a_{12}, \\
a_{41} &= \rho_2 x_1^{-2} + \rho_1 x_1^{-1} + 1 && = x_1 a_{13}, \\
a_{42} &= (T-2)\rho_2 x_1^{-2} + (T-1)\rho_1 x_1^{-1} + T && = Ta_{41} - (2\rho_2 x_1^{-2} + \rho_1 x_1^{-1}), \\
a_{43} &= \rho_2 x_1^2 + \rho_1 x_1 + 1 && = x_1^{-1} a_{11}, \\
a_{44} &= (T-2)\rho_2 x_1^2 + (T-1)\rho_1 x_1 + T && = Ta_{43} - (2\rho_2 x_1^2 + \rho_1 x_1).
\end{aligned}$$

By the rules of partitioned inversion,

$$\text{(B.4)}\qquad \begin{aligned}
\mathbf{B}_{22} &= x_1^T (\mathbf{A}_{22} - x_1^{2T} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12})^{-1} \\
&= \frac{x_1^T}{d_0} \left\{ \begin{bmatrix} a_{44} & -a_{34} \\ -a_{43} & a_{33} \end{bmatrix} + x_1^{2T} \begin{bmatrix} q_{44} & -q_{34} \\ -q_{43} & q_{33} \end{bmatrix} \right\},
\end{aligned}$$

where $(q_{ij}) = -\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, $i, j = 3, 4$, and

$$\text{(B.5)}\qquad \begin{aligned}
d_0 &= a_{33}a_{44} - a_{34}a_{43} + x_1^{2T}(a_{33}q_{44} + a_{44}q_{33} - a_{34}q_{43} - a_{43}q_{34}) \\
&\quad + x_1^{4T}(q_{33}q_{44} - q_{43}q_{34}).
\end{aligned}$$

Similarly

$$\text{(B.6)}\qquad \mathbf{B}_{12} = -\mathbf{A}_{11}\mathbf{A}_{12}\mathbf{B}_{22} = \frac{x_1^T}{d_0} \left\{ \begin{bmatrix} b'_{13} & b'_{14} \\ b'_{23} & b'_{24} \end{bmatrix} + x_1^{2T} \begin{bmatrix} b''_{13} & b''_{14} \\ b''_{23} & b''_{24} \end{bmatrix} \right\},$$

where the two 2×2 matrices are, respectively, the product of $-\mathbf{A}_{11}\mathbf{A}_{12}^{-1}$ times those in the last line of (B.4).

For column $T - 1$, the first two and last two equations in (2.1) are those in (B.1) with T replaced by $T - 1$ and 0 interchanged with 1 in the left-hand sides of the last two lines. All components $w_{i,T-1}$, $i = 1, 2, \cdots, T$, satisfy the homogeneous difference equation, and hence the vector of constants is like that in (B.2) with $\mathbf{e}' = (1, 0)$ instead of $\mathbf{u}'$. Summarizing, we have

$$\text{(B.7)}\qquad \mathbf{C}(T) = \frac{x_1^T}{d_0} \begin{bmatrix} b'_{14} + x_1^{2T} b''_{14} \\ b'_{24} + x_1^{2T} b''_{24} \\ -a_{34} - x_1^{2T} q_{34} \\ a_{33} + x_1^{2T} q_{33} \end{bmatrix}. \qquad \mathbf{C}(T-1) = \frac{x_1^T}{d_0} \begin{bmatrix} b'_{13} + x_1^{2T} b''_{13} \\ b'_{23} + x_1^{2T} b''_{23} \\ a_{44} + x_1^{2T} q_{44} \\ -a_{43} - x_1^{2T} q_{43} \end{bmatrix}.$$

Taking $|x_1| < 1$, one sees that for large T a reasonable approximation can be obtained discarding from the $C_k(s)$ obtained above, those parts with x_1^{2T} or x_1^{4T} as a factor. Hence one can approximate w_{iT}, $w_{i,T-1}$, and all w_{ij} in (3.6), by taking d_0 to be its leading part in (B.5), and omitting in (B.7) the second summand in the expression for each $C_k(s)$.

Acknowledgments. Appreciation is expressed to Professor Theodore W. Anderson for generous guidance and help while directing this work and to Professor Paul Shaman for corrections to a first draft.

References

T. W. Anderson (1971a). *The Statistical Analysis of Time Series*, John Wiley, New York.

——— (1971b). *Estimation of covariance matrices with linear structure and moving average processes of finite order*. Tech. Rep. 6. Dept. of Statistics, Stanford Univ., Stanford, Calif.

A. Calderón, F. Spitzer and H. Widom (1959). *Inversion of Toeplitz matrices*. Illinois J. Math., 3, pp. 490–498.

J. Durbin (1959). *Efficient estimation of parameters in moving average models*. Biometrika, 46, pp. 306–316.

B. G. Greenberg and A. E. Sarhan (1959). *Matrix inversion, its interest and application in analysis of data*. J. Amer. Statist. Assoc., 54, pp. 755–766.

L. Guttman (1955). *A generalized simplex for factor analysis*. Psychometrika, 20, pp. 173–192.

D. Kershaw (1969). *The explicit inverses of two commonly occurring matrices*. Math. Comp., 23, pp. 189–191.

V. Lovass-Nagy and D. L. Powers (1969). *Reduction of functions of some partitioned matrices*. Ibid., 23, pp. 127–133.

R. P. MENTZ (1972), *On the inverse of some covariance matrices of Toeplitz type*, Tech. Rep. 21, Dept. of Statistics, Stanford Univ., Stanford, Calif.

——— (1975), *Asymptotic properties of some estimators in moving average models*, Tech. Rep., Dept. of Statistics, Stanford Univ., Stanford, Calif.

CH. K. MUSTAFI (1967), *The inverse of a certain matrix, with an application*, Ann. Math. Statist., 38, pp. 1289–1292.

P. SHAMAN (1968), *On the inverse of the covariance matrix of a first order moving average*, Tech. Rep. 25, Dept. of Statistics, Stanford Univ., Stanford, Calif.

——— (1969), *On the inverse of the covariance matric of a first order moving average*, Biometrika, 56, pp. 595–600.

——— (1971), *On the inverse of the covariance matrix for an autoregressive-moving average process*, Tech. Rep. 56, Dept. of Statistics, Carnegie-Mellon Univ., Pittsburgh, Pa.

W. F. TRENCH (1964), *An algorithm for the inversion of finite Toeplitz matrices*, this Journal, 12, pp. 515–522.

——— (1967), *Weighting coefficients for the prediction of stationary time series from the finite past*, this Journal, 15, pp. 1502–1510.

Y. UKITA (1955), *Characterization of 2-type diagonal matrices with an application to order statistics*, J. Hokkaido College of Art and Literature, 6, pp. 66–75.

A. M. WALKER (1961), *Large sample estimation of parameters for moving average models*, Biometrika, 48, pp. 343–357.

J. WISE (1955), *The autocorrelation and the spectral density function*, Ibid., 42, pp. 151–159.

Estimation in the first order moving average model based on sample autocorrelations

by Raúl Pedro Mentz, University of Tucumán, Argentina

For the first order moving average we consider a proposal by Walker (*Biometrika*, 1961) to use k sample autocorrelations ($1 < k < T$, T sample size), to estimate the first autocorrelation of the model, and hence its basic parameter. When $k = k_T \to \infty$ as $T \to \infty$, the estimator is proved consistent and asymptotically normal and efficient, the latter provided k_T dominates $\log T$ and is dominated by $T^{\frac{1}{2}}$. An alternative form of the estimator facilitates the calculations and the analysis of the role of k, without changing the asymptotic properties.

1. Introduction. We consider the moving average time series model

$$y_t = \varepsilon_t + \alpha\varepsilon_{t-1}, \qquad t = \cdots, -1, 0, 1, \cdots, \tag{1.1}$$

where the ε_t are i.i.d. (independent identically distributed) normal $(0, \sigma^2)$, $0 < \sigma^2 < \infty$. Then (1.1) defines a stationary stochastic process with covariance sequence $\sigma_0 = \sigma^2(1 + \alpha^2)$, $\sigma_1 = \sigma_{-1} = \sigma^2\alpha$, $\sigma_j = 0$, $|j| > 1$, and autocorrelation sequence $\rho \equiv \rho_1 = \rho_{-1} = \alpha/(1 + \alpha^2)$, $\rho_j = 0$, $|j| > 1$. We further assume that $|\alpha| < 1$, which makes $|\rho| < \frac{1}{2}$. Only $|\alpha| \neq 1$ is important, since for $|\alpha^*| > 1$ the parameters $1/\alpha^*$ and σ^2/α^{*2} provide an equivalent parameterization. See, for example, Anderson (1971), Chapter 7.

To estimate α or ρ, we consider a sample $y_1, \cdots, y_T$ from (1.1). Consistent estimators of the σ_j and ρ_j are, respectively, the sample autocovariances $c_j = T^{-1}\sum_{t=1}^{T-|j|} y_t y_{t+|j|}$, and the sample autocorrelations $r_j = c_j/c_0$, $(r_1 \equiv r)$. Since we are interested in convergence in probability and in distribution, and $r \to \rho$ in probability as $T \to \infty$, we shall take $|r| < \frac{1}{2}$ throughout. In fact $r \to \rho$ a.s. as $T \to \infty$ (see, for example, Hannan (1970), Chapter IV), but we do not use this result here.

The moment estimator obtained by solving for $\hat{\alpha}$ the equation $r = \hat{\alpha}/(1 + \hat{\alpha}^2)$, namely $\hat{\alpha} = [1 - (1 - 4r^2)^{\frac{1}{2}}]/(2r)$, is consistent for α but Whittle (1953) proved it is inefficient compared with the maximum likelihood estimator. Its inefficiency can be ascribed to that of r as an estimator of ρ, and Walker (1961) proposed to improve the asymptotic efficiency by correcting it in terms of $r_2, \cdots, r_k$, for some k sufficiently large.

Let $\mathbf{W}(\rho)$ be the $k \times k$ covariance matrix of the limiting normal distribution

Received April 1976; revised February 1977.

[1] Work at Stanford was supported by a research contract with the Office of Naval Research (T. W. Anderson, Project Director), Contract N00014-75-C-0442 (NR-042-034).

AMS 1970 *subject classifications*. 62M10, 62M99.

Key words and phrases. First order moving average model, approximate maximum likelihood estimation, consistent and asymptotically efficient estimators.

of $T^{\frac{1}{2}}(\mathbf{r} - \boldsymbol{\rho})$, where $\mathbf{r} = (r_1, \cdots, r_k)'$, $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_k)' = (\rho, 0, \cdots, 0)'$, and let $\mathbf{W}(\mathbf{r})$ be the same with ρ_j replaced by r_j. If we partition $\mathbf{r} = (r, \mathbf{r}^{(2)})$, and

$$\mathbf{W}(\mathbf{r}) = \begin{pmatrix} w_{11} & \mathbf{w}_{12}' \\ \mathbf{w}_{12} & \mathbf{W}_{22} \end{pmatrix}, \tag{1.2}$$

then $w_{11} = 1 - 3r^2 + 4r^4$, $\mathbf{w}_{12} = (2r(1 - r^2), r^2, 0, \cdots, 0)'$, and $\mathbf{W}_{22}$ has $1 + 2r^2$, $2r$ and r^2 as components with indices i, j for $|i - j| = 0, 1, 2$, respectively, and 0 elsewhere.

Walker's estimator of ρ for model (1.1) is

$$\begin{aligned} \hat{\rho} &= \textstyle\sum_{j=0}^{k-1} m(j) r_{j+1} = r + \sum_{j=1}^{k-1} m(j) r_{j+1} = r - \mathbf{w}_{12}' \mathbf{W}_{22}^{-1} \mathbf{r}^{(2)} \\ &= r - 2r(1 - r^2) \textstyle\sum_{j=1}^{k-1} w^{1j} r_{j+1} - r^2 \sum_{j=1}^{k-1} w^{2j} r_{j+1}, \end{aligned} \tag{1.3}$$

where w^{ij} are the components of the inverse matrix $\mathbf{W}_{22}^{-1}$. Note that the $m(j)$ are random variables, functions of r.

Walker (1961) developed the asymptotic theory for his proposal when k is treated as fixed. In the following sections we present the corresponding theory when $k = k_T$, a function of the series length T, such that $\lim_{T\to\infty} k_T = \infty$. It was conjectured by Walker [(1961), page 353] that such a theory could be developed, essentially by means of the tools we use below, except that the w^{ij} will be evaluated explicitly.

Proofs will be simplified below; for full details see Mentz (1975). General comments are collected in Section 5.

2. The components in two rows of $\mathbf{W}_{22}^{-1}$. From Mentz (1976) we have that

$$w^{ij} = [C_1(j) + iC_2(j)]x^i + [C_3(j) + iC_4(j)]x^{-i}, \qquad i, j = 1, 2, \cdots, T, \tag{2.1}$$

where $x = [-1 + (1 - 4r^2)^{\frac{1}{2}}]/(2r)$, and for rows $i = 1, 2$,

$$\begin{aligned} C_1(1) &= a_{22}/\Delta + O_p(x^{2k}), & C_2(1) &= -a_{21}/\Delta + O_p(x^{2k}), \\ C_1(2) &= -a_{12}/\Delta + O_p(x^{2k}), & C_2(2) &= a_{11}/\Delta + O_p(x^{2k}), \end{aligned} \tag{2.2}$$

while $C_s(t) = O_p(x^{2k})$ for all other s and t, where $a_{11} = (r/2)[(1 - 4r^2)^{\frac{1}{2}} - 3]$, $a_{12} = -(r/2)[(1 - 4r^2)^{\frac{1}{2}} + 1]$, $a_{21} = -r^2$, $a_{22} = 0$, $\Delta = h_1 + O_p(x^{2k})$, $h_1 = a_{11}a_{22} - a_{12}a_{21} \neq 0$. Note that $|x| < 1$ for $|r| < \frac{1}{2}$.

3. Consistency. We now consider the following.

THEOREM 1. *Let $\{y_t\}$ satisfy equation (1.1), where $|\alpha| < 1$ and the ε_t are* i.i.d. *normal, $\mathscr{E}\varepsilon_t = 0$, $\mathscr{E}\varepsilon_t^2 = \sigma^2 (0 < \sigma^2 < \infty)$ for all t. Suppose that a set of observations of $\{y_t\}$ at times $t = 1, 2, \cdots, T$ is available, and that $k = k_T$ is a function of T $(T \geqq k + 1)$ satisfying $\lim_{T\to\infty} k_T = \infty$. Then if $\hat{\rho}$ is defined by (1.3), $p\lim_{T\to\infty} \hat{\rho} = \rho$.*

PROOF. Substituting (2.1) in (1.3) we delete terms having x^{2k} as a factor. Hence to prove that $\sum_{j=1}^{k-1} m(j) r_{j+1} \to_p 0$, we replace the $m(j)$ by $-\{2r(1 - r^2)[C_1(1) + jC_2(1)] + r^2[C_1(2) + jC_2(2)]\}x^j$. From (2.2) it suffices to prove that

$\sum_{j=1}^{k-1} x^j r_{j+1} \to_p 0$ as $T \to \infty$. Now

$$P\{|\textstyle\sum_{j=1}^{k-1} r_{j+1} x^j| > \varepsilon\} \leqq \sum_{j=2}^{n} P\{|r_j| > \varepsilon/2^n\} + P\left\{\frac{|x|^n}{1 - |x|} > \varepsilon/2\right\}. \tag{3.1}$$

In the second term $|x| \to_p |\alpha| < 1$, and by proper choice of n (independently of k and T) the term can be made arbitrarily small. The first term tends to 0 since $r_j \to_p 0$ for $j = 2, \cdots, n$. A similar argument can be used with $\sum_{j=1}^{k-1} j x^j r_{j+1}$.

4. Asymptotic normality. We now prove the following.

THEOREM 2. *Let the conditions of Theorem* 1 *hold together with*

$$\lim_{T\to\infty} k_T^{-1} \log T = 0, \qquad \lim_{T\to\infty} T^{-1} k_T^2 = 0. \tag{4.1}$$

Then as $T \to \infty$, $T^{\frac{1}{2}}(\hat{\rho} - \rho)$ has a limiting normal distribution with parameters 0 *and* $(1 - \alpha^2)^3/(1 + \alpha^2)^4$.

PROOF. The proof is done in three parts.

Part 1. (Replacement of r_j by c_j and simplification.)

$$\begin{aligned} \hat{\rho} - \rho &= \textstyle\sum_{j=1}^{k} m(j-1) r_j - \rho \\ &\equiv c_0^{-1} \textstyle\sum_{j=0}^{k} m(j-1)(c_j - \mathscr{E}c_j) - T^{-1} c_0^{-1} \sigma_1, \end{aligned} \tag{4.2}$$

where we introduced $m(-1) \equiv -\sigma_0^{-1}\sigma_1 = -\rho = -\alpha/(1 + \alpha^2)$. Since $c_0 \to \sigma_0$ and $T^{\frac{1}{2}} T^{-1} c_0^{-1} \sigma_1 \to 0$ in probability as $T \to \infty$, we see that $T^{\frac{1}{2}}(\hat{\rho} - \rho)$ has the same limiting distribution as $T^{\frac{1}{2}} \sigma_0^{-1} \sum_{j=0}^{k} m(j-1)(c_j - \mathscr{E}c_j)$.

From the argument in Section 2 we have that $m(j) = m_1(j) + x^{\lambda k} m_2(j)$, $j = 1, 2, \cdots, k - 1$, for some $0 < \lambda \leqq 2$ and functions m_1 and m_2. Hence

$$\begin{aligned} \hat{\rho} - \rho = \sigma_0^{-1} &\{\textstyle\sum_{j=0}^{k} m_1(j-1)(c_j - \mathscr{E}c_j) \\ &+ x^{\lambda k} \textstyle\sum_{j=0}^{k} m_2(j-1)(c_j - \mathscr{E}c_j) - x^{\lambda k} m_2(-1)(c_0 - \mathscr{E}c_0)\}, \end{aligned} \tag{4.3}$$

where $m_2(-1) = m_1(-1)$. The first two terms have sums of the same nature, and it will be shown below that the first term normalized by $T^{\frac{1}{2}}$ has a limiting normal distribution. Since the second term has a factor $x^{\lambda k} \to 0$ in probability, this part will be completed if we show that $T^{\frac{1}{2}}|x|^{\lambda k} \to_p 0$. But $\log T/k_T \to 0$ assumed in (4.1) implies this result.

Hence the limiting distribution is unchanged if $m(j)$ is replaced by $m_1(j)$, where

$$\begin{aligned} m_1(-1) &= -\alpha/(1 + \alpha)^2, \qquad m_1(j) = x^j[1 + j(1 - 4r^2)^{\frac{1}{2}}], \\ & \qquad\qquad\qquad\qquad\qquad\qquad j = 0, \cdots, k - 1. \end{aligned} \tag{4.4}$$

Part 2. (Substituting parameters for rv in the $m_1(j)$.) We now prove that if $\mu_1(j)$ denotes $m_1(j)$ with r replaced by ρ, then

$$p\lim_{T\to\infty} T^{\frac{1}{2}} \textstyle\sum_{j=1}^{k-1} [m_1(j) - \mu_1(j)] c_{j+1} = 0, \tag{4.5}$$

where we used that $\mathscr{E}c_j = 0$ for $j > 1$. Let $\tilde{x} \equiv x(\rho) = -\alpha$. Then

$$m_1(j) - \mu_1(j) = x^j[1 + j(1 - 4r^2)^{\frac{1}{2}}] - \tilde{x}^j[1 + j(1 - 4\rho^2)^{\frac{1}{2}}]$$

$$= [(1 - 4r^2)^{\frac{1}{2}} - (1 - 4\rho^2)^{\frac{1}{2}}]j\tilde{x}^j \tag{4.6}$$

$$+ [1 + j(1 - 4r^2)^{\frac{1}{2}}](x^j - \tilde{x}^j),$$

so that the rv in (4.5) will be taken to be formed by the corresponding two terms.

The sum on j of the first term in (4.6) gives a contribution of the form considered below, that is $T^{\frac{1}{2}} \sum_j \tilde{x}_j c_{j+1}$. Since $\tilde{x}^j = (-\alpha)^j$ is summable ($|\alpha| < 1$), such term converges in distribution to a normal rv with 0 expected value and finite variance. Further $(1 - 4r^2)^{\frac{1}{2}} \to (1 - 4\rho^2)^{\frac{1}{2}}$ in probability as $T \to \infty$, so that the contribution due to the first term of (4.6) converges stochastically to 0.

In the second term we are led to deal with $T^{\frac{1}{2}} \sum_{j=1}^{k-1} (x_j - \tilde{x}_j)c_{j+1}$, or a similar expression with weights $j(x_j - \tilde{x}_j)$. Using the same technique as in (3.1), together with the facts that $x \to_p \tilde{x}$ and $T^{\frac{1}{2}}(c_2, \cdots, c_n)$ is asymptotically normally distributed with 0 expectations and finite variances, this expression is shown to converge to 0 in probability as $T \to \infty$.

Part 3. (The asymptotic distribution.) The conclusion of the preceding argument is that we have to find the limiting distribution of

$$\Omega = T^{\frac{1}{2}}\sigma_0^{-1} \sum_{j=0}^{k-1} \mu_1(j-1)(c_j - \mathscr{E}c_j) \equiv T^{\frac{1}{2}} \sum_{t=1}^{T} W_t, \tag{4.7}$$

where $\mu_1(j) = \delta_j^*(-\alpha)^j[1 + j(1 - \alpha^2)/(1 + \alpha^2)]$, $j = 0, 1, \cdots, k-1$, $\delta_j^* = \frac{1}{2}$ for $j = -1$ and 1 otherwise, and we take $W_t = \sum_{j=0}^{k} \sigma_0^{-1}\mu_1(j-1)(y_t y_{t+j} - \mathscr{E}y_t y_{t+j})$, $t = 1, \cdots, T$. For $t = T - k + 1, \cdots, T$ the sums should range only up to $T - t$, but the simplification means adding a total of $k^2/2$ extra terms which is negligible compared with the existing Tk terms, because $k^2/T \to 0$ as $T \to \infty$. Taken as a stochastic process $\{W_t\}$ is stationary, finitely dependent of order $k + 1$, and finitely correlated of order 1.

We now proceed as in Anderson [(1971), pages 538–539]. Let $\{N\}$ be a sequence of integers such that $k/N \to 0$ as $T \to \infty$, and let $M = [T/N]$. Then (4.7) has the same limiting distribution as $M^{-\frac{1}{2}} \sum_{j=1}^{M} (Z_j + Y_j) + T^{-\frac{1}{2}}R$, where

$$Z_j = N^{-\frac{1}{2}} \sum_{t=1}^{N-k} W_{(j-1)N+t}, \qquad Y_j = N^{-\frac{1}{2}} \sum_{t=N-k+1}^{N} W_{(j-1)N+t},$$

$$j = 1, \cdots, M,$$

and $R = W_{NM+1} + \cdots + W_T$ ($R \equiv 0$ if $NM = T$). This random variable has the same limiting distribution as

$$\Omega^* = M^{-\frac{1}{2}} \sum_{j=1}^{M} Z_j, \tag{4.8}$$

which is proved by showing that the other terms converge to 0 in mean square. Next we have

$$\mathscr{E}Z_j^2 = N^{-1}[(N-k)\mathscr{E}W_1^2 + (N-k-1)2\mathscr{E}W_1W_2], \tag{4.9}$$

and let us write

$$\Omega^* = (\mathscr{E}Z_1^2) \sum_{j=1}^{M} \frac{Z_j}{(M\mathscr{E}Z_j^2)^{\frac{1}{2}}} \equiv (\mathscr{E}Z_1^2) \sum_{j=1}^{M} z_j. \tag{4.10}$$

Then $\mathscr{E} \sum z_j = 0$, $\mathscr{E} \sum z_j^2 = 1$. To use Liapounov's central limit theorem [Loève (1963), Chapter VI] it suffices to prove that for some $\delta > 0$, $\lim_{T\to\infty} \sum \mathscr{E}|z_j|^{2+\delta} = 0$. We choose $\delta = 2$. Then it suffices to prove that

$$\sum_{j=1}^{M} \mathscr{E}z_j^4 = \sum_{j=1}^{M} \mathscr{E} \frac{Z_j^4}{M^2(\mathscr{E}Z_j^2)^2} = \frac{\mathscr{E}Z_1^4}{M(\mathscr{E}Z_1^2)^2} \tag{4.11}$$

converges to 0 as $T \to \infty$, where $\mathscr{E}Z_1^4 = N^{-2} \sum_{s,t,q,v=1}^{N-k} \mathscr{E}W_sW_tW_qW_v$. Proceeding as in Anderson [(1971), page 539] or Berk [(1974), page 498], we conclude that we only need to show that those terms with $v = t$, $|t - s| \leqq k + 1$, $|s - q| \leqq k + 1$, $t < s < q$ provide contributions that tend to 0 as $T \to \infty$. There are at most $4(N-k)(k+1)^3$ such terms, and the total contribution to (4.11) is bounded by a constant times $4\sigma^8(N-k)(k+1)^3M^{-1}(\mathscr{E}Z_1^2)^{-2}\{\sum_{j=0}^{k} \mu_1(j-1)\}^4$. This completes the proof because this tends to 0 as $T \to \infty$, by (4.1), and means that $T^{\frac{1}{2}}(\hat{\rho} - \rho)$ is asymptotically normal with expectation 0 and variance

$$\lim_{T\to\infty} \mathscr{E}Z_1^2 = \lim_{T\to\infty} (\mathscr{E}W_1^2 + 2\mathscr{E}W_1W_2). \tag{4.12}$$

It remains to prove that (4.12) equals $(1 - \alpha^2)^2/(1 + \alpha^2)^4$. This is done by direct but rather laborious calculation, and the details are omitted here.

COROLLARY 1. *Under the conditions of Theorem* 2, *let* $\hat{\alpha}'$ *be the moment estimator with* r *replaced by* $\hat{\rho}$. *Then* $T^{\frac{1}{2}}(\hat{\alpha}' - \alpha)$ *has a limiting normal distribution with parameters* 0 *and* $1 - \alpha^2$.

5. Comments. We here collect several comments about Walker's proposal and our findings.

a. The estimator (1.3) can be interpreted as an improvement in the asymptotic efficiency of r, a consistent estimator of ρ, by approximating the maximum likelihood estimator by means of a linear combination of $r, r_2, \cdots, r_k$, with random coefficients. Since $T^{\frac{1}{2}}(r - \rho)$ tends in distribution to a normal with variance $1 - 3\rho^2 + 4\rho^4 = (1 - \alpha^2)^3(1 + \alpha^2)^{-4} + \{4\alpha^2 + \alpha^4(1 + \alpha^2)\}(1 + \alpha^2)^{-4}$, the improvement achieved by $\hat{\rho}$ is the second term, the first term being also the asymptotic variance of the maximum likelihood estimator of ρ. In terms of α, comparing the moment estimator with the same with r replaced by $\hat{\rho}$, the variance of the limiting distribution of the former is [cf. Whittle (1953)] $(1 + \alpha^2 + 4\alpha^4 + \alpha^6 + \alpha^8)(1 - \alpha^2)^{-2} = 1 - \alpha^2 + \alpha^2[4 + \alpha^2(1 + \alpha^2)^2](1 - \alpha^2)^{-2}$, where $1 - \alpha^2$ is also the asymptotic variance of the maximum likelihood estimator of α.

Walker's approach to derive (1.3) is to work with the limiting normal distribution of $T^{\frac{1}{2}}(\mathbf{r} - \boldsymbol{\rho})$ and use maximum likelihood ideas. Since some approximations are introduced, the final estimator comes closer to approximating the least-squares estimator, the Jacobian being omitted. These approximations have no relevance for the asymptotic theory we developed, but may be important in small samples.

b. The estimator (1.3) is consistent and asymptotically efficient, proved biased for small samples and "a priori" appears as robust to departures from normality in the distribution of the ε_t, for moderate sample sizes.

The conclusions for large samples follow from the studies by Walker (1961) [see also Anderson, (1971), Section 5.7.2] and the present paper. Small sample studies by Monte Carlo trials were made by Walker (1961) and more extensively by McClave (1974).

In the small-sample studies ($T = 100$), (1.3) showed considerable efficiency (i.e., agreement with the large-sample variances), but also rather important biases. Walker (1961) proposed a correction for bias that has not been studied empirically in detail.

The robustness argument arises because only the limiting normal distribution of $T^{\frac{1}{2}}(r - \rho)$ enters in the derivation of (1.3), and it is well known that the same limiting distribution holds for a wide class of distributions of the ε_t. Unfortunately no empirical results for small samples are available in this connection.

c. If in a practical situation $\mathscr{E}y_t$ were unknown, it would be estimated from the data. Then c_j would be replaced by $c_j^* = T^{-1}\sum_{t=1}^{T-|j|}(y_t - \bar{y})(y_{t+|j|} - \bar{y})$, where $\bar{y} = T^{-1}\sum_{t=1}^{T} y_t$; see, for example, Anderson (1971), for this and other types of mean corrections. It is easily proved that our results hold for the modified version of the estimators, since, for example $T^{\frac{1}{2}}\sum_{j=0}^{k}\{m(j-1)(c_j - \mathscr{E}c_j) - m^*(j-1)(c_j^* - \mathscr{E}c_j^*)\} \to_p 0$ as $T \to \infty$, where $m^*(j-1)$ is $m(j-1)$ with c_j replaced by c_j^*.

d. Our analysis has been restricted to the first-order moving average model. Extension of the approach to a model of order $q > 1$ seems quite feasible. The components of the W matrix in (1.2) are known for all q. W_{22} will be a Toeplitz matrix with equal elements along its central diagonals, and zeroes elsewhere; the components of the inverses of such matrices are given as functions of the roots of the associated polynomial equation in Mentz (1976). W_{22} is positive definite and can therefore be taken as the covariance matrix of a stationary moving average process; by the argument in Anderson [(1971), pages 224–225] half of the roots are larger and half are less than one in absolute value, as was the case in Section 2.

6. A modification to simplify the computations. From the argument in the proof of Theorem 2, it follows that

$$\hat{\rho}^* = \sum_{j=0}^{k-1} m_1(j) r_{j+1}, \tag{6.1}$$

where $m_1(j)$ was defined in (4.4), has the same large-sample properties of (1.3). It discards parts having x^k as dominating factor, and hence differs only slightly from $\hat{\rho}$ if k is moderately large.

From a practical point of view (6.1) makes easy the choice of k, guided by the degree of numerical approximation that is desired. Consider Table 1 where for r negative the values of $m_1(j)$ are those in the table all taken with positive

signs. Once the numerical value of r is available, the table can be used to see how many sample autocorrelations r_j, $2 \leqq j \leqq k$ to include in (6.1).

Using data generated in a computer, Walker (1961) studied in detail a set of $T = 100$ observations from (1.1) with $\alpha = 0.5$ ($\rho = 0.4$). He estimated $r = 0.35005$, $r_2 = -0.06174$, $r_3 = -0.08007$, $r_4 = -0.14116$ and $r_5 = -0.15629$. Then his estimates for $1 \leqq k \leqq 5$ are compared in Table 2 with $\hat{\rho}^*$. Only r_j up to $j = 5$ are provided by Walker. From Table 1 for $r = 0.35$ it is apparent that a larger k will be called for. However the behavior of $\hat{\rho}^*$ seems comparable to that of $\hat{\rho}$.

TABLE 1

Values of $m_1(j)$ for selected values of r

j	.05	.15	.25	.35	.45
1	−.1000000	−.3000000	−.5000000	−.7000000	−.9000000
2	.0075125	.0685482	.1961524	.4049504	.7353557
3	−.0005018	−.0139772	−.0692193	−.2140023	−.5682477
4	.0000314	.0026761	.0230114	.1072520	.4234477
5	−.0000018	−.0004922	−.0073620	−.0519085	−.3075804
6	.0000001	.0000880	.0022931	.0245097	.2192185
7	0.0000000	−.0000154	−.0007003	−.0113615	−.1539701
8	0.0000000	.0000026	.0002106	.0051919	.1068903
9	0.0000000	−.0000004	−.0000626	−.0023457	−.0735060
10	0.0000000	0.0000000	.0000184	.0010500	.0501521
11	0.0000000	0.0000000	−.0000053	−.0004664	−.0339916
12	0.0000000	0.0000000	.0000015	.0002058	.0229082
13	0.0000000	0.0000000	−.0000004	−.0000903	−.0153631
14	0.0000000	0.0000000	.0000001	.0000394	.0102590
15	0.0000000	0.0000000	0.0000000	−.0000171	−.0068249

TABLE 2

Estimates of ρ, $T = 100$, Walker's data

k	$\hat{\rho}$	$\hat{\rho}^*$
2	0.38051	0.39327
3	0.36879	0.36084
4	0.38498	0.39106
5	0.37934	0.37429

To compute (1.3) exactly, Walker (1961) proposed an iterative procedure; (6.1) is of course much simpler. In fact (2.2) could be written in detail to give the exact form of (1.3), and a table similar to Table 1 prepared for that case, but we omit these details here.

Acknowledgments. This study is a part of the author's Ph. D. dissertation in the Department of Statistics, Stanford University. Appreciation is expressed to Professor T. W. Anderson for generous guidance and help while directing this work. The author is also indebted to a referee, an associate editor and the editor for useful comments.

References

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.
BERK, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489–502.
HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.
LOÈVE, MICHEL (1963). *Probability Theory* (3rd ed.) Van Nostrand, New York.
MCCLAVE, JAMES T. (1974). A comparison of moving average estimation procedures. *Comm. Statist.* **3** 865–883.
MENTZ, RAÚL P. (1975). Asymptotic properties of some estimators in moving average models. Technical Report No. 21, Department of Statist., Stanford Univ.
MENTZ, RAÚL P. (1976). On the inverse of some covariance matrices of Toeplitz type. *SIAM J. Appl. Math.* **31** 426–437.
WALKER, A. M. (1961). Large sample estimation of parameters for moving average models. *Biometrika* **48** 343–357.
WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423–434.

Institute of Statistics
University of Tucumán
Casilla de correo 209
Tucumán, Argentina

Estimation in the first-order moving average model through the finite auteregressive approximation

by Raúl Pedro Mentz

University of Tucumán, Tucumán, Argentina

Received October 1976, final version received February 1977

To estimate α in the model $y_t = u_t + \alpha u_{t-1}$, we consider a proposal by Durbin (*Biometrika*, 1969). It consists in fitting an autoregression of order k to the data, and deriving from there an estimate $\hat{\alpha}$. The probability limit and the variance of the limiting normal distribution of $\hat{\alpha}$ are presented and discussed in detail, when the sample size $T \to \infty$, but k remains fixed. The differences between the resulting values and those corresponding to the maximum likelihood estimator are exponentially decreasing functions of k. Several modifications of the estimator are discussed and found consistent, but asymptotically inefficient.

1. Introduction

We consider the time series model

$$y_t = u_t + \alpha u_{t-1}, \tag{1.1}$$

where $0 < |\alpha| < 1$, and the u_t are independent, identically distributed, normal random variables with $\mathcal{E}u_t = 0$, $\mathcal{E}u_t^2 = \sigma^2$ $(0 < \sigma^2 < \infty)$, for all t. The autocovariance sequence of the process is given by

$$\sigma_0 = \sigma^2(1+\alpha^2), \quad \sigma_1 = \sigma^2\alpha = \sigma_{-1}, \quad \sigma_j = 0, \qquad |j| > 1, \tag{1.2}$$

and the autocorrelation sequence has

$$\rho_1 = \alpha(1+\alpha^2)^{-1} = \rho_{-1}, \quad \rho_j = 0, \qquad |j| > 1. \tag{1.3}$$

*This study is a part of the author's Ph.D. dissertation in the Department of Statistics, Stanford University. Appreciation is expressed to Professor T. W. Anderson for generous guidance and help while directing this work. Work at Stanford was supported by a Research Contract with the Office of Naval Research (Contract N00014-75-C-0442, NR-042-034, T. W. Anderson, Project Director).

On the basis of a sample $y_1, \ldots, y_T$, the autocovariances and autocorrelations can be estimated consistently but inefficiently (asymptotically and in comparison with the maximum likelihood estimators) by the sample autocovariances

$$c_j = \frac{1}{T}\sum_{t=1}^{T-j} y_t y_{t+j}, \tag{1.4}$$

and sample autocorrelations

$$r_j = c_j/c_0, \tag{1.5}$$

respectively.

This is the first-order case of the moving average model of order q, defined similarly by

$$y_t = u_t + \alpha_1 u_{t-1} + \ldots + \alpha_q u_{t-q}. \tag{1.6}$$

The importance of moving average models for time series analysis stems from several facts. Among them we note the following:

(a) In a variety of fields of application, the formulation of reasonable statistical models leads to moving average schemes, or more complicated versions of them. For several examples see Nicholls, Pagan and Terrell (1975). One may ascribe part of the potentiality of the moving average model in these situations to its structure, which postulates linear combinations of current and past error terms to explain the random part of the data.

(b) The autocovariance sequence has zero values for lag lengths exceeding q. This may be a reasonable hypothesis on which to model some empirical phenomena.

(c) The spectral density function is a real-valued trigonometric polynomial. As such it can approximate the spectral density function of a wide class of stochastic processes or time series. In fact, every second-order stationary stochastic process with a spectral density is a (possibly two-sided, and infinite) moving average of uncorrelated random variables. Hence, finite moving averages like (1.6) approximate most time series occurring in practice.

(d) Due to the relation between moving average and autoregressive models, which we consider in some detail below, the moving average model may on some occasions provide a competing framework with similar properties to that of the autoregressive model and less parameters to be studied statistically. This is important because the linear dependence of a time series on its own past values provides another empirically attractive model.

(e) The moving average model is a simple case of a mixed model, autoregressive with moving average residuals. Mixed models are very flexible tools to study time series empirically, and provide general approximation to many

stochastic processes, since they have rational spectral densities. However, their statistical analysis has proved very hard, due mainly to the presence of the moving average part of the model.

These reasons and others have witnessed in recent years the appearance of a wealth of work in solving statistical problems in moving average models (e.g. estimation of parameters), and in modelling empirical time series with members of the family (1.6). See for example, Anderson (1971a, 1975), Hannan (1969), Mentz (1975), Mentz and Antelo (1976), and Walker (1961), for reviews of estimation procedures.

Durbin (1959) proposed to estimate the moving average parameters by a rather simple procedure, that has attracted considerable attention in the literature. However, due to the difficulties encountered in the treatment of the mathematical aspects, little is known to date about the statistical properties of the estimator even for the simpler case of (1.1). For this case McClave (1973, 1974), and Nelson (1974) investigated Durbin's proposal by Monte Carlo trials; McClave (1973) proposed several modifications to reduce the small-sample biases, which made the estimated variances depart considerably from the desired values.

In this work we present a formalization of Durbin's and some related proposals for $q = 1$, under which explicit asymptotic results are proved. In our presentation mathematical details are kept at a minimum; for missing details see Mentz (1975).

2. The approximating-autoregression approach to estimation

Model (1.1) is equivalent to the finite autoregression of order m ($m \geqq 1$),

$$\sum_{j=0}^{m} (-\alpha)^j y_{t-j} = u_{tm}^*, \tag{2.1}$$

with autocorrelated residuals: $\mathcal{E}u_{tm}^* = 0$, $\mathcal{E}u_{tm}^{*2} = \sigma^2(1+\alpha^{2m+2})$, $\mathcal{E}u_{tm}^* u_{t+s,m}^* = (-1)\sigma^2\alpha^{m+1}$ if $|s| = m+1$, and otherwise equal to 0. Further, since $|\alpha| < 1$, (1.1) is also equivalent (in mean square) to the infinite autoregression,

$$\sum_{j=0}^{\infty} (-\alpha)^j y_{t-j} = u_t. \tag{2.2}$$

Durbin (1959) proposed to approximate (2.1) and (2.2) by

$$\sum_{j=0}^{k} \beta_j y_{t-j} = v_t, \tag{2.3}$$

where $\beta_0 = 1$, the v_t are taken to be uncorrelated with 0 expectations and finite common variance, and k is to be large enough to make the approximation useful for the purpose of estimation.

It should be pointed out, that the idea of approximating an arbitrary stationary time series by a finite autoregression has been used with considerable success in time series analysis, for example in problems of forecasting and of estimation of the spectral density function. For the latter see, for example, Akaike (1969), Berk (1974), and Parzen (1969).

Let $b_1, \ldots, b_k$ be the (ordinary) least squares estimators of $\beta_1, \ldots, \beta_k$ in (2.3), based on sample values $y_1, \ldots, y_T$, where $T > k$. Then Durbin (1959) argued that either one of the estimators,

$$\hat{\alpha}_T = -\left(\sum_{i=0}^{k-1} b_i b_{i+1} \Big/ \sum_{i=0}^{k} b_i^2\right), \qquad \hat{\alpha}'_T = -\left(\sum_{i=0}^{k-1} b_i b_{i+1} \Big/ \sum_{i=0}^{k-1} b_i^2\right), \tag{2.4}$$

approximates the maximum likelihood estimator (or more properly the least squares estimator) of α.

The procedure was developed by Durbin for $q \geqq 1$ along the same lines: In the first step a kth order autoregression is fitted, $q < k < T$ (which means that a $k \times k$ linear system has to be solved) and in the second step a set of q linear equations that generalize $(\sum b_i b_{i+1})\hat{\alpha}_T = -\sum b_i^2$ is to be solved explicitly.

The introduction of k as a constant to be specified in each case is of course a complication, since further k will in general depend on the unknown α. This, however, can be handled empirically, and in view of posterior developments by other writers it can be interpreted as the price paid to obtain a closed-form estimator of α related to maximum likelihood ideas. Incidentally, the moment estimator of α is $(2r_1)^{-1}\,(1-(1-4r_1^2)^{\frac{1}{2}})$: it is also a closed-form estimator, it is consistent but asymptotically inefficient compared with the maximum likelihood estimator.

Another attractive feature of Durbin's proposal [and that of Walker (1961), conceived along similar lines and based on k sample autocorrelations] is that apparently the dependence of the properties of the estimators on the normality of the error terms is small. This robustness argument is based on the fact that only some limiting normal distributions enter in the justification of (2.4), and they are valid for a wide range of distributions of the error terms. Unfortunately not only the statistical theory that is available (including the results presented below) rests on the assumption of normality, but so do the results based on Monte Carlo trials published so far.

The asymptotic theory we present here consists in letting $T \to \infty$ while k remains fixed; see, for example, Walker (1961, pp. 352–353). We derive the probability limits and the variances of the limiting normal distributions of (2.4) and of some variants discussed below. An alternative approach is to let $k = k(T)$, a function of T, such that $k(T) \to \infty$ as $T \to \infty$. Under this hypothesis one expects to prove, for example, that $T^{\frac{1}{2}}(\hat{\alpha}_T - \alpha)$, has a limiting normal distribution with variance $1-\alpha^2$, which is that corresponding to the maximum

likelihood estimator of α. See, for example, Hannan (1970, p. 375). Note however that some of our results are functions of k and as such facilitate the study of its role in the estimation problem. The present author, Mentz (1975), was able to treat satisfactorily Walker's (1961) estimator of ρ_1 for model (1.1) using the approach of $k = k(T) \to \infty$ as $T \to \infty$.

3. Results for Durbin`s original proposal

Under the hypotheses of sections 1 and 2 it can be proved that

$$\begin{aligned}\alpha_k^* \equiv \operatorname*{plim}_{T\to\infty} \hat{\alpha}_T &= \alpha \frac{(1-\alpha^{2k})(1+\alpha^{2k+4})-k\alpha^{2k}(1-\alpha^4)}{(1-\alpha^{2k+2})(1+\alpha^{2k+4})-2(k+1)\alpha^{2k+2}(1-\alpha^2)} \\ &= \alpha+\alpha^{2k+1}(1-\alpha^2) \\ &\quad \times \frac{\alpha^2(1-\alpha^{2k+2})-(k+1)(1-\alpha^2)}{(1-\alpha^{2k+2})(1+\alpha^{2k+4})-2(k+1)\alpha^{2k+2}(1-\alpha^2)} \\ &= \alpha+\alpha^{2k+1}(1-\alpha^2)[(2\alpha^2-1)-k(1-\alpha^2)]+0(k^2\alpha^{4k}),\end{aligned} \tag{3.1}$$

where by definition $|0(y)| \leqq My$ for all $y > 0$ and some fixed $M > 0$. Similar results hold for $\hat{\alpha}'_T$, but these will be omitted here; note that for k moderately large, and as $T \to \infty$, the difference between the behaviour of $\hat{\alpha}_T$ and $\hat{\alpha}'_T$ is very small. Further, as $T \to \infty$, $T^{\frac{1}{2}}(\hat{\alpha}_T - \alpha_k^*)$ has a limiting normal distribution with parameters 0, and

$$\begin{aligned}v_k^* &= (1-\alpha^2)\{1-\alpha^{2k}[1-8\alpha^2+14\alpha^4-8k\alpha^2(1-\alpha^2)]\} \\ &\quad +(1-\alpha^2)^2 B\alpha^{2k}+0(\alpha^{2k}),\end{aligned} \tag{3.2}$$

where

$$\begin{aligned}B = \tfrac{1}{3}\Bigg[&k^4(5\alpha^2+3)+k^3(6-8\alpha^2)+\frac{k^2(16\alpha^2-16\alpha^4-6)}{1-\alpha^2} \\ &+k\,\frac{74\alpha^2(1-2\alpha^2+\alpha^3)-12}{(1-\alpha^2)^2}+\frac{3(5\alpha^2+\alpha^4+2\alpha^6)}{(1-\alpha^2)^2}\Bigg].\end{aligned} \tag{3.3}$$

We observe that the differences between (3.1) and (3.2) and α and $1-\alpha^2$, respectively, are exponentially declining functions of α. For example, the factor of α in the first line of (3.1) is presented in table 1. The factor approaches 1 when $\alpha \to 0$ for given k, while it approaches $2k/(2k+3)$ as $\alpha \to 1$.

Table 1

Factors of α in (3.1), for selected values of α and k.

k	0.1	0.3	0.5	0.7	0.9
1	0.99009900	0.91743119	0.80000000	0.67114093	0.55248618
3	0.99999705	0.99819235	0.97347960	0.88987126	0.75932069
5	0.99999999	0.99997559	0.99729158	0.96309708	0.85290284
10	1.00000000	0.99999999	0.99999468	0.99801290	0.94549381
30	1.00000000	1.00000000	1.00000000	0.99999999	0.99824687
60	1.00000000	1.00000000	1.00000000	1.00000000	0.99999340
100	1.00000000	1.00000000	1.00000000	1.00000000	0.99999999

One practical implication for finite T is that for $|\alpha|$ close to 1, k should be set at a considerably high value, and for some range of k values observe if the estimates remain stable. This was done by McClave in his Monte Carlo studies, but apparently not by Nelson.

In the Monte Carlo studies by McClave and Nelson it was apparent that for $T = 100$ the agreement between the variances over replications and $T^{-1}(1-\alpha^2)$ was quite good for some values of α, while the agreement between the average over replications and α was quite poor for $|\alpha|$ close to 1. In view of (3.1) and table 1 it can be argued that not only should k be taken as large for these values of α, but T be chosen large enough to make the large sample approximations work well. This, however, was not considered by these authors.

In view of the structure of (3.1) and (3.2), it is clear that

$$\lim_{k\to\infty} \alpha_k^* = \lim_{k\to\infty} \operatorname{plim}_{T\to\infty} \hat{\alpha}_T = \alpha, \tag{3.4}$$

$$\lim_{k\to\infty} v_k^* = 1-\alpha^2, \tag{3.5}$$

which then provide a way to interpret formally the claim of approximate consistency and asymptotic efficiency in Durbin's (1959) original paper, and the statements in Hannan (1970, pp. 374–375) and Anderson (1971a, pp. 228–232).

4. Some modified versions of the estimators

There have been several attempts to modify the procedures described in section 2 to simplify the calculations or to correct for the small-sample biases.

The first step in Durbin's proposal can be approximated by the solution of the so-called Yule–Walker equations,

$$\begin{bmatrix} c_0 & c_1 & c_2 & \dots & c_{k-1} \\ c_1 & c_0 & c_1 & \dots & c_{k-2} \\ \vdots & \vdots & \vdots & & \vdots \\ c_{k-1} & c_{k-2} & c_{k-3} & \dots & c_0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = -\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}; \tag{4.1}$$

see Durbin (1959, p. 312) and note that (4.1) could also be written in terms of the sample autocorrelations. Since the population autocovariances are 0 for $|j| > 1$, Anderson (1971b) suggested replacing c_j by 0 for $|j| > 1$ in the left-hand side of (4.1), so that the $k \times k$ matrix has only three non-vanishing central diagonals; then the components of its inverse matrix can be written explicitly as

$$w_{ij} = \frac{(1-x_1^{2k-2j+2})(x_1^{j+i+1}-x_1^{j+i-1})}{r_1(1-x_1^2)(1-x_1^{2k+2})}, \qquad i \leqq j, \tag{4.2}$$

where $x_1 = (2r_1)^{-1}(-1+(1-4r_1^2)^{\frac{1}{2}})$; see, e.g., Mentz (1976). Using this result to find the solution of (4.1), and replacing the resulting estimators of the β_j in (2.4), we obtain the estimator of α

$$\tilde{\alpha}_T = -\left(\sum_{i=0}^{k-1}\left(\sum_{j=1}^{k} w_{ij}r_j\right)\left(\sum_{j=1}^{k} w_{i+1,j}r_j\right)\right)\Big/\left(\sum_{i=0}^{k}\left(\sum_{j=1}^{k} w_{ij}r_j\right)^2\right). \tag{4.3}$$

It is not difficult to see that $\operatorname{plim}_{T\to\infty}\tilde{\alpha}_T = \operatorname{plim}_{T\to\infty}\hat{\alpha}_T = \alpha_k^*$ in (3.1); further, it can be shown that as $T \to \infty$, $T^{\frac{1}{2}}(\tilde{\alpha}_T - \alpha_k^*)$ has a limiting normal distribution with parameters 0 and, say $\tilde{v}_k^*$, and that

$$\lim_{k\to\infty} \tilde{v}_k^* = \tilde{v} = (1-\alpha^2)+\alpha^6\frac{16+9\alpha^2+7\alpha^4}{1-\alpha^2}+\alpha^{12}\frac{8}{(1-\alpha^2)^2}. \tag{4.4}$$

Recently McClave (1973) considered modifications similar to the previous one, in that some c_j are replaced by 0 in both sides of (4.1). He considered replacing c_h by 0 in two ways: (a) $h = 2$; (b) $h = n+1, \dots, k$, for some integer n $(2 \leqq n < k)$ to be chosen simultaneously with k. These suggestions could be studied by means of the techniques used to derive results like (4.4); instead, and for the sake of simplicity, we consider a procedure defined by (b) for $n = 1$, that is, replacing c_j by 0 if $|j| > 1$ in both sides of (4.1). Let $\bar{\alpha}_T$ be the corresponding estimator, that is

$$\bar{\alpha}_T = -\left(\sum_{i=0}^{k-1} w_{i1}w_{i+1,1}r_1^2\Big/\sum_{i=0}^{k}(w_{i1}r_1)^2\right); \tag{4.5}$$

then it is not difficult to see that $\operatorname{plim}_{T\to\infty}\bar{\alpha}_T = \operatorname{plim}_{T\to\infty}\hat{\alpha}_T = \alpha_k^*$ in (3.1), and it can be shown that as $T \to \infty$, $T^{\frac{1}{2}}(\bar{\alpha}_T - \alpha_k^*)$ has a limiting normal distribution with parameters 0 and, say $\bar{v}_k^*$, and that

$$\lim_{k\to\infty} \bar{v}_k^* = \bar{v} = \frac{1+\alpha^2+4\alpha^4+\alpha^6+\alpha^8}{(1-\alpha^2)^2} = (1-\alpha^2)+\alpha^2\frac{4+\alpha^2(1+\alpha^2)^2}{(1-\alpha^2)^2}. \tag{4.6}$$

This is the asymptotic variance of the moment estimator of α that we introduced in section 2, and (4.6) was given for it by Whittle (1953). In fact, it can be shown that

$$\tilde{\alpha}_T = -x_1\left(\sum_{i=0}^{k-1}[x_1^{2i}+x_1^{-2i-2}-(1+x_1^2)]\right)\Big/\left(\sum_{i=0}^{k}(x_1^{2i}+x_1^{-2i}-2)\right), \tag{4.7}$$

and for large k this is very close to $-x_1$, the moment estimator.

For several values of α, $\tilde{\nu}$, $\bar{\nu}$ and $1-\alpha^2$ are compared in table 2.

Table 2

Values of $\tilde{\nu}$, $\bar{\nu}$ and $1-\alpha^2$ for different α.

α	$\tilde{\nu}$	$\bar{\nu}$	$1-\alpha^2$
0.1	0.990016	1.030916	0.99
0.2	0.961088	1.135488	0.96
0.3	0.923368	1.356351	0.91
0.4	0.923420	1.795849	0.84
0.5	1.118489	2.701388	0.75
0.6	2.028235	4.740849	0.64
0.7	5.962541	10.094951	0.51
0.8	30.477959	28.613550	0.36
0.9	362.098390	149.482220	0.19

The loss of (asymptotic) efficiency as $|\alpha| \to 1$ is striking and explains the kind of behavior observed empirically by McClave. Since McClave's proposals were considerably effective in reducing the small-sample biases, it results that the trade-off between bias and variance is very important.

It should be pointed out that for estimation in moving average models, no reduction of dimensionality is possible 'a priori' on the grounds of sufficiency, since the minimal sufficient statistic has the dimension T of the sample size. What is perhaps interesting is that in omitting from the sufficient statistic $c_0, c_1, \ldots, c_{T-1}$ the autocovariances for indices exceeding k (that is, $c_{k+1}, \ldots, c_{T-1}$) have little effect on the variances (asymptotic or small-sample), but considerable effects on the small-sample biases. If further from $c_0, c_1, \ldots, c_k$ we omit some c_j's in the ways indicated by the two proposals considered above (as well as by those given by McClave) the reverse seems to be true: gains can be obtained in terms of bias reduction, but with considerable losses in efficiency. Some justifications about why the modifications would reduce the small-sample biases are given by McClave (1973).

Appendix

The proof of (3.2) is considerably laborious and will not be reproduced. See Mentz (1975). However, a simplified argument was found to prove (3.5) directly, which was then used to derive (4.4) and (4.6) and could also be used to analyze similar modifications of Durbin's proposal. We sketch this argument here and note that a different argument to justify (3.5) was provided by Durbin (1959, sec. 4).

Direct Proof of (3.5)

Let the normal equations to estimate the β_j in (2.3) be

$$M_T\hat{\beta}_T = -m_T, \tag{A.1}$$

where $\hat{\beta}_T = (b_1, \ldots, b_k)'$, $M_T = T^{-1}\sum_{t=k+1}^{T} y_{t-1}y_{t-1}'$, $m_T = T^{-1}\sum_{t=k+1}^{T} y_{t-1}y_t$, and $y_t = (y_t, \ldots, y_{t-k-1})'$ for $t = k+1, \ldots, T$. We have that $\mathcal{E}y_t = 0$, $\mathcal{E}y_t y_t' \equiv \Sigma = (\sigma_{i-j})$. Σ is positive definite for any α, and for $T > k$, M_T is positive definite with probability 1. Further $\mathcal{E}M_T = T^{-1}(T-k)\Sigma$, $\mathcal{E}m_T = T^{-1}(T-k)q$, $q = (\sigma_1, 0, \ldots, 0)'$, $\operatorname{plim}_{T\to\infty} M_T = \Sigma$, $\operatorname{plim}_{T\to\infty} m_T = q$.

Since all needed results are homogeneous of degree 0 in σ, we take $\sigma^2 = 1$ without loss of generality.

It follows that $\operatorname{plim}_{T\to\infty}\hat{\beta}_T = \operatorname{plim}_{T\to\infty}(-M_T^{-1}m_T) = -\Sigma^{-1}q \equiv \beta^*$, whose components are $\beta_j^* = (-\alpha)^j(1-\alpha^{2k+2-2j})(1-\alpha^{2k+2})^{-1}$, $j = 1, 2, \ldots, k$. The explicit form of the components of Σ^{-1} is

$$\sigma^{ij} = (-\alpha)^{j-i}\frac{(1-\alpha^{2i})(1-\alpha^{2k-2j+2})}{(1-\alpha^2)(1-\alpha^{2k+2})}, \qquad j \geqq i; \tag{A.2}$$

see, for example, Shaman (1969). Using the β_j^* and $\alpha_k^* = -(\Sigma\beta_j^*\beta_{j+1}^*)(\Sigma\beta_j^{*2})^{-1}$ it is easy to derive (3.1). For details see Mentz (1975).

To study the limiting distribution of $T^{\frac{1}{2}}(\hat{\alpha}_T - \alpha_k^*)$ we first derive that of $T^{\frac{1}{2}}(\hat{\beta}_T - \beta^*)$. This random vector and $-T^{\frac{1}{2}}\Sigma^{-1}(m_T + M_T\beta^*)$ have the same limiting distribution. In effect

$$\begin{aligned} T^{\frac{1}{2}}(\hat{\beta}_T - \beta^*) &= -T^{\frac{1}{2}}(M_T^{-1}m_T - \Sigma^{-1}q) \\ &= -T^{\frac{1}{2}}\{\Sigma^{-1}[I+(M_T-\Sigma)\Sigma^{-1}]^{-1}[q+(m_T-q)] - \Sigma^{-1}q\}, \end{aligned} \tag{A.3}$$

where we can use that $(I+A)^{-1} = I - A + (I+A)^{-1}A^2$, provided $I+A$ is non-singular. For $A = (M_T-\Sigma)\Sigma^{-1}$, $I+A = M_T\Sigma^{-1}$ is non-singular with probability 1. We deduce that $\operatorname{plim}_{T\to\infty}(M_T-\Sigma) = 0$, $\operatorname{plim}_{T\to\infty} A = 0$, $\operatorname{plim}_{T\to\infty}(I+A)^{-1} = I$, and that $T^{\frac{1}{2}}A$ has asymptotically normal components [see e.g. Anderson

(1971a, sec. 8.4.2)]. Hence $\operatorname{plim}_{T\to\infty} T^{\frac{1}{2}}(I+A)^{-1}A^2 = 0$, and (A.3) has the same limiting distribution as

$$-T^{\frac{1}{2}}\{\Sigma^{-1}[I-(M_T-\Sigma)\Sigma^{-1}][q+(m_T-q)]-\Sigma^{-1}q\}$$
$$= -T^{\frac{1}{2}}\{\Sigma^{-1}[(m_T-q)-(M_T-\Sigma)\Sigma^{-1}q-(M_T-\Sigma)\Sigma^{-1}(m_T-q)]\}. \tag{A.4}$$

Since $T^{\frac{1}{2}}(M_T-\Sigma)$ has asymptotically normal components, and $\operatorname{plim}_{T\to\infty}(m_T-q) = 0$, (A.4) has the same limiting distribution as

$$-T^{\frac{1}{2}}\Sigma^{-1}[(m_T-q)+(M_T-\Sigma)\beta^*] = -T^{\frac{1}{2}}\Sigma^{-1}(m_T+M_T\beta^*). \tag{A.5}$$

Next: $\mathscr{E}(m_T+M_T\beta^*) = 0$, and $m_T+M_T\beta^*$ has components

$$T^{-1}\sum_{t=k+1}^{T}\sum_{h=0}^{k}\beta_h^* y_{t-i}y_{t-h} = T^{-1}\sum_{s=k+1-i}^{T-i} x_s, \qquad i = 1, \ldots, k, \tag{A.6}$$

where $x_s = \sum_{h=0}^{k}\beta_h^* y_s y_{s+i-h}$. By using (1.1) it is easily shown that the x_s are finitely dependent of order $k+1$, and hence that the limiting distribution of $T^{\frac{1}{2}}(m_T+M_T\beta^*)$ is normal with parameters 0, and

$$F = \lim_{T\to\infty} T\mathscr{E}(m_T+M_T\beta^*)(m_T+M_T\beta^*)', \tag{A.7}$$

while that of $T^{\frac{1}{2}}(\hat{\beta}_T-\beta^*)$ is normal with parameters 0 and $\Sigma^{-1}F\Sigma^{-1}$. See, for example, Anderson (1971a, theorem 7.7.5).

Using '$\sim$' to mean 'asymptotically (as $k\to\infty$) equivalent to', we have that

$$\sum_{h=0}^{k}\beta_h^* y_{t-h} = \sum_{h=0}^{k}(-\alpha)^h\frac{1-\alpha^{2k+2-2h}}{1-\alpha^{2k+2}}[\varepsilon_{t-h}-(-\alpha)\varepsilon_{t-h-1}]$$
$$\sim\sum_{h=0}^{k}(-\alpha)^h[\varepsilon_{t-h}-(-\alpha)\varepsilon_{t-h-1}]$$
$$= \varepsilon_t-(-\alpha)^{k+1}\varepsilon_{t-(k+1)}\sim\varepsilon_t,$$

so that (A.6) is approximately $T^{-1}\sum_{t=k+1}^{T}y_{t-i}\varepsilon_t$ for k large, and the components of F are approximated by

$$\lim_{T\to\infty} T^{-1}\sum_{s,t=k+1}^{T}\mathscr{E}(y_{s-i}\varepsilon_s y_{t-j}\varepsilon_t) = \sigma_{i-j}. \tag{A.8}$$

Hence $T^{\frac{1}{2}}(\hat{\beta}_T-\beta^*)$ converges in distribution as $T\to\infty$ to a normal with parameters given approximately by 0 and Σ^{-1}. Further

$$\sigma^{ij}\sim(-\alpha)^{j-i}\frac{1-\alpha^{2i}}{1-\alpha^2}, \qquad i\leqq j,$$
$$\partial\alpha_k^*/\partial\beta_j^*\sim-(1-\alpha^2)^2(-\alpha)^{j-1}. \tag{A.9}$$

Using these results we have that

$$v\sim\sum_{i,j=1}^{k}\sigma^{ij}\frac{\partial\alpha_k^*}{\partial\beta_i^*}\frac{\partial\alpha_k^*}{\partial\beta_j^*}$$
$$\sim\frac{(1-\alpha^2)^4}{\alpha^2}\cdot\frac{1}{1-\alpha^2}\sum_{i=1}^{k}\left[\sum_{j=1}^{i}(-\alpha)^{i+j}(-\alpha)^{i-j}(1-\alpha^{2j})\right.$$
$$\left.+\sum_{j=i+1}^{k}(-\alpha)^{i+j}(-\alpha)^{j-i}(1-\alpha^{2i})\right]$$
$$=\frac{(1-\alpha^2)^3}{\alpha^2}\sum_{i=1}^{k}\left(i\alpha^{2i}-\alpha^{2k+2}\frac{1-\alpha^{2i}}{1-\alpha^2}\right)$$
$$\sim\frac{(1-\alpha^2)^3}{\alpha^2}\sum_{i=1}^{\infty}i\alpha^{2i}$$
$$=1-\alpha^2. \tag{A.10}$$

Proofs of (4.4) and (4.6)

Let C_T have components $c_{|i-j|}$ if $|i-j|\leqq 1$ and otherwise equal to 0, and $c_T = (c_1, \ldots, c_k)'$. Then $\tilde{\beta}_T = -C_T^{-1}c_T\to\beta^*$ in probability, and $\operatorname{plim}_{T\to\infty}\tilde{\alpha}_T = \alpha_k^*$. By the argument used above $T^{\frac{1}{2}}(\tilde{\beta}_T-\beta^*)$ and $-T^{\frac{1}{2}}\Sigma^{-1}(c_T+C_T\beta^*)$ have the same limiting distribution. Here $c_T+C_T\beta^*$ has components $c_1+\beta_2^*c_1+\beta_1^*c_0$, $c_i+(\beta_{i-1}^*+\beta_{i+1}^*)c_1+\beta_i^*c_0$, $i = 2, \ldots, k-1$, and $c_k+\beta_{k-1}^*c_1+\beta_k^*c_0$.

The covariance matrix of the limiting normal distribution of $T^{\frac{1}{2}}(\tilde{\beta}_T-\beta^*)$ is approximated for large k by $\Sigma^{-1}A\Sigma^{-1}$, where $A = \lim_{T\to\infty}T\mathscr{E}(c_T+C_T\beta^*)\times(c_T+C_T\beta^*)'$. Using that

$$\lim_{T\to\infty} T\mathscr{E}(c_ic_j-\mathscr{E}c_ic_j) = 4\pi\int_{-\pi}^{\pi}\cos(vi)\cos(vj)$$
$$\times\frac{1+\alpha^2+2\alpha\cos v}{2\pi}\,dv, \tag{A.11}$$

[see, for example, Anderson (1971a)], the result (4.4) can be easily evaluated. We omit those details here.

The proof of (4.6) follows the same lines, because $\tilde{c}_T = (c_1, 0, \ldots, 0)'\to q$ in probability, and $\tilde{c}_T+C_T\beta^*$ has components $(\beta_{i-1}^*+\beta_{i+1}^*)c_1+\beta_i^*c_0$, $i = 1, 2, \ldots, k$, which are similar to those in the preceding case.

References

Akaike, H., 1969, Power spectrum estimation through autoregressive model fitting, Annals Institute of Statistical Mathematics 21, 407–419.

Anderson, T.W., 1971a, The statistical analysis of time series (Wiley, New York).

Anderson, T.W., 1971b, Estimation of covariance matrices with linear structure and moving average processes of finite order, Department of Statistics Technical Report no. 6 (Stanford University, Stanford, CA).

Anderson, T.W., 1975, Estimation by maximum likelihood in autoregressive moving average models in the time and frequency domains, Department of Statistics Technical Report no. 20 (Stanford University, Stanford, CA).

Berk, K.N., 1974, Consistent autoregressive spectral estimates, The Annals of Statistics 2, no. 3, 489–502.

Durbin, J., 1959, Efficient estimation of parameters in moving-average models, Biometrika 46, 306–316.

Hannan, E.J., 1969, The estimation of mixed moving average autoregressive systems, Biometrika 56, 579–593.

Hannan, E.J., 1970, Multiple time series (Wiley, New York).

McClave, J.T., 1973, On the bias of autoregressive approximations to moving averages, Biometrika 60, 599–605.

McClave, J.T., 1974, A comparison of moving average estimation procedures, Communications in Statistics 3, 865–883.

Mentz, R.P., 1975, Asymptotic properties of some estimators in moving average models, Department of Statistics Technical Report no. 21 (Stanford University, Stanford, CA).

Mentz, R.P., 1976, On the inverse of some covariance matrices of Toeplitz type, SIAM Journal on Applied Mathematics 31, 426–437.

Mentz, R.P. and R. Antelo, 1976, Estimation procedures for the first-order moving average model, submitted for publication.

Nelson, C.R., 1974, The first-order moving average process, Journal of Econometrics, 2, 121–141.

Nicholls, D.F., A.R. Pagan and R.D. Terrell, 1975, The estimation and use of models with moving average disturbance terms: A survey, International Economic Review 16, 113–134.

Parzen, E., 1969, Multiple time series modeling, in: P.R. Krishnaiah, ed., Multivariate analysis II (Academic Press, New York) 389–410.

Shaman, P., 1969, On the inverse of the covariance matrix of a first-order moving average, Biometrika 56, 595–608.

Walker, A.M., 1961, Large sample estimation of parameters for moving-average models, Biometrika 48, 343–357.

Whittle, P., 1953, Estimation and information in stationary time series, Arkiv för Matematik 2, 423–434.

The generalized variance of a stationary autoregressive process

T. W. Anderson and Raúl P. Mentz
Stanford University and University of Tucumán, Argentina

Communicated by the Editors

For a stationary autoregressive process of order p and disturbance variance σ^2 it is shown that the determinant of the covariance of $T\,(\geqslant p)$ consecutive random variables of the process is $(\sigma^2)^T \prod_{i,j=1}^{p} (1 - w_i w_j)^{-1}$, where $w_1, \ldots, w_p$ are the roots of the associated polynomial equation.

1. The generalized variance

An autoregressive process $\{y_t\}$ of order p with mean 0 is defined by

$$y_t + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} = u_t, \qquad t = \ldots, -1, 0, 1, \ldots, \tag{1.1}$$

where the u_t are independent random variables with $\mathcal{E}u_t = 0$, $\mathcal{E}u_t^2 = \sigma^2$, $0 < \sigma^2 < \infty$. The stochastic process is stationary and y_t is independent of $u_{t+1}, u_{t+2}, \ldots$ if and only if the β_j are such that the associated polynomial equation

$$b(w) = \sum_{j=0}^{p} \beta_j w^{p-j} = 0, \tag{1.2}$$

where $\beta_0 = 1$, has roots $w_1, \ldots, w_p$ less than 1 in absolute value. The purpose of this paper is to show that the generalized variance of the process is a power of the variance of u_t times $\prod_{i,j=1}^{p} (1 - w_i w_j)^{-1}$.

The covariance sequence of the process is composed of $\sigma_s = \mathcal{E}y_t y_{t+s} = \sigma_{-s}$, $s = 0, 1, \ldots$. Consider a sequence $y_1, \ldots, y_T$ for $T \geqslant p$ constituting a vector $\mathbf{y}_T = (y_1, \ldots, y_T)'$. The covariance matrix of this vector is

$$\mathcal{E}\mathbf{y}_T\mathbf{y}_T' = (\sigma_{i-j}) \equiv \mathbf{\Sigma}_T \equiv \sigma^2 \mathbf{Q}_T\,. \tag{1.3}$$

Received March 22, 1977.
AMS(MOS) subject classification numbers: 60G15 and 62H05.
Key words: Generalized variance, autoregressive process, covariance matrix.
* This research was supported by Office of Naval Research Contract N00014-75-C-0442, Department of Statistics, Stanford University. The U.S. Government's right to retain a nonexclusive royalty-free license in and to copyright covering this paper is acknowledged.

The determinant $|\Sigma_T| = (\sigma^2)^T |Q_T|$ is the *generalized variance* of y_T.

In the Gaussian case the joint density of y_p and $u_{p+1},\ldots,u_T$ $(T \geqslant p)$ is

$$\frac{|Q_p^{-1}|^{1/2}}{(2\pi)^{T/2}\sigma^T} \exp\left\{-\frac{1}{2\sigma^2}\left[y_p' Q_p^{-1} y_p + \sum_{t=p+1}^{T} u_t^2\right]\right\}, \tag{1.4}$$

where here y_p and $u_{p+1},\ldots,u_T$ denote the variables of integration. Since the Jacobian of the transformation from y_p, $u_{p+1},\ldots,u_T$ to y_T is 1, the constant of the density of y_T is the same as of (1.4) and hence $|Q_p^{-1}| = |Q_T^{-1}|$, $T \geqslant p$. (See Walker [9] and Siddiqui [8].) Since $|Q_T| = |Q_p|$ for $T \geqslant p$, we call $|Q_p|$ the *normalized generalized variance* of the process.

For $T \geqslant p$ substitution for u_t from (1.1), $t = p+1,\ldots,T$, into (1.4) to obtain the density of y_T yields the quadratic form $y_T' Q_T^{-1} y_T$, showing that every element of Q_T^{-1} is a second-degree polynomial in $\beta_1,\ldots,\beta_p$ except possibly elements of Q_p^{-1}. However, since the density of $y_1,\ldots,y_T$ is identical to the density of $y_T,\ldots,y_1$, the elements of Q_p^{-1} must be second-degree polynomials in $\beta_1,\ldots,\beta_p$. The components of Q_p^{-1} are therefore polynomials in the roots $w_1,\ldots,w_p$ of degree at most $2p$. Hence, the determinant $|Q_p^{-1}|$ is a polynomial in $w_1,\ldots,w_p$ of degree at most $(2p)^n$.

LEMMA 1. *If*

$$C = (c_{ij}) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ h_1 & h_2 & \cdots & h_p \\ h_1^2 & h_2^2 & \cdots & h_p^2 \\ \vdots & \vdots & & \vdots \\ h_1^{p-1} & h_2^{p-1} & \cdots & h_p^{p-1} \end{pmatrix}, \tag{1.5}$$

then

$$|C| = \prod_{i<j} (h_j - h_i) \tag{1.6}$$

and

$$C_{1k} = (-1)^{k-1}\left(\prod_{i\neq k} h_i\right) \prod_{\substack{i<j \\ i\neq k\neq j}} (h_j - h_i), \tag{1.7}$$

where C_{1k} denotes the cofactor of c_{1k} in C.

Proof. C is a Vandermonde matrix, and $|C|$ and C^{-1} are given, for example, by Hamming [6, Sects. 8.2 and 10.3]. A direct proof of (1.7) using (1.6) is as follows. To form C_{1k} delete row 1 and column k of $|C|$; in the cofactor, factor h_i out of the ith column $(i \neq k)$ to obtain a Vandermonde determinant of order $p-1$. Q.E.D.

LEMMA 2. *The determinant of order p,*

$$D_p \equiv \left|\frac{1}{a_i + b_j}\right| = \frac{\prod_{i<j}(a_j - a_i)(b_j - b_i)}{\prod_{i,j=1}^{p}(a_i + b_j)}. \tag{1.8}$$

Proof. This is Cauchy's determinant; see, for example, Bellman [2, Sect. 11.6, Exercise 1]. A direct proof is as follows. To convert into 0 each element in the first column, except for that in the first row, we subtract from each row an appropriate multiple of the first row. The i,jth element is thus converted into

$$\frac{1}{a_i + b_j} - \frac{1}{a_1 + b_j}\frac{a_1 + b_1}{a_i + b_1} = \frac{a_i - a_1}{a_i + b_1}\frac{b_j - b_1}{a_1 + b_j}\frac{1}{a_i + b_j}, \qquad i,j = 2,\ldots,p. \tag{1.9}$$

The first factor on the right-hand side is common to the ith row and the second to the jth column. Hence,

$$D_p = \frac{1}{a_1 + b_1}\left(\prod_{i=2}^{p}\frac{a_i - a_1}{a_i + b_1}\right)\left(\prod_{j=2}^{p}\frac{b_j - b_1}{a_1 + b_j}\right) D_{p-1}, \tag{1.10}$$

and the result follows. Q.E.D.

THEOREM. *For $\beta_1,\ldots,\beta_p$ such that the roots of* (1.2) *are less than* 1 *in absolute value for $T \geqslant p$*

$$|\Sigma_T| = (\sigma^2)^T \prod_{i,j=1}^{p} (1 - w_i w_j)^{-1}. \tag{1.11}$$

Proof. We first consider the case where $w_1,\ldots,w_p$ are different and different from 0. If $h_i = w_i^{-1}$, then in (1.5) $|C| \neq 0$. Anderson [1] in (24) of Sect. 5.3 gives an expression for the elements of Σ_p in terms of $w_1,\ldots,w_p$. (See also Problem 27 of Chap. 5.) Then

$$|Q_p| = \frac{C_{11}^2 \cdots C_{1p}^2}{|C|^{2p-2}} |V|, \tag{1.12}$$

where

$$|V| = \left|\frac{1}{1 - w_i w_j}\right| = \left|\frac{w_i^{-1}}{w_i^{-1} + (-w_j)}\right| = \frac{1}{\prod_{i=1}^{p} w_i}\left|\frac{1}{w_i^{-1} + (-w_j)}\right|$$

$$= \frac{1}{\prod_{i=1}^{p} w_i}\frac{\prod_{i<j}(w_j^{-1} - w_i^{-1})(w_i - w_j)}{\prod_{i,j=1}^{p}(w_i^{-1} - w_j)}$$

$$= \left(\prod_{i=1}^{p} w_i\right)^{p-1}\frac{\prod_{i<j}(w_i - w_j)^2 w_i^{-1} w_j^{-1}}{\prod_{i,j=1}^{p}(1 - w_i w_j)} = \frac{\prod_{i<j}(w_i - w_j)^2}{\prod_{i,j=1}^{p}(1 - w_i w_j)}. \tag{1.13}$$

Further,

$$\frac{C_{11}^2 \cdots C_{1p}^2}{|\mathbf{C}|^{2p-2}} = \left[\frac{(\prod_{i=1}^p h_i^{p-1}) \prod_{i<j}(h_j - h_i)^{p-2}}{\prod_{i<j}(h_j - h_i)^{p-1}}\right]^2$$

$$= \frac{\prod_{i=1}^p h_i^{2p-2}}{\prod_{i<j}(h_j - h_i)^2} = \frac{1}{\prod_{i<j}(w_i - w_j)^2}, \tag{1.14}$$

and (1.11) follows for the roots different and nonzero. The determinant $|\mathbf{Q}_p^{-1}|$ is the polynomial $\prod_{i,j=1}^p (1 - w_i w_j)$, which holds for all w_j such that $|w_j| < 1$, $j = 1,..., p$. Q.E.D.

2. Discussion

1. Grenander and Szegö [5] in effect showed that $\lim_{T\to\infty} |\mathbf{Q}_T| = \prod_{i,j=1}^p (1 - w_i w_j)^{-1}$ by use of an integral of $|b(w)|^{-2}$ (p. 78) and that $|\mathbf{Q}_T|$ for $T \geqslant p$ is equal to this limit (p. 71). However, they did not relate these results to the generalized variance of the autoregressive process. Also, Walker [9] noted that $|\mathbf{Q}_T| = |\mathbf{Q}_p| = 1/|\mathbf{Q}_p^{-1}|$ for $T \geqslant p$. (See also Finch [4].)

2. If the process is Gaussian, the normalizing constant in the normal density of $\mathbf{y}_T$ is $(2\pi)^{-T/2}$ times

$$|\mathbf{\Sigma}_T|^{-1/2} = (\sigma^2)^{-T/2} \prod_{i,j=1}^p (1 - w_i w_j)^{1/2}. \tag{2.1}$$

3. If one or more of the roots approaches 1 in absolute value, $|\mathbf{\Sigma}_T^{-1}| \to 0$ and $|\mathbf{\Sigma}_T| \to \infty$. These facts agree with the nonexistence of a nontrivial stationary process satisfying (1.1) if one or more roots are equal to 1 in absolute value.

4. A moving average model of order q is defined by

$$x_t = v_t + \alpha_1 v_{t-1} + \cdots + \alpha_q v_{t-q}, \tag{2.2}$$

where the v_t are independent random variables with $\mathcal{E}v_t = 0$, $\mathcal{E}v_t^2 = \tau^2$, $0 < \tau^2 < \infty$. The associated polynomial equation

$$z^q + \alpha_1 z^{q-1} + \cdots + \alpha_q = 0 \tag{2.3}$$

has roots $z_1,..., z_q$. Durbin [3] conjectured that if $\tau^2\mathbf{N}_T$ is the covariance matrix of $x_1,..., x_T$ generated by (2.2), and if all roots of (1.2) are less than 1 in absolute value, then

$$\lim_{T\to\infty} |\mathbf{N}_T| = |\mathbf{Q}_n|, \tag{2.4}$$

for some n sufficiently large compared with $p = q$ when $\alpha_j = \beta_j$, $j = 1,..., p$. Finch [4] showed that

$$\lim_{T\to\infty} |\mathbf{N}_T| = \lim_{T\to\infty} |\mathbf{Q}_T| \tag{2.5}$$

by use of some results of Grenander and Szego [5] and gave explicitly the limiting value of the generalized variance for an autoregressive moving average process. Walker [9] used more algebraic methods to show (2.4) for $n = p = q$. As an example, these results for $p = q = 1$ are $Q_1 = 1/(1 - \beta_1^2)$ and $|\mathbf{N}_T| = (1 - \alpha_1^{2T+2})/(1 - \alpha_1^2) \to 1/(1 - \alpha_1^2)$ as $T \to \infty$. Durbin [3] considered the case $p = q = 2$ in detail.

For further discussion, see the recent paper by Shaman [7].

References

[1] Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

[2] Bellman, R. (1960). *Introduction to Matrix Analysis.* McGraw-Hill, New York.

[3] Durbin, J. (1959). Efficient estimation of parameters in moving-average models. *Biometrika* **46** 306–316.

[4] Finch, P. D. (1960). On the covariance determinants of moving-average and autoregressive models. *Biometrika* **47** 194–196.

[5] Grenander, U. and Szegö, G. (1958). *Toeplitz Forms and Their Applications.* Univ. of California Press, Berkeley/Los Angeles.

[6] Hamming, R. W. (1962). *Numerical Methods for Scientists and Engineers.* McGraw-Hill, New York.

[7] Shaman, P. (1976). Approximations for stationary covariance matrices and their inverses with application to ARMA models. *Ann. Statist.* **4**, No. 2, 292–301.

[8] Siddiqui, M. M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. *Ann. Math. Statist.* **29** 585–588.

[9] Walker, A. M. (1961). On Durbin's formula for the limiting generalized variance of a sample of consecutive observations from a moving-average process. *Biometrika* **48** 197–199.

Printed by the St Catherine Press Ltd., Tempelhof 37, Bruges, Belgium.

Modelos de regresión para estimaciones estacionales

en Métodos Cuantitativos en las Ciencias Sociales

Raúl Pedro Mentz y Víctor Jorge Elías

Ensayos en memoria del Profesor Dr. José Raúl Souto

1979, pags. 77-92, Ediciones Machi

Modelos de regresión para estimaciones estacionales

Raúl Pedro Mentz y Víctor Jorge Elías

Introducción

El enfoque de regresión para tratar la estacionalidad es frecuente en el análisis econométrico, cuando el principal objetivo es estimar los parámetros de relaciones económicas en base a datos mensuales, trimestrales, etc. En tal caso el tratamiento de las estimaciones estacionales está generalmente condicionado por el deseo de lograr "buenas" estimaciones de los principales parámetros, los de la relación económica.

Cuando se quiere utilizar el enfoque de regresión para estimaciones de tendencia [1] y estacionalidad que sirvan para múltiples usos y que no estén vinculados por un modelo particular, surgen a menudo requerimientos distintos. Por ejemplo, se desea conseguir estimaciones distinguibles de los componentes estacionales y de la tendencia, o que los distintos componentes sean de

* Los autores son profesores de Estadística y de Econometría, respectivamente, en la Facultad de Ciencias Económicas, Universidad Nacional de Tucumán, República Argentina.

La versión en inglés del presente trabajo aparecerá en el número 113 de la revista *Estadística* del Instituto Interamericano de Estadística.

[1] Nuestro concepto de tendencia incluye a todos los movimientos suaves de la serie; en algunos casos puede ser de interés separar de este componente los "movimientos cíclicos", pero no consideraremos tal posibilidad.

fácil interpretación. Esto puede resultar importante para la elaboración de indicadores en un esquema de partes componentes, o para cotejar las estimaciones con las proporcionadas por otros métodos.

En este trabajo analizaremos los modelos de regresión desde el último punto de vista señalado y trataremos de establecer relaciones y diferencias entre varias alternativas. El objetivo es descubrir cuáles son equivalentes y fijar criterios prácticos para decidir sobre la adopción de alguna de ellas.

Dada una serie cronológica observada $\{y_t: t = 1, 2, \ldots, T\}$ con datos mensuales, una propuesta para estimar los componentes de tendencia y estacionalidad simultáneamente, se expresa mediante un modelo de regresión lineal múltiple como

$$(1)\quad y_t = \alpha_0 + \alpha_1 t + \ldots + \alpha_m t^m + \beta_1 D_{1t} + \beta_2 D_{2t} + \ldots + \beta_{12} D_{12,t} + u_t, \; t = 1, 2, \ldots, T,$$

donde el polinomio en t de grado m ($m \geqslant 1$) representa a la tendencia, $\alpha_0, \alpha_1, \ldots, \alpha_m, \beta_1, \beta_2, \ldots, \beta_{12}$ son los $m + 13$ parámetros desconocidos de interés, $\{u_t\}$ es un conjunto de variables aleatorias que para todos los valores de sus subíndices satisfacen $E(u_t) = 0$, $E(u^2_t) = \sigma^2$, $E(u_t u_s) = 0$ cuando $t \neq s$. D_{st} para $s = 1, 2, \ldots, 12$, es una variable cualitativa ("dummy" en inglés) que toma el valor 1 cuando el período t corresponde al mes s, y 0 cuando no corresponde. Se toma a m como conocido.

Por razones de simplicidad, tomaremos a $T = 12N$, notando que desde el punto de vista computacional o desde el teórico, el caso general sólo hace un poco más compleja la escritura de algunas fórmulas.

La expresión (1) se interpreta como el deseo de ajustar a los datos 13 polinomios en t de grado m, todos paralelos, uno que es el promedio y tiene como constante a α_0 y los restantes, uno para cada subconjunto de N datos correspondientes al mes s, tienen constantes $\alpha_0 + \beta_s$, $s = 1, 2, \ldots, 12$.

Notamos que éste es un caso de "estacionalidad constante", caracterizado porque los β_s no dependen del período t.

En el modelo (1) se verifica que.

$$(2) \qquad 1 = D_{1t} + D_{2t} + D_{12,t}, \qquad t = 1, 2, \ldots, T,$$

de manera que la matriz de regresores no tiene rango completo $m + 13$, sino rango $m + 12$.

Nuestro propósito es analizar y comparar varias propuestas para resolver este problema. Una posibilidad que no consideramos es resolver el sistema de ecuaciones normales obtenidas al intentar la estimación de (1) por mínimos cuadrados, utilizando alguna definición de matriz in ersa generalizada.

1. La solución como en el análisis de la varianza

Jorgenson (1964) propuso una solución paralela a la que se adopta habitualmente en el análisis de la varianza, esto es, imponer a (1) la restricción adicional

$$(3) \qquad \sum_{s=1}^{12} \beta_s = 0.$$

En tales condiciones (1) puede escribirse como

$$(4) \qquad y_t = A_0 + A_1 t + \ldots + A_m t^m + B_1 D^\circ_{1t} + \ldots + B_{11} D^\circ_{11,t} + u^\circ_t, \qquad t = 1, 2, \ldots, T,$$

donde D°_{st} toma el valor 1 cuando el período t corresponde al mes s, -1 cuando corresponde al mes 12 (diciembre) y 0 en los diez casos restantes.

Los sumandos $B_s D^\circ_{st}$ se interpretan como desvíos con respecto a la tendencia, y Jorgenson (pág. 694) hace notar que bajo la hipótesis (3) "las sumas anuales de los datos sin ajustar y ajustados por estacionalidad son iguales".

2. Omisión de la constante

Para resolver la estimación de (1), se propone a veces eliminar la constante, y en consecuencia escribir el modelo como

$$(5) \quad y_t = A^\circ_1 t + \ldots + A^\circ_m t^m + B^\circ_1 D_{1t} + \ldots + B^\circ_{12} D_{12,t} + v_t. \qquad t = 1, 2, \ldots, T.$$

Ver por ejemplo Johnston (1972).

Este modelo puede interpretarse como proponiendo doce regresiones simultáneas donde para el mes s la ecuación tiene los regresores $t, t^2, \ldots, t^m$ y la constante B°_s. No es posible estimar a la constante de la tendencia polinomial promedio, o interpretamos que se la toma en principio como arbitrariamente igual a cero. Una posibilidad sería declarar su valor $A^\circ_0 = y_1$, el primer valor observado; o bien

$$A^\circ_0 = (1/12) \sum_{s=1}^{12} B^\circ_s.$$

Cualquiera de estas hipótesis reemplazaría entonces a la (3) para poder estimar a los $13 + m$ parámetros de (1).

3. Omisión de un mes

En lugar de (5) a veces se estiman los parámetros de

$$(6) \quad y_t = A^{\circ\circ}_0 + A^{\circ\circ}_1 t + \ldots + A^{\circ\circ}_m t^m + B^{\circ\circ}_1 D_{1t} + \ldots + B^{\circ\circ}_{11} D_{11,t} + w_t, \qquad t = 1, 2, \ldots, T,$$

donde por conveniencia hemos omitido el mes doce. Por ejemplo se prefiere esta solución con respecto a (5) cuando se desea utilizar un programa de computación que exige el uso de una constante. Ver por ejemplo, Rao and Miller (1971), Sección 4.7.

4. Comparación de las tres alternativas

En esta sección comparamos los modelos (4), (5) y (6).

4.1. Relaciones algebraicas

Consideremos primeramente los modelos (4) y (5). Para el caso en que los respectivos parámetros se estiman por mínimos cuadrados Goldberger [(1964), Capítulo 5] propuso un método para comparar los estimadores resultantes de las ecuaciones normales. Las primeras 12 filas de la matriz de regresores para (4), cuando éstos se ordenan colocando a la constante entre t^m y D°_{1t}, forman la matriz

$$(7)\quad (\underset{\sim}{Q} \mid \underset{\sim}{J}) = \left(\begin{array}{cccc|ccccc} 1 & 1 & \dots & 1 & 1 & 1 & 0 & \dots & 0 \\ 2 & 2^2 & \dots & 2^m & 1 & 0 & 1 & \dots & 0 \\ 3 & 3^2 & \dots & 3^m & 1 & 0 & 0 & \dots & 0 \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot & & \cdot \\ 11 & 11^2 & \dots & 11^m & 1 & 0 & 0 & \dots & 1 \\ 12 & 12^2 & \dots & 12^m & 1 & -1 & -1 & \dots & -1 \end{array}\right).$$

mientras que para (5) la correspondiente matriz es $(\underset{\sim}{Q} \mid \underset{\sim}{I})$. Resulta entonces que

$$(8)\quad (\underset{\sim}{Q}_{12\times m}\ \underset{\sim}{J}_{12\times 12}) = (\underset{\sim}{Q}\ \underset{\sim}{I}_{12}) \begin{pmatrix} \underset{\sim}{I}_m & \underset{\sim}{Q}_{m\times 12} \\ \underset{\sim}{Q}_{12\times m} & \underset{\sim}{J} \end{pmatrix},$$

y

$$(9)\quad \begin{pmatrix} \hat{A}^\circ_1 \\ \cdot \\ \cdot \\ \hat{A}^\circ_m \\ \hat{B}^\circ_1 \\ \cdot \\ \cdot \\ \hat{B}^\circ_{12} \end{pmatrix} \begin{pmatrix} \underset{\sim}{I} & \underset{\sim}{Q} \\ \underset{\sim}{Q} & \underset{\sim}{J} \end{pmatrix} \begin{pmatrix} \hat{A}_1 \\ \cdot \\ \cdot \\ \hat{A}_m \\ \hat{A}_0 \\ \hat{B}_1 \\ \cdot \\ \cdot \\ \hat{B}_{11} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \hat{A}_1 \\ \cdot \\ \cdot \\ \hat{A}_m \end{pmatrix} \\ \begin{pmatrix} \hat{A}_0 \\ \hat{B}_1 \\ \cdot \\ \cdot \\ \hat{B}_{11} \end{pmatrix} \end{pmatrix}.$$

En términos de componentes (9) asegura que las estimaciones de los coeficientes de t, $t^2, \dots t^m$ por mínimos cuadrados son iguales en (4) y (5), mientras que las restantes estimaciones están relacionadas por

$$(10)\quad \begin{aligned} \hat{B}^\circ_1 &= \hat{A}_0 + \hat{B}_1, \\ &\cdot \qquad\quad \cdot \\ &\cdot \qquad\quad \cdot \\ &\cdot \qquad\quad \cdot \\ \hat{B}^\circ_{11} &= \hat{A}_0 + \hat{B}_{11}, \\ \hat{B}^\circ_{12} &= \hat{A}_0 - \hat{B}_1 - \dots - \hat{B}_{11} = \hat{A}_0 + \hat{B}_{12}, \end{aligned}$$

esto último pues la suma de los $\hat{B}_s$ satisface (3).

Sumando las ecuaciones en (10) se despeja $\hat{A}_0$, de manera que

$$(11)\quad \begin{aligned} \hat{A}_0 &= (1/12)\,(\hat{B}^\circ_1 + \dots + \hat{B}^\circ_{12}), \\ \hat{B}_1 &= \hat{B}^\circ_1 - \hat{A}_0 = (1/12)\,(11B^\circ_1 - B^\circ_2 - \dots - B^\circ_{12}), \\ &\cdot \qquad\qquad\qquad \cdot \\ &\cdot \qquad\qquad\qquad \cdot \\ &\cdot \qquad\qquad\qquad \cdot \\ \hat{B}_{11} &= \hat{B}^\circ_{11} - \hat{A}_0 = (1/12)\,(-B^\circ_1 - B^\circ_2 - \dots - B^\circ_{10} + \\ &\qquad + 11B^\circ_{11} - B^\circ_{12}). \end{aligned}$$

Esto demuestra que

$$(12)\quad J^{-1} = (1/12)\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 11 & -1 & -1 & \dots & -1 \\ -1 & 11 & -1 & \dots & -1 \\ & & \cdot & & \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ -1 & -1 & -1 & \dots & -1 \end{pmatrix}.$$

La primera igualdad en (11) expresa que si para recuperar la constante de la tendencia polinómica promedio en (5) utilizamos el promedio de los coeficientes de las variables estacionales, como se sugirió en la sección 2, estableceremos la equivalencia con el modelo (4), en cuanto al componente tendencial. La equivalencia será total si además ajustamos las componentes estacionales de acuerdo con (14).

Otra manera de expresar la relación de las partes tendenciales, es que (4), que reconoce la restricción (3), es equivalente a (5) con la estimación

$$\hat{A}^{\circ}_{0} = (1/12) \sum_{s=1}^{12} \hat{B}^{\circ}_{s}.$$

En vez de derivar la relación entre los estimadores por mínimos cuadrados, una alternativa interesante (y equivalente) es tratar de relacionar los respectivos parámetros.

Comenzamos observando que (2) es válida y que

$$(13)\quad D^{\circ}_{st} = D_{st} - D_{12,t},$$

para todos los valores de s y de t. Reemplazando en (4) obtenemos

$$\begin{aligned} y_t &= A_0 (D_{1t} + \dots + D_{12,t}) + A_1 t + \dots + A_m t^m + \\ &\quad + B_1 (D_{1t} - D_{12,t}) + \dots + B_{11} (D_{11,t} - D_{12,t}) + u^{\circ}_{t} \\ &= A_1 t + \dots + A_m t^m + (B_1 + A_0) D_{1t} + \dots + (B_{11} + A_0) \\ &\quad D_{11,t} + (A_0 - B_1 - \dots - B_{11}) D_{12,t} + u^{\circ}_{t}. \end{aligned} \tag{14}$$

Al igualar este modelo resultante con (5), se leen las relaciones (10) entre los respectivos parámetros. Recíprocamente, como

$D_{12,t} = (1/12)\,(1 - D^{\circ}_{1t} - \dots - D^{\circ}_{11,t})$, resulta que

$$D_{st} = D^{\circ}_{st} + D_{12,t} = D^{\circ}_{st} + (1/12)\,(1 - D^{\circ}_{1t} - \dots - D^{\circ}_{11,t}); \tag{15}$$

reemplazando en (5) e igualando a (4), se leen las relaciones (11) entre los respectivos parámetros.

Nótese que encontrar la relación (15) equivale a invertir la matriz J y recíprocamente, encontrar la relación (13) equivale a invertir la matriz J^{-1}.

Un subproducto interesante de este análisis es que la matriz Q de regresores correspondientes a las potencias de t, no participa en el análisis, de manera que los resultados finales son válidos para cualquier conjunto de m regresores linealmente independientes, y linealmente independientes de la constante y las variables estacionales incluidas. Claramente este resultado es válido para las comparaciones entre cualquier par de las propuestas que analizamos en esta sección.

Para comparar (5) y (6) procedemos de la misma manera. Sustituyendo (2) en (6) obtenemos

$$\begin{aligned} (16)\; y_t &= A^{\circ\circ}_{1} t + \dots + A^{\circ\circ}_{m} t^m + (B^{\circ\circ}_{1} + A^{\circ\circ}_{0})\, D_{1t} + \dots + \\ &\quad (B^{\circ\circ}_{11} + A^{\circ\circ}_{11}) D_{11,t} + A^{\circ\circ}_{0} D_{12,t} + w_t, \end{aligned}$$

que es de la forma (5). Por el contrario, si despejamos $D_{12,t}$ en (2) y la sustituimos en (5), obtenemos

$$\begin{aligned} (17)\quad y_t &= B^{\circ}_{12} + A^{\circ}_{1} t + \dots + A^{\circ}_{m} t^m + (B^{\circ}_{1} - B^{\circ}_{12})\, D_{1t} \\ &\quad + \dots + (B^{\circ}_{11} - B^{\circ}_{12})\, D_{11,t} + v_t, \end{aligned}$$

que es de la forma (6). En términos matriciales hemos obtenido

$$(18)\quad \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & 1 & \dots & 0 \\ . & . & . & . & & . \\ . & . & . & . & & . \\ . & . & . & . & & . \\ 1 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & 0 & \dots & 0 & -1 \\ 0 & 1 & 0 & \dots & & -1 \\ . & . & . & & . & . \\ . & . & . & & . & . \\ . & . & . & & . & . \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}$$

Las relaciones entre las estimaciones por mínimos cuadrados en (5) y (6) son por lo tanto las siguientes donde repetimos una parte de (10) y otra de (11) para facilitar la comparación:

$$(19)\quad \begin{aligned} &\hat{B}^{\circ}{}_1 = \hat{A}^{\circ\circ}{}_0 + \hat{B}^{\circ}{}_1 = \hat{A}_0 + \hat{B}_1 \\ &. \qquad\qquad . \qquad\qquad . \\ &. \qquad\qquad . \qquad\qquad . \\ &. \qquad\qquad . \qquad\qquad . \\ &\hat{B}^{\circ}{}_{11} = \hat{A}^{\circ\circ}{}_0 + \hat{B}^{\circ\circ}{}_{11} = \hat{A}_0 + \hat{B}_{11}, \\ &\hat{B}^{\circ}{}_{12} = \hat{A}^{\circ\circ}{}_0 \qquad\quad = \hat{A}_0 + \hat{B}_{12}, \end{aligned}$$

y

$$\begin{aligned} &\hat{B}^{\circ\circ}{}_0 = \hat{B}^{\circ}{}_{12} \qquad\quad = \hat{A}_0 + \hat{B}_{12}, \\ &\hat{B}^{\circ\circ}{}_1 = \hat{B}^{\circ}{}_1 - \hat{B}^{\circ}{}_{12} = \hat{B}_1 - \hat{B}_{12}, \end{aligned}$$

$$(20)\quad \begin{aligned} &. \qquad\qquad . \qquad\qquad . \\ &. \qquad\qquad . \qquad\qquad . \\ &. \qquad\qquad . \qquad\qquad . \\ &\hat{B}^{\circ\circ}{}_{11} = \hat{B}^{\circ}{}_{11} - \hat{B}^{\circ}{}_{12} = \hat{B}_{11} - \hat{B}_{12}, \end{aligned}$$

En (20) se interpreta que al incorporar la constante $A^{\circ\circ}{}_0$ en (6), sólo se pueden estimar las diferencias entre los efectos estacionales y el del mes omitido, que sirve entonces de base; se ve que la relación subsiste cuando se incorpora la constante y se impone la relación (3). Interpretaciones similares valen para (19).

4.2. Propiedades estadísticas

Del análisis de las propuestas de las secciones 1, 2 y 3 resulta que los parámetros de los modelos (4), (5) y (6) están relacionados unívocamente por expresiones lineales como (10), (11), (19) o (20).

Como los tres modelos tienen rango completo e igual a m + 12, los estimadores por mínimos cuadrados en cada caso son únicos y mejores lineales insesgados de los respectivos parámetros, por el Teorema de Gauss y Markov. Como los parámetros están además relacionados linealmente en la forma señalada en el párrafo precedente, utilizando cualquier formulación para el cómputo, los estimadores de cualquier conjunto de parámetros que se obtenga mediante las relaciones derivadas a tal efecto, siguen teniendo las propiedades mencionadas (of. Scheffé (1959), pág. 14, Teorema 2).

Desde el punto de vista de predecir valores de y, las relaciones del tipo (14), (16) o (17) muestran que los tres modelos, una vez estimados sus parámetros por mínimos cuadrados, proporcionarán idénticos predictores.

Finalmente, si suponemos que los errores aleatorios tienen distribuciones normales, los estimadores por mínimos cuadrados son los de máxima verosimilitud y son únicos, mejores insesgados en general. Como los estimadores son de máxima verosimilitud, la propiedad es extensible a los distintos conjuntos de parámetros, no importa con cuál se computen las estimaciones, pues los estimadores son invariantes a transformaciones lineales no-singulares (entre otras) de los parámetros.

Los procedimientos inferenciales, estimaciones de intervalo y tests de hipótesis, para el caso normal están detallados en Jorgenson (1964).

4.3. Conclusión

Resulta entonces de lo expuesto, que la elección de cuál de los tres modelos se utilizará para la estimación, es estadísticamente irrelevante. El modelo (4) tiene la ventaja de dar una estimación completa de la tendencia sin necesidad de nuevas elaboraciones, y de que los coeficientes estacionales resultan comparables de inmediato con los obtenidos por métodos como el de "diferencias de promedio móvil", cuando en este último la suma de los componentes estacionales se hace igual a cero. Resulta entonces recomendable el uso de este modelo.

Nótese por ejemplo que en (6), al omitir uno de los meses, $A^{\circ\circ}_0$ no es la constante de la tendencia. No es aconsejable tampoco utilizar a los B°_s del modelo (5) como componentes estacionales, pues $\sum_{s=1}^{12} B^{\circ}_s = 12A_0$, que en general no será igual a cero y en tal caso determinará que las sumas anuales de los datos corregidos por estacionalidad sean diferentes.

4. Observación

Una situación paralela a la que analizamos en las secciones precedentes, ocurre en el análisis de la varianza. Notamos que el modelo (1) es, desde este punto de vista, un modelo de "análisis de la covarianza".

Un enfoque clásico del problema del análisis de la varianza es mediante una formulación del tipo de (1) con la restricción (3); ésta es, por ejemplo la alternativa más utilizada en el libro de Scheffé (1959). Sin embargo cuando se quiere explotar la estrecha relación entre los modelos lineales del análisis de la varianza con los de regresión (que son habitualmente de rango completo) se ha utilizado a menudo la formulación (5) (por ejemplo en Rao (1965), secciones 4d y siguientes) o la formulación (6) (por ejemplo en Mendenhall [1968]).

Graybill ([1961], sección 11.2.3) demuestra expresamente que todo modelo lineal puede reparametrizarse de manera que el resultante tenga rango completo, y que para cualquier reparametrización los estimadores insesgados de las distintas combinaciones lineales de los parámetros, son idénticos. Desde este punto

de vista nuestro resultado consiste en haber encontrado la relación explícita entre varias reparametrizaciones de rango completo del modelo (1).

5. Estimación en dos etapas

Los procedimientos de las secciones 1, 2 y 3 serán considerados en esta sección como operando en una etapa, frente a alternativas en dos etapas como la que describimos a continuación.

5.1. Mínimos cuadrados en dos etapas

Para estimar los $13 + m$ parámetros de (1) se puede pensar en regresar

$$(21)\quad y_t = A'_0 + A'_1 t + \ldots + A'_m t^m + e_t \qquad T = 1, 2, \ldots, T,$$

y después ajustar a los residuos de la estimación por mínimos cuadrados las constantes estacionales:

(22)

$$y_t - \hat{A}'_0 - \hat{A}'_1 t - \ldots - \hat{A}'_m t^m = B'_1 D_{1t} + \ldots + B_{12} D_{12,t} + e'_t,$$
$$t = 1, 2, \ldots, T.$$

En (22) los regresores son ortogonales ($\sum_{t=1}^{T} D_{1t} D_{jt} = N\delta_{ij}$) y en consecuencia cada B'_s se estima por mínimos cuadrados como si se plantearan doce regresiones.

$$(23)\quad y_t - \hat{A}'_0 - \hat{A}'_1 t - \ldots - \hat{A}'_m t^m = B'_s D_{st} + e''_{st},$$
$$s = 1, 2, \ldots, 12,$$

basada cada una en los datos correspondientes al mes s. Resulta entonces que

$$(24)\quad \hat{B}'_s = \frac{1}{N} \sum_{(s)} (y_t - \hat{A}'_0 - \hat{A}'_1 t - \ldots - \hat{A}'_m t^m),$$
$$s = 1, 2, \ldots, 12,$$

el promedio aritmético simple de los residuos para el mes, correspondientes a los N años respectivos.

Como en (4) las variables D°_{st} no son ortogonales con respecto a todos los restantes regresores, (24) diferirá en general de B_s estimado en (4) por mínimos cuadrados. Obsérvese sin embargo que $\Sigma^{12}_{s=1} B'_s = 0$ cuando $A'_0 \neq 0$, pues la suma de (24) para todos los t, $t = 1, 2, \ldots, T$, es entonces la ecuación normal correspondiente a derivar con respecto a A'_0 la función de criterio $\sum_{t=1}^{T} (y_t - A'_0 - A'_1 t - \ldots - A'_m t^m)^2$, y aquélla es igual a cero como consecuencia del método de estimación utilizado.

Bajo esta propuesta el planteo de (1) como un verdadero problema de regresión para tendencia y estacionalidad simultáneamente es en alguna medida sólo aparente. Sin embargo Jorgenson (op. cit. págs. 697-698) demostró que si la remoción de la tendencia en la primera etapa es total y se efectúa por un "filtro" o procedimiento lineal (matricial), el respectivo procedimiento en dos etapas, similar al que describimos precedentemente, es equivalente a la solución de la sección 1, y por lo tanto a las otras dos propuestas en una etapa.

En conclusión: si el modelo (1) fuese el verdadero, si se deseara utilizar una formulación simultánea para tendencia y estacionalidad, y se juzgaran relevantes las propiedades estadísticas de los estimadores detalladas en la sección 4.2, el método en una etapa de la sección 1 (o sus equivalentes) sería preferible a la propuesta en dos etapas de esta sección.

5.2. El método de diferencias del promedio móvil

La formulación en dos etapas que hemos presentado en la sección precedente puede utilizarse para clarificar y reinterpretar el método tradicional de "diferencias del promedio móvil". En este esquema se parte de una propuesta general del tipo

$$(25)\quad y_t = T_t + \beta'_1 D_{1t} + \ldots + \beta'_{12} D_{12,t} + z_t, \quad t = 1, 2, \ldots, T,$$

donde T_t es la tendencia de la serie, no necesariamente parametrizada como en (1). Se propone estimar a T_t por suavizado de

la serie observada mediante promedios móviles (que es también un procedimiento no-paramétrico), habitualmente "centrado de doce meses":

$$(26)\quad \hat{T}_t = (1/24)\,(y_{t-6} + 2\sum_{j=-5}^{5} y_{t+j} + y_{t+6}), \quad t = 7, 8, \ldots, T-6,$$

y luego a los β'_s por una fórmula equivalente a (24),

$$(27)\quad \hat{\beta}'_s \frac{1}{N-1} \Sigma'_{(s)} (y_t - \hat{T}_t), \qquad s = 1, 2, \ldots, 12,$$

donde se omiten de las sumas el primer valor para los meses correspondientes a los períodos $t = 1, 2, \ldots, 6$ y el último para $t = T-5, T-4, \ldots, T$ (en caso que el período considerado conste de años enteros, éstos son los primeros y últimos seis meses, respectivamente).

En primer lugar, la propiedad (3) no se cumple en general para los β'_s, por lo que habitualmente se propone reajustarlos previamente para que la satisfagan.

El método de estimación de la tendencia (26) ha demostrado empíricamente ser bastante apropiado y se espera que la alternativa (21) conduzca a valores de m comparativamente elevados. Por ejemplo Johnston [(1972), sección 6.2] encontró que $m = 6$ fue necesario en un caso concreto. Para un análisis crítico del uso de promedios móviles en este contexto ver Durbin (1963).

Con relación al resultado de Jorgenson ya mencionado, puede interpretarse a este método como una aproximación a los requerimientos de que el "filtrado" de la tendencia sea total y producido por una tranformación lineal.

5.3. Los modelos de regresión y el método de diferencias del promedio móvil en la práctica

La formulación de la sección 5.2 puede tomarse como un esfuerzo pedagógico para relacionar el método de diferencias del promedio móvil, muchas veces presentado sin justificación estadística alguna, con el modelo de regresión lineal múltiple.

Sin embargo esta explicación deja sin aclarar la popularidad de aquel método, frente a los procedimientos aparentemente exactos de las secciones 1, 2 y 3. La clave está en el tratamiento empírico de m, el grado del polinomio utilizado en (1) [1]. Desde el punto de vista práctico m no es un dato adicional del problema, sino un parámetro destinado a estimarse en base a los mismos datos disponibles. Si consideramos un grupo de series cronológicas a las que debe ajustarse el modelo (1), debe resolverse la elección del grado m en cada caso y una mala elección puede conducir a errores de especificación importantes.

Una posibilidad es aplicar procedimientos como el descripto en Anderson [(1971), sección 3.2.2] para decidir sobre el valor de m, que gocen por su parte de buenas propiedades estadísticas. Sin embargo, si consideramos en conjunto al procedimiento consistente en decidir el valor de m y luego estimar los $12 + m$ parámetros de (4), no es inmediato que los estimadores resultantes tengan las propiedades estadísticas de la sección 4.2.

Una alternativa práctica es entonces aproximar no-paramétricamente a la tendencia y confiar que para muchas de las series observadas el procedimiento resultante tenga propiedades por lo menos tan buenas como el procedimiento paramétrico combinado aludido en el párrafo anterior, aún cuando las series respondan efectivamente a modelos como (1), con m desconocido.

Este razonamiento puede contribuir a aclarar en alguna medida el aparente éxito logrado por métodos como el de diferencias del promedio móvil y otras variantes y refinamientos suyos, muchos de evidente extracción empírica y que carecen de una justificación estadística explícita aceptable. Además el argumento contribuye a anticipar que no es imperioso que en la práctica estos métodos de extracción empírica sean superados netamente por los basados en el modelo de regresión lineal múltiple.

[1] Consideramos poco importante la restricción impuesta por la forma polinomial, siempre que m quede libre.

Referencias

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, Wiley

Durbin J. (1963). "Trend elimination for the purpose of estimating seasonal and periodic components of time series", en *Proceedings of the Symposium on Time Series Analysis*, M. Rosemblatt, editor, Wiley, pp. 3-16.

Goldberger, A. S. (1964), *Econometric Theory*, Wiley.

Graybill, F. A. (1961), *An Introduction to Linear Statistical Models, Volume I*, McGraw-Hill.

Johnston, J. (1972), *Econometric Methods, 2nd. Edition*, McGraw-Hill.

Jorgenson, D. W. (1964), "Minimum linear, unbiased seasonal adjustment of economic time series". *Journal of the American Statistical Association*. Vol. 59, pp. 681-724.

Mendenhall, W. (1968). *Introduction to Linear Models and the Design and Analysis of Experiments*. Duxbury Press.

Rao, C. R. (1965), *Linear Statistical Inference and Its Applications*. Wiley

Rao, P. y R. L. Miller (1971), *Applied Econometrics*, Wadssworth.

Scheffé, H. (1959), *The Analysis of Variance*. Wiley.

Análisis of lead-lag relations by cross-spectral methods

Raúl Pedro Mentz

Contributed Papers, International Statistical Institute
42nd. session, December 4-14, 1979, pp. 341-344

Analysis of lead-lag relations by cross-spectral methods

Raúl Pedro Mentz
University of Tucumán, Argentina

1. Introduction

The joint spectral analysis of pairs of time series has been developed with considerable detail in the literature; see, for example, Hannan (1970), Jenkins and Watts (1968), or Koopmans (1974). For two jointly (weakly) stationary stochastic processes x_t and y_t, two functions of interest in the frequency domain are the coherence $\gamma_{xy}(\lambda)=\gamma_{yx}(\lambda)$, that measures the degree of correlation between the random components of the processes associated with frequency λ, and the phase $\phi_{xy}(\lambda)= -\phi_{yx}(\lambda)$, that measures by how much the random components of the two processes corresponding to frequency λ are in the average out of phase, that is, one is leading or lagging the other.

If two processes are closely related in all respects, then $\gamma_{xy}(\lambda)$ is near 1 at all frequencies, since γ is the square of a correlation coefficient. If none of the two processes leads or lags the other, then $\phi_{xy}(\lambda)$ is close to 0 at all frequencies.

In economic applications it is often the case that not all components of the series are in phase or out of phase in the same manner. One has to expect, for example, that the low-frequency components of some pairs of series (e.g. those associated with the cyclical movements) may tend to have a different lag structure than the seasonal components, or the purely random ("irregular") components. One advantage of the empirical joint spectral approach is that it allows for a refinement of the analysis on agreement with this idea, since it tends to maintain the sources of variation of the series identified with different portions of the frequency range.

In applying the spectral methods care has to be given to the preliminary elimination of departures from stationarity. The underlying probabilistic theory has the standing assumption that the series in a pair are jointly (weakly) stationary. It is typically the case that many economic series present trends whose presence affects the stationary assumption, and hence have to be removed at a preliminary stage. This operation, that is called in general a filtering operation, can be done in several ways, not all of which affect the spectral analysis in the same way. An important objective of the present work is to explore several alternative techniques often considered in the literature. Failure to remove an important trend component, for example, produces biases in the estimation of spectral quantities, if the standard procedures are used; one such phenomenon is called "leakeage", and means that the presence of a nonstationary component concentrated in a certain frequency range (e.g. in the low

frequencies) affects (i.e. biases) the estimates at other frecuencies. For strong trend components such as those present in Argentine monetary series, leakeage tends to be very important.

2. The empirical study

Our main purpose in this paper is to present some empirical evidence concerned with the use of joint spectral methods to study the lead-lag structure of money and prices. The approach is to consider pairs of monthly series in Argentina for the period 1960-1973, and to see how far the spectral methods allow us to progress in the analysis. A related investigation for money and income, with quarterly data, has been developed by Elías (1979) along econometric lines that resort to regression equations.

The series we used are currency (C_t), demand deposits (D_t), total amount of money (definitions $M1_t$ and $M2_t$), wholesale price index (PW_t), and consumer price index (PC_t); 168 monthly observations, 1960-1973, for each series.

Some series have a seasonal, and most have a strong trend component, which means that they have to be filtered out. Three techniques were attempted to pre-process the series for the spectral analysis: (a) Differencing. The form $\log y_t - \log y_{t-12}$ was found to control changes in variance through time, to control quite satisfactorily the trend and seasonal components, and to exhibit a smooth cyclical component; (b) Residuals from a regression. Following Hsiao (1977) [see Granger (1969), Sims (1972)] the joint spectral analysis of $\hat{v}_t$ and $\hat{u}_t$ was performed, where in our context $\hat{v}_t = M2_t - (\hat{\alpha}_0 + \hat{\alpha}_1 M2_{t-1} + \ldots + \hat{\alpha}_k M2_{t-k})$, $\hat{u}_t = PW_t - (\hat{\beta}_0 + \hat{\beta}_1 M2_t + \ldots + \hat{\beta}_{p-1} M2_{t-p})$, and also of residuals defined with the roles of M2 and PW interchanged; (c) The X-11 seasonal and trend filter (See Shiskin, Young and Musgrave (1967)). This widely used procedure provides series of which the seasonal, trend, or trend and seasonal components are eliminated. To agree with standard economic practice the procedures were applied to each original series (y_t), their logarithms ($\log y_t$), and their logarithmic rates of change ($y_t^* = \log y_t - \log y_{t-1}$).

For each pair of series arising from a common filtering operation, we estimate $\hat{\gamma}_{xy}(\lambda)$ and $\hat{\phi}_{xy}(\lambda)$, and plot them together with a confidence band for the phase, that provides a way to evaluate approximately the validity of the hypothesis of 0 phase.

Some very brief comments about our findings are: (i) The coherence diagrams are in general easy to interpret in terms of the traditional components (trend-cycle, seasonal, and irregular) and their corresponding frequency ranges; (ii) The results of the filtering operations are in general to produce considerably low coherences at most frequencies; (iii) As a consequence of (ii), the confidence band on the phase tends to be very wide and to include very often the 0 value; (iv) These remarks are particularly true for the X-11 procedure, which seems to be due to two facts: The procedure is quite complicated and adds difficulty to the interpretation of the

TABLE 1. Estimated Coherence and Phase Between Money (M2) and Prices (PW) in Argentina, using differenced series: log y_t - log y_{t-12}

1960-1973 T=156, m=48		1960-1969 T=108, m=30		1968-1973 T=60, m=18	
Months	Phase (Coherence)	Months	Phase (Coherence)	Months	Phase (Coherence)
96	-0.69 (0.74)	60	1.28 (0.40)	36	-0.45 (0.61)
48	-1.03 (0.76)	30	1.13 (0.46)	18	-0.87 (0.60)
32	-1.35 (0.62)	20	0.57 (0.32)	12	-1.35 (0.43)

Note: T is the available number of observations; m is the total number of frequencies at which estimation takes place.

phase estimates, and the estimation of the low frequency component includes trend and cycle together, which is contrary to what is being looked for in our type of analysis; (v) Better results are provided by the differencing approach, which amounts to same-month comparisons. It was found that a rather strong out-of-phase movement at low frequencies was accompanied by coherences that were not too small. See Table 1 for the period 1960-1973; (vi) Further, the exploratory analysis of several differences led to a separation into two periods, for which the out-of-phase movements are important, but there is a reversal of the direction from one period to the other; (vii) Much less clear is the picture of what occurs at seasonal frequencies, for which no systematic scheme was discovered; (viii) Results of the same type as those reported in Table 1 were provided by the analysis of residuals $\hat{v}_t$ and $\hat{u}_t$ (see above); however the interesting movements at low frequencies were accompanied by considerably lower coherences; this was also found out by Hsiao (1977) in his analysis.

3. References

Elías, V.J. (1979), "Dinero e Ingreso Nominal en la Argentina", Ensayos Económicos, forthcoming.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric and Cross-Spectral Methods", Econometrica, 37, 424-38.

Hannan, E.J. (1970), Multiple Time Series, Wiley, New York.

Hsiao, C. (1977), "Money and Income Causality Detection", Working paper N° 167, N.B. of Economic Research, Cambridge, Mass.

Jenkins, G.M. and D.G. Watts (1968), Spectral Analysis and its Applications, Holden-Day, San Francisco.

Koopmans, L.H. (1974), The Spectral Analysis of Time Series, Academic Press, New York.

Shiskin, J., A.H. Young and J.C. Musgrave (1967), "The X-11 Variant of the Census Method II Seasonal Adjustment Program", Technical Paper N° 15, Bureau of the Census, Washington, D.C.

Sims, C.A. (1972), "Money, Income, and Causality", American Economic Review, LXII, 540-52.

Acknowledgements. The work in joint spectral analysis is one of several projects in Time Series Analysis developed by a team formed by J.C. Abril, Z.N. Cardozo, M.B. Ceballos and the present author. Some calculations were also done by N.M. Jarma and C.I. Martínez. Thanks are due to Professor V.J. Elías for helpful comments. The largest share of data processing was done at the Central Bank in Buenos Aires, under a program of "Research Residencies", in October 1977 and February 1979; comments by statisticians and economists of the Bank's staff are gratefully acknowledged. The author is member of the "Carrera del Investigador Científico", CONICET.

Summary. Empirical results in the joint spectral analysis of money and price series are analyzed, in the context of monthly series of Argentina, 1960-1973. The effects of three filters are studied: Differencing, taking residuals from a regression, and the X-11 seasonal program.

Résumé. L'auteur analyse les résultats empiriques de l'analyse spectrale conjointe de la monnaie et des prix dans le contexte des séries mensuelles de l'Argentine dès 1960 jusqu'à 1973. On étudie les effets de trois filtres: par des différences, au moyen des résidus d'une regression et avec le programme stationnaire X-11.

Journal of time series analysis

A journal sponsored by the Bernoulli Society for Mathematical statistics and probability

On the structure of the likelihood function of autoregressive and moving average models

by T. W. Anderson and Raúl P. Mentz

Stanford University and University of Tucumán, Argentina

Abstract. In finite order normal moving average models the maximum likelihood estimates always exist. For finite order normal autoregressive models sufficient conditions for the existence of maximum likelihood estimates is given. Some cases not satisfying the conditions are studied.

Keywords. autoregressive model; moving average; maximum likelihood; existence of maximum.

1. Introduction

This is a study of the likelihood function under normality of autoregressive and moving average models with finite numbers of parameters. For the moving average model it is shown that for all numerical sequences that are not identically 0 the maximum likelihood estimates of the parameters exist, and the estimates have some desired properties analogous to those often required for the parameters.

For the autoregressive model the situation is more complicated. The approach consists in analysing the behaviour of the likelihood function near points in the parameter space for which the roots of the associated polynomial equation are equal to 1 in absolute value. A critical question is whether the observed series satisfies a homogeneous difference equation with certain properties. We find a sufficient condition for the existence of maximum likelihood estimates. We study some cases where this condition is not satisfied. An example of such a case is a sample sequence satisfying a linear relation $a+bt$; the likelihood function does not have a maximum.

The results for the autoregressive model can be used in several ways: (a) to list the kinds of data sets that can lead to difficulties in the estimation of the parameters; (b) to design numerical examples to analyse in detail the performance of computational algorithms to compute the maximum likelihood estimates; and (c) to check data sets to be used for estimation when they come from examples generated artificially or arise from empirical measurements subjected to rounding.

2. The moving average model

The Gaussian moving average process $\{x_t\}$ of order q can be written

$$x_t = u_t + \alpha_1 u_{t-1} + \cdots + \alpha_q u_{t-q}, \tag{2.1}$$

[1] Research supported by the U.S. Office of Naval Research under Contract Number N00014-75-C-0442.

where the x_t are observable, the u_t are unobservable independent normal random variables, $\mathcal{E}u_t = 0$, $\mathcal{E}u_t^2 = \sigma^2$, and $\mathcal{E}u_t u_s = 0$, $t \neq s$, and the α_j are parameters. Then (2.1) for $t = \ldots, -1, 0, 1, \ldots$ defines a stationary stochastic process for all choices of $\boldsymbol{\alpha}_q = (\alpha_1, \ldots, \alpha_q)'$ and $0 < \sigma^2 < \infty$.

The associated polynomial equation

$$a_q(v) = \sum_{j=0}^{q} \alpha_j v^{q-j} = 0, \tag{2.2}$$

where $\alpha_0 = 1$, has roots v_j, $j = 1, \ldots, q$, and the covariance sequence of the process has components

$$\sigma_s = \mathcal{E}x_t x_{t+s} = \sigma^2 \sum_{j=0}^{q-s} \alpha_j \alpha_{j+s} = \sigma_{-s}, \qquad s = 0, 1, \ldots, q,$$

$$= 0, \qquad s > q. \tag{2.3}$$

The σ_s satisfy the inversion formula

$$\sigma_s = \int_{-\pi}^{\pi} e^{i\lambda s} f(\lambda)\, d\lambda, \tag{2.4}$$

where $f(\lambda) = (\sigma^2/2\pi)|a_q(e^{i\lambda})|^2 = (\sigma^2/2\pi)\prod_{r=1}^{q} |e^{i\lambda} - v_r|^2$ is the spectral density function of the process.

If $x_1, \ldots, x_T$ is a sample from (2.1), the likelihood function is

$$L(\boldsymbol{\alpha}_q, \sigma^2) = (2\pi)^{-T/2} |\boldsymbol{\Sigma}_T|^{-1/2} \exp\left\{-\frac{1}{2} \boldsymbol{x}_T' \boldsymbol{\Sigma}_T^{-1} \boldsymbol{x}_T\right\}$$

$$= (2\pi)^{-T/2} (\sigma^2)^{-T/2} |\boldsymbol{P}_T|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} \boldsymbol{x}_T' \boldsymbol{P}_T^{-1} \boldsymbol{x}_T\right\}, \tag{2.5}$$

where $\boldsymbol{x}_T = (x_1, \ldots, x_T)'$, $\boldsymbol{\Sigma}_T = \mathcal{E}\boldsymbol{x}_T \boldsymbol{x}_T' \equiv \sigma^2 \boldsymbol{P}_T$. It is often convenient to maximise (2.5) in two steps, first with respect to σ^2 at

$$\sigma^2 = \frac{1}{T} \boldsymbol{x}_T' \boldsymbol{P}_T^{-1} \boldsymbol{x}_T, \tag{2.6}$$

and then, substituting back in (2.5), with respect to $\boldsymbol{\alpha}_q$ in

$$n(\boldsymbol{\alpha}_q) = \frac{1}{|\boldsymbol{P}_T| (\boldsymbol{x}_T' \boldsymbol{P}_T^{-1} \boldsymbol{x}_T)^T}; \tag{2.7}$$

the "concentrated likelihood function" is a constant times the square root of (2.7).

We want to show that for $\boldsymbol{x}_T \neq \boldsymbol{0}$, the maximum likelihood estimates of $\boldsymbol{\alpha}_q$ and σ^2 exist, that of σ^2 is positive and finite, and that of $\boldsymbol{\alpha}_q$ can be taken so that the roots of (2.2) with α_j's replaced by their maximum likelihood estimates satisfy

$$|v_j| \leq 1, \qquad j = 1, \ldots, q. \tag{2.8}$$

Condition (2.8) need not be imposed on the α_j to make (2.1) stationary. However, we show below in Proposition 2 that there may exist equivalent parameterisations of the model, depending on the choice of some of the v_j to be

smaller or greater than 1 in absolute value; see also Anderson (1971), section 7.5.2. Hence, for the sake of definiteness we choose the defining set of α_j's in (2.1) to be such that (2.8) is satisfied, and we shall prove that the maximum likelihood estimates can be required to have this peroperty. Note that $|v_j|<1$, $j=1,\ldots,q$, is the 'invertibility condition' that allows (2.1) to be represented as an infinite autoregression.

PROPOSITION 1. *Σ_T is positive definite for any α_q, $\sigma^2>0$, $T\geq 1$.*

PROOF. For any T and q consider the matrix

$$A=\begin{pmatrix} \alpha_q & \alpha_{q-1} & \ldots & \alpha_1 & 1 & \ldots & 0 & 0 \\ 0 & \alpha_q & \ldots & \alpha_2 & \alpha_1 & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & \alpha_1 & 1 \end{pmatrix}, \tag{2.9}$$

of order $T\times(T+q)$ and of rank T. Then $\Sigma_T=\sigma^2 AA'$, which is positive definite for $\sigma^2>0$. Q.E.D.

PROPOSITION 2. *For each set of values $\alpha_1,\ldots,\alpha_q$, $\sigma^2>0$ there exists a set $\alpha_1^*,\ldots,\alpha_q^*$, $\sigma^{*2}>0$ such that $L(\alpha_q,\sigma^2)=L(\alpha_q^*,\sigma^{*2})$ and the roots of (2.2) with α_j replaced by α_j^* are less than or equal to 1 in absolute value.*

PROOF. If the α_j's are such that (2.8) holds, take $\alpha_j^*=\alpha_j$, $\sigma^{*2}=\sigma^2$. Otherwise, suppose that $|v_j|>1$, $j=1,\ldots,n$, $|v_j|\leq 1$, $j=n+1,\ldots,q$. Define α_j^* by

$$\sum_{r=0}^{q}\alpha_r^* v^{q-r}=\prod_{j=1}^{n}\left(v-\frac{1}{v_j}\right)\prod_{s=n+1}^{q}(v-v_s), \tag{2.10}$$

where $\alpha_0^*=1$, and

$$\sigma^{*2}=\sigma^2\left(\prod_{j=1}^{n}|v_j|^2\right)^{-1}. \tag{2.11}$$

Since the roots for which $|v_j|>1$ are either real or occur in conjugate pairs, the coefficients in (2.10) are real. The spectral density for α_r^*, σ^{*2} is identical to $f(\lambda)$ in (2.4), and hence defines the same covariance matrix and likelihood function. Q.E.D.

PROPOSITION 3. *The set of α_j for which (2.8) holds is compact.*

PROOF. The set of $v_1,\ldots,v_p$ for which (2.8) holds is compact. The p-tuple $\alpha_1,\ldots,\alpha_p$ is a continuous mapping of $v_1,\ldots,v_p$ as

$$\alpha_s=(-1)^s\sum_{i_1>i_2>\cdots>i_s}v_{i_1}v_{i_2}\cdots v_{i_s},\qquad s=1,\ldots,q. \tag{2.12}$$

Since the image of a compact set under a continuous mapping is compact, the proposition follows. Q.E.D.

PROPOSITION 4. *For $x_T\neq 0$ the function (2.5) is maximized by a set of real $\alpha_1,\ldots,\alpha_q$, $\sigma^2>0$ such that (2.8) holds.*

PROOF. For any set of real α_j's,

$$\lim_{\sigma^2\to 0}L(\alpha_q,\sigma^2)=\lim_{\sigma^2\to\infty}L(\alpha_q,\sigma^2)=0, \tag{2.13}$$

so that the maximum with respect to σ^2 has to occur for $0<\sigma^2<\infty$. The set of α_j's for which (2.8) is satisfied is compact (proposition 3), nonempty since it contains **0**, and n is continuous on that set. It follows that n attains its maximum value on that set. In fact, the maximum value is larger than or equal to $n(\mathbf{0})=(\sum_{t=1}^{T}x_t^2)^{-T}$. We then estimate σ^2 according to (2.6).

Suppose there exists a maximum of (2.5) at a set α_q, $\sigma^2>0$ for which (2.8) does not hold. Then by proposition 2 there is another set of α_q, $\sigma^2>0$ for which (2.8) holds and such that the corresponding value of (2.5) is the maximum. Q.E.D.

Some comments follow.

1. The present argument is not valid for estimates obtained by minimising the exponent in (2.5), as has been noted repeatedly in the literature; there the compensating effect of the Jacobian determinant is missing. Also, the present argument will in general not extend to approximations of the exact maximum likelihood procedure.

2. We shall show that $|\boldsymbol{P}_T|\to\infty$ as $|\alpha_j|\to\infty$. Each component of $\boldsymbol{P}_T$ is a polynomial in $\alpha_1,\ldots,\alpha_q$ and $|\boldsymbol{P}_T|$ is also a polynomial in $\alpha_1,\ldots,\alpha_q$. Since $\boldsymbol{P}_T$ is positive definite, $|\boldsymbol{P}_T|>0$. The coefficient of the highest degree in α_j must therefore be positive. Since $|\boldsymbol{P}_T|>0$ for α_j negative and arbitrarily large in absolute value, that term must be of even degree; hence $|\boldsymbol{P}_T|\to\infty$ as $\alpha_j\to-\infty$.

3. In the moving average model the components of Σ_T are given by (2.3); however, the determinant $|\Sigma_T|$ and the components of Σ_T^{-1} are not known in convenient closed form for general q. This contrasts with the autoregressive model to be considered below, for which $|\Sigma_T|$ and Σ_T^{-1} are known in detail, but the components of Σ_T are not known in a convenient closed form for the general case. Comment 2 above is an example of how to operate with $|\Sigma_T|$ of the moving average model for general $q\geq 1$.

4. The argument in this section holds for all samples sizes $T\geq 1$. For example, if $q\geq 1$, $T=1$, let $x\neq 0$ be observed. Then $\boldsymbol{P}_T=|\boldsymbol{P}_T|=\sigma_0=1+\alpha_1^2+\cdots+\alpha_q^2$, a scalar, (2.6) becomes $\sigma^2=x^2/(1+\alpha_1^2+\cdots+\alpha_q^2)$, and $n(\alpha_q)=x^{-2}$, a constant. The maximum likelihood estimates can then be taken to be any set of α_j's for which (2.8) holds, and for this set estimate σ^2 by the expression given above.

In a practical situation the condition $T>q$ will be adopted.

3. The autoregressive model

3.1. Introduction

The autoregressive process of order p satisfies

$$y_t+\beta_1 y_{t-1}+\cdots+\beta_p y_{t-p}=u_t, \tag{3.1}$$

where the y_t are observable, the u_t are unobservable independent normal random variables with $\mathcal{E}u_t = 0$, $\mathcal{E}u_t^2 = \sigma^2$, $\mathcal{E}u_t u_s = 0$, $t \neq s$, and the β_j's are parameters. For $t = \ldots, -1, 0, 1, \ldots$ (3.1) defines a stochastic process that is stationary with y_t independent of $u_{t+1}, u_{t+2}, \ldots$ if and only if the β_j's are such that the roots of the associated polynomial equation

$$b_p(w) = \sum_{j=0}^{p} \beta_j w^{p-j} = 0 \tag{3.2}$$

(where $\beta_0 = 1$) are less than 1 in absolute value.

Letting $\boldsymbol{\beta}_p = (\beta_1, \ldots, \beta_p)'$, we define $B_p = \{\boldsymbol{\beta}_p : b_p(w) = 0$ has roots $|w_j| < 1$, $j = 1, \ldots, p\}$, $\bar{B}_p = \{\boldsymbol{\beta}_p : b_p(w) = 0$ has roots $|w_j| \leq 1$, $j = 1, \ldots, p\}$, and $B_p^* = \{\boldsymbol{\beta}_p : b_p(w) = 0$ has roots $|w_j| = 1, j = 1, \ldots, p\}$.

Under stationarity of (3.1) (that is, $\boldsymbol{\beta}_p \in B_p$), the process has a strictly positive spectral density function $f(\lambda) = (\sigma^2/2\pi)|b_p(e^{i\lambda})|^{-2} = (\sigma^2/2\pi)\{\prod_{r=1}^{p} |e^{i\lambda} - w_r|^2\}^{-1}$, and the inversion formula (2.4) holds for the covariance sequence $\sigma_s = \mathcal{E}y_t y_{t+s}$, $s = 0, \pm 1, \ldots$.

If $y_1, \ldots, y_T$ is a sample from $\{y_t\}$ satisfying (3.1) and if $\boldsymbol{\beta}_p \in B_p$, the likelihood function is

$$L(\boldsymbol{\beta}_p, \sigma^2) = (2\pi)^{-T/2}|\boldsymbol{\Sigma}_T|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{y}_T'\boldsymbol{\Sigma}_T^{-1}\mathbf{y}_T\right\}$$

$$= (2\pi)^{-T/2}(\sigma^2)^{-T/2}|\mathbf{Q}_T|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}_T'\mathbf{Q}_T^{-1}\mathbf{y}_T\right\}, \tag{3.3}$$

where $\mathbf{y}_T = (y_1, \ldots, y_T)'$, $\boldsymbol{\Sigma}_T = \mathcal{E}\mathbf{y}_T\mathbf{y}_T' \equiv \sigma^2\mathbf{Q}_T$. Then (3.3) can be maximized with respect to σ^2 at

$$\sigma^2 = \frac{1}{T}\mathbf{y}_T'\mathbf{Q}_T^{-1}\mathbf{y}_T, \tag{3.4}$$

and with respect to $\boldsymbol{\beta}_p$ by maximising a multiple of the square of the concentrated likelihood function, namely,

$$n(\boldsymbol{\beta}_p) = \frac{|\mathbf{Q}_T^{-1}|}{(\mathbf{y}_T'\mathbf{Q}_T^{-1}\mathbf{y}_T)^T}. \tag{3.5}$$

3.2. Properties of the Covariance Matrix and the Likelihood Function

PROPOSITION 5. *The sets $\bar{B}_p$ are closed and bounded, that is, compact.*

PROOF. See proposition 3. Note $\bar{B}_p$ is the closure of B_p.

PROPOSITION 6. *For $\boldsymbol{\beta}_p \in B_p$ and $T > p$ (a) $\mathbf{Q}_r$ and $\mathbf{Q}_r^{-1}$ are positive definite for $r = 1, \ldots, T$; (b)*

$$\mathbf{y}_T'\mathbf{Q}_T^{-1}\mathbf{y}_T = \mathbf{y}_p'\mathbf{Q}_p^{-1}\mathbf{y}_p + \sum_{t=p+1}^{T}\left(\sum_{j=0}^{p}\beta_j y_{t-j}\right)^2; \tag{3.6}$$

(c) $|\mathbf{Q}_p| = |\mathbf{Q}_T|$; (d) the components of $\mathbf{Q}_p^{-1}$ are given by

$$q_p^{ij} = \sum_{s=0}^{i-1}\beta_s\beta_{s+j-i} - \sum_{s=p+1-j}^{p+i-j}\beta_s\beta_{s+j-i}, \qquad i \leq j; \tag{3.7}$$

(e) the components of $\mathbf{Q}_T^{-1}$ are given by

$$q_T^{ij} = \sum_{s=\max(j,p+1)}^{\min(i+p,T)}\beta_{s-i}\beta_{s-j} + q_p^{ij}, \qquad i \leq j \leq p,$$

$$= \sum_{s=\max(j,p+1)}^{\min(i+p,T)}\beta_{s-i}\beta_{s-j}, \qquad i \leq j, j > p,$$

$$= 0, \qquad i + p \leq j \leq T, i \leq T - p. \tag{3.8}$$

PROOF. Results in (b) through (*e*) were given by Siddiqui (1958), who assumed, $T > 2p$. However, (3.6) holds for $T \geq p$, and the fact that (*c*), (*d*) and (*e*) hold more generally follows from Anderson and Mentz (1977), Walker (1961), etc. Q.E.D.

PROPOSITION 7. *For $\boldsymbol{\beta}_p \in B_p$,*

$$|\mathbf{Q}_p^{-1}| = \prod_{i,j=1}^{p}(1 - w_i w_j). \tag{3.9}$$

PROOF. See Anderson and Mentz (1977).

PROPOSITION 8. *For $\boldsymbol{\beta}_p \in B_p$, as $\boldsymbol{\beta}_p \to \bar{\boldsymbol{\beta}}_p \in \bar{B}_p$, $\mathbf{Q}_p^{-1} \to \mathbf{0}$ if and only if $\bar{\boldsymbol{\beta}}_p \in B_p^*$.*

PROOF. Let $\mathbf{x} = (x_1, \ldots, x_p)' \neq \mathbf{0}$ and let x_s be the nonzero component of $\mathbf{x}$ with highest index. Then $\sum_{t=1}^{p} x_t z^{t-1} = x_s\prod_{t=1}^{s-1}(z - a_t)$, say. By the inversion formula (2.4),

$$\mathbf{x}'\mathbf{Q}_p\mathbf{x} = \sum_{g,h=1}^{s} x_g x_h \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{e^{i\lambda(g-h)}}{\left|\sum_{r=0}^{p}\beta_r e^{i\lambda(p-r)}\right|^2}d\lambda$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\left|\sum_{t=1}^{s} x_t e^{i\lambda t}\right|^2}{\left|\sum_{r=0}^{p}\beta_r e^{i\lambda(p-r)}\right|^2}d\lambda = \frac{|x_s|^2}{2\pi}\int_{-\pi}^{\pi}\frac{\left|\prod_{t=1}^{s-1}(e^{i\lambda} - a_t)\right|^2}{\left|\prod_{r=1}^{p}(e^{i\lambda} - w_r)\right|^2}d\lambda. \tag{3.10}$$

As $\boldsymbol{\beta}_p \to \bar{\boldsymbol{\beta}}_p$, $w_r \to \nu_r$, say, for which $|\nu_r| = 1$, $r = 1, \ldots, p$. Then the integral approaches ∞ because at most $s - 1$ $(1 \leq s \leq p)$ of $\nu_1, \ldots, \nu_p$ agree with $a_1, \ldots, a_{s-1}$.

The fact that $\mathbf{x}'\mathbf{Q}_p\mathbf{x} \to \infty$ for every $\mathbf{x} \neq \mathbf{0}$ implies that $\mathbf{Q}_p^{-1} \to \mathbf{0}$ is proved in Anderson and Taylor (1979).

Conversely, suppose that $\mathbf{Q}_p^{-1} \to \mathbf{0}$. Then each component in (3.7) approaches 0. In particular, $q_p^{11} = 1 - \beta_p^2 \to 0$. Thus $\beta_p^2 = \prod_{j=1}^{p} w_j^2 \to 1$. Since $|w_j| \leq 1$, this implies that $|w_j| \to 1$, $j = 1, \ldots, p$. Q.E.D.

PROPOSITION 9. *For any β_p such that the roots of $b_p(w)=0$ are different from 1 in absolute value and $\sigma^2>0$ there exists a $\tilde{\beta}_p$ and $\tilde{\sigma}^2>0$ such that $\tilde{\beta}_p \in B_p$ and the likelihood function based on observations $y_1, \ldots, y_T$ has the same value for (β_p, σ^2) as for $(\tilde{\beta}_p, \tilde{\sigma}^2)$.*

PROOF. See proposition 2.

3.3. Existence of Maximum Likelihood Estimates

The question of existence of the maximum likelihood estimate is emphasized by the following result.

PROPOSITION 10. *Let p and $y_T \neq \mathbf{0}$ be given $(T>p)$. If there exists a vector $\beta_k \in B_k^*$ for which*

$$y_t+\beta_1 y_{t-1}+\cdots+\beta_k y_{t-k}=0, \qquad t=k+1,\ldots,T, \tag{3.11}$$

for some k, $1 \leq k \leq p$, then

$$\inf_{\beta_p \in B_p} y_T' Q_T^{-1} y_T = 0. \tag{3.12}$$

PROOF. For any $\beta_k \in B_k$, $1 \leq k \leq p$, we have

$$y_T' Q_T^{-1} y_T = y_k' Q_k^{-1} y_k + \sum_{t=k+1}^{T} (y_t+\beta_1 y_{t-1}+\cdots+\beta_k y_{t-k})^2 \tag{3.13}$$

by proposition 6. As β_k approaches any vector in B_k^*, Q_k^{-1} approaches the $\mathbf{0}$ matrix by proposition 8, and hence $y_k' Q_k^{-1} y_k$ approaches 0 for every y_k. The second term in (3.13) approaches 0 by (3.11). Q.E.D.

We consider in this section vectors of observations y_T for which no equation of the form (3.11) is satisfied exactly outside the sets B_k where Q_k and Q_T are positive definite.

CONDITION I. *$y_T \neq \mathbf{0}$ and $y_1, \ldots, y_T$ do not satisfy (3.11) with $\beta_k \neq 0$ for any $k=1,2,\ldots,p$ and β_k in $B_k^c \cap \bar{B}_k$.*

Note that $B_k^c \cap \bar{B}_k$ contains those vectors β_k for which $b_k(w)=0$ has at least one root with absolute value equal to 1.

PROPOSITION 11. *If y_T satisfies condition I for given $p<T$, then there exist maximum likelihood estimates of σ^2 and β_p corresponding to (3.1) with $\beta_p \in B_p$ and $0<\sigma^2<\infty$.*

PROOF[1]. By hypothesis $y_T' Q_T^{-1} y_T > 0$ for $\beta_p \in \bar{B}_p$. By proposition 7 $|Q_p^{-1}| \to 0$ as $\beta_p \to \bar{\beta}_p$ in $B_p^c \cap \bar{B}_p$. It follows that $n(\beta_p) \to 0$ as $\beta_p \to \bar{\beta}_p$ in $B_p^c \cap \bar{B}_p$. Thus n is continuous on $\bar{B}_p$, which is nonempty $(\mathbf{0} \in \bar{B}_p)$ and compact (proposition 5). Hence n is bounded on $\bar{B}_p$ and attains its maximum value there. Further, since $n(\beta_p) \to 0$

as $\beta_p \to \bar{\beta}_p$ in $B_p^c \cap \bar{B}_p$, the maximum value is attained in B_p. In fact, the maximum value of n is larger than or equal to $n(\mathbf{0}) = (\sum_{t=1}^{T} y_t^2)^{-1} > 0$.

The maximum of (3.3) is attained for $0<\sigma^2<\infty$, because $\lim_{\sigma^2 \to \infty} L(\beta_p, \sigma^2) = \lim_{\sigma^2 \to 0} L(\beta_p, \sigma^2) = 0$ for $\beta_p \in B_p$. Further, for $\bar{\beta}_p$ in $B_p^c \cap \bar{B}_p$

$$\lim_{\beta_p \to \bar{\beta}_p, \sigma^2 \to 0} L(\beta_p, \sigma^2) = 0, \tag{3.14}$$

because the exponential function dominates $(\sigma^2)^{-T/2}$, and the quadratic form in (3.3) is positive in $\bar{B}_p$.

By proposition 9 it follows that if a maximum occurs at $\beta_p, \sigma^2>0$ for which one or more roots of (3.2) are greater than 1 in absolute value, there is another set of $\beta_p, \sigma^2>0$ for which β_p belongs to B_p and the value of the likelihood function is the same. Q.E.D.

3.4. The Case Where the Observations Do Not Satisfy Condition 1.

CONDITION II. *$y_T \neq \mathbf{0}$ and $y_1, \ldots, y_T$ satisfy (3.11) for some $k=1,\ldots,p$ and β_k in B_k^*.*

Comparing with condition I, we see that in condition II we consider data sets y_T that satisfy exactly equations as those in (3.11) for parameter values for which all roots of $b_k(w)=0$ are equal to 1 in absolute value. For a given y_T we take k to be the smallest integer for which condition II is met.

Under condition II $\inf_{\beta_p \in B_p} y_T' Q_T^{-1} y_T = 0$, and hence the maximum likelihood estimate of β_p may not exist. To see the difficulties we face in trying to make a general statement, we consider the following example.

For $p=2$ let $y_t = a+bt$, $t=1,\ldots,T$, for some constants a and $b \neq 0$. Then y_t satisfies condition II for $k=2$ and roots $\nu_1=\nu_2=1$. We have

$$Q_2^{-1} = \begin{pmatrix} 1-\beta_1^2 & \beta_1(1-\beta_2) \\ \beta_1(1-\beta_2) & 1-\beta_2^2 \end{pmatrix} = (1-w_1w_2)\begin{pmatrix} 1+w_1w_2 & -(w_1+w_2) \\ -(w_1+w_2) & 1+w_1w_2 \end{pmatrix}. \tag{3.15}$$

For roots $w_1=w_2=x$, $0<|x|<1$, we have

$$y_T' Q_T^{-1} y_T = y_2' Q_2^{-1} y_2 + \sum_{t=3}^{T} (y_t-(w_1+w_2)y_{t-1}+w_1w_2y_{t-2})^2$$

$$= (1-x)[(1+x)2xb^2+O(1-x)], \tag{3.16}$$

where $O(1-x)$ indicates a term that divided by $1-x$ is bounded as $x \to 1$. Since $|Q_2^{-1}| = (1-x^2)^4$,

$$n(\beta_2) = \frac{(1+x)^4(1-x)^{4-T}}{[(1+x)2xb^2+O(1-x)]^T}. \tag{3.17}$$

As $x \to \nu=1$, $n(\beta_2)$ approaches $(2b^2)^{-4}>0$ if $T=4$, it approaches 0 if $T<4$, and it diverges to ∞ if $T>4$. Hence, if $T>4$ the likelihood function attains no maximum for $\beta_2 \in B_2$.

PROPOSITION 12. *For $\beta_k^* \in B_k^*$ let $\nu_1, \ldots, \nu_k$ be the roots of $\sum_{r=0}^{k} \beta_r^* \nu^{p-r} = 0$, and let l, h, m_j be the multiplicities of the roots equal to $1, -1$, and $e^{i\theta_j}$ $(0 < \theta_j < \pi)$, $j = 1, \ldots, m$, respectively. Let $w_j = x\nu_j$, $j = 1, \ldots, k$, and Q_k defined by the w_j, where $|x| < 1$. Then, as $x \to 1$,*

$$|Q_k^{-1}| = (1-x)^K f(x), \tag{3.18}$$

where $K = l^2 + h^2 + 2\sum_{j=1}^{m} m_j^2$, and f is such that $0 < \lim_{x\to 1} f(x) \leq M < \infty$.

PROOF. $|Q_k^{-1}|$ is a polynomial in $w_1, \ldots, w_k$ of degree $2k^2$, which implies it is a polynomial in x of degree $2k^2$ when $w_j = x\nu_j$. From (3.9) it is seen that in this case $x = 1$ is a root of the polynomial. The product $\prod_{i,j}(1 - x^2\nu_i\nu_j)$ over i, j such that $\nu_i = \nu_j = 1$ is $(1-x^2)^{l^2} = (1-x)^{l^2}(1+x)^{l^2}$. Similarly, the product over i, j such that $\nu_i = \nu_j = -1$ is $(1-x)^{n^2}(1+x)^{n^2}$. The product over h, j such that $(\nu_h, \nu_j) = (e^{i\theta_g}, e^{i\theta_g})$, $(e^{-i\theta_g}, e^{-i\theta_g})$, $(e^{i\theta_g}, e^{-i\theta_g})$, or $(e^{-i\theta_g}, e^{i\theta_g})$ is

$$(1-x^2 e^{2i\theta_g})^{m_g^2}(1-x^2 e^{-2i\theta_g})^{m_g^2}(1-x)^{2m_g^2}(1+x)^{2m_g^2}. \tag{3.19}$$

The other factors of $|Q_k^{-1}|$ are $1+x^2$, $1+x^2 e^{\pm i\theta_g}$, and $1-x^2 e^{\pm i\theta_g \pm i\theta_f}$. These factors are all bounded by 2 and bounded away from 0. Q.E.D.

PROPOSITION 13. *Let y_T satisfy condition II for given k, $1 \leq k \leq p$, $p < T$, and $\beta_k^* \in B_k^*$. Then if $T > K$, where K was defined in Proposition 12,*

$$\sup_{\beta_p \in B_p} n(\beta_p) = \infty. \tag{3.20}$$

PROOF. Let $\nu_1, \ldots, \nu_k$ be the roots associated with β_k^*, so that $|\nu_j| = 1$, $j = 1, \ldots, k$. To prove (3.20) it suffices to exhibit a sequence of roots $w_1, \ldots, w_p$ (or the associated sequence of β_p in B_p) such that as they approach $\nu_1, \ldots, \nu_p$, respectively, $n(\beta_p)$ approaches ∞. We can restrict attention to $\beta_k \in B_k$, since $\beta_p' = (\beta_k', 0')$ can be used as the argument of $n(\beta_p)$ as β_p' approaches $(\beta_k^{*\prime}, 0')$. As in proposition 12, let $w_j = x\nu_j$, $j = 1, \ldots, k$, where $0 < |x| < 1$, and let β_k be determined by $w_1, \ldots, w_k$. Then (3.18) holds and $y_T' Q_T^{-1} y_T$ is a polynomial in x with a zero of $x = 1$. This implies that $y_T' Q_T^{-1} y_T = (1-x)^r f(x)$, for some multiplicity $r \geq 1$. Hence, if $T > K$, as $x \to 1$, $n \to \infty$. Q.E.D.

It can be shown that if we consider a set of roots $\nu_1 = \ldots = \nu_k = 1$ with associated parameter vector β_k^*, and another set $w_1 = \cdots = w_k = x$, for $0 < |x| < 1$, with associated parameter vector β_k, then $x = 1$ is a simple root of $y_k' Q_k^{-1} y_k$ when the quadratic form is evaluated at the w_j, $y_T' Q_T^{-1} y_T = (1-x) f(x)$ with $f(x)$ such that $0 < \lim_{x\to 1} f(x) \leq M < \infty$, and $n(\beta_p) = (1-x)^{K-T} f(x)$ as $\beta_k \to \beta_k^*$. This result generalizes that considered at the beginning of this section.

3.5. Comments

1. The present argument is in general not valid for approximations to the exact maximum likelihood procedure, including the suggestion to minimise the

exponent $y_T' Q_T^{-1} y_T$ in (3.3) with respect to the β_j. By other direct procedures it can be shown that the estimates of the β_j in (3.1) coming from the (sample) Yule–Walker equations defined by

$$\sum_{j=1}^{p} c_{i-j}\hat{\beta}_j = c_i, \qquad i = 1, 2, \ldots, p, \tag{3.21}$$

where $c_i = c_{-i} = T^{-1}\sum_{t=1}^{T-i} y_{t+i} y_t$, $i = 0, 1, \ldots, p$, belong to B_p for $y_T \neq \mathbf{0}$. See Anderson (1971a) and Pagano (1973).

2. A basic difficulty in the analysis of the likelihood function (3.3) is that Σ_T^{-1} is not positive definite unless all roots of (3.2) are less than 1 in absolute value. This contrasts with the situation for the moving average model (proposition 1).

3. Q_p^{-1} is positive definite if and only if no root of (3.2) has absolute value equal to 1. In effect, if $\beta_p \in B_p$, let $x \neq \mathbf{0}$; then

$$\begin{aligned}
x' Q_p x &= \sum_{g,h=1}^{p} x_g x_h \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\lambda(g-h)}}{\left|\sum_{j=0}^{p} \beta_j e^{i\lambda(p-j)}\right|^2} d\lambda \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\left|\sum_{g=1}^{p} x_g e^{i\lambda g}\right|^2}{\left|\sum_{j=0}^{p} \beta_j e^{i\lambda(p-j)}\right|^2} d\lambda \\
&\geq \frac{1}{2\pi}\frac{1}{\left(1+\sum_{j=1}^{p}|\beta_j|\right)^2}\int_{-\pi}^{\pi}\left|\sum_{g=1}^{p} x_g e^{i\lambda g}\right|^2 d\lambda \\
&= \frac{1}{2\pi}\frac{1}{\left(1+\sum_{j=1}^{p}|\beta_j|\right)^2}\sum_{g,h=1}^{p} x_g x_h \int_{-\pi}^{\pi} e^{i\lambda(g-h)} d\lambda \\
&= \frac{\sum_{g=1}^{p} x_g^2}{\left(1+\sum_{j=1}^{p}|\beta_j|\right)^2} > 0.
\end{aligned} \tag{3.22}$$

This is a proof of part (*a*) of proposition 6. Conversely, if Q_p^{-1} is positive definite, the fact that the roots of (3.2) are all less than 1 in absolute value follows from propositions 9 and 7.

4. According to our results, an algorithm should consider (3.3) and maximise it for $\sigma^2 > 0$ and $\beta_p \in B_p$. This is important, since the likelihood function may have different forms depending upon the nature of the roots of (3.2), that is, their being smaller or larger than 1 in absolute value. If y_T is generated by the model $y_t + \beta y_{t-1} = u_t$ with $|\beta| < 1$, $\mathcal{E}u_t^2 = \sigma^2$, $0 < \sigma^2 < \infty$, then

$$L(\beta, \sigma^2) = (2\pi)^{-T/2}(\sigma^2)^{-T/2}\sqrt{1-\beta^2}\exp\left\{-\frac{1}{2\sigma^2}\left[(1-\beta^2)y_1^2 + \sum_{t=2}^{T}(y_t + \beta y_{t-1})^2\right]\right\}. \tag{3.23}$$

However, if y_T is generated by $y_t+\beta^* y_{t-1}=u_t^*$, with $|\beta^*|>1$, $\mathcal{E}u_t^{*2}=\sigma^{*2}$, $0<\sigma^{*2}<\infty$, then (cf. Anderson (1971), section 5.2)

$$L_1(\beta^*, \sigma^{*2}) = (2\pi)^{-T/2}(\sigma^{*2})^{-T/2}\sqrt{\beta^{*2}-1}|\beta^*|^{T-1} \times \exp\left\{-\frac{1}{2\sigma^{*2}}\left[\sum_{t=2}^{T}(y_t+\beta^* y_{t-1})^2+(\beta^{*2}-1)y_T^2\right]\right\}. \tag{3.24}$$

A consequence of proposition 9 is that $L(\beta^{*-1}, \sigma^{*2}\beta^{*-2}) = L_1(\beta^*, \sigma^{*2})$, for the same data set y_T, which can be checked directly.

5. Even under condition I, the existence of maximum likelihood estimates satisfying the desired conditions does not imply that they are unique. Uniqueness for β_p in B_p and $0<\sigma^2<\infty$ requires further study of the likelihood function.

6. At the beginning of section 3.3 we argued that if the observations satisfy $y_t+\beta_1 y_{t-1}+\cdots+\beta_k y_{t-k}=0$, $t=k+1,\ldots,T$ for some k, $1\leq k\leq p$, and for some vector in B_k^*, then $\inf_{\beta_p\in B_p} y_T' Q_T^{-1} y_T = 0$. An interesting question is whether the converse is true. It can be shown that (3.12) implies that (3.11) holds for β_i's such that at least one root of the associated polynomial equation has absolute value equal to 1.

By direct argumentations the cases $p=1$ and $p=2$ can further be shown to hold for parameter sets in B_p^* ($p=1, 2$); it remains to see if this result extends to higher orders.

7. One application of our analysis is in the exploratory fitting of autoregressions to time series data. For given T and once p is chosen, the analysis shows the kind of difficulties that can arise.

For example, if $p=2$ we find that the sequences $a+bt$, $(a+bt)(-1)^t$, $a+b(-1)^t$ and $a\cos(\nu t-b)$, that correspond to roots $(1, 1)$, $(-1, -1)$, $(1, -1)$ and $(e^{i\nu}, e^{-i\nu})$, respectively, imply that the maximum likelihood estimates do not exist if T is large enough.

For observed sequences that are 'close' to those associated with roots of absolute value equal to 1, it seems safe to conjecture that even if mathematically the likelihood function admits a maximizing set with the required conditions, the numerical procedures to find the values of the estimates may be quite unstable.

8. The question discussed in the last part of the previous comment is connected with the fact that our mathematical analysis should also be viewed in connection with the computational algorithms needed to perform the calculations.

There seems to be little evidence in this respect and it would be useful to know how 'close' an observed sequence has to be to one satisfying exactly Condition II, to produce computational complications for a given algorithm.

Also it would be useful to relate this behavior to that observed using 'a priori' more stable approximations to the likelihood function, like those for which the covariance matrix is always positive definite.

Notes

[1] The argument to prove Lemma 2 of Deistler, Dunsmuir, and Hannan (1978) could be used here.

References

ANDERSON, T. W. (1971) *The Statistical Analysis of Time Series*. John Wiley, New York.

ANDERSON, T. W. (1971a), The Stationarity of an Estimated Autoregressive Process, Technical Report No. 7, ONR Contract N00014-67-A-0112-0030, Department of Statistics, Stanford University.

ANDERSON, T. W. and R. P. MENTZ (1977) The Generalized Variance of a Stationary Autoregressive Process. *Journal of Multivariate Analysis*, 7, 584–588.

ANDERSON, T. W. and J. B. TAYLOR (1979) Strong Consistency of Least Squares Estimates in Dynamic Models. *Annals of Statistics*, 7, 484–489.

DEISTLER, M., W. DUNSMUIR and E. J. HANNAN (1978) Vector Linear Time Series Models: Corrections and Extensions. *Advances in Applied Probability* 10, 360–372.

PAGANO, M. (1973) When is an Autoregressive Scheme Stationary? *Communications in Statistics*, 1, 533–544.

SIDDIQUI, M. M. (1958) On the Inversion of the Sample Covariance Matrix in a Stationary Autoregressive Process. *The Annals of Mathematical Statistics* 29, 585–588.

WALKER, A. M. (1961) Large-Sample Estimation of Parameters for Moving-Average Models. *Biometrika* 48, 343–357.

Encuesta a centros estadísticos latinoamericanos

by Raúl Pedro Mentz and Nora M. Jarma de Cortés

Invited Paper 43rd session of the International Statistical Institute
Buenos Aires, Argentina
November 30th, December 11th, 1981

Encuesta a centros estadísticos latinoamericanos

Raúl Pedro Mentz and Nora M. Jarma de Cortés

Instituto de Investigaciones Estadísticas (INIE)
Universidad Nacional de Tucumán, Argentina

INTRODUCCION. El principal propósito de este trabajo es informar los resultados de una encuesta a centros estadísticos latinoamericanos realizada en 1981.

Los antecedentes de la encuesta son los siguientes. El COLPRIE (Comité Latinoamericano para Promover Relaciones Internacionales en Estadística) definió en 1978 como uno de sus temas de interés, el apoyo a la difusión en los órganos apropiados de las investigaciones y los trabajos estadísticos realizados en la región. Un documento preparado a tal efecto fué presentado al Instituto Interamericano de Estadística (IASI), fué considerado por un grupo de trabajo reunido en Washington en abril de 1979, y las principales recomendaciones incorporadas en un documento presentado al Comité Ejecutivo del IASI en setiembre de 1979. En el curso de estos trabajos y como elemento de apoyo informativo, el COLPRIE inició un relevamiento del potencial latinoamericano para producir y publicar trabajos estadísticos, el que condujo a una encuesta preliminar llevada a cabo entre 1979 y 1980 (Mentz y Jarma, 1980). Para esta encuesta por correo se interrogaron a 27 centros, de los que se obtuvieron 14 respuestas. Se enfatizaron los aspectos del potencial humano, habiéndose logrado formar tablas con datos del personal superior clasificado por nivel del grado o título universitario, área de especialización, dedicación, categoría docente, y otros conceptos.

La existencia de la 43a. Sesión Bienal del Instituto Internacional de Estadística (ISI) en Buenos Aires a fines de 1981 y los interesantes resultados de la mencionada encuesta preliminar, determinaron que a partir de 1980 se decidiera realizar dos trabajos paralelos: (1) Formalizar y ampliar la cobertura de la encuesta; (2) Complementarla con un trabajo de cuantificación de las publicaciones de estadísticos latinoamericanos, realizado en base a la revista Current Index in Statistics.

En este informe se resumirán los principales aspectos de estos dos trabajos: los de la encuesta en las secciones 1 y 2 y los del otro en las secciones 3 y 4. Los detalles completos aparecerán en (Mentz y Jarma, 1981) y (Mentz y Lopez, 1981).

1. Objetivos y procedimientos de la encuesta

1.1. Objetivos de la encuesta

El objetivo principal de la encuesta es detectar a los centros (grupos humanos institucionalizados) de nivel superior existentes en los diversos países latinoamericanos, que realizan investigaciones y trabajos estadísticos, los publican, editan libros y revistas, etc., y describirlos en términos de sus recursos humanos y de su producción.

El interés está sobre todo en centros que son grandes o que tienen una producción importante. Se trata de definir el tamaño por la cantidad y calidad de los recursos humanos y de los productos finales alcanzados. El tema de los presupuestos y los elementos físicos y de equipo disponibles no ha sido considerado en la encuesta.

Una idea central subyacente es que en años recientes se ha producido en América Latina una creciente institucionalización de las actividades estadísticas, sobre todo en el campo académico. El progreso ha sido más señalado en algunos paises. Han surgido en universidades y en otros lugares, grupos altamente especializados, con un interesante caudal de recursos y a menudo bien orientados académicamente, que ofrecen a sus respectivos paises y también en el plano regional, la posibilidad de servir de base para un desarrollo importante de la disciplina y la profesión estadística en el presente o el futuro inmediato.

La encuesta no está diseñada para verificar con sentido histórico la importancia relativa del momento que vive la estadística en la región. Sin embargo, al aportar una descripción cuantitativa de los mejores esfuerzos que se están realizando en un momento actual (el año 1981), se aspira a que los resultados actúen como complemento de otras informaciones para iluminar la verdadera situación, sus características actuales, sus perspectivas y los puntos claves a utilizar o mejorar.

No se enfatiza en la encuesta el tema de la enseñanza, ni se trata el de las estadísticas oficiales. Debe advertirse que paralelamente con este trabajo, otras personas preparan informes específicos sobre la enseñanza de la estadística, las estadísticas oficiales y la profesión de estadístico, todos referidos a América Latina. Ellos serán presentados en forma conjunta en una sesión sobre "Desarrollos Recientes en el Campo Estadístico en América Latina", en la reunión del ISI de 1981.

1.2. Procedimientos de la encuesta

La encuesta estuvo basada en un formulario que aparece en el Apéndice 1. El cuestionario comprende una serie de preguntas referidas al centro y otro conjunto referido a sus integrantes. Las preguntas 1-4 tratan de describir institucionalmente al centro según su nombre, domicilio, dirección postal, dependencia y director. Las preguntas 5 y 6 procuran caracterizar sus actividades, si bien el tratamiento en la pregunta 5 es sólo de resumen: los datos concretos sobre actividades surgen de otras preguntas posteriores. Las preguntas 7-10 se refieren a la producción del centro: informes, publicaciones, tesis, actividades editoriales diversas y revistas editadas. La pregunta 11 es sobre datos detallados de los integrantes del centro: título, área del título, actividades, dedicación, categoría y área de interés.

La encuesta fué enviada a un total de 106 centros. Con relación a los 27 centros de la encuesta preliminar de 1979-80, se incorporaron todas las instituciones invitadas al SELENES (Seminario Latinoamericano sobre la Ense-

ñanza de la Estadística, Santiago de Chile, diciembre de 1978), y otras. Debe advertirse que muchas de las instituciones incluídas entre las 106 no son grandes, ni estables, ni están dedicadas al trabajo o a la investigación estadística en ninguna de sus formas típicas.

La encuesta fué despachada inicialmente en carácter de "encuesta piloto" a 6 centros en el mes de octubre de 1980. Ellos la llenaron y como no surgieron comentarios desfavorables, el mismo cuestionario se envió a la nómina completa en el mes de febrero de 1981. Posteriormente se reiteró el envío del cuestionario y el pedido de datos a los centros que no habían respondido, lo que se hizo en el mes de mayo de 1981.

Para algunos centros disponemos entonces de información de uno o más de los siguientes tipos: (a) Encuesta preliminar 1979-80; (b) Comentarios y correcciones a la encuesta preliminar 1979-80; (c) Encuesta piloto de 1980; (d) Encuesta definitiva de 1981. Los errores y respuestas incompletas fueron tratados mediante el intercambio de correspondencia, habiendo completado la recepción de datos en el mes de julio de 1981.

Desde un punto de vista estricto, existen por lo expuesto algunas diferencias temporales en los datos disponibles. Sin embargo, a los fines de este trabajo el efecto de esta complicación es menor. Debe advertirse que no existió ninguna posibilidad de visitar a los centros encuestados, lo que habría facilitado el perfeccionamiento del trabajo.

En resumen, disponemos de información total o parcial sobre 29 centros latinoamericanos. Ellos pertenecen a Argentina (6), Brasil (11), Colombia (3), Chile (4), Guatemala (1), México (3) y Venezuela (1). La nómina de respondentes constituye el Apéndice 2.

2. Principales resultados de la encuesta

Como se anticipó en la Sección 1.1 , el principal objetivo de la encuesta es describir los recursos humanos y la producción de los centros estadísticos respondentes. Los cuadros I, II, III y IV contienen la información básica disponible sobre estos dos temas.

El Cuadro I exhibe al personal de investigación y docencia de cada centro, clasificado por título. Son 549 personas en total, entre las cuales hay 146 doctores y 178 masters. De los primeros 90 obtuvieron sus títulos en Estados Unidos, Europa o lugares similares, y de los segundos 30. Nótese que para un centro no tenemos datos sobre su personal.

Además de la información sobre los títulos de las personas, disponemos de datos sobre las áreas de sus títulos, las actividades que realizan, el tipo de docencia que ejercen, la dedicación y las áreas de interés. Por razones de espacio no presentaremos en detalle tablas completas con estos datos. El Cuadro II ilustra un aspecto importante: al clasificar a las 549 personas por título y área del título, se encuentra que hay 240 graduados en estadística y 111 en las áreas de matemática, computación e investigación operativa.

Hay también resultados interesantes para los otros atributos: (1) De las 549 personas hay 266 con dedicación exclusiva, 213 con dedicación no exclusiva y 70 que no tienen especificada su dedicación; (2) De las 549 personas hay 86 con categoría de profesor titular o equivalente, 215 con otras categorías de profesor, 104 auxiliares y 144 que no tienen especificada su categoría; (3) Las 19 áreas de interés (ver Apéndice 1) fueron indicadas para personas ocupadas en los centros; sobre este tema formularemos algún comentario en la Sección 4.

El Cuadro III muestra los productos principales de los centros respon-

CUADRO I

Encuesta a Centros Estadísticos Latinoamericanos

Personal de Investigación y Docencia de los Centros, Clasificado por Título

Países	Centros	Total	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Argentina	U.N.Buenos Aires	14	1	1	-	-	12	-	-
	U.N.Córdoba	18	1	1	1	-	13	2	-
	U.N.Comahue	10	-	-	-	1	7	2	-
	U.N.Rosario	72	-	3	-	-	64	4	1
	U.N.Tucumán	18	4	1	1	1	9	2	-
	INTA,Castelar	9	1	-	1	1	6	-	-
Brasil	U.Brasilia	33	9	1	-	19	4	-	-
	USP,Estadística	41	12	2	-	17	10	-	-
	USP,Medicina	9	-	6	-	2	-	-	1
	USP,Agric.L.Queiroz	11	-	9	-	2	-	-	-
	USP,Matem.Sao Carlos	5	2	-	-	3	-	-	-
	U.E.Campinas	26	4	2	2	17	1	-	-
	U.Fed.Pernambuco	29	-	1	-	12	16	-	-
	U.Fed.Rio Janeiro (a)	25	9	5	1	7	3	-	-
	U.Fed.Sao Carlos	14	3	1	-	3	7	-	-
	U.Fed.R.G.do Sul	23	-	5	3	7	8	-	-
	EMBRAPA, Brasilia (a)	27	4	12	1	10	-	-	-
Colombia	U.Andes,Ingeniería	12	5	-	2	3	2	-	-
	C.C.Reg.Población	12	-	5	4	2	-	-	1
	Instituto SER	27	4	1	3	3	14	2	-
Chile	U.Austral de Chile	7	-	-	1	3	3	-	-
	U.Tecn.F.Sta.María	15	2	-	2	8	3	-	-
	CELADE (b)	nd	nd	nd	nd	nd	nd	nd	nd
	CIENES	11	3	-	3	1	4	-	-
Guatemala	Inst.Nutrición	5	2	-	-	2	1	-	-
México	U.N.Autónoma,IIMAS	13	7	-	-	2	4	-	-
	Chapingo,Est.y Cal. (a)	24	10	-	-	14	-	-	-
	Sec.Agric.,INIA	26	1	-	1	6	18	-	-
Venezuela	U.de los Andes (a)	13	6	-	4	2	1	-	-
Total		549	90	56	30	148	210	12	3

(1) Doctor, centros de Estados Unidos, Europa, etc.; (2) Doctor, centros Latinoamericanos u otros; (3) Master, centros de Estados Unidos, Europa, etc.; (4) Master, centros Latinoamericanos u otros; (5) Profesional y pregrado; (6) Estudiante no graduado; (7) No especifica el título.

(a) Datos de la encuesta preliminar 1979-80.

(b) No dispusimos de datos sobre su personal.

CUADRO II

Encuesta a Centros Estadísticos Latinoamericanos
Personal de Investigación y Docencia Clasificado por Título y Area del Título

Area del Título	Total	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Estadística	240	45	20	12	91	71	-	1
Matemática, Computación o Investigación Operativa	111	15	9	5	24	57	-	1
Otras Areas	137	16	10	11	8	80	12	-
No Especifican Area	61	14	17	2	25	2	-	1
Total	549	90	56	30	148	210	12	3

Los encabezamientos (1)-(7) figuran en el Cuadro I.

dentes en 1977, 1978 y 1979. Nótese que para 3 centros, que son importantes por la cantidad y calidad de los recursos humanos con que cuentan, no tenemos datos sobre sus publicaciones. A pesar de ello, los totales del Cuadro III muestran una intensa actividad en el trienio. Hay 183 trabajos publicados en revistas dentro de los respectivos países, 63 en revistas fuera de los países y 22 libros. El tema de las publicaciones del mejor nivel será tratado en detalle en las secciones 3 y 4. Los informes técnicos suman 265 en el trienio. Hay además 17 tesis doctorales y 90 tesis de master realizadas por alumnos regulares de los centros. Finalmente, hay 148 publicaciones de otros tipos: notas para cursos y seminarios, etc.

En general, los cuadros I y II proporcionan dos ideas centrales; (1) Los recursos humanos de nivel superior agrupados en los centros respondentes son muy importantes, y también lo es la producción estadística lograda por ellos en el trienio 1977-79; (2) Existen diferencias considerables entre los centros respondentes en cuanto a su tamaño y a su producción.

Resulta difícil con los resultados disponibles lograr un buen ordenamiento de los centros respondentes en términos de la cantidad y de la calidad de sus recursos humanos, de su producción y de otros elementos considerados en la encuesta. Además tal intento nos alejaría del propósito básico de nuestro trabajo. Sin embargo, para destacar el esfuerzo institucional realizado en algunos casos y enfatizar el potencial desarrollado hasta el presente, en el Cuadro IV presentamos con sentido exploratorio una colección de nueve indicadores.

Los siete primeros indicadores se refieren al personal. El indicador (1) es el total de personas, reproducido del Cuadro I, y vemos que oscila entre 5 y 72 personas, con un promedio de casi 20 personas para los 28 centros que tienen datos. Los indicadores (2) y (3) registran la cantidad de personas con título de postgrado, doctorado o master, cualquiera sea su origen. El primero se refiere al total de personas en tales condiciones, oscila entre 1 y 31, con un promedio de 11,6 personas para los 28 centros con datos; el segundo se refiere al total de personas con título de postgrado en estadística, matemática, computación o investigación operativa: el indicador varía entre 0 y 30 personas, con un promedio de 8,5 para los 26 centros con datos.

Los indicadores (4), (5) y (6) totalizan las personas con títulos de postgrado (doctor o master) recibidos de instituciones de Estados Unidos, Europa, etc.. El primero muestra el total de estos casos, varía entre 0 y 12, con un promedio de 4,7 para los 28 centros respondentes con datos. El indica

CUADRO III

Encuesta a Centros Estadísticos Latinoamericanos

Publicaciones y Trabajos Realizados en 1977, 1978 y 1979

Países	Centros	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Argentina	U.N.Buenos Aires	-	5	-	10	-	-	-
	U.N.Córdoba	-	-	-	-	-	-	12
	U.N.Comahue	-	-	-	-	-	-	-
	U.N.Rosario	18	2	-	9	8	-	-
	U.N.Tucumán	4	5	1	17	-	-	-
	INTA,Castelar	-	2	-	40	-	-	3
Brasil	U.Brasilia	9	3	-	-	-	18	-
	USP,Estadística	32	7	3	1	2	18	1
	USP,Medicina	32	6	2	-	5	8	-
	USP,Agric.L.Queiroz	50	3	8	-	-	26	46
	USP,Matem.Sao Carlos	1	3	-	-	-	4	11
	U.E.Campinas	8	21	-	4	2	10	-
	U.Fed.Pernambuco	-	-	-	-	-	4	-
	U.Fed.Rio Janeiro (a)	nd	nd	nd	nd	nd	nd	nd
	U.Fed.Sao Carlos	-	-	-	-	-	-	-
	U.Fed.R.G.do Sul	-	-	-	-	-	-	-
	EMBRAPA,Brasilia (a)	nd	nd	nd	nd	nd	nd	nd
Colombia	U.Andes,Ingeniería	-	-	-	-	-	-	28
	C.C.Reg.Población	-	-	4	17	-	-	-
	Instituto SER	-	-	-	34	-	-	8
Chile	U.Austral de Chile	-	-	-	-	-	-	-
	U.Tecn.F.Sta.María	-	-	-	-	-	-	3
	CELADE	8	-	2	43	-	-	19
	CIENES	1	3	1	50	-	-	17
Guatemala	Inst.Nutrición	-	-	-	-	-	-	-
México	U.N.Autónoma,IIMAS	5	3	1	19	-	2	-
	Chapingo,Est.y Cal. (a)	nd	nd	nd	nd	nd	nd	nd
	Sec.Agric.,INIA	15	-	-	-	-	-	-
Venezuela	U.de los Andes	-	-	-	21	-	-	-
Total		183	63	22	265	17	90	148

(1) Publicaciones y comunicaciones en revistas nacionales, incluyendo las editadas por el propio centro, o contribuciones en libros; (2) Publicaciones en revistas extranjeras o internacionales, o contribuciones en libros; (3) Libros, individuales o conjuntos; (4) Informes o memorias técnicas; (5) Tesis doctorales realizadas por alumnos regulares del centro; (6) Tesis de master realizadas por alumnos regulares del centro; (7) Otras.

(a) No dispusimos de datos sobre sus publicaciones en el período.

CUADRO IV

Encuesta a Centros Estadísticos Latinoamericanos

Algunos Indicadores del Personal y la Producción de los Centros

Países	Centros	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Argentina	U.N.Buenos Aires	14	2	2	1	1	1	6	5	-
	U.N.Córdoba	18	3	-	2	-	-	3	-	-
	U.N.Comahue	10	1	1	-	-	-	3	-	-
	U.N.Rosario	72	3	2	-	-	-	2	2	8
	U.N.Tucumán	18	7	4	5	3	3	9	5	-
	INTA,Castelar	9	3	3	2	2	2	9	2	-
Brasil	U.Brasilia	33	29	29	9	9	6	14	3	18
	USP,Estadística	41	31	30	12	11	10	28	7	20
	USP,Medicina	9	8	7	-	-	-	8	6	13
	USP,Agric.L.Queiroz	11	11	11	-	-	-	10	3	26
	USP,Matem.Sao Carlos	5	5	5	2	2	1	5	3	4
	U.E.Campinas	26	25	25	6	6	3	26	21	12
	U.Fed.Pernambuco	29	13	13	-	-	-	17	-	4
	U.Fed.Rio Janeiro (a)	25	22	20	10	9	8	14	nd	nd
	U.Fed.Sao Carlos	14	7	7	3	3	3	12	-	-
	U.Fed.R.G.do Sul	23	15	9	3	2	2	12	-	-
	EMBRAPA,Brasilia (a)	27	27	nd	5	nd	nd	nd	nd	nd
Colombia	U.Andes,Ingeniería	12	10	2	7	2	-	4	-	-
	C.C.Reg.Población	12	11	1	4	-	-	-	-	-
	Instituto SER	27	11	3	7	3	-	10	-	-
Chile	U.Austral de Chile	7	4	4	1	1	1	7	-	-
	U.Tecn.F.Sta.María	15	12	12	4	4	-	15	-	-
	CELADE	nd	nd	nd	nd	nd	nd	nd	-	-
	CIENES	11	7	6	6	5	4	11	3	-
Cuatemala	Inst.Nutrición	5	4	4	2	2	1	5	-	-
México	U.N.Autónoma,IIMAS	13	9	8	7	6	6	12	3	2
	Chapingo,Est.y Cal. (a)	24	24	nd	10	nd	nd	nd	nd	nd
	Sec.Agric.,INIA	26	8	7	2	2	2	24	-	-
Venezuela	U.de los Andes (a)	13	12	6	10	4	4	nd	-	-
Total		549	324	221	120	77	57	266	63	107

(1) Total del personal, ver Cuadro I; (2) Títulos de postgrado: doctor y master; (3) Id, en estadística, matemática, computación e investigación operativa; (4) Títulos de postgrado en Estados Unidos, Europa, etc.; (5) Id, en estadística, matemática, computación e i.o.; (6) Id, en estadística; (7) Dedicación exclusiva al centro; (8) Publicaciones extranjeras, ver Cuadro III; (9) Tesis de doctor y master, ver Cuadro III.

(a) Datos de la encuesta preliminar 1979-80. nd=no disponible.

cador (5) se limita a los títulos en estadística, matemática, computación e investigación operativa: oscila entre 0 y 11, con un promedio de 3 personas

para los 26 centros con datos. El indicador (6) se limita a las personas con títulos en estadística: varía entre 0 y 10 y representa un promedio de 2,2 personas para 26 centros.

El indicador (7) registra la cantidad de personas que tienen dedicación exclusiva al centro respondente: oscila entre 0 y 28 personas, con un promedio de 10,6 para 25 centros.

Los indicadores (8) y (9) son sobre publicaciones. El primero reproduce los datos sobre publicaciones en revistas extranjeras del Cuadro III y el segundo es el total de tesis de doctor y master realizadas por los alumnos regulares del centro, ambos indicadores referidos a los años 1977, 1978 y 1979.

Si consideramos como grandes centros de nivel superior en América Latina, dentro de los respondentes, a aquellos que tienen las mayores dotaciones de personal en estadística y sus áreas conexas, con formación superior y dedicación exclusiva, y que además exhiben cantidades importantes de productos del mejor nivel en el lapso considerado, podemos concluir que un primer grupo debería incluír, en el orden arbitrario utilizado en los cuadros I, III y IV, a los siguientes centros: Universidad de Brasilia, Universidad de San Pablo (Estadística), Universidad de Campinas, Universidad Federal de Rio de Janeiro, Universidad Autónoma de México (IIMAS), Centro de Estadística y Cálculo de Chapingo y la Universidad de los Andes, Mérida. Son siete centros, los cuatro primeros de Brasil, los dos siguientes de México y el último de Venezuela. Es de lamentar que la información referente a la Universidad Federal de Río de Janeiro, al Centro de Chapingo y a la Universidad de los Andes, Mérida, esté incompleta en nuestro trabajo.

Después de este primer grupo existen otros centros con buena calidad de recursos humanos y buena producción de nivel superior, que se distinguen de aquellos por su menor tamaño en términos de las cantidades de personas quelos integran. Existen además otros centros interesantes e importantes pero que no poseen recursos humanos con formación superior en el extranjero, o que no han informado sobre producción del mejor nivel a que hacemos referencia. Finalmente existen centros muy pequeños, algunos de los cuales no son propiamente centros específicos de investigación o trabajo en estadística, por lo menos a los niveles que hemos enfatizado en nuestro análisis.

Para EMBRAPA de Brasilia y el CELADE de Chile, disponemos de insuficiente información para intentar un análisis detallado, si bien se conoce que son centros importantes por su tamaño y su producción.

Antes de dejar este análisis, debe enfatizarse que es solamente una primera aproximación al problema, basada exclusivamente en los datos provenientes de la encuesta. Para un análisis a fondo faltaría especificar con más precisión algunos aspectos cualitativos de mucha significación. Debe recordarse que esta es una encuesta por correo y que no hemos tenido oportunidad de visitar a los centros de interés. En realidad, los autores sólo conocen a una minoría de los centros respondentes. En trabajos de este tipo se ha encontrado frecuentemente que el contacto o las visitas personales son un complemento indispensable para perfeccionar el análisis.

La encuesta cubrió otros temas interesantes que comentamos en forma abreviada a continuación: (1) De los 29 centros respondentes 16 son públicos (de ellos 11 son nacionales, 5 provinciales o locales), 8 son privados y 5 no especificaron su dependencia; (2) De los respondentes 14 otorgan títulos de grado o profesionales, 12 otorgan títulos de postgrado y 6 otorgan certificados de postgrado; (3) 10 centros publican informes periódicos de su labor, 15 no los publican y 4 no responden a la pregunta respectiva; (4) La participación de miembros del centro en actividades editoriales en estadísti

ca y sus áreas conexas en el trienio 1977-79 muestran las siguientes cifras: 5 personas fueron editores o editores asociados de revistas estadísticas, 48 se desempeñaron como árbitros y 22 como revisores de trabajos de estadística y sus áreas conexas, todo en el lapso considerado; (5) Finalmente, en los centros respondentes se editan 3 revistas en estadística y sus áreas conexas.

3. Publicaciones de estadísticos latinoamericanos: objetivos y procedimientos

El propósito de este trabajo es cuantificar las publicaciones realizadas por estadísticos latinoamericanos en el lapso 1975-1979, con énfasis en las publicaciones del mejor nivel.

Desde 1975 la "American Statistical Association" (ASA) y el "Institute of Mathematical Statistics" (IMS) publican la revista Current Index to Statistics (CIS) de periodicidad anual, que contiene una nómina de los trabajos de estadística y sus áreas conexas en el mundo. El Volúmen 5 de 1979 contiene aproximadamente 6.700 artículos, extraídos de más de 600 fuentes originales, principalmente revistas, actas de congresos y libros. De cada trabajo se da la cita bibliográfica completa: autor, año, título y fuente; luego se lo clasifica por las palabras claves que (generalmente) aparecen en el título.

En el momento de encarar este trabajo habían aparecido los volúmenes 1 al 5, años 1975 al 1979, que cubren aproximadamente las publicaciones aparecidas en esos mismos años.

El procedimiento seguido consistió en formar una lista de nombres de estadísticos latinoamericanos y buscar en los cinco volúmenes del CIS las publicaciones incluídas bajo esos nombres.

Las listas utilizadas contienen un total general de 2186 nombres, según consta en el Cuadro V. Las listas se originaron como sigue: (1) CIENES, Centro Interamericano de Enseñanza de Estadística, Santiago, Chile: "Egresados

CUADRO V

Lista con Nombres de Estadísticos Latinoamericanos, con Clasificación por Países

País	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Total
Argentina	127	26	141	22	3	13	3	3	-	338
Brasil	59	52	243	52	8	4	4	5	-	427
Bolivia	76	26	-	1	1	-	-	-	-	104
Chile	195	50	33	32	42	8	1	2	-	363
Colombia	92	44	51	27	6	7	-	1	-	228
Guatemala	-	-	5	-	-	-	-	-	-	5
México	55	54	63	41	6	5	2	9	79	314
Paraguay	34	23	-	1	1	-	-	-	-	59
Perú	90	32	-	11	3	-	-	-	-	136
Venezuela	56	26	13	24	3	4	18	4	-	148
Uruguay	44	19	-	-	1	-	-	-	-	64
Total	828	352	549	211	74	41	28	24	79	2186

(1) CIENES; (2) IASI; (3) Encuesta; (4) ASA; (5) SELENES; (6) COLPRIE; (7) Bernoulli; (8) ISI; (9) Otras.

por Cursos, Paises e Indicación del Año Cursado", 1962-1976; (2) IASI, Instituto Interamericano de Estadística: "Latin Americans Receiving some Statistical Training in the United States", 1942-1964; (3) ENCUESTA: "Encuesta a Centros Estadísticos Latinoamericanos", ver secciones 1 y 2 de este informe; (4) ASA: "Directory of Statisticians", 1978; (5) SELENES, Seminario Latinoamericano sobre Enseñanza de Estadística", Santiago, Chile: "Informe Final", 1978; (6) COLPRIE, Comité Latinoamericano para Promover Relaciones Internacionales en Estadística: lista de envíos del "Boletín Informativo", 1979; (7) BERNOULLI, Regional Latinoamericana de la Sociedad Bernoulli para la Estadística Matemática y la Probabilidad: "Boletín Informativo N°1", 1980; (8) ISI, Instituto Internacional de Estadística: "List of Honorary and Ordinary Members by Countries", 1979; (9) Otras fuentes diversas.

Las listas presentan alguna superposición, pues hay personas que aparecen en más de una de ellas. Un cálculo del total de personas, sin repeticiones, es 1813.

Algunos de los problemas generales que pueden originarse con este tipo de diseño del trabajo son los siguientes: (a) Definición de "Latinoamericano". La nacionalidad de las personas en general no estuvo disponible. No conocemos la de origen ni la del momento de formar la lista o la actual. Tampoco dispusimos del domicilio, pasado o actual. Desde estos puntos de vista la información está menos definida por país que para el total; (b) Tratamiento de autores conjuntos. Un caso es el de un autor latinoamericano y otro que no lo es: se optó por incluír todos los trabajos en los que interviene un autor de nuestra nómina.

Existen además algunos problemas que pueden conducir a que el total de trabajos detectados sea menor de lo que podría considerarse la producción de los estadísticos latinoamericanos. Algunos de los más importantes son los siguientes: (a) La nómina de personas es incompleta. No disponemos de censos de los estadísticos latinoamericanos existentes en los años 1975 a 1979. Por otra parte, nuestras listas tienden a concentrarse en estadísticos profesionales por estudio, formación o lugar de trabajo, y tienden a excluir a los especialistas en áreas sustantivas: agrónomos, biólogos, economistas, ingenieros, sociólogos, etc. Igualmente tienden a excluír a los matemáticos, expertos en computación, etc.; (b) La nómina de revistas y otras fuentes tratadas por el CIS, en general excluye muchas de las revistas de los paises latinoamericanos en las que sus estadísticos pueden publicar sus trabajos. De alguna manera puede decirse que el CIS se concentra en fuentes internacionales y de diversos paises, que pueden ser de mayor interés para la disciplina y sus especialistas; (c) La lista de trabajos se refiere en general a los artículos, libros y similares publicados en forma definitiva. Excluye por lo tanto los informes técnicos, las monografías, las tesis y otros trabajos terminados, pero no publicados; (d) Omisiones de trabajos por errores de ejecución.

Un problema que puede actuar en el sentido contrario a los mencionados precedentemente, es que hemos incluído tanto artículos como libros, comentarios a artículos o libros, y toda otra referencia que apareció en el CIS: no realizamos ningún ajuste por la naturaleza del trabajo. Otro problema es que algunos autores, si bien latinoamericanos por su origen, estaban en el momento de escribir sus trabajos fuera del continente, o viven permanentemente fuera de él. Sin embargo, un análisis cualitativo de la información nos indica que ambos problemas no tienen una magnitud suficiente como para invalidar la significación de los hallazgos que comentamos a continuación.

4. Principales resultados del trabajo sobre publicaciones

CUADRO VI

Trabajos de Estadísticos Latinoamericanos Listados en los Volúmenes 1 a 5 del "Current Index to Statistics", Años 1975 a 1979

Clasificación por País y Año de Publicación

País	Año de Publicación					Total
	1975	1976	1977	1978	1979	
Argentina	7	5	4	4	6	26
Brasil	8	8	7	4	-	27
Chile	1	-	2	-	1	4
Colombia	1	-	3	2	-	6
México	4	2	2	-	2	10
Uruguay	-	1	2	1	-	4
Venezuela	1	-	2	3	1	7
Total	22	16	22	14	10	84

Un resultado básico del estudio está en el Cuadro VI. Un total de 84 trabajos fué encontrado en el lapso 1975-1979. La distribución por paises muestra mayores cantidades en Brasil (27), Argentina (26) y México (10), y cantidades menores en Venezuela, Colombia, Chile y Uruguay.

Existen diferencias entre los totales de los años considerados, con un máximo de 22 trabajos en 1975 y 1977 y un mínimo de 10 en 1979. No es fácil explicar esta cantidad baja en 1979.

El Cuadro VII presenta una clasificación por temas, utilizando el esquema de la American Mathematical Society (AMS) de 1979. La asignación de temas fué hecha en base al título del trabajo, lo que podría dar lugar a algunos cambios si se conociera el contenido de la publicación. Hay 59 trabajos de estadística, 24 de probabilidades y 1 de computación. Las áreas estadísticas mayoritarias son inferencia en procesos estocásticos (10), aplicaciones (10), inferencia paramétrica (9) y regresión y correlación (9). Las áreas principales en probabilidades son teoremas límites (6) y procesos estocásticos (6).

En general el interés de los estadísticos latinoamericanos aparece como repartido entre los principales temas. Omisiones importantes son teoría del muestreo y encuestas por muestreo (62D) y estadística en la ingeniería (62N). Se podría conjeturar que estas omisiones se deben a que la mayoría de los que publican están asociados con instituciones de tipo académico, mientras que los temas faltantes son especificos de los sectores gubernamentales y de las empresas, respectivamente.

Otra información disponible es la fuente, es decir la revista o el editor del libro donde apareció la publicación. Los 84 trabajos aparecieron en un total de 43 fuentes distintas. De ellas 4 son total o parcialmente en español, 2 en alemán, 1 en francés, 1 en italiano y el resto en inglés. De las fuentes internacionales hay 14 trabajos en The Annals of Statistics y The Annals of Probability, 5 en Estadística y 5 en publicaciones del Instituto Internacional de Estadística (ISI).

Con respecto a la revista Estadística del IASI, una revisión de los volúmenes publicados entre 1975 y 1979 mostró 7 trabajos publicados por estadísticos latinoamericanos que no figuran en el CIS. Ellos corresponden a estadísticos de Brasil (2), Chile (1), Colombia (1), México (1), República Dominicana (1) y Uruguay (1). Respetando el enfoque seguido en este estudio, los

CUADRO VII

Trabajos de Estadísticos Latinoamericanos Listados en los Volúmenes 1 a 5 del "Current Index to Statistics". Años 1975 a 1979

Clasificación por Año de Publicación y Tema (AMS 1979)

Tema		Año de Publicación					Total
Código	Nombre	1975	1976	1977	1978	1979	
62	ESTADISTICA	-	1	1	-	-	2
62A	Fundamentos	1	3	-	-	-	4
62B	Suficiencia	-	-	-	-	-	-
62C	Teoría de la decisión	-	1	2	-	-	3
62D	Teoría del muestreo, encuestas por muestreo	-	-	-	-	-	-
62E	Teoría de las distribuciones	-	-	-	-	-	-
62F	Inferencia paramétrica	4	-	3	1	1	9
62G	Inferencia no paramétrica	-	1	-	-	-	1
62H	Análisis multivariado	1	-	1	-	1	3
62J	Regresión y correlación	4	1	-	3	1	9
62K	Diseños experimentales	1	-	1	3	-	5
62L	Métodos secuenciales	1	-	1	1	-	3
62M	Inferencia en procesos estocásticos	3	1	3	1	2	10
62N	Estadística en la ingeniería	-	-	-	-	-	-
62P	Aplicaciones	3	-	7	-	-	10
62Q	Tablas estadísticas	-	-	-	-	-	-
	Total ESTADISTICA	18	8	19	9	5	59
60	PROBABILIDAD	-	-	-	-	-	-
60A	Fundamentos de la teoría de la probabilidad	-	1	-	-	-	1
60B	Teoría de la probabilidad en estructuras algebraicas y topológicas	-	1	-	2	1	4
60C	Probabilidad combinatoria	-	1	-	-	-	1
60D	Probabilidad geométrica; geometría estocástica; conjuntos aleatorios	-	1	-	1	1	3
60E	Teoría de las distribuciones	1	-	-	-	-	1
60F	Teoremas límites	2	2	1	-	1	6
60G	Procesos estocásticos	1	1	1	1	2	6
60H	Análisis estocástico	-	-	-	-	-	-
60J	Procesos de Markov	-	-	-	1	-	1
60K	Procesos especiales	-	-	1	-	-	1
	Total PROBABILIDAD	4	7	3	5	5	24
68J	Simulación (computación)	-	1	-	-	-	1
	Total General	22	16	22	14	10	84

trabajos no fueron agregados a los 84 extraídos del CIS.

Es indudable que existe una asociación fuerte entre la publicación en revistas norteamericanas o editadas en Norteamérica, y el hecho de que muchos de los estadísticos latinoamericanos que publicaron hayan estudiado en este país.

Otra clasificación intentada fué por país y fuente (además de por año). No encontramos observaciones de interés en los datos resultantes.

Al analizar las publicaciones de las distintas personas, encontramos que varias personas publicaron más de un trabajo: una publicó 7, dos publicaron 5, cinco publicaron 4 (dos de ellas como coautores), cinco publicaron 3 y 3 publicaron 2 cada uno. Un total de 61 personas distintas, que tomamos como estadísticos latinoamericanos, aparecen como autores o coautores de uno o más trabajos en la nómina que formamos. Esto representa en el orden del 3% del total de personas incluídas en nuestra indagación, lo que aparentemente se alejaría bastante de la experiencia tenida en estudios de este tipo. Para los Estados Unidos se estimó que "para la profesión de economista como un todo, la producción de artículos es probablemente uno por persona cada 20 años" (G.J. Stigler).

5. Resumen y conclusiones

Encuesta a Centros Latinoamericanos

La encuesta de 1981 condujo a un total de 29 respuestas sobre 106 cuestionarios enviados por correo. Sin embargo, es posible aseverar que la mayoría de los principales centros académicos superiores de América Latina que operan al nivel de postgrado y que se dedican en forma intensiva al trabajo o a la investigación en temas estadísticos, está entre los respondentes.

Para justificar esta aseveración tenemos en cuenta la relación existente entre la encuesta y el trabajo presentado en las secciones 3 y 4 de este informe. En efecto, mientras las publicaciones de estadísticos latinoamericanos incluídas en el CIS suman 46 para los años 1977, 1978 y 1979 (y 52 cuando agregamos 6 de Estadística que no están en el CIS hasta 1979), los centros respondentes informan un total de 63 publicaciones en revistas extranjeras de esos años (Cuadro III). Es decir que los centros respondentes publicaron en revistas que se presuponen del mejor nivel, un total superior al consignado en el CIS del período correspondiente.

Existen diferencias importantes entre las definiciones que dan lugar a estas dos cifras, algunas de las cuales fueron exploradas en la Sección 3. A ellas debe agregarse que desde el punto de vista de un centro, las revistas de otro país latinoamericano son extranjeras. Debe notarse que ninguna de las dos colecciones de trabajos es un subconjunto de la otra, aunque existe una considerable superposición.

Algunos de los principales indicadores que pueden construirse con los datos disponibles sobre el personal y la producción de los centros respondentes, constituyen el Cuadro IV. Es importante destacar que los centros que informaron sobre su dotación de personal tienen un total de 549 personas dedicadas a docencia e investigación, de los cuales 324 tienen títulos de postgrado. Una cifra interesante es que los centros respondentes tienen un total de 45 doctorados en estadística que recibieron sus títulos en centros de Estados Unidos, Europa, etc.

Un aspecto que debe enfatizarse es que la encuesta no estuvo ligada al tema de la enseñanza de la estadística, sobre todo al de la enseñanza de grado. Es conocido que en varios países existen centros importantes cuya principal actividad es la enseñanza de grado; de ellos sólo ocasionalmente quedó alguno incluído entre nuestros respondentes. En realidad, nuestro tratamiento de la propia enseñanza de postgrado es incompleto, pues nos limitamos a considerar las tesis publicadas como trabajo del centro: hay centros que no las exigen dentro de los programas de master, por ejemplo.

Otro aspecto que no se tuvo en cuenta en forma directa es la actividad en el área de las estadísticas oficiales de los países y en el campo internacional. Se advierte que posiblemente la mayor fuente de ocupación para los estadísticos profesionales es precisamente este sector gubernamental.

Publicaciones de Estadísticos Latinoamericanos

La información obtenida en este trabajo puede evaluarse con un sentido más o menos optimista. Si se compara el total de 84 (ó 91 si agregamos los de la revista Estadística) trabajos publicados en 5 años en 43 fuentes distintas, con los 6.700 trabajos provenientes de más de 600 fuentes que figuran en el CIS del año 1979 solamente, queda en descubierto el modesto desarrollo de la estadística en América Latina, en comparación con otras partes del mundo.

Sin embargo, una visión más optimista debe señalar lo siguiente: (1) La cantidad y calidad de los trabajos publicados equivale a casi una revista periódica de buen nivel, que podría imputarse a America Latina como si todos estos esfuerzos hubieran estado consolidados; (2) Es posible conjeturar que si esta indagación se hubiera realizado en el pasado, digamos hace 10 años, los resultados hubieran sido significativamente menores.

Con respecto al primer comentario, no debe interpretárselo como una propuesta para que tal revista se intente, por lo menos no necesariamente. Es indudable que una de las razones para la existencia de los trabajos detectados es la presencia de fuentes prestigiosas que motivan a los estadísticos para publicar en ellas. Sin embargo, la existencia de estos trabajos del nivel más elevado deja en claro la existencia de una infraestructura estadística, sobre todo en los sectores académicos, que sirve de base para otras publicaciones de interés menos general, y que puede explorarse para posibles desarrollos futuros. Esta aseveración está confirmada por los datos sobre producción de los centros que respondieron a la encuesta, Cuadro III.

Conclusiones

Una persona interesada en conocer el desarrollo de la disciplina y la profesión estadística en América Latina podría válidamente preguntarse: (1) Se registra un progreso significativo a través del tiempo, por ejemplo en los últimos 20 o 30 años?; (2) En el momento actual, está la estadística suficientemente desarrollada en los distintos países y en la región como un todo, en comparación con otras disciplinas de interés para la sociedad, o de acuerdo con el nivel político, económico y social alcanzado?

Los trabajos empíricos que hemos presentado no ofrecen respuestas directas a estos interrogantes: son aportes para que en conjunto con otras fuentes de información, se tienda a dar respuestas bien justificadas. Un primer signo de su valor será si consiguen generar motivación suficiente para que aquellas y otras preguntas similares se analicen exhaustivamente.

Además los trabajos han logrado reunir un conjunto interesante de datos estadísticos. A pesar de sus limitaciones, es indudable que la encuesta ha mostrado con claridad que en el sector académico de varios países latinoamericanos existe un proceso (que tiene distintos estados de avance según los

paises), que ha logrado un destacable grado de institucionalización. Existen centros estadísticos bien dotados de personal con formación superior, dedicados a la producción de estudios e investigaciones del mejor nivel.

Parece indudable que la cantidad y calidad de los recursos existentes en los centros estadísticos latinoamericanos, hasta donde nos permiten inferir los resultados de esta sencilla encuesta, abren una perspectiva auspiciosa para el desarrollo integral de la estadística en los paises de la región.

Igual tipo de conclusión surge al considerar la información del trabajo sobre publicaciones de los estadísticos latinoamericanos.

Con respecto a las tendencias en el tiempo, a pesar de la falta de información detallada, se observa en la América Latina actual una cierta efervecencia en el campo estadístico, liderada por los sectores académicos. Los resultados de estos trabajos contribuyen a confirmar esta idea. Lo que seguramente falta es continuar vigorizando este esfuerzo, dentro de cada país y en los planos regional e internacional, tratando de extenderlo a sectores de aplicación que han resultado claves en otras partes del mundo: las aplicaciones de estadística en los sectores gubernamentales, científicos dedicados a áreas sustantivas, y de las empresas.

BIBLIOGRAFIA. R.P.Mentz y N.M.Jarma de Cortés (1980), "Encuesta a Centros Estadísticos Latinoamericanos, Informe Preliminar", mimeografiado. R.P.Mentz y N.M. Jarma de Cortés (1981), "Encuesta a Centros Estadísticos Latinoamericanos, Informe Final", aparecerá como Cuaderno del INIE; R.P.Mentz y M.B.Lopez (1981), "Publicaciones de Estadísticos Latinoamericanos", aparecerá como Cuaderno del INIE. G.J.Stigler, "Do Economist Matter?", mimeografiado.

Agradecimientos. El procesamiento de los datos sobre publicaciones de estadísticos latinoamericanos (secciones 3 y 4) estuvo a cargo de María Beatriz López. Los resultados de la Encuesta Preliminar de 1979-1980 fueron analizados en una reunión interna del INIE y luego en reuniones del COLPRIE: en esos casos se recibieron una serie de sugerencias y comentarios de mucho valor. Agradecemos al Profesor Víctor Jorge Elías por sus comentarios y el interés puesto en nuestro trabajo.

Summary. A total of 29 Latin American centers provided detailed information about their teaching and research personnel as of 1981, and about their output (publications and other) for 1977, 1978 and 1979. These centers represent a good coverage of large, academic, statistical centers devoted to research and graduate-level activities in the region. A parallel investigation is reported in which the Current Index in Statistics volumes for 1975-1979 were searched for publications made by Latin Americans. Information from these sources show a reasonable agreement. It is argued that the quantity and quality of the human resources and of their statistical production, together with the interesting degree of institutionalization that has been achieved, can serve as a basis for a strong development of statistics in Latin American in the near future.

Résume. Un total de 29 centres en Amérique du Sud ont donné une information detaillée sur leur personnel d'enseignement et d'investigation en 1981, et aussi sur leur production en 1977, 1978, et 1979. Ces centres ci, représentent une couverture des centres statistiques du genre académique de grand volume, et travaillent sur l'investigation et les activités au niveau du doctorat dans toute cette région. Parallèlement, les volumes 1975-1979 revue "Current Index Statistics", ont été examinés pour extraire les publications faites par les sud américains. L'information qui vient de ces sources nous montre un raisonnable degré de fermeté. On soutient que la quantité de resources humaines et leur production statistique avec le degré tres intéressant d'institutionalisation atteint, ils peuvent servir de base a un dévellopement statistique en Amérique du Sud dans un futur prochain.

Apéndice 1: Encuesta a Centros Estadísticos Latinoamericanos
Cuestionario Utilizado en la Encuesta

ENCUESTA A CENTROS ESTADISTICOS LATINOAMERICANOS

Auspiciada por el COLPRIE, Comité Latinoamericano para Promover Ralaciones Internacionales en Estadística

- 1980 -

1. Nombre Completo y Sigla (Abreviatura) del Centro .....
.....
.....

2. Ubicación o Sede Principal
.....
.....
.....

Dirección Postal
.....
.....
.....

3. Nombre de la Institución de la que depende: .....
.....
3.1 ☐ Gobierno Nacional o Federal 3.3 ☐ Institución o Empresa
3.2 ☐ Gobierno Provincial o Local 3.4 ☐ Otros
(marcar todas las casillas que correspondan)

4. Nombre del Director actual.....

5. Objetivos dentro del campo de la Estadística y sus Areas Conexas
Nota: En Estadística y sus Areas Conexas incluímos los siguientes temas: Estadística Teórica y Aplicada, Probabilidades, Bioestadística, Econometría, Demografía, Investigación Operativa, Computación en Estadística y temas similares.
5.1 ☐ Enseñanza a nivel de Grado o Profesional: Estadístico, Licenciado
5.2 ☐ Enseñanza a nivel de Postgrado: Master, Doctor
5.3 ☐ Trabajo en el Area de las Estadísticas Gubernamentales
5.4 ☐ Consultoría, Apoyo a la Investigación Aplicada, Trabajos para Empresas
5.5 Otros (especificar)
.....
.....

6. Otorga en forma regular Títulos o Certificados en Estadística y sus Areas Conexas? 6.1 ☐ si 6.2 ☐ no
6.3 ☐ Títulos de Grado o Nivel Profesional: Estadístico, Licenciado
6.4 ☐ Títulos de Postgrado: Master, Doctor
6.5 ☐ Certificados de Postgrado
6.6 Otros (especificar)
.....
.....

7. Se Publican informes periódicos de la labor del Centro? 7.1 ☐ si 7.2 ☐ no
Nombre de la publicación .....
Periodicidad .....
Desde cuando se publica? .....
Nota: Se ruega acompañar un ejemplar del último informe.

8. Cantidad de publicaciones y trabajos terminados en Estadística y sus Areas Conexas, realizados por miembros del Centro en 1977, 1978, 1979.

	1977	1978	1979
8.1 Publicaciones y comunicaciones en revistas nacionales, incluyendo las editadas por el propio Centro, ocontribuciones en libros			
8.2 Publicaciones en revistas extranjeras o internacionales, o contribuciones en libros			
8.3 Libros, individuales o conjuntos			
8.4 Informes o memorias técnicas			
8.5 Tesis doctorales realizadas por alumnos regulares del Centro			
8.6 Tesis de master realizadas por alumnos regulares del Centro			
8.7 Otras (especificar el tipo)			
TOTAL			

Nota: Acompañar una nómina detallada de los trabajos y publicaciones si está disponible

9. Cantidad de Miembros del Centro que tuvieron actividades editoriales en Estadística y sus Areas Conexas en 1977, 1978, 1979.

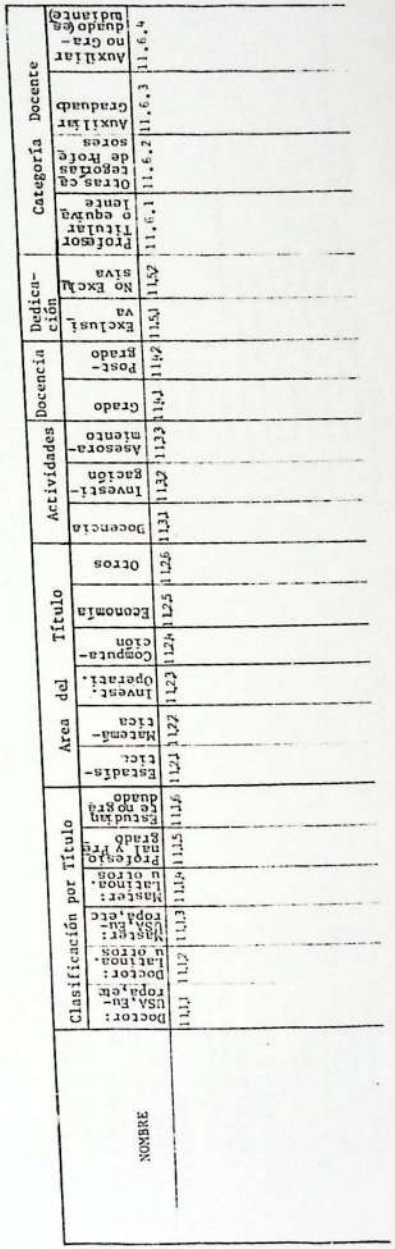
	1977	1978	1979
9.1 Editor o editor asociado de una revista estadística			
9.2 Referee (árbitro)			
9.3 Reviewer (revisor)			
9.4 Otras actividades (especificar).....			
TOTAL			

10. Se editan actualmente en el Centro una o más revistas periódicas en Estadística y sus Areas Conexas? 10.1 ☐ si 10.2 ☐ no

Nombre .....
Periodicidad .....
Desde cuándo se edita en el Centro? .....

11. Personal del Centro (ver planilla adjunta)

Uno de los objetivos de esta encuesta es mostrar el potencial de los Centros Latinoamericanos dedicados a la estadística y sus áreas conexas, para editar revistas estadísticas, publicar artículos teóricos y aplicados, etc. Por esta razón nos será de suma utilidad conocer la formación académica de cada uno de los integrantes del Centro como así también sus áreas de interés en investigación, etc.



11. PERSONAL DEL CENTRO: Instrucciones

Para llenar las planillas, coloque la nómina completa de los miembros del Centro dedicados a docencia e investigación. Para cada persona marque con una cruz (x) la columna correspondiente.

Las clasificaciones siguientes deberán contener una sola cruz (x) por persona:

11.1 Clasificación por Título
11.2 Area del Título
11.5 Dedicación
11.6 Categoría Docente

Las clasificaciones siguientes pueden contener más de una cruz (x) por persona:

11.3 Actividades
11.4 Docencia
11.7 Areas de Interés

CLAVES

11.1 Clasificación por Título
- 11.1.1 Doctor: Centros de Estados Unidos, Europa, etc
- 11.1.2 Doctor: Centros Latinoamericanos u Otros
- 11.1.3 Master: Centros de Estados Unidos, Europa, etc.
- 11.1.4 Master: Centros Latinoamericanos u Otros
- 11.1.5 Profesional y Pregrado
- 11.1.6 Estudiante no Graduado

11.2 Area del Título
- 11.2.1 Estadística
- 11.2.2 Matemática
- 11.2.3 Investigación Operativa
- 11.2.4 Computación
- 11.2.5 Economía
- 11.2.6 Otros

11.3 Actividades
- 11.3.1 Docencia
- 11.3.2 Investigación
- 11.3.3 Asesoramiento

11.4 Docencia
- 11.4.1 Grado
- 11.4.2 Postgrado

11.5 Dedicación
- 11.5.1 Exclusiva
- 11.5.2 No Exclusiva

11.6 Categoría Docente
- 11.6.1 Profesor Titular o Equivalente
- 11.6.2 Otras Categorías de Profesores
- 11.6.3 Auxiliar Graduado
- 11.6.4 Auxiliar no Graduado (Estudiante)

11.7 Areas de Interés
- 11.7.1 Matemática
- 11.7.2 Probabilidades y Procesos Estocásticos
- 11.7.3 Teoría de las Distribuciones, Inferencia Paramétrica
- 11.7.4 Inferencia no Paramétrica
- 11.7.5 Análisis Multivariado
- 11.7.6 Regresión y Correlación
- 11.7.7 Diseño de Experimentos
- 11.7.8 Técnicas de Muestreo
- 11.7.9 Métodos Secuenciales
- 11.7.10 Inferencia en Procesos Estocásticos
- 11.7.11 Aplicaciones
- 11.7.12 Computación
- 11.7.13 Demografía
- 11.7.14 Métodos Numéricos, Investigación Operativa
- 11.7.15 Análisis Numérico
- 11.7.16 Tablas
- 11.7.17 Simulación
- 11.7.18 Teoría Estadística
- 11.7.19 Otros

Apéndice 2: Encuesta a Centros Estadísticos Latinoamericanos
Nómina de Respondentes y Abreviaturas Usadas en este trabajo

U.N.Buenos Aires	Laboratorio de Estadística, Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.
U.N.Córdoba	Instituto de Estadística, Facultad de Ciencias Económicas, Universidad Nacional de Córdoba.
U.N.Comahue	Departamento de Matemática y Estadística, Facultad de Economía y Administración, Universidad Nacional del Comahue, Neuquen.
U.N.Rosario	Escuela de Estadística, Facultad de Ciencias Económicas, Universidad Nacional de Rosario.
U.N.Tucumán	Instituto de Investigaciones Estadísticas, Facultad de Ciencias Económicas, Universidad Nacional de Tucumán.
INTA, Castelar	Departamento de Estadística, Instituto Nacional de Tecnología Agropecuaria, Castelar.
U. Brasilia	Departamento de Estadística, Universidad de Brasilia.
USP, Estadística	Instituto de Matemática y Estadística, Universidad de San Pablo, San Pablo.
USP, Medicina	Departamento de Epidemiología (Area de Estadística), Facultad de Salud Pública, Universidad de San Pablo, San Pablo.
USP,Agric.L.Queiroz	Departamento de Matemática y Estadística, Escuela Superior de Agricultura Luiz Queiroz, Universidad de San Pablo, Piracicaba.
USP,Matem.Sao Carlos	Instituto de Ciencias Matemáticas de San Carlos, Universidad de San Pablo, San Carlos.
U.E.Campinas	Instituto de Matemática, Estadística y Ciencias de la Computación, Universidad Estatal de Campinas.
U.Fed.Pernambuco	Departamento de Estadística e Informática, Universidad Federal de Pernambuco, Recife.
U.Fed.Rio Janeiro	Instituto de Matemática, Universidad Federal de Rio de Janeiro.
U.Fed. Sao Carlos	Departamento de Computación y Estadística, Universidad Federal de San Carlos.
U.Fed.R.G. do Sul	Departamento de Estadística, Instituto de Matemática, Universidad Federal de Río Grande do Sul, Porto Alegre.
EMBRAPA, Brasilia	Empresa Brasileira de Pesquisa Agropecuaria, Brasilia.
U.Andes, Ingeniería	Departamento de Ingeniería Industrial, Universidad de los Andes, Bogotá.
C.C.Reg. Población	Corporación Centro Regional de Población, Bogotá.
Instituto SER	Instituto SER de Investigación, Bogotá.
U.Austral de Chile	Escuela de Estadística, Universidad Austral de Chile, Valdivia.
U.Tecn.F.Sta.María	Departamento de Matemática, Facultad de Ciencias, Universidad Técnica Federico Santa María, Valparaiso.
CELADE	Centro Latinoamericano de Demografía, Naciones Unidas, Santiago.
CIENES	Centro Interamericano de Enseñanza de Estadística, OEA, Santiago.
Inst.Nutrición	División de Estadística, Instituto de Nutrición de Centro América y Panamá, Guatemala.
U.N.Autónoma, IIMAS	Instituto de Investigaciones en Matemática Aplicada y en Sistemas, Universidad Nacional Autónoma de México, México D.F.
Chapingo, Est. y Cal.	Centro de Estadística y Cálculo, Colegio de Postgraduados, Chapingo.
Sec.Agric., INIA	Instituto Nacional de Investigaciones Agrícolas, Secretaría de Agricultura y Ganadería, México D.F.
U.de los Andes	Facultad de Economía, Universidad de los Andes, Mérida.

Notes on the estimation of parameters in vector autoregressive models

T. W. Anderson, Stanford University
Raúl P. Mentz, University of Tucumán

1. Introduction

In the current literature there exists interest in studying vector models for time series that have a linear structure and a finite number of parameters. In particular, the vector autoregressive model is attractive for applications in various areas.

In the present paper we consider the estimation of parameters in vector autoregressive models. In the estimation by means of the sample Yule-Walker equations, we prove that under general conditions the estimators have properties analogous to those satisfied by the parameters of a stationary process; this is an extension of a result found useful in the study of scalar time series. In the area of maximum likelihood estimation, we present some results that are extensions of those found useful in the study of the structure of the likelihood function for the scalar case; see Anderson and Mentz (1980).

2. The model and its basic properties

Let us consider the m-component vector autoregressive process of order p

$$y_t + B_1 y_{t-1} + \dots + B_p y_{t-p} = u_t, \qquad t=\dots,-1,0,1,\dots , \quad (2.1)$$

where the y_t are observable random vectors of finite size $m \geq 1$, the B_j are $m \times m$ matrices of parameters, and the u_t are

unobservable, independent random vectors with $Eu_t = 0$, $Eu_t u_s' = \delta_{st} V$, where δ_{st} is Kronecker's delta function. We assume that V is positive definite. This model has $m^2 p$ parameters in the B_j and $\frac{1}{2}m(m+1)$ parameters in V.

Let

$$B(z) = \sum_{j=0}^{p} B_j z^{p-j} , \qquad (2.2)$$

where $B_0 = I$. The process is stationary and y_t independent of $u_{t+1}, u_{t+2}, \ldots$ provided all roots $z_1, \ldots, z_{mp}$ of the determinantal equation

$$|B(z)| = 0 \qquad (2.3)$$

are less than one in absolute value. In this case the process is invertible into an infinite moving average

$$y_t = \sum_{j=0}^{\infty} A_j u_{t-j} , \qquad (2.4)$$

where $A_0 = I$. See, for example, Hannan (1970). Anděl (1971) analyzed alternative forms of the condition for stationarity.

Let $\{\Sigma_j\}$ denote the covariance sequence of the process,

$$\Sigma_j = E y_t y_{t-j}', \qquad j=0, \pm 1, \pm 2, \ldots , \qquad (2.5)$$

where $\Sigma_h = \Sigma_{-h}'$. These satisfy the Yule-Walker equations

$$\begin{aligned} &\Sigma_0 + B_1 \Sigma_1' + \ldots + B_p \Sigma_p' = V, \\ &\Sigma_s + B_1 \Sigma_{s-1} + \ldots + B_p \Sigma_{s-p} = 0, \qquad s=1,2,\ldots . \end{aligned} \qquad (2.6)$$

The spectral density matrix is

$$f(\lambda) = \frac{1}{2\pi} B(e^{i\lambda})^{-1} V \overline{B}'(e^{i\lambda})^{-1} = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\lambda h} \Sigma_h , \qquad (2.7)$$

where the bar denotes complex conjugate, and the second equality holds provided the series converges. The inversion formula is then

$$\Sigma_h = \int_{-\pi}^{\pi} e^{i\lambda h} f(\lambda) \, d\lambda . \qquad (2.8)$$

Some special cases may be illustrative. If all B_j and V are diagonal, (2.1) represents m scalar autoregressive processes of order p, with independent probabilistic structures. If $p = 1$, we have the first-order vector autoregressive process of size m,

$$y_t + B y_{t-1} = u_t, \qquad t=\ldots,-1,0,1,\ldots . \qquad (2.9)$$

In this case the Yule-Walker equations give

$$\Sigma_s = (-B)^s \Sigma_0, \qquad \Sigma_0 - B \Sigma_0 B' = V. \qquad (2.10)$$

The m-component vector autoregressive process of order p can also be written as an mp-component vector autoregressive process of order 1. In effect, letting $Y_t^* = (y_t', y_{t-1}', \ldots, y_{t-p+1}')'$, $U_t^* = (u_t', 0', \ldots, 0')'$,

$$B^* = \begin{bmatrix} B_1 & B_2 & B_3 & \ldots & B_{p-1} & B_p \\ -I & 0 & 0 & \ldots & 0 & 0 \\ 0 & -I & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & & -I & 0 \end{bmatrix} , \qquad (2.11)$$

the model (2.1) can be written as

$$Y_t^* + B^* Y_{t-1}^* = U_t^*, \qquad t=\ldots,-1,0,1,\ldots , \qquad (2.12)$$

and the roots of (2.3) are those of

$$|B^* + \lambda I| = 0. \qquad (2.13)$$

See Anderson (1971a, Section 5.3) for an analysis of this transformation in the scalar case. General references for the vector

autoregressive process are Hannan (1970), Kashyap and Rao (1976), and Quenouille (1968).

With observations $y_1,\ldots,y_T$ we want to estimate the B_j and V. Two frequently used procedures are to estimate the B_j by means of the sample Yule-Walker equations, or to estimate all parameters by maximum likelihood under the additional assumption of normality.

3. Estimation by means of the sample yule-walker equations

We consider estimation of the B_j by the sample analog of (2.6) for s=1, ..., p; namely,

$$\sum_{j=1}^{p} B_j C_{s-j} = - C_s, \qquad s=1, \ldots, p, \tag{3.1}$$

where

$$C_s = C_{-s}' = \sum_{t=s+1}^{T} y_t y_{t-s}', \qquad s=0,1, \ldots, p. \tag{3.2}$$

We want to prove that the estimators of the B_j in (3.1) arising from these sample Yule-Walker equations are such that the corresponding roots of (2.3) are those of a stationary process. The proof generalizes that in Anderson (1971b). Whittle (1963) used a similar argument to show that the relation holds in terms of the parameters of the underlying process defined by (2.1). De Jong (1976) considered solving recursively systems like (3.1) for increasing values of p.

Proposition 1. Let $(y_1, \ldots, y_T)$ *be an* mxT *matrix of real numbers,* $T \geq mp-p+1$ *and define*

$$C^* = \begin{bmatrix} C_0 & C_1 & \cdots & C_{p-1} \\ C_1' & C_0 & \cdots & C_{p-2} \\ \vdots & \vdots & & \vdots \\ C_{p-1}' & C_{p-2}' & \cdots & C_0 \end{bmatrix}. \tag{3.3}$$

If C^* *is nonsingular, then the roots of (2.3) where* $B_1, \ldots, B_p$ *satisfy (3.1) are less than one in absolute value.*

Proof. Let B^* be as in (2.11), and

$$\bar{C} = \begin{bmatrix} C_1 & C_2 & \cdots & C_p \\ C_0 & C_1 & \cdots & C_{p-1} \\ C_1' & C_0 & \cdots & C_{p-2} \\ \vdots & \vdots & & \vdots \\ C_{p-2}' & C_{p-3}' & \cdots & C_1 \end{bmatrix}. \tag{3.4}$$

Then (3.1) is identical to

$$B^* C^* = -\bar{C}, \tag{3.5}$$

and the roots of (2.3) are identical to the characteristic roots of $B^* = -\bar{C}\, C^{*-1}$. A characteristic root λ and corresponding left-sided characteristic vector $u^{*'} = (u_1', \ldots, u_p')$ satisfy

$$-u^{*'} \bar{C} = \lambda\, u^{*'} C^*. \tag{3.6}$$

Let $y_t^{*'} = (y_t', y_{t-1}', \ldots, y_{t-p+1}')$, t=0,1,...,T+p, where $y_{-p+1} = y_{-p+2} = \ldots = y_0 = 0$ and $y_{T+1} = y_{T+2} = \ldots = y_{T+p} = 0$. Normalize u^* so

$$\begin{aligned} 1 = u^{*'} C^* \overline{u^*} &= \sum_{t=1}^{T+p} (u^{*'} y_t^*) \overline{(u^{*'} y_t^*)} \\ &= \sum_{t=1}^{T+p} (u^{*'} y_{t-1}^*) \overline{(u^{*'} y_{t-1}^*)}, \end{aligned} \tag{3.7}$$

where the bars denote complex conjugation. Then multiplication of (3.6) on the right by $\overline{u^*}$ gives

$$\lambda = -u^{*'} \bar{C}\, \overline{u^*} = - \sum_{t=1}^{T+p} (u^{*'} y_t^*) \overline{(u^{*'} y_{t-1}^*)}. \tag{3.8}$$

By the Cauchy-Schwarz inequality, $|\lambda| \leq 1$. Equality occurs only if $u^{*'} y_t^* = k\, u^{*'} y_{t-1}^*$, t=1, ..., T+p, for some $k \neq 0$. Let τ

be the smallest value of t for which $u^{*'} y_t^* \neq 0$. But $u^{*'} y_{\tau-1}^* = 0$, which leads to a contradiction. Q.E.D.

Proposition 2. When $m=1$, *the conclusion of the proposition itself holds if* $y_t \neq 0$ *for at least one value of* t.

Proof. Let τ be the smallest value of t for which $y_t \neq 0$. Then the matrix with y_{t-i+1} in the tth row and ith column is

$$\begin{bmatrix} 0 & 0 & \cdots & 0 \\ y_\tau & 0 & \cdots & 0 \\ y_{\tau+1} & y_\tau & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ y_{\tau+p-1} & y_{\tau+p-2} & \cdots & y_\tau \\ \vdots & \vdots & & \vdots \\ y_T & y_{T-1} & \cdots & y_{T-p+1} \\ 0 & y_T & \cdots & y_{T-p+2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & y_T \end{bmatrix}. \qquad (3.9)$$

The submatrix with $y_\tau I$ as its main diagonal is nonsingular and hence has full row rank and then C^* is nonsingular. Q.E.D.

Remark 1. Whittle (1963) assumes that (in distribution terms) no linear combination of $y_t, y_{t-1}, \ldots, y_{t-p+1}$ is identically 0 in order that (3.1) has a solution. In the sample here this latter assumption is replaced by C^* being nonsingular. To rule out $|\lambda| = 1$, he assumes no linear combination of $y_t, y_{t-1}, \ldots, y_{t-p+1}, y_{t-p}$ is identically 0. Our proof of Proposition 1 shows that the condition is unnecessary.

Remark 2. By (3.7) and (3.8), $|\lambda| = 1$ if and only if $u^{*'} y_t^* = k\, u^{*'} y_{t-1}^*$. The condition is

$$u_1' y_t + (u_2' - k\, u_1')y_{t-1} + \ldots + (u_p' - k\, u_{p-1}')y_{t-p+1} - k\, u_p' y_{t-p} = 0. \qquad (3.10)$$

But these are not completely general linear combinations. Hence, Whittle's comment (p. 134) that there can be linear combinations of $y_t, \ldots, y_{t-p}$ identically 0 without the root being 1 in absolute value.

Remark 3. In order to obtain an estimate of V that is nonsingular (with probability 1), we would assume $T \geq m(p+1)$, thus insuring the condition of Proposition 1.

Whittle's condition on the y_t can be replaced by some condition on the parameters of the underlying process, as we now comment.

Let y_t be a stationary stochastic process. Define

$$\Gamma_i = \Gamma_{-i}' = E y_t y_{t-i}', \qquad (3.11)$$

$$\Gamma^* = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{p-1} \\ \Gamma_1' & \Gamma_0 & \cdots & \Gamma_{p-2} \\ \vdots & \vdots & & \vdots \\ \Gamma_{p-1}' & \Gamma_{p-2}' & \cdots & \Gamma_0 \end{bmatrix}. \qquad (3.12)$$

If a one-sided moving average representation holds for y_t,

$$y_t = \sum_{s=0}^{\infty} H_s \varepsilon_{t-s}, \qquad H_0 = I, \qquad (3.13)$$

where the ε's are independent with mean 0 and covariance V, then

$$\Gamma^* = E \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} [y_t' \; y_{t-1}' \cdots y_{t-p+1}'] \qquad (3.14)$$

$$= \begin{bmatrix} I & H_1 & H_2 & \dots \\ 0 & I & H_1 & \dots \\ \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \dots \end{bmatrix} \begin{bmatrix} V & 0 & \dots \\ 0 & V & \dots \\ \vdots & \vdots & \end{bmatrix} \begin{bmatrix} I & \dots & 0 \\ H_1' & \dots & 0 \\ \vdots & & \vdots \end{bmatrix}.$$

Proposition 3. V *nonsingular implies* Γ^* *nonsingular.*

Proof. The rows of

$$\begin{bmatrix} I & H_1 & \dots \\ 0 & I & \dots \\ \vdots & \vdots & \\ 0 & 0 & \dots \end{bmatrix} \tag{3.15}$$

are linearly independent. Hence, if V is nonsingular, Γ^* is nonsingular. Q.E.D.

4. Some properties of the likelihood function

When the process defined by (2.1) is Gaussian, the likelihood function of the sample can be written as

$$L(B_1,\dots,B_p,V) = (2\pi)^{-\frac{mT}{2}} |\Sigma_T^*|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2} Y_T' \Sigma_T^{*-1} Y_T\}, \tag{4.1}$$

where $Y_T = (y_1', \dots, y_T')'$, $\Sigma_T^* = EY_TY_T'$.

For a recent review and analysis of several proposals to obtain the maximum likelihood estimators, see Anderson (1978) and references therein.

Proposition 4. For V *positive definite,* $T \geq p$, *and* B_j *such that the roots of (2.3) have absolute value less than* 1,

$$Y_T'\Sigma_T^{*-1}Y_T = Y_p'\Sigma_p^{*-1}Y_p + \sum_{t=p+1}^{T} (y_t+B_1y_{t-1} + \dots + B_py_{t-p})' V^{-1}(y_t+B_1y_{t-1} + \dots + B_py_{t-p}), \tag{4.2}$$

$$|\Sigma_T^*||V^{-1}|^T = |\Sigma_p^*||V^{-1}|^p, \tag{4.3}$$

where $Y_p = (y_1', \dots, y_p')'$, $\Sigma_p^* = E\, Y_pY_p'$.

Proof. The joint density of the random variables $y_1,\dots,y_p$ is

$$(2\pi)^{-\frac{pm}{2}} |\Sigma_p^*|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2} Y_p'\Sigma_p^{*-1}Y_p\}, \tag{4.4}$$

and that of $u_{p+1}, \dots, u_T$ is

$$(2\pi)^{-\frac{(T-p)m}{2}} |V|^{-\frac{T-p}{2}} \exp\{-\tfrac{1}{2} \sum_{j=p+1}^{T} u_j'V^{-1}u_j\}. \tag{4.5}$$

Since $y_1,\dots, y_p$ are independent of $u_{p+1}, \dots, u_T$ and the transformation given by (2.1) for $t=p+1, \dots, T$ has Jacobian 1, (4.2) and (4.3) follow by comparing the joint density with the form given by (4.1). Q.E.D.

Proposition 5. For V *positive definite and* $T \geq p$, *the* m×m *submatrix of* Σ_p^{*-1} *in positions* i,j *is given by*

$$\Sigma_p^{*ij} = \sum_{s=0}^{i-1} B_s'V^{-1}B_{s+j-i} - \sum_{s=p+1-j}^{p+i-j} B_s'V^{-1}B_{s+j-i}, \quad i \leq j, \tag{4.6}$$

where $B_0 = I$, *and by the transpose of (4.6) for* $i \geq j$, i,j=1, ..., p. *The* m×m *submatrix of* Σ_T^{*-1} *in positions* i,j *is given by*

$$\begin{aligned} \Sigma_T^{*ij} &= \sum_{s=\max(j,p+1)}^{\min(i+p,T)} B_{s-i}'V^{-1}B_{s-j} + \Sigma_p^{*ij}, && i \leq j \leq p, \\ &= \sum_{s=\max(j,p+1)}^{\min(i+1,T)} B_{s-i}'V^{-1}B_{s-j}, && i \leq j,\ j > p, \\ &= 0, && i+p \leq j \leq T, \end{aligned} \tag{4.7}$$

and by the transpose of (4.7) for $i \geq j$, i,j=1, ..., T.

Proof. The proposition follows by an argument similar to that in Siddiqui (1958), obtained by expanding the sum in (4.2). Siddiqui's argument requires that $T > 2p$, but (4.6) is independent of T and the results hold for all $T \geq p$. Q.E.D.

Anděl (1971) studied (4.7) when $V = I$.

5. Concluding remarks

Vector autoregressive models for time series provide an approach that is attractive empirically. To illustrate this point let us consider a simple example. Let x_t and z_t be two observable time series for which we propose the two-component first-order autoregressive model

$$\begin{bmatrix} x_t \\ z_t \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ z_{t-1} \end{bmatrix} = \begin{bmatrix} v_t \\ w_t \end{bmatrix} , \qquad (5.1)$$

which is a case of (2.9) written in terms of components. A preliminary estimate of B can be obtained from the Yule-Walker equations. We have that

$$C_0 = \sum_{t=1}^{T} y_t y_t' = \sum_{t=1}^{T} \begin{bmatrix} x_t x_t & x_t z_t \\ z_t x_t & z_t z_t \end{bmatrix} = \begin{bmatrix} c_{xx}(0) & c_{xz}(0) \\ c_{zx}(0) & c_{zz}(0) \end{bmatrix} , \qquad (5.2)$$

$$C_1 = \sum_{t=2}^{T} y_t y_{t-1}' = \sum_{t=2}^{T} \begin{bmatrix} x_t x_{t-1} & x_t z_{t-1} \\ z_t x_{t-1} & z_t z_{t-1} \end{bmatrix} = \begin{bmatrix} c_{xx}(1) & c_{xz}(1) \\ c_{zx}(1) & c_{zz}(1) \end{bmatrix} , \qquad (5.3)$$

where $T^{-1}c_{xx}(j)$ estimates the covariance of the x_t series at lag j ($c_{xx}(j) = c_{xx}(-j)$), while $T^{-1}c_{xz}(j)$ estimates the cross covariance between the x_t and z_t series at lag j ($c_{xz}(j) = c_{zx}(-j)$). Then (3.1) becomes $BC_0 = -C_1$, and $B = -C_1 C_0^{-1}$, provided C_0 is nonsingular. In terms of components,

$$\begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} \\ \hat{\beta}_{21} & \hat{\beta}_{22} \end{bmatrix} = -\frac{1}{D} \begin{bmatrix} c_{xx}(1)c_{zz}(0) - c_{xz}(1)c_{zx}(0) & -c_{xx}(1)c_{xz}(0) + c_{xz}(1)c_{xx}(0) \\ c_{zx}(1)c_{zz}(0) - c_{zz}(1)c_{zx}(0) & -c_{zx}(1)c_{xz}(0) + c_{zz}(1)c_{xx}(0) \end{bmatrix} \qquad (5.4)$$

where $D = c_{xx}(0)c_{zz}(0) - (c_{xz}(0))^2$.

If the series x_t and z_t are independent, one expects that $c_{xz}(0)$ and $c_{xz}(1)$ are close to 0. If they were in fact equal to 0, of course

$$\hat{\beta}_{11} = -\frac{c_{xx}(1)}{c_{xx}(0)} = -\frac{\sum_{t=2}^{T} x_t x_{t-1}}{\sum_{t=1}^{T} x_t^2} , \quad \hat{\beta}_{22} = -\frac{c_{zz}(1)}{c_{zz}(0)} = -\frac{\sum_{t=2}^{T} z_t z_{t-1}}{\sum_{t=1}^{T} z_t^2} , \qquad (5.5)$$

$$\hat{\beta}_{12} = \hat{\beta}_{21} = 0.$$

We note that the $\hat{\beta}_{ij}$ in (5.4) are in correspondence with the ordinary least squares estimators that are obtained by considering separately the two scalar equations

$$x_t = \beta_{11}x_{t-1} + \beta_{12}z_{t-1} + v_t \qquad (5.6)$$

and

$$z_t = \beta_{21}x_{t-1} + \beta_{22}z_{t-1} + w_t , \qquad (5.7)$$

while those in (5.5) correspond to considering separately the two scalar autoregressions

$$x_t = \beta_{11}x_{t-1} + h_t , \quad z_t = \beta_{22}z_{t-1} + g_t . \qquad (5.8)$$

This occurs independently of the structure of $V = Eu_t u_t'$, which is due to the nature of the Yule-Walker estimation procedure: The relation between x_t and z_t at different lags enters only through their measures of correlation, and not through the relation between the corresponding error terms.

Equations (5.5) show that the estimators are very similar to those of the ordinary (unweighted) least squares procedure, which will differ from (5.5) in that all sums will be over the range $2 \leq t \leq T$. In general, (3.1) is related to the ordinary least squares procedure in that for the latter the sums in (3.2) range

over the same set of values, namely that corresponding to C_p, $p+1 \leq t \leq T$.

The method of maximum likelihood takes care of the nature of the dependence among the components of the u_t vector in (2.1), which is made clear by the appearance of V in (4.2), (4.3), (4.6), and (4.7).

For the scalar first-order autoregression, Hasza (1980) gave a close form expression for the maximum likelihood estimators of the parameters, say β and $\sigma_u^2 = \text{Var}(u_t)$. The former is to be compared with (5.5).

References

Andĕl, J. (1971), "On the Multiple Autoregressive Series," *The Annals of Mathematical Statistics*, 42, 755-759.

Anderson, T.W. (1971a), *The Statistical Analysis of Time Series*, New York: Wiley.

———— (1971b), "The Stationarity of an Estimated Autoregressive Process," Technical Report No. 7, ONR Contract NO0014-67-A-0112-0030, Department of Statistics, Stanford University.

———— (1978), "Maximum Likelihood Estimation for Vector Autoregressive Moving Average Models," in *Directions in Time Series*, Edited by D. R. Brillinger and G. C. Tiao, Institute of Mathematical Statistics.

————, and R. P. Mentz (1980), "On the Structure of the Likelihood Function of Autoregressive and Moving Average Models," *Journal of Time Series Analysis*, 1, 83-94.

De Jong, P. (1976), "The Recursive Fitting of Autoregressions," *Biometrika*, 63, 525-530.

Hannan, E.J. (1970), *Multiple Time Series*, New York: Wiley.

Hasza, D.P. (1980), "A Note on Maximum Likelihood Estimation for the First-Order Autoregressive Process," *Communications in Statistics*, A9, No. 13, 1411-1415.

Kashyap, R. L., and A. Ramachandra Rao (1976), *Dynamic Stochastic Models from Empirical Data*, New York: Academic.

Quenouille, M.H. (1968), *The Analysis of Multiple Time-Series*, London: Charles Griffin.

Siddiqui, M.M. (1958), "On the Inversion of the Sample Covariance Matrix in a Stationary Autoregressive Process," *Annals of Mathematical Statistics*, 29, 343-357.

Whittle, P. (1963), "On the Fitting of Multivariate Autoregressions, and the Approximate Canonical Factorization of a Spectral Density Matrix," *Biometrika*, 50, 129-134.

A note on the use of the sample autocorrelogram

Raúl Pedro Mentz
Instituto de Investigaciones Estadísticas
Universidad Nacional de Tucumán
Casilla de Correo 209, 4000 Tucumán, Argentina

Summary

Sample autocorrelograms are used frequently in empirical time series analysis, often from the beginning stages of a modelling procedure. Departure from stationarity may render quite inappropriate the use of the sample autocorrelogram, under various possible definitions of the sample quantities. In this paper we study for some of these standard definitions, the departures that can be expected in the presence of linear trends or of a random walk structure of the underlying process.

Key words: Linear trends; random walk; sample autocorrelogram; stationarity; time series.

1. Introduction

The correlation coefficient in the joint distribution of the random variable X and Y is defined by $\rho(X,Y) = \mathcal{E}(X-\mathcal{E}X)(Y-\mathcal{E}Y)/[\mathcal{E}(X-\mathcal{E}X)^2(Y-\mathcal{E}Y)^2]^{1/2}$. Its estimator based on a sample (x_i, y_i), $i = 1,\ldots,n$ is usually taken to be Pearson's correlation coefficient

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\left[\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{j=1}^{n}(y_j-\overline{y})^2\right]^{1/2}} \tag{1.1}$$

In applying the notion of correlation to the analysis of a time series or stochastic process $\{y_t: t = \ldots,-1,0,1,\ldots\}$, the autocorrelation coefficients for

lags $s = \pm 1, \pm 2, \ldots$ are defined by $\rho_t(s) = \mathcal{E}(y_t - \mathcal{E}y_t)(y_{t+s} - \mathcal{E}y_{t+s})/[\mathcal{E}(y_t - \mathcal{E}y_t)^2\mathcal{E}(y_{t+s} - \mathcal{E}y_{t+s})^2]^{1/2}$. Under the assumptions of stationarity or weak stationarity of the process, $\mathcal{E}y_t = \mu$ (a constant) for all t, and the variances do not change with time, so that one is led to define the autocorrelation function at lag s by $\rho(s) = \mathcal{E}(y_t - \mu)(y_{t+s} - \mu)/\mathcal{E}(y_t - \mu)^2$, where now $\rho(s) = \rho(-s)$. The graph of $\rho(s)$ against s is the (theoretical) autocorrelogram of the process.

With a finite record $y_1, \ldots, y_T$ the $\rho(s)$ are usually estimated by the sample autocorrelation coefficients

$$r_1(s) = \frac{\sum_{t=1}^{T-s}(y_t - \overline{y})(y_{t+s} - \overline{y})}{\sum_{t=1}^{T}(y_t - \overline{y})^2}, \qquad s = 1, \ldots, T-1, \tag{1.2}$$

where $\overline{y} = T^{-1}\sum_{t=1}^{T} y_t$, the mean of all available observations. The graph of $r_1(s)$ (or other variants to be considered below) against s is the sample autocorrelogram. It is easy to show that $|r_1(s)| \leq 1$ for each s. In empirical time series analysis less than $T-1$ sample autocorrelations are usually computed and analyzed.

The sums in the numerator and denominator of (1.2) are over different numbers of terms. Several ideas to account for this fact have been considered in the literature, among others the following ones.

a) Let both sums range over $t = 1, \ldots, T-s$, that is, define

$$r_2(s) = \frac{\sum_{t=1}^{T-s}(y_t - \overline{y})(y_{t+s} - \overline{y})}{\sum_{t=1}^{T-s}(y_t - \overline{y})^2}, \qquad s = 1, \ldots, T-1. \tag{1.3}$$

A direct application of Schwarz's inequality shows that $|r_1(s)| \leq 1$, and also $|r_2(s)| \geq |r_1(s)|$ holds.

b) Let both sums range over $t = 1, \ldots, T$ by enlarging the sample record by means of a circular definition: $y_{T+1} = y_1, \ldots, y_{T+s} = y_s$. Then

$$r_3(s) = \frac{\sum_{t=1}^{T}(y_t - \overline{y})(y_{t+s} - \overline{y})}{\sum_{t=1}^{T}(y_t - \overline{y})^2}, \qquad s = 1, \ldots, T-1. \tag{1.4}$$

Again we have that $|r_3(s)| \leq 1$, by Schwarz's inequality.

c) Correct numerator and denominator of (1.2) as follows:

$$r_4(s) = \frac{\frac{1}{T-s}\sum_{t=1}^{T-s}(y_t - \overline{y})(y_{t+s} - \overline{y})}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2}, \qquad s = 1, \ldots, T-1. \tag{1.5}$$

This sample quantity is not bounded by -1 and 1, as is noted, for example, by Kendall and Stuart (1966) or Priestley (1981).

The definition of the sample autocovariances in the numerator and denominator of (1.5) with divisors $T-s$ and T, respectively, is not recommended in the literature, since the desired property of positive definiteness of the sample autocovariance sequence is lost. See, for example, Parzen (1971, Sec. 3.3.2) or Priestley (1981, Sec. 5.3.3).

The statistical properties of the sample autocovariances and autocorrelations introduced above, as well as of those to be discussed below, are usually known through large-sample results. Finite-sample results are known for definition (1.4). Large-sample results are often derived on the assumption of stationarity of the y_t process, and further that the process is linear. A standard assumption is that $y_t = \mu + \sum_{j=-\infty}^{\infty}\gamma_j u_{t-j}$, where the u_t are independent, identically distributed, with zero expected value and finite fourth-order moment, and $\sum|\gamma_j| < \infty$. The class of processes admitting such a representation is considerably large in view of most applications. Under these assumptions the asymptotic (normal) distributions of the various definitions of the sample autocovariances are the same, and hence the corresponding asymptotic (normal) distributions of the sample autocorrelation coefficients defined in terms of them are the same. See, for example, Anderson (1971, Sec. 8.4.2 and 8.4.5.)

2. Linear Trends

From the preceding discussion we see that the sample autocorrelations should be used only in those cases where stationarity is clear. The current

practice differs often from this idea, in that the statistical analysis of a sample record is done even when stationarity is not inherent in the data. For example, in trying to choose a time series model for the observations at hand, analysts will at times process sample records with trends or other departures from stationarity. The purpose of this note is to warn against the problems that may arise when using (1.2) or (1.3) with observations of this type.

As an extreme example, consider "data" $y_t = a + bt$, $t = 1, \ldots, T$, that is, "data" that correspond exactly to a linear trend. It can then be shown that

$$r_1(s) = \frac{T-s}{T}\left(1 - 2s\frac{T+s}{T^2-1}\right), \qquad s = 1, \ldots, T-1. \tag{2.1}$$

$$r_2(s) = 1 - \frac{6s^2}{T^2 - 1 - 2Ts + 4s^2}, \qquad s = 1, \ldots, T-1. \tag{2.2}$$

For this particular case it is clear that the use of the circular definition (1.4) is against intuition.

For $T \geq 50$, Figures 1 and 2 contain the graphs of (2.1) and (2.2), respectively. These results are contrary to expectation, since one wants to have $r_1(s) = r_2(s) = 1$ at all relevant lags. In practice, a series may be generated by $y_t = a + bt + u_t$, where u_t has variance $\sigma^2 > 0$. Then the sample autocorrelogram will have only approximately the shapes in Figure 1 or 2, but clearly the interpretation of the diagram may induce error.

The linear trend is just the simplest case to analyze as is done above. It is important since in practical data analysis data sets often have linear trends, or else are previously transformed so that in the new scale they achieve (near) linearity in the trend; see, for example, Anscombe (1981, Chapter 10), or Tukey (1977).

A practical suggestion is to define the sample autocorrelations so that they are effectively equal to 1 in the presence of a linear trend. Since for lag s we have available the pairs (y_t, y_{t+s}), $t = 1, \ldots, T-s$, it is reasonable to consider

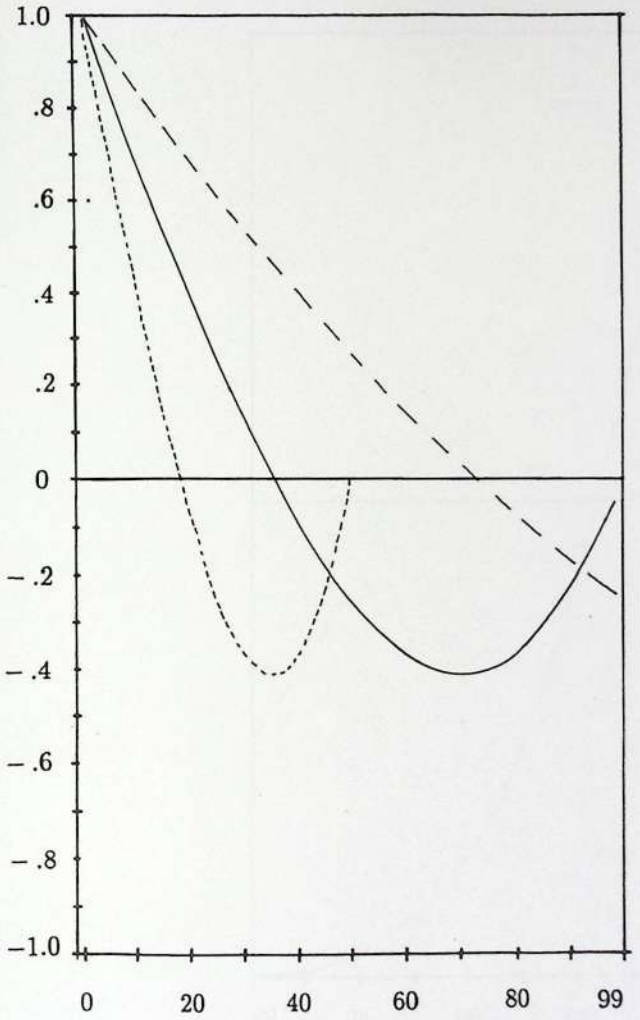


Figure 1

Graphs of $r_1(s)$ for "data" $y_t = a + bt$, for sample sizes $T = 50$ (lags $0 \leq s \leq 49$), $T = 100$ and (lags $0 \leq s \leq 99$). Values are expected to be 1 at all lags. For small lags (say, $1 \leq s \leq 30$), for which the sample autocorrelogram in usually computed, the effect of increasing sample sizes is important. (– – for $T = 200$, — for $T = 100$, - - - for $T = 50$).

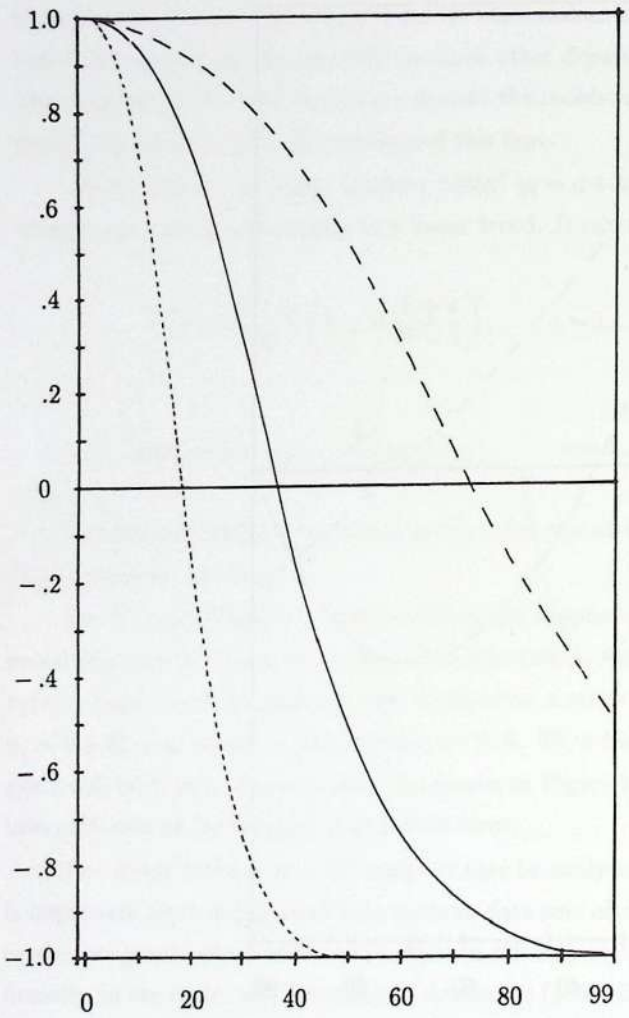


Figure 2

Graphs of $r_2(s)$ computed under the same conditions of Figure 1. One effect of changing the denominator of the sample autocorrelation has been that for lags $1 \leq s \leq 30$ the curves are closer to 1, in particular that for $T = 200$. (- - for $T = 200$, — for $T = 100$, - - - for $T = 50$).

$$r_5(s) = \frac{\sum_{t=1}^{T-s}(y_t - \overline{y}_s)(y_{t+s} - \overline{y}_{s+})}{\left[\sum_{t=1}^{T-s}(y_t - \overline{y}_s)^2 \sum_{u=1}^{T-s}(y_{u+s} - \overline{y}_{s+})^2\right]^{1/2}}, \quad s = 1, \ldots, T-2. \quad (2.3)$$

where $\overline{y}_s = (T-s)^{-1}\sum_{t=1}^{T-s} y_t$, $\overline{y}_{s+} = (T-s)^{-1}\sum_{t=1}^{T-s} y_{t+s}$, averages of the first and last $T-s$ observations, respectively. Such a definition has been considered, for example, by Kendall and Stuart (1966), Jenkins and Watts (1968), Priestley (1981), and Tintner (1952).

The numerator of (2.3) is related to the sample autocovariances

$$C_s = \frac{1}{T-s}\sum_{t=1}^{T-s}(y_t - \overline{y}_s)(y_{t+s} - \overline{y}_{s+}), \qquad s = 0, 1, \ldots, T-2. \quad (2.4)$$

that were studied, for example, by Anderson (1971, Sec. 8.2.1).

Since $r_5(s)$ is Pearson's correlation coefficient for the pairs (y_t, y_{t+s}), $t = 1, \ldots, T-s$, it is clear that $|r_5(s)| \leq 1$. Further, $r_5(s)$ is the maximum likelihood estimator of $\rho(s)$ when y_t and y_{t+s} are assumed to have a joint normal density with $\rho(s)$ as one of its parameters. However, the sequence formed by the $r_5(s)$ loses the property of positive definiteness; see Jenkins and Watts (1968). In the case of $y_t = a + bt$, we have that $r_5(s) = 1$ for all lags.

Whittle (1978) suggests the alternative definition (in our notation)

$$r_6(s) = \frac{\frac{1}{T-s}\sum_{t=1}^{T-s}(y_t - \overline{y}_s)(y_{t-s} - \overline{y}_{s+})}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2}. \quad (2.5)$$

In the case of $y_t = a + bt$, we have that

$$r_6(s) = 1 - s\frac{2T - s}{T^2 - 1}, \quad (2.6)$$

which appears in Figure 3.

$$= b^2\left[\frac{(T-k)(T-k+1)(2T-2k+1)}{6} - (T+1-s)\frac{(t-k)(T-k+1)}{2}\right.$$
$$\left. +\frac{1}{4}(T+1)(T+1-2s)(T-k)\right]$$
$$= \frac{b^2(T-k)}{12}[2(T+1-k)(2T+1-2k)$$
$$-6(T+1-k)(T+1-s)+3(T+1)(T+1-2s)]$$
$$= \frac{b^2(T-k)}{12}(T^2-1-2Tk+4k^2-6ks).$$

When $k = s$ we are left with

$$\sum_{t=1}^{T-s}(y_t-\overline{y})(y_{t+s}-\overline{y}) = \frac{b^2(T-s)}{12}(T^2-1-2Ts-2s^2),$$

and when $k = s = 0$,

$$\sum_{t=1}^{T}(y_t-\overline{y})^2 = \frac{b^2T}{12}(T^2-1),$$

so that

$$r_1(s) = \frac{T-s}{T}\frac{T^2-1-2Ts-2s^2}{T^2-1} = \frac{T-s}{T}\left[1-2s\frac{T+s}{T^2-1}\right].$$

When $s = 0$ we have that

$$\sum_{t=1}^{T-k}(y_t-\overline{y})^2 = \frac{b^2(T-k)}{12}(T^2-1-2Tk+4k^2),$$

so that

$$r_2(s) = \frac{T^2-1-2Ts-2s^2}{T^2-1-2Ts+4s^2} = 1-\frac{6s^2}{T^2-1-2Ts+4s^2},$$

which is (2.2)

Proof of Equation (2.6)

$$\overline{y}_s = \frac{1}{T-s}\sum_{t=1}^{T-s}(a+bt) = \frac{1}{T-s}\left[(T-s)a+b\frac{(T-s)(T-s+1)}{2}\right] = a+b\frac{T-s+1}{2}.$$

$$\overline{y}_{s+} = \frac{1}{T-s}\sum_{t=1}^{T-s}(a+bt+bs) = \frac{1}{T-s}\left[(T-s)(a+bs)+b\frac{(T-s)(T-s+1)}{2}\right]$$
$$= a+bs+b\frac{T-s+1}{2}.$$

$$\sum_{t=1}^{T-s}(y_t-\overline{y}_s)(y_{t+s}-\overline{y}_{s+}) = b^2\sum_{t=1}^{T-s}(b-\frac{T-s+1}{2})^2 = \frac{b^2(t-s)(T-s+1)(T-s-1)}{12}.$$

Estadística española
núm. 116, 1988, págs. 87 a 106

Estimación en los modelos autoregresivos y de promedios móviles

por Raúl Pedro Mentz

Resumen

El propósito de esta presentación es seguir la trayectoria de algunas de las principales ideas aparecidas en la literatura, en el tema de la estimación de los parámetros de los modelos autorregresivos y de promedios móviles, enfatizando un enfoque en el dominio del tiempo.

En cada modelo se define la estructura probabilística y se mencionan algunas propiedades básicas. Luego se consideran sugerencias para estimar los parámetros por el método de los momentos o por mínimos cuadrados. Se consideran después los resultados de investigaciones para demostrar las propiedades asintóticas de estas propuestas y se llega al estudio de métodos de estimación por máxima verosimilitud (exacta o aproximada) en el caso normal.

Se considera brevemente el modelo mixto y algunas propuestas recientes para implementar la estimación máximo verosímil en las computadoras. Se concluye que en los últimos 65 años se han producido desarrollos muy importantes en el área.

Palabras clave: Autorregresivo, promedios móviles, estimación por mínimos cuadrados, estimación por máxima verosimilitud, propiedades asintóticas.

1. Introducción

El propósito de esta presentación es seguir la trayectoria de algunas de las principales ideas aparecidas en la literatura, en el tema de la estimación de los parámetros de los modelos autorregresivos y de promedios móviles, enfatizando un enfoque en el dominio del tiempo.

Se establece el origen de estos modelos en los trabajos de G. U. Yule (inglés, 1871-1951) y E. Slutzky (ruso, 1880-1948) aparecidos entre 1921 y 1937; ver Wold (1954). Antes de estos trabajos el enfoque predominante entre los analistas de datos cronológicos era la búsqueda de las "periodicidades ocultas", con herramientas como el periodograma de Schuster (1898, 1900). En la actualidad decimos que éste es un enfoque no paramétrico, y su versión moderna es el análisis espectral empírico basado en estimadores consistentes de la densidad espectral del proceso estocástico estacionario subyacente.

Un mérito importante de los trabajos de Yule y Slutzky fué insistir en la formulación de modelos estocásticos o estadísticos para el mecanismo generador de los datos. Además los modelos propuestos fueron definidos inicialmente en el dominio del tiempo, lo que sin duda contribuyó a su difusión entre los economistas y otros investigadores sociales.

Incidentalmente, en el momento actual es válido aseverar que las técnicas del análisis de series cronológicas en el dominio del tiempo, tipificadas quizás por el enfoque de Box y Jenkins (1976), y las propias del análisis en el dominio de la frecuencia, interpretadas como consecuencia de una transformada de Fourier, deben pertenecer al acervo de quienes se dedican a las series cronológicas, como investigadores metodológicos o como analistas de datos. A pesar de esta aseveración, nuestro enfoque en esta exposición estará casi exclusivamente basado en el dominio del tiempo.

2. El modelo autoregresivo

2.1. Definición y Propiedades

Un proceso estocástico $\{ z_t \}$ con índice temporal discreto se dice estacionario si las distribuciones conjuntas de probabilidad asociadas con un vector $(z_{t_1},z_{t_2},...,z_{t_k})$ son idénticas a las asociadas con el vector $(z_{t_1+h},z_{t_2+h},...,z_{t_k+h})$ obtenido por una traslación temporal, y ésto para todo conjunto $(t_1,t_2,...,t_k)$ de índices, para todo k y para todo h. Un proceso estacionario tiene todos sus momentos invariantes a cambios en el tiempo. Un proceso se dice "estacionario débil" si sus momentos de primer y segundo orden (esperanzas matemáticas, varianzas, covarianzas) son invariantes a cambios en el tiempo.

Un proceso estocástico estacionario $\{y_t\}$ se dice que es un proceso autorregresivo (o que corresponde a un modelo autorregresivo) de orden finito p, $p \geq 1$, denotado AR(p).

si tiene esperanza matemática constante (digamos, $Ey_t=\mu$ para todo t), varianza constante, positiva y finita, y si para coeficientes $\beta_0=1$, β_1, β_2,...,β_p satisface la ecuación estocástica en diferencias finitas.

$$\sum_{j=0}^{p} \beta_j(y_{t-j}-\mu) = (y_t-\mu) + \beta_1(y_{t-1}-\mu) + ... + \beta_p(y_{t-p}-\mu) = u_t, \quad t = ...,-1,0,1,..., \qquad (2.1.1.)$$

con u_t independiente de $y_{t-1}, y_{t-2}, ...$. Las u_t a su vez forman un proceso estocástico de variables aleatorias independientes, $Eu_t = 0, Eu_t^2 = \sigma^2$, $0 < \sigma^2 < \infty$ (1). Las y_t son variables aleatorias observables, mientras que las u_t son inobservables; $\mu, \beta_1, ..., \beta_p$ y σ^2 son los $p+2$ parámetros del proceso.

El nombre de autorregresión y la propiedad señalada de que los u_t son independientes de los y_t pasados, se vuelven sugestivos frente a la siguiente interpretación: en la regresión lineal múltiple entre y_t y los regresores estocásticos $y_{t-1},...,y_{t-p}$, que escribimos

$$y_t = \mu - \beta_1 (y_{t-1}-\mu) - \beta_2 (y_{t-2}-\mu) - ... - \beta_p (y_{t-p}-\mu) + u_t, \qquad (2.1.2.)$$

los "errores" o "innovaciones" contemporáneos son independientes de los regresores usados y de los pasados, es decir, de $y_{t-1}, y_{t-2}, ...$.

La ecuación polinomial asociada,

$$w^p + \beta_1 w^{p-1} + ... + \beta_p = 0 \qquad (2.1.3.)$$

tiene bajo estas condiciones p raíces $w_1,...,w_p$ menores que 1 en valor absoluto. A su vez, podemos "invertir" la expresión (2.1.1.), es decir, expresar a y_t como función de los u_t contemporáneo y pasados mediante sucesivos reemplazos, hasta obtener

$$y_t = \mu + \sum_{j=0}^{\infty} \gamma_j u_{t-j}, \qquad (2.1.4.)$$

para ciertos coeficientes γ_j determinados unívocamente por los β_j, donde $\gamma_0 = 1$. La representación (2.1.4.) se dice que es un promedio móvil infinito (que denotamos MA(∞), como se verá en la Sección 3); (2.1.1.) y (2.1.4.) resultan equivalentes en media cuadrática (y por lo tanto en probabilidad), en el sentido de que la serie infinita en (2.1.4.) converge en media cuadrática y $\sum_{j=0}^{\infty} \gamma_j^2 < \infty$.

Las autocovarianzas del proceso forman una sucesión, se definen como $\sigma_s = E(y_t-\mu)(y_{t+s}-\mu)$ y son tales que $\sigma_s=\sigma_{-s}$. Multiplicando a (2.1.1.) sucesivamente por $y_t-\mu$, $y_{t-1}-\mu$, ... y tomando esperanzas matemáticas, se obtienen

$$\sigma_0 + \beta_1 \sigma_1 + ... + \beta_p \sigma_p = \sigma^2, \qquad (2.1.5.)$$

(1) Si fuera $Eu_t^2=0$, u_t sería la constante 0 con probabilidad 1 y la relación lineal (2.1.1) indicaría que y_t es "determinístico", lo que queremos descartar. La posibilidad de que el proceso tenga varianza infinita también es descartada, aunque es posible un análisis estadístico de este caso; ver, por ejemplo, Yohai y Maronna (1977).

y

$$\sigma_s + \beta_1 \sigma_{s-1} + ... + \beta_p \sigma_{s-p} = 0, \quad s=1,2,... . \qquad (2.1.6.)$$

Estas pueden resolverse como una ecuación lineal en diferencias finitas (no estocástica), con condiciones de contorno $\sigma_s=\sigma_{-s}$ y condición de escala dada por (2.1.5.).

La transformada de Fourier de la sucesión de autocovarianzas da lugar a la aparición de la densidad espectral

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{p} e^{i\lambda j} \beta_j \right|^2, \quad -\pi \le \lambda \le \pi, \qquad (2.1.7.)$$

que es tal que vale la fórmula de inversión

$$\sigma_s = \int_{-\pi}^{\pi} e^{i\lambda s} f(\lambda)\, d\lambda, \quad s=0, \pm 1, \pm 2, ... \qquad (2.1.8.)$$

2.2. Orígenes

Wold (1954) atribuye a Yule (1927) la introducción del modelo AR(p) para el análisis empírico de las series cronológicas. Yule (1927) utilizó el modelo para el análisis de los datos de Wolfer sobre las manchas solares.

Wold (1954) propuso que el proceso (2.1.1.) fuera reconocido con el nombre de "proceso de Yule", designación que sin embargo no ha tenido aceptación generalizada en la literatura.

En el libro de Wold (1954) se aplica el modelo AR(2) para analizar empíricamente un índice de costo de la vida construido por G. Myrdal, Premio Nobel de Economía.

2.3. Primeras Sugerencias para estimar los parámetros

Con un registro finito $y_1,y_2,...,y_T$ (que denotamos con la misma letra que al proceso teórico) podemos estimar a μ mediante la media muestral

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t, \qquad (2.3.1.)$$

y a los σ_s mediante las autocovarianzas muestrales

$$c_s = \frac{1}{T} \sum_{t=s+1}^{T} (y_t-\bar{y})(y_{t-s}-\bar{y}), \quad s=0,1,...,T-1, \qquad (2.3.2.)$$

de manera que $c_{-s} = c_s$.

Antes de continuar, haremos una breve digresión sobre estos estimadores. Las propiedades estadísticas de estos y otros estimadores, se analizan en general a través de resultados asintóticos, cuando $T \to \infty$, por lo difícil que resultan los cálculos para muestras finitas. Es claro que $E(\bar{y}) = \mu$ y que si μ fuese conocida,

$$E\{\frac{1}{T-s}\sum_{t=s+1}^{T}(y_t-\mu)(y_{t-s}-\mu)\}=\sigma_s,\quad s=0,1,...,T-1.$$

Cuando μ es desconocida, c_s y $c_s T/(T-s) = (T-s)^{-1}\sum_{t=s+1}^{T}(y_t-\bar{y})(y_{t-s}-\bar{y})$ tienen similares propiedades asintóticas, para cualquier s fijo. Por ejemplo, son asintóticamente insesgados. Se prefieren los estimadores (2.3.2.) pues matrices de covarianza muestrales formadas con ellos tienen la propiedad de ser definidas positivas; esto es importante, entre otras cosas, pues las correspondientes matrices de parámetros son también definidas (o por lo menos semi-definidas) positivas.

El uso de $\bar{y}$ en los dos factores $y_t-\bar{y}$ e $y_{t-s}-\bar{y}$ de (2.3.2.) tiene el inconveniente de que si hubiera cambios en el promedio de la serie observada (posiblemente incompatibles con el supuesto de estacionariedad, pero bastante frecuentes en la práctica) la inadecuada estimación del momento de primer orden μ, complicaría la estimación de la estructura de segundo orden: desde hace mucho tiempo se conoce la alternativa de reemplazar a estos factores por $y_t-\bar{y}_s^*$ e $y_{t-s}-\bar{y}_{s+}^*$, respectivamente, donde $\bar{y}_s^* = (T-s)^{-1}\sum_{t=1}^{T-s} y_t$ e $\bar{y}_{s+}^* = (T-s)^{-1}\sum_{t=1}^{T-s} y_{t+s}$, es decir, las medias aritméticas de los datos $y_1,...,y_s$ e $y_{T-s+1},...,y_T$, respectivamente. Para esta y otras alternativas, consultar, por ejemplo, Anderson (1971), Capítulo 8. Bajo el supuesto de estacionariedad, todos estos estimadores tienen las mismas propiedades asintóticas.

Para estimar a los β_j en (2.1.1.), un estimador del tipo de los momentos se obtiene al reemplazar a σ_s por c_s en las primeras p ecuaciones de (2.1.6.). El sistema resultante, escrito aquí en notación matricial,

$$\begin{bmatrix} c_0 & c_1 & \dots & c_{p-1} \\ c_1 & c_0 & \dots & c_{p-2} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ c_{p-1} & c_{p-2} & \dots & c_0 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ . \\ . \\ . \\ \beta_p^* \end{bmatrix} = - \begin{bmatrix} c_1 \\ c_2 \\ . \\ . \\ . \\ c_p \end{bmatrix} \qquad (2.3.3.)$$

se conoce con el nombre de ecuaciones muestrales de Yule y Walker, en honor a Yule y G. Walker. Por supuesto que (2.1.6.) son las ecuaciones teóricas de Yule y Walker.

Se sabe que los estimadores de los β_j provistos por este método son tales que la ecuación polinomial asociada (2.1.3.) con β_j reemplazado por su estimación, tiene todas sus raíces menores que 1 en valor absoluto; Pagano (1973), Anderson (1971a).

Estos estimadores están estrechamente vinculados a los por mínimos cuadrados simples que corresponden al modelo (2.1.1.). En efecto, las ecuaciones normales provenientes de minimizar $\sum_{t=p+1}^{T}\{(y_t-\mu)+\beta_1(y_{t-1}-\mu)+...+\beta_p(y_{t-p}-\mu)\}^2$ con respecto a $\mu, \beta_1,...,\beta_p$ son

$$\sum_{t=p+1}^{T}\{(y_t-\tilde{\mu})+\tilde{\beta}_1(y_{t-1}-\tilde{\mu})+...+\tilde{\beta}_p(y_{t-p}-\tilde{\mu})\}=0\,, \qquad (2.3.4.)$$

y

$$\sum_{t=p+1}^{T}(y_t-\tilde{\mu})(y_{t-s}-\tilde{\mu}) = -\tilde{\beta}_1\sum_{t=p+1}^{T}(y_{t-1}-\tilde{\mu})(y_{t-s}-\tilde{\mu}) - ... - \tilde{\beta}_p\sum_{t=p+1}^{T}(y_{t-p}-\tilde{\mu})(y_{t-s}-\tilde{\mu}),\quad s=1,...,p. \qquad (2.3.5.)$$

Este sistema de $p+1$ ecuaciones en $p+1$ incógnitas es no lineal en los parámetros, problema que puede obviarse mediante una aproximación proveniente del siguiente razonamiento: la ecuación (2.3.4.) conduce a la estimación de μ mediante

$$\tilde{\mu} = \frac{\bar{y}_1+\tilde{\beta}_1\bar{y}_2+...+\tilde{\beta}_p\bar{y}_p}{1+\tilde{\beta}_1+...+\tilde{\beta}_p}\,; \qquad (2.3.6.)$$

si T es grande con respecto a p, podemos reemplazar a $\tilde{\mu}$ por $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$. Reemplazando este valor en el sistema (2.3.5.) llegamos a las ecuaciones

$$\sum_{t=p+1}^{T}(y_t-\bar{y})(y_{t-s}-\bar{y}) = -\hat{\beta}_1\sum_{t=p+1}^{T}(y_{t-1}-\bar{y})(y_{t-s}-\bar{y}) - ... - \hat{\beta}_p\sum_{t=p+1}^{T}(y_{t-p}-\bar{y})(y_{t-s}-\bar{y}),\quad s=1,...,p, \qquad (2.3.7.)$$

que difieren de las ecuaciones muestrales de Yule y Walker (2.3.3.) en que todas las sumas van de $p+1$ a T, en lugar de contener cantidades variables de sumandos como en (2.3.2.).

El reemplazo de $\tilde{\mu}$ por $\bar{y}$, e incluso el del parámetro desconocido μ por uno de sus estimadores, $\tilde{\mu}$ ó $\bar{y}$, no afectará (en general) los desarrollos asintóticos ni las correspondientes propiedades, que son las que se analizan en estos casos; en efecto, tanto $\tilde{\mu}$ como $\bar{y}$ son estimadores consistentes de μ y en el análisis de las distribuciones límites (en general) es válido reemplazar a un estimador consistente por otro, o a un parámetro por uno de sus estimadores consistentes. Esto es particularmente cierto en el caso del Teorema Central del Límite, que es una herramienta típica de análisis asintótico en estos casos.

Los estimadores obtenidos de (2.3.7.) son los que se obtienen de un programa regular de regresión lineal múltiple, cuando se calcula la regresión de y_t en $y_{t-1},...,y_{t-p}$ con $T-p$ conjuntos completos de datos. Sin embargo, las propiedades tradicionales del método de mínimos cuadrados, por ejemplo las que provienen del Teorema de Gauss y Marcov, no son válidas en este caso puesto que los regresores no son fijos sino estocásticos.

Una vez disponibles los estimadores de los β_j podemos formar un estimador de σ^2 usando los "residuos"; por ejemplo, para (2.3.7.) tendremos

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=p+1}^{T}[(y_t-\bar{y})+\hat{\beta}_1(y_{t-1}-\bar{y})+...+\hat{\beta}_p(y_{t-p}-\bar{y})]^2\,. \qquad (2.3.8.)$$

2.4. El Análisis de las Propiedades

Un análisis formal de las propuestas de la sección precedente implica considerar explícitamente a y_t como un proceso estocástico estacionario y a (2.1.1.) como una ecuación estocástica en diferencias finitas. Esto fué formulado a partir del trabajo de Mann y Wald (1943). Ellos comenzaron considerando la posibilidad de estimar a los β_j y σ^2 por máxima verosimilitud cuando el proceso es Gaussiano y además se supone que los valores iniciales $y_{-p+1},...,y_0$ son fijos.

Con este enfoque resultó que las propuestas de la sección precedente son asintóticamente equivalentes a la del método de máxima verosimilitud y que los resultados propios del análisis de un modelo de regresión lineal múltiple en variables explicativas no aleatorias son válidos en el sentido asintótico, cuando $T \rightarrow \infty$ mientras p permanece fijo.

2.5. Estimación por Máxima Verosimilitud

De lo dicho en la sección precedente se desprende que cuando $T \rightarrow \infty$ es razonable usar a (2.3.3.) o (2.3.7.) como equivalentes al método de máxima verosimilitud para estimar a los β_j. Un análisis detallado de estas propuestas y de sus propiedades aparece en Anderson (1971), Capítulo 5. Con la aparición del libro de Box y Jenkins (primera edición de 1970, impresión revisada en 1976) se enfatizó la posibilidad práctica de lograr estimaciones exactas σ aproximadas de los parámetros del modelo (2.1.1.) por el método de máxima verosimilitud, cuando se supone normalidad del proceso y_t.

Deben distinguirse dos casos. En uno se considera que el proceso se inició en y_{-p+1}, tuvo un estado (posiblemente) evolutivo hasta y_0, y de allí en adelante $y_1,...,y_T$ provienen de un proceso estacionario. La distribución (conjunta) condicional de $y_1,...,y_T$ dados $y_{-p+1},...,y_0$ es

$$(2\pi\sigma^2)^{-T/2} \exp\{ -\frac{1}{2\sigma^2} \sum_{t=1}^{T} [(y_t-\mu) + \beta_1 (y_{t-1}-\mu) +...+ \beta_p (y_{t-p}-\mu)]^2 \}. \quad (2.5.1.)$$

Llamaremos a ésta la función de verosimilitud condicional y a los valores que la maximizan estimadores por máxima verosimilitud condicionados, pues dependen de los valores iniciales $y_{-p+1},...,y_0$.

Las ecuaciones provenientes de minimizar a (2.5.1.) con respecto a $\mu, \beta_1,...,\beta_p$ son nuevamente no lineales en los parámetros, como lo eran (2.3.4.) y (2.3.5.). Consideramos entonces la simplificación consistente en estimar previamente a μ por $\bar{y}$, para luego obtener las ecuaciones normales

$$\sum_{t=1}^{T} (y_t-\bar{y})(y_{t-s}-\bar{y}) = -\hat{\beta}_1 \sum_{t=1}^{T} (y_{t-1}-\bar{y})(y_{t-s}-\bar{y}) -...- \hat{\beta}_p \sum_{t=1}^{T} (y_{t-p}-\bar{y})(y_{t-s}-\bar{y}), \quad s=1,...,p. \quad (2.5.2.)$$

Estas son similares a las ecuaciones (2.3.5.), excepto que las sumas van en todos los casos desde 1 hasta T. Si al resolver este sistema hacemos $y_{-p+1} = ... = y_0 = \bar{y}$ obtendremos soluciones idénticas a las de (2.3.7.).

Reemplazando estos estimadores en (2.5.1.) y maximizando la expresión resultante con respecto a σ^2, obtendremos el correspondiente estimador

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} [(y_t-\bar{y}) + \hat{\beta}_1 (y_{t-1}-\bar{y}) +...+ \hat{\beta}_p (y_{t-p}-\bar{y})]^2 . \quad (2.5.3.)$$

Un segundo enfoque consiste en suponer que el proceso estocástico $\{ y_t \}$ que genera los datos es Gaussiano, estacionario y estable, de manera que las observaciones $y_1,...,y_T$ tienen una densidad normal conjunta dada por

$$(2\pi)^{-T/2} |\underset{\sim}{\Sigma}|^{-1/2} \exp\{ -\frac{1}{2} (\underset{\sim}{y}_T-\underset{\sim}{\mu})' \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{y}_T-\underset{\sim}{\mu}) \}, \quad (2.5.4.)$$

donde $\underset{\sim}{y}_T = (y_1,...,y_T)'$ es el vector de observaciones, $\underset{\sim}{\mu} = E\, \underset{\sim}{y}_T = (\mu,...,\mu)'$ el vector de esperanzas matemáticas, y $\underset{\sim}{\Sigma} = E\, (\underset{\sim}{y}_T-\underset{\sim}{\mu})(\underset{\sim}{y}_T-\underset{\sim}{\mu})'$ es la matriz de varianzas y covarianzas. Por comparación con el caso anterior, llamamos a (2.5.4.) como función de los parámetros, la función de verosimilitud incondicional y a los valores que la maximizan estimadores por máxima verosimilitud incondicionados.

Si bien los elementos σ_s de la matriz $\underset{\sim}{\Sigma}$, es decir, las autocovarianzas del proceso que satisfacen (2.1.5.) y (2.1.6.), no tienen formas explícitas fáciles de escribir en el caso general, los componentes de $\underset{\sim}{\Sigma}^{-1}$ fueron dados explícitamente por Siddiqui (1958) en función de los parámetros σ^2 y β_j, mientras que $|\underset{\sim}{\Sigma}|$ está dado explícitamente en Anderson y Mentz (1980) como función de las raíces w_j de (2.1.3.) y de σ^2.

La maximización de (2.5.4.) (independientemente del tratamiento que se de a μ) conduce a problemas no lineales, algunos de los cuales serán considerados en la Sección 4.

Desde el punto de vista de las propiedades estadísticas de los estimadores de los parámetros $\mu,\beta_1,...,\beta_p$ y σ^2, asintóticamente todos los procedimientos sugeridos en esta sección para el método de máxima verosimilitud tienen las mismas propiedades: los estimadores son consistentes y se pueden realizar las inferencias como en el caso de regresores no estocásticos. A su vez, estos resultados asintóticos coinciden con los de los estimadores considerados en la Sección 2.3.

El trabajo Anderson y Mentz (1980) contiene un resumen de las propiedades del modelo AR(p) y un análisis del problema de existencia de los estimadores correspondientes a la función (2.5.4.).

2.6. Ejemplo: El Modelo AR(1)

Para aclarar algunos de los detalles de la presentación, consideraremos el caso más sencillo del modelo AR(1). Para simplificar la escritura consideraremos para la esperanza matemática que, o bien es conocida, en cuyo caso podemos tomarla como μ=0, o si es desconocida estimarla por $\bar{y}$ y en todas las expresiones donde corresponde, reemplazar a y_t por $y_t-\bar{y}$. Esta es a menudo la manera como se opera en la práctica, por ejemplo (internamente) en muchos de los programas de cómputo disponibles.

El modelo autorregresivo de primer orden, AR(1), postula que

$$y_t = -\beta\, y_{t-1} + u_t\,, \qquad t=\ldots,\ -1,0,1,\ldots\,. \tag{2.6.1.}$$

La ecuación polinomial asociada es $w+\beta=0$, de donde se deduce que $|\beta| < 1$ es la condición para que y_t sea estacionario e independiente de $u_{t+1}, u_{t+2}, \ldots$, e invertible en $y_t = \sum_{j=0}^{\infty} (-\beta)^j\, u_{t-j}$. Las funciones de autocovarianzas y de autocorrelaciones son,

$$\sigma_s = (-\beta)^s\, \frac{\sigma^2}{1-\beta^2}\,, \qquad \rho_s = (-\beta)^s\,, \tag{2.6.2.}$$

respectivamente, y la densidad espectral es

$$f(\lambda) = \frac{\sigma^2}{2\pi}\, \frac{1}{1+\beta^2+2\beta\cos\lambda}\,, \qquad -\pi \le \lambda \le \pi\,. \tag{2.6.3.}$$

Para derivar la función de verosimilitud en el caso normal, consideramos a $u_1,\ldots,u_T$ normales independientes, con densidad

$$(\sigma^2 2\pi)^{-T/2}\, exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{T} u_t^2\right]. \tag{2.6.4.}$$

Suponemos que $y_1 = u_1 - \beta y_0$, $y_t + \beta y_{t-1} = u_t$, $t=2,\ldots,T$. En resumen

$$\begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ \beta & 1 & 0 & \ldots & 0 & 0 \\ 0 & \beta & 1 & \ldots & 0 & 0 \\ . & . & . & & . & . \\ . & . & . & & . & . \\ 0 & 0 & 0 & \ldots & \beta & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ y_T \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ . \\ . \\ u_T \end{bmatrix} + \begin{bmatrix} -\beta y_0 \\ 0 \\ 0 \\ . \\ . \\ 0 \end{bmatrix}. \tag{2.6.5.}$$

De esta transformación se desprende que $y_1,\ldots,y_T$ tienen densidad

$$(\sigma^2 2\pi)^{-T/2}\, exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{T} (y_t+\beta y_{t-1})^2\right], \tag{2.6.6.}$$

condicional en y_0. El máximo de (2.6.6.) con respecto a β conduce a

$$\hat{\beta} = -\frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}\,. \tag{2.6.7.}$$

Poniendo $y_0=0$ construímos el estimador por mínimos cuadrados simples,

$$\tilde{\beta} = -\frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = -\frac{\sum_{t=1}^{T-1} y_t y_{t+1}}{\sum_{t=1}^{T-1} y_t^2}\,. \tag{2.6.8.}$$

El criterio de Yule-Walker conduce a estimar a β por

$$\beta^* = -\frac{\sum_{t=1}^{T-1} y_t y_{t+1}}{\sum_{t=1}^{T} y_t^2} = -\frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_t^2}\,, \tag{2.6.9.}$$

que utiliza a todas las observaciones disponibles en el denominador.

En el caso en que suponemos a $y_1,\ldots,y_T$ normal multivariante, puede demostrarse que la función de verosimilitud es

$$(\sigma^2 2\pi)^{-T/2}\sqrt{1-\beta^2}\; exp\left\{-\frac{1}{2\sigma^2}\left[(1-\beta^2)\, y_1^2 + \sum_{t=2}^{T} (y_t+\beta y_{t-1})^2\right]\right\}. \tag{2.6.10.}$$

Los valores de β de σ^2 que hacen máxima esta expresión fueron escritos en forma explícita por Hasza (1980).

3. El modelo de promedios moviles

3.1. Definición y Propiedades

El modelo de promedios móviles de orden $q \ge 1$, denotado MA(q), se define por

$$y_t - \mu = u_t + \alpha_1 u_{t-1} + \ldots + \alpha_q u_{t-q}\,, \tag{3.1.1.}$$

donde los y_t son observables y los u_t son como en el modelo AR(p) ya considerado. Los $q+2$ parámetros del modelo son $\mu, \alpha_1, \ldots, \alpha_q$ y $\sigma^2 = Eu_t^2 (0 < \sigma^2 < \infty)$. Aquí q es finito.

Para $t=\ldots,-1,0,1,\ldots$ (3.1.1.) define un proceso estacionario débil para cualquier valor de los α_j. Si además la ecuación polinomial asociada

$$z^q + \alpha_1 z^{q-1} + \ldots + \alpha_q = 0 \tag{3.1.2.}$$

tiene raíces $z_1,\ldots,z_q$ menores que 1 en valor absoluto, el proceso es invertible,

$$u_t = \sum_{j=0}^{\infty} \delta_j\, (y_{t-j} - \mu)\,, \tag{3.1.3.}$$

donde $\delta_0=1$. Las representaciones (3.1.1.) y (3.1.3.) son equivalentes (en media cuadrática y en probabilidad) y permiten escribir MA(q) = AR(∞), advirtiendo que en (3.1.3.) $\delta_j=\delta_j(\alpha_1,...,\alpha_q)$, funciones de los α_r.

El proceso tiene autocovarianzas

$$\sigma_s=\sigma^2 \sum_{j=0}^{q-|s|} \alpha_j \alpha_{j+|s|}, \qquad s=0,\pm 1,\pm 2,...,\pm q, \tag{3.1.4.}$$

$$=0, \qquad |s|>q.$$

La densidad espectral es

$$f(\lambda)=\frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{q} \alpha_j e^{i\lambda j}\right|^2, \qquad -\pi \le \lambda \le \pi, \tag{3.1.5.}$$

y es tal que vale la fórmula de inversión (2.1.7.).

3.2. Orígenes

Yule (1921, 1926) y Slutzky (1927, 1937) consideraron esquemas de este tipo. Slutzky (1937) consideró combinaciones lineales de términos puramente aleatorios (nuestros u_t) como fuentes de procesos cíclicos.

Wold (1954) propuso que el proceso (3.1.1.) fuera reconocido con el nombre de "proceso de Slutzky", designación que sin embargo no ha tenido aceptación generalizada en la literatura.

En el libro de Wold (1954) se aplican los modelos MA(1), MA(2) y MA(4) a los índices de precios del trigo en Europa Occidental, 1518-1869, recopilados por W. Beveridge.

3.3. Primeras Sugerencias para Estimar los Parámetros

Wold (1954) aparece como el primer autor que estimó concretamente los parámetros de un modelo de promedios móviles con datos observados. Su estimación es por lo que llamamos actualmente el método de los momentos. Por ejemplo, si $q=1$, tenemos que $\sigma_0=\sigma^2(1+\alpha^2)$, $\sigma_1=\sigma^2\alpha$ y por lo tanto el coeficiente de autocorrelación (teórico) de primer orden es $\rho_1=\alpha/(1+\alpha^2)$. Formando el análogo muestral

$$r_1=\frac{\sum_{t=1}^{T-1}(y_t-\bar{y})(y_{t+1}-\bar{y})}{\sum_{t=1}^{T}(y_t-\bar{y})^2}, \tag{3.3.1.}$$

y resolviendo para α la ecuación $r_1=\alpha/(1+\alpha^2)$, obtenemos como estimador admisible (menor que 1 en valor absoluto cuando $|r_1|<\frac{1}{2}$)

$$\alpha^*=\frac{1-\sqrt{1-4r_1^2}}{2r_1}. \tag{3.3.2.}$$

De manera similar se puede proceder para modelos de orden $q>1$.

3.4. El Análisis de las Propiedades

Correspondió a Whittle (1953), un discípulo de Wold, analizar las propiedades asintóticas de los estimadores considerados en la sección precedente. Whittle argumentó que α^* definido en (3.3.2.) es consistente para $\alpha(plim_{T\to\infty}\,\alpha^*=\alpha)$, pero que es asintóticamente ineficiente comparado con el estimador por máxima verosimilitud en el caso normal.

Whittle encontró que la varianza de la distribución normal límite de α^* definido en (3.3.2.) es

$$\frac{1+\alpha^2+4\alpha^4+\alpha^6+\alpha^8}{T(1-\alpha^2)^2}=\frac{1-\alpha^2}{T}+\alpha^2\frac{4+\alpha^2(1+\alpha^2)^2}{T(1-\alpha^2)^2}, \tag{3.4.1.}$$

mientras que para el estimador por máxima verosimilitud la varianza de la distribución normal límite es sólo el primer sumando del miembro de (3.4.1.), es decir,

$$\frac{1-\alpha^2}{T}. \tag{3.4.2.}$$

Los argumentos de Whittle fueron en general de tipo informal, pero luego se proveyeron demostraciones rigurosas de las aseveraciones citadas.

Como conclusión de este análisis surgió la necesidad de mejorar la estimación sugerida por el método de los momentos. Una posibilidad consistía (en el caso $q=1$) en mejorar la estimación de r_1 que entra en (3.3.2.); otra consistía en explotar la relación MA(q)=AR(∞), pues hemos visto que para el modelo AR(p) las ideas intuitivas del tipo de las ecuaciones de Yule-Walker o mínimos cuadrados simples, condujeron a estimadores asintóticamente equivalentes a los por máxima verosimilitud.

3.5. Aproximaciones a la Estimación por Máxima Verosimilitud

Whittle (1951) sugirió un enfoque iterativo para resolver un sistema de ecuaciones que eran asintóticamente equivalentes al método de máxima verosimilitud. Este autor utilizó sus ideas para estimar los parámetros con datos simulados. Esta sugerencia no tuvo un gran impacto práctico hasta la difusión del trabajo de Durbin que comentamos a continuación.

Durbin (1959, 1961) logró mejorar la estimación en el modelo MA(q) explotando la relación MA(q)=AR(∞). Su razonamiento es que si aproximamos al esquema MA(q) por uno AR(k), para un orden k suficientemente grande, y estimamos los parámetros del modelo AR(k) por máxima verosimilitud o por mínimos cuadrados (sin restricciones), esos estimadores pueden combinarse y proporcionar estimadores de los parámetros del modelo MA(q) que "heredarán" entonces (aproximadamente) las propiedades asintóticamente óptimas.

En el caso concreto del modelo AR(1), si $\hat{\beta}_1,...,\hat{\beta}_k$ con los estimadores de los parámetros del modelo AR(k), mostró que

$$\tilde{\alpha} = - \frac{\sum_{i=0}^{k-1} \hat{\beta}_i \hat{\beta}_{i+1}}{\sum_{i=0}^{k-1} \hat{\beta}_i^2} \qquad (3.5.1.)$$

es el estimador deseado, donde $\hat{\beta}_0=1$. Las propiedades de este estimador fueron analizadas por Durbin (1959) y más formalmente por Mentz (1977). Varios autores realizaron comparaciones para muestras finitas en pruebas Monte Carlo.

Nótese que en (3.5.1.) el valor de k debe determinarse explícitamente para una aplicación práctica de la propuesta.

Walker (1961) siguió un enfoque similar, pero basado en la estimación de los coeficientes de autocorrelación. En consecuencia, su propuesta puede considerarse como un mejoramiento de (3.3.2.) basado en un mejoramiento de r_l como estimador de ρ_l (para q=1), o como una formulación alternativa de la propuesta de Durbin (1959). Si $r_1,...,r_k$ son estimadores de las primeras k autocorrelaciones del proceso, definidos por

$$r_s = \frac{\sum_{t=1}^{T-s} (y_t-\bar{y})(y_{t+s}-\bar{y})}{\sum_{t=1}^{T} (y_t-\bar{y})^2} \quad , \quad s = 1,..., k, \qquad (3.5.2.)$$

el estimador de ρ_l propuesto por Walker (1961) es

$$\hat{\rho}_l = r_l + \sum_{j=2}^{k} m_j r_j \, , \qquad (3.5.3.)$$

donde los m_j son también funciones de r_l . Al igual que en la propuesta de Durbin (1959), el valor de k debe establecerse en cada aplicación práctica de la propuesta.

Walker (1961) analizó algunas propiedades de su propuesta y Mentz (1977a) formalizó un análisis cuando T $\to \infty$ y $k=k(T) \to \infty$

Las propuestas de Durbin y de Walker originaron a partir de 1960 un gran caudal de estudios. Se analizaron matemáticamente las propiedades asintóticas y se propusieron modificaciones a los métodos, y se realizaron además comparaciones Monte Carlo entre éstas y otras propuestas.

De estos estudios surgió que la varianza asintótica (3.4.2.) es alcanzada por estas propuestas (y por otras derivadas de ellas), cuando no sólo consideramos que T $\to \infty$, sino que la cantidad k de valores muestrales es función de T, $k=k(T) \to \infty$, pero a una velocidad menor que T; por ejemplo, en algunos casos se requiere que $k(T)^2 / T \to 0$ cuando T $\to \infty$.

3.6. Estimación por Máxima Verosimilitud

El libro de Box y Jenkins (1976) enfatizó para este modelo la posibilidad práctica de lograr estimaciones exactas o aproximadas de los parámetros por el método de máxima verosimilitud bajo el supuesto de normalidad. Un excelente resúmen es el trabajo de Godolphin y de Gooijer (1982).

Nuevamente la función de verosimilitud para el vector $\underset{\sim}{y}_T=(y_1,...,y_T)'$ de observaciones del modelo es

$$(2\pi)^{-T/2} |\underset{\sim}{V}|^{-\frac{1}{2}} exp\left[-\frac{1}{2}(\underset{\sim}{y}_T-\underset{\sim}{v})' \underset{\sim}{V}^{-1} (\underset{\sim}{y}_T-\underset{\sim}{v})\right], \qquad (3.6.1.)$$

donde $\underset{\sim}{v} = E\, \underset{\sim}{y}_T$, $\underset{\sim}{V} = E\, (\underset{\sim}{y}_T-\underset{\sim}{v})(\underset{\sim}{y}_T-\underset{\sim}{v})'$. Aquí se desconoce la forma general de $\underset{\sim}{V}^{-1}$ y de $|\underset{\sim}{V}|$, por lo que las posibilidades de operar con la función de verosimilitud exacta son escasas. En el caso q=1 Godolphin y de Gooijer derivaron un procedimiento iterativo para obtener los estimadores (exactos) de α y σ^2 usando expresiones conocidas de los componentes de $\underset{\sim}{V}^{-1}$, y $|\underset{\sim}{V}| = (1-\alpha^{2T+2}) / (1-\alpha^2)$.

Un resúmen de una gran cantidad de propuestas para estimar el parámetro α de un modelo MA(1) aparece en el trabajo de Mentz y Antelo (1977). En el trabajo Anderson y Mentz (1980) se demuestra que si $\underset{\sim}{y}_T \neq \underset{\sim}{0} = (0,...,0)'$ los estimadores correspondientes a (3.6.1.) existen.

4. El modelo mixto

Las ideas de una autorregresión que liga a las variables aleatorias observables, definida por el miembro de la izquierda de (2.1.1.), y la de una combinación lineal de variables aleatorias inobservables, definida por el miembro de la derecha de (3.1.1.), se pueden combinar para definir un proceso lineal con un número finito de parámetros que generaliza simultáneamente al AR(p) y al MA(q). Se trata del modelo mixto o ARMA(p,q) (autorregresivo con errores promedios móviles, de orden (p,q)), que con las mismas letras definimos por

$$\sum_{j=0}^{p} \beta_j (y_{t-j}-\mu) = \sum_{k=0}^{q} \alpha_k u_{t-k} . \qquad (4.1.)$$

Para la estacionariedad se requiere que los β_j hagan que las raíces de (2.1.3.) sean todas menores que 1 en valor absoluto; en cambio las α_k pueden tomar valores reales sin restricciones. Si las raíces de (2.1.3.) son todas menores que 1 en valor absoluto, (4.1.) se puede invertir para obtener un promedio móvil infinito: simbólicamente, ARMA(p,q)=MA(∞); si las raíces de (3.1.2.) son todas menores que 1 en valor absoluto, (4.1.) se puede invertir para obtener una autorregresión infinita, ARMA(p,q)=AR(∞). En cada caso tendremos que deducir las relaciones que ligan a los coeficientes de las representaciones.

La posibilidad de usar este tipo de modelos en la práctica fué enfatizada por Box y Jenkins en su libro publicado inicialmente en 1970 (1976) y desde entonces atrajo considerable interés teórico y aplicado.

No es fácil en este caso operar con la sucesión de autocovarianzas, definida nuevamente por $E(y_t-\mu)(y_{t+s}-\mu)$. Resulta interesante destacar que la densidad espectral es

$$\frac{\sigma^2}{2\pi} \frac{\left| \sum_{k=0}^{q} \alpha_k e^{i\lambda k} \right|^2}{\left| \sum_{j=0}^{p} \beta_j e^{i\lambda j} \right|^2} , \qquad (4.2.)$$

o sea una función racional que generaliza a (2.1.7.) y (3.1.5.).

Un tratamiento de este nuevo modelo en sus aspectos probabilísticos y estadísticos nos llevaría mas allá del propósito de esta presentación. En esta sección nos limitaremos a dar una descripción concisa de algunos aspectos de trabajos contemporáneos destinados a la estimación de los parámetros de (4.1.) por máxima verosimilitud, exacta o aproximada, en el caso Gaussiano. Como ya lo comentamos antes (Sección 2.5.) existe mucho interés en este tipo de estimación, con preferencia a otros tipos de aproximaciones que se pueden considerar en el caso del modelo mixto.

Aparte de limitar la longitud del trabajo, esta discusión servirá para el siguiente propósito: Está claro que la familia de modelos ARMA(p,q) para $p \geq 0$ y $q \geq 0$ (finitos) incluye a los modelos que tratamos en las secciones 2 y 3, pues AR(p) = ARMA(p,0) y MA(q) = ARMA(0,q). Una de las consecuencias del interés práctico en estos modelos, ha sido el desarrollo de enfoques, algoritmos y programas de cómputo eficaces. Nuestras referencias a estos importantes problemas fueron limitadas y en esta sección presentaremos una discusión, que al considerar el caso de los modelos ARMA, incluye a los "puros" AR y MA ya estudiados.

El problema de obtener estimadores de los parámetros aplicando exactamente el método de máxima verosimilitud es complicado, pues las matrices de covarianzas, o sus

En general, derivar procedimientos iterativos tiende a implicar mas elaboraciones matemáticas con la función de verosimilitud y mas conocimiento de ella, con relación a los enfoques considerados en la sección precedente.

Anderson y Mentz (1987) analizaron en detalle una serie de procedimientos iterativos para el modelo MA(1). Algunos de estos procedimientos se pueden relacionar con uno conocido en la literatura como el "algoritmo EM" (Dempster *et al*, 1977) que fué propuesto inicialmente para resolver problemas de datos faltantes. La conexión la proporciona el hecho de que podemos considerar a los valores iniciales como datos faltantes, en cuyo caso el algoritmo EM sugerirá una sucesión de pasos alternando el cálculo de esperanzas condicionales (E) y maximizaciones (M), lo que conducirá a un procedimiento iterativo.

Por otra parte, varios autores han relacionado estos problemas con el "filtro de Kalman", que es otro procedimiento iterativo que ha encontrado múltiples aplicaciones en el campo de las series cronológicas. Ver, por ejemplo, Harvey (1981).

5. Conclusiones

El progreso alcanzado en el campo de la estimación de los parámetros de los modelos lineales con un número finito de parámetros (AR(p), MA(q) y ARMA(p,q)), desde su difusión en el campo de la estadística hasta el presente, es muy importante. En los 65 años que van desde 1921 hasta el presente, se ha logrado progresar en el conocimiento teórico de las propiedades de los procesos, en el análisis inferencial basado en teoría asintótica y en algunos resultados para muestras finitas, en comparaciones Monte Carlo para tamaños muestrales pequeños, en el desarrollo de algoritmos computacionales y en el uso concreto de los modelos para análisis, predicción y control.

Los primeros trabajos de estimación enfocaron el problema a través del método de los momentos o intentaron la estimación por mínimos cuadrados. El método de máxima verosimilitud fué considerado en forma teórica o a través de aproximaciones, pues los problemas de cómputos eran difíciles o imposibles de solucionar. Al fin de esta etapa los trabajos de Durbin (1959) y Walker (1961) intentaron aproximar la estimación máximo verosímil. A fines de la década de 1960 y comienzos de la siguiente, aparecieron una serie de trabajos importantes, entre otros los de Hannan (1969), Parzen (1971) y Anderson (1975), los que fueron reconsiderados y comparados desde un punto de vista teórico por Anderson (1975a, 1977); la colección de trabajos editada por Brillinger y Tiao (1980) es importante.

La aparición de los libros de Box y Jenkins (1976) y Anderson (1971) a comienzos de la década de 1970 coincidió con un gran incremento en el interés en los temas de series cronológicas en general, y en la estimación de parámetros en particular. Algunos de los desarrollos recientes fueron considerados en nuestra Sección 4. A comienzos de la

década de 1980 debe destacarse la aparición del texto de Priestley (1981), entre otros de lo que es ahora una bibliografía muy extensa en el campo de las series cronológicas.

El área de la inferencia en procesos estocásticos es comparativamente difícil y quedan muchas cosas para investigar y descubrir. La cantidad de especialistas dedicados al tema ha aumentado de unos pocos alrededor de 1970, a una cantidad importante en 1980. Es de confiar que con el transcurso de pocos años más, los problemas importantes reconocidos en la actualidad sean resueltos y se desarrollen nuevas y atrayentes áreas de aplicación.

Bibliografía

ALI, M. M. (1977), Analysis of autoregressive-moving average models: Estimation and prediction, *Biometrika*, 64, 535-45.

ANDERSON, T. W. (1971), *The Statistical Analysis of Time Series*, New York: John Wiley.

ANDERSON, T. W. (1971a), "The Stationarity of an Estimated Autoregressive Process", Technical Report N.º 7, Department of Statistics, Stanford University.

ANDERSON, T. W. (1975), "Estimation by Maximum Likelihood in Autoregressive Moving Average Models in the Time and Frequency Domains", Technical Report N.º 20, Department of Statistics, Stanford University.

ANDERSON, T. W. (1975a), "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure", *Annals of Statistics*, 3, 1283-04.

ANDERSON, T. W. (1977), "Estimation for Autoregressive Moving Average Models in the Time and Frequency Domains", *Annals of Statistics*, 5, 842-865.

ANDERSON, T. W. and R. P. MENTZ (1980), "On the Structure of the Likelihood Function of Autoregressive and Moving Average Models", *Journal of Time Series Analysis*, 1, 83-94.

ANDERSON, T. W. and R. P. MENTZ (1987), Some results on exact maximum likelihood estimation in time series models, *Bulletin of the ISI*, 19-20.

ANSLEY, C. F. (1979), "An Algorithm for the Exact Likelihood of a Mixed Autoregressive Moving Average Process", *Biometrika*, 66, 59-65.

ANSLEY, C. F. and P. NEWBOLD (1980), "Finite Properties of Estimators for Autoregressive Moving Average Models", *Journal of Econometrics*, 13, 159-183.

BMDP (1985), BMDP Statistical Software, Los Angeles, California.

BOX, G. E. P. and G. M. JENKINS (1976), *Time Series Analysis, Forecasting and Control*, Revised Edition, San Francisco: Holden-Day.

BRILLINGER, D. R. and G. C. TIAO (editors) (1980), *Directions in Time Series*, Institute of Mathematical Statistics.

DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN (1977), Maximum Likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, Series B, 39, 1-38.

DENT, W. (1977), Computation of the exact Likelihood function of an ARIMA process, *J. Statist. Comp. and Simul.* 5, 193-206.

DURBIN, J. (1959), "Efficient Estimation of Parameters in Moving Average Models", *Biometrika*, 46, 306-316.

DURBIN, J. (1961), "Efficient Fitting of Linear Models for Continuous Stationary Time Series from Discrete Data", *Bulletin of the International Statistical Institute*, 38, 273-282.

GODOLPHIN, E. J. and J. G. de GOOIJER (1982), "On the Maximum Likelihood Estimation of the Parameters of Gaussian Moving Average process", *Biometrika*, 69, 443-451.

HANNAN, E. J. (1969), "The Estimation of Mixed Moving Average Autoregressive Systems", *Biometrika*, 56, 579-593.

HARVEY, A. C. (1981), *Time Series Models*, Philip Allan Pub. Ltd., Oxford.

HASZA, D. F. (1980), "A Note on Maximum Estimation for the First-Order Autoregressive Process", *Communications in Statistics*, A9, 1411-1415.

HILLMER, S. G. and G. C. TIAO (1979), Likelihood function of stationary multiple autoregressive moving average models, *J. Amer Statist. Assoc.*, 74, 652-61.

IMSL (1979), The IMSL Library, Houston, Texas.

LJUNG, G. M. and G. E. P. BOX (1979), "The Likelihood Function of Stationary Autoregressive Moving Average Models", *Biometrika*, 66, 265-270.

MANN, H. B. and A. WALD (1943), "On the Statistical Treatment of Linear Stochastic Difference Equations", *Econometrica*, 11, 173-220.

MENTZ, R. P. (1977), "Estimation in the First Order Moving Average Model Through the finite Autoregressive Approximation", *J. of Econometrics*, 6, 225-236.

MENTZ, R. P. (1977a), "Estimation in the First Order Moving Average Model Based on Sample Autocorrelations", *Annals of Statistics*, 5, 1250-57.

MENTZ, R. P. and R. ANTELO (1977), "Estimation Procedures for the First-Order Moving Average Model", *Proceedings of the Business and Economic Statistics Section*, Part 1, 329-334, American Statistical Association.

NEWBOLD, P. (1974), The exact likelihood function for a mixed autoregressive-moving average process, *Biometrika*, 61, 423-26.

NICHOLLS, D. F. and A. D. HALL (1979), The exact likelihood function of multivariate autoregressive moving average models, *Biometrika*, 66, 259-64.

OSBORN, D. R. (1977), Exact and approximate maximum likelihood estimators for vector moving average processes, *J. R. Statist. Soc. B*, 39, 114-18.

PAGANO, M. (1973), When is an Autoregressive Scheme Sationary?, *Communications in Statistics*, 1, 533-544.

PARZEN, E. (1971), "Efficient Estimation of Stationary Time Series Mixed Schemes", *Bulletin of the International Statistical Institute*, 44, 315-319.

PHADKE M. S. and G. KEDEM (1978), Computation of the exact likelihood function of multivariate moving average models, *Biometrika*, 65, 511-20.

PRIESTLEY, M. R. (1981), *Spectral Analysis and Time Series*, Vol. 1, London: Academic Press.

SCHUSTER, A. (1898), "On the Investigation of Hidden Periodicities with Application to a Supposed 26 Day Period of Meteorological Phenomena", *Terrestrial Magnetism*, March, 13-41.

SCHUSTER, A. (1900), "The Periodogram of Magnetic Declination as obtained from the record of the Greenwich Observatory during the years 1871-1895", *Transactions of the Cambridge Philosophical Society*, 18, 107-135.

SIDDIQUI, M. (1958), "On the Inversion of the Sample Covariance Matrix in a Stationary Autoregressive Process", *The Annals of Mathematical Statistics*, 29, 585-588.

SLUTZKY, E. (1927), "The Summation of Random Causes as the Source of Cyclic Processes". (Russian with an English summary). *Problems of Economics Conditions*, 3, N.º 1.

SLUTZKY, E. (1937), "The Summation of Random Causes as the Source of Cyclic Processes", *Econometrica*, 5, 105.

WALKER, A. M. (1961), "Large Sample Estimation of Parameters for Moving Average Models", *Biometrika*, 48, 343-357.

WHITTLE, P. (1951), *Hypothesis Testing in Time Series Analysis*, Uppsala: Almqvist and Wicksells.

WHITTLE, P. (1953), "Estimation and Information in Stationary Time Series", *Arkiv für Mathematik*, 2, 423-424.

WOLD, H. (1954), *A Study in the Analysis of Stationary Time Series, 2nd. edition* (first edition of 1938), Stockholm: Almqvist and Wicksells.

YOHAI, V. J. and R. A. MARONNA (1977), Asymptotic behavior of least squares estimates for autoregressive processes with infinite variances, *Annals of Statistics*, Vol 5, 554.

YULE, G. U. (1921), "On the Time-Correlation Problem, with Special Reference to the variate-Difference Correlation Method", *Journal of the Royal Statistical Society*, 84, 497-526.

YULE, G. U. (1926), "Why do we Sometimes get Nonsense-Correlations Between Time-Series? - A Study in Sampling and the Nature of Time-Series", *Journal of the Royal Statistical Society*, 89, 1-64.

YULE, G. U. (1927), "On a Method of Investigating Periodicities in Disturbed Series, with special reference to Wolfer's Sunspot Numbers", *Philosophical Transactions*, A 226, 267-298.

Estimation of autoregressive moving average models

Summary

This work summarizes the present status of the estimation of autoregressive moving average models in the time domain.

For every model studied, the probability distributions are defined and the most important properties are mentioned. Then the suggestions for moment and least square estimation are analyzed as well as maximum likelihood for normal errors.

Finally, some recent ideas for maximum likelihood estimation of mixed models are revised. A conclusion of this survey is that in the last 65 years many important developments have taken place in this area.

Teaching of statistics in Latin America

Raúl P. Mentz
Universidad Nacional de Tucumán, C.C. 209, Tucumán, Argentina

Pedro A. Morettin
Universidad de São Paulo, C.P. 20570, São Paulo, Brasil

Abstract

The paper summarizes information about education and training in statistics in Latin America. Material assembled by others is reviewed, information is provided about undergraduate, master's and doctoral programs in statistics operating in the various countries, and also about training in international centers, publications, meetings in statistics and related fields, and other topics. University programs in statistics started in Latin America around the 1950s, and expanded in the 1960s and 1970s to include many countries. It is concluded that important changes occurred in the statistical field in the Latin American region in recent years.

Key words

Education; undergraduate; graduate; training; publications; meetings.

1. Introduction

This paper summarizes a certain amount of information available about education and training in statistics in the Latin American (LA) region. Education includes the activities of formal programs granting academic degrees at the undergraduate or graduate levels. The educational area also includes problems of teaching statistics as service courses to other disciplines, and at secondary and elementary school levels. By training we mean practical or theoretical instruction directed to impart skills required for specific tasks, as a part of on-the-job training.

With such a wide scope, we do not attempt to cover all topics in equal depth. We pay more attention to university programs, and other aspects related to this level.

The amount of information available on these topics has improved in recent years. In 1978 the CIENES (Interamerican Center for Teaching Statistics) convened in Santiago, Chile, a "LA Seminar on the Teaching of Statistics" (SELENES), in which representatives from 15 countries participated. Country reports from 10 countries were presented together with other papers and a preliminary survey of degree-granting institutions in the region (SELENES, 1978). A second SELENES was held in Caracas, Venezuela, in September of 1989, organized by the School of Statistics, Universidad Central de Venezuela. In 1981 the 43rd. Session of the International Statistical Institute (ISI) was held in Buenos Aires: An invited-papers meeting on "Recent Developments in the Field of Statistics in LA" was held in which four papers with comments and discussion were presented (ISI, 1981b), and a "Round Table on Statistical Education" with six papers and discussion (ISI, 1981a). These and other sources were used by Morettin *et al* (1985). In Loynes (1987) there are reports on statistical education and training in Argentina and Brazil, together with nine other reports covering all parts of the world; the book in an undertaking of the Taskforce on Technical and Tertiary Education in Statistics of ISI's Education Committee.

In spite of these improvements in the quantity of publications dealing with the teaching of statistics in LA, there is still a severe deficit of concrete figures to support some of the thoughts and ideas currently under discussion. As a consequence, there are "empty cells" in many of our statistical records.

2. University level programs in statistics in LA countries

We consider separately the programs working at undergraduate, master's

and doctoral levels. In each category we consider degrees in statistics and in other related disciplines.

2.1. Undergraduate level

Table 1 contains a summary of undergraduate programs granting degrees in statistics. Some observations follow.

TABLE 1

Programs Granting Undergraduate Degrees in Statistics in Latin American Countries

Country	Number of programs	Earliest known years	Average duration at present (ys.)	Program operates within area of: (1)	(2)	(3)	(4)	Largest annual enrollment Year	Number
Argentina	1	1948	5		1			'79-93	199
Bolivia	2	1972	4				2	'78(?)	100
Brazil	18	1953	4	15	2	1		1983	190*
Chile	4	1970	4	1	1		2		
Colombia	3	1958	5		2		1		
Costa Rica	1	1978	4		1			1977	73
Dominican R.	1	1972	4		1			1978	646
Equador	1	1971	4		1			1978	435
Mexico	4	1975	4	1	2		1		
Panama	1	1975	4				1	'78(?)	82
Paraguay	1	1978	4	1					
Peru	6	1966	5	1	3	2		1978	300
Venezuela	1	1953	5		1				

(1) Mathematics, Physical Sciences; (2) Economics, Administration, Social Sciences; (3) Other; (4) Not available.
* Average (estimated) of the 18 programs.

There are 44 undergraduate programs, spread over 13 countries. Brazil has the lead with 18 programs. There are LA countries that have no undergraduate program in statistics. The oldest program is that of the (now) University of Rosario, Argentina, established in 1948. It started as a 3-year program and is now a 5-year program. Most programs in Table 1 last for 4 years.

In Brazil most undergraduate programs in statistics are connected with departments of statistics, or with departments of mathematics in the physical sciences, or engineering. In some of the remaining countries there are large undergraduate programs in statistics connected with administration, economics or other social sciences. Some cases are in Argentina (U. of Rosario, 199 enrolled students), Costa Rica (U. of Costa Rica), Dominican Republic (U. Nacional Autónoma, 646), Equador (U. Central, 435), and Venezuela (U. Central).

In general, undergraduate programs in statistics based on mathematics departments have the potential difficulty that not enough applied statistics may reach the majority of students, in particular if they do not follow further (graduate) studies of statistics. Undergraduate programs in statistics based on economics, administration of similar disciplines have found a basic difficulty in developing sound curricula dealing efficiently with competing needs: mathematics, theoretical statistics, applied statistical methods (including data processing), and a sound introduction to the corresponding social science.

Besides these programs, undergraduate teaching of statistics is done as part of degrees in mathematics, computers and other disciplines, with "emphasis" on or "specialization" in statistics. We have no detailed information about these programs.

2.2. Graduate level: Master's degree

Table 2 contains a summary of graduate programs granting master's degrees (or equivalent) in statistics in LA countries. There are 17 such programs in 6 countries. In Morettin *et al* (1985) Peru was also mentioned, but the program at the U. Nacional de Ingeniería grants a degree in mathematics with statistical emphasis.

There are more programs operating in Brazil and in Mexico than in the other countries, and many countries have no graduate program in statistics. Not all programs in Mexico are equally well implemented (Mendez Ramirez, 1982).

TABLE 2

Programs Granting Graduate (Master's) Degrees in Statistics in Latin American Countries

Country	Number of programs	Earliest known year	Average duration at present (years)
Argentina	2	1973	2
Brazil	6	1964	3
Chile	2		2
Colombia	1		2
Mexico	4	1964	2
Venezuela	2		2

We do not have information about the role of master's degrees in mathematics, computer science or other fields with specialization in statistics.

2.3. Doctoral level

The U. of São Paulo in Brazil started in 1979 a doctoral program in statistics that grants an average of two degrees per year. The possibility of starting doctoral programs in statistics has been considered in Argentina, Chile, Mexico and Venezuela.

Doctoral programs in mathematics also consider specialization in statistics or probability for the thesis work. One example is the U. of Buenos Aires in Argentina, that has granted a few such degrees in recent years.

3. Training

An important part of training is that developed in national statistical offices of the various LA countries. Papers on this topic are included in IASI (1979), and in country reports prepared following a request by the IASI, at least by Chile, Colombia, Panama and the Dominican Republic. Of importance is the training provided by international programs, but this will be considered in the following section.

4. Education and training in international centers

CIENES was established in 1962 in Santiago, Chile, under the aegis of the Organization of American States (OAS). It maintained two training programs of (approximately) one year and half a year, respectively, oriented to the needs of official and socio-economic statistics. In the lapse 1962-1976 they trained 559 and 404 persons belonging to 20 and 21 countries, respectively. It also offered a teaching program at master's level, that initially lasted for two academic years; between 1963 and 1976, 245 persons from 18 LA countries graduated (IASI, 1979).

CELADE (Centro Latinoamericano de Demografía) was established in 1957 in Santiago, Chile, under the aegis of the United Nations Economic Commission for LA (CEPAL). It offers teaching programs in demography at undergraduate and graduate levels. CELADE also conducts research in demography and maintains regular publications in this field.

There have been several training programs of the USA Bureau of the Census to which LA statisticians had access. The Bureau's International Statistical Programs Center started in 1986 the operations of a "School in Applied Statistics and Computer Techniques" (ESAYTEC) that offers programs in Spanish. Two separate curriculum were presented, Computer Data Systems and Population Censuses and Surveys. A total of 42 participants from 17 countries attended part or all of the 7-month programs in the first three years of operations. Plans for the fourth round of ESAYTEC call for presentation of an Economic Statistics for Development curriculum. Activities are held in Washington, D.C..

5. Other topics related to education and training in statistics

One question of importance is how well staffed and equipped the statistical centers conducting teaching programs are in LA. A survey of related interest

was conducted in 1979-80 and presented at the ISI Session in Buenos Aires (ISI, 1981b). The survey emphasized research capabilities of the centers. Answers were received from 29 centers, with a total of 549 members on their teaching and research staffs. Of these, 146 had doctorates and 178 master's degrees. Further, 130 had majored in statistics and 111 in mathematics, computer science, or operations research.

As a complement of this survey the *Current Index to Statistics* for 1975-79 was screened for publications by LA statisticians. A total of 84 publications were found, more than 60% belonging to statisticians from Argentina or Brazil (ISI, 1981b).

Journal based in LA are scarce. Brazil's National Statistical Institute (IBGE) publishes the *Revista Brasileira de Estatística* devoted mainly to applied and official statistics. Mexico's National Institute of Statistics, Geography and Informatics started in 1985 the publication of *Revista de Estadística* devoted to problems in statistics, demography, economics, and related fields. The Chilean Statistical Association has started publishing the journal *Revista de la Sociedad Chilena de Estadística* in 1984, and the Brazilian Statistical Association publishes the *Revista Brasileira de Probabilidade e Estatística* since 1987. The IASI has published *Estadística* for many years, since 1942; there were difficulties and delays in some recent years, but starting with 1989 two issues per year are planned, under a new editorial arrangement that includes a much enlarged editorial board. We already mentioned CELADE's publications in demography.

National statistical meetings occur regularly in some countries, as a key activity of the corresponding statistical societies. One objective is to publish regularly the proceedings of the meetings, which has been accomplished at least in Argentina and Brazil. There are several international groups operating in the regions. The LA committee of the Econometric Society, the Brazilian Region of the Biometric Society, and the LA Committee of the Bernoulli Society organize meetings with certain periodicities, and the meetings are quite successful. In pre

vious years there were also activities organized by the Iberoamerican Committee for Sampling (CIM), by the LA Committee to Promote International Relations in Statistics (COLPRIE), and by the Conference of Government Statisticians of the Americas (CEGA); they are not operating at the present, and the latter has been replaced by the Inter-American Statistical Conference (CIE) and the activities of its Permanent Executive Committee (COM/CIE). The IASI started a program of Seminars in Applied Statistics, of which two already took place (in 1987 in Argentina and in 1989 in Chile), and one is scheduled for 1991 in Mexico; they include the publications of their proceedings.

6. Final comments

An important question refers to the absorption of statistically-trained persons by the labor market. In most LA countries employment opportunities are more often available in the academic and in the official-statistics sectors; see Morettin *et al* (1985). Not many regular teaching programs are oriented to the needs of official statistics, and a great deal of training is provided by in-house programs. Detailed information on these topics is scarce.

The inclusion of statistics in high school curricula is not spread in LA. This may have a negative effect on the future development of the discipline. At this stage in many countries it seems premature to deal in a systematic way with this problem, but perhaps opportunities to do so are not too remote.

As in other parts of the world, LA is in the middle of an "information boom", to which statistics and statisticians have been reacting rather slowly. In some LA countries a decline in university enrollment in both mathematics and statistics has been experienced vis-a-vis computer science.

A large portion of the teaching of statistics is conducted in close association with mathematics and the physical sciences. In some LA countries the identity of statistics as a discipline may be at stake, and an open-minded attitude from the traditional disciplines is called for. This situation may be not unlike that

experienced in other countries in the past, and decisions which are important for the future of statistics are taken by the LA scientific communities of the present.

From tables 1 and 2 it follows that university programs leading to degrees in statistics started in LA around the 1950s, and that they expanded in the 1960s and 1970s to cover many countries. Not all existing programs are comparable in size or in quality, and it is unfortunate that more detailed information about them is not available. The graduate and the undergraduate levels need improvement, and a key question currently under debate is which level to emphasize initially; see the discussion in Morettin *et al* (1985). At any rate, it is clear that important changes have occurred throughout LA in the direction of improving the teaching capabilities in the statistical field.

(Received September 1989)

References

IASI (1979). Documentos presentados en la reunión técnica sobre la evolución de la profesión estadística en el hemisferio occidental. República Dominicana, 1977.

ISI (1981a). Round Table on statistical education in Latin America. Invited papers and discussion. Buenos Aires Session, mimeo.

ISI (1981b). Recent developments in the field of statistics in Latin America. Invited papers, comments, discussion. *Proceedings of the 43rd Session*, Book 1, 309–425.

LOYNES, R.M., ed. (1987). *The Training of Statisticians Round the World*. International Statistical Institute.

MENDEZ RAMIREZ, I. (1982). La enseñanza de la estadística en México. Memoria del V Congreso Nacional sobre Estadísticas Agropecuarias y Forestales, 41–44, México: S.A.R.H.

MORETTIN, P.A., C.A. PERES, S.C. NARULA and R.P. MENTZ (1985). Statistics in Latin America. *The American Statistician*, **39**, 274–278.

SELENES (1978). Informe Final plus invited talks, country reports, preliminary survey of LA institutions, and others. CIENES, mimeo.

The inter-american statistical institute: work program for regional change

Ma. Elena González
US Office of Management and Budget

Raúl P. Mentz
U. of Tucumán, Argentina

Pedro A. Morettin
U. of São Paulo, Brazil

1. Introduction

The Inter-American Statistical Institute (IASI) is an organization of professional statisticians that focuses on statistical activities in the Americas. The work program of IASI has four basic areas of interest: Publications, meetings, the relation between IASI and official statistics in the Americas, and IASI's future assistance to national statistical societies.

2. LASI publications

Estadistica. the official journal of IASI since 1943, is the medium for a technical statistical interchange among the statisticians of the Americas. A Joint Agreement between the Council of the Organization of American States (OAS) and IASI, signed in 1950 and revised in 1955, stipulated that IASI be provided financial support to cover the costs of preparation, publication, and distribution of *Estadistica*. Over the years this support has lessened considerably, affecting the frequency of publication of the journal. In 1981, financial support was halted altogether.

Publication of *Estadistica* experienced two periods of interruption: from 1982 to 1984 and from 1986 to 1987. Issue number 125, December 1981, was the last produced with OAS financing. In 1985, volume 36 (1984), numbers 126

and 127, were published toghether in a single issue under a new editorial policy that had been established in 1979. In 1988, two double volumes were published: volumes 37-38, 1985-86 containing numbers 128 through 131, and volumes 39-40, 1987-88, containing numbers 132 through 135. It is expected that publication of the journal will again become regular in 1989, with semiannual (June and December) issues.

In December 1977, in its fifth period of sessions, the General Assembly of the OAS approved a study of the statistical publications of the General Secretariat, including *Estadistica*. However, the evaluating group decided that the editorial responsability of the journal was the intellectual property of IASI (not the OAS), and hence, it did not in the end evaluate *Estadistica*. Because of this decision, the Executive Committee of IASI decided to establish a Working Group to examine the following issues: (1) Policy matters for *Estadistica*; (2) means to increase the number of readers of the journal; (3) means to ensure the collaboration of Latin-American authors; (4) matters related to the costs of publication, distribution, subscriptions, and so forth.

Chaired by Jorge Arias de Blois of Guatemala, the Working Group met in Washington, D.C., from April 23-27, 1979. Robert Ferber (USA) participated as a member of the Publications Committee of the American Statistical Associatioon, and Walter Duffett (Canada), Raul P. Mentz (Argentina), and Evelio O. Fabbroni (CIENES) were also members of the group. A summary of the conclusions and recommendations of the Committee is found in the document, "Report on the Meeting of the Working Group on the Journal *Estadistica*, Washington, D.C., April 23-27, 1979", published by IASI.

Based on this report, a new editorial policy was established for *Estadistica*, and two editors, one for "Theory and Methods" and the other for "Official Statistics", were appointed. Volumes 36 to 40 were published under the new policy. The authors published in these volumes represented a wide range of countries: Argentina (six), Brazil (six), Canada (four), Chile, India (six), Jordan, Mexico, and United States of America (seven).

The newly elected Executive Committee of IASI met in Aguascalientes, Mexico, in February 1988, and decided to give high priority to *Estadistica*. An annual funding of $10,000 was allocated to cover the costs of publication of the journal. The Committee also decided to appoint a Board of Associate Editors for each section, to help process articles and to maintain contact with potential contributors to the journal. Several internationally known experts in the field have agreed to participate on this editorial board, and it is expected that the journal will receive new impetus in the next few years.

Newsletter. At its April 1979 session, the Working Group recommended the

publication of a newsletter which would have the role of disseminating information on national and international meetings, activities of the national statistical associations, questions related to the practice of the statistical profession, and activities related to the organization and administration of official statistics. Following this recommendation, a bilingual first issue of the *Newsletter* (*Boletin Informativo*) was distributed in August 1987. Issue No. 2 and 3 were then published in June and December 1988, respectively. The first two issues were edited in Mexico by Clara Judisman; the IASI Secretariat took responsibility for issue No. 3. The goal is to publish at least two issues per year. The members of the Executive Committee and the Technical Secretary will try to identify and appoint national correspondents from both the official and academic communities to assist in the preparation of the *Newsletter*.

Proceedings. The *Proceedings* is a compilation of papers presented at IASI seminars. IASI intends to sponsor seminars in different countries annually or biennially, in order to disseminate current topics of interest to the academic and official communities in those countries. The *Proceedings* will be distributed to libraries and other interested institutions and individuals, with the seminar participants receiving them free of charge.

The *Proceedings* will be reviewed by experts in the field, under the direction of a program committee appointed for each seminar. Both solicited and unsolicited papers will be included. The *Proceedings* of the Seminar on Applied Statistics, held in Mar del Plata, Argentina, August 31-September 2, 1987, are now being completed.

Meetings sponsored by IASI

Seminars in applied statistics. At a meeting held in May 1986, IASI's Executive Committee approved the establishment of a Program of Seminars in Applied Statistics, as part of the Institute's permanent work program. Under the seminar program a set of regional statistical meetings will be held periodically in the countries of Latin America and the Caribbean. The meetings will center on topics of current interest to academic and official statisticians of the Americas, in an attempt to bring to their attention standard and modern statistical methods and how they are applied. Each seminar will focus on a specific topic; for example, the first seminar focused on cyclical and seasonal analyses. By attending the seminars, statisticians and other professionals can share their experiences and convey what they have learned to their various institutions.

According to the procedures established by the permanent committee set up to organize the seminars, first, one must obtain an invitation from one or more institutions in a Latin American or Caribbean country. The participation of more than one institution is encouraged, and the participation of a national

statistical association is of particular interest. A topic is then selected, according to the general objetive of the program, and three, four, or five "keynote speakers", chosen because of their expertise in the field, are invited to prepare papers containing selections of topics of current interest. These speakers may belong to institutions located outside the Latin American and Caribbean region. After their selection, statisticians and others both inside and outside the region are invited to contribute papers on the seminar topics. The meeting consists of a 3-day session at which all papers are presented, and roundtable discussions form part of the activities. A Program Committee composed of experts mainly from the region is in charge of the technical aspects of the meeting, as well as the publication of the *Proceedings*, an important part of the project. An Organizing Committee and a Local Arrangements Committee also exist, and function as their names indicate.

Seed funds on the order of $5,000 per seminar are made available to the organizers by IASI. The remaining resources are raised by the host country organizers. Travel expenses for statisticians from the region contributing papers are expected to by financed, at least in part, with funds obtained by the organizers.

The topic of the first seminar held in Mar del Plata, August 31-September 2, 1987, was "Statistical Methods for Cyclical and Seasonal Analyses". A second seminar, on "Statistics in the Improvement of Quality and Productivity", is being organized in Santiago, Chile, for July 1989, jointly by the Catholic University of Chile and the OAS Inter-American Statistical Training Center (CIENES). Two further seminars are in their early stages of planning: one for 1990 in Brazil, with a large share of the organizational effort in the hands of statisticians of the University of Sao Paulo, the other to be held in Mexico City in 1991, for which an invitation was extended by the Instituto Tecnológico Autónomo de México (ITAM).

The first seminar followed an invitation by the National University of Mar de Plata and was held jointly with the annual meeting of the Argentinian Statistical Society. The joint sponshorship proved quite effective: during the three days, the Argentinian Statistical Society met in the mornings and accepted papers on a variety of topics without restricting the subject area, while the seminar took place in the afternoons and evenings and concentrated on the specialized topic of cyclical and seasonal analyses.

Four keynote speakers participated in the seminar: Estela Bee Dagum from Statistics Canada, David F. Findley from the U.S. Bureau of the Census, Geoffrey H. Moore from Columbia University, and David Pierce from the U.S. Federal Reserve Board. Thirty-one papers were contributed and 30 of these were presented at the seminar. Eighteen were chiefly on seasonality, and 12

chiefly on cycles. The authors represented a wide range of countries in the region: Argentina (nine), Brazil (six), Colombia (four), Spain (four), Mexico (three), Chile, Ecuador, Peru, and the United States.

Financial support was obtained by the organizers (in many cases with IASI's collaboration) from the Third World Academy of Sciences, the International Statistical Institute, the American Statistical Association, the institutions of the keynote speakers, and several Argentine academic institutions. The number of registrants to the joint event was 385, with more than 30 coming from outside Argentina.

The first seminar was deemed a considerable success by IASI, because of the large attendance and high quality of the papers presented. In particular, state-of-the-art methodology and applications were discussed in the invited papers.

Collaboration with other sponsors. IASI has a long tradition of participating with other institutions in the organization of statistical meetings. In 1987 alone, IASI contributed to: (1) a workshop on the development of survey capability, organized by the Caribbean community Secretariat and held in Port of Spain, Trinidad and Tobago; (2) a course on household sample surveys, organized by the Central Office of Statistics and Informatics of Venezuela; (3) the Ninth Conference of Commonwealth Caribbean Government Statisticians, held in Kingston, Jamaica; (4) a course on decision theory, organized by the National Institute of Statistics of Argentina; and (5) a workshop for statisticians of developing countries, organized by the International Statistical Institute (ISI) in Tokyo.

In the past, IASI has shared responsibilities in meetings organized by the Inter-American Statistical Conference (CIE) convened by OAS, the U.N. Economic Commission for Latin America and the Caribbean (ECLAC), the Food and Agriculture Organization (FAO) and other United Nations organizations, ISI and its sections, and several academic institutions of the region. For the future, IASI has decided to organize and finance activities chiefly from its own work program, due to restrictions in the availability of resources.

IASI's quinquennial meeting. At the beginning of its operations, IASI attempted to organize statistical congresses periodically. However, the practice was soon abandoned when the OAS statistical office organized and sponsored statistical meetings for the official statistical sector of the American countries. Now, the Executive Committee of IASI has reconsidered the idea of periodic statistical congresses, but along the following clearly spelled-out lines. After the program of seminars in applied statistics is consolidated, IASI could once again organize Inter-American Statistical Congresses, say, every five years, or even at shorter intervals. Such congresses should cover a wide spectrum of top-

ics of interest to the region and thereby attract the attention of statisticians with different backgrounds and interests. The aim is for them to become a permanent and useful activity of IASI.

The organization of these congresses would demand a great deal of effort. Countries would have to be approached to host the congresses and would also have to be willing to contribute a substantial portion of the effort and cost involved in their organization. High-quality academic and professional participation would be demanded, and attractive programs would have to be designed. Funds would have to be provided to the many statisticians from the region to travel to the site of the congress. These travel funds may come from their home institutions or from other sources. Among the benefits of these efforts would be the furtherance of statistical of the region who work in the academic, business, and official sectors.

Educational activities. The educational activities of IASI are aimed at establishing new programs and developing existing programs that are devoted to teaching and training activities in statistics and related fields. These programs operate at various levels and have several directions or orientations.

The field of teaching statistics is expanding globally. One sign of this is the International Conference in the Teaching of Statistics, held every 4 years under the auspices of the ISI. Other indications are the recent introduction of statistics at the primary and secondary school levels, the analysis of particular instructional needs in various statistical fields, the development of adequate teaching material on statistics, and the widespread use of computers in the field. Although Latin America and the Carribbean region have experienced some progress in all of these areas, much remains to be done. It is the role of IASI to help in these endeavors, both through its own action and by fostering international cooperation.

One activity that has been under consideration is the establishment of a program of visiting professors, researchers, and technical personnel in statistics and related fields. Under this program, statisticians from both inside and outside the region will stay in the educational centers of the region for various lengths of time, depending on the purpose of their visit. This purpose may be varied: teaching up-to-date courses and seminars, helping in the research or training programs of the center, discussing organizational or academic matters, contributing to the evaluation of existing programs, and helping in the organization of consulting services. Funding such a program would be a key concern.

Other educational activities that IASI has organized in the past, alone or in conjunction with other agencies, could play an important role in the future

as well. In the CIE meeting mentioned earlier, the area of statistical training has always been of importance. In 1978, the First Latin American Seminar on the Teaching of Statistics (SELENES) organized by CIENES with support from OAS and IASI, took place in Santiago, Chile. Programs like these could serve as the basis for future conferences on statistical education sponsored by IASI.

3. Official statistics

Participation in CIE. The *Inter-American Statistical Conference* (CIE) sponsored by the OAS held its ninth conference in Rio de Janeiro in 1986. The major objective of CIE is to facilitate the continued development of official statistics in the Western Hemisphere. All member countries of OAS are entitled to representation on CIE, as is the Government of Canada. The purpose of CIE is to promote the development of governmental statistical services in the Americas. An important topic considered at the Rio meeting and recommended for further reserach was the planning of the 1990 censuses of population and housing in Latin America and the Caribbean.

IASI has participated in the CIE meetings because they provide an apportunity to exchange information with the principal statisticians of the countries of the Americas, determine the needs of the various countries, and disseminate information to them on IASI's programs, plans, and activities.

The Sixth Permanent Executive Committee of CIE met with the IASI Executive Committee in February 1988 in Mexico. The Conference supported close cooperation between CIE and IASI. The president a IASI expressed IASI's interest in preserving its role as a catalyst in furthering interaction between the different sectors involved in statistical activities (i.e., producers and users of basic statistics, academic statisticians, independent professionals, and so forth). The Seventh meeting of the executive committee took place in New York in February 1989, at the time of the meeting of the UN Statistical Commission. At that meeting, IASI discussed the schedule and agenda of future meetings of statisticians of the Americas.

Participation in ECLAC's meetings of Directors of Statistics. The first meeting of directors of statistics sponsored by the Economic Commission for Latin America and the Caribbean was held in Santiago, Chile, in 1987. (In 1986, the commission and the OAS had agreed to convene such statistical meetings in alternate years). The second meeting of directors is scheduled for September 1989. The technical secretary of IASI will participate.

Specialized activities with international agencies. In 1989, IASI established an Inter-American Household Survey Program to improve this form of data collection. Activities were coordinated with similar ones of the UN Statistical Office, ECLAC, OAS, and other agencies through the creation of the Regional Program for the Development of Household Surveys. As a part of this program, IASI organized a Central American Seminar-Workshop on Household Income and Expenditure Surveys in 1987 in San José, Costa Rica, conducted with the cooperation of ECLAC and the UN Statistical Office. In 1988, IASI distributed two technical documents pertaining to this regional program.

A Working Group on Agricultural Statistics is jointly maintained by IASI and the UN Food and Agriculture Organization (FAO). The last session of the group was held in September 1986 in Santiago, Chile, another is scheduled for 1989 or 1990.

Technical assistance to national statistical offices. IASI often receives requests from statistical offices of various countries of the Americas to provide technical assistance on topics such as planning sample surveys. In the past, some of these requests have been granted. However,IASI no longer has the resources to provide this assistance. Such requests are consequently channelled to the OAS, ECLAC, or countries with official technical assistance programs, such as Spain, Canada, or the United States.

4. Assistance to national statistical societies

National statistical societies are important to the development of statistics as a discipline and a profession. By having regular meetings, a program of publications, and other activities, national statistical associations contribute effectively to the progress of statistics in the Americas. Meetings can be programmed so that not only is there presentation of papers, but also, short courses, conferences, exhibits, roundtable discussions, and other activities are offered. In addition, meetings provide an opportunity for exchanging information with other statisticians. The publications of national statistical societies can include one or more newsletters that keep statisticians in the country up to date regarding developments.

In spite of all these advantages,countries in Latin America and the Caribbean region often find it difficult to establish and maintain national satistical societies. One reason for this is the small number of working statisticians in some

countries. Another reason is the difficulty of attracting academic, business, and government statisticians: some statisticians who might be attracted to the activities of a national statistical society are uninterested because they have limited their own activities to their particular area of expertise.

IASI has been interested in national statistical societies for many years. Its present work program views such societies as key to the development of statistics in the countries of the region, and a strategy is being formulated for IASI to assist in promoting their growth. The Seminars in Applied Statistics described earlier provide an excellent opportunity for IASI to coordinate its activities with those of national statistical societies of the country in which a seminar takes place. As noted, the next seminar will be held in Chile in 1989.

Future quinquennial meetings also provide an opportunity for joint labor. Without active participation of local statistical societies and other institutions, it would be very difficult for IASI to program an attractive and useful congress in a given country. A crucial role must be played by the appropriate national statistical societies.

International cooperation can be used to improve the quality of the regular meetings of the national statistical societies. Funding may become an important issue, but even with limited financial support, IASI could contribute its know-how and membership participation to help national statistical societies in their activities.

On a more grandiose scale, IASI could become a forum for coordination among national statistical societies in the entire Latin American and Caribbean region. When the program of seminars becomes well established, or as soon as one of the proposed quinquennial meetings occurs, a meeting of presidents of statistical societies operating within the various countries of the region could be convened, and activities programmed. This kind of project is actually taking place in some fields, for example, mathematics, economics, and computer science-albeit to a limited extent.

A permanent focus of IASI should be those countries in which no national statistical society is currently in operation. Through its regular members, IASI could look at ways to encourage statisticians to work toward establishing and maintaining stable statisticial societies in such countries. Various statistical activities that take place in the country could be used to promote the establishment of a statistical society. When more than one statistical society exists in a country, IASI could function as a liaison for them.

Evaluation of quadratic forms and traces for iterative estimation in first-order moving average models

R. P. Mentz

Institute of Statistics, University of Tucumán, 4000 Tucumán, Argentina

T. W. Anderson

Department of Statistics, Stanford University, Stanford, C.A. 94305

Key Words and Phrases: moving average model: iterative estimation: quadratic forms: traces: successive elimination: Cholesky decomposition.

Abstract

This paper deals with the evaluation of certain quadratic forms and traces associated with the first-order moving average model. The problem arose while considering the maximum likelihood estimation under normality of the parameters of this model. The quadratic forms are $y'R^{-j}y$, where y is a vector of observations generated by the model, and R is the correlation matrix of the model; the traces are $\mathrm{tr}R^{-j}$; j can be any natural number, but emphasis is placed on small values, $j = 1, 2, 3$.

Procedures in the time and frequency domains are studied, and the amount of computations needed in each case are considered and compared, from which a preferred approach emerges. The computations are compared with several alternative procedures suggested in the literature.

1. Introduction

The first-order Gaussian moving average model, denoted MA(1), with mean zero is defined by

$$y_t = u_t + \alpha u_{t-1}, \qquad t = \ldots, -1, 0, 1, \ldots, \tag{1.1}$$

where the y_t's are observable, the u_t's are unobservable independent $N(0,\sigma^2)$ random variables, $0 < \sigma^2 < \infty$, and α and σ^2 are parameters. This process is stationary for any value of α. The only nonzero covariances (or autocovariances) of the process are $\sigma_0 = \sigma^2(1+\alpha^2)$ and $\sigma_1 = \sigma_{-1} = \sigma^2\alpha$; then $\rho_0 = 1$ and $\rho_1 = \rho_{-1} = \rho = \alpha/(1+\alpha^2)$ are the only nonzero correlations (or autocorrelations) of the process. The process can be parametrized by the pairs (α, σ^2), (σ_0, σ_1), or (σ_0, ρ); for an analysis of the relations among these parametrizations see Anderson and Takemura (1986) and Anderson and Mentz (1990).

If $y_1, \ldots, y_T$ is a sample from (1.1), $y = (y_1, \ldots, y_T)'$ has a multivariate normal distribution with parameters $\mathcal{E}y = 0$, and $\mathcal{E}yy' = \Sigma = (\sigma_{|i-j|})$. We can write the $T \times T$ matrix Σ in the following ways:

$$\begin{aligned} \Sigma &= \sigma_0 I + \sigma_1 G = \sigma^2(1+\alpha^2)I + \sigma^2\alpha G \\ &= \sigma^2\{(1+\alpha^2)I + \alpha G\} \equiv \sigma^2 P \\ &= \sigma^2(1+\alpha^2)\{I + \rho G\} \equiv \sigma_0 R. \end{aligned} \tag{1.2}$$

where G has components $g_{ij} = 1$ for $|i-j| = 1$ and $g_{ij} = 0$ otherwise, and R is the $T \times T$ autocorrelation matrix. Since $\sigma^2 > 0$, Σ is positive definite for $-a < \rho < a$, where $a = 1/\{2\cos[\pi/(T+1)]\}$; see Anderson and Takemura (1986).

The likelihood function can be written as a function of (α, σ^2), or (σ_0, ρ). In the latter case it is

$$L(\sigma_0, \rho) = (2\pi)^{-T/2}(\sigma_0)^{-T/2}|R|^{-1/2}\exp\left\{-\frac{1}{2\sigma_0}y'R^{-1}y\right\}. \tag{1.3}$$

Instead of maximizing (1.3) with respect to σ_0 and ρ simultaneously, we can first maximize (1.3) with respect to σ_0 at $\sigma_0 = (1/T)y'R^{-1}y$, and then maximize with respect to ρ the function

$$n(\rho) = \frac{1}{|R|(y'R^{-1}y)^T}; \tag{1.4}$$

the "concentrated likelihood function" is a constant times the square root of (1.4). We call this the concentrated likelihood procedure or approach.

In this paper we consider the evaluation of quadratic forms

$$q_{jk} = y'R^{-(j+1)}G^k y, \qquad j = -1, 0, 1, \ldots, \; k = 0, 1, \ldots, \tag{1.5}$$

and traces

$$t_{jk} = \mathrm{tr}R^{-j}G^k, \qquad j, k = 0, 1, \ldots, \tag{1.6}$$

that occur in iterative maximum likelihood procedures. When y is a random vector, q_{jk} is a random variable and

$$\begin{aligned} \mathcal{E}q_{jk} &= \mathcal{E}y'R^{-(j+1)}G^k y = \mathcal{E}\mathrm{tr}y'R^{-(j+1)}G^k y = \mathrm{tr}R^{-(j+1)}G^k\mathcal{E}yy' \\ &= \mathrm{tr}R^{-(j+1)}G^k\sigma_0 R = \sigma_0\mathrm{tr}R^{-j}G^k = \sigma_0 t_{jk}; \end{aligned} \tag{1.7}$$

this explains our indexing of q_{jk} and t_{jk}.

To estimate by maximum likelihood the covariances of the moving average part and the coefficients of the autoregressive part of an ARMA(p,q) model, Anderson (1977), Section 4.1, derived the equations that in general correspond to the iterative procedures in the Gaussian case. Using these results Anderson and Mentz (1990), operating in the time domain and using either the scoring or the Newton-Raphson procedures, derived explicit iterative procedures for the MA(1) case, and these involve the quadratic forms and traces introduced in (1.5) and (1.6), respectively. For example, an iterative procedure for estimating ρ based on the likelihood function (1.3) and the scoring procedure is written

$$\left\{\tilde{t}_{22}^{(i-1)}\tilde{q}_{10}^{(i-1)} - \tilde{t}_{21}^{(i-1)}\tilde{q}_{11}^{(i-1)}\right\}\hat{\rho}^{(i)} = \tilde{t}_{20}^{(i-1)}\tilde{q}_{11}^{(i-1)} - \tilde{t}_{21}^{(i-1)}\tilde{q}_{10}^{(i-1)}. \tag{1.8}$$

However, if (1.4) is used with the scoring procedure, an iterative procedure to estimate ρ is written

$$\tilde{t}_{22}^{(i-1)}\tilde{q}_{00}^{(i-1)}\hat{\rho}^{(i)} = T\tilde{q}_{11}^{(i-1)} - \tilde{t}_{21}^{(i-1)}\tilde{q}_{00}^{(i-1)}, \tag{1.9}$$

or, under an alternative approach,

$$\begin{aligned} &\tilde{q}_{00}^{(i-1)}\left\{\tilde{t}_{22}^{(i-1)} - \frac{1}{T}\left[\tilde{t}_{11}^{(i-1)}\right]^2\right\}\hat{\rho}^{(i)} \\ &\quad = T\tilde{q}_{11}^{(i-1)} - \tilde{t}_{11}^{(i-1)}\tilde{q}_{00}^{(i-1)} + \tilde{q}_{00}^{(i-1)}\left\{\tilde{t}_{22}^{(i-1)} - \frac{1}{T}\left[\tilde{t}_{11}^{(i-1)}\right]^2\right\}\hat{\rho}^{(i-1)}. \end{aligned} \tag{1.10}$$

If the likelihood function is used with the Newton-Raphson procedure, an iterative procedure for estimating σ_0 and σ_1 is written

$$\begin{pmatrix} \tilde{q}_{20}^{*(i-1)} - \frac{1}{2}\tilde{t}_{20}^{(i-1)} & \tilde{q}_{21}^{*(i-1)} - \frac{1}{2}\tilde{t}_{21}^{(i-1)} \\ \tilde{q}_{21}^{*(i-1)} - \frac{1}{2}\tilde{t}_{21}^{(i-1)} & \tilde{q}_{22}^{*(i-1)} - \frac{1}{2}\tilde{t}_{22}^{(i-1)} \end{pmatrix}\begin{pmatrix} \hat{\sigma}_0^{(i)} \\ \hat{\sigma}_1^{(i)} \end{pmatrix} = \begin{pmatrix} \frac{3}{2}\tilde{q}_{10}^{*(i-1)} - \tilde{t}_{10}^{(i-1)} \\ \frac{3}{2}\tilde{q}_{11}^{*(i-1)} - \tilde{t}_{11}^{(i-1)} \end{pmatrix}, \tag{1.11}$$

where q_{jk}^* and t_{jk}^* are quadratic forms and traces defined as in (1.5) and (1.6), respectively, with R replaced by Σ. If the concentrated likelihood function is used with the Newton-Raphson approach, then an iterative procedure for estimating ρ is given by

$$\begin{aligned} &\left\{2T\tilde{q}_{22}^{(i-1)}\tilde{q}_{00}^{(i-1)} - \tilde{t}_{22}^{(i-1)}\left[\tilde{q}_{00}^{(i-1)}\right]^2 - T\left[\tilde{q}_{11}^{(i-1)}\right]^2\right\}\hat{\rho}^{(i)} \\ &\quad = T\tilde{q}_{11}^{(i-1)}\tilde{q}_{00}^{(i-1)} - \tilde{t}_{11}^{(i-1)}\left[\tilde{q}_{00}^{(i-1)}\right]^2 \\ &\qquad + \left\{2T\tilde{q}_{22}^{(i-1)}\tilde{q}_{00}^{(i-1)} - \tilde{t}_{22}^{(i-1)}\left[\tilde{q}_{00}^{(i-1)}\right]^2 - T\left[\tilde{q}_{11}^{(i-1)}\right]^2\right\}\hat{\rho}^{(i-1)}. \end{aligned} \tag{1.12}$$

We note that all quadratic forms q_{jk} for $k \le j$ can be expressed as functions of $q_{j0} = y'R^{-(j+1)}y$. In effect, since $R = I + \rho G$, we can substitute $G = \rho^{-1}(R - I)$ in q_{jk}, provided $\rho \neq 0$, to obtain

$$q_{jk} = y'R^{-(j+1)}\rho^{-k}(R-I)^k y = \rho^{-k} y'R^{-(j+1)} \sum_{s=0}^{k} (-1)^{k-s}\binom{k}{s} R^s y \qquad (1.13)$$

$$= \rho^{-k}\sum_{s=0}^{k}(-1)^{k-s}\binom{k}{s} y'R^{-(j-s+1)}y = \rho^{-k}\sum_{s=0}^{k}(-1)^{k-s}\binom{k}{s} q_{j-s,0}, \qquad k \le j.$$

For example,

$$q_{11} = \frac{1}{\rho}(q_{00} - q_{10}), \quad q_{21} = \frac{1}{\rho}(q_{10} - q_{20}), \quad q_{22} = \frac{1}{\rho^2}(q_{00} - 2q_{10} + q_{20}). \qquad (1.14)$$

Similarly, all traces t_{jk}, where $k \le j$, can be expressed in terms of $t_{j0} = \mathrm{tr}R^{-j}$. We have $t_{jk} = \rho^{-k}\sum_{s=0}^{k}(-1)^s\binom{k}{s}t_{j-s,0}$, $k \le j$. For example,

$$t_{11} = \frac{1}{\rho}(T - t_{10}), \quad t_{21} = \frac{1}{\rho}(t_{10} - t_{20}), \quad t_{22} = \frac{1}{\rho^2}(T - 2t_{10} + t_{20}). \qquad (1.15)$$

In the remainder of this paper we study the evaluation of q_{00}, q_{10}, q_{20}, t_{10}, and t_{20}. From the mathematical point of view the problem is to evaluate $y'(I + \rho G)^{-(j+1)}y$, $j = 0, 1, 2$, for a fixed value of ρ and a fixed vector y and $\mathrm{tr}(I + \rho G)^{-j}$, $j = 1, 2$, for a fixed value of ρ.

In Sections 2 and 3 we consider the evaluation of quadratic forms and traces, respectively, in the time domain. For the former the main idea is that the needed computations correspond to solving certain linear systems which can be computed efficiently by the method of successive elimination. These computations can be extended to compute traces (Section 3.1), but for traces we have alternative computational approaches, in particular those leading to rational expressions (Section 3.3). In Section 4 the computationally powerful Fourier approach is considered for computing quadratic forms (Section 4.3) and traces (Section 4.4). In Section 5 the main proposals in the time and frequency domains are compared in terms of the numbers of operations needed to do the calculations. Finally, in Section 6 the computational procedures are compared with several other statistical approaches and methods that appeared in recent literature. Some proofs will be simplified and others omitted: for full details see Anderson and Mentz (1990).

2. Evaluation of quadratic forms in the time domain

2.1. Introduction

Let us define $x = R^{-1}y$ and $u = R^{-1}x$. Then $q_{00} = y'R^{-1}y = y'x$, $q_{10} = y'R^{-2}y = x'x$, and $q_{20} = y'R^{-3}y = x'u$. We see that it suffices to solve for x in the linear system

$$y = Rx. \qquad (2.1)$$

and having done that to solve for u in the linear system $x = Ru$. In the following sections these linear systems will be analyzed from a computational point of view.

2.2. Successive elimination

In this section we consider the method of successive elimination to solve a linear system like (2.1). There is an interesting connection between the "forward solution" of the method of successive elimination and the Cholesky decomposition of the matrix of the linear system. Our presentation will emphasize successive elimination because we want to stress questions related to the numerical computions (for example, those considered in Section 5). We do this in Sections 2.2 and 2.3, and then relate the results we obtain to the Cholesky decomposition in Section 2.4. In effect, our arguments provide a derivation of the Cholesky decomposition of the corresponding matrices.

Let us consider (2.1) in detail: we have to solve it for x, given a vector of observations y and a matrix R evaluated during an iterative procedure. The method of successive elimination corresponds to multiplying the system on the left by the matrix F that is lower triangular with diagonal elements 1 so that in the resulting system

$$FRx = Fy \qquad (2.2)$$

FR is upper triangular. This upper triangular system is called the "forward solution" of the method of successive (or Gaussian) elimination or pivotal condensation.

The product FR corresponds to successive left products by elementary matrices F_j so that

$$F = F_T F_{T-1} \dots F_2 F_1 \qquad (2.3)$$

with $F_1 = I$. Let us define R_j matrices by $R_1 = R = F_1 R$ and

$$R_j = F_j R_{j-1}, \qquad j = 2,\ldots,T. \tag{2.4}$$

Each matrix F_j for $j = 2,\ldots,T$ introduces the changes in the j-th row of R_{j-1} only. Let $F_j = (f_{ij}^{(j)})$, where F_j has a nonzero component $f_{j,j-1}^{(j)}$ in row j and column $j-1$. 1's along its main diagonal, and 0's elsewhere. Suppose that at the j-th step R_{j-1} has $v_1,\ldots,v_{j-1}$ followed by $T-j+1$ 1's on the main diagonal; the diagonal above the main diagonal has $T-1$ ρ's; the diagonal below the main diagonal has $j-2$ 0's followed by $T-j+1$ ρ's, and 0's elsewhere.

The product $F_j R_{j-1}$ is

$$F_j R_{j-1} = \begin{bmatrix} v_1 & \rho & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & v_2 & \rho & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & v_3 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & v_{j-1} & \rho & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & f_{j,j-1}^{(j)} v_{j-1} + \rho & f_{j,j-1}^{(j)}\rho + 1 & \rho & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & \rho \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \rho & 1 \end{bmatrix} \tag{2.5}$$

For (2.5) to have the form of R_{j-1} we require that

$$f_{j,j-1}^{(j)} v_{j-1} + \rho = 0, \qquad f_{j,j-1}^{(j)}\rho + 1 = v_j, \qquad j = 2,\ldots,T. \tag{2.6}$$

Thus

$$v_j = 1 - \frac{\rho^2}{v_{j-1}}, \qquad f_{j,j-1}^{(j)} = -\frac{\rho}{v_{j-1}}, \qquad j = 2,\ldots,T, \tag{2.7}$$

Finally $R_T = FR$ has $v_1,\ldots,v_T$ on its main diagonal, ρ's on the diagonal above the main diagonal and 0's elsewhere, with $v_1 = 1, v_j > 0$, $j = 2,\ldots,T$. Note that the v_j are nonincreasing, $0 < v_j \leq 1$, $j = 1,\ldots,T$, and $v_j > \rho^2$, $j = 1,\ldots,T-1$.

We now use these results to compute the components of the vectors x and u introduced in Section 2.1. Let us define $w = Fy$, the forward solution of the linear system (2.1). In view of (2.3)

$$w = F_T F_{T-1} \ldots F_2 F_1 y, \tag{2.8}$$

so that using expression (2.7) for the $f_{j,j-1}^{(j)}$ component of F_j, the components w_j are calculated in sequence by

$$w_1 = y_1, \qquad w_j = y_j - \frac{\rho w_{j-1}}{v_{j-1}}, \qquad j = 2,\ldots,T. \tag{2.9}$$

With the w_j we compute the x_j, using that $Fy = FRx = R_T x = w$. In terms of components the last equality is

$$v_j x_j + \rho x_{j+1} = w_j, \qquad j = 1,2,\ldots,T-1, \qquad v_T x_T = w_T, \tag{2.10}$$

from which

$$x_T = \frac{w_T}{v_T}, \qquad x_j = \frac{w_j - \rho x_{j+1}}{v_j}, \qquad j = T-1,\ldots,1, \tag{2.11}$$

completing the "backward solution" of (2.1).

Since $x = Ru$ is similar to (2.1) in that they involve the same matrix R, we have that the forward solution of $x = Ru$ is

$$h_1 = x_1, \qquad h_j = x_j - \frac{\rho h_{j-1}}{v_{j-1}}, \qquad j = 2,\ldots,T, \tag{2.12}$$

and the backward solution is

$$u_T = \frac{h_T}{v_T}, \qquad u_j = \frac{h_j - \rho u_{j+1}}{v_j}, \qquad j = T-1,\ldots,1. \tag{2.13}$$

With the y, x and u vectors we compute q_{00}, q_{10} and q_{20} as indicated in Section 2.1, and we also have available $q_{30} = y'R^{-1}y = u'u$.

2.3. Remarks on computation

1. From our analysis in Section 2.2 we concluded that we calculate $q_{00} = y'x$ by using the vector of observations and the backward solution of (2.1). However, we now show that the forward solution of (2.1) is sufficient to compute q_{00}. To show this we first prove that

$$FRF' = V, \tag{2.14}$$

where V is diagonal with diagonal elements $v_1,\ldots,v_T$. In effect, F is lower triangular and has 1's on its main diagonal, $R_T = FR$ is upper triangular and has the v_j's on its main diagonal, F' is also upper triangular, and hence $R_T F'$ is upper triangular. However, $R_T F' = FRF'$ is symmetric, and hence diagonal. The structure of FR and F' imply (2.14).

Using (2.14) and the definition $w = Fy$ we have

$$q_{00} = y'R^{-1}y = y'(F^{-1}VF'^{-1})^{-1}y = (Fy)'V^{-1}(Fy) = w'V^{-1}w$$
$$= \sum_{j=1}^{T} \frac{w_j^2}{v_j}. \tag{2.15}$$

as desired. Similarly, we have $q_{20} = y'R^{-3}y = x'F'V^{-1}Fx = h'V^{-1}h = \sum_{i=1}^{T} h_i^2/v_i$. We conclude that if only q_{00} is needed, it suffices to find the forward solution of (2.1) and use (2.15). If q_{00} and q_{10} are needed, the backward solution x is needed: then $q_{10} = x'x$, and we have two ways to compute: $q_{20} = y'x = w'V^{-1}w$. The argument extends to q_{20}, q_{30}, etc.

2. If only the quadratic forms are required, the calculations can be done as indicated above: we start with: $v_1 = 1$, $w_1 = y_1$, and compute $v_j = 1 - \rho^2/v_{j-1}$, $w_j = y_j - \rho w_{j-1}/v_{j-1}$, $j = 2,\ldots,T$, to obtain the forward solution, and then proceed as indicated above. However, if some traces are also to be calculated, the expressions for the traces that we give in Section 3.1 are more naturally expressed in terms of the $f_{j,j-1}^{(j)}$ components. It will then be more convenient computationally to store these components and to use, for example, expressions like $w_j = y_j + f_{j,j-1}^{(j)} w_{j-1}$. At the end of Section 3.1 we summarize the calculations needed for quadratic forms and traces, and hence use this kind of notation.

3. From the definition of F_j it follows that F_j^{-1} has a nonzero component $-f_{j,j-1}^{(j-1)}$ in row j column $j-1$, 1's along its main diagonal, and 0's elsewhere, so that, in view of (2.3)

$$F^{-1} = F_1^{-1}F_2^{-1}\cdots F_{T-1}^{-1}F_T^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -f_{21}^{(2)} & 1 & 0 & \cdots & 0 & 0 \\ 0 & -f_{32}^{(3)} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -f_{T,T-1}^{(T)} & 1 \end{bmatrix}. \tag{2.16}$$

We can check (2.9) by using $F^{-1}w = y$ and the simple structure of F^{-1} given in (2.16).

4. Anderson (1971a) gave this procedure in detail for the case of (2.2): see also Anderson (1984), Appendix A, Theorem A.1.2.

2.4. Relation to the Cholesky decomposition

Expression (2.14), or its form

$$R = UVU' \tag{2.17}$$

with $U = F^{-1}$, is a case of the Cholesky decomposition, and can also be written as $R = TT'$ with $T = F^{-1}V^{\frac{1}{2}}$, where $V^{\frac{1}{2}}$ is diagonal with diagonal elements $\sqrt{v_1},\ldots,\sqrt{v_T}$. In effect, from general results we know that for all ρ for which R is positive definite (namely, for $-a < \rho < a$), its *unique* Cholesky decomposition exists: when it is written as $R = TT'$, $T = (t_{ij})$ is bidiagonal (because R is tridiagonal), lower triangular, $t_{ii} > 0$, and $t_{ij} = 0$ for $i < j$ and $i > j+1$: in view of the properties of F^{-1} and V, these results follow. When the decomposition is written as in (2.17), $U = (u_{ij})$ is bidiagonal, lower triangular, $u_{ii} = 1$, $u_{ij} = 0$ for $i < j$ and $i > j+1$, and V is diagonal with positive diagonal elements, which is also readily verified.

Expression $R = TT'$ is often called the Cholesky decomposition of R, and T is called the Cholesky triangle. The procedure to obtain T is sometimes called the square root method. Setting $VU' = S$ say, we see that S is bidiagonal, upper triangular and that $R = US$, which is a case of the so-called LR decomposition.

We have then verified that in this case the general results hold: the Cholesky decomposition is equivalent to the forward part of the method of successive elimination. To obtain the Cholesky decomposition of P we use that $P = (1-\rho^2)R$, and hence

$$P = F^{-1}[(1+\rho^2)V](F^{-1})'; \tag{2.18}$$

the only change in relation to (2.17) is that the diagonal elements of the diagonal matrix in the decomposition are now $(1+\rho^2)v_j$ for $j = 1,\ldots,T$.

Useful references for these topics, and also for the material in the preceding sections, are Anderson (1984), in particular Appendix A, Golub and Van Loan (1983), and Graybill (1983).

2.5. Expressions in terms of determinants

The results so far were presented in computational forms. Interesting analytical expressions (not necessarily useful for computations) can be given in terms of the R (and P) matrices. Let Δ_s denote the determinant of R of order $s \times s$. Then

$$\Delta_s = \frac{1}{2^s}\sum_{j=0}^{[s/2]} \binom{s+1}{2j+1}(1-4\rho^2)^j, \qquad j = 0,1,\ldots, \tag{2.19}$$

so that, for example

$$\Delta_0 = \Delta_1 = 1, \quad \Delta_2 = 1-\rho^2, \quad \Delta_3 = 1-2\rho^2, \quad \Delta_4 = 1-3\rho^2+\rho^4. \tag{2.20}$$

These determinants are known to satisfy the difference equation $\Delta_s = \Delta_{s-1} - \rho^2\Delta_{s-2}$; see, for example, Shaman (1969).

To translate the expressions given in Sections 2.2 and 2.3 in terms of the v_i and $f^{(j)}_{j,j-1}$ into expressions invoiving Δ_s, it suffices to show that $v_s = \Delta_s/\Delta_{s-1}, s = 1,\ldots,T$, which can be proved easily by induction.

A similar analysis can be done for the determinants $\tilde{\Delta}_s$ of the P matrices, which are given by

$$\tilde{\Delta}_s = \frac{1-\alpha^{2(s+1)}}{1-\alpha^2}, \qquad s = 0,1,\ldots, \tag{2.21}$$

and satisfy the difference equation $\tilde{\Delta}_s = (1+\alpha^2)\tilde{\Delta}_{s-1} - \alpha^2\tilde{\Delta}_{s-2}$. In fact these results are readily deduced from those for the R matrices, because $\tilde{\Delta}_s = (1+\alpha^2)^s\Delta_s$.

3. Evaluation of traces in the time domain

3.1. Using the solutions of linear systems

The calculations in Section 2.2 were presented as part of the computations needed for quadratic forms, but can also be used for traces.

From (2.17) we have that $R = UVU'$, so that the "forward" solution (2.2) $FRx = Fy = w$ is $VU'x = U^{-1}y = w$, and simultaneously $U'x = V^{-1}U^{-1}y = V^{-1}w$. Note that $F = U^{-1}$.

To compute $t_{10} = \mathrm{tr}R^{-1}$ we set $RX = I$, where X and I are of orders $T\times T$. In successive elimination the forward solution is $FRX = F$ or $VU'X = U^{-1}$. We get U^{-1} and (diagonal) V by recording the steps of the forward solution of $Rx = y$. Then

$$t_{10} = \mathrm{tr}R^{-1} = \mathrm{tr}X = \mathrm{tr}(U')^{-1}V^{-1}U^{-1} = \mathrm{tr}F'V^{-1}F \tag{3.1}$$
$$= \mathrm{tr}V^{-1}FF' = 1 + \sum_{s=2}^{T}\frac{1+f_{s1}^2+f_{s2}^2+\cdots+f_{s,s-1}^2}{v_s},$$

where $F = (f_{ij})$. Using (2.3) and the structure of the F_j given in Section 2 we obtain

$$f_{ij} = f^{(i)}_{i,i-1}f^{(i-1)}_{i-1,i-2}\cdots f^{(j+1)}_{j+1,j}, \quad j = 1,\ldots,T-1;\ i = j+1,\ldots,T. \tag{3.2}$$

In more detail, the elements of F below its main diagonal are

$$\begin{aligned}
&f_{21} = f^{(2)}_{21},\\
&f_{31} = f^{(3)}_{32}f^{(2)}_{21}, && f_{32} = f^{(3)}_{32},\\
&f_{41} = f^{(4)}_{43}f^{(3)}_{32}f^{(2)}_{21}, && f_{42} = f^{(4)}_{43}f^{(3)}_{32}, && f_{43} = f^{(4)}_{43},\\
&\vdots && \vdots\\
&f_{T1} = f^{(T)}_{T,T-1}\cdots f^{(2)}_{21}, && f_{T2} = f^{(T)}_{T,T-1}\cdots f^{(3)}_{32}, \ldots, && f_{T,T-1} = f^{(T)}_{T,T-1}.
\end{aligned} \tag{3.3}$$

We next compute t_{20} as

$$t_{20} = \mathrm{tr}\,R^{-2} = \mathrm{tr}(F'V^{-1}F)(F'V^{-1}F) = \mathrm{tr}\,FF'V^{-1}FF'V^{-1} \tag{3.4}$$
$$= \mathrm{tr}(V^{-\frac{1}{2}}FF'V^{-\frac{1}{2}})(V^{-\frac{1}{2}}FF'V^{-\frac{1}{2}})' = \mathrm{tr}\,HH' = \sum_{i=1}^{T}\sum_{j=1}^{T}h_{ij}^2,$$

where $H = V^{-\frac{1}{2}}FF'V^{-\frac{1}{2}}$ is symmetric, and we have used the circular property of the trace. The components of FF' are

$$\sum_{s=1}^{T}f_{is}f_{js} = \sum_{s=1}^{\min(i,j)}f_{is}f_{js}, \quad i,j = 1,\ldots,T, \tag{3.5}$$

so that the components of H are

$$h_{ij} = \frac{\sum_{s=1}^{\min(i,j)}f_{is}f_{js}}{\sqrt{v_i}\sqrt{v_j}} = \sum_{s=1}^{\min(i,j)}\frac{f_{is}}{\sqrt{v_i}}\frac{f_{js}}{\sqrt{v_j}}, \quad i,j = 1,\ldots,T, \tag{3.6}$$

and hence

$$\begin{aligned}
t_{20} &= \sum_{i=1}^{T}\sum_{j=1}^{T}\left[\sum_{s=1}^{\min(i,j)}\frac{f_{is}}{\sqrt{v_i}}\frac{f_{js}}{\sqrt{v_j}}\right]^2\\
&= \sum_{i=1}^{T}\left[\sum_{s=1}^{i}\frac{f_{is}^2}{v_i}\right]^2 + 2\sum_{\substack{i=1\\ i<j}}^{T}\sum_{j=1}^{T}\left[\sum_{s=1}^{i}\frac{f_{is}}{\sqrt{v_i}}\frac{f_{js}}{\sqrt{v_j}}\right]^2\\
&= \sum_{i=1}^{T}\frac{1}{v_i^2}\left[\sum_{s=1}^{i}f_{is}^2\right]^2 + 2\sum_{\substack{i=1\\ i<j}}^{T}\sum_{j=1}^{T}\frac{1}{v_iv_j}\left[\sum_{s=1}^{i}f_{is}f_{js}\right]^2.
\end{aligned} \tag{3.7}$$

These computations for a given value of ρ can be added to those presented in Section 2.2 to compute q_{00} and q_{10}. We now present the computations needed for q_{00}, q_{10}, q_{20}, t_{10}, and t_{20}, followed by some comments to facilitate the interpretation. Define $\Phi = FF'$.

Starting values ($s = 1$)

$$v_1 = 1, \quad f_{jj} = 1, \quad j = 1,\ldots,T, \quad t_{10}^{(1)} = 1, \quad t_{20}^{(1)} = 1. \tag{3.8}$$

$$w_1 = y_1. \tag{3.9}$$

Step s, $s = 2,\ldots,T$

$$v_s = \frac{v_{s-1} - \rho^2}{v_{s-1}}, \quad f_{s,s-1}^{(s)} = -\frac{\rho}{v_{s-1}}. \tag{3.10}$$

$$f_{sj} = f_{s,s-1}^{(s)} f_{s-1,j}, \qquad j = 1,\ldots,s-1. \tag{3.11}$$

$$o_{sj} = \sum_{k=1}^{j} f_{sk} f_{jk}, \qquad j = 1,\ldots,s, \tag{3.12}$$

$$t_{10}^{(s)} = t_{10}^{(s-1)} + \frac{o_{ss}}{v_s}, \quad t_{20}^{(s)} = t_{20}^{(s-1)} + \frac{o_{ss}^2}{v_s^2} + 2\left[\frac{o_{s,s-1}^2}{v_s v_{s-1}} + \cdots + \frac{o_{s1}^2}{v_s v_1}\right], \tag{3.13}$$

$$w_s = y_s + f_{s,s-1}^{(s)} w_{s-1}. \tag{3.14}$$

After compieting these computations we have

$$t_{10} = t_{10}^{(T)}, \qquad t_{20} = t_{20}^{(T)}, \tag{3.15}$$

$$x_T = \frac{w_T}{v_T} \qquad x_j = \frac{w_j}{v_j} + f_{j+1,j}^{(j+1)} x_{j+1}, \qquad j = T-1,\ldots,1. \tag{3.16}$$

$$q_{00} = \sum_{i=1}^{T} y_i z_i, \qquad q_{10} = \sum_{i=1}^{T} x_i^2. \tag{3.17}$$

Finally, to compute q_{20} we have

$$h_1 = x_1, \qquad h_s = x_s + f_{s,s-1}^{(s)} h_{s-1}, \qquad s = 2,\ldots,T, \tag{3.18}$$

$$q_{20} = \sum_{s=1}^{T} \frac{h_s^2}{v_s}. \tag{3.19}$$

Note the following points. 1. Formulas (3.10), (3.14), and (3.16) were already given in Section 2.2 to compute q_{00} and q_{10} in (3.17). 2. Similar computations in (3.18) produce q_{20} in (3.19). This was discussed in Sections 2.2 and 2.3. 3. Hence, the superscripts in $f_{s,s-1}^{(s)}$, and the indices in v_s and w_s, correspond to the calculations being done in sequence. 4. To calculate t_{10} and t_{20} we need to compute the components of F and $\Phi = FF'$. One row of F is computed at each step, namely, for row s, the elements f_{sj} for $j = 1,\ldots,s-1$. Further $f_{ss} = 1$ and $f_{sj} = 0$ for $s < j$. The calculations in (3.11) correspond to the structure of f_{ij} given in (3.2) or (3.3). 5. The components of $\Phi = FF'$ are scalar products of the rows of F, and are calculated in (3.12), where the sums can also reach $k = T$ in each case, because for $k > j$ ($j \le s$) at least one of the factors f_{sk} is 0. 6. In $t_{10}^{(s)}$ and $t_{20}^{(s)}$ the superscripts denote partial sums, so that $t_{10}^{(T)} = t_{10}$ and $t_{20}^{(T)} = t_{20}$, which is (3.15).

3.2. Series expressions

For $|\rho| < 1/2$ we have

$$t_{10} = \operatorname{tr} R^{-1} = \operatorname{tr}(I - G)^{-1} = \operatorname{tr} \sum_{j=0}^{\infty} (-\rho)^j G^j = \sum_{k=0}^{\infty} \rho^{2k} \operatorname{tr} G^{2k}, \tag{3.20}$$

since $\operatorname{tr} G^j = 0$ for j odd. Note that (3.20) converges because the characteristic roots of G are less than 2 in absolute value and $|\rho| \le \frac{1}{2}$. Similarly,

$$t_{11} = \operatorname{tr} R^{-1} = \operatorname{tr}(I + \rho G)^{-1} = \operatorname{tr} \sum_{j=1}^{\infty} j(-\rho)^{j-1} G^{j-1} = \sum_{k=0}^{\infty} (2k+1)\rho^{2k} \operatorname{tr} G^{2k}. \tag{3.21}$$

In an appendix we prove that

$$t_{20} = \frac{T+1}{\sqrt{1-4\rho^2}} - \frac{1}{1-4\rho^2} - 2(T+1) \sum_{j=1}^{\infty} \sum_{k=g(T+1)}^{\infty} \binom{2k}{k-g(T+1)} \rho^{2k}. \tag{3.22}$$

See also Section 4.4.

3.3. Rational expressions

To compute t_{11} and t_{12} we consider an expression for $\log|R|$ and use

$$\frac{d}{d\rho} \log|R| = \operatorname{tr} R^{-1} G = t_{11}, \quad \frac{d^2}{d\rho^2} \log|R| = -\operatorname{tr} R^{-1} G R^{-1} G = -t_{22}. \tag{3.23}$$

Anderson (1971), Lemma 6.7.9, shows that

$$I - \theta G^j = (1-4\theta^2)^{-\frac{1}{2}} \left(\frac{1}{2}\right)^{T+1} \left\{\left(1+\sqrt{1-4\theta^2}\right)^{T+1} - \left(1-\sqrt{1-4\theta^2}\right)^{T+1}\right\}. \tag{3.24}$$

Hence, we identify $\theta = -\rho$ and use this result directly. Let us denote $a = \sqrt{1-4\rho^2}$, so that $da/d\rho = -4\rho/a$. Then

$$|R| = |I + \rho G| = \frac{1}{a}\left(\frac{1}{2}\right)^{T+1}\left\{(1+a)^{T+1} - (1-a)^{T+1}\right\}, \tag{3.25}$$

$$\log|R| = -\log a - (T+1)\log 2 + \log\left\{(1+a)^{T+1} - (1-a)^{T+1}\right\} \tag{3.26}$$

$$\frac{d}{d\rho}\log|R| = \frac{4\rho}{a^2} + \frac{(T+1)\frac{4\rho}{a}\left\{-(1+a)^T - (1-a)^T\right\}}{(1+a)^{T+1} - (1-a)^{T+1}} \tag{3.27}$$

$$= \frac{4\rho}{a^2} - \frac{4(T+1)\rho}{a}\,\frac{(1+a)^T + (1-a)^T}{(1+a)^{T+1} - (1-a)^{T+1}}.$$

$$\frac{d^2}{d\rho^2}\log|R| = \frac{4a^2+32\rho^2}{a^4} - \frac{4(T+1)(a^2+4\rho^2)}{a^3}\,\frac{(1+a)^T + (1-a)^T}{(1+a)^{T+1} - (1-a)^{T+1}}$$
$$+ \frac{16(T+1)\rho^2}{a^2}\left[\frac{T\left\{(1+a)^{T-1} - (1-a)^{T-1}\right\}\left\{(1+a)^{T+1} - (1-a)^{T+1}\right\}}{\left\{(1+a)^{T+1} - (1-a)^{T+1}\right\}^2}\right.$$
$$\left. - \frac{(T+1)\left\{(1+a)^T + (1-a)^T\right\}^2}{\left\{(1+a)^{T+1} - (1-a)^{T-1}\right\}^2}\right]. \tag{3.28}$$

Simplifying slightly this last expression we have

$$\frac{d^2}{d\rho^2}\log|R| = \frac{4+16\rho^2}{a^4} - \frac{4(T+1)\left\{(1+a)^T + (1-a)^T\right\}}{a^3\left\{(1+a)^{T+1} - (1-a)^{T+1}\right\}} \tag{3.29}$$
$$- \frac{16(T+1)\rho^2}{a^2}\,\frac{(1+a)^{2T} + (1-a)^{2T} + 4T + 8\rho^2)(4\rho^2)^{T-1}}{\left\{(1+a)^{T+1} - (1-a)^{T+1}\right\}^2}.$$

4. Evaluation of quadratic forms and traces in the frequency domain

4.1. Introduction

We presented our results so far in the time domain. We can consider the effect of a Fourier transformation. Let K be the orthogonal $T \times T$ matrix with components

$$\sqrt{\frac{2}{T+1}}\sin\frac{\pi jk}{T+1}, \qquad j,k = 1,\ldots,T. \tag{4.1}$$

and let D be the diagonal matrix with diagonal elements

$$d_j = 2\cos\frac{\pi j}{T+1}, \qquad j = 1,\ldots,T. \tag{4.2}$$

Then $K'K = KK' = I$. $K'GK = D$. We have a way to diagonalize the matrix G appearing in (1.2). If $y = Kz$, then $z = K'y$, that is.

$$z_j = \sqrt{\frac{2}{T+1}}\sum_{k=1}^{T} y_k \sin\frac{\pi jk}{T+1}, \qquad j = 1,\ldots,T; \tag{4.3}$$

then z is multivariate normal with $\mathcal{E}z = 0$ and $\mathcal{E}zz' = \sigma_0 I + \sigma_1 D = \sigma_0(I + \rho D)$.

This then provides an alternative approach that we may call a "frequency domain approach": any expression in terms of G can be translated into an expression in terms of D, and any method formally presented in terms of y can be translated into a method presented in terms of z.

Since $I = KK'$ and $G = KDK'$, it follows that $R = I + \rho G = KK' + \rho KDK' = K(I + \rho D)K'$, so that R is also diagonalized by the orthogonal matrix K, and $I + \rho D$ has diagonal elements $1 + 2\rho\cos[\pi j/(T+1)]$. It then follows that $R^{-s} = K(I + \rho D)^{-s}K'$, $G^s = KD^sK'$, $s = 0, 1, \ldots$.

Putting these results together we find that the quadratic forms introduced in (1.5) can be written in terms of the z_j defined in (4.3) as

$$q_{jk} = y'R^{-(j+1)}G^k y = y'K(I+\rho D)^{-(j+1)}K'KD^kK'y$$
$$= (K'y)'(I+\rho D)^{-(j+1)}D^k(K'y) = z'(I+\rho D)^{-(j+1)}D^k z$$
$$= \sum_{s=1}^{T}\frac{d_s^k}{(1+\rho d_s)^{j+1}}z_s^2, \qquad j = -1, 0, 1, \ldots, \quad k = 0, 1, \ldots, \tag{4.4}$$

while the traces introduced in (1.6) become

$$t_{jk} = \operatorname{tr} R^{-j}G^k = \operatorname{tr} K(I+\rho D)^{-j}D^kK'$$
$$= \operatorname{tr}(I+\rho D)^{-j}D^kK'K = \operatorname{tr}(I+\rho D)^{-j}D^k$$
$$= \sum_{s=1}^{T}\frac{d_s^k}{(1+\rho d_s)^j}, \qquad j,k = 0, 1, \ldots . \tag{4.5}$$

4.2. Calculation of Fourier coefficients

The Fourier transformation of the observations (4.3) is different from the usual transformation

$$\sqrt{\frac{2}{T}}\sum_{k=1}^{T} y_k \cos\frac{2\pi jk}{T}, \qquad j = 0, 1, \ldots, \left[\frac{T}{2}\right], \tag{4.6}$$

$$\sqrt{\frac{2}{T}}\sum_{k=1}^{T} y_k \sin\frac{2\pi jk}{T}, \qquad j = 1, \ldots, \left[\frac{T-1}{2}\right]. \tag{4.7}$$

Since the matrix (4.1) diagonalizes G and that implied by (4.6) and (4.7) does not diagonalize G, the former yields simpler results, as indicated in Section 4.1.

For large T, the fast Fourier transform can be used for efficient computation of (4.3). We write for $j = 1, \ldots, T$

$$z_j = \sqrt{\frac{2}{T+1}} \sum_{k=1}^{T} y_k \sin \frac{\pi jk}{T+1} = \sqrt{\frac{2}{T+1}} \sum_{k=1}^{T+1} y_k \sin \frac{\pi jk}{T+1} \qquad (4.8)$$

for arbitrary y_{T+1} since $\sin [\pi jk/(T+1)] = 0$ for $k = T+1$. Further we have

$$z_j = \sqrt{\frac{1}{2(T+1)}} \sum_{k=1}^{2(T+1)} y_k \sin \frac{2\pi jk}{2(T+1)}, \qquad (4.9)$$

where $y_k = -y_{2(T+1)-k}$, $k = T+2, \ldots, 2T+1$, and y_{2T+2} is arbitrary. Then (4.9) has the usual form of the sine-transform for $2(T+1)$ observations and the usual computations for the fast Fourier transform are available.

4.3. Evaluation of quadratic forms

We want to calculate

$$y' R^{-(j+1)} y = q_{j0} = \sum_{s=1}^{T} \frac{z_s^2}{(1+\rho d_s)^{j+1}}, \qquad j = 0, 1, 2. \qquad (4.10)$$

Using $d_s = d_{T+1-s}$ and $d_{(T+1)/2} = 0$ for T odd, we can prove that:

$$q_{00} = \sum_{s=1}^{T/2} \frac{z_s^2 + z_{T+1-s}^2 - \rho d_s (z_s^2 - z_{T+1-s}^2)}{1-\rho^2 d_s^2}, \quad T \text{ even}, \qquad (4.11)$$

$$= z_{(T+1)/2}^2 + \sum_{s=1}^{(T-1)/2} \frac{z_s^2 + z_{T+1-s}^2 - \rho d_s (z_s^2 - z_{T+1-s}^2)}{1-\rho^2 d_s^2}, \quad T \text{ odd}.$$

$$q_{10} = \sum_{s=1}^{T/2} \frac{(1+\rho^2 d_s^2)(z_s^2 + z_{T+1-s}^2) - 2\rho d_s (z_s^2 - z_{T+1-s}^2)}{(1-\rho^2 d_s^2)^2}, \quad T \text{ even}. \qquad (4.12)$$

$$= z_{(T+1)/2}^2 + \sum_{s=1}^{(T-1)/2} \frac{(1+\rho^2 d_s^2)(z_s^2 + z_{T+1-s}^2) - 2\rho d_s (z_s^2 - z_{T+1-s}^2)}{(1-\rho^2 d_s^2)^2}, \quad T \text{ odd}.$$

$$q_{20} = \sum_{s=1}^{T/2} \frac{(1+3\rho^2 d_s^2)(z_s^2 + z_{T+1-s}^2) - (3\rho d_s + \rho^3 d_s^3)(z_s^2 - z_{T+1-s}^2)}{(1-\rho^2 d_s^2)^3}, \quad T \text{ even}.$$

$$= z_{(T+1)/2}^2 \qquad (4.13)$$

$$+ \sum_{s=1}^{(T-1)/2} \frac{(1+3\rho^2 d_s^2)(z_s^2 + z_{T+1-s}^2) - (3\rho d_s + \rho^3 d_s^3)(z_s^2 - z_{T+1-s}^2)}{(1-\rho^2 d_s^2)^3}, \quad T \text{ odd}.$$

The series form can be used to obtain

$$q_{j0} = \sum_{k=0}^{\infty} \frac{\Gamma(j+k+1)}{k!\Gamma(j+1)} (-\rho)^k \sum_{s=1}^{T} d_s^k z_s^2. \qquad (4.14)$$

4.4. Evaluation of traces

In the Appendix we prove that (3.22) holds. Further, by using the hypergeometric function we are able to express t_{10} as follows:

$$t_{10} = \frac{T+1}{\sqrt{1-\rho^2}} - \frac{1}{1-4\rho^2} + 2(T+1) \sum_{s=1}^{\infty} \frac{(2\rho)^{2s(T+1)}}{\sqrt{\pi}} \frac{\Gamma[s(T+1)+1/2]\Gamma[s(T+1)+1]}{\Gamma[2s(T+1)+1]} F[s(T+1)+1/2, s(T+1)+1; 2s(T+1)+1; 4\rho^2]. \qquad (4.15)$$

The argument in Section 4.3 can also be used for traces. In effect,

$$t_{10} = \sum_{s=1}^{T} \frac{1}{1+\rho d_s} = 2 \sum_{s=1}^{T/2} \frac{1}{1-\rho^2 d_s^2}, \quad T \text{ even}.$$

$$= 1 + 2 \sum_{s=1}^{(T-1)/2} \frac{1}{1-\rho^2 d_s^2}, \quad T \text{ odd}. \qquad (4.16)$$

$$t_{20} = \sum_{s=1}^{T} \frac{1}{(1+\rho d_s)^2} = 2 \sum_{s=1}^{T/2} \frac{1+\rho^2 d_s^2}{(1-\rho^2 d_s^2)^2}, \quad T \text{ even}.$$

$$= 1 + 2 \sum_{s=1}^{(T-1)/2} \frac{1+\rho^2 d_s^2}{(1-\rho^2 d_s^2)^2}, \quad T \text{ odd}. \qquad (4.17)$$

5. Operations needed to do the calculations

Quadratic forms and traces were given (in the frequency domain) in Sections 4.3 and 4.4, respectively, for T even and for T odd. To simplify the analysis in this section we consider the case of T even, since we are interested in orders of magnitude of the numbers of operations.

The traces t_{10} and t_{20} are given in (4.16) and (4.17). Assuming that $d_s = 2\cos \pi s/(T+1)$ is available in the computer, we calculate d_s^2 for $s = 1,\ldots,T/2$ once and for all, and use that $d_s = -d_{T+1-s}$ for $s = (T/2)+1,\ldots,T$: we also calculate ρ^2 once for each iteration. Then (4.16) requires $T/2$ multiplications $\rho^2 d_s^2$, $T/2$ subtractions $1-\rho^2 d_s^2$, $T/2$ inversions and $T/2$ additions. Formula (4.17) invoives additionally $T/2$ additions $1+\rho^2 d_s^2$, $T/2$ multiplications to obtain the square in the denominator, $T/2$ divisions, and $T/2$ additions. For t_{10} and t_{20} we have altogether $2T$ additions or subtractions, T multiplications, and T divisions. See Table I.

We next consider the calculation of the quadratic forms given in equations (4.11) to (4.13). The calculation of the z_j in (4.9) is about $T\log_2 T$ multiplications and additions, but that is done once and for all. For T even, q_{00} is given in (4.11). The sums $z_s^2 + z_{T+1-s}^2$ and differences $z_s^2 - z_{T+1-s}^2$ are calculated only once. The additional computations for one iteration is T multiplications to obtain ρd_s and then $\rho d_s(z_s^2 - z_{T+1-s}^2)$, $T/2$ additions, $T/2$ divisions, and $T/2$ additions. For T even, q_{10} is given by (4.12). This is additionally T multiplications, $T/2$ subtractions, $T/2$ divisions, and $T/2$ additions. Thus for q_{00} and q_{10} we have $2T$ additions or subtractions, $2T$ multiplications, and T divisions. For T even, q_{20} is given by (4.13). This is additionally $2T$ additions or subtractions, $5T/2$ multiplications, and $T/2$ divisions. Finally, for q_{00}, q_{10}, and q_{20} we have $4T$ additions or subtractions, $9T/2$ multiplications, and $3T/2$ divisions.

The calculations in the time domain were summarized in equations (3.8) to (3.19). To compute q_{00} and q_{10} in (3.17), we use formulas (3.10), (3.14), and (3.16), which were also given in Section 2.2. Considering as if we had T steps instead of the $T-1$ actually considered there, they involve $5T$ additions or subtractions, $4T$ multiplications, and $3T$ divisions. To compute q_{20} in (3.19) we use formula (3.18) that involves additionally $2T$ additions, $2T$ multiplications, and T divisions.

The traces t_{10} and t_{20} are calculated in (3.11), (3.12) and (3.13). In (3.11) there are $\sum_{s=2}^{T}(s-1) = \sum_{r=1}^{T-1} r = \frac{1}{2}T(T-1)$ multiplications. Then (3.12) involves

$$\sum_{s=2}^{T}\sum_{j=1}^{s} j = \sum_{s=2}^{T}\frac{1}{2}s(s+1) = \frac{1}{2}\sum_{s=1}^{T-1}(s+1)(s+2) = \frac{1}{2}\sum_{s=1}^{T-1}(s^2+3s+2)$$
$$= \frac{1}{2}\left\{\frac{(T-1)T(2T-1)}{6} + \frac{3T(T-1)}{2} + 2(T-1)\right\} \qquad (5.1)$$
$$= \frac{1}{6}(T-1)(T^2+4T+6)$$

multiplications and additions. To obtain $t_{10} = t_{10}^{(T)}$ in (3.13) we need additionally $T-1$ divisions and $T-1$ additions, and to obtain $t_{20} = t_{20}^{(T)}$ we need additionally

Table I

Number of Operations

Quantity Computed	Time Domain +/−	Time Domain ×	Time Domain ÷	Freq. Domain +/−	Freq. Domain ×	Freq. Domain ÷
q_{00}	$4T$	$3T$	$3T$	T	T	$\frac{T}{2}$
q_{10}	T	T	—	T	T	$\frac{T}{2}$
q_{00}, q_{10}	$5T$	$4T$	$3T$	$2T$	$2T$	T
q_{20}	$2T$	$2T$	T	$2T$	$\frac{5T}{2}$	$\frac{T}{2}$
q_{00}, q_{10}, q_{20}	$7T$	$6T$	$4T$	$4T$	$\frac{9T}{2}$	$\frac{4T}{2}$
t_{10}	$\frac{(T-1)(T^2+4T+12)}{6}$	$\frac{(T-1)(T^2+7T+6)}{6}$	$T-1$	T	$\frac{T}{2}$	$\frac{T}{2}$
t_{20}	$\frac{(T-1)(T+2)}{2}$	$\frac{(T-1)(T+2)}{2}$	$\frac{T(T-1)}{2}$	T	$\frac{T}{2}$	$\frac{T}{2}$
t_{10}, t_{20}	$\frac{(T-1)(T^2+7T+18)}{6}$	$\frac{(T-1)(T^2+10T+12)}{6}$	$\frac{(T-1)(T+2)}{2}$	$2T$	T	T

$\sum_{s=2}^{T} s = \frac{1}{2}(T+2)(T-1)$ additions, the same number of multiplications, and $\frac{1}{2}T(T-1)$ divisions. Thus t_{10} and t_{20} require altogether $(T-1)(T^2+7T+18)/6$ additions and subtractions, $(T-1)(T^2+10T+12)/6$ multiplications, and $(T-1)(T+2)/2$ divisions.

The number of operations can be summarized in Table I.

We can compare the different procedures by comparing the number of computations per iteration. The scoring procedures 1 and 3 require the computation of $\tilde{q}_{00}^{(i)}$, $\tilde{q}_{10}^{(i)}$, $t_{10}^{(i)}$, $t_{20}^{(i)}$, while the Newton-Raphson procedures require in addition the computation of $\tilde{q}_{20}^{(i)}$. It will be seen in the table above that except for the computation of the Fourier coefficients $z_1, \ldots, z_T$ the number of computations carried out in the frequency domain is substantially less than in the time domain. In particular, the number of operations for $t_{10}^{(i)}$ and $t_{20}^{(i)}$ is of the order $T^3/3$ in the time domain, but of the order $4T$ in the frequency domain. Since the advantage of the frequency domain is in the calculation of the traces, which do not require the Fourier transform of the data, the efficient calculation by any of the procedures is to compute the quadratic forms in the time domain and the traces in the frequency domain.

Of course, counting the number of operations is only one aspect of the evaluation of these methods. Also relevant are the speed of convergence and the behavior in medium-sized samples.

6. Relation to other procedures

6.1. Box-Jenkins procedures

In this section we consider the approach of Box and Jenkins (1976) for computing the quadratic form $\tilde{q}_{00}(\alpha) = y'P^{-1}y$ and its derivative $d\tilde{q}_{00}(\alpha)/d\alpha$ for any given value of α. Box and Jenkins (1976) proposed to estimate α by minimizing $\tilde{q}_{00}(\alpha)$; operating with this objective function is different from maximizing the likelihood function or minimizing the concentrated likelihood with respect to α because the determinant $|P|$ is ignored. See Box and Jenkins (1976), Chapter 7.

Let us consider the transformation from $(u_0, u')'$, which is $N(0, \sigma^2 I_{T+1})$ to $(u_0, y')'$, defined by

$$\begin{pmatrix} u_0 \\ u \end{pmatrix} = B^{-1}\begin{pmatrix} u_0 \\ y \end{pmatrix}, \qquad B = B_{T+1} = I_{T+1} + \alpha L_{T+1}, \tag{6.1}$$

where L_{T+1} has 1's along the diagonal immediately below its main diagonal and 0's elsewhere. Let

$$M = (B^{-1})'B^{-1} = \begin{pmatrix} m_{00} & m_{01} \\ m_{10} & M_{11} \end{pmatrix}. \tag{6.2}$$

where $m_{10} = m_{01}'$. Then the quadratic form in the exponent of the normal density of $(u_0, u')'$ is $-1/(2\sigma^2)$ times

$$(u_0, u')\begin{pmatrix} u_0 \\ u \end{pmatrix} = (u_0, y')\begin{pmatrix} m_{00} & m_{01} \\ m_{10} & M_{11} \end{pmatrix}\begin{pmatrix} u_0 \\ y \end{pmatrix} \tag{6.3}$$

$$= m_{00}\left(u_0 + \frac{1}{m_{00}}m_{01}y\right)^2 + y'\left(M_{11} - \frac{1}{m_{00}}m_{10}m_{01}\right)y.$$

Since the Jacobian of the transformation (6.1) is 1, in the normal density of $(u_0, u')'$ we can substitute $B^{-1}(u_0, y')'$ directly to get the normal density of $(u_0, y')'$, and this in turn can be expressed as the product of the marginal density of y times the conditional density of u_0 given y. The quadratic form in the exponent of the marginal normal density of y is $-1/(2\sigma^2)$ times

$$\tilde{q}_{00}(\alpha) = y'\left(M_{11} - \frac{1}{m_{00}}m_{10}m_{01}\right)y, \tag{6.4}$$

and the quadratic form in the exponent of the conditional normal density of u_0 given y is $-m_{00}\left(u_0 + \frac{1}{m_{00}}m_{01}y\right)^2/(2\sigma^2)$. Thus

$$\mathcal{E}(u_0|y) = -\frac{1}{m_{00}}m_{01}y, \qquad \mathcal{E}\left[\begin{pmatrix} u_0 \\ y \end{pmatrix}\middle| y\right] = \begin{pmatrix} -\frac{1}{m_{00}}m_{01} \\ I \end{pmatrix}y \tag{6.5}$$

because $y = \mathcal{E}(y|y)$, and hence

$$\left\{\mathcal{E}\left[\begin{pmatrix} u_0 \\ u \end{pmatrix}\middle| y\right]\right\}'\left\{\mathcal{E}\left[\begin{pmatrix} u_0 \\ u \end{pmatrix}\middle| y\right]\right\} = y'\left(-\frac{1}{m_{00}}m_{10}, I\right)B^{-1\prime}B^{-1}\begin{pmatrix} -\frac{1}{m_{00}}m_{01} \\ I \end{pmatrix}y$$

$$= y'\left(M_{11} - \frac{1}{m_{00}}m_{10}m_{01}\right)y = \tilde{q}_{00}(\alpha). \tag{6.6}$$

Thus, we have shown that

$$\tilde{q}_{00}(\alpha) = y'P^{-1}y = y'\left(M_{11} - \frac{1}{m_{00}}m_{10}m_{01}\right)y = \sum_{t=0}^{T}[\mathcal{E}(u_t|y)]^2. \tag{6.7}$$

To compute $\mathcal{E}(u_t|y)$ for $t = 0, 1, \ldots, T$ given y we use the process u_t introduced in (1.1) and a process of independent normal $(0, \sigma^2)$ random variables v_t for which

$$y_t = v_t + \alpha v_{t+1}, \tag{6.8}$$

that is, has the "time" reversed. Box and Jenkins (1976), Chapter 6, call this the "backward" form of the process.

From model (1.1) we have for a given value of α the recursive relations

$$\mathcal{E}(u_t|y) = \mathcal{E}(y_t|y) - \alpha\mathcal{E}(u_{t-1}|y), \qquad t = \ldots, -1, 0, 1, \ldots. \tag{6.9}$$

Since $\mathcal{E}(y_t|y) = y_t, t = 1,\ldots,T$, (6.9) would provide all needed conditional expectations if $\mathcal{E}(u_0|y)$ were known. Note that since the u_t's are independent with $\mathcal{E}u_t = 0$, and y depends only on $u_0, u_1, \ldots u_T$, it follows that $\mathcal{E}(u_{-1}|y) = \mathcal{E}(u_{-2}|y) = \ldots = 0$ and $\mathcal{E}(y_{-1}|y) = \mathcal{E}(y_{-2}|y) = \ldots = 0$. Hence, $\mathcal{E}(u_0|y) = \mathcal{E}(y_0|y)$. To find or approximate $\mathcal{E}(y_0|y)$ we use the recursive relation derived from (6.8), namely

$$\mathcal{E}(v_t|y) = \mathcal{E}(y_t|y) - \alpha\mathcal{E}(v_{t+1}|y), \qquad t = \ldots, -1, 0, 1, \ldots \tag{6.10}$$

We start by approximating $\mathcal{E}(v_{T+1}|y)$ by $\mathcal{E}v_{T+1} = 0$. In turn (6.10) is used to calculate $\mathcal{E}(v_T|y), \ldots \mathcal{E}(v_1|y)$. Since $\mathcal{E}(v_0|y) = 0, \mathcal{E}(y_0|y) = \alpha\mathcal{E}(v_1|y)$, which is the desired starting value.

If more accuracy in $\sum_{t=0}^{T}[\mathcal{E}(u_t|y)]^2$ is desired, one can calculate $\mathcal{E}(y_{T+1}|y) = \alpha\mathcal{E}(u_T|y)$ and $\mathcal{E}(v_{T+1}|y) = \mathcal{E}(y_{T+1}|y)$; note that $\mathcal{E}(v_{T+2}|y) = \mathcal{E}(v_{T+3}|y) = \cdots = 0$ and $\mathcal{E}(y_{T+2}|y) = \mathcal{E}(y_{T+3}|y) = \cdots = 0$. Now use (6.10) for another round of recursions. Thus we compute $\tilde{q}_{00}(\alpha) = \sum_{t=0}^{T}[\mathcal{E}(u_t|y)]^2$ for any given value of α. The analysis of $\tilde{q}_{00}(\alpha)$ is illustrated in Box and Jenkins (1976), Section 7.1, where it is denoted by $S(\theta)$ in general.

Suppose that α_0 is an initial value of α, and let $\mathcal{E}(u_t|y, \alpha_0)$ denote the value of the conditional expectation of u_t given y calculated for this value α_0. For any α we can approximate $\mathcal{E}(u_t|y, \alpha)$ as

$$\mathcal{E}(u_t|y, \alpha) \sim \mathcal{E}(u_t|y, \alpha_0) + \left.\frac{d\mathcal{E}(u_t|y, \alpha)}{d\alpha}\right|_{\alpha=\alpha_0} (\alpha - \alpha_0), \tag{6.11}$$

and $\tilde{q}_{00}(\alpha)$ as

$$\tilde{q}_{00}(\alpha) \sim \sum_{t=0}^{T}\left\{\mathcal{E}(u_t|y, \alpha_0) + \left.\frac{d\mathcal{E}(u_t|y, \alpha)}{d\alpha}\right|_{\alpha=\alpha_0} (\alpha - \alpha_0)\right\}^2. \tag{6.12}$$

Minimization of (6.12) with respect to α occurs at

$$\alpha - \alpha_0 = \frac{\sum_{t=0}^{T}\mathcal{E}(u_t|y, \alpha_0)\left.\frac{d\mathcal{E}(u_t|y, \alpha)}{d\alpha}\right|_{\alpha=\alpha_0}}{\sum_{t=0}^{T}\left[\left.\frac{d\mathcal{E}(u_t|y, \alpha)}{d\alpha}\right|_{\alpha=\alpha_0}\right]^2}. \tag{6.13}$$

From (6.9) and (6.10) we obtain

$$\frac{d\mathcal{E}(u_t|y)}{d\alpha} = -\mathcal{E}(u_{t-1}|y) - \alpha\frac{d\mathcal{E}(u_{t-1}|y)}{d\alpha}, \qquad t = 1, \ldots, T, \tag{6.14}$$

Table II

Quantity Computed	+/−	×
$\mathcal{E}(v_t\|y)$	$T+1$	$T+1$
$\mathcal{E}(u_t\|y)$	$T+1$	$T+1$
$\tilde{q}_{00}(\alpha)$	T	$T+1$
$\frac{d\mathcal{E}(v_t\|y)}{d\alpha}$	$T+1$	$T+1$
$\frac{d\mathcal{E}(u_t\|y)}{d\alpha}$	$T+1$	$T+1$
$\frac{d\tilde{q}_{00}(\alpha)}{d\alpha}$	$T+1$	$T+1$
Total	$6T+5$	$6T+6$

$$\frac{d\mathcal{E}(v_t|y)}{d\alpha} = -\mathcal{E}(v_{t+1}|y) - \alpha\frac{d\mathcal{E}(v_{t+1}|y)}{d\alpha}, \qquad t = 1, \ldots, T, \tag{6.15}$$

since y_t, $t = 1, \ldots, T$ does not depend on α. Further, we need

$$\frac{d\mathcal{E}(u_0|y)}{d\alpha} = \frac{d\mathcal{E}(y_0|y)}{d\alpha} \tag{6.16}$$

since $\mathcal{E}(u_{-1}|y) = \mathcal{E}(u_{-2}|y) = \cdots = 0$. To calculate (6.16) we use an approximation obtained by calculating (6.15) recursively from T replacing $d\mathcal{E}(v_{T+1}|y)/d\alpha$ by 0. This leads to

$$0 = \frac{d\mathcal{E}(v_0|y)}{d\alpha} = \frac{d\mathcal{E}(y_0|y)}{d\alpha} - \alpha\frac{d\mathcal{E}(v_1|y)}{d\alpha} - \mathcal{E}(v_1|y), \tag{6.17}$$

which is solved for $\alpha\mathcal{E}(y_0|y)/d\alpha$. Thus we obtain the constituents of (6.13).

The Taylor-series expansion is considered by Box and Jenkins (1976) in Section 7.2.

As in Section 5 we can now count the number of operations needed to do the calculations. These are summarized in Table II.

It should be emphasized that the minimization of $\tilde{q}_{00}(\alpha)$ is not the same as the maximization of the loglikelihood because of the factor $\log|P| = \log(1-\alpha^{2T+2})/(1-\alpha^2)$. Even for T so large that α^{2T+2} is negligible the term $\log(1-\alpha^2) \sim -\alpha^2$ may not be small enough to ignore. Each procedure studied in detail in this paper is exactly maximum likelihood in the sense that the iteration is meant to converge to a local maximum. Therefore, these iterative maximum likelihood procedures are not directly comparable to the Box-Jenkins procedures.

6.2. Using Woodbury's formula

We consider an approach that permits the calculation of quadratic forms of the type $y'P^{-j}y$. We illustrate the ideas with the cases $j = 1$ and 2. These quadratic forms can be used to implement iterative procedures in terms of α or to compute quadratic forms in an iterative procedure for ρ by using $\alpha = (1-\sqrt{1-4\rho^2})/(2\rho)$.

As in Section 6.1 let $B = I + \alpha L$, and let also $a = (\alpha, 0, \ldots, 0)'$. Then B is nonsingular, and if we denote $Q = BB'$, $P = Q + aa'$,

$$P^{-1} = Q^{-1} - \frac{1}{1+a'Q^{-1}a}\,Q^{-1}aa'Q^{-1}. \tag{6.18}$$

This is a simple case of a general formula called Woodbury's formula by some authors; see, for example, Phadke and Kedem (1978) and Press (1982).

Calculation of $y'P^{-1}y$

$$\begin{aligned} y'P^{-1}y &= y'Q^{-1}y - \frac{1}{1+a'Q^{-1}a}\,y'Q^{-1}aa'Q^{-1}y \\ &= y'B'^{-1}B^{-1}y - \frac{1}{1+a'B'^{-1}B^{-1}a}\,y'B'^{-1}B^{-1}aa'B'^{-1}B^{-1}y \\ &= z'z - \frac{z'kk'z}{1+k'k} = z'z - \frac{(z'k)^2}{1+k'k}, \end{aligned} \tag{6.19}$$

where z and k are $T \times 1$ vectors defined in the linear systems $y = Bz$, $a = Bk$, and the problem is to find z and k for given B, y and a. To solve $y = Bz$ for z we have

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = (I+\alpha L)z = z + \alpha Lz = \begin{pmatrix} z_1 \\ z_2 + \alpha z_1 \\ \vdots \\ z_T + \alpha z_{T-1} \end{pmatrix} \tag{6.20}$$

and hence $z_1 = y_1, z_j = y_j - \alpha z_{j-1}$, $j = 2, \ldots, T$. Solving by repeated substitution we obtain $z_j = \sum_{s=1}^{j}(-\alpha)^{j-s}y_s, j = 1, \ldots, T$. Proceeding in a similar manner, we solve $a = Bk$ for k, obtaining $k_j = -(-\alpha)^j, j = 1, \ldots, T$, so that to use in (6.19) we have $1 + k'k = (1-\alpha^{2T+2})/(1-\alpha^2)$.

We then proceed as follows: start with $z_1 = y_1$, $S_{11} = z_1^2$, $S_{21} = \alpha z_1$; then compute in succession

$$z_j = y_j - \alpha z_{j-1},\ S_{1j} = S_{1,j-1} + z_j^2,\ S_{2j} = S_{2,j-1} - (-\alpha)^j z_j, \quad j = 2, \ldots, T. \tag{6.21}$$

Then

$$y'P^{-1}y = S_{1T} - \frac{1-\alpha^2}{1-\alpha^{2T+2}}S_{2T}^2. \tag{6.22}$$

To calculate $y'P^{-2}y$ we first consider

$$\begin{aligned} P^{-1}y &= B'^{-1}(B^{-1}y) - \frac{1}{1+k'k}B'^{-1}(B^{-1}a)(B^{-1}a)'(B^{-1}y) \\ &= B'^{-1}z - \frac{1}{1+k'k}B'^{-1}kk'z = m - \frac{k'z}{1+k'k}n, \end{aligned} \tag{6.23}$$

and it suffices to find m and n in the linear systems $z = B'm$, $k = B'n$, where B is known and z and k are available from the calculation of $y'P^{-1}y$.

The components of n are $n_j = -(-\alpha)^j(1-\alpha^{2T-2j+2})/(1-\alpha^2)$, $j = 1, 2, \ldots, T$, from which $\sum_{j=1}^{T} n_j^2 = [\alpha^2 - (2T+1)(1-\alpha^2)\alpha^{2T+2} - \alpha^{4T+4}]/(1-\alpha^2)^3$. The calculations can be summarized as follows: start with $m_T = z_T$, $S_{3T} = z_T^2$, $S_{4T} = -(-\alpha)^T z_T$; then compute in succession

$$m_j = z_j - \alpha m_{j+1},\ S_{3j} = S_{3,j+1} + m_j^2,\ S_{4j} = S_{4,j+1} + m_j n_j, \quad j = T-1, \ldots, 1. \tag{6.24}$$

Then

$$\begin{aligned} y'P^{-2}y = {} & S_{31} + \frac{\alpha^2 - (2T+1)(1-\alpha^2)\alpha^{2T+2} - \alpha^{4T+4}}{(1-\alpha^2)(1-\alpha^{2T+2})^2}S_{2T}^2 \\ & - 2\frac{1-\alpha^2}{1-\alpha^{2T+2}}S_{2T}S_{41}. \end{aligned} \tag{6.25}$$

Tiao and Ali (1971) considered the ARMA (1,1) model; when it is specialized to the MA (1) model, the iterative calculations in their Section 2.1.1 are similar to those above for $y'P^{-1}y$. However, they do not consider $y'P^{-2}y$.

6.3. Estimation using the EM algorithm.

The analysis in the preceding section can be related to the EM algorithm for computing maximum likelihood estimates, as described for example in Dempster,

Laird and Rubin (1977). The generating equations for $y_1, \ldots, y_T$ coming from the MA(1) model (1.1) can be written as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} u_1 + \alpha u_0 \\ u_2 + \alpha u_1 \\ \vdots \\ u_T + \alpha u_{T-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ \alpha & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \alpha & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix} + u_0 \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{6.26}$$

In terms of the notation used in Section 6.2 we write this as

$$y = (I + \alpha L)u + u_0 a = Bu + u_0 a, \tag{6.27}$$

which in turn can be written as the transformation

$$\begin{pmatrix} u_0 \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0' \\ a & B \end{pmatrix} \begin{pmatrix} u_0 \\ u \end{pmatrix}. \tag{6.28}$$

We take $(u_0, u')'$ as $N(0, \sigma^2 I_{T+1})$. The transformation (6.28) has Jacobian equal to 1. and hence $(u_0, y')'$ is normal with expectation 0 and covariance matrix

$$\sigma^2 \begin{pmatrix} 1 & 0' \\ a & B \end{pmatrix} \begin{pmatrix} 1 & a' \\ 0 & B' \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & a' \\ a & aa' + BB' \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & a' \\ a & P \end{pmatrix}. \tag{6.29}$$

The determinant of this covariance matrix is 1.

To use the EM algorithm we augment the observations $y_1, \ldots, y_T$ by the unobserved u_0 and consider it as a "missing observation." The EM algorithm is an iterative procedure. Given preliminary values of the parameters α and σ^2 we obtain an estimate of u_0 say $u_0^{(1)}$ as the conditional expectation of u_0 given y and preliminary values of α and σ^2. Next we obtain maximum likelihood estimates of α and σ^2 on the basis of $(u_0^{(1)}, y')$. Because α appears only in the exponent of the normal distribution of (u_0, y'), this step amounts to minimizing the quadratic form in the exponent of the normal distribution of (u_0, y') and then maximizing the resulting concentrated likelihood with respect to σ^2. However, since the value of σ^2 is irrelevant to maximizing the likelihood with respect to α, one can carry out the iteration with respect to α and after its completion find the estimate of σ^2.

To study the joint density $f(u_0, y)$ we use $f(u_0, y) = g(u_0|y)h(y)$. From the covariance matrix (6.29) we find

$$\begin{aligned} \mathcal{E}(u_0|y) = a'P^{-1}y &= a'\left(Q^{-1} - \frac{1}{1 + a'Q^{-1}a} Q^{-1}aa'Q^{-1}\right) y \\ &= \frac{a'Q^{-1}y}{1 + a'Q^{-1}a}, \end{aligned} \tag{6.30}$$

$$\begin{aligned} \operatorname{Var}(u_0|y) = 1 - a'P^{-1}a &= 1 - a'\left(Q^{-1} - \frac{1}{1 + a'Q^{-1}a} Q^{-1}aa'Q^{-1}\right) a \\ &= \frac{1}{1 + a'Q^{-1}a}, \end{aligned} \tag{6.31}$$

while $\mathcal{E}y = 0$. $\operatorname{Var}(y) = P$. Hence. the exponent in the joint density of (u_0, y') is $-\frac{1}{2}$ times

$$\begin{aligned} &y'P^{-1}y + (1 + a'Q^{-1}a)\left(u_0 - \frac{1}{1 + a'Q^{-1}a} a'Q^{-1}y\right)^2 = \\ &\begin{pmatrix} u_0 \\ y \end{pmatrix}' \left\{ \begin{pmatrix} 0 & 0 \\ 0 & P^{-1} \end{pmatrix} + \begin{pmatrix} 1 + a'Q^{-1}a & -a'Q^{-1} \\ -Q^{-1}a & (1 + a'Q^{-1}a)^{-1} Q^{-1}ac'Q^{-1} \end{pmatrix} \right\} \begin{pmatrix} u_0 \\ y \end{pmatrix}. \end{aligned} \tag{6.32}$$

We now apply the EM algorithm.

E-step. For a given value of α calculate

$$\hat{u}_0 = \mathcal{E}(u_0|y) = \frac{a'Q^{-1}y}{1 + a'Q^{-1}a}. \tag{6.33}$$

M-step. Minimize with respect to α the quadratic form

$$\begin{aligned} &(\hat{u}_0, y') \left[\begin{pmatrix} 1 & 0' \\ a & B \end{pmatrix} \begin{pmatrix} 1 & a' \\ 0 & B' \end{pmatrix} \right]^{-1} \begin{pmatrix} \hat{u}_0 \\ y \end{pmatrix} \\ &\quad = (\hat{u}_0, y') \begin{pmatrix} 1 + a'Q^{-1}a & -a'Q^{-1} \\ -Q^{-1}a & Q^{-1} \end{pmatrix} \begin{pmatrix} \hat{u}_0 \\ y \end{pmatrix} \\ &\quad = \hat{u}_0^2(1 + a'Q^{-1}a) - 2\hat{u}_0 a'Q^{-1}y + y'Q^{-1}y. \end{aligned} \tag{6.34}$$

With the minimizing value of α repeat the E and M steps.

Since $Q = BB'$, $z = B^{-1}y$, $k = B^{-1}a$ as used in Section 6.2, we find that in this notation (6.33) is

$$\hat{u}_0 = \frac{a'B'^{-1}B^{-1}y}{1 + a'B'^{-1}B^{-1}a} = \frac{z'k}{1 + k'k}, \tag{6.35}$$

and (6.34) is

$$\hat{u}_0(1 + k'k) - 2\hat{u}_0 z'k + z'z. \tag{6.36}$$

If in (6.36) we substitute for $\hat{u}_0$ expression (6.35), we obtain

$$\frac{(z'k)^2}{1 + k'k} - 2\frac{(z'k)^2}{1 + k'k} + z'z = z'z - \frac{(z'k)^2}{1 + k'k} = y'P^{-1}y \tag{6.37}$$

in view of (6.19). Hence, we are minimizing $y'P^{-1}y$ with respect to α, but doing the iterations via the EM algorithm.

6.4. Use of the explicit components of the inverse covariance matrix

As indicated in Section 1, the likelihood function can be written as a function of α and σ^2 in terms of the determinant (2.21) and the components of P^{-1}. In effect,

$$y'P^{-1}y = \frac{1}{(1-\alpha^2)(1-\alpha^{2(T+1)})}\left\{\sum_{s=1}^{T} y_s^2(1-\alpha^{2s})(1-\alpha^{2(T-s+1)}) + 2\sum_{s=1}^{T-1}\sum_{t=1}^{T-s} y_s y_{s+t}(-\alpha)^t(1-\alpha^{2s})(1-\alpha^{2(T-s-t+1)})\right\}. \tag{6.38}$$

Godolphin and de Gooijer (1982) derived from the likelihood function, expressed in these terms, an iterative procedure for α.

6.5. Relations to optimal prediction

The likelihood function considered in Section 1 can be written as a function of α and σ^2 as

$$L^*(\alpha,\sigma^2) = (2\pi)^{-T/2}(\sigma^2)^{-T/2}|P|^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}y'P^{-1}y\right\}. \tag{6.39}$$

where P was defined in (1.2), $P = (1+\alpha^2)R$. The Cholesky decomposition of P derived in (2.18) is $P = UVU'$, where $U = F^{-1}$ as in (2.17), and $\tilde{V}$ is diagonal with diagonal elements $\tilde{v}_j = (1+\alpha^2)v_j$, $j = 1,\ldots,T$. Then $|P| = |U||\tilde{V}||U'| = \prod_{s=1}^{T}\tilde{v}_s$, because U is lower triangular with diagonal elements equal to 1, and $y'P^{-1}y = (U^{-1}y)'\tilde{V}^{-1}(U^{-1}y) = w'\tilde{V}^{-1}w = \sum_{s=1}^{T} w_s^2/\tilde{v}_s$. We then have

$$L^*(\alpha,\sigma^2) = (2\pi)^{-T/2}(\sigma^2)^{-T/2}\left(\prod_{s=1}^{T}\tilde{v}_s\right)^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{s=1}^{T}\frac{w_s^2}{\tilde{v}_s}\right\}. \tag{6.40}$$

Expression (6.40) can be related to the problem of minimum mean square error prediction (and hence to Kalman filtering), as several writers have recently emphasized. In effect, the optimal minimum mean square error predictor of y_s based on $y_{s-1},\ldots,y_1$ is the conditional expectation $\mathcal{E}(y_s|y_{s-1},\ldots,y_1)$, $s = 2,\ldots,T$, and we can prove in relation to (6.40) that in our case

$$w_s = y_s - \mathcal{E}(y_s|y_{s-1},\ldots,y_1), \qquad s = 2,\ldots,T, \tag{6.41}$$

is the error of the optimal prediction of y_s, while

$$\sigma^2\tilde{v}_s = \mathrm{Var}(w_s) = \mathcal{E}\left\{\left[y_s - \mathcal{E}(y_s|y_{s-1},\ldots,y_1)\right]^2|y_{s-1},\ldots,y_1\right\}. \tag{6.42}$$

Brockwell and Davis (1987) in their Section 5.2 and 8.6 provide an analysis by means of which (6.41) and (6.42) can be proved. In fact in their Section 5.2, while considering the MA (1) model, they provide iterative expressions that are equivalent to the interations for the v_j and w_j that we presented in Section 3.1. In our case (6.41) and (6.42) can also be proved by operating with conditional expectations and covariance marices of multivariate normal distributions.

Harvey (1981), while considering the Kalman filtering approach to the problem of estimation in the MA(1) model, gave (6.40), (6.41), and (6.42), and wrote the recursions (in our notation) as

$$\tilde{v}_s = 1 + \frac{\alpha^{2s}}{1+\alpha^2+\cdots+\alpha^{2(s-1)}} = \frac{1-\alpha^{2(s+1)}}{1-\alpha^{2s}}, \tag{6.43}$$

$$w_s = y_s - \alpha\frac{w_{s-1}}{\tilde{v}_{s-1,s-1}} = y_s + (-\alpha)\frac{1-\alpha^{2(s-1)}}{1-\alpha^{2s}}w_{s-1}. \tag{6.44}$$

Appendix

We prove that (3.22) and (4.15) hold and present approximations to the traces occuring in our procedures. Since the characteristic roots of $\boldsymbol{G}$ are $d_s = 2\cos\pi s/(T+1) = e^{i\pi s/(T+1)} + e^{-i\pi s/(T+1)}$, the characteristic roots of G^{2k} are

$$d_s^{2k} = \left(e^{i\frac{\pi s}{T+1}} + e^{-i\frac{\pi s}{T+1}}\right)^{2k} = \sum_{j=0}^{2k}\binom{2k}{j}e^{i\frac{\pi s}{T+1}j}e^{-i\frac{\pi s}{T+1}(2k-j)} = \sum_{j=0}^{2k}\binom{2k}{j}e^{i\frac{2\pi s}{T+1}(k-j)} \tag{A.1}$$

and

$$\mathrm{tr}\, G^{2k} = \sum_{j=0}^{2k}\binom{2k}{j}\sum_{s=1}^{T}e^{i\frac{2\pi(k-j)}{T+1}s}. \tag{A.2}$$

Since

$$\sum_{s=0}^{T}e^{i\frac{2\pi(k-j)}{T+1}s} = T+1 \quad \text{if } (k-j) = 0, \pm(T+1), \pm 2(T+1), \ldots, \tag{A.3}$$

and otherwise equal to 0, we have

$$\sum_{s=1}^{T} e^{i\frac{2\pi(k-j)}{T+1}s} = T \qquad \text{if } j = k, k \pm (T+1), k \pm 2(T+1), \ldots . \tag{A.4}$$

and otherwise equal to -1. Then

$$\operatorname{tr} G^{2k} = \sum_{j=0}^{2k} \binom{2k}{j} (-1)^j + (T+1) \sum_{\substack{j=0 \\ j=k, k\pm(T+1), \ldots}}^{2k} \binom{2k}{j}. \tag{A.5}$$

The first term in $\operatorname{tr} G^{2k}$ is -1 times $\sum_{j=0}^{2k}\binom{2k}{j} = 2^{2k}$. The second term is $T+1$ times

$$\begin{aligned} &\binom{2k}{k}, && k = 0, 1, \ldots, T, \\ &\binom{2k}{k} + 2\binom{2k}{k+(T+1)}, && k = T+1, \ldots, 2T+1, \\ &\vdots \end{aligned} \tag{A.6}$$

and so on. Then

$$\begin{aligned} t_{10} = \sum_{k=0}^{\infty} \left\{ \binom{2k}{k}(T+1) - 2^{2k} \right\} \rho^{2k} + 2(T+1) \sum_{k=T+1}^{\infty} \binom{2k}{k-(T+1)} \rho^{2k} \\ + 2(T+1) \sum_{k=2(T+1)}^{\infty} \binom{2k}{k-2(T+1)} \rho^{2k} + \ldots . \end{aligned} \tag{A.7}$$

The first term in the first sum in (A.7) is $T+1$ times

$$\begin{aligned} \sum_{k=0}^{\infty} \frac{(2k)!}{(k!)^2} \rho^{2k} &= \sum_{k=0}^{\infty} \frac{\Gamma(2k+1)}{k!\Gamma(k+1)} \rho^{2k} \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(k+\frac{1}{2})\Gamma(k+1)2^{2k}}{k!\Gamma(k+1)\sqrt{\pi}} \rho^{2k} \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(k+\frac{1}{2})}{k!\Gamma(\frac{1}{2})} (2\rho)^{2k} = \left[1-(2\rho)^2\right]^{-\frac{1}{2}}. \end{aligned} \tag{A.8}$$

We have used $n! = \Gamma(n+1)$ and the duplication formula for the gamma function. The first sum in t_{10} is

$$\frac{T+1}{\sqrt{1-4\rho^2}} - \frac{1}{1-4\rho^2}, \tag{A.9}$$

which is a good approximation to t_{10} because the neglected terms are $O(\rho^{2(T+1)})$. Thus

$$t_{10} = \frac{T+1}{\sqrt{1-4\rho^2}} - \frac{1}{1-4\rho^2} + 2(T+1) \sum_{j=1}^{\infty} \sum_{k=g(T+1)}^{\infty} \binom{2k}{k-g(T+1)} \rho^{2k}, \tag{A.10}$$

which is (3.22).

The sum on k in (A.10) can be related to the hypergeometric function. For each fixed $g = 1, 2, \cdots$

$$\begin{aligned} &\sum_{k=g(T+1)}^{\infty} \binom{2k}{k-g(T+1)} \rho^{2k} \\ &= \sum_{k=g(T+1)}^{\infty} \frac{(2k)!}{[k-g(T+1)]![k+g(T+1)]!} \rho^{2k} \\ &= \sum_{h=0}^{\infty} \frac{[2(h+g(T+1))]!}{h![h+2g(T+1)]!} \rho^{2h+2g(T+1)} \\ &= \rho^{2g(T+1)} \sum_{h=0}^{\infty} \frac{\Gamma[2h+2g(T+1)+1]}{h!\Gamma[h+2g(T+1)+1]} \rho^{2h} \\ &= \frac{\rho^{2g(T+1)} 2^{2g(T+1)}}{\sqrt{\pi}} \sum_{h=0}^{\infty} \frac{\Gamma[h+g(T+1)+1/2]\Gamma[h+g(T+1)+1]}{h!\Gamma[h+2g(T+1)+1]} (2\rho)^{2h} \\ &= \frac{(2\rho)^{2g(T+1)}}{\sqrt{\pi}} \frac{\Gamma[g(T+1)+1/2]\Gamma[g(T+1)+1]}{\Gamma[2g(T+1)+1]} \\ &\qquad F[g(T+1)+1/2, g(T+1)+1; 2g(T+1)+1; 4\rho^2], \end{aligned} \tag{A.11}$$

from which (4.15) follows. Here we used the definition of the hypergeometric function

$$F(a, b; c; z) = \sum_{j=0}^{\infty} \frac{\Gamma(a+j)}{\Gamma(a)} \frac{\Gamma(b+j)}{\Gamma(b)} \frac{\Gamma(c)}{\Gamma(c+j)} \frac{z^j}{j!}. \tag{A.12}$$

The expression in (A.9) is an approximation to t_{10}. This value can be obtained by approximating the sum defining t_{10} by a corresponding integral, as was done in Anderson (1971b). Besides (A.9) this procedure provides the following approximations:

$$t_{11} \sim -\frac{1}{\rho} \left\{ (T+1) \frac{1-\sqrt{1-4\rho^2}}{\sqrt{1-4\rho^2}} - \frac{4\rho^2}{1-4\rho^2} \right\}, \tag{A.13}$$

$$t_{20} \sim \frac{(T+1)\sqrt{1-4\rho^2}-(1+4\rho^2)}{(1-4\rho^2)^2}. \qquad (A.14)$$

$$t_{21} \sim \frac{-4(T+1)\rho\sqrt{1-4\rho^2}+8\rho}{(1-4\rho^2)^2}, \qquad (A.15)$$

$$t_{22} \sim \frac{T+1}{\rho^2}\left\{1-\frac{1-8\rho^2}{(1-4\rho^2)^{3/2}}\right\}-\frac{4+16\rho^2}{(1-4\rho^2)^2}. \qquad (A.16)$$

These five approximations also satisfy relations similar to those derived in Section 1 among the needed traces, and in particular they satisfy equations similar to (1.15).

Acknowledgements

The research of T.W. Anderson was supported by U.S. Army Research Office Contract No. DAAL03-89-K-0033. The research of R. P. Mentz was partially supported by the Argentine Research Council (CONICET).

Bibliography

Anderson, T. W. (1971a), *The Statistical Analysis of Time Series*. New York: John Wiley and Sons.

Anderson, T. W. (1971b). Estimation of covariance matrices with linear structure and moving average processes of finite order. Technical Report No. 6. Office of Naval Research Office Contract N00014-67-A-0112-0030. Stanford University, October 1971.

Anderson, T. W. (1977). Estimation for autoregressive moving average models in the time and frequency domains, *Ann. of Statis.*, 5, 842–865.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Second Edition, New York: John Wiley and Sons.

Anderson, T. W., and R. P. Mentz (1990). Iterative procedures for exact maximum likelihood estimation in the first-order Gaussian moving average model. Technical Report No. 25, U. S. Army Research Office Contract DAAL03-89-K-0033, Stanford University, November 1990.

Anderson, T. W., and Akimichi Takemura (1986). Why do noninvertible estimated moving averages occur? *J. Time Ser. Anal.*, 7, 235–254.

Box, G. E. P., and G. M. Jenkins (1976). *Time Series Analysis, Forecasting and Control*. Revised Edition. San Francisco: Holden-Day.

Brockwell, P. J., and R. A. Davis (1987). *Time Series: Theory and Methods*. New York: Springer-Verlag.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Statist. Soc. B*, 39, 1–38.

Godolphin, E. J., and J. G. de Gooijer (1982). On the maximum likelihood estimation of the parameters of Gaussian moving average process, *Biometrika*, 69, 443–451.

Golub, G. H., and C. F. Van Loan (1983). *Matriz Computations*. Oxford: North Oxford Academic.

Graybill, F. A. (1983). *Matrices with Applications in Statistics*. Second Edition. California: Wadsworth & Brooks/Cole.

Harvey, A. C. (1981). *Time Series Models*. Oxford: Philip Allan.

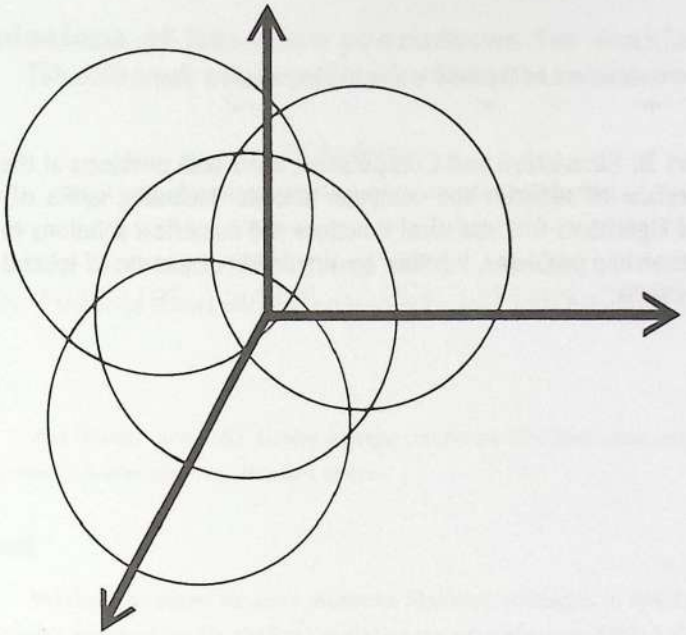
Phadke, M. S., and G. Kedem (1978). Computation of the exact likelihood function of multivariate moving average models, *Biometrika*, 65, 511–520.

Press, S. James (1982). *Applied Multivariate Analysis*. Second Edition, R. E. Krieger, Malibar, Florida.

Shaman, P. (1969). On the inverse of the covariance matrix of a first-order moving average, *Biometrika*, 56, 595–600.

Tiao, G. C. and M.M. Ali (1971). Analysis of correlated random effects: linear model with two random components, *Biometrika*, 58, 37-51.

Received July 1991; Revised November 1992



Communications in statistics

Simulation and computation

Communications in statistics
Part B: simulation and computation

Communications in Statistics is a multi-part journal.

Part B: Simulation and Computation deals with problems at the interface of statistics and computer science- including tables of, and algorithms for, statistical functions and numerical solutions to outstanding problems, whether by simulation or the use of special functions.

For subscription information write the **Promotion Department:**

MARCEL DEKKER, INC.
270 MADISON AVENUE
NEW YORK, NEW YORK 10016

Simulations of iterative procedures for maximum likelihood estimation in MA(1) models

T. W. Anderson
Department of Statistics, Stanford University, Stanford, CA 94305

R. P. Mentz, N. M. Jarma and C. I. Martinez
Institute of Statistical Studies, National University of Tucumán, 4000 Tucumán, Argentina

Key Words: first-order moving average; maximum likelihood estimates; scoring; Newton-Raphson; standard errors

Abstract

Iterative procedures for exact maximum likelihood estimation in MA(1) models were considered by the first two authors in a series of papers (1993a,b,c). The analysis was organized around four dichotomies: model parameters vs. autocorrelations, likelihood vs. concentrated likelihood, Newton-Raphson vs. scoring, and time vs. frequency domains. In this paper we study via simulations the behavior of the iterative procedures and how the choices between the indicated dichotomies relate to this behavior. One main conclusion is that for various values of the parameter and reasonable choices of the sample size, the procedures based on the scoring method are effective and the corresponding asymptotic theory can be used for inference. More difficulties were encountered with the Newton-Raphson method. Likelihood or concentrated

likelihood were found to make no important difference; some evidence for time vs. frequency domain calculations is similar.

1. Introduction

The first-order moving average model of time series, abbreviated here MA(1), is defined by $y_t = u_t + \alpha u_{t-1}$, where the u_t are i.i.d. $N(0,\sigma^2)$. We assume that $|\alpha| < 1$; that is, that the model is invertible. We often parameterize the model by using its first-order autocovariance $\sigma_0 = \sigma^2(1+\alpha^2)$ and its first-order autocorrelation $\rho = \alpha/(1+\alpha^2)$. In terms of these parameters the likelihood function of a vector $y = (y_1, \cdots, y_T)'$ of observations is

$$L(\sigma_0, \rho) = (2\pi\sigma_0)^{-T/2}|R|^{-1/2} \exp\{-y'R^{-1}y/(2\sigma_0)\}, \qquad (1)$$

where R is the $T \times T$ matrix of autocorrelations: $R = I + \rho G$, where I is the identity matrix and G has 1's along its two diagonals adjacent to the main diagonal, and 0's elsewhere.

Instead of maximizing (1) simultaneously with respect to its two parameters, we can maximize it with respect to σ_0 at $\sigma_0 = y'R^{-1}y/T$, and maximize with respect to ρ the function

$$n(\rho) = \{|R|(y'R^{-1}y)^T\}^{-1}; \qquad (2)$$

the "concentrated likelihood function" is a constant times the square root of (2). We call this the concentrated likelihood procedure.

In Anderson and Mentz (1993a) it is shown that, based on either (1) or (2) as indicated and using either the scoring or Newton-Raphson procedures, iterative estimation expressions are derived, that depend on quadratic forms and traces

$$q_{jk} = y'R^{-(j+1)}G^k y, \qquad t_{jk} = \mathrm{tr}R^{-j}G^k, \qquad (3)$$

whose evaluation is considered in Anderson and Mentz (1993b). It suffices to compute q_{j0} for $j = 0,1,2$ and t_{j0} for $j = 1,2$ since, for $\rho \neq 0$, $q_{jk} = \rho^{-k}\sum_{s=0}^{k}(-1)^{k-s}\binom{k}{s}q_{j-s,0}$ for $k \leq j$, and a similar relation holds for traces.

The purpose of this paper is to present some results obtained by simulation designed to evaluate the performance of three of the iterative expressions derived in the indicated ways. They are

$$\text{FORM1}: \ \{t_{22}q_{10} - t_{21}q_{11}\}\hat{\rho}^{(i)} = t_{20}q_{11} - t_{21}q_{10}, \qquad (4)$$

where the quadratic forms and traces are evaluated at the values of $\hat{\rho}^{(i-1)}$ in succession, starting from an initial value $\hat{\rho}^{(0)}$. FORM 1 arises in maximizing (1) with respect to ρ and using the scoring procedure. The next two procedures arise in maximizing (2) with respect to ρ, and both use the scoring procedure:

$$\text{FORM2}: \ t_{22}q_{00}\hat{\rho}^{(i)} = Tq_{11} - t_{21}q_{00}, \qquad (5)$$

$$\text{FORM3}: \ q_{00}[t_{22} - (t_{11})^2/T]\hat{\rho}^{(i)} = Tq_{11} - t_{11}q_{00} + q_{00}[t_{22} - (t_{11})^2/T]\hat{\rho}^{(i-1)}. \qquad (6)$$

Two other iterative estimating expressions were derived from (1) and (2) by using the Newton-Raphson procedure, and these will be considered below.

For the calculations used in this paper, the quadratic forms and traces were evaluated in the time domain by considering linear systems $y = Rx$ and solving them for x recursively by the method of successive elimination. A different approach consists in diagonalizing G (and hence of $R = I + \rho G$) to derive $q_{jk} = \sum_{s=1}^{T} d_s(1+\rho d_s)^{-(j+1)}z_s^2$, where $d_s = 2\cos[\pi s/(T+1)]$ and

$$z_s = [2(T+1)]^{-\frac{1}{2}}\sum_{k=1}^{2(T+1)} y_k \sin\{2\pi sk/[2(T+1)]\} \qquad (7)$$

with $y_k = -y_{2(T+1)-k}$, $k = T+2, \cdots, 2T+1$, and y_{T+1} and y_{2T+2} are arbitrary. Then (7) has the usual form of the sine transform for $2(T+1)$ observations, and the usual computations for the fast Fourier transform are available. Similarly, $t_{jk} = \sum_{s=1}^{T} d_s(1+\rho d_s)^{-j}$. These expressions were used in our analysis mostly to check those in the time domain; see below.

In the simulations pseudo-random $N(0,1)$ u_t's were generated with the BMDP statistical package. When a run was generated, the first 100 numbers were discarded, and also a set of numbers between the series that were used (to simplify the numbering of the simulated observations). Series of lengths $T = 100$ and 250 were generated with $\alpha = 0.30$ or 0.90 to obtain $y_1, \cdots, y_T$ according to the MA(1) model. For each choice of parameter, sample size, and initial value (see below), 100 replications were done.

A standard choice of initial value for each iteration is $\hat{\rho}^{(0)} = r$, the first-order sample autocorrelation; however, in general, better results are obtained by using r_c , a corrected value, defined by $r_c = r$ if $|r| \leq 0.5$, $1 - r$, if $r > 0.5$, $-1 - r$ if $r < -0.5$. Other choices of initial values were considered in some cases: (a) The (final) estimate of α provided by the BMDP package;

(b) An estimate of α obtained by using the "long autoregression procedure" of Durbin (1959), in which $\hat{\alpha} = -\sum_{i=0}^{k} \hat{\beta}_i \hat{\beta}_{i+1} / \sum_{i=0}^{k} \hat{\beta}_i^2$, the $\hat{\beta}_i$ are fitted to the data, and only $k = 4$ was used in our case; (c) A fixed value $\alpha = 0.10$, which corresponds to $\rho = 0.09900$. These alternative choices will be considered below.

Results of the simulation studies will be summarized here; for full details see the technical report by the same authors (1994).

2. Description and analysis of results

2.1. The structure of Tables I and II

The main results of the simulation study are presented in Tables I and II. Only results for ρ are studied, since those for α represent a simple transformation. The notation used here and elsewhere is T for sample size, n for number of replications (hence, $n = 100$), and k_i number of iterations needed to produce a final estimate $\hat{\rho}_i$.

In Table I column 1 reports the mean of the estimated ρ's, that is $m(\hat{\rho}) = m = \sum_{i=1}^{n} \hat{\rho}_i / n$, whose standard error is estimated by $ste(m) = s/\sqrt{n}$, where s is the sample standard deviation reported in column 4, $s(\hat{\rho}) = s = [\sum_{i=1}^{n} (\hat{\rho}_i - m)^2/(n-1)]^{1/2}$. Column 2 reports the estimated bias for ρ, defined by $bi(\hat{\rho}) = m - \rho$, and its estimated standard error which is also $s/\sqrt{n}$. Column 3 reports the estimated relative bias, defined by $rbi(\hat{\rho}) = bi(\hat{\rho})/\rho$, and its estimated standard error which is $s/(\rho\sqrt{n})$.

In column 4 we use the approximation for the standard error of

$$ste(s) \approx s/[2(n-1)]^{\frac{1}{2}}, \tag{8}$$

which we justify as follows: s^2 is based on n replications; the random variable $(n-1)s^2/\sigma^2(\hat{\rho})$ is chi square with $n-1$ degrees of freedom, and also asymptotically normal with parameters $n-1$ and $2(n-1)$; its square root is asymptotically normal with variance $2(n-1)/[2\sqrt{n-1}]^2$, and s is asymptotically normal with variance $\sigma^2(\hat{\rho})/[2(n-1)]$, from which (8) follows.

Column 5 reports the asymptotic standard deviation of the maximum likelihood estimator of ρ, evaluated at $\alpha = \alpha(m)$, obtained by inverting $m = a/(1+\alpha^2)$. It is given by

TABLE I

Figures of merit in the estimation of ρ.

Part a. $\alpha = 0.30$ ($\rho = 0.27523$), $T = 100$, $\hat{\rho}^{(0)} = r$

	$m(\hat{\rho}) = m$	$bi(\hat{\rho})$	$rbi(\hat{\rho})$	$s(\hat{\rho})$	$as_m(\hat{\rho})$	$rs_\rho(\hat{\rho})$	$mse(\hat{\rho})$
	1	2	3	4	5	6	7
FORM1	0.26792	−0.00731	−0.0266	0.07023	0.07450	0.961	0.004986
St.Errors		(0.00702)	(0.0255)	(0.00499)		(0.068)	
FORM2	0.26794	−0.00729	−0.0265	0.07024	0.07450	0.961	0.004987
St.Errors		(0.00702)	(0.0255)	(0.00499)		(0.068)	
FORM3	0.26795	−0.00728	−0.0265	0.07024	0.07450	0.961	0.004987
St.Errors		(0.00702)	(0.0255)	(0.00499)		(0.068)	
r	0.25010	−0.02513		0.08725			0.008244
BMDP	0.27022	−0.00501	−0.0182	0.07061	0.07405	0.966	0.005011

Part b. $\alpha = 0.90$ ($\rho = 0.49724$), $T = 100$ $\hat{\rho}^{(0)} = r_c$

FORM1	0.49682	−0.00042	−0.00084	0.00406	0.00282	1.605	0.0000166
St.Errors		(0.00041)	(0.00082)	(0.00029)		(0.115)	
FORM2	0.49704	−0.00020	−0.00039	0.00465	0.00266	1.838	0.0000216
St.Errors		(0.00047)	(0.00095)	(0.00033)		(0.131)	
FORM3	0.49773	0.00049	0.00098	0.00510	0.00217	2.016	0.0000263
St.Errors		(0.00051)	(0.00103)	(0.00036)		(0.144)	
r_c	0.43963	−0.05761		0.04889			0.0057095
BMDP	0.49750	0.00026	0.00052	0.00367	0.00234	1.452	0.0000135

Part c. $\alpha = 0.90$ ($\rho = 0.49724$), $T = 250$, $\hat{\rho}^{(0)} = r_c$

FORM1	0.49689	−0.00035	−0.00070	0.00198	0.00175	1.238	0.00000404
St.Errors		(0.00020)	(0.00040)	(0.00009)		(0.089)	
r_c	0.46371	−0.03353		0.03041			0.00204903

TABLE II

Average (standard deviations) numbers of iterations in the estimation of ρ.

α	T	FORM1	FORM2	FORM3
0.30	100	3.89	3.89	3.84
		(1.62)	(1.66)	(1.60)
0.90	100	4.28	4.25	4.58
		(2.18)	(2.17)	(2.47)
0.90	250	3.33		
		(1.23)		

$$as_m(\hat{\rho}) = \{T^{-1}[1-\alpha^2]^3/[1+\alpha^2]^4\}^{\frac{1}{2}}, \tag{9}$$

and is obtained from the asymptotic variance of the maximum likelihood estimator of α, which is $(1-\alpha^2)/T$. The sample and asymptotic standard deviations are compared in column 6 by means of the ratio $rs_\rho(\hat{\rho}) = s/as_\rho(\hat{\rho})$, whose denominator is evaluated at the theoretical value of ρ; its estimated standard error is $ste[rs_\rho(\hat{\rho})] \approx ste(\hat{\rho})/as_\rho(\hat{\rho})$.

Finally, column 7 reports the mean square error, $mse(\hat{\rho})$, whose standard error (not included in the table) can be approximated as follows:

$$ste[mse(\hat{\rho})] \approx \{2s^4[(n-1)^{-1}+n^{-2}]\}^{\frac{1}{2}}. \tag{10}$$

Table II contains the average numbers of iterations used by the three procedures to produce the estimates of ρ, followed by the corresponding estimated standard deviations.

2.2. Numerical results of the simulations

Table I, Part a, shows the results for series generated with $\alpha = 0.30$, which corresponds to $\rho = 0.27523$. The initial value is r; in no case did it exceed 0.50 in absolute value.

The results obtained when FORM1, FORM2 and FORM3 were used are very similar. The biases and relative biases are of the same order of magnitude as their standard errors. We computed the standard errors of differences between pairs, and found that the differences are far from significant. In effect, standardized differences in bias are as follows:

pair (1,2), $-0.00002/0.0000562 \approx -.4$,
pair (1,3), $-0.00003/0.0000592 \approx -.5$,
pair (2,3), $-0.00001/0.0000309 \approx -.3$.

Standard deviations are in column 4 with their standard errors, the value predicted by the asymptotic theory in column 5, and their comparison through the ratio $rs_\rho(\hat{\rho})$ in column 6. The latter shows, for each method, no significant difference. A likelihood ratio test was used to compare pairs of variances, and no significant differences were found; for details see the appendix.

The next to last line of Part a, contains results when r is considered as an estimator of ρ. Biases are much larger than for the three FORM's. Since $s/\sqrt{T} = 0.00872$, the bias of -0.02513 is significantly different from 0. The standard deviation in column 4 is larger than those of the three FORM's. These two facts make the estimated $mse(\hat{\rho})$ larger.

The last line of Part a contains results for the estimator provided by the BMDP time series program. The BMDP package that we used estimates the parameters of Gaussian ARMA models by the procedures suggested originally by Box and Jenkins (1970); a description of the procedures and their comparison with the "exact" approach we use is given in Anderson and Mentz (1993b). It is found to have bias smaller than the three FORM's, and standard deviation slightly larger, so that the $mse(\hat{\rho})$ turns out being coincident when rounded to four decimal places.

Table II shows that less than 4 iterations are needed on the average to produce a final estimate by using any of the FORM's, with a standard deviation of approximately 1.6 iterations.

A similar analysis is performed in Table I, Part b, with $T = 100$ and $\alpha = 0.90$, which corresponds to $\rho = 0.49724$, values close to the boundary of the region of invertibility of the MA(1) model into an infinite autoregression.

The corrected r was used as initial value here. For FORM1, FORM2, and FORM3 the conclusions about biases (columns 1, 2 and 3) are similar to those in the previous case: biases and relative biases are small, of the same order of magnitude as their standard errors, or smaller; differences between pairs of FORM's are not significant: the standardized differences in biases are

$$\text{pair (1,2)}, -0.00022/0.000143 \approx -1.54,$$
$$\text{pair (1,3)}, -0.00081/0.001742 \approx -0.46,$$
$$\text{pair (2,3)}, -0.00069/0.001757 \approx -0.39.$$

However, a different picture emerges in analyzing standard deviations (columns 4, 5 and 6): estimated values are larger than predicted by the asymptotic theory, and hence the ratios in column 6 are significantly different from 1.

Results for r_c as an estimator of ρ show a performance that in relation to the three FORM's is poorer than for $\alpha = 0.30$. Biases are larger and highly significantly different from 0; the standardized bias is $-\sqrt{100}\ 0.05761/0.04889 \approx -11.78$. The estimated standard deviation is of the order of 10 times that of the FORM's.

Results for the estimator originated in the BMDP program are in general better than those of any of the FORM's in terms of bias, standard deviation and mean square error. However, column 6 shows that the fit of the standard deviation to the predicted asymptotic value is not satisfactory.

Table II shows that the average number of iterations has experienced a small increase in going from $\alpha = 0.30$ to 0.90, as was to be expected.

When $\alpha = 0.90$ and $T = 100$, the agreement of the simulated results with the asymptotic theory is not good in the sense that the empirical standard deviations are much larger than the value predicted by the asymptotic theory. We take this as an indication that for that value of α $T = 100$ is too small for using the asymptotic approximations. To check this, a new simulation was done under the same conditions, but with $T = 250$. Only FORM1 was used, and r_c was taken as the initial value. The results are reported in Part c of Table I. We now find that the bias and relative bias are not significantly different from 0 and that the empirical standard deviation is closer to the value predicted by the asymptotic theory than when T was 100.

For comparison, the results obtained by using r_c as an estimator of ρ are included in Part c: bias, standard deviation and mean square error are considerably larger than those of FORM1. Standardized bias is $-\sqrt{250}\ 0.03353/0.03041 \approx -17.40$.

Table II shows that the average number of iterations has diminished in comparison to the case where $T = 100$, and also in comparison to the case of $\alpha = 0.30$ and $T = 100$.

3. Some other questions of interest

3.1. Initial values

Throughout Section 2 r_c was used as initial value to start the iterative procedures aimed at estimating ρ in agreement with common practice, since r is usually available at an early stage of the empirical analysis of a time series. Since the iterative procedures may depend on the initial value, other choices were explored when $T = 100$.

The estimate generated by the time series program of the BMDP package (see Table I), was used also as an initial value for the three iterative procedures. Its use did not affect the fit of the empirical results to the asymptotic theory and reduced considerably the required number of iterations. Two other choices, the estimate coming from the "long autoregression" approach mentioned at the end of Section 1 and the use of a fixed value $\hat{\rho}^{(0)} = 0.09900$ (corresponding to $\alpha = 0.10$), had a different effect; they did not affect the fit to the asymptotic theory, but increased the number of iterations.

We conclude that for the three proposed iterative procedures, when series were generated with either $\alpha = 0.30$ or 0.90 and $T = 100$ was the sample size, there is little effect if a "reasonable" choice of initial value is used for the iterations. Hence, selection of r_c is safe for these iterative procedures.

For a different case of the effect of the choice of initial value, see Section 3.4.

3.2. Shape analysis

Besides the figures of merit studied in Table I, the estimates $\hat{\rho}_i$ were

analyzed graphically for $\alpha = 0.30$ and 0.90, $T = 100$ and 250, and the various choices of initial values. For $\alpha = 0.30$ and $T = 100$, the histograms were approximately normal, and differences of shapes among graphs were small. Further, dispersion diagrams for pairs of methods gave results mostly along straight lines. This is in part due to the fact that the same simulated numbers were used for all variants of a method, a design chosen to control variability.

For $\alpha = 0.90$ and $T = 100$ the distributions were asymmetrical; there were values falling above 0.50, and the differences among graphs were most noticeable. Dispersion diagrams showed in a clear form the presence of the exceptional values.

For $\alpha = 0.90$, $T = 250$, and $\hat{\rho}^{(0)} = r_c$, better approximations to normal shapes were clear, in comparison with the case of $T = 100$.

The numbers of iterations were also studied graphically, and we concluded that the shapes of their distributions follow the same type of behavior that we observed for the estimates.

Pairs $(\hat{\rho}_i, k_i)$ were also studied, for all α's and T's, and correlation coefficients were less than 0.02 in all cases.

3.3. Other parameter values

Two other parameter values were considered, $\alpha = 0.50$ and 0.70. The three iterative procedures, in particular FORM1, were used to process series simulated with these values. Some of the analysis was done in relation to the subject of Section 3.4. In all cases it was found that the behavior of the iterative procedures was in agreement with that for $\alpha = 0.30$ and 0.90.

3.4. Scoring vs. Newton Raphson

In Anderson and Mentz (1993a) two other iterative estimating expressions are derived for the main parameter of a MA(1) model, coming from the maximization of (1) or (2) and using the Newton-Raphson procedure. These procedures were investigated in detail with series simulated with $\alpha = 0.30$, 0.50, 0.70 and 0.90, $T = 100$, and r_c as initial value. It was found that

for the smallest α the numerical estimates of the parameter and the average number of iterations required to produce the final estimates were comparable to the three methods originated in the scoring procedure, except for a few cases of nonconvergence. However, as the underlying parameter was increased, many cases of lack of convergence appeared. The main reason was found to be the strong dependence of these procedures upon the choice of initial value. Even for α as large as $\alpha = 0.90$, numerical results comparable to those of FORM1 were obtained in some cases when adequate (i.e. close to the solution) initial values were provided for each series.

In conclusion, the two procedures stemming from the Newton-Raphson approach cannot be recommended as comparable to those stemming from the scoring approach due to their dependence on the choice of initial value. Our experience in this case has been similar to what is usually considered in similar contexts, see, for example, Johnson and Dean Riess (1977, Section 4.3.3) or Burden and Faires (1985, Section 2.3).

3.5. Time vs. frequency domain calculations

In doing the studies of the preceding section, it was found useful to check the numerical calculations of quadratic forms and traces by using the two procedures mentioned in Section 1. As we wrote there, the bulk of the computations were done in the time domain, one reason being that these methods were stressed in Anderson and Mentz (1993b) for analytic purposes. However, in the more detailed analysis and checking of computations, those described for the frequency domain were used, and agreement between them was found to be complete.

4. Conclusions

We analyzed the behavior of a practical implementation of three iterative procedures proposed by Anderson and Mentz (1993a, b) to estimate by (exact) maximum likelihood the main parameter of a Gaussian MA(1) time series model. In the implementation the number of iterations was limited to a

maximum of 20, after which either a numerical value was available as a final estimate or the procedure was stopped and the estimation declared nonfeasible. The iterations were stopped before 20 if the absolute value of the difference between two consecutive values fell below 0.0005.

Parameter values of $\alpha = 0.30$ and 0.90 were considered. It is expected that 0.30 will produce little difficulties, while 0.90 can produce considerably more problems: 0.90 is closer to the boundary of the "invertibility region," since the MA(1) process can be inverted into an infinite autoregression if $|\alpha| < 1$. We also studied in less detail $\alpha = 0.50$ and 0.70.

For $\alpha = 0.30$ only $T = 100$ was studied, and the three iterative procedures worked well in terms of bias and variability, irrespective of which type of initial value was used.

For $\alpha = 0.90$ $T = 100$ and 250 were studied. The first sample size turned out to be too small to produce results in agreement with the asymptotic theory. The larger sample size produced results in good agreement with the asymptotic theory.

For $\alpha = 0.90$ the procedures may produce final estimates of ρ larger than 0.50 in absolute value. The adequate bound for ρ is $|\rho| < 1/\{2\cos[\pi/(T+1)]\} \approx 0.50024$ when $T = 100$ (Anderson and Takemura (1986). For general use of a procedure, a modification may be introduced to force the final estimates to be less than or equal to 0.50 in absolute value, and we know that this will occur only infrequently.

Our conclusions can be stated as follows: The scoring iterative procedures (i.e., FORM1, FORM2, FORM3) with r_c as initial value yield rather unbiased estimates at $\alpha = 0.30$ and 0.90 and $T = 100$; for this sample size the standard error is estimated adequately by the asymptotic formula at $\alpha = 0.30$, but is underestimated at $\alpha = 0.90$. At $T = 250$ and $\alpha = 0.90$ the standard error is estimated adequately by the asymptotic formula. The number of iterations is fewer than 4 at $T = 100$ and $\alpha = 0.30$, and only a little more than 4 at $T = 100$ and $\alpha = 0.90$. Further, these conclusions are not very sensitive to the initial value as long as it is not too different from the true parameter value.

Referring now specifically to the four dichotomies mentioned in the summary and elsewhere, we have: (a) The difference between the scoring and the Newton-Raphson procedures is important, since the latter was found to be dependent upon the choice of starting value for the iterations, and in some cases it demanded that this initial value be close to that to be estimated. (b) The choice between using the likelihood or the concentrated likelihood functions appeared as not important. In Section 2 FORM1, arising from the likelihood function, and FORM2, from the concentrated likelihood function (both under scoring), exhibited similar behaviors; in Section 3.4 two other methods arising from the different approaches (both under Newton-Raphson) also behaved in similar ways. (c) The choice between time and frequency domains for the calculations, was not explored in detail, since the bulk of our calculations were done in the time domain; however, when the frequency domain calculations were used as checks, no differences existed between them. (d) Finally, the choice of parametrizations was not explored in this paper since we choose to set our iterative procedures in terms only of ρ.

Appendix
Likelihood ratio test of equality of variances

The same pseudo-random numbers were used for all of the methods. If X denotes the estimator generated by one FORM and Y by another, we take the pair (X, Y) as bivariate normal with unknown means, variances and "intraclass" correlation coefficient R. The unrestricted maximum likelihood estimators of these parameters are the sample means, sample variances $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$, and Pearson's sample correlation coefficient $\hat{R} = \hat{\sigma}_{xy}/\hat{\sigma}_x\hat{\sigma}_y$. Under the null hypothesis of equality of variances, the estimators are the sample means, $\hat{\sigma}_0^2 = \frac{1}{2}(\hat{\sigma}_x^2 + \hat{\sigma}_y^2)$ and $\hat{R}_0 = \hat{\sigma}_{xy}/\hat{\sigma}_0^2$. Hence the likelihood ratio statistic λ to test equality of variances satisfies

$$\lambda^{2/n} = \frac{\hat{\sigma}_x^2\hat{\sigma}_y^2(1-\hat{R}^2)}{[\frac{1}{2}(\hat{\sigma}_x^2+\hat{\sigma}_y^2)]^2(1-\hat{R}_0^2)} \ , \tag{11}$$

and we reject $\sigma_x^2 = \sigma_y^2$ if λ is small. For $T = 100$ we can use that $\phi = -2\log_e \lambda$ is approximately chi square with one degree of freedom, and hence we have:

$$\begin{aligned} \alpha = 0.30,\ \text{pair}(1,2), \phi &= 2.26, \\ \text{pair}(1,3), \phi &= 1.75, \\ \text{pair}(2,3), \phi &= 0.04, \end{aligned}$$

and in comparison with the 0.05 values of 3.84, we accept the equality of the variances. Also,

$$\alpha = 0.90, \quad \text{pair}(1,2)\phi = 4.35,$$
$$\text{pair}(1,3),\phi = 5.63,$$
$$\text{pair}(2,3),\phi = 0.76,$$

where the 0.01 value is 6.63. Again, we notice a difference in the behavior of the results for $T = 100$, according to whether $\alpha = 0.30$ or 0.90.

Acknowledgements

The research of the last three authors was partially supported by the Argentine Research Council (CONICET).

Bibliography

Anderson, T. W., and R. P. Mentz (1993a). Iterative procedures for exact maximum likelihood estimation in the first-order Gaussian moving average model, in *Statistics and Probability: A Raghu Raj Bahadur Festchrift* (J. K Ghosh, S. K. Mitra, R. R. Parthasarathy, and B. L. S. Prakasa Rao, eds), Wiley Eastern Limited, 19-35.

Anderson, T. W., and R. P. Mentz (1993b). Evaluation of quadratic forms and traces for iterative estimation in first-order moving average models, *Communications in Statistics: Theory and Methods*, 22, 931-963.

Anderson, T. W., and R. P. Mentz (1993c). A note on maximum likelihood estimation in the first-order Gaussian moving average model, *Statistics and Probability Letters*, 16, 205-211.

Anderson, T.W., R. P. Mentz, N. M. Jarma, and C. I. Martinez (1994). Simulations of iterative procedures for maximum likelihood estimation in MA(1) models, Technical Report 296, NSF Grant DMS 93-01366. Department of Statistics, Stanford University.

Anderson, T. W., and A. Takemura (1986). Why do noninvertible estimated moving averages occur? *Journal of Time Series Analysis*, 7, 235-254.

BMDP (1990), *BMDP Statistical Software*, Los Angeles, California.

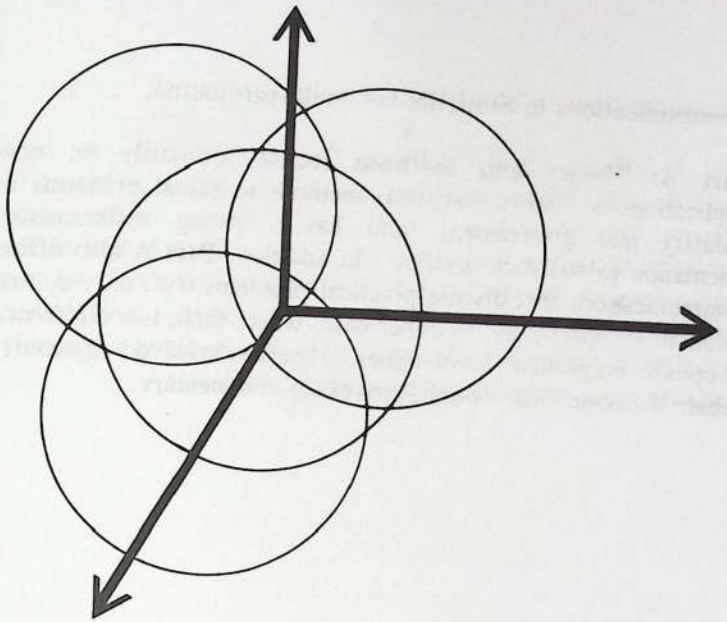
Box, G. E. P., and G. M. Jenkins (1970). *Time Series Analysis Forecasting and Control*, Holden Day, San Francisco, California.

Burden, R. L., and J. D. Faires (1985). *Numerical Analysis*, PWS, Boston, Massachusetts.

Durbin, J. (1959). Efficient estimation of parameters in moving-average models, *Biometrika*, 46, 306-316.

Johnson, L. W., and R. Dean Riess (1977) *Numerical Analysis*, Addison-Wesley, Reading, Massachusetts.

Received February, 1996; Revised March, 1996.



Communications in statistics

Theory and methods

Communications in statistics
Part A: theory and methods

Communications in Statistics is a multi-part journal.

Part A: Theory and Methods focuses primarily on new applications of known statistical methods to actual problems in industry and government, and has a strong mathematical orientation to statistical studies. In addition, **Part A** also offers communications that discuss practical problems with only ad hoc solutions or none at all; in either case, where there is a difference of opinion on particular techniques, all parties involved vigorously debate the issue with thought-provoking commentary.

For subscription information write the **Promotion Department:**

MARCEL DEKKER, INC.
270 MADISON AVENUE
NEW YORK, NEW YORK 10016

Residual variance estimation in moving average models

Raúl Pedro Mentz, Pedro A. Morettin and Clélia M.C. Toloi

University of Tucumán and CONICET, 4000 Tucumán, Argentina
University of São Paulo, 05315-970 - São Paulo, Brazil

Key words: bias; maximum likelihood; method of moments; time series.

Abstract

We consider time series models of the MA (moving average) family, and deal with the estimation of the residual variance. Results are known for maximum likelihood estimates under normality, both for known or unknown mean, in which case the asymptotic biases depend on the number of parameters (including the mean), and do not depend on the values of the parameters. For moment estimates the situation is different, because we find that the asymptotic biases depend on the values of the parameters, and become large as they approach the boundary of the region of invertibility. Our approach is to use Taylor series expansions, and the objective is to obtain asymptotic biases with error of $o(1/T)$, where T is the sample size. Simulation results are presented, and corrections for bias suggested.

1. Introduction

The moving average time series models of order q, denoted $MA(q)$, is defined by

$$X_t - \mu = \sum_{k=0}^{q} \alpha_k a_{t-k}, \quad t = \ldots, -1, 0, 1, \ldots, \tag{1.1}$$

where $\alpha_0 = 1$, X_t is the observable time series, a_t is a white noise residual with zero mean and constant variance σ^2, and the α_j's, μ and σ^2 are parameters, $0 < \sigma^2 < \infty$. We call σ^2 the residual variance of the process.

The process is stationary for any choice of parameters. If the roots of the associated polynomial equation $\sum_{k=0}^{q} \alpha_k w^{q-w} = 0$ satisfy $|w_k| < 1, k = 1, \ldots, q$, then (1.1) can be inverted into an infinite autoregression, $a_t = \sum_{s=0}^{\infty} \delta_s (X_{t-s} - \mu)$, $t = \ldots, -1, 0, 1, \ldots$, with $\delta_0 = 1$, where the δ_s are determined by the α_k.

The covariance sequence of the process is

$$\gamma_s = E(X_t - \mu)(X_{t+s} - \mu) = \sigma^2 \sum_{k=0}^{q-s} \alpha_k \alpha_{k+s} = \gamma_{-s}, \quad s = 0, 1, \ldots, q, \tag{1.2}$$

and equal to 0 if $|s| > q$, and the correlation sequence has

$$\rho_0 = 1, \quad \rho_s = \frac{\gamma_s}{\gamma_0} = \frac{\alpha_s + \alpha_1 \alpha_{s+1} + \cdots + \alpha_{q-s}\alpha_q}{1 + \alpha_1^2 + \cdots + \alpha_q^2} = \rho_{-s}, \quad s = 1, 2, \ldots, q. \tag{1.3}$$

The covariance and correlation sequences are two of the tools of the time domain analysis. The covariance sequence satisfies the inversion formula $\gamma_s = \int_{-\pi}^{\pi} e^{i\lambda s} f(\lambda) \mathrm{d}\lambda$ where $f(\lambda) = (\sigma^2/2\pi) \left| \sum_{k=0}^{q} \alpha_k e^{i\lambda k} \right|^2$, $-\pi < \lambda < \pi$, is the spectral density of the process, defined in the frequency domain.

We must specify the theoretical nature of the innovations a_t. One case is when they are i.i.d. normal with parameters 0 and σ^2. This assumption is used to define the likelihood function, and hence the maximum likelihood estimators. For purposes of comparison, we assume this property when defining and investigating other types of estimators; they, however, are sometimes studied under more general conditions, but this question will not be addressed here.

In this paper we consider the statistical problem of estimating σ^2. This is important because estimates of σ^2 (or of σ) enter, for example, in confidence

sets for the other parameters, in the estimation of the spectrum, and in expressions for the estimated prediction error. They are also used in criteria for order determination, like AIC or BIC.

For purposes of inference we consider a sample $X_1, X_2, \ldots, X_T$ from (1.1). Estimates of σ^2 come from the method of moments (MM), and maximum likelihood under normality (ML). Some authors have also considered least squares (LS) procedures, or deriving estimators from frequency domain arguments. In spite of their inferential role, not many papers have been written about determination of large-sample biases of residual variance estimators in MA models. Our purposes in this paper are to review the literature, and to present new material. There are more known results about AR models; some of these, and new results, are presented in Mentz, Morettin and Toloi (1995).

2. Review of the literature

The object of the inference may be taken to be the covariances or correlations introduced in (1.2) and (1.3), respectively. For these functions, large-sample expectations, variances, covariances and distributions are available, for several standard definitions of the sample quantities. This point will be briefly considered in Section 3.

When the object of the inference is taken to be the α_j and μ appearing in (1.1), early results available in the literature are surveyed in several sources. For example, the inefficiency of the moment estimator of α_j has been known for many years.

Tanaka (1984) suggests a technique for obtaining the Edgeworth-type asymptotic expansion associated with ML estimators in ARMA models. He obtained biases up to order $1/T$ for AR(1), AR(2), MA(1), MA(2) and ARMA(1,1) models, with and without constant terms. Biases for the residual variance estimators are also derived.

Cordeiro and Klein (1994) present a general procedure to obtain the biases of ML estimators in ARMA models. It turns out that their formula is difficult to obtain for models other than the lower order ones, but it can be easily computed numerically.

Other related papers are Davis (1977) and Porat and Friedlander (1986). Good references for techniques and results for asymptotic analysis in time series are Anderson (1971), and Fuller (1996).

3. Estimation of covariances and correlations

We consider estimating γ_j by

$$c_j = \frac{1}{T}\sum_{t=1}^{T-j}(X_t - \overline{X})(X_{t+j} - \overline{X}) = c_{-j}, \quad j = 0, 1, \ldots, T-1. \tag{3.1}$$

Other estimators are considered in the literature, for example, by changing in (3.1) the denominator, the value to be subtracted from the X's, or the range of the sums; see, for example, Anderson (1971, Chapter 8), or Fuller (1996, Chapter 6). We use (3.1) because for $T > p$, a covariance matrix with elements $c_{|i-j|}$ is positive definite, a fact that we shall use below.

With these estimators of the covariances, we form estimators of the correlations, $r_j = c_j/c_0$, $j = 1, 2, \ldots, T-1$.

From the above indicated sources, we deduce large-sample moments of these estimators. They are derived in the general case of a linear model $Y_t = \sum_{-\infty}^{\infty}\omega_j e_{t-j}$, under some conditions on the random variables e_t and coefficients ω_j. Our model (1.1) is a special case of such linear models. For example, when $q = 1$, for further reference we note that, with $r = r_1$,

$$\begin{aligned}
&E(c_0 - \gamma_0) \approx -\frac{\sigma^2}{T}(1+\alpha)^2, \quad E(c_1 - \gamma_1) \approx -\frac{\sigma^2}{T}(1+3\alpha+\alpha^2),\\
&E(c_0 - \gamma_0)^2 \approx \frac{2\sigma^4}{T}(1+4\alpha^2+\alpha^4), \quad E(c_1 - \gamma_1)^2 \approx \frac{\sigma^4}{T}(1+5\alpha^2+\alpha^4),\\
&E(c_0 - \gamma_0)(c_1 - \gamma_1) \approx \frac{4\sigma^4}{T}\alpha(1+\alpha^2),\\
&E(r-\rho) \approx -\frac{1}{T}\frac{1+4\alpha+\alpha^2+4\alpha^3+\alpha^4+4\alpha^5+\alpha^6}{(1+\alpha^2)^3},\\
&E(r-\rho)^2 \approx \frac{1}{T}\frac{1+\alpha^2+4\alpha^4+\alpha^6+\alpha^8}{(1+\alpha^2)^4},
\end{aligned} \tag{3.2}$$

where the expressions have errors $o(1/T)$. These results follow from general expressions in, for example, Anderson (1971) or Fuller (1996), and will be studied further in Section 4.2.3.

4. Parameter estimation by the method of moments

4.1. Estimation of the Coefficients

The parametric relation (1.3) between α_j's and ρ_j's can be inverted to express the former in terms of the latter, leading to a form of moment estimators of the coefficients. In the simple case of $q = 1$, $\rho_1 = \rho = \gamma_1/\gamma_0 = \alpha/(1+\alpha^2)$, and hence

$$\alpha = \frac{1-\sqrt{1-4\rho^2}}{2\rho}, \quad 0 < |\rho| < 0.5. \tag{4.1}$$

The relation for general q is more easily seen by factoring the spectral density function:

$$\frac{\sigma^2}{2\pi}\left|\sum_{k=0}^{q}\alpha_k e^{i\lambda k}\right|^2 = \frac{\sigma^2}{2\pi}\left(\sum_{k=0}^{q}\alpha_k e^{i\lambda k}\right)\left(\sum_{j=0}^{q}\alpha_j e^{-i\lambda j}\right) = \frac{\gamma_0}{2\pi}\sum_{s=-q}^{q}\rho_s e^{-i\lambda s}, \tag{4.2}$$

so that the moment estimators of the α_j are those satisfying

$$\frac{\left(\sum_{k=0}^{q}\hat{\alpha}_k z^k\right)\left(\sum_{j=0}^{q}\hat{\alpha}_j z^{-j}\right)}{1+\hat{\alpha}_1^2+\cdots\hat{\alpha}_q^2} = \sum_{s=-q}^{q} r_s z^{-s}. \tag{4.3}$$

Asymptotic biases of these estimators can be obtained as follows. Let $\dot{\alpha}_{jk} = \partial\alpha_j/\partial\rho_k$, $\ddot{\alpha}_{jks} = \partial^2\alpha_j/\partial\rho_k\partial\rho_s$, etc. Then

$$\begin{aligned}
E(\hat{\alpha}_j - \alpha_j) \approx &\sum_{k=1}^{q}\dot{\alpha}_{jk}E(r_k - \rho_k) + \frac{1}{2}\sum_{k=1}^{q}\ddot{\alpha}_{jkk}E(r_k-\rho_k)^2\\
&+\sum_{k=1}^{q}\sum_{s=1}^{q}\ddot{\alpha}_{jks}E(r_k-\rho_k)(r_s-\rho_s),
\end{aligned} \tag{4.4}$$

where this is valid with an error $o(1/T)$.

In the case of $q = 1$, $\partial\rho/\partial\alpha = (1-\alpha^2)/(1+\alpha^2)^2$, $\partial^2\rho/\partial\alpha^2 = -2\alpha(3-\alpha^2)/(1+\alpha^2)^3$, and hence

$$\dot{\alpha} = \frac{\partial\alpha}{\partial\rho} = \frac{(1+\alpha^2)^2}{1-\alpha^2}, \quad \ddot{\alpha} = \frac{\partial^2\alpha}{\partial\rho^2} = \frac{2\alpha(1+\alpha^2)^3(3-\alpha^2)}{(1-\alpha^2)^3}, \tag{4.5}$$

and using (3.2),

$$E(\hat{\alpha}-\alpha) = -\frac{1}{T}\frac{1+\alpha-2\alpha^2-7\alpha^3+2\alpha^4-4\alpha^5-2\alpha^6+\alpha^7+\alpha^8+\alpha^9}{(1-\alpha^2)^3} + o(1/T). \tag{4.6}$$

Similarly,

$$E(\hat{\alpha}-\alpha)^2 \approx \mathrm{Var}(\hat{\alpha}) \approx \frac{(1+\alpha^2)^4}{(1-\alpha^2)^2}E(r-\rho)^2 \approx \frac{1}{T}\frac{1+\alpha^2+4\alpha^4+\alpha^6+\alpha^8}{(1-\alpha^2)^2}, \quad (4.7)$$

a known result, see, for example, Fuller (1996, (8.3.3)). These approximations have errors $o(1/T)$. See Section 4.2.3 below.

4.2. Estimation of the Residual Variance

A moment estimator of σ^2 can be deduced by making $s = 0$ in (1.3), that is,

$$\hat{\sigma}^2_{MM} = \frac{c_0}{1+\hat{\alpha}_1^2+\cdots+\hat{\alpha}_q^2}, \quad (4.8)$$

where the $\hat{\alpha}_j$ are those defined in Section 4.1.

To study the asymptotic bias, we derive an expansion valid for arbitrary $q \geq 1$, and then evaluate this explicity for the special case of $q = 1$.

4.2.1. An Expansion for the MS(q) Model

Applying to (4.8) a Taylor expansion that retains terms contributing up to second order, and expanding again terms involving $\hat{\alpha}_j - \alpha_j$ as functions of $r_k - \rho_k$, as we did to obtain (4.4), we prove that

$$\begin{aligned} E(\hat{\sigma}^2_{MM}-\sigma^2) &= \frac{\sigma^2}{\gamma_0}E(c_0-\gamma_0)+\frac{2\sigma^4}{\gamma_0^4}A_1E(c_0-\gamma_0)^2 \\ &-\frac{2\sigma^4}{\gamma_0^3}\sum_{k=1}^{q}A_{2k}E(c_0-\gamma_0)(c_k-\gamma_k)-\frac{2\sigma^4}{\gamma_0}\sum_{k=1}^{q}A_{2k}E(r_k-\rho_k) \\ &+\sum_{k=1}^{q}A_{3k}E(r_k-\rho_k)^2+\sum_{\substack{k=1\\k>s}}^{q}\sum_{s=1}^{q}A_{4ks}E(r_k-\rho_k)(r_s-\rho_s)+o(1/T). \quad (4.9) \end{aligned}$$

where

$$A_1 = \sum_{j=1}^{q}\sum_{k=1}^{q}\alpha_j\gamma_k\dot{\alpha}_{jk}, \qquad A_{2k} = \sum_{j=1}^{q}\alpha_j\dot{\alpha}_{jk}, \quad (4.10)$$

$$A_{3k} = \sum_{j=1}^{q}\left[-\frac{\sigma^4}{\gamma_0}\alpha_j\ddot{\alpha}_{jkk}-\frac{\sigma^4}{\gamma_0}\dot{\alpha}_{jk}^2+\frac{4\sigma^6}{\gamma_0^2}\alpha_j^2\dot{\alpha}_{jk}^2\right]+\frac{8\sigma^6}{\gamma_0^2}\sum_{\substack{j=1\\j>i}}^{q}\sum_{i=1}^{q}\alpha_j\alpha_i\dot{\alpha}_{jk}\dot{\alpha}_{ik}.$$

$$A_{4ks} = \sum_{j=1}^{q}\left[-\frac{2\sigma^4}{\gamma_0}\alpha_j\ddot{\alpha}_{jks}-\frac{2\sigma^4}{\gamma_0}\dot{\alpha}_{jk}\dot{\alpha}_{js}+\frac{8\sigma^6}{\gamma_0^2}\alpha_j^2\dot{\alpha}_{jk}\dot{\alpha}_{js}\right]+\frac{16\sigma^6}{\gamma_0^2}\sum_{\substack{j=1\\j>i}}^{q}\sum_{i=1}^{q}\alpha_j\alpha_i\dot{\alpha}_{jk}\dot{\alpha}_{is}.$$

4.2.2. The Special Case of MA(1)

When $q = 1$, (4.10) becomes

$$A_1 = \sigma^2\alpha^2\frac{(1+\alpha^2)^2}{1-\alpha^2}, \quad A_{2k} = \alpha\frac{(1+\alpha^2)^2}{1-\alpha^2}, \quad A_{3k} = -\sigma^2\frac{(1+\alpha^2)^4}{(1-\alpha^2)^3} \quad (4.11)$$

while the necessary asymptotic expectations are given in (3.2). Substitution leads to

$$E(\hat{\sigma}^2_{MM}-\sigma^2) = -\frac{\sigma^2}{T}\frac{2-6\alpha^2-2\alpha^3+15\alpha^4+4\alpha^5-4\alpha^6-2\alpha^7+\alpha^8}{(1-\alpha^2)^3}+o(1/T). \quad (4.12)$$

In this special case we can check (4.12) by using directly an expansion of its left-hand side stemming from (4.8), plus $E(c_0-\gamma_0)$, $E(\hat{\alpha}-\alpha)$, $E(\hat{\alpha}-\alpha)^2$ given, respectively, in (3.2), (4.6), (4.7), and $E(c_0-\gamma_0)(\hat{\alpha}-\alpha) \approx (\sigma^2/T)[2\alpha(1+\alpha^4)]/(1-\alpha^2)$.

In figure 1 we graph T/σ^2 times (4.12). The graph is flat, at the level of -2, for α small in absolute value: -2 is the value corresponding to $\alpha = 0$, and in Section 5 we will find its connection with ML estimation. As α increases in absolute value(i.e. tends to leave the region of invertibility) the values increase substantially their negative values, since in fact, for fixed α and T, (4.12) approaches $-\infty$ as $|\alpha|$ approaches 1.

The depicted values, times σ^2/T, are the negative biases in the estimation of the residual variance by the MM, as a function of the parameter α. Hence, the proposed estimator tends to lead to values that, in the average and approximately, are smaller than desired. A consequence will be that confidence intervals with scale parameter estimated by MM will tend to be unduly optimistic, that is, short. Further, this defect tends to become more important for values of α approaching 1 in absolute value.

Note, however, that these biases in the estimation of the residual variance must be divided by the sample size T, so that, as expected, larger sample sizes will lead to better results.

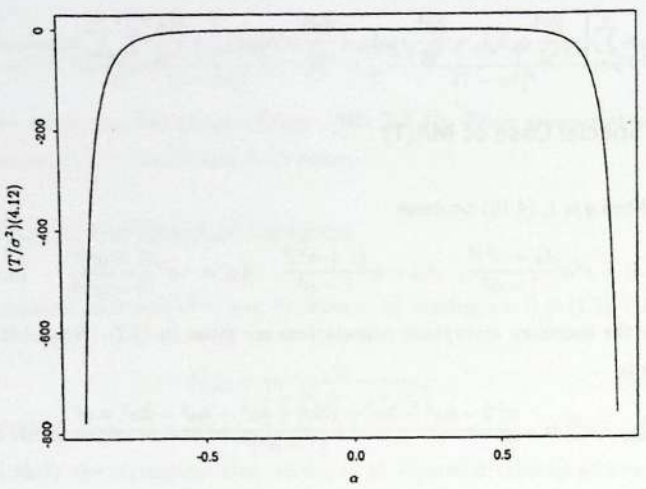


FIGURE 1: Representation of T/σ^2 times (4.12).

4.2.3. A Simulation Study for the MA(1) Model

Expression (4.12) is to be added to comparable and previously known asymptotic biases for the MA(1) model: for moment estimators, the next to last line of (3.2) corresponds to r and (4.6) to α; the first line of (3.2) has the asymptotic biases of the usual estimators of γ_0 and γ_1. A graphical analysis of these functions shows, in general, dependence on the values of α: the α-effect illustrated in Figure 1 for (4.12), tends to hold for several of the other expressions, and anticipates difficulties near the boundary of the region of invertibility.

To study the behaviour of these estimation procedures, a simulation study was performed. Two of the main purposes of such a study were to investigate the sample sizes needed for reasonable fits of the asymptotic theory to the empirical results, and to relate the analysis to values of α.

In the application of these procedures we found that the sample correlation r not always satisfies $|r| < 0.50$ as suggested by the parametric relation (4.1), so that the corresponding estimator of α is not real. For small values of α, no value of r was found to violate this restriction, but as we considered larger values of α, replications with $|r| > 0.5$ were found. The following definition was introduced: use the sample analog of (4.1) only for $0 < |r| \leq 0.50$, and then set $\hat{\alpha} = -1$ if $r < -0.5$, $\hat{\alpha} = 1$ if $r > 0.5$, and $\hat{\alpha} = 0$ if $r = 0$ (Fuller, 1996, Section 8.3). The problem of noninvertibility of sample estimators in MA models has been studied in other contexts, for example, in ML estimation under normality: see, for example, Cryer and Ledolter (1981), Pesaran (1983), Anderson and Takemura (1986).

The results of the simulations are reported as follows. Besides the estimation of the residual variance σ^2 by (4.8), we also consider the estimation of $\gamma_0 = \sigma^2(1+\sigma^2)$, $\gamma_1 = \sigma^2\alpha$, $\rho = \gamma_1/\gamma_0$, and α, with estimators introduced in Section 3, and (4.3) (that for $q = 1$ leads to the sample analog of (4.1)), respectively. The objective of this design is to facilitate the comparison of (4.12) with existing similar expressions. We simulated T values corresponding to (1.1), with the a_t pseudorandom, independent normal (0,1).

Several values of α ranging from 0.2 to 0.9 were considered and various sample sizes starting at $T = 50$. To simplify, and to stress the nature of our findings, we restrict our presentation to $\alpha = 0.4$, 0.6 and 0.8, and $T = 400$, 600, 800 and 1,000. For each choice of α and T, 1,000 replications were done.

Table I summarizes the results of the simulation study, by presenting the averages over the 1,000 replications, of the estimates obtained by using the indicated procedures. The signs of the biases correspond to those of the theoretical expressions, except for the small variability of $\hat{\alpha}$ around 0.40 for $\alpha = 0.40$. The last line contains the numbers of cases where $r > 0.50$, and we see that occurences are frequent for $\alpha = 0.80$. This tends to raise the values of $\hat{\alpha}$, and hence to lower those of $\hat{\sigma}^2$.

Table II presents an inferential analysis of the results of the simulations done for $\alpha = 0.40$ and $T = 400$. For this combination of values (as well as for smaller $|\alpha|$ or for larger T) no case of $r > 0.5$ was observed. Column 2

TABLE I: Average estimates over 1,000 replications, usual estimators of γ_0, γ_1, MM estimators of ρ, α, σ^2.

Parameter to be estimated	$\alpha = 0.40$ Parameter value	Sample sizes 400	600	800	1,000
γ_0	1.160	1.154	1.156	1.160	1.160
γ_1	0.400	0.395	0.397	0.401	0.398
ρ	0.345	0.341	0.343	0.345	0.343
α	0.400	0.400	0.401	0.402	0.399
σ^2	1.000	0.992	0.994	0.997	0.999
$\lvert r\rvert > 0.5$		(0)	(0)	(0)	(0)

Parameter to be estimated	$\alpha = 0.60$ Parameter value	Sample sizes 400	600	800	1.000
γ_0	1.360	1.354	1.357	1.357	1.359
γ_1	0.600	0.595	0.598	0.598	0.598
ρ	0.441	0.437	0.440	0.440	0.440
α	0.600	0.618	0.614	0.612	0.606
σ^2	1.000	0.978	0.984	0.986	0.993
$\lvert r\rvert > 0.5$		(52)	(19)	(9)	(7)

Parameter to be estimated	$\alpha = 0.80$ Parameter value	Sample sizes 400	600	800	1,000
γ_0	1.640	1.631	1.637	1.637	1.638
γ_1	0.800	0.792	0.797	0.796	0.799
ρ	0.488	0.476	0.481	0.481	0.483
α	0.800	0.800	0.816	0.815	0.823
σ^2	1.000	0.998	0.986	0.987	0.980
$\lvert r\rvert > 0.5$		(340)	(315)	(287)	(283)

TABLE II: Inferential analysis of simulations, usual estimators of γ_0, γ_1, MM estimators of ρ, α, σ^2. $\alpha = 0.40$, $T = 400$

1. Parameter to be estimated	2. Average estimate over 1,000 replications	3. Empirical bias	4. Asymptotic bias	5. Empirical standard error	6. Studentized bias
γ_0	1.1545	-0.0055	-0.0049	0.0045	-0.1438
γ_1	0.3953	-0.0047	-0.0059	0.0034	0.3437
ρ	0.3410	-0.0038	-0.0077	0.0022	1.7847
α	0.3998	-0.0002	-0.0027	0.0036	0.6962
σ^2	0.9919	-0.0081	-0.0056	0.0036	-0.6954

contains the averages over replications (also reported in Table I), and column 3 the corresponding estimated biases; these should be compared with theoretical values in column 4. To judge the significance of the differences between columns 3 and 4, column 5 contains the empirical standard errors, from which the studentized differences are formed in column 6.

We observe the following: (1) Columns 3 and 4 visually show good concordance; (2) The studentized differences are small for all rows. We conclude that for $\alpha = 0.40$, $T = 400$ is large enough to lead to a good fit of the asymptotic theory to the simulated results.

Tables similar to Table II were constructed for each combination of α and T in the indicated ranges. The studentized differences of the biases in these tables are presented in Table III. We observe the following: (1) When the estimation of γ_0 and γ_1 is considered, all values are small, with a maximum of 2.07 for $\alpha = 0.40$ and $T = 800$; (2) Most values corresponding to the estimation of ρ are large, and increase as α gets larger; (3) Studentized biases in the estimation of α are smaller than 3 for $\alpha = 0.40$ and 0.60, and very large negative for $\alpha = 0.80$; (4) Studentized biases in the estimation of σ^2 are smaller than 3 (all but one are smaller than 2) for $\alpha = 0.40$ and 0.60, and very large and positive for $\alpha = 0.80$.

TABLE III: Studentized biases, usual estimators of γ_0, γ_1, MM estimators of ρ, α, σ^2.

Parameter to be estimated	$\alpha = 0.40$ Sample sizes				$\alpha = 0.60$ Sample sizes				$\alpha = 0.80$ Sample sizes			
	400	600	800	1000	400	600	800	1000	400	600	800	1000
γ_0	-0.14	-0.26	1.15	1.07	0.10	0.31	-0.08	0.53	-0.06	0.43	0.17	0.46
γ_1	0.34	0.59	2.07	0.59	0.55	1.07	1.04	0.89	0.31	1.12	0.23	1.53
ρ	1.78	2.21	3.45	1.39	4.53	5.70	5.77	5.19	7.23	8.78	6.76	8.35
α	0.70	1.08	2.16	0.24	1.62	2.18	2.55	1.38	-22.64	-18.30	-16.07	-12.60
σ^2	-0.70	-1.08	-0.14	0.93	-1.45	-1.79	-2.62	-0.87	23.36	19.21	17.46	13.59

Our conclusions are: (1) Moment estimators of ρ, α and σ^2 have biases, both asymptotic and for finite samples, that depend on the true value of α that generated the MA(1) series; (2) For a sample size of $T = 400$ or larger, the finite-sample biases are well predicted by the asymptotic theory if $\alpha = 0.40$ or smaller; (3) For $\alpha = 0.80$, even for T as large as 1,000, the finite-sample biases are not well predicted by the asymptotic theory: in some cases the approximations tend to improve as sample size increases, but clearly larger values of T will be needed to observe good fits; (4) For values of α intermediate between 0.40 and 0.80, and values of T intermediate between 400 and 1,000, results of the simulations tend to fill in the trends suggested by the extreme values chosen for the simulations; (5) These situations contrast with the analysis of the estimation of the covariances by the standard sample quantities, since for them the simulated biases are in agreement with those predicted by the asymptotic theory, for all chosen values of α and T.

In terms of the method of moments estimation of the residual variance in MA(1) models, if α is close to the boundary of the region of invertibility, very large sample sizes will be required to make the asymptotic theory for its bias useful in practice. This contrasts with the analysis for the method of ML, to be considered below. Users of the "preliminary estimators" of the residual variance given by the method of moments, should be aware of the difficulties in predicting the biases associated with such a choice, and the differences that may exist with

ML estimation. The situation is similar when MM estimation of ρ and α is considered, a fact that has not been sufficiently stressed in the literature, even when asymptotic results are known. On the other side, the situation differs markedly from that of estimating the covariances of the process, for which the asymptotic results are useful for all α and sample sizes. An important role in these findings is played by the fact that the autocorrelation estimator r tends to give values $r > 0.50$ with high frequency, when $\alpha > 0$ is large.

5. Parameter estimation by Maximun Likelihood

When the residuals in (1.1) are independent $N(0, \sigma^2)$, the likelihood function of the observations $X_1, X_2, \ldots, X_T$ is the function of μ, σ and the α_j's given by

$$L = (2\pi)^{-\frac{1}{2}T}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{X}-\mu)'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\mu)\right\}, \tag{5.1}$$

where $\mathbf{X} = (X_1, \ldots, X_T)'$, $\mu = (\mu, \ldots, \mu)'$, and $\boldsymbol{\Sigma}$ is the covariance matrix with components $\gamma_{|i-j|}$. This can also be written as

$$L = (2\pi\sigma^2)^{-\frac{1}{2}T}|\mathbf{P}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{X}-\mu)'\mathbf{P}^{-1}(\mathbf{X}-\mu)\right\}, \tag{5.2}$$

where $\boldsymbol{\Sigma} = \sigma^2\mathbf{P}$. Maximum likelihood estimators of μ, $\alpha_1, \ldots, \alpha_q$ and σ^2 are obtained by maximizing (5.2) over its parameter space. No explicit formulas for these estimators are known, not even in the case of $q = 1$. However, asymptotic expressions for some of the biases are known: Tanaka (1984) and Cordeiro and Klein (1994) give, for the MA(1) model with $\alpha_1 = \alpha$,

$$E(\hat{\alpha} - \alpha) = \frac{1+2\alpha}{T} + o(1/T). \tag{5.3}$$

Maximizing (5.2) with respect to σ^2 only, shows that

$$\hat{\sigma}^2_{ML} = \frac{1}{T}(\mathbf{X}-\hat{\mu})'\hat{\mathbf{P}}^{-1}(\mathbf{X}-\hat{\mu}). \tag{5.4}$$

For this estimator, Tanaka (1984) and Cordeiro and Klein (1994) give, for the MA(1) model

$$E(\hat{\sigma}^2_{ML} - \sigma^2) = -\frac{2\sigma^2}{T} + o(1/T). \tag{5.5}$$

5.1. An Approximated Closed-form Expression for in MA(1)

We operate with (5.4). If $\mathbf{G}$ denotes the $T \times T$ matrix that has components equal to 1 in its two diagonals adjacent to the main diagonal, and 0 elsewhere, then $\mathbf{P} = (1+\alpha^2)\mathbf{I} + \alpha\mathbf{G} = (1+\alpha^2)[\mathbf{I}+\rho\mathbf{G}]$, and $\mathbf{P}^{-1} = (1+\alpha^2)^{-1}[\mathbf{I}+\rho\mathbf{G}]^{-1}$, so that it suffices to invert the $\mathbf{I}+\rho\mathbf{G}$ matrix. This matrix and its inverse are symmetric Toplitz matrices, and the inverse was studied in detail in Mentz (1976) and Shaman (1969). In the former the components of $[\mathbf{I}+\rho\mathbf{G}]^{-1}$ are given as

$$\frac{\left(1-x_1^{2T-2j+2}\right)\left(x_1^{j+i+1}-x_1^{j-i+1}\right)}{\rho\left(1-x_1^2\right)\left(1-x_1^{2T+2}\right)} \tag{5.6}$$

where x_1 is a root of the associated polynomial equation $\rho x^2 + x + \rho = 0$, and hence $x_1 = -\alpha$.

In Shaman (1969) it is suggested that components of the inverse matrix can be approximated by $x_1^k(1-4\rho^2)^{-\frac{1}{2}}$ that in turn, substituting $\rho = \alpha(1+\alpha^2)^{-1}$, can be expressed as $(-\alpha)^k(1+\alpha^2)(1-\alpha^2)^{-1}$, $k = 0, 1, \ldots, T-1$. The effect of this approximation is as follows: $\mathbf{P}$ times a matrix with these components, produces a matrix that differs from the identity matrix in the components of the first row and column, and in the component in row T, column T.

Denoting by $p^{ij} = p^{|i-j|}$ the components of $\mathbf{P}^{-1}$, $p^k \approx (-\alpha)^k(1-\alpha^2)^{-1}$, and hence (5.4) becomes

$$\hat{\sigma}^2_{ML} = \sum_{i=1}^{T}\sum_{j=1}^{T} p^{ij}\frac{1}{T}(X_i-\hat{\mu})(X_j-\hat{\mu}) \approx \frac{c_0 + 2\sum_{j=1}^{T-1}(-\hat{\alpha})^j c_j}{1-\hat{\alpha}^2}. \tag{5.7}$$

Hence, the maximum likelihood estimator of σ^2 has an approximate representation as a weighted sum of c_j's, weights involving powers of $-\hat{\alpha}$, where $\hat{\alpha}$ is the ML estimator of α.

It can be shown that we can arrive at this expression (with the corresponding least squares estimator of α) by starting from a definition of σ^2 in terms of a residual sum of squares (see, for example, Brockwell and Davis, 1991a), and using a "long autoregression" approximation to be able to write the defining expression in a regression type of format (see the technical report by the

authors, 1995). The fact that the LS and ML estimators, under some approximations, coincide, has been noted elsewhere. Brockwell and Davis (1991a, Section 8.7) note that if the determinant of the likelihood function is asymptotically negligible compared with the sum of squares in the exponent, as in the case when the parameter is constrained to be invertible, then minimization of the sum of squares will be equivalent to minimization of the likelihood and the least squares and maximum likelihood estimators will have similar asymptotic properties.

For comparison, the MM estimator can be expressed in terms of powers of $-\hat{\alpha}^2$ as follows,

$$\hat{\sigma}^2_{MM} = \frac{c_0}{1+\hat{\alpha}^2} \approx c_0 + \sum_{j=1}^{T-1} c_0(-\hat{\alpha}^2)^j, \tag{5.8}$$

where now $\hat{\alpha}$ is the MM estimator of α.

5.2. Simulations for ML Estimation of the Residual Variance

To illustrate the biases in the ML estimation of the residual variance of a MA(1) model, a small simulation study was performed. The values $\alpha = -0.40$, 0.20 and 0.80 were used to generate the simulated observations, and sample sizes $T = 50$, 100, 200 and 400 were considered. Only 100 replications were done for each pair of α and T. The ML procedure in Brockwell and Davis (1991b) was used, in the understanding that it represents accurately the theory presented in this section, and further that the procedure is one that is used frequently in empirical studies. The main results are collected in Table IV.

In column 3 there are some sign changes in relation to what is expected. However, column 6 of studentized biases contains only one value as large as 3, which occurs for the smallest sample size $T = 50$. The assertion that the biases do not depend upon the values of α is well supported by these figures.

In spite of the small number of replications, the differences between this study for ML estimation and that in Section 4.2.3 for the MM are clear. The asymptotic theory is seen to be a valid approximation for a wider range of values of the basic parameter σ, and for smaller sample sizes.

TABLE IV: Inferential analysis of simulations, ML estimators of σ^2.

1. Value of α	2. Sample size	3. Empirical bias	4. Asymptotic bias	5. Empirical standard error	6. Studentized bias
-0.4	50	-0.022	-0.040	0.019	0.947
-0.4	100	-0.002	-0.020	0.013	1.385
-0.4	200	-0.012	-0.010	0.009	-0.222
-0.4	400	-0.004	-0.005	0.006	0.167
0.2	50	0.014	-0.040	0.018	3.000
0.2	100	0.007	-0.020	0.013	2.077
0.2	200	-0.007	-0.010	0.010	0.300
0.2	400	-0.009	-0.005	0.006	-0.667
0.8	50	-0.004	-0.040	0.022	1.636
0.8	100	0.013	-0.020	0.016	2.063
0.8	200	0.007	-0.010	0.011	1.545
0.8	400	0.007	-0.005	0.007	1.714

6. Concluding remarks

We considered the estimation of the variance of the white noise component of the MA(q) model defined by (1.1). This is a nuisance parameter, and is important because estimates enter into prediction and confidence intervals, tests of hypotheses, spectral estimates, and other inferential procedures.

In spite of the indicated usefulness, not many results are available about properties of estimators of the residual variance in MA models, except for maximum likelihood estimators under normality. The situation is better for AR models, as indicated in Mentz, Morettin and Toloi (1995).

Available results for ML estimates under normality can be interpreted as follows: with $X_1, X_2, \ldots, X_T$ i.i.d. $N(\mu, \sigma^2)$, the ML estimator of σ^2 is $\hat{\sigma}^2 = \sum_{i=1}^{T}(X_i - \overline{X})^2/T$, and it is biased, it underestimates σ^2, $E(\hat{\sigma}^2 - \sigma^2) = -\sigma^2/T$, the reason being that the denominator T is too large. Intervals formed with $\hat{\sigma}$ tend to be too short, too optimistic. Note, however, that the bias is $O(1/T)$. In MA(q) models, the ML estimators under normality of $\sigma^2 = \mathrm{Var}(a_t)$, satisfy $E(\hat{\sigma}_{ML} - \sigma^2) \approx -k\sigma^2/T$, where k is the number of parameters in the model,

including $\mu = E(X_t)$; hence, $k = 2$ for the MA(1) model, $k = 3$ for the MA(2) models, etc., when μ is unknown.

In the analytical part of our study we concentrate in the study of the asymptotic biases of the estimators by the method of moments (MM). We rely on Taylor-type expansions, and search for results to $O(1/T)$, where T is sample size. Incidentally, for simplicity we also tried to rely on first-order Taylor expansions, but found that for MA models they can be misleading, in that for certain ranges of the parameter space, they gave results with the wrong sign.

Section 4.2.1 contains an expansion valid for $q \geq 1$. However, our explicit results, analyses and simulations are for $q = 1$.

For the MA(1) model, Figure 1 shows the behaviour of the asymptotic bias for the MM estimator, and it can be compared with the corresponding value for the ML estimator, namely -2 for all values of α: (1) $\hat{\sigma}^2_{MM}$ underestimates σ^2 for all α; (2) For approximately $|\alpha| \leq 0.5$, $\hat{\sigma}^2_{MM}$ and $\hat{\sigma}^2_{ML}$ have approximately the same (asymptotic) bias; (3) For approximately $|\alpha| > 0.5$, the negative bias of $\hat{\sigma}^2_{MM}$ is larger than that of $\hat{\sigma}^2_{ML}$; (4) The asymptotic bias of $\hat{\sigma}^2_{MM}$ tends to $-\infty$ as $|\alpha|$ tends to 1.

Our simulations tend to confirm some of our expectations, for example, that the ML estimator works better than the MM, or that better fits are obtained for larger sample sizes. The fit of the simulated results to the asymptotic theory is much better for ML than for MM procedures.

Correction for bias when using ML estimators of the residual variance is simple, since it does not depend on the parameter values. A word of caution should be expressed when MM estimators are used for MA(1) models whose parameters tend to be near the region of invertibility, since for the usual sample sizes, we expect to find difficulties in using the available asymptotic reults for the biases.

7. Acknowledgements

We express our appreciation to a referee for comments that led to a substantial improvement of the paper, to Professor Carlos E. Harle (São Paulo)

for his help with some mathematical problems, and to Eng. Carlos I. Martinez (Tucumán) for his help with the computations. Interchange of visits by the authors were facilitated by a research subsidy from "Fundación Antorchas".

8. Bibliography

Anderson, T.W. (1971). *The Statistical Analysis of Time Series.* New York: John Wiley and Sons.

Anderson, T.W. and Takemura, A. (1986). Why do noninvertible, estimated moving averages occur? *J. of Time Series Analysis*, **7**, 235–254.

Brockwell, P.J. and Davis, R.A. (1991a). *Time Series: Theory and Methods.* New York: Springer Verlag.

—— (1991b). *ITSM: An Interactive Time Series Modelling Package for the PC.* New York: Springer Verlag.

Cordeiro, G.M. and Klein, R. (1994). Bias correction of maximum likelihood estimates for ARMA models. *Stat. and Prob. Letters*, 169–176.

Cryer, J.D. and Ledolter, J. (1981). Small-sample properties of the maximum likelihood estimator in the first-order moving average model. *Biometrika*, **68**, 691–694.

Davis, W.W. (1977). Robust interval estimation of the innovation variance of an ARMA model. *The Annals of Statistics*, **5**, 700–708.

Fuller, W.A. (1996). *Introduction to Statistical Time Series.* New York: John Wiley and Sons.

Mentz, R.P. (1976). On the inverse of some covariance matrices of Toeplitz type. *SIAM J. Appl. Math.*, **31**, 326–337.

Mentz, R.P., Morettin, P.A. and Toloi, C.M.C. (1995). On residual variance estimation in autoregressive models. Forthcoming.

—— (1995). Residual variance estimation in moving average models. Cuaderno 68, INIE, U. Nacional de Tucumán.

Pesaran, M.H. (1983). A note on the maximum likelihood estimation of regression models with first order moving average errors with roots on the unit circle. *Australian J. of Stat.*, **25**, 442–448.

Porat, B. and Friedlander, B. (1986). On the estimation of variance for autoregressive and moving average processes. *IEEE Trans. on Inf. Theory*, IT-32, No.1, 120–125.

Priestley, M.B. (1981). *Spectral Analysis and Time Series.* London: Academic Press.

Shaman, P. (1969). On the inverse of the covariance matrix of a first-order moving average. *Biometrika*, **60**, 193–196.

Tanaka, K. (1984). An asymptotic expansion associated with the maximum likelihood estimator in ARMA models. *Journal of the Royal Stat. Soc.*, Series B, **46**, 58–67.

Received August, 1995; Revised April, 1997

Reprinted from

COMPUTATIONAL STATISTICS & DATA ANALYSIS

On least-squares estimation of the residual variance in the first-order moving average model

R. P. Mentz, P. A. Morettin, C. M. C. Toloi

University of Tucumán and CONICET, Argentina
University of São Paulo, Brazil

Received 1 November 1997, revised 1 October 1998

COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Instructions to Authors

A. Journal Portion (Sections I-III)

Four copies of contributions (including one original of the figures) should be submitted to the appropriate Co-Editor. Manuscripts from Europe, Asia and Africa should be sent to Peter Naeve, Ph.D., Department of Statistics and Computing, University of Bielefeld, Post Box 100131, D-33501 Bielefeld, Germany. Manuscripts from North-, Central and South America, as well as the rest of the world should be sent to Stanley Azen, Ph.D., Department of Preventive Medicine, University of Southern California, 1540 Alcazar St., CHP 218, Los Angeles, CA 90033, USA. Although the submission of papers is regionally divided, the reviewing process is standardized between the CoEditors.

Submission of accepted papers as *electronic manuscripts*, i.e., on disk with accompanying manuscipt, is encouraged. Electronic manuscripts have the advantage that there is no need for rekeying of text, thereby avoiding the possibility of introducing errors and resulting in reliable and fast delivery of proofs. The preferred storage medium is a 5.25 or 3.5 inch disk in MS-DOS format, although other systems are welcome, e.g., Macintosh (in this case, save your file in the usual manner; do not use the option "save in MS-DOS format"). Do not submit your original paper as electronic manuscript but hold on the disk until asked for this by the Editor (in case your paper is accpeted without revisions). Do submit the accepted version of your paper as electronic manuscript. Make absolutely sure that the file on the disk and the printout are identical. Please use a new and correctly formatted disk and label this with your name; also specify the software and hardware used as well as the title of the file to be processed. Do not convert the file to plain ASCII. Ensure that the letter "l" and digit "1", also the letter "O" and digit "0" are used properly, and format your article (tabs, indents, etc.) consistently. Characters not available on your word processor (Greek letters, mathematical symbols, etc.) should not be left open but indicated by a unique code (e.g., gralpha, ⟨alpha⟩, @, etc., for the Greek letter α). Such codes should be used consistently throughout the entire text; a list of codes used should accompany the electronic manuscript. Do not allow your word processor to introduce word breaks and do not use a justified layout. Please adhere strictly to the general instructions below on style, arrangement and, in particular, the reference style of the journal.

1. The focus of the papers submitted to CSDA *must* include either a computational or data analysis component. Papers which are purely theoretical are not appropriate for CSDA, and will be returned to the authors. Whenever appropriate the manuscript should present an illustrative example or application.

2. Manuscripts describing simulation studies must a) be thorough with regard to the choice of parameter settings, b) not overgeneralize the conclusions, c) carefully describe the limitations of the simulations studies, d) and should guide the user regarding when the recommended methods are appropriate. In addition, it is recommended that the author(s) indicate why comparisons cannot be made theoretically and why therefore simulations are necessary.

3. Papers reporting results based on computations should provide enough information so that readers can evaluate the quality of the results, as well as desciptions of pseudo-random-number generators, numerical algorithms, computer(s), programming language(s), and major software components that were used.

4. In some cases, articles submitted to CSDA may be more appropriate for the SSN. In this case, a recommendation may be made by the Associate Editor or Co-Editor that a modified (e.g., revised and shortened) version of the manuscript should be submitted to the SSN. The author of course has the right to decline this recommendation.

5. All contributions should be written in English and contain an abstract of approximately 200 words plus a list of key words.

6. The manuscript must be typed on one side of the paper in double spacing (including footnotes, references and abstracts) with wide margins.

7. All mathematical symbols which are not typewritten should be listed and explained on a separate sheet. The numbers identifying displayed mathematical expressions should be placed in parentheses at the right margin. Parts of text should not be subject to this numbering.

8. Special care should be given to the preparation of the drawings for the figures and diagrams. Except for a reduction in size, they will appear in the final printing in exactly the same form as submitted by the author; normally they will not be redrawn by the printer. In order to make a photographic reproduction possible, all drawings should be on separate sheets, with wide margins, drawn large size, in India ink, and carefully lettered. Exceptions are diagrams only containing formulae and a small number of straight lines (or arrows): these can be typeset by the printer.

9. References should be listed alphabetically in the same way as the following examples:

Rao, C.R., *Linear statistical inference and its applications* (Wiley, New York, 1973) 14–32.

Bock, R.D. and D. Brandt, Comparison of some computer programs for univariate and multivariate analysis of variance. in: P.R. Krishnaiah (Ed.), *Handbook of statistics*, Vol. 1 (North-Holland, Amsterdam, 1980) 703–744.

Ghosh, S., Robustness of BIBD against the unavailability of data, *J. Statist. Plann. Inference*, 6(1) (Dec. 1982) 29–32.

Pillai, K.C.S. and B. Saweris, Asymptotic formulae for the distribution of Hotelling's trace for equality for two covariance matrices, Mimeo. (Dept. of Statistics, Purdue University, Lafayette, IN, 1983).

Submission of a paper will be held to imply that it contains original unpublished work and is not being submitted for publication elsewhere.

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher. This tansfer will ensure the widest possible dissemination of information. Please ensure that the paper is submitted in its final form. Correction in the proof stage, other than printer's errors, should be avoided; cost arising from such extra corrections will be charged to the authors. Otherwise, no page charges will be made. A reprint order form will be sent to the principal author together with the proofs. Twenty-five reprints of each contribution to Sections I-III are available free of charge. Additional reprints can be ordered from the Publisher.

B. Statistical Software Newsletter (Section IV)

Three copies of informational articles and news items (including one original of the figures) should be submitted to one of the SSN Editors: Allmut Hörmann (Managing Editor), GSF – Forschungszentrum für Umwelt und Gesundheit, Institut für Medizinische Informatik und Systemforschung (Medis), Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany; Joyce Niland, Ph.D., Department of Biostatistics, City of Hope National Medical Centre, 1500 E. Duarte Road, Duarte, CA 91010, USA; Andrew Westlake, Department of Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

Books and other relevant literature for consideration should be submitted to the Book Review Editor, Walter Krämer, Lehrstuhl für Wirtschafts- und Sozialstatistik, FB Statistik der Universität Dortmund, Postfach 500, D-44221 Dortmund, Germany.

1. All contributions must be in English. Informational articles must follow the specifications for Sections I-III.

2. In order to expedite publication, the editors would welcome material on diskette, e.g. texts favouritely as Word for Windows files. Word for MS-DOS, Word for Apple Macintosh (only DOS formatted), WordPerfect, and ASCII files will always be accepted. Additional specific details can be obtained in advance from the Managing Editor.

3. If figures, tables, and computer output cannot be included on the diskette, material must be presented in a format to permit direct reproduction.

4. The Managing Editor will have the final responsibility for each issue of the SSN.

ELSEVIER Computational Statistics & Data Analysis 29 (1999) 485–499

COMPUTATIONAL STATISTICS & DATA ANALYSIS

On least-squares estimation of the residual variance in the first-order moving average model

R. P. Mentz, P. A. Morettin, C. M. C. Toloi

University of Tucumán and CONICET, Argentina
University of São Paulo, Brazil

Abstract

In the first-order moving average model we analyze the behavior of the estimator of the variance of the random residual coming from the method of least squares. This procedure is incorporated into some widely used computer programs. We show through simulations that the asymptotic formulas for the bias and variance of the maximum likelihood estimator, can be used as approximations for the least-squares estimator, at least when the model parameter is far from the region of non-invertibility. Asymptotic results are developed using the "long autoregression" idea, and this leads to a closed-form expression for the least-squares estimator. In turn this is compared with the maximum likelihood estimator under normality, both in its exact and in an approximated version, which is obtained by approximating the matrix in the exponent of the Gaussian likelihood function. This comparison is illustrated by some numerical examples. The dependency of the results about biases on the values of the model parameter is emphasized.

Keywords: Moving average model; Residual variance estimation; Least squares; Asymptotic bias; Asymptotic mean square error

* Correspondence address. Facultad de Ciencias Economicas, Inst. de Investigaciones Estadisticas, Universidad Nacional de Tucuman, Casilla de Correo 209, 4000 Tucuman, Argentina.

1. Introduction

1.1. The model

The first-order moving average time series model, denoted MA(1), is defined by

$$X_t - \mu = \varepsilon_t + \alpha\varepsilon_{t-1}, \quad t = \ldots, -1, 0, 1, \ldots, \tag{1.1}$$

where X_t is the observable time series, ε_t is a white noise residual with zero mean and constant variance σ^2, and α, μ and σ^2 are parameters, $0 < \sigma^2 < \infty$. We call σ^2 the residual variance of the process. Further specification of the ε_t will be made in Section 1.2.

The process is stationary for any choice of α. If α is less than one in absolute value (1.1) can be inverted into an infinite autoregression,

$$\varepsilon_t = \sum_{s=0}^{\infty} (-\alpha)^s (X_{t-s} - \mu), \quad t = \ldots, -1, 0, 1, \ldots . \tag{1.2}$$

The covariance sequence of the process is $\gamma_0 = \sigma^2(1 + \alpha^2)$, $\gamma_1 = \gamma_{-1} = \sigma^2\alpha$, and $\gamma_j = 0$ for $|j| > 1$. The correlation sequence in turn has $\rho_1 = \alpha(1 + \alpha^2)^{-1}$ and $\rho_j = 0$ for $|j| > 1$.

Moving average models in general are important tools both in theoretical and empirical time-series analyses. They enter into the definition of ARMA models for stationary time series and ARIMA models for some non-stationary ones (Box and Jenkins, 1970, 3rd ed. of 1994), and these proved to be important empirical tools, as witnessed by the applications arising in a variety of fields. Some of the difficulties encountered in the statistical study of these models are due to the nature of the MA part, for example in that it postulates a linear combination of unobservable random variables. The MA(1) is the simplest case of these models and as will be shown below, serves to exhibit some of the difficulties.

1.2. Objectives of this paper

For purposes of inference we consider a sample $X_1, \ldots, X_T$ from Eq. (1.1). Estimation procedures frequently used are the methods of moments (MM), least squares (LS) and maximum likelihood (ML) under normality. Some authors have also considered estimation from frequency domain considerations. We shall not study MM in this paper (see Mentz et al., 1997) and ML will be only briefly considered in Section 3.1.

The method of LS is not particularly suited for MA models. If we interpret the method as requiring to minimize $\sum \hat{\varepsilon}_t^2$ with respect to μ and α, then substitution of ε_t by Eq. (1.2) will "contain an infinite number of terms so that we cannot express (the sum) as a finite linear function of the observations" (Priestley, 1981, Section 5.4.2). However, LSE is frequently considered in practice, in treatises on time series and in many standard computer programs commercially available. In our case we used BMDP Statistical Software (BMDP, 1990), ITSM (Brockwell and

Davis, 1991b) and SCA (Scientific Computing Associates, 1993). These programs consider the estimation of parameters like α and μ in Eq. (1.1) by minimizing a properly defined sum of squares, as we shall see in Section 2. An estimate of σ^2 is then defined as proportional to the minimized sum of squares. We call this the LS estimator of σ^2 and denote it by $\hat{\sigma}^2_{LS}$.

The main objectives of this paper are: (1) to show through simulations that the asymptotic formulas for the bias and variance of the MLE estimator of σ^2 apply also to the LSE, at least for values of α far from the region of non-invertibility; (2) to derive an approximate expression for the bias of the LSE estimator; (3) to derive a closed-form approximation of the LSE, and (4) to relate this approximation to a well-known approximation used in MLE, and to illustrate its behavior with simulations.

Different assumptions are made for the ε_t: to deal with (2) and (3) above we assume that they are independent identically distributed $(0, \sigma^2)$; to deal with (1) and (4) we add that they are normal. The latter is used to define the likelihood function and hence to compare the LSE with the MLE. It follows that our simulations are made with pseudonormal random numbers.

It is relevant to ask under what assumptions can an asymptotic theory for LSE in the MA(1) model be developed. Existing theorems for ARMA models (see e.g. Brockwell and Davis, 1991, Ch. 8 and 10, or Fuller, 1996, Ch. 8) give conditions for the asymptotic theory of MLE under which the LSE has the same properties. A question is whether the LSE are consistent under more general conditions than the MLE. Another question concerns robustness, whether the LSE is more robust than the MLE to deviations from normality, since it is not based on the Gaussian likelihood. We shall not deal with these questions here.

We pay considerable attention to bias. It has been argued (Simonoff, 1993) that in the estimation of the variance of a statistic, attention should concentrate on the bias: a misleadingly optimistic result is obtained from a negatively biased estimator. In spite of this emphasis, we also consider the variance of the variance estimator, its standard error and mean square error.

In spite of its inferential role, not many papers have been written about large-sample biases and variances of residual variance estimator in MA models; for MM and ML see the authors' paper of 1997. There are more known results about AR models, some of these, as well as some new results, are presented in the authors' paper of 1998. For an interesting discussion see Harvey (1993, Ch. 3).

1.3. Asymptotic results for sample covariances and correlations

We can estimate the covariances γ_j of the model by

$$c_j = \frac{1}{T}\sum_{t=1}^{T-j}(X_t - \bar{X})(X_{t+j} - \bar{X}) = c_{-j}, \quad j = 0, 1, \ldots, T-1. \tag{1.3}$$

Other estimators are considered in the literature, for example, by changing in Eq. (1.3) the denominator, the value to be subtracted from the X's, or the range

of the sums; see, for example, Anderson (1971) or Fuller (1996). We use Eq. (1.3) because for $T > p$, a covariance matrix with elements $c_{|i-j|}$ is positive definite, a property that we shall use below. With these estimators of the covariances we form the estimators of the correlations ρ_j,

$$r_j = c_j/c_0, \quad j = 1,2,\ldots,T-1. \tag{1.4}$$

From the above indicated sources we deduce large-sample moments of these estimators. They are derived in the general case of a linear model $Y_t = \sum_{-\infty}^{\infty} \delta_j e_{t-j}$, under some conditions on the random variables e_t and coefficients δ_j; Eq. (1.1) is a linear model. For further reference we note that, with $r = r_1$,

$$\begin{aligned}
&E(c_0-\gamma_0)\approx -\frac{\sigma^2}{T}(1+\alpha)^2, && E(c_1-\gamma_1)\approx -\frac{\sigma^2}{T}(1+3\alpha+\alpha^2),\\
&E(c_0-\gamma_0)^2\approx \frac{2\sigma^4}{T}(1+4\alpha^2+\alpha^4), && E(c_1-\gamma_1)^2\approx \frac{\sigma^4}{T}(1+5\alpha^2+\alpha^4),\\
&E(c_0-\gamma_0)(c_1-\gamma_1)\approx \frac{4\sigma^4}{T}\alpha(1+\alpha^2), && E(r-\rho)^2\approx \frac{1+\alpha^2+4\alpha^4+\alpha^6+\alpha^8}{T(1+\alpha^2)^4},\\
&E(r-\rho)\approx -\frac{1+4\alpha+\alpha^2+4\alpha^3+\alpha^4+4\alpha^5+\alpha^6}{T(1+\alpha^2)^3}.
\end{aligned} \tag{1.5}$$

These expressions have errors $o(1/T)$.

2. Estimation by least squares

2.1. Estimation of the coefficients and residual variance

One procedure that is often used in practice, is to consider that the LS criterion function is the quadratic form appearing in the exponent of the Gaussian likelihood function. This function will be written down in Section 3.1. It is often preferred to use an alternative expression using elements of the minimum mean square error prediction (MMSE) of a time series. Letting now X_j denote a zero mean MA(1) (in our case we can take the mean-corrected values to approximate them), we take

$$S(\alpha) = \sum_{j=1}^{T} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} \tag{2.1}$$

as the criterion, where $\hat{X}_j$ is the MMSE predictor of X_j and $E(X_j - \hat{X}_j)^2 = \sigma^2 r_{j-1}$; these quantities are defined recursively by

$$\begin{aligned}
&r_0 = 1+\alpha^2, \quad r_{j+1} = 1+\alpha^2 - \frac{\alpha^2}{r_j}, \quad j = 0,1,2,\ldots,\\
&\hat{X}_1 = 0, \quad \hat{X}_{j+1} = \frac{\alpha(X_j - \hat{X}_j)}{r_{j-1}}, \quad j = 1,2,3,\ldots .
\end{aligned} \tag{2.2}$$

See, for example, Brockwell and Davis (1991a) that discuss this theory in detail. In their analysis they restrict attention to $|\alpha| < 1$, in which case the estimators arising

from the minimization of Eq. (2.1) have the same asymptotic properties of the MLE procedure. They suggest using $\hat{\sigma}^2_{LS} = S(\hat{\alpha}_{LS})/(T-1)$, which is incorporated in the software ITSM.

2.2. A simulation study

To study the behavior for finite samples of the bias in the estimation procedure described in the preceding section, a simulation study was performed, which is here briefly described. Model (1.1) with $\mu = 0$ was simulated with pseudorandom normal $(0,1)$ numbers, and eight values of α ranging from -0.90 to 0.95. Sample sizes were set at $T = 50, 100, 200$ and 400, and for each case 1000 replications were done. Estimation was done by using SCA.[1] Table 1 reports the numerical results, and is similar to Table IV presented in Mentz et al. (1997) for the MLE of the model. The empirical biases are averages over replications and their empirical standard errors are also based on these replications. The reported asymptotic biases are those of the MLE (see Section 3) and they are used to compute the "Studentized values": the values in column 6 come from column 3 minus 4 divided into 5. Column 7 reports the empirical mean square error, computed as the sum of the variance (obtained as 1000 times the square of column 5) and the square of the bias reported in column 3.

The main observations stemming from Table 1 can be summarized as follows: (1) *For all values of* α. (a) Empirical biases decrease in absolute values as T increases, as expected; (b) Empirical standard errors decrease monotonically as T increases, as expected; (c) Empirical standard errors for a given T have similar values for different values of α. They are well predicted by the MLE asymptotic value of $\sigma^2[(2/T)/1000]^{1/2}$ (cf. Harvey, 1993, Ch. 3) for $\sigma^2 = 1$; (d) Empirical mean square errors are well approximated by the MLE asymptotic value of $\sigma^4(2/T + 4/T^2)$ for $\sigma^4 = 1$. (2) *For* $-0.40 \le \alpha \le 0.60$. (a) Empirical biases have the same signs as the asymptotic biases of MLE; (b) Studentized biases are less than 3 in absolute value, except for one case ($\alpha = -0.40$, $T = 50$). Of the remaining 19 cases only 2 are larger than 2 in absolute values; (3) No studentized biases are larger than 2 in absolute values for $T = 200$ or 400; (d) For a fixed α, studentized values for $T = 200$ and 400 are smaller in absolute values than for $T = 50$ and 100, except in one case ($\alpha = 0.20$); (3) *For* $\alpha = -0.90, 0.80$ *and* 0.95. (a) Empirical biases differ in signs from those of the MLE; (b) Studentized biases are larger than 3 in absolute values; (d) Studentized values decrease as T increases.

We conclude that behavior of the LSE in terms of bias is comparable to that of the MLE for moderately large samples sizes and for values of the parameter removed from the boundary of the region of invertibility; for values closer to this boundary, even for sample sizes as large as $T = 400$ the fit to the MLE asymptotic theory is not good, and only for larger sample sizes the results of this theory are expected to be useful in practice. Hence, while the MLE behaves without dependence on the values

[1] A preliminary simulation study was done by using ITSM with some of the indicated values of α, the same sample sizes, and 100 replications for each case. The findings of that study were refined and confirmed by those reported here.

Table 1
Analysis of simulations, LSE estimation of σ^2

Value of α	Sample size	Empirical bias	Asymptotic bias MLE	Empirical standard error	Studentized bias	Empirical mean square error
1	2	3	4	5	6	7
−0.9	50	0.01397	−0.04000	0.00636	8.48585	0.04063
	100	0.00599	−0.02000	0.00433	6.00231	0.01876
	200	0.00313	−0.01000	0.00311	4.22186	0.00969
	400	0.00134	−0.00500	0.00217	2.92166	0.00469
−0.4	50	−0.06078	−0.04000	0.00603	− 3.44610	0.04007
	100	−0.02317	−0.02000	0.00431	− 0.73550	0.01908
	200	−0.00900	−0.01000	0.00312	0.32051	0.00979
	400	−0.00574	−0.00500	0.00219	− 0.33790	0.00482
−0.1	50	−0.02624	−0.04000	0.00612	2.24837	0.03812
	100	−0.01280	−0.02000	0.00435	1.65517	0.01907
	200	−0.00943	−0.01000	0.00307	0.18567	0.00949
	400	−0.00288	−0.00500	0.00226	0.93801	0.00514
0.1	50	−0.03052	−0.04000	0.00643	1.47434	0.04232
	100	−0.01059	−0.02000	0.00472	1.99364	0.02241
	200	−0.00872	−0.01000	0.00333	0.38438	0.0112
	400	−0.00183	−0.00500	0.00233	1.36051	0.00544
0.2	50	−0.05130	−0.04000	0.00593	− 1.90556	0.03776
	100	−0.02148	−0.02000	0.00442	− 0.33484	0.01996
	200	−0.01018	−0.01000	0.00315	− 0.05714	0.01004
	400	−0.00679	−0.00500	0.00230	− 0.77826	0.00533
0.6	50	−0.04931	−0.04000	0.00601	− 1.54909	0.03858
	100	−0.03097	−0.02000	0.00413	− 2.65618	0.01804
	200	−0.01376	−0.01000	0.00310	− 1.21290	0.00978
	400	−0.00522	−0.00500	0.00216	− 0.10185	0.00472
0.8	50	0.03069	−0.04000	0.00615	11.49431	0.03880
	100	0.02403	−0.02000	0.00430	10.23953	0.01908
	200	0.01225	−0.01000	0.00317	7.01893	0.01017
	400	0.00425	−0.00500	0.00227	4.07489	0.00517
0.95	50	0.05743	−0.04000	0.00624	15.61378	0.04221
	100	0.02998	−0.02000	0.00453	11.03311	0.02144
	200	0.01754	−0.01000	0.00316	8.71519	0.01032
	400	0.00894	−0.00500	0.00227	6.14097	0.00522

of the parameter in the true model, the LSE shows dependence on this value. This phenomenon is in agreement with our findings in other situations, some of which are pointed out in our (1997) paper on the MA model, and will also be illustrated in Section 3.2.

Another interesting finding is that for the variance of the variance estimator, the approximation provided by the MLE theory behaves better than that of the bias, in particular is not affected by the values of the main parameter α. This explains why the values of the empirical mean square error reported in column 7 are consistently

close to those coming from the asymptotic theory, namely 0.04160, 0.02040, 0.01010 and 0.00503 for $T = 50, 100, 200$ and 400, respectively.

2.3. Asymptotic theory for the residual variance estimator

To develop the statistical theory for the LSE we use the idea of a "long autoregression" (Durbin, 1959), which amounts to approximating the series in Eq. (1.2) by the finite sum

$$\varepsilon_t^* = \sum_{s=0}^{B_T} (-\alpha)^s (X_{t-s} - \mu), \tag{2.3}$$

where in the theory $B_T \to \infty$ as $T \to \infty$ in such a way that B_T (or a power of it) is dominated by T, $B_T/T \to 0$ as $T \to \infty$. Using Eq. (2.3) together with the substitution of $\bar{X}$ for μ, the following approximation can be justified,

$$\hat{\sigma}^2_{\mathrm{LS}} \approx \frac{1}{T - B_T - 2} \sum_{t=B_T+1}^{T} \left[\sum_{j=0}^{B_T} (-\hat{\alpha})^j (X_{t-j} - \bar{X}) \right]^2. \tag{2.4}$$

Under these conditions in the appendix we show that the following asymptotic representation of Eq. (2.4) holds:

$$\hat{\sigma}^2_{\mathrm{LS}} \approx \frac{c_0}{1 - \hat{\alpha}^2} + 2 \sum_{j=1}^{T-1} c_j \frac{(-\hat{\alpha})^j}{1 - \hat{\alpha}^2}, \tag{2.5}$$

and we arrive at the expression

$$E(\hat{\sigma}^2_{\mathrm{LS}} - \sigma^2) = -\frac{\sigma^2(1 - 3\alpha^2)}{T(1 - \alpha^2)} + \frac{2\sigma^2(1 + 2\alpha^2 + 21\alpha^4)}{(1 - \alpha^2)^3} E(\hat{\alpha} - \alpha)^2 + C + o(1/T), \tag{2.6}$$

where $\hat{\alpha}$ is the LSE of α, and

$$C = \frac{2\alpha E(c_0 - \gamma_0)(\hat{\alpha} - \alpha)}{(1 - \alpha^2)^2} - \frac{2(1 + \alpha^2) E(c_1 - \gamma_1)(\hat{\alpha} - \alpha)}{(1 - \alpha^2)^2} + 2 \sum_{j=2}^{\infty} \frac{-j(-\alpha)^{j-1}(1 - \alpha^2) + 2\alpha(-\alpha)^j}{(1 - \alpha^2)^2} E c_j (\hat{\alpha} - \alpha). \tag{2.7}$$

The first term on the right-hand side of Eq. (2.6) is the contribution coming from the biases of the covariance estimators, that is, $E(c_j - \gamma_j)$ for $j \geq 0$. Terms with factors $E(\hat{\alpha} - \alpha)^s$ do not contribute to the order $0(1/T)$ when $s = 1$ or $s \geq 3$. The expected value $E(\hat{\alpha} - \alpha)^2$ appearing in Eq. (2.6) can be approximated by the asymptotic variance of the MLE of α, which to $0(1/T)$ is $(1 - \alpha^2)/T$. Hence,

$$E(\hat{\sigma}^2_{\mathrm{LS}} - \sigma^2) = \frac{\sigma^2}{T} \frac{6\alpha^2 + 18\alpha^4}{(1 - \alpha^2)^2} + C + o(1/T). \tag{2.8}$$

The component C includes covariances between the LSE $\hat{\alpha}$ and the c_j, for which we have no exact or approximate expressions. In a practical situation a data analyst can approximate them numerically by using the bootstrap procedure described, for example, in Efron and Tibshirani (1993), (Ch. 8).

As simple examples we took two series of lengths $T=100$ each, of the kind we employed in the simulations reported in Table 1, one for $\alpha=-0.40$ and one for $\alpha=0.20$, that is, values of α far from the region of non-invertibility. 100 bootstrap replications were done for each series, 10 covariances were included in the sums, and the results were as follows:

$$\alpha=-0.40,\quad \hat{\alpha}=-0.2643,\quad \hat{C}=-0.0244,\quad \hat{E}(\hat{\sigma}^2_{\mathrm{LS}}-\sigma^2)=-0.0230,$$

$$\alpha=0.20,\quad \hat{\alpha}=0.3054,\quad \hat{C}=-0.0182,\quad \hat{E}(\hat{\sigma}^2_{\mathrm{LS}}-\sigma^2)=-0.0161.$$

We use $\hat{C}$ and $\hat{E}$ to indicate that we used the estimated values in place of the parameters σ^2 and α. The bias of the MLE estimator is -0.02, and the approximations are seen to be reasonable. Other quantities were estimated via the bootstrap procedure, to control the accuracy: estimates of $\sigma^2=1$ we obtained as 1.0076 and 1.0036, respectively.

3. Estimation by maximum likelihood

3.1. Estimation of the coefficients and residual variance

When the residuals in Eq. (1.1) are independent $N(0,\sigma^2)$ the likelihood function of the observations $X_1,X_2,\ldots,X_T$ is the function of μ,σ and α given by

$$L=(2\pi)^{-T/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\mu)'\Sigma^{-1}(\boldsymbol{X}-\mu)\right\}, \tag{3.1}$$

where $\boldsymbol{X}=(X_1,\ldots,X_T)'$, $\mu=(\mu,\ldots,\mu)'$, and Σ is the covariance matrix with components $\gamma_{|i-j|}$. This can also be written as

$$L=(2\pi\sigma^2)^{-T/2}|\boldsymbol{P}|^{-1/2}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{X}-\mu)'\boldsymbol{P}^{-1}(\boldsymbol{X}-\mu)\right\}, \tag{3.2}$$

where $\Sigma=\sigma^2\boldsymbol{P}$.

Maximum likelihood estimators of μ,α and σ^2 are obtained by minimizing Eq. (3.2) over its parameter space. No explicit formulas for these estimators are known, not even in the case of $q=1$. However, asymptotic expressions for some of the biases are known: Tanaka (1984) and Cordeiro and Klein (1994) give for the MA(1) model,

$$E(\hat{\alpha}-\alpha)=\frac{1+2\alpha}{T}+\mathrm{o}(1/T). \tag{3.3}$$

Minimizing Eq. (3.2) with respect to σ^2 only, shows that

$$\hat{\sigma}^2_{\mathrm{ML}}=\frac{1}{T}(\boldsymbol{X}-\hat{\mu})'\hat{P}^{-1}(\boldsymbol{X}-\hat{\mu}). \tag{3.4}$$

For this estimator in the MA(1) model Tanaka (1984) and Cordeiro and Klein (1993) give

$$E(\hat{\sigma}^2_{\mathrm{ML}}-\sigma^2)=-\frac{2\sigma^2}{T}+\mathrm{o}(1/T). \tag{3.5}$$

3.2. A numerical analysis of an approximation

In the present authors' paper (1997) the exact form of Eq. (3.4) was considered, as well as an approximation consisting in using instead of the components of $\boldsymbol{P}^{-1}$ along diagonals,

$$(-\alpha)^k\frac{1}{1-\alpha^2},\quad k=0,1,\ldots,T-1. \tag{3.6}$$

The effect of this approximation is that $\boldsymbol{P}$ times a matrix with these components produces a matrix that differs from the identity matrix in the components of the first and Tth rows.

With this approximation plus estimating μ by $\bar{X}$, it can be shown that the MLE of the residual variance becomes identical to Eq. (2.5), the approximate representation of the LSE. The fact that the LS and ML estimators coincide under some approximations has been noted elsewhere. Brockwell and Davis (1991a, Section 8.7) note that if the determinant of the covariance matrix appearing in the Gaussian likelihood function is asymptotically negligible compared with the sum of squares in its exponent, as is the case when the parameter is constrained to the region of invertibility, then minimization of the sum of squares is equivalent to maximization of the likelihood and the LS and ML estimator have the same asymptotic properties.

Table 2 reports the findings of a simulation study aimed at illustrating the effects of the indicated approximation. Model (1.1) with $\mu=0$ was simulated with pseudo-random normal (0,1) numbers, and eight values of α ranging from -0.40 to 0.95. Sample size was set at $T=100$ and for each case 1000 replications were done.

Columns 2 and 3 contain the averages over replications of the MLE of α and σ^2, respectively, obtained by processing the simulated series with BMDP.[2] In parentheses below the averages are the calculated standard deviations. Similar values are reported in column 4 when σ^2 is estimated by Eq. (3.4) with $\mu=0$ and the components of $\hat{\boldsymbol{P}}^{-1}$ obtained by replacing $\hat{\alpha}$ by α in the exact form of the inverse.[3] Columns 5 and 6 contain the values obtained by using the approximation (2.5) with covariances c_j ranging up to $K=10$ or 20, respectively.

[2] A preliminary simulation study was done by using ITSM with a few of the indicated values of α, the same sample size, and 100 replications for each case. The findings of that study were refined and confirmed by those reported here.

[3] In the preliminary study mentioned in footnote (2) the quadratic form was computed directly as a double sum using the exact form of the component of the inverse matrix. For the study reported here a much faster algorithm was used. Linear systems of the form $\boldsymbol{X}=\boldsymbol{P}\boldsymbol{Y}$ were solved recursively for $\boldsymbol{Y}$ by the forward solution of the method of successive elimination. For the case of the MA(1) model, see Anderson and Mentz (1993) for details.

Table 2
Analysis of an approximation by simulations

Value of α	$\hat{\alpha}$ by BMDP	$\hat{\sigma}^2_{(1)}$ by BMDP	$\hat{\sigma}^2_{(2)}$ Exact	$\hat{\sigma}^2_{(3)}$ Appr. $K=10$	$\hat{\sigma}^2_{(4)}$ Appr. $K=20$
1	2	3	4	5	6
−0.40	−0.4100	1.0006	0.9911	0.9873	0.9871
	(0.0996)	(0.1376)	(0.1368)	(0.1370)	(0.1372)
−0.10	−0.0984	0.9978	0.9882	0.9791	0.9791
	(0.1080)	(0.1418)	(0.1412)	(0.1397)	(0.1397)
0.10	0.1066	0.9978	0.9881	0.9793	0.9793
	(0.1098)	(0.1417)	(0.1409)	(0.1396)	(0.1396)
0.20	0.2092	0.9974	0.9877	0.9800	0.9800
	(0.1089)	(0.1416)	(0.1406)	(0.1397)	(0.1397)
0.40	0.4140	0.9959	0.9864	0.9832	0.9833
	(0.1022)	(0.1413)	(0.1402)	(0.1405)	(0.1404)
0.60	0.6190	0.9931	0.9842	0.9918	0.9907
	(0.0924)	(0.1409)	(0.1399)	(0.1508)	(0.1468)
0.80	0.8166	0.9901	0.9826	1.1646	1.0375
	(0.0700)	(0.1407)	(0.1397)	(1.3911)	(0.8363)
0.95	0.9473	1.0008	0.9934	3.9730	3.4136
	(0.0367)	(0.1460)	(0.1452)	(9.8494)	(13.3612)

The observations stemming from Table 2 can be summarized as follows: (1) *For all values of* α. (a) The MLE of α with BMDP was included as a check, and the results show a reasonable agreement with the asymptotic theory, under which the bias satisfies Eq. (3.3) and $\mathrm{Var}(\hat{\alpha}) \approx (1-\alpha^2)/T$; (b) The MLE of σ^2 and that using the exact form of $\boldsymbol{P}^{-1}$ show comparable results; it is worth noting that the empirical standard deviations cluster close to $0.1414 = \sqrt{0.02} = \sqrt{2/T}$, since $2\sigma^4/T$ can be taken as the asymptotic variance of the MLE of σ^2 (cf. Harvey, 1993, Ch. 3); (2) *For* $-0.40 \le \alpha \le 0.60$. The estimation of σ^2 by the approximation (2.5) with $K=10$ or 20, show results comparable to BMDP or exact MLE. Negative estimates were only obtained in one replication for $\alpha = 0.40$ and one for $\alpha = 0.60$. (3) *For* $\alpha = 0.80$ *and* 0.95. (a) Negative estimates of $\sigma^2 = 1$ appear with increasing frequencies, there were negative estimates in 43 of the 1000 replications for $\alpha = 0.80$ and 525 for $\alpha = 0.95$; (b) Even discarding the negative estimates and averaging over the positive values, Table 2 shows clear differences in the behavior of the approximations.

We conclude that the approximation used for the components of $\boldsymbol{P}^{-1}$ works well in the estimation by ML of the residual variance, for values of α small in absolute value for a sample size of $T = 100$. As α approaches the boundary of the region of invertibility, the use of the approximation requires larger sample sizes to be useful in the indicated way.

4. Summary and coclusions

4.1. Summary

We studied the estimation of the residual variance of the MA(1) model by means of the estimator $\hat{\sigma}^2_{\mathrm{LS}}$ arising from the method of least squares. In Section 2.3 and the appendix we developed an approach to study the asymptotic behavior of $\hat{\sigma}^2_{\mathrm{LS}}$ as sample size $T \to \infty$ that does not require the assumption of normal error terms. We obtained a rather simple closed-form approximation to this estimator as a linear combination of sample covariances with weights given by functions of the LSE of α (the approximation is also valid for $\hat{\sigma}^2_{\mathrm{ML}}$ only that the weights will be functions of the MLE or α). This approximation can safely be used provided the underlying α is smaller than 0.60 in absolute value and T is at least 100: our simulations in Table 2, done with normal data, show that in these ranges its behavior is well predicted by the asymptotic behavior of the exact MLE in terms of bias and variance.

In Section 2.2 we studied by simulations the finite-sample behavior of $\hat{\sigma}^2_{\mathrm{LS}}$ (Table 1). With normal data and using a standard computer program we concentrated in the analysis of bias and variance. We concluded that the bias in the LSE depends on the value of α, opposite to the MLE which has a constant asymptotic bias. The bias of the LSE is well approximated by $-2\sigma^2/T$ for $T = 50$ or larger, provided that the underlying α is smaller than approximately 0.60 in absolute value. If $|\alpha| > 0.60$ the approximation fails, at least for $T \le 400$ and tends to get worse as $|\alpha| \to 1$. Finally, in Section 3, after a brief presentation of $\hat{\sigma}^2_{\mathrm{ML}}$ we relate the closed-form approximation to the LSE mentioned above to an often-used approximation to the likelihood function, and illustrate it by simulations (Table 2).

4.2. Concluding remarks

Estimation in ARMA models is justified through asymptotic results, consistency (weak or strong) and asymptotic normality. The assumptions considered in this paper for the MA(1) model are sufficient for the MLE: i.i.d. $(0, \sigma^2)$ errors with σ^2 positive and finite, and $|\alpha| < 1$ (invertibility). These assumptions are also sufficient to prove that the LSE has the same properties (see for example, Brockwell and Davis 1991a, Theorems 10.8.1 and 10.8.2, or Fuller, 1996, Theorems 8.4.1 and 8.4.2). Hence, under these conditions the asymptotic theory provides no guidance to distinguish between MLE and LSE, and it follows that differences, if any, should come from

more refined asymptotic results, or from finite-sample considerations, either through theory or simulations.

In connection with the focus of our paper Harvey (1993, p. 69) writes: "As usual, the simple MA(1) models provides insight into the behavior of models other than the pure AR. The question of estimates on or outside the invertibility region is now of some importance".

A question that simulations tend to answer is how large has to be the finite sample size for the asymptotic theory to be a valid approximation. Another is how results depend on their choice of parameter values. For simulations to allow for comparisons with MLE, pseudonormal error terms are called for. Simulations with other error distributions will address the important but different question of robustness.

If a data analyst plans to use the assumption of normality in his analyses, he will find useful the results in a 1997 study by the present authors; he can construct the estimator $\hat{\sigma}^2_{\mathrm{ML}}(1+2/T)$ corrected (approximately) for bias. Our results in Section 2.2, Table 1, show the differences that arise if he uses the LSE instead, and the risks he faces if he still uses the MLE theory, in particular if $|\alpha|>0.60$ and T is small. If with his data he uses the LSE procedure, then our results in Section 2.3 are useful, and our simulation study in Section 2 serves as a reference: if $|\alpha|>0.60$, even large normal data sets tend to give unexpected results in terms of biases, existing theory for MLE fails to provide adequate guidance, the true correction for bias is complicated since it depends on the value of α.

5. For Further Reading

The following references are also of interest to the reader: Mentz et al. 1998.

Acknowledgements

Appreciation is expressed to Eng. Carlos Martinez for help in conducting the simulations leading to Table 2. We thank two referees and the editor for critiques and comments that led to substantial improvements in our presentation.

Appendix. Derivation of Eqs. (2.5)-(2.8)

Eq. (2.4) can be written as

$$\hat{\sigma}^2_{\mathrm{LS}} \approx \sum_{j=0}^{B_T} \hat{\alpha}^{2j} c^*_{jj} + 2\sum_{j=1}^{B_T} \sum_{0\le k<j} (-\hat{\alpha})^{j+k} c^*_{jk}, \tag{A.1}$$

where

$$c^*_{jk} = \frac{1}{T-B_T-2} \sum_{t=B_T+1}^{T} (X_{t-j}-\bar{X})(X_{t-k}-\bar{X}). \tag{A.2}$$

These quantities are not the same as those defined in Eq. (1.3), they all have the same number of terms and they do not form a Toeplitz matrix, since, for example, c_{11} and c_{22} are not equal. However, we have that, for fixed B_T

$$\begin{aligned}
&\lim_{T\to\infty} \mathrm{TE}(c^*_{jk} - \gamma_{j-k}) \\
&= \lim_{T\to\infty} \mathrm{TE}\left\{ \frac{1}{T-B_T-2} \sum_{B_T+1}^{T} [(X_{t-j}-\mu)-(\bar{X}-\mu)][(X_{t-k}-\mu)-(\bar{X}-\mu)] \right. \\
&\qquad \left. -E[(X_{t-j}-\mu)(X_{t-k}-\mu)] \right\} \\
&= \lim_{T\to\infty} \mathrm{TE}\left\{ \frac{1}{T-B_T-2} \sum_{t=B_T+1}^{T} [-(X_{t-j}-\mu)(\bar{X}-\mu) \right. \\
&\qquad \left. -(X_{t-k}-\mu)(\bar{X}-\mu)+(\bar{X}-\mu)^2] \right\} \\
&= \lim_{T\to\infty} \mathrm{TE}[-(\bar{X}-\mu)^2] = \lim_{T\to\infty} -T\,\mathrm{Var}(\bar{X}) = -2\pi f(0) = -\sigma^2(1+\alpha)^2,
\end{aligned} \tag{A.3}$$

where f is the MA(1) spectral density function $f(\lambda)=(\sigma^2/2\pi)(1+\alpha^2+2\alpha\cos\lambda)$, $-\pi\le\lambda\le\pi$; see, for example, Anderson (1971) Theorem 8.3.1. It then follows that the c^*'s and the c's have the same asymptotic biases, except when $|j-k|=1$; see the second formula in the first line of Eq. (1.5). To the degree of approximation used here, it follows that we can approximate (A.1) by using c's instead of c^*'s, which after summing the series involving $\hat{\alpha}$ leads to

$$\hat{\sigma}^2_{\mathrm{LS}} \approx \frac{c_0}{1-\hat{\alpha}^2} + 2\sum_{k=1}^{\infty} c_k \frac{(-\hat{\alpha})^k}{1-\hat{\alpha}^2}. \tag{A.4}$$

Expression (2.5) is derived from Eq. (A.4) by letting the sum range up to $T-1$, since with T observations we obtain sample covariances up to this order only.

In terms of parameters we have that

$$\frac{\gamma_0}{1-\alpha^2} + 2\sum_{j=1}^{\infty} \gamma_j \frac{(-\alpha)^j}{1-\alpha^2} = \frac{\gamma_0+2(-\alpha)\gamma_1}{1-\alpha^2} = \sigma^2, \tag{A.5}$$

so that

$$\hat{\sigma}^2_{\mathrm{LS}} - \sigma^2 \approx \left(\frac{c_0}{1-\hat{\alpha}^2} - \frac{\gamma_0}{1-\alpha^2}\right) + \left(c_1\frac{-2\hat{\alpha}}{1-\hat{\alpha}^2} - \gamma_1\frac{-2\alpha}{1-\alpha^2}\right) + 2\sum_{j=2}^{\infty} c_j \frac{(-\hat{\alpha})^j}{1-\hat{\alpha}^2}, \tag{A.6}$$

where we used that $\gamma_j=0$ for $j\ge 2$.

We want a Taylor expansion that retains terms up to second order, when the variables in the expansion are the c's and $\hat{\alpha}$. The coefficientes of $(c_j-\gamma_j)^s$ vanish for $s>1$, and the coefficient of $(\hat{\alpha}-\alpha)$ is

$$\frac{2\alpha\gamma_0}{(1-\alpha^2)^2} - \frac{2(1+\alpha^2)\,\gamma_1}{(1-\alpha^2)^2} \tag{A.7}$$

which equals 0 because $\gamma_0 = \sigma^2(1+\alpha^2)$, $\gamma_1 = \sigma^2\alpha$. The coefficient of $(\hat{\alpha}-\alpha)^2$ is

$$\frac{2\sigma^2(1+2\alpha^2+21\alpha^4)}{(1-\alpha^2)^3}. \tag{A.8}$$

Hence,

$$\begin{aligned}\hat{\sigma}^2_{LS} - \sigma^2 \approx{}& \frac{c_0-\gamma_0}{1-\alpha^2} - \frac{2\alpha(c_1-\gamma_1)}{1-\alpha^2} + \frac{2}{1-\alpha^2}\sum_{j=2}^{\infty} c_j(-\hat{\alpha})^j \\ &+ \frac{2\sigma^2(1+2\alpha^2+21\alpha^4)}{(1-\alpha^2)^3}(\hat{\alpha}-\alpha)^2 \\ &+ \frac{2\alpha(c_0-\gamma_0)}{(1-\alpha^2)^2}(\hat{\alpha}-\alpha) - \frac{2(1+\alpha^2)(c_1-\gamma_1)}{(1-\alpha^2)^2}(\hat{\alpha}-\alpha) \\ &+ 2\sum_{j=2}^{\infty} c_j(\hat{\alpha}-\alpha)\frac{-j(-\alpha)^{j-1}(1-\alpha^2)+2\alpha(-\alpha)^j}{(1-\alpha^2)^2}.\end{aligned} \tag{A.9}$$

The expected value of the fourth term in Eq. (A.9) is the second term on the right-hand side of Eq. (2.6), and the expected value of the last three terms is the component C in Eq. (2.7). To evaluate the expected value of the sum of the first three terms we use results in Eq. (1.5) and

$$Ec_j \approx -\mathrm{Var}(\bar{X}) = -\frac{2\pi f(0)}{T} = -\frac{\sigma^2(1+\alpha)^2}{T}, \quad j \geq 2, \tag{A.10}$$

and Eq. (2.6) follows.

References

Anderson, T.W., 1971. The Statistical Analysis of Time Series. Wiley, New York.

Anderson, T.W., Mentz, R.P., 1993. Evaluation of quadratic forms and traces for iterative estimation in first-order moving average models. Comm. Statist. Theory Methods 22, 931–963.

BMDP, 1990. BMDP Statistical Software. Los Angeles, California.

Box, G.E.P., Jenkins, G.M., 1970. Time Series Analysis, Forecasting and Control. San Francisco, Holden Day, 3rd ed. with Gregory C. Reinsel, 1994, Prentice-Hall, NJ.

Brockwell, P.J., Davis, R.A., 1991a. Time Series: Theory and Methods, 2nd ed. Springer, New York.

Brockwell, P.J., Davis, R.A., 1991b. ITSM: An Interactive Time Series Modelling Package for the PC. Springer, New York.

Cordeiro, G.M., Klein, R., 1994. Bias correction of maximum likelihood estimates for ARMA models. Statist. Probab. Lett. 169–176.

Durbin, J., 1959. Efficient estimation of parameters in moving average models. Biometrika 46, 306–316.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

Fuller, W.A., 1996. Introduction to Statistical Time Series, 2nd ed. Wiley, New York.

Harvey, A.C., 1993. Time Series Models, 2nd ed. MIT Press, Cambridge, MA.

Mentz, R.P., Morettin, P.A., Toloi, C.M.C., 1997. Residual variance estimation in moving average models. Comm. Statist. Theory Methods 26 (8), 1905–1923.

Mentz, R.P., Morettin, P.A., Toloi, C.M.C., 1998. On residual variance estimation in autoregressive models. J. Time Series Anal. 19 (2), 187–208.

Priestley, M.B., 1981. Spectral Analysis of Time Series. Academic Press, London.

Scientific Computing Associates, 1993. The SCA Statistical System (Release V.2). Lisle, Illinois.

Simonoff, J.S., 1993. The relative importance of bias and variability in the estimation of the variance of a statistic. Statistician 42, 3–7.

Tanaka, K., 1984. An asymptotic expansion associated with the maximum likelihood estimator in ARMA models. J. Roy. Statist. Soc. Ser. B 46, 58–67.

Aims and Scope of CSDA

Computational Statistics & Data Analysis ("CSDA") the official journal of the International Association of Statistical Computing ("IASC"), is an international journal dedicated to the dissemination of methodological research and applications in the areas of computational statistics and data analysis. The Journal consists of three refereed sections, and a fourth section dedicated to news on statistical computing. The refereed sections are divided into the following subject areas: I) Computational Statistics — Manuscripts dealing with the explicit impact of computers on statistical methodology (e.g., algorithms, computer graphics, computer intensive inferential methods, data exploration, evaluation of statistical software, expert systems, neural networks, parallel computing, statistical databases, statistical systems); II) Statistical Methodology for Data Analysis — Manuscripts dealing with data analysis strategies and methodologies (e.g., classification, data exploration, density estimation, design of experiments, model free data exploration, pattern recognition/image analysis, robust procedures); and III) Special Applications — Manuscripts at the interface of statistics and computers (e.g., comparison of statistical methodology, computer-assisted instruction for statistics, simulation experiments). The fourth section, IV) the Statistical Software Newsletter ("SSN") — The rapid exchange of informational articles and news items (e.g., articles related to the development, usage and validation of statistical software; software reviews; review of books related to computational statistics/data analysis; announcements of new software products/releases; comparison of software products; software tutorials). Announcements and meetings, and news items from the IASC and the ASA Section of Computing and Graphics may also be contributed to this section. Although not peer-reviewed, contributions to the SSN are screened for appropriateness and edited for accuracy. On occasion, special events or topics will be published as a Special Issue of CSDA, prepared by a Guest Editor.

Publication information: COMPUTATIONAL STATISTICS & DATA ANALYSIS (ISSN 0167-9473). For 1999 volumes 29–31 are scheduled for publication. Subscription prices are available upon request from the publisher. Subscriptions are accepted on a prepaid basis only and are entered on a calendar year basis. Issues are sent by surface mail except to the following countries where air delivery via SAL is ensured: Argentina, Australia, Brazil, Canada, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, PR China, Singapore, South Africa, South Korea, Taiwan, Thailand, USA. For all other countries airmail rates are available upon request. Claims for missing issues must be made within six months of our publication (mailing) date. For orders, claims, product enquiries (no manuscript enquiries) please contact the Customer Support Department at the Regional Sales Office nearest to you:

New York, Elsevier Science, P.O. Box 945, New York, NY 10159-0945, USA. Tel: (+1) 212-633-3730, [Toll Free number for North American Customers: 1-888-4ES-INFO (437-4636)], Fax: (+1) 212-633-3680, E-mail: usinfo-f@elsevier.com

Amsterdam, Elsevier Science, P.O. Box 211, 1000 AE Amsterdam, The Netherlands. Tel: (+31) 20-485-3757, Fax: (+31) 20-485-3432, E-mail: nlinfo-f@elsevier.nl

Tokyo, Elsevier Science, K.K., Higashi-Azabu 1-chome, Minato-ku, Tokyo 106, Japan. Tel: (+81) 3-5561-5033, Fax: (+81) 3-5561-5047, E-mail: info@elsevier.co.jp

Singapore, Elsevier Science, No. 1 Temasek Avenue, #17-01 Millenia Tower, Singapore 039192. Tel: (+65) 434-3727, Fax: (+65) 337-2230, E-mail: asiainfo@elsevier.com.sg

US mailing info-Computational Statistics & Data Analysis (ISSN 0167-9473) is published monthly by Elsevier Science NL (Postbus 211, 1000 AE Amsterdam). Annual subscription price in the USA US$ 1276.00 (US$ price valid in North, Central and South America only), including air speed delivery. Periodicals postage rate is paid at Jamaica, NY 11431. USA POSTMASTERS: Send address changes to Computational Statistics & Data Analysis Publications Expediting, Inc., 200 Meacham Avenue, Elmont, NY 11003. Airfreight and mailing in the USA by Publications Expediting. Abstracted or indexed in Current Index to Statistics, Mathematical Reviews, INSPEC, ISI, ACM Computing Reviews, STMA, QCAS, and OR/MS.

Journal of time series analysis

A journal sponsored by the Bernoulli Society
for Mathematical Statistics and Probability

P. Burman and R. H. Shumway: Semiparametric modeling of seasonal time series 127

P. H. Franses and M. McAleer: Testing for unit roots and non-linear transformations 147

R. J. Kulperger and R. A. Lockhart: Tests of independence in time series 165

R. P. Mentz, P. A. Morettin and C. M. C. Toloi: On residual variance estimation in autoregressive models 187

A. C. Monti: A proposal for estimation of the parameters of multivariate moving-average models 209

J.-M. Poggi and M.-C. Viano: An estimate of the fractal index using multiscale aggregates 221

Z. Qiou and N. Ravishanker: Bayesian inference for time series with stable innovations 235

Book Review 251

The **Journal of Time Series Analysis** is published six times a year in January, March, May, July, September and November by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK or 350 Main Street, Malden, MA 02148, USA.

Editor: M. B. Priestley, Department of Mathematics, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD.

The Journal is sponsored by the BERNOULLI SOCIETY for Mathematical Statistics and Probability which is an international society for the advancement of the theory and application of mathematical statistics and probability. It is a Section of the International Statistical Institute. Further information can be obtained from the Society's Executive Secretary, Dr. Zoltan E. Kenessey, International Statistical Institute, 428 Prinses Beatrixlaan, 2270 AZ Voorburg, Netherlands.

Information for Subscribers: New orders and sample copy requests should be addressed to the Journals Marketing Manager at the publisher's address above (or by email to jnlsamples@blackwellpublishers.co.uk, quoting the name of the journal). Renewals, claims and all other correspondence relating to subscriptions should be addressed to Blackwell Publishers Journals, PO Box 805, 108 Cowley Road, Oxford OX4 1FH, UK (tel: +44 (0)1865 244083, fax: +44 (0)1865 381381 or email: jnlinfo@blackwell-publishers.co.uk). Cheques should be made payable to Blackwell Publishers Ltd. All subscriptions are supplied on a calendar year basis (January to December).

Internet: For information on all Blackwell Publishers books, journals and services, log onto URL: http://www.blackwellpublishers.co.uk.

Annual Subscription Rate (1998): six issues. For Libraries and other multiple reader Institutions: £285 (UK and Europe), £314 (Rest of World), $521 (US)*. Individuals may subscribe at the reduced rate of £58 (UK and Europe), £76 (Rest of World), $126 (US) on the understanding that such a subscription is only for their personal use. The special student rate is available at £30 (UK and Europe), £41 (Rest of World), $68 (US) per year for a minimum of three years. Some evidence of student status should be enclosed. *Canadian customers/residents please add 7% for GST.

Papers appearing in the **Journal of Time Series Analysis** are abstracted in **Statistical Theory and Method Abstracts (STMA).**

Back Issues: Single issues from the current and previous two volumes are available from Blackwell Publishers Journals at the current single issue price. Earlier issues may be obtained from Swets & Zeitlinger, Back Sets, Heereweg 347, PO Box 810, 2160 SZ Lisse, Holland.

Microform: The Journal is available on microfilm (16mm or 35mm) or 105mm microfiche from the Serials Acquisitions Department, University Microfilms Inc, 300 North Zeeb Road, Ann Arbor, MI 48106, USA.

US Mailing: Periodicals postage paid at Rahway NJ. Postmaster: send address corrections to Journal of Time Series Analysis, c/o Mercury Airfreight International Ltd Inc, 2323 E-F Randolph Ave, Avenel, NJ 07001, USA (US mailing agent).

Advertising: For details contact Andy Patterson, Wheatsheaf House, Woolpit Heath, Bury St Edmunds, IP30 9RN, UK (Tel. 01359 242375, Fax: 01359 242837) or write to the Publisher.

On residual variance estimation in autoregressive models

by Raúl P. Mentz, Pedro A. Morettin, Clélia M. C. Toloi

University of Tucumán and CONICET, Argentina

and University of São Paulo, Brazil

First version received June 1995

Abstract. In this paper we consider time series models belonging to the autoregressive (AR) family and deal with the estimation of the residual variance. This is important because estimates of the variance are involved in, for example, confidence sets for the parameters of the model, estimation of the spectrum, expressions for the estimated error of prediction and sample quantities used to make inferences about the order of the model. We consider the asymptotic biases for moment and least squares estimators of the residual variance, and compare them with known results when available and with those for maximum likelihood estimators under normality. Simulation results are presented for finite samples.

Keywords. AR models; bias; least squares estimator; maximum likelihood estimator; moment estimator; residual variance; time series.

1. Introduction

We consider time series models belonging to the AR(p) family in which the observable stationary process $\{X_t\}$ has $E\{X_t\} = \mu$ and finite variance, and satisfies

$$\sum_{j=0}^{p} \beta_j (X_{t-j} - \mu) = a_t \qquad t = \ldots, -1, 0, 1, \ldots \tag{1.1}$$

for $\beta_0 = 1$, $\beta_1, \ldots, \beta_p$ real parameters, where $\{a_t\}$ is white noise with $E\{a_t^2\} = \sigma_a^2 > 0$ and a_t is independent of X_{t-1}, $X_{t-2}, \ldots$. The roots of the polynomial equation

$$\sum_{j=0}^{p} \beta_j z^{p-j} = 0 \tag{1.2}$$

satisfy $|z_j| < 1$, $j = 1, \ldots, p$, and (1.1) is invertible into an infinite moving average.

In this paper we consider the estimation of the parameter σ_a^2. This is important because estimates of σ_a^2 are involved in, for example, confidence sets for the parameters, estimation of the spectrum and expressions for the

estimated error of prediction. Also, to fit the correct order requires good estimates of σ_a^2.

Estimates of σ_a^2 come from the method of moments (MM), least squares (LS) or maximum likelihood (ML) under normality, and also from frequency domain arguments.

The problem of finding asymptotic biases for the estimators of the parameters in AR models has been considered by several authors. The biases for Yule–Walker (YW) estimators (or MM estimators) and LS estimators of the coefficients were considered by Tjøstheim and Paulsen (1983) and Shaman and Stine (1988). Shaman (1983) derived the asymptotic biases of the estimators of the residual variance of autoregressive (AR) models with known mean, assuming that the order is not known: the underlying model is an infinite autoregression, which is approximated by AR(p) models with finite p, and YW and LS estimators are considered. Paulsen and Tjøstheim (1985) studied also Burg-type estimators. For ML estimators the main references are Tanaka (1984), who suggested a technique for obtaining the Edgeworth-type asymptotic expansion associated with ML estimators in autoregressive moving-average (ARMA) models, and Cordeiro and Klein (1994), who presented a general procedure for obtaining the biases of ML estimators in ARMA models.

The estimation of the innovation variance through frequency domain considerations has been studied by Pukkila and Nyquist (1985). This approach will not be considered in this paper.

Further references are Marriott and Pope (1954), Walker (1962), Ansley and Newbold (1981), Bhansali (1981), Kunitomo and Yamamoto (1985), Lysne and Tjøstheim (1986), Stine and Shaman (1990), Tuan (1992) and De Gooijer and Pukkila (1994). Techniques and results for asymptotic analysis in time series models are well covered by Anderson (1971) and Fuller (1996).

In this paper we derive new results on the subject and, for comparison, state and review other results existing in the literature. Some preliminary results are given in Section 2, MM (YW) estimators are discussed in Section 3, LS estimators in Section 4 and ML estimators in Section 5. Section 6 presents some simulation results, and a brief discussion on asymptotic distribution is given in Section 7. Some concluding remarks are provided in the final section.

2. Preliminaries

The autocovariance sequence of the AR(p) process is given by

$$\gamma_s = \gamma_{-s} = E(X_t - \mu)(X_{t+s} - \mu) \qquad s = 0, 1, 2, \ldots \tag{2.1}$$

and satisfies the (theoretical) YW equations

$$\sum_{j=0}^{p} \beta_j \gamma_j = \sigma_a^2 \tag{2.2}$$

and

$$\sum_{j=0}^{p} \beta_j \gamma_{j-s} = 0 \qquad s = 1, 2, \ldots. \tag{2.3}$$

Setting $\boldsymbol{\beta}_p = (\beta_1, \ldots, \beta_p)'$, $\boldsymbol{\gamma}_p = (\gamma_1, \ldots, \gamma_p)'$, $\boldsymbol{\Gamma}_p = [\gamma_{ij}]$, with $\gamma_{ij} = \gamma_{|i-j|}$, $i, j = 1, \ldots, p$, (2.2) can be written as

$$\sigma_a^2 = \gamma_0 + \boldsymbol{\beta}_p' \boldsymbol{\gamma}_p \tag{2.4}$$

and the first p equations in (2.3) as

$$\boldsymbol{\Gamma}_p \boldsymbol{\beta}_p = -\boldsymbol{\gamma}_p. \tag{2.5}$$

The autocovariance sequence satisfies the 'inversion formula'

$$\gamma_s = \int_{-\pi}^{+\pi} \exp(i\lambda s) f(\lambda)\, d\lambda$$

where $f(\lambda)$ is the spectral density of the process,

$$f(\lambda) = \frac{\sigma_a^2}{2\pi} |B(\lambda)|^{-2} \qquad -\pi \leq \lambda \leq \pi \tag{2.6}$$

where $B(\lambda) = \sum_{j=0}^{p} \beta_j \exp(-i\lambda j)$. The correlation sequence is $\rho_s = \gamma_s/\gamma_0$, $s = 0, \pm 1, \ldots$.

For purposes of inference we have a sample $X_1, \ldots, X_T$ from (1.1). We shall denote by $\hat{\sigma}_{\mathrm{R}k}^2$, $k = 1, 2, \ldots$, the various estimators that we consider in the case of AR(p) models. We omit the identifier p, for simplicity.

3. Moment (yule walker) estimators

A YW or MM estimator $\hat{\sigma}_{\mathrm{R1}}^2$ is provided by the sample analogue of (2.2), namely

$$\hat{\sigma}_{\mathrm{R1}}^2 = \sum_{j=0}^{p} \hat{\beta}_j^{(1)} c_j \tag{3.1}$$

where $\hat{\beta}_0^{(1)} = 1$ and

$$c_j = \frac{1}{T} \sum_{t=1}^{T-j} (X_t - \overline{X})(X_{t+j} - \overline{X}) = c_{-j} \qquad j = 0, 1, \ldots, p \tag{3.2}$$

with $\overline{X} = \hat{\mu}_1 = (1/T)\sum_{t=1}^{T} X_t$. In (3.1) the $\hat{\beta}_j^{(1)}$ come from the sample analogue of (2.5), namely

$$\hat{\boldsymbol{\Gamma}}_p \hat{\boldsymbol{\beta}}_p^{(1)} = -\mathbf{c}_p \tag{3.3}$$

where $\hat{\boldsymbol{\Gamma}}_p = [c_{|i-j|}]$, $\mathbf{c}_p = (c_1, \ldots, c_p)'$ and $\hat{\boldsymbol{\beta}}_p^{(1)} = (\hat{\beta}_1^{(1)}, \ldots, \hat{\beta}_p^{(1)})'$. Equations (3.1) and (3.3) are the sample YW equations.

3.1. Asymptotic biases of coefficient estimators

Assume that the order p is known. In deriving the bias of $\hat{\sigma}^2_{R1}$, the biases of the YW estimators of the components of $\boldsymbol{\beta}_p$ will be needed. They are obtained as follows. Following Shaman and Stine (1988), let $\{X_t\}$ be given by (1.1) and assume that the a_t have finite moment of order 16 and that their Equation (2.5) is satisfied. These conditions hold if the a_t are independent and identically distributed (i.i.d.) N(0, σ_a^2), $0<\sigma_a^2<\infty$. Then

$$\lim_{T\to\infty} TE(\hat{\boldsymbol{\beta}}_p^{(1)} - \boldsymbol{\beta}_p) = \boldsymbol{b}_1 + \boldsymbol{b}_2 + \boldsymbol{b}_3 + \boldsymbol{b}_4 \tag{3.4}$$

where the vectors $\boldsymbol{b}_i$ are such that the following hold.

(i) The jth component of $\boldsymbol{b}_1$ is given by

$$\sum_{r=0}^{j-1}(\beta_r - \beta_{p-r}) \qquad j = 1, \ldots, p.$$

(ii) $\boldsymbol{b}_2 = -(\beta_1, 2\beta_2, \ldots, p\beta_p)'$.

(iii) $\boldsymbol{b}_3 = \begin{cases} \sum_{j=0}^{p/2-1}(\beta_j - \beta_{p-j})\boldsymbol{u}_j & \text{if } p \text{ is even} \\ \sum_{j=0}^{(p-1)/2}(\beta_{p-1} - \beta_{p-j})\boldsymbol{v}_j & \text{if } p \text{ is odd} \end{cases}$

where $\boldsymbol{u}_j$ is a $p\times 1$ vector with 1s in rows $j+2$, $j+4, \ldots, p-j$ and 0s eleswhere, and $\boldsymbol{v}_j$ is a $p\times 1$ vector with 1s in rows $j+1$, $j+3, \ldots, p-j$ and 0s elsewhere. Note that $\beta_{-1} = 0$.

(iv) $\boldsymbol{b}_4 = \boldsymbol{\Gamma}_p^{-1}\boldsymbol{d}$, where $\boldsymbol{d}$ is a $p\times 1$ vector with jth element

$$d_j = \sum_{k=0}^{p} |j-k|\gamma_{j-k}\beta_k \qquad j = 1, \ldots, p.$$

Note that if $\mu = 0$ then $\boldsymbol{b}_1 = \boldsymbol{0}$.

For $p = 1$ (3.4) gives

$$E(\hat{\beta}_1^{(1)} - \beta_1) \approx \frac{1}{T}(1 - 4\beta_1). \tag{3.5}$$

For the case $p = 2$ and other details, see Shaman and Stine (1988).

These approximations have errors $o(T^{-1})$. If in model (1.1) we use the parameterization $\phi_j = -\beta_j$ (as in Box *et al.*, 1994, for example), the expressions for the biases will change.

3.2. Asymptotic biases of variance estimators

We now consider finding the asymptotic bias for the YW estimator of σ_a^2. In what follows we assume that p is known and the $\{a_t\}$ are i.i.d. N(0, σ_a^2).

3.2.1. Expansions for the AR(ρ) model

From (2.4) we have, with $\hat{\boldsymbol{\beta}}_p^{(1)} = \hat{\boldsymbol{\beta}}_p$,

$$\hat{\sigma}^2_{R1} - \sigma_a^2 = (c_0 - \gamma_0) + (\hat{\boldsymbol{\beta}}_p'\boldsymbol{c}_s - \boldsymbol{\beta}_p'\boldsymbol{\gamma}_p). \tag{3.6}$$

If in the second term of (3.6) we use a Taylor expansion to $O(T^{-1})$ and take expected values, we obtain (because all other contributions vanish)

$$E(\hat{\sigma}^2_{R1} - \sigma_a^2) = E(c_0 - \gamma_0) + \sum_{j=1}^{p}\gamma_j E(\hat{\beta}_j - \beta_j) + \sum_{j=1}^{p}\beta_j E(c_j - \gamma_j)$$

$$+ \sum_{j=1}^{p} E(\hat{\beta}_j - \beta_j)(c_j - \gamma_j) + o(T^{-1}). \tag{3.7}$$

The $E(\hat{\beta}_j - \beta_j)$ are given asymptotically by (3.4), and

$$E(c_j - \gamma_j) = \frac{-|j|}{T}\gamma_j - \frac{\sigma_a^2}{T}\frac{1}{(1+\sum_{j=1}^{p}\beta_j)^2} + o(T^{-1}) \tag{3.8}$$

(Fuller, 1996) so that

$$E(\hat{\sigma}^2_{R1} - \sigma_a^2) = -\frac{\sigma_a^2}{T}\frac{1}{1+\sum_{j=1}^{p}\beta_j} + \sum_{j=1}^{p}\gamma_j E(\hat{\beta}_j - \beta_j) - \frac{1}{T}\sum_{j=1}^{p}|j|\beta_j\gamma_j$$

$$+ \sum_{j=1}^{p} E(\hat{\beta}_j - \beta_j)(c_j - \gamma_j) + o(T^{-1}). \tag{3.9}$$

The disadvantage of (3.7) or (3.9) is that we mix c_js with $\hat{\beta}_j$s. Another possibility is to express σ_a^2 as a function of $(\gamma_0, \beta_1, \ldots, \beta_p)$, by working with the YW equations. A third option is to express everything in terms of covariances. For example, in the case $p = 1$ we would have

$$\sigma_a^2 = \gamma_0 + \beta_1\gamma_1 = (1-\beta_1^2)\gamma_0 = g_1(\gamma_0, \beta_1)$$

or

$$\sigma_a^2 = \gamma_0 + \beta_1\gamma_1 = \gamma_0 - \gamma_1^2/\gamma_0 = g_2(\gamma_0, \gamma_1)$$

respectively. We explore the third possibility now.

From (3.3)

$$\hat{\boldsymbol{\beta}}_p = \hat{\boldsymbol{\Gamma}}_p^{-1}\boldsymbol{c}_p = -\begin{pmatrix} c_0 & c_1 & \cdots & c_{p-1} \\ c_1 & c_0 & \cdots & c_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p-1} & c_{p-2} & \cdots & c_0 \end{pmatrix}^{-1}\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{pmatrix} \tag{3.10}$$

so that $\hat{\beta}_j = \hat{\beta}_j(c_0, \ldots, c_p)$, $j = 1, 2, \ldots, p$.

From (2.4) the YW estimator is

$$\hat{\sigma}^2_{R1} = c_0 + \sum_{j=1}^{p} c_j \hat{\beta}_j = c_0 + \sum_{j=1}^{p} c_j \hat{\beta}_j(c_0, \ldots, c_p)$$

which can be expanded in terms of the c_j. Then

$$E(\hat{\sigma}^2_{R1} - \sigma^2_a) = \sum_{j=0}^{p} E(c_j - \gamma_j) g_j + \tfrac{1}{2} \sum_{j=0}^{p} E(c_j - \gamma_j)^2 g_{jj} + \sum_{j=1}^{p} \sum_{k=0}^{j-1} E(c_j - \gamma_j)(c_k - \gamma_k) g_{kj} + o(T^{-1}) \tag{3.11}$$

where $g_j = \partial \sigma^2_a / \partial \gamma_j$, $g_{jj} = \partial^2 \sigma^2_a / \partial \gamma^2_j$ and $g_{kj} = \partial^2 \sigma^2_a / \partial \gamma_k \partial \gamma_j$. In fact, the partial derivatives are those of $\hat{\sigma}^2_a$ with respect to the c_j, evaluated at the parameter values. Hence we obtain an expression for the asymptotic bias of the YW estimator of the variance in terms of asymptotic expressions for the biases of the c_j, their variances and covariances.

3.2.2. The special case of an AR(1) model

From (3.11) we obtain, for $p = 1$,

$$E(\hat{\sigma}^2_{R1} - \sigma^2_a) = E(c_0 - \gamma_0) g_0 + E(c_1 - \gamma_1) g_1 + \tfrac{1}{2} E(c_0 - \gamma_0)^2 g_{00} + \tfrac{1}{2} E(c_1 - \gamma_1)^2 g_{11} + E(c_1 - \gamma_1)(c_0 - \gamma_0) g_{01} + o(T^{-1}). \tag{3.12}$$

It is easy to verify that

$$E(c_0 - \gamma_0) \approx \frac{-\sigma^2_a}{T} \frac{1}{(1+\beta_1)^2}$$

$$E(c_1 - \gamma_1) \approx \frac{\beta_1 \gamma_0}{T} - \frac{-\sigma^2_a}{T} \frac{1}{(1+\beta_1)^2}$$

$$E(c_0 - \gamma_0)^2 \approx \operatorname{cov}(c_0, c_0) \approx \frac{2}{T} \gamma_0^2 \frac{1+\beta_1^2}{1-\beta_1^2}$$

$$E(c_0 - \gamma_0)(c_1 - \gamma_1) \approx \operatorname{cov}(c_0, c_1) \approx \frac{4}{T} \gamma_0^2 \frac{-\beta_1}{1-\beta_1^2}$$

$$E(c_1 - \gamma_1)^2 \approx \operatorname{cov}(c_1, c_1) \approx \frac{\gamma_0^2}{T} \frac{1 + 4\beta_1^2 - \beta_1^4}{1-\beta_1^2}$$

and $g_0 = 1 + \beta_1^2$, $g_1 = 2\beta_1$, $g_{00} = -2\beta_1^2/\gamma_0$, $g_{11} = -2/\gamma_0$, $g_{01} = -2\beta_1/\gamma_0$.
Finally, substituting in (3.12) we have

$$E(\hat{\sigma}^2_{R1} - \sigma^2_a) = \frac{-2\sigma^2_a}{T} \frac{1 - 2\beta_1^2}{1 - \beta_1^2} + o(T^{-1}). \tag{3.13}$$

This can also be written in terms of the root z_1 of (1.2), since $z_1 = -\beta_1$ in this case.

3.2.3. The special case of an AR(2) model

Operating with (3.11) in the case of $p = 2$ we derive

$$E(\hat{\sigma}^2_{R1} - \sigma^2_a) = \frac{-\sigma^2_a}{T} \left\{ \frac{2(1 + z_1^2 z_2^2)(1 + z_1 z_2)}{(z_2^2 - 1)(z_2^2 - 1)(z_1 z_2 - 1)} + 5 \right\} + o(T^{-1}) \tag{3.14}$$

where z_1 and z_2 are the distinct roots of (1.2). This is the same result as that obtained by Paulsen and Tjøstheim (1985).

3.2.4. Available results for known mean

For completeness and to facilitate comparison with our results, the results derived by Shaman (1983) will be briefly reviewed. They were obtained assuming that X_t is an autoregressive process of infinite order and known mean (supposed to be zero). It is not assumed that the $\{a_t\}$ are normally distributed. Approximate X_t by an AR(p) model

$$\sum_{j=0}^{p} \beta_{jp} X_{t-j} = u_t \qquad \beta_{0p} = 1 \tag{3.15}$$

where the $\beta_{1p}, \ldots, \beta_{pp}$ are such that

$$E(u_t^2) = \sigma_p^2 = E\left(\sum_{j=0}^{p} \beta_{jp} X_{t-j} \right)^2 \tag{3.16}$$

is a minimum. Let $B_p(\lambda) = \sum_{j=0}^{p} \beta_{jp} \exp(-i\lambda j)$. Call S_{1p} the YW estimator of σ^2_a. Then Shaman (1983) proves that

$$E(S_{1p} - \sigma^2_a) = \frac{-2}{T} \sum_{j=1}^{p} j \beta_{jp} \gamma_j - \frac{1}{T} \sum_{j,k=1}^{p} |j-k| \beta_{jp} \beta_{kp} \gamma_{j-k} - \frac{2\pi}{T} \sum_{j,k=1}^{p} \gamma_p^{jk} \int_{-\pi}^{\pi} [\exp\{i\lambda(j-k)\} |B_p(\lambda)|^2 + \exp\{i\lambda(j+k)\} |B_p(\lambda)|^2] f^2(\lambda)\, d\lambda + o(T^{-1}) \tag{3.17}$$

where $\Gamma_p^{-1} = [\gamma_p^{jk}]$ and

$$\gamma_p^{jk} = \sum_{l=\max(j,k)}^{p} \frac{\beta_{l-j,l-1}\beta_{l-k,l-1}}{\sigma_{l-1}^2} \qquad j, k = 1, \ldots, p.$$

Note that $\beta_{00} = 1$, and if X_t is truly an AR(m), $m \leq p$, with $\beta_{jp} = \beta_j$, $j = 1, \ldots, m$, and $\beta_{jp} = 0$, $j = m+1, \ldots, p$, then the last sum in (3.17) reduces to $-\sigma_a^2 p/T$ and we get

$$E(S_{1p} - \sigma_a^2) = \frac{-\sigma_a^2 p}{T} - \frac{2}{T}\sum_{j=1}^{p} j\beta_j\gamma_j - \frac{1}{T}\sum_{j,k=1}^{p} |j-k|\beta_j\beta_k\gamma_{j-k} + o(T^{-1}). \tag{3.18}$$

In the special case $m = p = 1$ (3.18) gives

$$E(S_{11} - \sigma_a^2) = -\frac{\sigma_a^2}{T}\frac{1-3\beta_1^2}{1-\beta_1^2} + o(T^{-1}) \tag{3.19}$$

and for $m = p = 2$ we derive

$$E(S_{12} - \sigma_a^2) = -\frac{2\sigma_a^2}{T}\frac{1+\beta_2-2\beta_1^2-3\beta_2^2-3\beta_2^3+2\beta_1^2\beta_2}{(1-\beta_2)(1+\beta_2+\beta_1)(1+\beta_2-\beta_1)} + o(T^{-1}) \tag{3.20}$$

In Shaman's approach it is not required that the underlying model be correctly specified. See also Bhansali (1981).

3.3. Comments

We considered the model (1.1) with unknown mean μ, estimated by $\overline{X}$. The assumption that μ is known (e.g. equal to zero) has an effect over some of the results. For example, (3.13) and (3.19) are not equal, and in fact they are related by an 'add 1' rule: the factor of $-\sigma_a^2/T$ in (3.13) is one plus that in (3.19).

In Section 3.2.4, with $\mu = 0$, we started from an infinite autoregression and then passed to a finite-order p. We can reproduce final results like (3.19) for $p = 1$ by starting with a finite p and $\mu = 0$ and using approaches parallel to those in Sections 3.2.1 and 3.2.2. Note that, when $\mu = 0$, terms involving the spectral density disappear in (3.8). For the AR(2) model with $\mu = 0$, calculations parallel to those used in Section 3.2.3 produce

$$-\frac{\sigma_a^2}{T}\left\{\frac{2(1+z_1z_2)(1+z_1^2z_2^2)}{(z_1^2-1)(z_2^2-1)(z_1z_2-1)} + 4\right\} + o(T^{-1}) \tag{3.21}$$

which is the same as (3.20), expressed in terms of the parameters of the model. Again, we obtain (3.14) from (3.21) by just 'adding 1' to the expression inside braces.

The autocovariances γ_j are estimated by the c_j defined in (3.2). Several other estimators are often encountered in the literature. The c_j have the property that $\hat{\Gamma}_p$ in (3.3) is positive definite, with probability one, and that the $\hat{\beta}^{(1)}$ also

defined in (3.3) are such that (1.2) with β_j replaced by $\hat{\beta}_j^{(1)}$ has all roots less than one in absolute value, so that the fitted model is causal.

In most cases, MM estimators are less efficient than LS or ML estimators. For an AR(p) model, with p known, the YW estimator $\hat{\beta}^{(1)}$ has the same asymptotic distribution as the ML estimator. Moreover, the Durbin–Levinson recursions can be used to compute $\hat{\beta}^{(1)}$, avoiding the matrix inversions in (3.3). See Morettin (1984) for details.

One of our main findings, for the AR(p) model with unknown mean, is the asymptotic expansion (3.11), which gives the bias of the YW estimator of the variance in terms of the asymptotic evaluations of $E(c_j - \gamma_j)$, $E(c_j - \gamma_j)(c_k - \gamma_k)$, all of which are available in the literature. The error term is $o(T^{-1})$.

As a final remark we should say that a corrected estimator of the form

$$\hat{\sigma}_{\mathrm{R1c}}^2 = \frac{T}{T-p-1}\hat{\sigma}_{\mathrm{R1}}^2$$

could be used, where the correcting factor is motivated by the finite-sample bias-correcting factor used in standard variance estimation, assuming that the mean is estimated.

4. Least squares estimators

The LS procedure for an AR(p) model can be taken to consist of estimating $\mu, \beta_1, \ldots, \beta_p$ by minimizing

$$\sum_{t=p+1}^{T} a_t^{*2} = \sum_{t=p+1}^{T} \{(X_t - \mu^*) + \beta_1^*(X_{t-1} - \mu^*) + \ldots + \beta_p^*(X_{t-p} - \mu^*)\}^2 \tag{4.1}$$

among all relevant choices of $\mu^*, \beta_1^*, \ldots, \beta_p^*$. An estimate of μ can be taken to be

$$\hat{\mu}_2 = \frac{\overline{X}_1 + \hat{\beta}_1^{(2)}\overline{X}_2 + \ldots + \hat{\beta}_p^{(2)}\overline{X}_{p+1}}{1 + \hat{\beta}_1^{(2)} + \ldots + \hat{\beta}_p^{(2)}} \tag{4.2}$$

where

$$\overline{X}_{j+1} = \frac{1}{T-p}\sum_{t=p+1-j}^{T-j} X_t. \tag{4.3}$$

Let $\hat{\boldsymbol{\beta}}_p^{(2)} = (\hat{\beta}_1^{(2)}, \ldots, \hat{\beta}_p^{(2)})'$ be the resulting estimator for $\boldsymbol{\beta}$; then

$$\hat{\sigma}_{\mathrm{R2}}^2 = \frac{1}{T-2p-1}\sum_{t=p+1}^{T} \{(X_t - \hat{\mu}_2) + \hat{\beta}_1^{(2)}(X_{t-1} - \hat{\mu}_2) + \ldots + \hat{\beta}_p^{(2)}(X_{t-p} - \hat{\mu}_2)\}^2 \tag{4.4}$$

is an estimator for σ_a^2. The divisor $T-2p-1$ comes from $T-p$ summands with $p+1$ parameters (see Priestley, 1981). A different formulation (see below) was used by Shaman and Stine (1988) and Stine and Shaman (1990). We here concentrate on a simplified approach often used in practice, in which μ is replaced by its standard estimate $\overline{X}$, and the β_j are estimated by minimizing the resulting sum of squares, giving

$$0 = \sum_{t=p+1}^{T} \left\{ \sum_{j=0}^{p} \tilde{\beta}_j (X_{t-j} - \overline{X}) \right\} (X_{t-k} - \overline{X}) = (T-p) \sum_{j=0}^{p} \tilde{\beta}_j c_{jk}^* \qquad k = 1, 2, \ldots, p \tag{4.5}$$

where $\beta_0 = \tilde{\beta} = 1$, the $\tilde{\beta}_j$ are the corresponding LS estimators of the β_j and

$$c_{jk}^* = \frac{1}{T-p} \sum_{t=p+1}^{T} (X_{t-j} - \overline{X})(X_{t-k} - \overline{X}) \qquad j, k = 0, 1, \ldots, p. \tag{4.6}$$

These sample covariances are different from (3.2) in that they all have $T-p$ summands and they do not form Toeplitz matrices, since, for example, c_{11} and c_{22} are different. Clearly $c_{jk}^* = c_{kj}^*$, and so the resulting matrices are symmetric. It is not difficult to verify that

$$\lim_{T\to\infty} TE(c_{jk}^* - \gamma_{j-k}) = \lim_{T\to\infty} -TE(\overline{X} - \mu)^2 = \lim_{T\to\infty} -T \operatorname{var}(\overline{X}) = -2\pi f(0). \tag{4.7}$$

This result has to be used instead of (3.8), valid for YW estimators, i.e. all c_{jk}^* have the same asymptotic biases, namely

$$E(c_{jk}^* - \gamma_{j-k}) = -\frac{\sigma_a^2}{T} \frac{1}{(1+\beta_1+\ldots+\beta_p)^2} + o(T^{-1}). \tag{4.8}$$

This asymptotic result is also valid for the covariances considered by Shaman and Stine (1988, p. 844).

The resulting LS estimator of σ_a^2 is

$$\hat{\sigma}_{R3}^2 = \frac{1}{T-2p-1} \sum_{t=p+1}^{T} \{(X_t - \overline{X}) + \tilde{\beta}_1 (X_{t-1} - \overline{X}) + \ldots + \tilde{\beta}_p (X_{t-p} - \overline{X})\}^2. \tag{4.9}$$

Stine and Shaman (1990) use the divisor $T-2p$ in (4.9), besides replacing $\overline{X}$ by the $\hat{\mu}_j$ given above.

4.1. Asymptotic biases of coefficient estimators

Under the assumptions of Section 3.1 we have that

$$\lim_{T\to\infty} TE(\tilde{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_p) = \mathbf{b}_1 + \mathbf{b}_2 + \mathbf{b}_3 \tag{4.10}$$

where $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ are as in (3.4). See Shaman and Stine (1988) for details.

For $p = 1$

$$E(\tilde{\beta}_1 - \beta_1) \approx \frac{1 - 3\beta_1}{T} \tag{4.11}$$

and for $p = 2$

$$E\{(\tilde{\beta}_1, \tilde{\beta}_2)' - (\beta_1, \beta_2)'\} \approx \frac{1}{T}\{(1 - \beta_1 - \beta_2, 2 - 4\beta_2)'\}. \tag{4.12}$$

These approximations have errors $o(T^{-1})$.

See also Kendall (1954), Yamamoto and Kunitomo (1984) and Tjøstheim and Paulsen (1983).

4.2. Asymptotic biases of variance estimators

In this and the following two sections, we assume that the a_t are i.i.d. $N(0, \sigma_a^2)$, $0 < \sigma_a^2 < \infty$.

From (4.9) we have, for $p < T$,

$$(T-2p-1)\hat{\sigma}_{R3}^2 = (T-p) \sum_{j=0}^{p} \tilde{\beta}_j c_{j0}^* + (T-p) \sum_{j=0}^{p} \sum_{k=0}^{p} \tilde{\beta}_j \tilde{\beta}_k c_{jk}^*.$$

The second term on the right-hand side vanishes, and we can write

$$(T-2p-1)\hat{\sigma}_{R3}^2 = (T-p)c_{00}^* + (T-p) \sum_{j=0}^{p} \tilde{\beta}_j c_{j0}^*. \tag{4.13}$$

Therefore,

$$\hat{\sigma}_{R3}^2 - \sigma_a^2 \approx (c_{00}^* - \gamma_0) + \sum_{j=1}^{p} (\tilde{\beta}_j c_{j0}^* - \beta_j \gamma_j). \tag{4.14}$$

If we now use a Taylor expansion to second order and take expectations, we obtain that the bias of the LS estimator is

$$E(\hat{\sigma}_{R3}^2 - \sigma_a^2) = E(c_{00}^* - \gamma_0) + \sum_{j=1}^{p} \gamma_j E(\tilde{\beta}_j - \beta_j)$$

$$+ \sum_{j=1}^{p} \beta_j E(c_{j0}^* - \gamma_j) + \sum_{j=1}^{p} E\{(c_{j0}^* - \gamma_j)(\tilde{\beta}_j - \beta_j)\} + o(T^{-1}). \tag{4.15}$$

4.2.1. The case of an AR(1) process

For an AR(1) model, (4.15) with $\beta_1 = \beta$, $\tilde{\beta}_1 = \tilde{\beta}$ reduces to

$$E(\hat{\sigma}^2_{R3} - \sigma_a^2) = E(c^*_{00} - \gamma_0) + \beta E(c^*_{10} - \gamma_1) + \gamma_1 E(\tilde{\beta} - \beta)$$

$$+ E(c^*_{10} - \gamma_1)(\tilde{\beta} - \beta) + o(T^{-1}). \tag{4.16}$$

We can prove that

(i) $E(c^*_{10} - \gamma_1)(\tilde{\beta} - \beta) = (-1/\gamma_0)E(c^*_{10} - \gamma_1)^2 + (\gamma_1/\gamma_0^2)E(c^*_{10} - \gamma_0) + O(T^{-2})$;
(ii) $E(c^*_{10} - \gamma_1)^2 = E(c_1 - \gamma_1)^2 + O(T^{-2})$;
(iii) $E(c^*_{10} - \gamma_1)(c^*_{11} - \gamma_0) = E(c_1 - \gamma_1)(c_0 - \gamma_0) + O(T^{-2})$.

Using these results together with (4.8) and (4.11) we have, after substitution in (4.16),

$$E(\hat{\sigma}^2_{R3} - \sigma_a^2) = -\frac{2\sigma_a^2}{T} + o(T^{-1}). \tag{4.17}$$

4.2.2. Available results for known mean

Let us return to the situation of Section 3.2.4. Call S_{2p} the LS estimator of σ_p^2, using the notation of Shaman (1983). Then

$$E(S_{2p} - \sigma_p^2) = -\frac{2}{T}\sum_{j=1}^{p} j\beta_{jp}\gamma_j - \frac{1}{T}\sum_{j,k=1}^{p} \max(j, k)\beta_{jp}\beta_{kp}\gamma_{j-k}$$

$$-\frac{2\pi}{T}\sum_{j,k=1}^{p} \gamma_p^{jk}\int_{-\pi}^{\pi} [\exp\{i\lambda(j-k)\}|B_p(\lambda)|^2$$

$$+ \exp\{i\lambda(j+k)\}|B_p(\lambda)|^2]f^2(\lambda)d\lambda + o(T^{-1}). \tag{4.18}$$

If X_t is truly an AR(1) we obtain

$$E(S_{21} - \sigma_a^2) = \frac{\sigma_a^2}{T}\frac{1 - 2\beta_1^2}{1 - \beta_1^2} + o(T^{-1}). \tag{4.19}$$

We remember that $\mu = 0$ here.

4.3. Comments

For the case of $\mu = 0$, the bias of the LS variance estimator is given by $-\sigma_a^2/T + o(T^{-1})$. For arbitrary (fixed) p we have

$$E(\hat{\sigma}^2_{R3} - \sigma_a^2) = \frac{-p\sigma_a^2}{T} + o(T^{-1}) \tag{4.20}$$

for the case $\mu = 0$ and

$$E(\hat{\sigma}^2_{R3} - \sigma_a^2) = \frac{-(p+1)\sigma_a^2}{T} + o(T^{-1}) \tag{4.21}$$

for the case of unknown mean.

In the case of infinite autoregression, the summation to be minimized is different from ours, since it ranges from 1 to T, with $X_0 = X_{-1} = \ldots = X_{1-p} = 0$, so the result (4.18) obtained by Shaman (1983) is different from (4.20). Stine and Shaman (1990) in their analysis of the LS procedure use divisor $T - 2p$ instead of the $T - 2p - 1$ that we considered; for fixed p, the choice of divisors, as well as the use of $\overline{X}$ or $\hat{\mu}_j$, does not alter the asymptotic results to order $1/T$.

The asymptotic bias of the LS estimator of the residual variance will be seen to coincide with that of the ML estimator, under normality. This is consistent with the analysis in, for example, Brockwell and Davis (1991, Section 8.7). They define the LS procedure as minimizing the weighted sum of squares

$$S = \sum_{j=1}^{T} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} \tag{4.22}$$

where $\hat{X}_j$ are the one-step predictors and $E(X_j - \hat{X}_j)^2 = \sigma_a^2 r_{j-1}$. It can be shown that (4.22) is the sum of squares in the exponent of the Gaussian likelihood.

It follows then that it is reasonable to assume that the LS estimator for σ_a^2 is given by

$$\hat{\sigma}^2_{LS} = \frac{1}{T-p}S \tag{4.23}$$

where this differs from (4.9) by the denominator and the number of terms in the sum. For small p and reasonably large T this will have no essential effect.

Therefore, it would be reasonable to assume that

$$\hat{\sigma}^2_{LS} \approx \frac{T}{T-p}\hat{\sigma}^2_{ML}. \tag{4.24}$$

Hence, for an AR(p) model we have

$$E(\hat{\sigma}^2_{LS} - \sigma_a^2) \approx E\left(\frac{T}{T-p}\hat{\sigma}^2_{ML} - \sigma_a^2\right) = \frac{T}{T-p}E(\hat{\sigma}^2_{ML} - \sigma_a^2) + \sigma_a^2\frac{p}{T-p}. \tag{4.25}$$

In the simulations, we have taken $\sigma_a^2 = 1$; hence for an AR(1) model, for example,

$$\text{BIAS(LS)} = \frac{T}{T-1}\text{BIAS(ML)} + \frac{1}{T-1} \approx \text{BIAS(ML)} + \frac{1}{T}. \tag{4.26}$$

This result can be checked in Table II (later), where we can see that it holds with a great degree of accuracy. The same is true for $p = 2$, with $2/T$ as the last term in (4.26).

5. Maximum likelihood estimators

In this section we consider ML estimation when the underlying process is assumed to be Gaussian.

Let $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_p, \sigma_a^2, \mu)'$ be the vector of $p+2$ unknown parameters. The likelihood for $\boldsymbol{\theta}$ given $\boldsymbol{X} = (X_1, \ldots, X_T)'$ is

$$L(\boldsymbol{\theta}) = (2\pi)^{-T/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left[\frac{-1}{2}\{\boldsymbol{X}-\mu)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\mu)\}\right] \qquad (5.1)$$

where $\mu = E(\boldsymbol{X}) = (\mu, \ldots, \mu)'$ is the vector of expected values and $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{X}$. Let $\mu^{(4)}$, $\beta_j^{(4)}$ and $\hat{\sigma}_{R4}^2$ be the corresponding ML estimators. Asymptotic biases for the ML estimators of coefficients and residual variance have been considered by Tanaka (1984) and Cordeiro and Klein (1994) for the ARMA(p,q) model.

Cordeiro and Klein use a general expression for the bias of ML estimators given by Cox and Snell (1968) to derive the bias correction of the parameter estimators for any ARMA model.

Tanaka operates with Edgeworth-type expansions for the joint density and for the marginal distributions of the ML estimators, from which the asymptotic biases of the estimators of the parameters are derived.

In both approaches, it is difficult to obtain explicit expressions for higher orders p and q. But at least numerically the approach of Cordeiro and Klein can be implemented quite easily.

For the asymptotic biases of the variance estimators the results are, for unknown μ,

$$E(\hat{\sigma}_{R4}^2 - \sigma_a^2) = -\frac{2\sigma_a^2}{T} + o(T^{-1}) \qquad (5.2)$$

for the AR(1) model and

$$E(\hat{\sigma}_{R4}^2 - \sigma_a^2) = -\frac{3\sigma_a^2}{T} + o(T^{-1}) \qquad (5.3)$$

for the AR(2) model.

For the case $\mu = 0$ the biases are

$$E(\hat{\sigma}_{R4}^2 - \sigma_a^2) = -\frac{\sigma_a^2}{T} + o(T^{-1}) \qquad (5.4)$$

and

$$E(\hat{\sigma}_{R4}^2 - \sigma_a^2) = -\frac{2\sigma_a^2}{T} + o(T^{-1}) \qquad (5.5)$$

respectively. We see that, up to order T^{-1}, the biases only depend on the variance σ_a^2. The general results are given by (4.20) and (4.21); therefore LS and ML estimators have the same asymptotic biases, using the approach of Section 4.2.

De Gooijer and Pukkila (1994), using a method of Whittle (1954), consider

approximating the maximization of (5.1). On expression they use involves the spectrum of the process and the sample periodogram, while another is expressed in terms of covariances.

6. Simulations

In this section we present a simulation experiment. See also Ansley and Newbold (1981) and Paulsen and Tjøstheim (1985).

Table I gives the models that were used in order to check empirically the conclusions of the theoretical analyses presented in the paper. In all cases the a_t are i.i.d. N(0, 1) and $\mu = 0$. Four different sample sizes were considered: $T = 50$, 100, 200, 400. For each sample size, 1000 replicates were taken, for each model. For the computations, we used the software S-PLUS for MM and LS and SCA for ML.

Tables II and III give the main results of the simulations. A total of eight models was used: four for each of the AR(1) and AR(2) models.

For each model, the following definitions are used: EST.BIAS, estimated bias, the average of the 1000 replicates; ST.ERROR, standard error of the estimated bias, computed from the 1000 replicates; ASYM.BIAS, asymptotic bias, given by the theoretical formulae, namely (3.19) and (3.20) for MM estimators, (5.4) and (5.5) for LS and ML estimators.

The standard error is given be $s/(1000)^{1/2}$, where

$$s^2 = \sum_{i=0}^{1000} \frac{(\hat{b}_i - \bar{b})^2}{1000} \qquad \hat{b}_i = \hat{\sigma}_{a,i}^2 - 1 \qquad i = 1, \ldots, 1000 \qquad \bar{b} = \sum_{i=1}^{1000} \frac{\hat{b}_i}{1000}$$

and $\hat{\sigma}_{a,i}^2$ is the variance estimate for each method.

To facilitate the interpretation of the tables of results, we computed the intervals $\bar{b} \pm 2s/(1000)^{1/2}$, and marked with an asterisk those estimates for which the interval does not include the corresponding asymptotic bias. The results of this analysis is summarized in Table V.

The columns of Table III are ordered in increasing order of the modulus of the roots of (1.2).

Some conclusions are as follows.

TABLE I

GENERATED MODELS: $a_t \sim$ i.i.d. N(0, 1), $\mu = 0$

Parameters		Model I	Model II	Model III	Model IV
AR(1)	β_1	−0.60	−0.30	0.90	0.98
AR(2)	β_1	−1.20	−0.40[a]	−0.95[a]	−1.95[a]
	β_2	0.30	0.70	0.90	0.96

Note: [a]Complex roots.

TABLE II

ESTIMATED BIAS (WITH STANDARD ERROR) AND ASYMPTOTIC BIAS FOR MOMENT, LEAST SQUARES AND MAXIMUM LIKELIHOOD ESTIMATORS: AR(1) MODEL

		$\beta = -0.6$			$\beta = -0.3$			$\beta = 0.9$			$\beta = 0.98$		
T		MM	LS	ML	MM	LS	ML	MM	LS	ML	MM	LS	ML
50	EST.BIAS	0.03589*	−0.00546*	−0.02200	0.01822*	−0.00635*	−0.00900	0.18200*	−0.00684*	−0.01000	1.01907	−0.00545*	−0.02127
	ST.ERROR	0.00649	0.00621	0.00623	0.00636	0.00620	0.00636	0.00923	0.00620	0.00639	0.03787	0.00621	0.00639
	ASYM.BIAS	0.00250	−0.02000	−0.02000	−0.01604	−0.02000	−0.02000	0.15053	−0.02000	−0.02000	0.95010	−0.02000	−0.02000
100	EST.BIAS	0.01483*	−0.00552	−0.00300	0.00586*	−0.00589	−0.01458	0.09138*	−0.00609	−0.01382	0.47986	−0.00452	−0.01731
	ST. ERROR	0.00457	0.00447	0.00439	0.00452	0.00447	0.00443	0.00553	0.00446	0.00444	0.01554	0.00448	0.00442
	ASYM.BIAS	0.00125	−0.01000	−0.01000	−0.00802	−0.01000	−0.01000	0.07526	−0.01000	−0.01000	0.47505	−0.01000	−0.01000
200	EST.BIAS	0.00556	−0.00454	−0.00300	0.00114	−0.00452	−0.00200	0.04368	−0.00441	−0.00627	0.24401	−0.00281	−0.00685
	ST.ERROR	0.00323	0.00320	0.00304	0.00322	0.00320	0.00313	0.00358	0.00318	0.00314	0.00836	0.00319	0.00310
	ASYM.BIAS	0.00063	−0.00500	−0.00500	−0.00401	−0.00500	−0.00500	0.03763	−0.00500	−0.00500	0.23753	−0.00500	−0.00500
400	EST.BIAS	0.00108	−0.00395	−0.00064	−0.00097	−0.00381	−0.00393	0.02033	−0.00381	−0.00055	0.11803	−0.00203	−0.00655
	ST.ERROR	0.00226	0.00224	0.00224	0.00226	0.00225	0.00227	0.00242	0.00224	0.00224	0.00456	0.00224	0.00220
	ASYM.BIAS	0.00031	−0.00250	−0.00250	−0.00201	−0.00250	−0.00250	0.01882	−0.00250	−0.00250	0.11876	−0.00250	−0.00250

TABLE III

ESTIMATED BIAS (WITH STANDARD ERROR) AND ASYMPTOTIC BIAS FOR MOMENT, LEAST SQUARES AND MAXIMUM LIKELIHOOD ESTIMATORS: AR(2) MODEL

		$\beta_1 = -1.2, \beta_2 = 0.3$			$\beta_1 = -0.4, \beta_2 = 0.7$			$\beta_1 = -0.95, \beta_2 = 0.90$			$\beta_1 = -1.95, \beta_2 = 0.96$		
T		MM	LS	ML	MM	LS	ML	MM	LS	ML	MM	LS	ML
50	EST.BIAS	0.27697*	−0.02481*	−0.03000	0.10939*	−0.02499*	−0.02900	0.49478*	−0.02276*	−0.04291	72.61510*	−0.04993	−0.05224
	ST.ERROR	0.01203	0.00625	0.00636	0.00751	0.00629	0.00626	0.01566	0.00633	0.00629	2.21734	0.00620	0.00614
	ASYM.BIAS	0.24389	−0.04000	−0.04000	0.04371	−0.04000	−0.04000	0.42807	−0.04000	−0.04000	96.24570	−0.04000	−0.04000
100	EST.BIAS	0.13920	−0.01717	−0.01400	0.04247*	−0.01792	−0.02600	0.22698	−0.01711	−0.01700	31.90908*	−0.02665	−0.02186
	ST.ERROR	0.00695	0.00462	0.00446	0.00495	0.00463	0.00439	0.00797	0.00464	0.00443	0.91131	0.00430	0.00445
	ASYM.BIAS	0.12194	−0.02000	−0.02000	0.02186	−0.02000	−0.02000	0.21404	−0.02000	−0.02000	48.12285	−0.02000	−0.02000
200	EST.BIAS	0.07299*	−0.00921	−0.00500	0.02082*	−0.00971	−0.00600	0.11582	−0.00941	−0.01000	17.34470*	−0.00907	−0.01116
	ST.ERROR	0.00398	0.00321	0.00313	0.00330	0.00323	0.00310	0.00442	0.00322	0.00313	0.49605	0.00312	0.00306
	ASYM.BIAS	0.06097	−0.01000	−0.01000	0.01093	−0.01000	−0.01000	0.10702	−0.01000	−0.01000	24.06143	−0.01000	−0.01000
400	EST.BIAS	0.03457	−0.00690	−0.00200	0.00877	−0.00688	−0.00200	0.05686	−0.00684	−0.00200	9.81424*	−0.00291	−0.00415
	ST.ERROR	0.00254	0.00226	0.00228	0.00229	0.00227	0.00224	0.00278	0.00226	0.00224	0.27366	0.00228	0.00212
	ASYM.BIAS	0.03049	−0.00500	−0.00500	0.00546	−0.00500	−0.00500	0.05351	−0.00500	−0.00500	12.03071	−0.00500	−0.00500

TABLE IV

ESTIMATES OF $\text{var}(\hat{\sigma}_a^2)$; $S^2(\hat{\sigma}_a^2) = 100(\text{ST.ERROR})^2$

T	MM	LS	ML	MM	LS	ML	MM	LS	ML	MM	LS	ML	i.i.d $N(\mu, \sigma^2)$ $\frac{2\sigma^4}{T} \sim \frac{2}{T}$
(*a*) AR(1)		$\beta = -0.6$			$\beta = -0.3$			$\beta = 0.9$			$\beta = 0.98$		
50	0.04215	0.03852	0.03895	0.04041	0.03848	0.04038	0.08526	0.03839	0.04096	1.43442	0.03862	0.04077	0.04
100	0.02085	0.02001	0.01945	0.02046	0.01999	0.01963	0.03055	0.01992	0.01975	0.24152	0.02003	0.01951	0.02
200	0.01041	0.01022	0.00930	0.01036	0.01025	0.00982	0.01283	0.01012	0.00986	0.06989	0.01021	0.00963	0.01
400	0.00509	0.00504	0.00501	0.00509	0.00505	0.00516	0.00584	0.00503	0.00500	0.02078	0.00504	0.00483	0.005
(*b*) AR(2)		$\beta_1 = -1.2$, $\beta_2 = 0.3$			$\beta_1 = -0.4$, $\beta_2 = 0.7$			$\beta_1 = -0.95$, $\beta_2 = 0.90$			$\beta_1 = -1.95$, $\beta_2 = 0.96$		
50	0.14481	0.03907	0.04040	0.05633	0.03952	0.03920	0.24539	0.03998	0.03957	4916.60601	0.03848	0.03774	0.04
100	0.04826	0.02133	0.02000	0.02450	0.02148	0.01940	0.06346	0.02156	0.01950	830.47762	0.01848	0.01984	0.02
200	0.01581	0.01032	0.00975	0.01089	0.01041	0.00961	0.01953	0.01039	0.00975	246.06431	0.00970	0.00937	0.01
400	0.00643	0.00510	0.00519	0.00525	0.00514	0.00499	0.00771	0.00513	0.00497	74.89172	0.00520	0.00450	0.005

1. MM estimators tend to have larger biases than LS estimators, and these perform about the same as ML estimators, as expected.
2. Estimated biases tend to come closer to the corresponding asymptotic values, and hence to become smaller as T increases, also in agreement with what is expected.
3. Estimated biases can differ considerably from the theoretical values for small sample sizes.
4. For given sample sizes, the variability of the estimated biases, as measured by the estimated standard errors, tend to have similar values across methods.
5. More cases of lack of fit occurred when the method of moments was used. In the case of the AR(2) model with roots very close to unity, the method performed poorly, independently of the sample size. Larger sample sizes will be needed to render the asymptotic approximations useful.
6. For MM estimators and AR(1) with roots very close to unity, the performance was poor, but there was agreement between estimated and asymptotic biases.
7. The fact that the biases of MM estimators depend on the parameter values, while those of ML are constant, is well supported by the simulation study.

7. Variance and asymptotic distribution of estimators

We know that, if $X_1, \ldots, X_T$ are i.i.d $N(\mu, \sigma^2)$, then the variance of the ML estimator and the variance of the unbiased estimator of σ^2 are approximated by

$$\text{var}(\hat{\sigma}^2) \approx \frac{2}{T}\sigma^4. \tag{7.1}$$

Shaman (1983), using the approach that we discussed before (see, for example, Section 3.2.4) was able to prove for YW and LS estimators of σ_a^2 in the AR(p) model that

TABLE V

ANALYSIS OF THE RESULTS

Table	Model	Modulus of roots	Analysis of results
II	AR(1)	0.30	* for MM at $T = 50$, 100 and for LS at $T = 50$
		0.60	* for MM at $T = 50$, 100 and for LS at $T = 50$
		0.90	* for MM at $T = 50$, 100 and for LS at $T = 50$
		0.98	* for LS at $T = 50$
III	AR(2)	0.85; 0.36	* for MM at $T = 50$, 200 and for LS at $T = 50$
		0.84; 0.84	* for MM at $T = 50$, 100, 200 and for LS at $T = 50$
		0.95; 0.95	* for MM at $T = 50$ and for LS at $T = 50$
		0.98; 0.98	* for MM at $T = 50$, 100, 200, 400

$$\text{var}(\hat{\sigma}_a^2) = \frac{1}{T}(2\sigma_a^4 + \kappa_4) + o(T^{-1}) \qquad (7.2)$$

where κ_4 is the fourth-order cumulant of the a_t. In the case of normal errors we obtain (7.1).

To check on the usefulness of this asymptotic result, in Table IV we present the values of estimates of $\text{var}(\hat{\sigma}_a^2)$ defined by $S^2(\hat{\sigma}_a^2) = 1000(\text{ST.ERROR})^2$, where the latter was introduced in Section 6. We see that the agreement between $S^2(\hat{\sigma}_a^2)$ and $2/T$ for AR(1) and AR(2) models is good for LS and ML estimators.

Shaman (1983) also showed that, under suitable conditions, the asymptotic distribtion of $T^{1/2}(\hat{\sigma}_a^2 - \sigma_a^2)$ is normal, with mean zero and variance given by $2\sigma_a^4 + \kappa_4$.

8. Concluding remarks

In this paper we considered the estimation of the residual variance in AR models, i.e. the variance of the white noise component of the models defined by (1.1). This variance is a nuisance parameter, and its estimation is important because estimates enter into prediction and confidence intervals, tests of hypotheses, spectral estimates, and other inferential procedures.

It has been argued (see, for example, Simonoff, 1993) that in the estimation of the variance of a statistic attention should concentrate on the bias: a negatively biased estimator will lead to misleadingly short prediction or confidence intervals, which is a dangerous situation in an inferential context.

We considered estimation by three standard methods, namely moments, least square and maximum likelihood under normality. In the analytical part of our study we concentrated on the study of the asymptotic biases of the estimators, by using certain approximation procedures. In general, our approach is to rely on Taylor-type expansions and use asymptotic results in the literature for means, variances, autocovariances or autocorrelations of linear processes, under different definitions of these sample quantities.

The large-sample results were reasonably well supported by our simulations presented in Section 6. For example, the connection between LS and ML estimators established in Equations (4.24)–(4.26), or the parameter independence of the LS and ML results, were well fitted for all models, choices of parameters and sample sizes.

In general we found that

$$E(\hat{\sigma}_a^2 - \sigma_a^2) = -\frac{\sigma_a^2}{T} h(\theta) + o(T^{-1})$$

where $\theta = (\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)'$.

Then,

$$E\left\{\hat{\sigma}_a^2 + \frac{\sigma_a^2}{T} h(\theta)\right\} = \sigma_a^2 + o(T^{-1})$$

and hopefully also

$$E\left[\hat{\sigma}_a^2\left\{1 + \frac{1}{T} h(\hat{\theta})\right\}\right] = \sigma_a^2 + o(T^{-1}).$$

Hence

$$\hat{\sigma}^2_{\text{corrected}} = \hat{\sigma}_a^2\left\{1 + \frac{1}{T} h(\hat{\theta})\right\}$$

is suggested as an approximately bias-free estimator of σ_a^2 up to $O(T^{-1})$.

Finally, some remarks about assumptions. In taking expectations in Taylor expansions, some conditions must hold (see Fuller, 1996, p. 244). These are satisfied in the normal case. For the non-normal case, under suitable assumptions given by Bhansali (1981) and Brillinger (1969), the results of Shaman (1983) and Shaman and Stine (1988) hold.

Acknowledgement

The authors would like to thank two referees for comments that led to a substantial improvement of the paper. This work was done with partial support of Fundación Antorchas, Argentina. R.P. Mentz acknowledges partial support from the Argentine Research Council, CONICET.

References

ANDERSON, T. W. (1971) *The Statistical Analysis of Time Series*. New York: Wiley.
ANSLEY, C. F. and NEWBOLD, P. (1981) On the bias in estimates of forecast mean square error. *J. Am. Stat. Assoc.* 76, 569 –78.
BHANSALI, R. J. (1981) Effects of not knowing the order of an autoregressive process on the mean squared error of prediction—I. *J. Am. Stat. Assoc.* 76, 588–97.
BOX, G. E. P., JENKINS, G. M. and REINSEL, G. C. (1994) *Time Series Analysis: Forecasting and Control*, 3rd Ed. Englewood Cliffs, NJ: Prentice-Hall.
BRILLINGER, D. R. (1969) Asymptotic properties of spectral estimates of second order. *Biometrika* 56, 375–90.
BROCKWELL, P. J. and DAVIS, R. A. (1991) *Time Series: Theory and Methods*. New York: Springer.
CORDEIRO, G. M. and KLEIN, R. (1994) Bias correction of maximum likelihood estimates for ARMA models. *Probab. Stat. Lett.* 169–76.
COX, D. R. and SNELL, E. J. (1968) A general definition of residuals. *J. R. Stat. Soc. B* 30, 248–75.
DE GOOIJER, J. and PUKKILA, T. (1994) On the expectation of estimators for general ARMA processes. *Statistica*, 59, 39–50.
FULLER, W. A. (1996) *Introduction to Statistical Time Series*, 2nd Edn. New York: Wiley.
KENDALL, M. G. (1954) Note on bias in the estimation of autocorrelation. *Biometrika* 41, 403–4.
KUNITOMO, N. and YAMAMOTO, T. (1985) Properties of predictors in misspecified autoregressive time series models. *J. Am. Stat. Assoc.* 80, 941–50.
LYSNE, D. and TJØSTHEIM, D. (1987) Loss of spectral peaks in autoregressive spectral estimation. *Biometrika* 74, 200–6.

MARRIOTT, F. H. and POPE, J. A. (1954) Bias in the estimation of autocorrelation. *Biometrika* 41, 390–402.

MORETTIN, P. A. (1984) The Levinson algorithm and its applications in time series analysis. *Int. Stat. Rev.* 52, 83–92.

PAULSEN, J. and TJØSTHEIM, D. (1985) On the estimation of residual variance and order in autoregressive time series. *J. R. Stat. Soc., Ser. B* 47, 216–28.

PRIESTLEY, M. B. (1981) *Spectral Analysis and Time Series*. London: Academic Press.

PUKKILA, T. and NYQUIST, H. (1985) On the frequency domain estimation of the innovation variance of a stationary univariate time series. *Biometrika* 72, 317–23.

SHAMAN, P. (1983) Properties of estimates of the mean square error of prediction in autoregressive models. In *Studies in Econometrics, Time Series and Multivariate Statistics* (eds S. Karlin, T. Amemyia and L. A. Goodman). London: Academic Press, pp. 331–42.

—— and STINE, R. A. (1988) The bias of autoregressive coefficient estimators. *J. Am. Stat. Assoc.* 83, 842–48.

SIMONOFF, J. S. (1993) The relative importance of bias and variability in the estimation of the variance of a statistic. *Statistician* 42, 3–7.

STINE, R. A. and SHAMAN, P. (1990) Bias of autoregressive spectral estimators. *J. Am. Stat Assoc.* 85, 1091–98.

TANAKA, K. (1984) An asymptotic expansion associated with the maximum likelihood estimators in ARMA models. *J. R. Stat. Soc., Ser. B* 46, 58–67.

TJØSTHEIM, D. and PAULSEN, J. (1983) Bias of some commonly-used time series estimates. *Biometrika* 70, 389–99.

TUAN, D. P. (1992) Approximate distribution of parameter estimators for first-order autoregressive models. *J. Time Series Anal.* 13, 147–70.

WALKER, A. M. (1962) Large-sample estimation of parameters for autoregressive processes with moving-average residuals. *Biometrika* 49, 117–31.

WHITTLE, P. (1954) Some recent contributions to the theory of stationary processes. In *A Study in the Analysis of Stationary Time Series*, 2nd Edn (ed. H. Wold). Uppsala: Almqvist and Wiksell, Appendix 2.

YAMAMOTO, T. and KUNITOMO, N. (1984) Asymptotic bias of the least squares estimators for multivariate autoregressive models. *Ann. Inst. Stat. Math.* 36, 419–30.

Notes on the submission of manuscripts

1. **Papers** should normally be no longer than 10,000 words (or, 20 A4 typed pages).
2. They should be typed, double-spaced, on A4 paper, with ample left- and right-hand margins. A cover page should contain only the title, author's name and affiliation, and the address to which proofs should be sent.
3. Submission of a paper is held to imply that it is original and unpublished work, that it has not been submitted for publication elsewhere, and that full copyright clearance has been obtained for any material quoted in it.
4. An **abstract** should be included. This should not exceed 200 words.
5. To facilitate the production of the annual subject index, please provide a list of **key words** (not more than six) under which your paper should be indexed.
6. **Two** copies of the article should be submitted.
7. **Footnotes** should be avoided. Essential **notes** should be numbered in the text and grouped together at the end of the article. **Diagrams** and **Figures**, if they are considered essential, should be clearly related to the section of text to which they refer.
8. **Authors** of papers involving extensive use of **mathematical** symbols and formulae should take special note of the guidelines on the back cover.
9. **References** should be set out in alphabetical order of the author's name in a list at the end of the article. They should be given in standard form, as in the following examples.

   HANNAN, E. J. (1969) A note on an exact test for trend and serial correlation. *Econometrica* 37, 485–89.

   HANNAN, E. J. and TERRELL, R. D. (1973) Multiple equation systems with stationary errors. *Econometrica* 41, 299–305.

   JOHNSON, H. G. (1964a) The international competitive positions of the United States and the balance of payments prospects for 1968. *Rev. Economics and Statistics* 46, 14–32.

   —— (1964b) *Money, Trade and Economic Growth* (2nd edn). London: Allen and Unwin.

   References in the text of an article should be by author's name and year of publication as in these examples:

   Jones (1977) in a paper on . . .
   Jones (1978c, p.25) states that . . .
   Evidence is given by Smith *et al.* (1974) . . .
   Further exploration of this aspect may be found in many sources (e.g. White, 1971a; Brown and Green, 1972; Jackson, 1973, ch. 2).
10. Authors are urged to supply their article on **disk** as well as in hard copy. **Tex** and **LaTex** files on disk are especially welcome. The disk version must be identical to the hard copy and all notes on manuscript submission should still be followed.
11. A typescript which is well prepared in accordance with these guidelines will greatly assist the ease and speed of publication.
12. The author of an article accepted for publication will receive a page proof for correction. This stage must not be used as an opportunity to revise the paper, because alterations are extremely costly; extensive changes will be charged to the author and will probably result in the article being delayed to a later issue. **Speedy return of corrected proofs is important.**
13. Authors will receive 50 offprints, free of charge. Additional copies may be ordered when returning corrected proofs and a scale of charges will be sent to you at the appropriate time.
14. Contributions and queries should be sent to the Editor:

    Professor M. B. Priestley, Department of Mathematics, University of Manchester Institute of Science and Technology, PO Box 88, Manchester M60 1QD.

Special Notes for Mathematical Papers

It is important to distinguish symbols that cannot be typed, and in particular to distinguish carefully between (a) capitals and small letters, (b) ordinary and bold-faced letters, (c) certain Greek letters and similar Roman letters, (d) subscripts, superscripts, and 'ordinary' symbols, (e) numbers 0 and 1 and letters O and I where the same symbol is used. Bold-faced symbols should be underlined with a wiggly line in pencil. Mathematical symbols are automatically set in italics, and need not be underlined except to prevent ambiguity, e.g. when an isolated letter, such as a, occurs in the text. However, certain standard abbreviations are set in roman fount, not italic fount, e.g. log, lim, exp (but not e), max, min, sup, var, cov, sin, cos.

Small fount setting is technically difficult, expensive, and sometimes impossible. To reduce the use of small founts:

(a) Avoid elaborate notations involving multiple suffices.
(b) Use the expression 'exp' for the exponential function when the argument is longer than a single compact group of symbols, e.g. $\exp(a+bt+ct^2)$ but e^t.

Arrangement of formulae

Equations and long formulae should be displayed (i.e. shown on a separate line), where necessary being numbered at the right of the page. Short isolated formulae should usually be left in the text and then must be arranged so as not to be more than one line high. For example $\sum_{i=1}^{n} x_i$ should, if possible, be written $\sum x_i$ when the limits of summation are obvious.

The solidus sign (/) should be used for fractions in the text. Note that it is essential to bracket a group of symbols to the right of the solidus if they are to be included in the denominator; $(a+b)/(c+d)(h+k)$ is wrong, being ambiguous without a special convention. However, simple fractions $\frac{1}{2}$, . . . should be written as one-line fractions, thus $\frac{1}{2}t$ is preferred to $t/2$, while $\frac{t}{2}$ must not be used in the text.

Equations involving complicated expressions should, where possible, be avoided by introducing abbreviating symbols, e.g. $\omega = (1-\varepsilon_h - \varepsilon_f)$, or $\mu = \frac{1}{(1=\varepsilon_h - \varepsilon_f)}$, which saves a fraction line as well.

Equations must be punctuated in the usual way.

Some further details

(a) Where several sets of brackets occur inside one another in the same formula, the order should be [{0}].
(b) Square roots should be denoted by the sign $\surd$ or the superscript $\frac{1}{2}$, the former being used in simple expressions, the latter in complicated ones.
(c) The range of running variables should be given as in the following example:

$$p_{i+1} = ap_i + bp_{i-1} \qquad (i = 1, \ldots, n).$$

Cyclical components of local rainfall data

R. P. Mentz, M. A. D'Urso, N. M. Jarma and G. B. Mentz

Universidad Nacional de Tucumán and CONICET, Facultad de Ciencias Económicas

Casilla de Correo 209, 4000 Tucumán, Argentina

Received 9 September 1998, Revised 24 April 1999, Accepted 9 May 1999

Abstract

This paper reports on the use of a comparatively simple statistical methodology to study local short time series rainfall data. The objective is to help in agricultural planning, by diminishing the risks associated with some uncertainties affecting this business activity.

The analysis starts by assuming a model of unobservable components, trend, cycle, seasonal and irregular, that is well known in many areas of application. When series are in the realm of business and economics, the statistical methods popularized by the US Census Bureau and US National Bureau of Economic Research are used for seasonal and cyclical estimation, respectively. The flexibility of these methods makes them good candidates to be applied in the meteorological context, and this is done in this paper for a selection of monthly rainfall time series.

Use of the results to help in analysing and forecasting cyclical components is emphasized. The results are interesting. An agricultural entrepreneur, or a group of them located in a single geographical region, will profit by systematically collecting information (monthly in our work) about rainfall, and adopting the scheme of analysis described in this paper.

KEY WORDS: rainfall in Argentina; unobservable components; X-12-ARIMA seasonal procedure; US National Bureau of Economic Research; dating cyclical turning points

1. Introduction

1.1. General

The purpose of this paper is to report on the use of a simple statistical methodology to describe local records of rainfall data coming in short series. The aim is to facilitate agricultural planning by helping to reduce some of the risks associated with the activity.

The records are local since it is desirable to differentiate among geographical locations which are comparatively close to each other in space, in so far as they differ in their climatic experiences. The series to be studied are short, in the order of 15 or 20 years of monthly data, which are shorter than those used in other types of meteorological studies.

The statistical methods that we use arose historically in the treatment of empirical time series. For them, it was found useful to propose models of components associated with distinctive systems of causes: trend, cycles, seasonal and irregular. It is postulated that these unobservable components are combined to form an observable time series, in a multiplicative way, $T \cdot C \cdot S \cdot I$, or in an additive way, $T' + C' + S' + I'$. Other types of combinations have also been considered in the literature. See, for example, Nerlove *et al.* (1979) for an interesting review, that includes early references to applications in astronomy, meteorology, the study of diseases, and others.

Some authors have studied the presence of long-term trends in rainfall data, and tried to relate this to more general issues of global climate change; see, for example, the analysis of annual data in Karl *et al.*

* Correspondence to: Universidad Nacional de Tucuman and CONICET, Facultad de Ciencias Economicas, Casilla de Correo 209, 4000 Tucuman, Argentina.

(1996). However, for rainfall series as short as those considered here, we can safely disregard the effect of trend components. Our rainfall series have large seasonal and random components, and cyclical components. The seasonal component is large, and the variability of the data is different in the different seasons, in the extreme case the dry season contains zeros.

Simple statistical methods to study these components are, for example, moving averages to estimate a trend component, and the so-called 'ratio (or difference) to moving average' to estimate a seasonal component. For series arising in the economics and business areas, these ideas led to the development in the USA of computer programs identified as the X-11 family, which have been used for many years, in various countries and by numerous researchers.

In a manner that parallels these developments, in the US National Bureau of Economic Research (NBER) a procedure was developed and made computationally available, to assign dates to business cycles. It is also true that these methods have been in use for many years, and a considerable amount of experience has been accumulated.

In this paper we analyse monthly rainfall series by using modern versions of the X-11 seasonal procedure and of the NBER business cycle analysis program. One main purpose of our study is to explore how useful these methods are to deal with time series other than those of the economic area, for which they were originally developed. Our basic argument is that for rainfall data (as well as for other meteorological variables), the indicated decomposition has full meaning, and that flexible methods can work reasonably well, in spite of the different nature of the component sources.

We use the seasonal and cyclical procedures in a descriptive way, in the sense that no attribution of physical causes is attempted.

1.2. Some alternative statistical approaches

In the analyses we present in the following sections we show that our choice of statistical procedures is adequate for our purposes. Several other statistical methods have been used in the literature, to deal with various aspects of the description and analysis of rainfall and similar types of series. It is useful to keep in mind two dichotomies: (a) deterministic versus stochastic components or models; (b) time versus frequency domain analyses. Note that those we discuss are not causal procedures; as such, they will differ from causal models developed by meteorologists.

To start our presentation, let us consider a very simple deterministic regression model for monthly rainfall data, formed by the addition of the following components: (i) trend, modelled by a low degree polynomial; (ii) cycle, modelled by a sinusoidal term with a given period; (iii) a stable seasonal pattern represented by 12 constants; and (iv) random independent error terms. This seasonal component can also be written as the sum of sine and cosine functions. In general this approach will prove to be too rigid. Variants to allow for more flexibility include using other smoothing functions of time to model trend, a sum of sinusoidal terms to model the presence of more than one cycle, and time trending seasonal constants.

Non-deterministic trends can be dealt with by smoothing or filtering procedures, one of the most widely used being moving averages, which can also be used in cyclical analysis. A stochastic formulation that is frequently used is that of seasonal autoregressive integrated moving average (ARIMA) models, as described in Box *et al.* (1994); they will take into account trends and seasonal effects. ARMA models can be used in the regression context to give more flexibility to the error terms.

The structure of an observed time series can often be seen more clearly in the frequency domain, by studying the estimated spectral density of a series, or of some of its transformations. There are several statistical procedures to obtain these estimates. The analysis of seasonality and the effect of seasonal adjustments can be evaluated in the frequency domain: spectral density estimates show peaks clearly at seasonal frequencies, even in series with only moderately regular seasonal components (Nerlove, 1964).

Combinations of some of these elements have been used in the literature to study monthly series of rainfall or similar questions. Kashyap and Ramachandra Rao (1976) considered 'the class of autoregressive models with deterministic sinusoidal terms added to them' (p. 249) to model monthly river flows.

They compare their forecasting performances with ARMA and seasonal ARIMA models. These authors present other case studies, some related to rainfall, and compare results of the analysis of daily, monthly and yearly data. However, they do not emphasize cyclical analysis. Percival and Walden (1993) also study river flows among other series, and emphasize the use of spectral methods, including tests of the significance of periodicities. Burroughs (1992) considers studies of cycles based on monthly data, by 'using both spectral analysis and filtering techniques' (p. 39); the results he studies are directly relevant to our purposes, as will be commented on in Section 6.3.

There are many studies of trends and cycles based on annual data, that therefore do not have to deal with the intra-annual seasonal variation. They use several of the types of models and methods we discussed in the section. In general they search for cycles longer than those treated in our work.

Two useful sources of information are the book by Murphy and Katz (1985) and also the article by Gani (1985).

2. Statistical methods

2.1. Seasonal analysis

To perform a seasonal analysis of a given series, we used the X-12-ARIMA (Beta Version 1.0) program of the US Bureau of the Census, made available in 1996 on the Internet (see also Findley *et al.*, 1998). This program uses an iterative semi-automatic procedure that works in stages. In the first stage, an ARIMA model can be attempted for the series, so that if a reasonably good fit is achieved, 1 year of predicted observations are added at one or both ends of the series. Stages two to four consist of the following: a 'trend–cycle' component (trend and cycle are not separated in these programs) is estimated by means of moving averages which are centred at a given observation for central data; this estimate is used to define a detrended series, and this in turn to estimate a seasonal component responding to an additive or multiplicative model, which must be specified by the user of the program. The estimated seasonal component is used to define a seasonally adjusted series. In the second and third stages, the seasonally adjusted series is the input of the following stage, and leads to new estimates of trend–cycle and seasonal with procedures similar to those of the preceding stage. In the fourth and last stage, the program uses the output of the third stage to automatically obtain the final estimates of trend–cycle and seasonal components. In the third and fourth stages, trend–cycle moving averages can be of lengths 9, 13 or 23, a choice is made automatically by the program, but the user can alter this choice.

Program X-12-ARIMA incorporates the so-called 'sliding spans' methodology, which is a diagnostic procedure that compares seasonal estimates obtained from partially overlapping subseries of the given series, and aims at evaluating the reliability of the final estimates by computing and analysing a set of statistical measures.

The complete output of the X-12-ARIMA program for monthly series, contains more than 50 tables, each with $12n$ figures, where n is the number of years that are processed, plus a set of graphs and a set of additional statistical measures. Some of the main results appear in the following tables: B1, original series; D10, final seasonal factors; D11, final seasonally adjusted series; D12, final trend–cycle; D13, final irregular series; F, summary measures; S, sliding spans. It is very convenient for the user that the program allows that the output tables be transferred to independent archives, so that other tables or graphs can be developed using any kind of software.

Processing a series with the X-12-ARIMA program leads to a set of summary measures (table F) to be used to evaluate the quality of the seasonal fit. There are 11 quality control measures, the M-statistics, with values between 0 and 3, and acceptance regions between 0 and 1. Each statistic evaluates a certain characteristic of the seasonal analysis, and gives the user a measure of success along this direction. All measures are combined to define a Q-statistic, a global measure of the quality achieved by the program with the given series.

The X-11 method was described in detail in Shiskin *et al.* (1967). The ARIMA variant of X-11 is described, for example, in Dagum (1983), and the sliding spans methodology in Findley *et al.* (1989). Useful information about seasonal (and cyclical) adjustment procedures can be found in Mentz *et al.* (1989), and in the recent paper with discussion (Ghysels *et al.*, 1996). For a recent detailed description of the final version of the new X-12-ARIMA package, see Findley *et al.* (1998).

The monthly rainfall series used in our study have some values equal to zero, which correspond to months without rain. (However, they do not have long strings of zero values, as they might have in a very dry region.) Under these circumstances, the computer program automatically chooses an additive model. Since the multiplicative option tends to be more popular, we tried it with data transformed to avoid the zero values, by adding a constant. However, the results were poor and are not included in this report.

An analysis of other common transformations was also performed. Since logarithms are not feasible due to the zero values, power transformations were explored. Powers 1/4, 1/2 and 3/4 of the observations were considered, and some results are given in Table I. It is apparent that not much is gained in general by using these transformations. Also, graphs of the estimated seasonal components were studied. The conclusion is that the original observations could be retained without much loss.

One reason for our preoccupation with the quality of the seasonal fit, is that it may affect the cyclical analysis that is carried out with deseasonalized observations. Our conclusion is that if an X-11 type of seasonal adjustment is used, not much is expected to be gained by standard data transformations.

2.2. Cyclical analysis

Once the seasonal analysis is completed, series D11 and D12 are used to perform cyclical analyses, for which we used the program 'Turning Points Determination' made available by the Center for International Business Cycle Research (CIBCR) of Columbia University. It is based on the NBER methodology. This program identifies the maxima and minima of a seasonally adjusted series, and assigns dates to them. In the standard or 'classical' business cycle analysis of economic time series (the option being the growth cycle analysis), this operation is performed with series D11, final seasonally adjusted series; with our rainfall data we tend to rely more on the dates coming from the D12 series, which is 'irregular free'.

Cycles identified by these procedures can be qualified as short cycles, as compared with other lengths considered in the literature for meteorological data. On the other hand, oscillations shorter than 15 months are deemed irrelevant, and are not included into the dated cycles. For a description of these procedures, see, for example, Bry and Boschan (1971) and for applications Mentz *et al.* (1989). Considerations about the duration of the rainfall cycles will be made in Section 6.

After completing the cyclical analysis, the main results appear in tables, some of which facilitate comparisons among locations.

3. Data used in the study

Our initial motivation for this study was to generate information useful for some local agricultural firms, and this explains the nature of the available data. We have rainfall series for four locations in the State of Tucuman, Argentina, three in the central-northern part of the state, chosen with reference to the Agricultural Experimental Station providing the information, one (Las Cejas) is usually regarded as drier than the other (La Ramada). The fourth location in Tucuman (La Cocha) is in the southern-most part of the state. For purposes of comparison, data from two other states are considered, Santa Fe and La Pampa, which are part of the traditional Argentine prairies.

Table I. Seasonal analysis, rainfall data, additive model, X-12-ARIMA: figures of merit

Locations	Period	Transformation	*F*-statistics analysis			Month of cyclical dominance	M-statistics analysis				Sliding spans analysis			
			Stable seasonality	Moving seasonality	Seasonal fit is:		Number of M>1	Detail	*Q*-statistic	Seasonal fit is:	*S* (%)	MM (%)	Seasonal fit is:
Agricultural Experimental Station Obispo Colombres (Tucuman)	1975–1995	Original	69.36	1.28	Stable 0.1%	12	1	M5	0.71	Accepted	91.7	98.2	Unlikely
	1975–1995	1/2 power	71.90	0.99	Stable 0.1%	12	2	M3, M5	0.91	Cond. accep.	52.8	77.8	Unlikely
	1975–1995	1/4 power	71.90	0.99	Stable 0.1%	12	2	M3, M5	0.91	Cond. accep.	55.6	77.8	Unlikely
	1975–1995	3/4 power	63.83	1.06	Stable 0.1%	12	2	M3, M5	0.94	Cond. accep.	92.6	87	Unlikely
La Ramada (Tucuman)	1975–1989	Original	39.02	1.25	Stable 0.1%	12	4	M1, M2, M3, M5	1.01	Cond. reject.	91.7	84.7	Unlikely
	1975–1989	1/2 power	57.44	0.88	Stable 0.1%	12	4	M1, M2, M3, M5	1.05	Cond. reject.	77.1	95.1	Unlikely
	1975–1989	1/4 power	57.44	0.88	Stable 0.1%	12	4	M1, M2, M3, M5	1.05	Cond. reject.	77.1	95.1	Unlikely
	1975–1989	3/4 power	49.78	1.05	Stable 0.1%	12	4	M1, M2, M3, M5	1.01	Cond. reject.	96.5	88.9	Unlikely
Las Cejas (Tucuman)	1975–1989	Original	36.54	2.24	Stable 0.1%	12	2	M3, M5	0.83	Cond. accep.	90.7	83.3	Unlikely
	1975–1989	1/2 power	52.36	1.89	Stable 0.1%	12	5	M1, M2, M3, M5, M6	0.99	Cond. reject.	71.1	86.8	Unlikely
	1975–1989	1/4 power	57.44	0.88	Stable 0.1%	12	4	M1, M2, M3, M5	1.05	Cond. reject.	75.7	95.1	Unlikely
	1975–1989	3/4 power	42.48	2.31	Stable 0.1%	12	5	M1, M2, M3, M5, M6	0.99	Cond. reject.	96.5	82.6	Unlikely
Agricultural Experimental Station Obispo Colombres (Tucuman)	1982–1995	Original	39.67	0.78	Stable 0.1%	12	1	M5	0.71	Accepted	92.6	89.8	Unlikely
	1982–1995	1/2 power	51.62	0.32	Stable 0.1%	12	2	M1, M3	0.93	Cond. accep.	77.1	77.8	Unlikely
	1982–1995	1/4 power	51.47	0.33	Stable 0.1%	12	2	M3, M5	0.99	Cond. accep.	62.0	65.7	Unlikely
	1982–1995	3/4 power	43.55	0.54	Stable 0.1%	12	2	M3, M5	0.96	Cond. accep.	92.6	87.0	Unlikely
La Cocha (Tucuman)	1982–1995	Original	13.18	0.98	Stable 0.1%	12	4	M1, M2, M3, M5	1.23	Rejected	95.8	95.1	Unlikely
	1982–1995	1/2 power	23.22	1.06	Stable 0.1%	12	6	M1, M2, M3, M5, M10, M11	1.25	Rejected	90.7	94.4	Unlikely
	1982–1995	1/4 power	20.11	1.89	Stable 0.1%	12	6	M1, M2, M3, M5, M10, M11	1.23	Rejected	62.0	91.7	Unlikely
Santa Fe	1982–1995	3/4 power	19.82	9.6	Stable 0.1%	12	4	M1, M2, M3, M5	1.31	Rejected	97.9	93.8	Unlikely
	1975–1989	Original	7.18	1.01	Stable 0.1%	12	7	M1, M2, M3, M5, M8, M10, M11	1.50	Rejected	94.4	95.4	Unlikely
	1975–1989	1/2 power	9.70	1.01	Stable 0.1%	12	7	M1, M2, M3, M5, M8, M10, M11	1.52	Rejected	86.1	94.4	Unlikely
	1975–1989	1/4 power	9.01	1.16	Stable 0.1%	12	7	M1, M2, M3, M5, M8, M10, M11	1.35	Rejected	21.3	88.9	Unlikely
	1975–1989	3/4 power	8.53	0.95	Stable 0.1%	12	7	M1, M2, M3, M5, M8, M10, M11	1.49	Rejected	94.4	95.4	Unlikely
La Pampa	1975–1989	Original	13.47	0.71	Stable 0.1%	12	4	M1, M2, M3, M5	1.28	Rejected	95.1	92.4	Unlikely
	1975–1989	1/2 power	16.34	0.65	Stable 0.1%	12	4	M1, M2, M3, M5	1.30	Rejected	83.3	93.8	Unlikely
	1975–1989	1/4 power	12.64	0.78	Stable 0.1%	12	4	M1, M2, M3, M5	1.37	Rejected	10.4	84.0	Unlikely
	1975–1989	3/4 power	15.42	0.66	Stable 0.1%	12	4	M1, M2, M3, M5	1.30	Rejected	98.6	94.4	Unlikely

A summary of the available information is as follows:

State	Location	Period	Average monthly rainfall (mm)
Tucuman	Agricultural Experimental Station (EEAOC)	1975–1995	93.2
	Agricultural Experimental Station (EEAOC)	1982–1995	88.8
	La Ramada	1975–1989	96.0
	Las Cejas	1975–1989	69.5
	La Cocha	1982–1995	81.1
Santa Fe		1975–1989	74.5
La Pampa		1975–1989	70.5

The first series has 21 years of monthly data, while the remaining series are 14- or 15-years long. The second series is just a segment of the first one.

4. Main findings on seasonal and cyclical analyses

1. Results of the exploration of certain power transformations under the additive version of the X-12-ARIMA seasonal program, are given in Table I. For each series, the results of processing the original and the power transformations for 1/2, 1/4 and 3/4 are compared. This analysis leads us to retain the original data for analysis, since differences are small.

2. As indicated in Section 1.1, rainfall data are known to be highly variable, so that in our terms, controlling or eliminating the irregular component is expected to be a source of difficulty. Our results, in terms of the goodness of fit statistics, confirm this general idea.

3. About the goodness of fit statistics mentioned in Section 2, high values for the *F*-statistic for the presence of stable seasonality, and low values for the *F*-statistic for the presence of moving seasonality, are an indication of a strong and comparatively regular seasonal component. Hence, the part of Table I containing these *F*-statistics shows in general that our series are satisfactory in this respect.

An often-used indicator of the quality of a seasonal adjustment of a series, is the so-called 'month of cyclical dominance' (MCD), which comes from relating the size of the irregular to that of the trend–cycle component: a large MCD means that the movements of the irregular component are large enough to make difficult a reliable estimation of trend–cycle. In our work we often find values higher than 6 for MCD, when small values are 1 or 2, with 12 being the largest possible. In Table I, all reported MCDs are equal to 12, which is the least favourable value.

We note that the number of M-statistics that exceed the threshold value of 1 varies from 1 to 7. The summary measure *Q*, a weighted average of the 11 M-statistics, has values either smaller than 1 or slightly larger. Further, the M7 statistic is never found to be larger than 1: 'Values of M7 larger than 1 are usually interpreted as evidence that the seasonal component is evolving too rapidly for the X-11 procedure' (Findley, 1996).

Finally, Table I shows the measures associated with the sliding spans, which indicate uniformly that the seasonal fits are considered poor. This points to difficulties in trying to predict or extrapolate seasonal effects.

4. In view of the preceding comments, we processed with the turning points program, not only the seasonally adjusted series (D11), but also the final trend–cycle series (D12) in which the irregular has been smoothed out.

5. A graphical analysis associated with the first (and longest) series mentioned in Section 3 is incorporated as an example in Figures 1–4. Graphs are, in terms of the definitions introduced in Section 2, respectively, those of series B1, D10, D11, and D12 with its dated turning points and superimposed average (mean) line.

6. For the same series, Table II summarizes the distances among dates presented in the graph of series D12. The mean, median and standard deviation of the cyclical durations are presented. Cycle durations are measured peak–trough–peak and trough–peak–trough, which in cases of asymmetries need not coincide.

7. Table III facilitates the comparison among the available series, of the average durations of cycles estimated in the indicated way.

8. For the series reported in Table II, Table IV summarizes the information in the graphs of series D12 in a different way: cycles are now defined as comprising consecutive periods of permanence of the series over and under the average line, which is shown in all graphs. It is thought that this is a manner reasonably easy to understand, to exhibit the results of the analysis.

9. Besides the preceding analysis, there is a good deal of information in the graph of series D12. For example, comparing the graphs for La Cocha (Tucuman) in Figure 5, where the risks associated with the behaviour of rainfall are well known, with La Pampa in Figure 6, part of the traditional Argentine

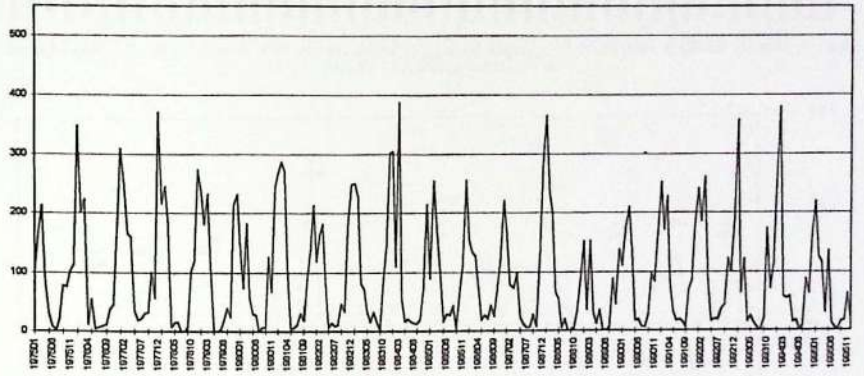


Figure 1. Monthly rainfall in Agricultural Experimental Station (Tucuman). January 1975–December 1995. Original series

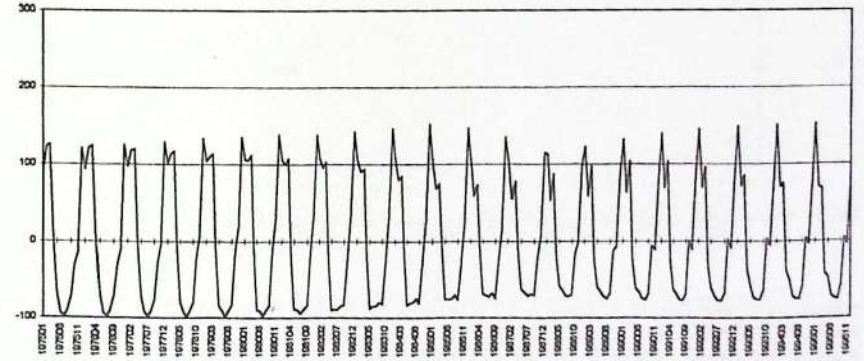


Figure 2. Monthly rainfall in Agricultural Experimental Station (Tucuman). January 1975–December 1995. Seasonal factor, additive model, X-12-ARIMA

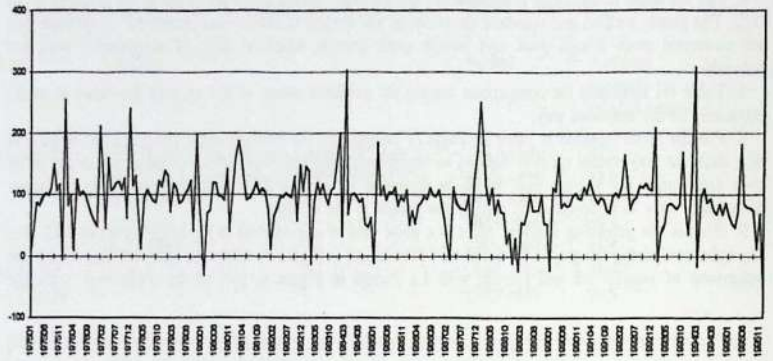


Figure 3. Monthly rainfall in Agricultural Experimental Station (Tucuman). January 1975–December 1995. Seasonal adjusted series, additive model, X-12-ARIMA

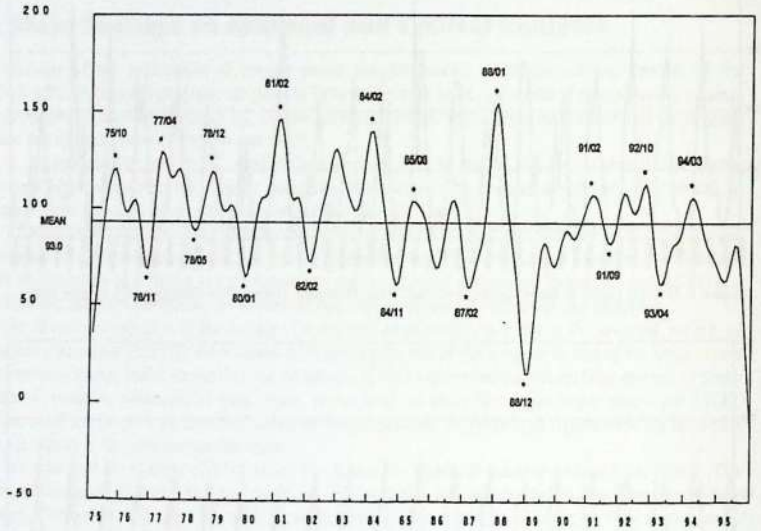


Figure 4. Monthly rainfall in Agricultural Experimental Station (Tucuman). January 1975–December 1995 trend–cycle component, additive model, X-12-ARIMA, dated by CIBCR procedure, and superimposed mean line

prairies, we observe that the difference is not so much in the duration of the observed cycles as in the amplitude of the movements of the series D12. Since the two graphs are drawn in the same scale, we conclude that the risks associated with the cyclical behaviour of rainfall are much higher in La Cocha than in La Pampa, in agreement with experience. In fact, La Pampa is, in a large proportion, a part of an important area known precisely as 'la pampa húmeda' (humid prairies).

5. Forecasting

In trying to forecast one of our rainfall series, attention must be paid to its seasonal, trend and cyclical components. The estimated seasonal components comprise table D10 of the output in the X-12-ARIMA program, and appears in Figure 2 for one of the series. The corresponding analysis was presented in Section 4. The X-12-ARIMA program also provides a 'year ahead' set of 12 estimates of the seasonal component, that could be used in forecasting. In our case we recall a warning coming from the sliding spans measures of fit (see Section 3).

More interesting is trying to forecast the other components. Trends and cycles are treated jointly in the X-12-ARIMA program as the 'trend–cycle' component. The estimates appear in table D12, and are shown in Figures 4–6 and 11 for various series. We recall that we disregard the contribution of trend, in view of the limited lengths of our series. Hence, we take the indicated estimates as representing cyclical components only. We aim at forecasting these by extending or extrapolating the corresponding series.

Table II. Monthly rainfall in Agricultural Experimental Station Obispo Colombres (Tucuman), January 1975–December 1995. Final turning points of trend–cycle component, additive model, X-12 ARIMA dated by CIBCR procedure

(a) Peak–trough–peak analysis

Peaks		Troughs		Peaks		Duration in months		
Year	Months	Years	Months	Years	Months	Trough–peak	Peak–trough	Total cycle
1975	10	1976	11	1977	4	13	5	18
1977	4	1978	5	1978	12	13	7	20
1978	12	1980	1	1981	2	13	13	26
1981	2	1982	2	1984	2	12	24	36
1984	2	1984	11	1985	6	9	7	16
1985	6	1987	2	1988	1	20	11	31
1988	1	1988	12	1991	2	11	26	37
1991	2	1991	9	1992	10	7	13	20
1992	10	1993	4	1994	3	6	11	17
1994	3							
Means						11.6	13.0	24.6
Medians						12.0	11.0	20.0
Standard deviations						4.1	7.4	8.2

(b) Trough–peak–trough analysis

Troughs		Peaks		Troughs		Duration in months		
Years	Months	Year	Months	Years	Months	Peak–trough	Trough–peak	Total cycle
1976	11	1977	4	1978	5	5	13	20
1978	5	1978	12	1980	1	7	13	26
1980	1	1981	2	1982	2	13	12	36
1982	2	1984	2	1984	11	24	9	16
1984	11	1985	6	1987	2	7	20	31
1987	2	1988	1	1988	12	11	11	37
1988	12	1991	2	1991	9	26	7	20
1991	9	1992	10	1993	4	13	6	17
1993	4	1994	3			11		
Means						13.0	11.4	25.4
Medians						11.0	11.5	23.0
Standard deviations						7.4	4.4	8.4

Table III. Final turning points of trend-cycle components, rainfall data, additive model, X-12-ARIMA. Dated by CIBCR procedure

Locations	Period	Peak–trough–peak			Trough–peak–trough		
		Trough–peak	Peak–trough	Total cycle	Peak–trough	Trough–peak	Total cycle
Agricultural Experimental Station Obispo Colombres	1975–1995	11.4	15.0	24.6	13.0	11.4	25.4
La Ramada (Tucuman)	1975–1989	15.6	15.8	32.8	15.8	16.5	32.3
Las Cejas (Tucuman)	1975–1989	10.7	13.1	23.9	13.1	11.2	25.5
Agricultural Experimental Station Obispo Colombres	1982-1995	11.3	16.7	24.4	13.4	11.3	26.5
La Cocha (Tucuman)	1982–1995	14.2	14.2	28.4	14.2	15.3	30.3
Santa Fe	1975–1989	12.8	11.0	24.8	12.8	12.8	25.7
La Pampa	1975–1989	16.4	16.5	28.5	14.8	16.4	31.2

Table IV. Monthly rainfall in Agricultural Experimental Station Obispo Colombres (Tucuman), January 1975–December 1995. Final turning points of trend-cycle component, additive model, X-12-ARIMA dated by CIBCR procedure. Peak–trough–peak analysis referred to the mean

Over the mean				Under the mean				Months of rainfall		
Start		End		Start		End		Over the mean	Under the mean	Total cycle
Year	Month	Year	Month	Year	Month	Year	Month			
1975	5	1976	7	1976	8	1976	12	14	4	18
1977	1	1978	3	1978	4	1978	6	14	2	16
1978	7	1979	9	1979	10	1980	4	14	6	20
1980	5	1981	10	1981	11	1982	5	17	6	23
1982	6	1984	7	1984	8	1985	3	25	7	32
1985	4	1985	10	1985	11	1986	6	6	7	13
1986	7	1986	10	1986	11	1987	8	3	9	12
1987	9	1988	5	1988	6	1990	9	8	27	35
1990	10	1991	6	1991	7	1991	11	8	4	12
1991	12	1992	12	1993	1	1993	11	12	10	22
1993	12	1994	5	1994	6	1995	10	5	16	21
Mean								11.5	8.9	20.4
Medians								12.0	7.0	20.0
Standard deviations								6.3	7.1	7.6

To do this, we use the information given, for example, in Tables II and IV for one of the series. We take as our main result the estimated average duration of the cycles. To agree with our purposes, we resort to the duration reported in Table IV, which is based on periods of rainfall over or under the mean. This average duration of cycles may differ slightly from those in Table II. At the end of a period of observation (in our examples it is always the month of December), if the last dated turning point is a

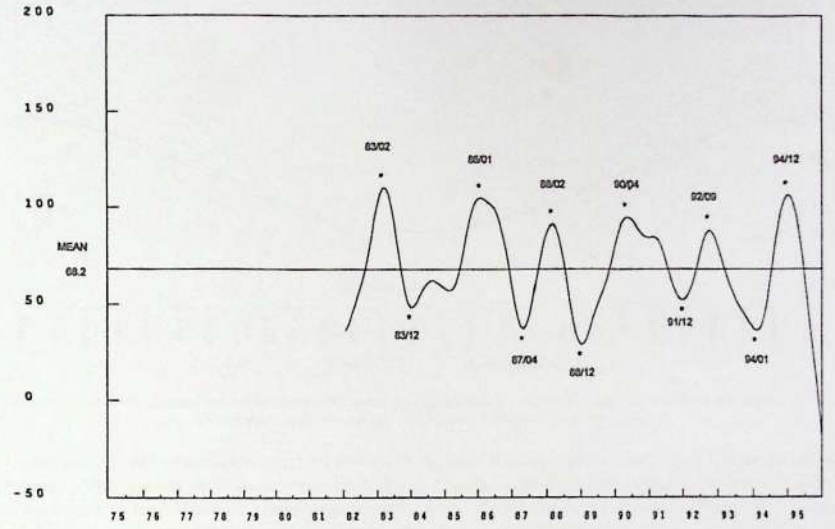


Figure 5. Monthly rainfall in La Cocha (Tucuman). January 1982–December 1995. Trend–cycle component, additive model, X-12-ARIMA, dated by CIBCR procedure, and superimposed mean line

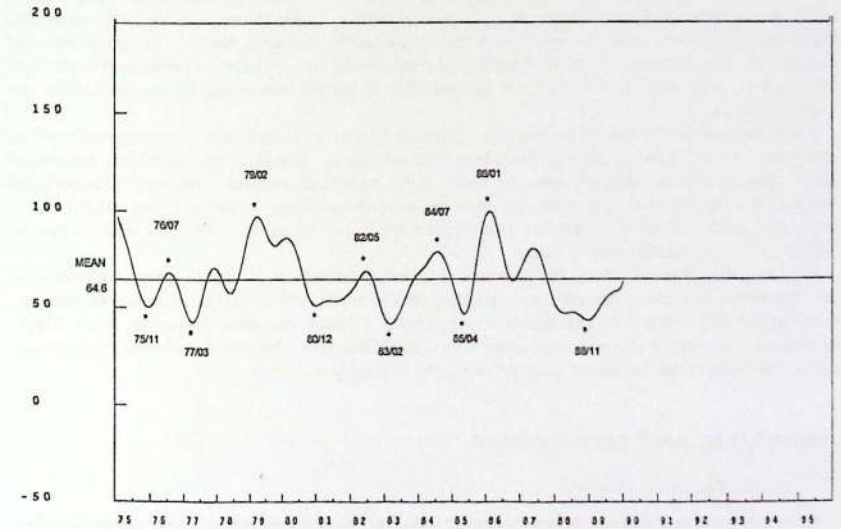


Figure 6. Monthly rainfall in La Pampa. January 1975–December 1989. Trend–cycle component, additive model, X-12-ARIMA, dated by CIBCR procedure, and superimposed mean line

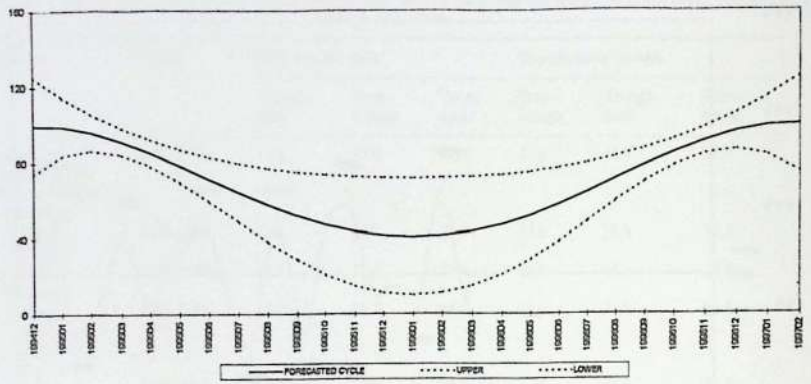


Figure 7. Monthly rainfall in La Cocha (Tucuman), January 1982–December 1995, forecasted cycle December 1994–February 1997 and upper and lower 90% confidence limits (fourth degree polynomial)

peak, we extend from there for the estimated average duration of a cycle, and consider that in this period a full cycle occurs ending in another dated peak. A parallel argument is used when the last dated turning point is a trough. In another example, for the series appearing in Figure 5 we add a new cycle: this is shown in Figure 8, and is calculated by the procedure we describe below. This will then indicate in which months to expect extra high or low levels of rain, in addition to the seasonal pattern. This is what we claim is of use in agricultural planning.

A 'forecasted cycle' is defined by the following steps: (i) two points are computed by averaging amplitudes at peak dates and trough dates; (ii) one of these is considered twice, as end of a cycle and beginning of the next one; (iii) two points are computed by averaging dates at which periods of permanence over the mean are started and ended, respectively; (iv) to these five points polynomials of degrees four, three and two are fitted and analysed. The forecasted cycle is that showing best fit on an inspection basis.

A further question is how to measure the reliability of this forecasted cycle. A confidence band is computed by the following steps: (i) confidence intervals are set at each of the five points mentioned above, considering that three of them are means of the amplitudes of certain quantities of peaks and troughs, and the other two are means of durations as those calculated in Tables II and III; (ii) for a confidence coefficient of 90%, the Bonferroni procedure is used to set the confidence band, with the t-table used to extract the needed values.

For La Cocha, Figure 7 shows the forecasted cycle which extends for 27 months (26.6), beginning at the observed peak of December 1994, and extending until February 1997, with a fourth degree fit. We also show upper and lower 90% confidence limits. Figure 8 shows this same forecasted cycle drawn immediately after the trend–cycle component values reach December 1995. We note that we took a very simple approach to the important issue of timing the beginning of a cycle.

6. Analysis and conclusions

6.1. Comparing locations

An interesting question is how to compare results obtained for different locations. Two attempts in this direction appear in Figures 9 and 10. Figure 9 is based upon the following idea: to stress the oscillatory nature of the cycle component, graphs associated with it (e.g. Figure 4) are often presented with

superimposed shaded strips covering the periods from one dated peak to the following dated trough. In these periods we have decreasing parts of the graph. This is usually performed in business cycle analysis, and will be used here for purposes of comparison.

In Figure 9 the dates assigned by the CIBCR cycle procedure to the turning points are compared for six locations. The shaded areas for a given location, correspond to periods of decreasing rainfall, from peak to trough. For example, for the series EEAOC (1975–1995), its diagram is an abridgement of Figure 4: most details are omitted, and only the dates appearing in Table II are retained.

Two series are said to coincide in a given month, if both are going from (dated) peaks to troughs, or both from troughs to peaks. Figure 9 shows a good level of coincidence among the four locations in Tucuman in the period when we have data for all of them (1982–1989), and among three locations for another period (1975–1982). Santa Fe tends to coincide to some extent, but La Pampa shows larger differences.

Figure 10 is based on the number of months of coincidence for the series in Figure 9, except La Pampa. The measure reported is the proportion of the absolute difference between the number of series that coincide and those that differ in a given month. For five series these differences can be 1 (three of a kind and two of the other), 3 (four of a kind and one of the other) or 5 (all of the same kind), and hence the proportions can be 0.20, 0.60 or 1.

6.2. Reference to El Niño

In studying meteorological variables, it pays to consider what is known as the El Niño phenomenon. We have information about the Southern Oscillation Index (SOI), which is defined as the normalized difference (i.e. difference divided into its standard deviation) between the pressure anomalies (monthly means minus long-term means) in two locations, Tahiti and Darwin, as defined by Troup (1965).

Information about El Niño is frequently used in meteorological studies in the place of a possible causal factor. In our case, and in view of the descriptive approach used throughout, we will only apply to the SOI series the same kind of methods we have discussed so far.

Figure 11 presents the final result of applying to the SOI monthly series, from January 1975 to September 1997, the seasonal and cyclical methods described in the previous sections, and therefore it is the estimated trend–cycle component dated by the CIBCR procedure. The mean duration of the cycle is

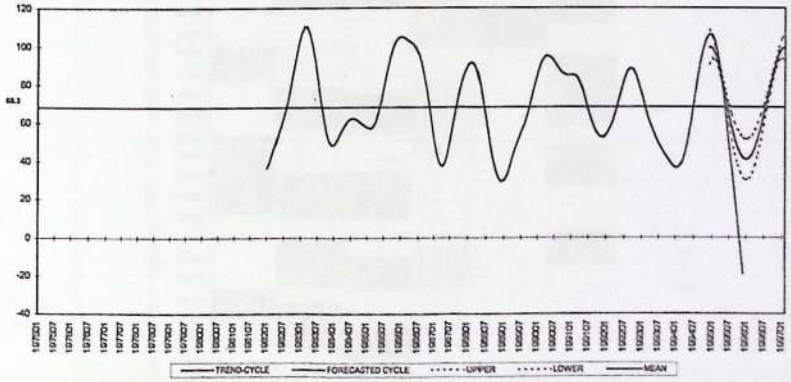


Figure 8. Monthly rainfall in La Cocha (Tucuman). January 1982–December 1995. Trend–cycle component, additive model, X-12-ARIMA, dated by CIBCR procedure. Superimposed mean line, forecasted cycle (fourth degree polynomial). December 1994–February 1997

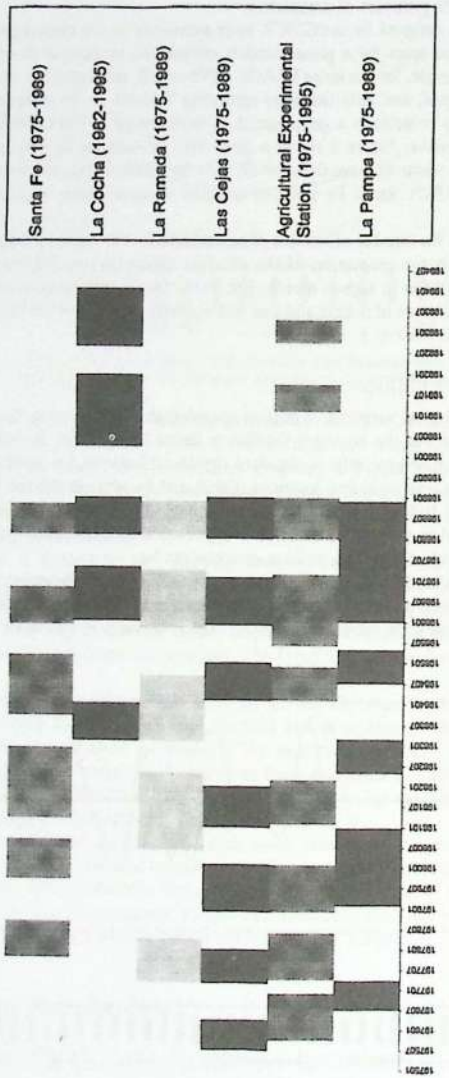


Figure 9. Final turning points of trend-cycle component of six locations, January 1975-December 1995; comparison of dates assigned by CIBCR procedure. Shaded areas correspond to periods from peak to trough

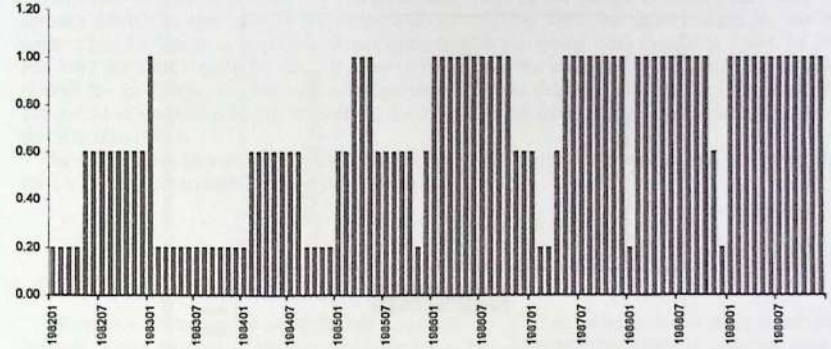


Figure 10. Proportion of coincidences among five series in Figure 9

35.7 months for peak-trough-peak and 34.2 months for trough-peak-trough, and the corresponding median durations are 32 months in both cases. The means are comparable with those in Table III, being very slightly larger.

In general, and in relation to our cyclical analysis of rainfall in the locations that we considered, we expect some degree of relation between the SOI and rainfall in La Pampa. The direction of this relation is that the lower value of the SOI should coincide with periods of high rainfall in La Pampa.

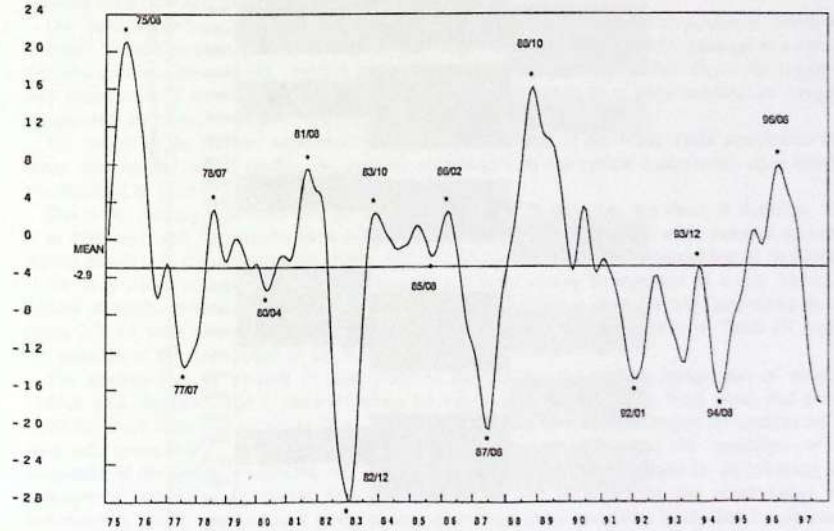


Figure 11. Monthly values of SOI January 1975-September 1997. Trend-cycle component, additive model, X-12-ARIMA, dated by CIBCR procedure, and superimposed mean line

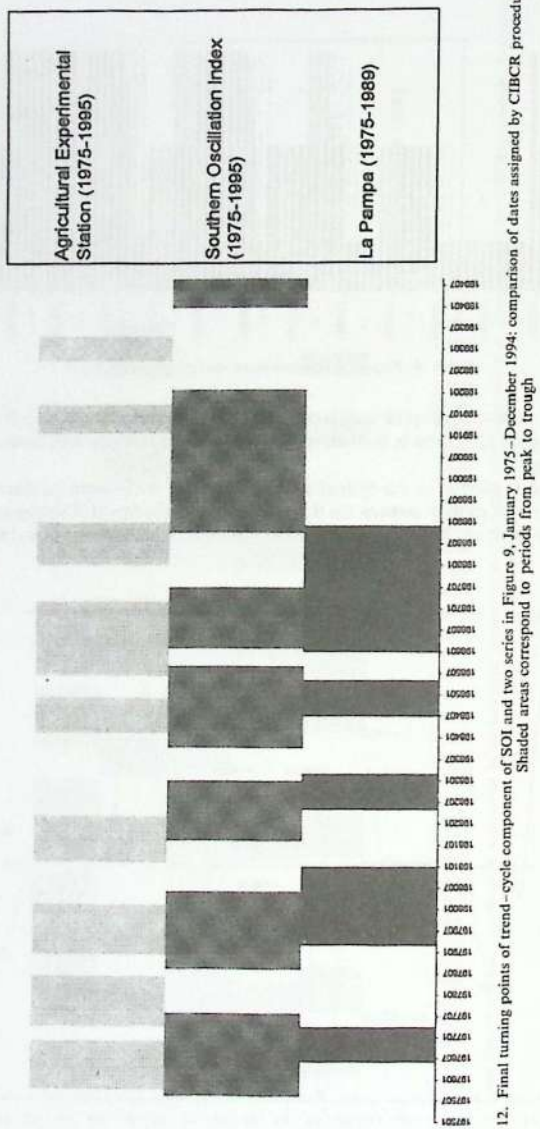


Figure 12. Final turning points of trend–cycle component of SOI and two series in Figure 9, January 1975–December 1994: comparison of dates assigned by CIBCR procedure. Shaded areas correspond to periods from peak to trough

The dates assigned to the SOI are compared with those of La Pampa in Figure 12, where Tucuman's EEAOC is also included for comparison. In 1975 and 1985, the dated troughs for the SOI occur when La Pampa is in periods of increasing rainfall (i.e. going from trough to peak). In 1980 and 1982 the dated troughs for the SOI occur shortly before the beginning of the period of increasing rainfall for La Pampa. We find reasonable agreement of these findings with what was expected. The last period of decreasing rainfall reported for La Pampa covers from 1986 to 1988, and its relation to the SOI is less clear.

The coincidences between the SOI and Tucuman's EEAOC is much less regular and interesting than for La Pampa, in agreement with what is expected.

6.3. Final comments

Information about weather variables and conditions are of great importance for agricultural management. Good forecasts of weather conditions are of real value since they help in reducing some of the risks associated with the activity. Good forecasts may be used to program yearly agricultural activities in such a way, that over a period of several years this anticipatory behaviour of the programmers leads to important benefits, in comparison with a traditional program of some sort of average behaviour over the years.

One important component of weather variables is the amount of rainfall, which in our case has been observed monthly. Other (shorter) periods could also be considered, as has often been done in practice. In our case we consider the description and analysis of monthly rainfall data for local areas, understanding that, farms or groups of them may differ in weather conditions, even when they are not too far apart. This has also been considered by other investigators, leading to the definition of areas of certain radius, so that homogeneity of weather conditions holds within areas, and heterogeneity between them. We have not explored this here.

The methods we have used deal, separately, with the seasonal and cyclical components, which are assumed to enter as unobservables in a given time series. The new X-12-ARIMA package is a flexible procedure for seasonal analysis, that can be very useful in the analysis of rainfall. Given the regularity and importance of the seasonal components, even in the presence of a large random or irregular component, we expect reasonable results of the proposed seasonal analysis.

The output of the seasonal adjustment program is an estimate of the trend–cycle component of a series, that for the case of rainfall can be safely associated with the cyclical component, since trend is not expected to be of practical significance.

The cycles that are considered through the use of the CIBCR program, are short in duration. This is in agreement with the intuitive ideas behind our work, and is in agreement with standard meteorological analysis. In effect, Burroughs (1992), section 3.12, writes: 'The QBO (quasi-biennial oscillation) is the most widely observed feature in the records, and must clearly be regarded as a real feature of almost all meteorological records'. The identification of the QBO is used for any periodicity in the range 2.2–2.8 years (26.4–33.6 months), and is then consistent with our findings in Table IV and in the duration of the cycles found in the SOI.

The analysis that we propose is then useful in interpreting the cyclical component of monthly rainfall data. We have come to the point where we believe that for one of the local areas, and having available short series (say, of up to 20 years of monthly data), a systematic statistical analysis of the data will prove useful. The average duration of the estimated cycles, and the inspection of the amplitude of the cyclical oscillations, will be useful indications for those engaged in the planning and management aspects of agriculture. One drawback lies in the variability of the observations: our indicators of quality tend to detect in the present cases even more variability than that for economic series processed for Argentina. Hence, a word of caution in the interpretation of the results should be expressed, as we pointed out in various parts of this work.

Acknowledgements

This study was started with data for La Cocha collected by Lic. Antonio Benito, whose interest in interpreting the results of our calculations we greatly appreciate. The rainfall series for the other locations were provided by Eng. Cesar Lamelas, meteorologist from EEAOC, whose interest in our work we sincerely appreciate. We thank Dr Martin Grondona for helpful comments and information about El Niño, and Professor Pedro Morettin for help with the references. We also thank two referees and the editors for many comments that we used to improve our presentation.

References

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. 1994. *Time Series Analysis: Forecasting and Control*, 3rd Edition. Prentice Hall, New Jersey.

Bry, G. and Boschan, C. 1971. *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*. US National Bureau of Economic Research, New York.

Burroughs, W.J. 1992. *Weather Cycles Real or Imaginary?* Cambridge University Press, Cambridge.

Dagum, E.B. 1983. *The X-11-ARIMA Seasonal Adjustment Method*. Statistics Canada, Ottawa.

Findley, D.F. 1996. 'Comment', *J. Business Econ. Stat.*, 14(3), 389–393.

Findley, D.F., Monsell, B.C. and Shulman, H.B. 1989. 'Sliding spans diagnostics for seasonal and related adjustments', in Mentz, R.P. *et al.* (eds), *Statistical Methods for Cyclical and Seasonal Analyses*. IASI, Panama.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C. and Chen, B.-C. 1998. 'New capabilities and methods of the X-12-ARIMA seasonal-adjustment program', *J. Business Econ. Stat.*, 16(2), 127–177 (with discussion).

Gani, J. 1985. 'Statistics in meteorology', in Kotz, S. and Johnson, N.L. (eds-in-chief), *Encyclopedia of Statistical Sciences*, volume 5. Wiley, New York, pp. 453–460.

Ghysels, E., Granger, C.W.J. and Siklos, P.L. 1996. 'Is seasonal adjustment a linear or nonlinear data-filtering process?', *J. Business Econ. Stat.*, 14(3), 374–397 (with discussion).

Karl, T.R., Knight, R.W., Easterling, D.R. and Quayle, R.G. 1996. 'Indices of climate change for the United States', *Bull. Meteorol. Soc.*, 77(2), 279–292.

Kashyap, R.L and Ramachandra Rao, A. 1976. *Dynamic Stochastic Models from Empirical Data*. New York, Academic Press.

Mentz, R.P., de Alba, E., Espasa, A. and Morettin, P.A. (eds) 1989. *Statistical Methods for Cyclical and Seasonal Analyses*. Inter-American Statistical Institute (IASI), Panama.

Murphy, A.H. and Katz, R.W. 1985. *Probability, Statistics and Decision Making in the Atmospheric Sciences*. Westview Press, Boulder, CO.

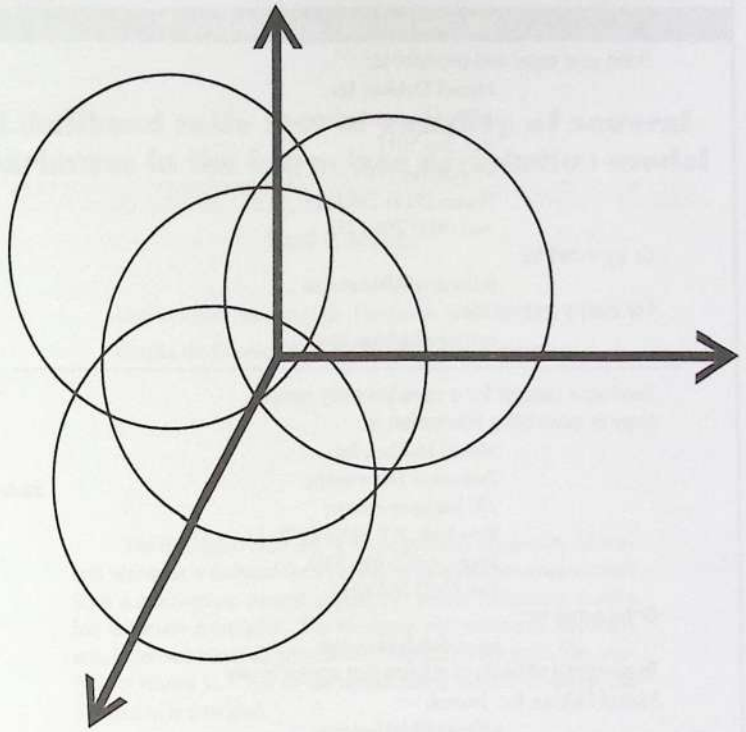
Nerlove, M. 1964. 'Spectral analysis of seasonal adjustment procedures', *Econometrica*, 32, 241–286.

Nerlove, M., Grether, D.M. and Carvalho, J.L. 1979. *Analysis of Economic Time Series*. Academic Press, New York.

Percival, D.B. and Walden, A.T. 1993. *Spectral Analysis for Physical Applications: Multitaper and Coventional Univariate Techniques*. Cambridge University Press, Cambridge.

Shiskin, J., Young, A.H. and Musgrave, J.C. 1967. *The X-11 Variant of the Census Method. II. Seasonal Adjustment Program*. Bureau of the Census, Washington, DC.

Troup, A.J. 1965. 'The southern oscillation', *Q. J. R. Meteorol. Soc.*, 91, 490–506.



Communications in Statistics

Theory and methods

Likelihood ratio test of equality of several variances in the intraclass correlation model

Raúl P. Mentz

Universidad Nacional de Tucumán and CONICET
Casilla de Correo 209, 4000 Tucumán, Argentina

Abstract

The likelihood ratio test of the hypothesis of equality of several variances is deduced for a sample of independent observations from a multivariate normal population whose covariance matrix has intraclass correlation. The resulting test, related to Bartlett's test, is derived when all population correlations are zero. The analysis is related to a test in the mixed linear model. A numerical illustration is provided.

Key Words: Likelihood ratio test; Equality of variances; Intraclass correlation; Bartlett's test; Asymptotic χ^2 distribution.

1. Introduction

The well-known Bartlett's test of equality of several variances in the case of independent univariate sampling from normal populations, is a modification of the likelihood ratio procedure, and except for constants, amounts to comparing the

*E-mail: rmentz@herrera.unt.edu.ar

arithmetic and geometric means of the unbiased estimators of the variances. In this paper a generalization of the likelihood ratio test is presented, in which samples are not independent among themselves, but respond to an intraclass correlation structure. This means that while observations within each sample are independent, pairs of observations in any two samples, with the same index, are correlated, and the correlation coefficients for the various indices are equal.

An application of this situation was done in a simulation study: several statistical methods were to be compared, and in order to control variation, the same sample was used with all methods in a given replication. It then follows that for a given number indexing the replication, the observations are dependent, and it is reasonable to postulate that the correlation between pairs of methods are equal. See (1), where the case of testing for the equality of two variances is presented.

The intraclass correlation model, in particular with homogeneous variances, has a long history in statistics. Referring to problems in educational research, Olkin (2) says that it is "...perhaps the fundamental model in test construction." He presents several inferential techniques about the intraclass correlation coefficient, including a "conventional estimate" going back to (3,4). The model with homogeneous variances appears in the mixed-effects linear procedure, as will be discussed hereafter.

For further general information, see the entries of "Bartlett's test of homogeneity of variances" (5), and "Intraclass correlation coefficient" (6) in the *Encyclopedia of Statistics*.

2. Main results

We consider the following set up: a k-variate random vector $\mathbf{X}$ has a normal distribution with mean vector $\mu = (\mu_1, \ldots, \mu_k)'$ and positive definite covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \cdots & \rho\sigma_1\sigma_k \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho\sigma_2\sigma_k \\ \vdots & \vdots & & \vdots \\ \rho\sigma_1\sigma_k & \rho\sigma_2\sigma_k & \cdots & \sigma_k^2 \end{pmatrix} \tag{1}$$

that is, variances σ_j^2, $j = 1, \ldots, k$, and correlation coefficients $\rho_{ij} = \rho, i, j = 1, \ldots, k, i \neq j$. With a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of i.i.d. observations from this distribution, it is desired to test

$$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2, \tag{2}$$

against the alternative that at least one of the variances is different.

Lemma 1. *For Σ defined in Equation (1) with $-(1-k)^{-1} < \rho < 1$ and $\sigma_j > 0$,*

$$\Sigma = \mathbf{D}^{1/2}\,\mathbf{P}\,\mathbf{D}^{1/2}, \tag{3}$$

where $\mathbf{D}$ is diagonal with diagonal elements $\sigma_1^2, \ldots, \sigma_k^2$, $\mathbf{D}^{1/2}$ is diagonal with diagonal elements $\sigma_1, \ldots, \sigma_k$,

$$\mathbf{P} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} = (1-\rho)\mathbf{I} + \rho\,\mathbf{J}, \tag{4}$$

and $\mathbf{J}$ is a $k \times k$ matrix of 1's. Then,

$$|\mathbf{P}| = (1-\rho)^{k-1}[1 + (k-1)\rho], \tag{5}$$

$$\mathbf{P}^{-1} = \frac{1}{1-\rho}\left[\mathbf{I} - \frac{\rho}{1+(k-1)\rho}\mathbf{J}\right] \tag{6}$$

and $\mathbf{P}$ has characteristic roots $1 + (k-1)\rho$ and $(1-\rho)$ with multiplicity $k-1$.

These results follow, for example, from (7) or (8).
Note that $\mathbf{P}$ is a Toeplitz matrix, that can be written as

$$\mathbf{P} = \mathbf{I} + \sum_{j=1}^{k-1} \rho\,\mathbf{G}_j, \tag{7}$$

where $\mathbf{G}_j$ has 1's along the two diagonals parallel to the main diagonal at distances equal to j. Also note that $\mathbf{J}$ is singular, of rank 1, with nonzero characteristic root equal to k.

Theorem 1. *Under the stated conditions on μ and Σ given by Equation (1), the unrestricted maximum likelihood estimators are*

$$\hat{\mu}_j = \bar{X}_j = \frac{1}{n}\sum_{\alpha=1}^{n} X_{\alpha j}, \quad \hat{\sigma}_j^2 = \frac{1}{n}\sum_{\alpha=1}^{n}(X_{\alpha j} - \bar{X}_j)^2,$$

$$\hat{\rho} = \frac{1}{k(k-1)}\sum_{i=1}^{k}\sum_{\substack{j=1 \\ i\neq j}}^{k}\frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i\hat{\sigma}_j}, \tag{8}$$

where

$$\hat{\sigma}_{ij} = \frac{1}{n}\sum_{\alpha=1}^{n}(X_{\alpha i} - \bar{X}_i)(X_{\alpha j} - \bar{X}_j), \quad i, j = 1, \ldots, k, \ i \neq j, \tag{9}$$

and the maximum likelihood estimators under H_0 are

$$\hat{\mu}_{j0} = \bar{X}_j, \quad \hat{\sigma}_0^2 = \frac{1}{k}\sum_{j=1}^{k}\hat{\sigma}_j^2, \quad \hat{\rho}_0 = \frac{1}{k(k-1)}\sum_{i=1}^{k}\sum_{\substack{j=1\\ i\neq j}}^{k}\frac{\hat{\sigma}_{ij}}{\hat{\sigma}_0^2} \tag{10}$$

Under the unrestricted condition, μ_j and σ_j^2 are estimated by the usual MLE of the normal case; further, $\hat{\sigma}_{ij}/\hat{\sigma}_i\hat{\sigma}_j$ estimates ρ_{ij}, and because $\rho_{ij} = \rho$ for each pair (i, j), $i \neq j$, its estimator is the average over all such estimators. Under H_0, the common variance σ^2 is estimated by the average of the $\hat{\sigma}_j^2$'s, and then this estimator is used to form the avrage of ρ.

The proof of Theorem 1 appears in Section 6.

Theorem 2. *Under the stated conditions, the likelihood ratio test rejects if and only if*

$$\lambda_1^{2/n} = \left[\frac{\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2\right)^{1/k}}{1/k\left(\hat{\sigma}_1^2 + \cdots + \hat{\sigma}_k^2\right)}\right]^k \frac{(1-\hat{\rho})^{k-1}[1+(k-1)\hat{\rho}]}{(1-\hat{\rho}_0)^{k-1}[1+(k-1)\hat{\rho}_0]} < c, \tag{11}$$

where c is chosen to satisfy a level of significance α.

It is known that in multivariate independent normal sampling, the likelihood ratio procedure for a given hypothesis, rejects if and only if

$$\lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}\right)^{n/2} < c', \tag{12}$$

where Σ is the sample covariance matrix estimated without restrictions, and Σ_0 is the sample covariance matrix estimated with the restrictions defined by H_0. To test H_0 in Equation (2) against the hypothesis defined by Equation (1), we have from Lemma 1,

$$|\Sigma| = \sigma_1^2, \ldots, \sigma_k^2|\mathbf{P}| = \sigma_1^2, \ldots, \sigma_k^2(1-\rho)^{k-1}[1+(k-1)\rho], \tag{13}$$

$$|\Sigma_0| = (\sigma^2)^k|\mathbf{P}| = (\sigma^2)^k(1-\rho)^{k-1}[1+(k-1)\rho], \tag{14}$$

and hence Equation (11) will be proved when we prove Theorem 1.

The $k \times n$ data matrix formed by the (independent) $\mathbf{X}_1, \ldots, \mathbf{X}_n$ can also arise from the mixed linear model without interaction and with one observations per cell, namely $X_{j\alpha} = \mu + \tau_j + \beta_\alpha + e_{j\alpha}$, $j = 1, \ldots, k, \alpha = 1, \ldots, n$, where μ is a constant mean, $\tau_1, \ldots, \tau_k$ are constants (the main effects for the fixed factor), $\beta_1, \ldots, \beta_n$ are independent $N(0, \sigma_\beta^2)$ random variables (the main effects for the random factor), and $e_{j\alpha}$ are independent $N(0, \sigma_e^2)$ errors, independent of the β_α's. Then $E(X_{j\alpha} - \mu - \tau_j)^2 = E(\beta_\alpha + e_{j\alpha})^2 = \sigma_\beta^2 + \sigma_e^2$, $E(X_{j\alpha} - \mu - \tau_j)(X_{i\alpha} - \mu - \tau_i) = E(\beta_\alpha + e_{j\alpha})(\beta_\alpha + e_{i\alpha}) = \sigma_\beta^2$ if $i \neq j$, whereas $E(X_{j\alpha} - \mu - \tau_j)(X_{i\beta} - \mu - \tau_i) = 0$ if $\alpha \neq \beta$. The covariance matrix associated with this model

has $\sigma_\beta^2 + \sigma_e^2$ along its main diagonal and σ_β^2 elsewhere, so that in the notation previously introduced it can be written as $\sigma_e^2\mathbf{I} + \sigma_\beta^2\mathbf{J} = (\sigma_e^2 + \sigma_\beta^2)[(1-\rho^*)\mathbf{I} + \rho^*\mathbf{J}]$, where $\rho^* = \sigma_\beta^2/(\sigma_\beta^2 + \sigma_e^2)$. In this model the nuisance parameter $\sigma_e^2 > 0$ corresponds to the assumption of homoscedasticity, and interest is switched from Equation (2) to the null hypothesis that $\sigma_\beta = 0$, which occurs if and only if $\rho^* = 0$: ρ^* is the intraclass correlation coefficient; see for example, Donner (9), who reviews the random effects model. Anderson (10) used this model as an illustration of a patterned matrix corresponding to Equation (7). The example also follows from ((11), Example 1.3.3), as will be discussed in Section 3.2.

3. Discussion

Once the results in Section 2 are established, it pays to consider some other results related to them. The matrix in Equation (1) is a "patterned" matrix, and we can consider other patterned matrices that are used frequently in multivariate statistical analysis and elsewhere. In Table 1 we present several such matrices, with the names they are frequently given in the literature. We use the notation introduced in the Lemma of Section 2, plus $\hat{\mathbf{R}} = \|\sigma_{ij}\| = \|\sigma_{ij}/\sigma_i\sigma_j\|$ (which corresponds to $\mathbf{R} = \|\sigma_{ij}\| = \|\sigma_{ij}/\sigma_i\sigma_j\|$), and the notation for the maximum likelihood estimators in Equation (8), (9), and (10).

In the case of independent sampling from a multivariate normal population, these results can be used, in conjunction with Equation (12), to design various tests, as will now be discussed.

3.1. Tests of Equality of Several Variances

1. In practice, a test procedure such as Equation (12) will be used with the transformation $-2\log\lambda$, for which the asymptotic distribution is

Table 1. Hypotheses About Covariance Matrices, and Corresponding Maximum Likelihood Estimators Under Independent Normal Sampling

Hypothesis	Covariance Matrix	Determinant of MLE of Covariance Matrix
H_s	$\sigma^2\mathbf{I}$	$(\hat{\sigma}_0^2)^k$
H_c	$\sigma^2\mathbf{P}$	$(\hat{\sigma}_0^2)^k(1-\hat{\rho}_0)^{k-1}[1+(k-1)\hat{\rho}_0]$
H_d	$\mathbf{D}$	$\hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2$
H_{ic}	$\mathbf{D}^{1/2}\mathbf{P}\mathbf{D}^{1/2}$	$\hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2(1-\hat{\rho})^{k-1}[1+(k-1)\hat{\rho}]$
H_g	$\mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2}$	$\hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2\lvert\hat{\mathbf{R}}\rvert$

χ^2 with $k-1$ degrees of freedom. We have that Equation (11) corresponds to testing the null hypothesis H_e clliptical against H_{ic} (intraclass correlation) in Table 1.

2. For the case of $k = 2$, the test statistic given in Equation (11) was presented in (1), as

$$\lambda_1^{2/n} = \frac{\hat{\sigma}_1^2\hat{\sigma}_2^2(1-\hat{\rho}^2)}{\left(\frac{\hat{\sigma}_1^2+\hat{\sigma}_2^2}{2}\right)^2(1-\hat{\rho}_0^2)} \tag{15}$$

3. If $\rho = 0$ in Equation (1), we are in the case of i.i.d. sampling from a $N(\mu, \mathbf{D})$ population, and H_0 is again given by Equation (2). Then the likelihood ratio is a power of the first factor in Equation (11), namely

$$\left[\frac{\left(\prod_{j=1}^k \hat{\sigma}_j^2\right)^{1/k}}{\frac{1}{k}\sum_{j=1}^k \hat{\sigma}_j^2}\right]^{\frac{1}{2}nk} = \left(\frac{|\hat{\mathbf{D}}|^{1/k}}{\frac{1}{k}(\mathrm{tr}\,\hat{\mathbf{D}})}\right)^{\frac{1}{2}nk}, \tag{16}$$

where $\hat{\mathbf{D}}$ is diagonal with diagonal elements $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2$. This agrees with an expression used in (12, Sec. 10.7) when analyzing the likelihood ratio test for sphericity. We see that the acceptance or rejection of H_0 depends on the relation between the geometric (G) and arithmetic ($\bar{X}$) means of $\hat{\sigma}_j^2$: since for nonnegative quantities, $G < \bar{X}$, and equality if and only if all quantities are equal, Equation (16) will be equal to 1 if and only if H_0 is true, and otherwise smaller than 1, which means that the rejection region has a clear interpretation.

Bartlett (13) proposed to modify Equation (16) by using $s_j^2 = n\hat{\sigma}_j^2/(n-1)$ and modifying further the resulting statistic to achieve closeness to the asymptotic χ^2 distribution. In the case of equal sample sizes, his test statistic can be written, after taking logarithms and multiplying by -2 , as

$$B = \frac{1}{C}\left(nk\log s_0^2 - n\sum_{i=j}^k \log s_j^2\right)$$

$$= \frac{n}{C}\log\left[\frac{\frac{1}{k}\sum_{j=1}^k s_j^2}{\left(\prod_{j=1}^k s_j^2\right)^{1/k}}\right]^k, \tag{17}$$

where

$$C = 1 + \frac{(k/n)-(1/nk)}{3(k-1)}. \tag{18}$$

Bartlett's procedure is available for the case of unequal sample sizes.

4. A multivariate analog of the problems treated here, is that of testing the hypothesis that $\Sigma_1 = \cdots = \Sigma_k$ with samples $\mathbf{x}_\alpha^{(g)}$, $\alpha = 1, \ldots, N_g$, $g = 1, \ldots, k$ (e.g., (12), Section 10.2). Bartlett's procedure is related to the solution of this case in the univariate case.
5. We conclude that we have two likelihood ratio test of homogeneity of variances, one for independent samples (point 3), and another (Eq. (11)) for dependent samples under the intraclass correlation model. Stigler (14) recalled that "To Galton, correlation meant what we might call today intraclass correlation—two variables are correlated because they share a common set of influences."

3.2. Tests on Correlations

6. The information in Table 1 can also be used to design tests of correlations. For example, taking H_s (sphericity) as a null hypothesis and testing it against H_e, we have a test of $\rho_{ij} = 0$ against a constant (unspecified) ρ_{ij}, in both cases with homogeneous variances. The likelihood ratio test rejects if and only if λ_2 is small, where

$$\lambda_2 = \{(1-\hat{\rho}_0)^{k-1}[1+(k-1)\hat{\rho}_0]\}^{n/2} \tag{19}$$

Testing that the intraclass correlation coefficient is 0 was discussed briefly at the end of Section 2, as arising in the analysis of data generated by a mixed linear model. Khuri et al. (11, p. 14) state that "in the balanced case, the standard F-test based on the ratio of the treatment sum of squares to the error sum of squares easily turns out to be an optimum test." This procedure was given, for example, by (2) that we cited in Section 1. Khuri et al. (11, p. 16) also write, "Except in some simple cases, the standard likelihood ratio tests (LRT's) are not readily applicable in the context of mixed linear models, and often such LRT's are neither exact nor easily available. For example, even in the simplest case of a one-way balanced random effects model, the exact distribution of the LRT statistic under the null hypothesis of no treatment effect is not known and can only be simulated."

3.3. Tests of Variances and Correlations

7. Another test situation is to consider the general covariance matrix with components $\sigma_i\sigma_j\rho_{ij}$, corresponding to H_g (general), and testing the null hypothesis H_e: $\sigma_1 = \cdots = \sigma_k$, $\rho_{12} = \cdots = \rho_{k-1,k}$. In Table 1, we tested the ellipticity hypothesis against the general case of Σ, whereas

in Theorem 2 we tested the ellipticity hypothesis against the less general case designated as H_{ic} in Table 1, in which the variances are allowed to vary but the correlations are the same.

Under the general hypothesis, the maximum likelihood estimators are the sample means and variances in Equation (8), and $\hat{\rho}_{ij} = \hat{\sigma}_{ij}/\hat{\sigma}_i\hat{\sigma}_j$, whereas under the ellipticity hypothesis they are the same as given in Equation (10). Then the likelihood ratio test rejects if and only if,

$$\lambda_3^{2/n} = \left[\frac{\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2\right)^{1/k}}{1/k\left(\hat{\sigma}_1^2 + \cdots + \hat{\sigma}_k^2\right)}\right]^k \frac{|\hat{\mathbf{R}}|}{(1-\hat{\rho}_0)^{k-1}[1+(k-1)\hat{\rho}_0]} < c''' \tag{20}$$

This procedure was given by (15, Sec. 8.6.2) as

$$(k-1)^{k-1}\frac{|\mathbf{V}|}{u v^{k-1}} < \text{constant}, \tag{21}$$

where $\mathbf{V} = n\hat{\Sigma}$, $u = n\hat{\tau}_1$ and $\tau_1 = \sigma^2[1+(k-1)\rho]$, and $v = n(k-1)\hat{\tau}_2$ and $\tau_2 = \sigma^2(1-\rho)$, and all estimators are by maximum likelihood. If $k = 2$, Equation (20) coincides with Equation (11) because there is only one correlation for the given pair of variables.

8. We also have what is known as the "sphericity test," namely to test the null hypothesis H_s against H_g. Using the results in Table 1, we find that the likelihood ratio test statistic is

$$\lambda_4^{2/n} = \left[\frac{\left(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_k^2\right)^{1/k}}{1/k\left(\hat{\sigma}_1^2 + \cdots + \hat{\sigma}_k^2\right)}\right]^k |\hat{\mathbf{R}}|. \tag{22}$$

This expression coincides, except for the choice of notation, with one given, for example, in (12, Sec. 10.7.2).

It can be shown that under the null hypothesis of sphericity, the two factors on the right-hand side of Equation (22) are independent, because they depend on the sample variances and correlations, respectively (cf (12), Sec. 10.7.3). We note that the test statistics given in Equations (11) and (20) are also products of two factors, one involving a ratio of functions of the sample variances, and the other a ratio of functions of the estimators of ρ. However, the underlying null hypotheses in both cases are not that of sphericity, and it is not immediate or intuitive whether the factors are independent under the corresponding null hypotheses.

9. Finally we comment on the nested nature of some of these hypotheses. In effect, referring to Table 1, H_s is nested in H_e and in H_d (diagonal), which are nested in H_{ic}, which is nested in H_g. Hypothesis H is nested

in hypothesis K if the parameter subspace defined by H is a subset of that defined by K.

4. Using tests of equality of several variances in a simulation study

In (1) the estimation of the (scalar) parameter of a Gaussian first order moving average time series model [denoted MA(1)] was studied by means of simulations, as mentioned in Section 1 (Introduction). This model postulates that a real-valued, observable, time series $\{y_t\}$ is generated by $y_t = \varepsilon_t + \alpha\varepsilon_{t-1}$, with unobservable errors ε_t independent with zero expected value and constant variance, and $|\alpha| < 1$. For the simulation, pseudorandom independent normal errors were generated, and from them y_t values responding to the model with given α's. Various sample sizes were explored and 100 replications were done in each case.

Three procedures were subject to detailed analyses, namely those identified as FORM1, FORM2, and FORM3. They are iterative procedures for estimating, by the exact maximum likelihood procedure, the main parameter, which in all cases was taken to be the parametric function $\alpha(1+\alpha^2)^{-1}$, the first order autocorrelation of the process. They differ in that FORM1 is derived from the likelihood function whereas FORM2 and FORM3 are derived, by using two different arguments, from the concentrated likelihood function. Detailed discussion of the origin of these procedures appears in (16).

When a given sample of 100 observations was generated, all estimation procedures were used with it to compute the estimates, in a design aimed at controlling the noise in the experiment. This leads to a correlation among the values obtained with this sample. Since the three procedures are mathematically equivalent, even when they may differ computationally, it is reasonable to postulate the intraclass correlation model, in which all correlations between FORMs are assumed to be equal.

It was found that the results for $\alpha = 0.30$ and $\alpha = 0.90$ differ considerably when they were compared for sample size $T = 100$, and this led to an experiment with FORM1 and $T = 250$ for the larger value of the parameter. In view of this fact, here we will only analyze the results for $\alpha = 0.30$ and $T = 100$, which are complete.

In (1) the means of the three estimators were compared in detail. Also the variances of pairs of procedures were analyzed. We now include the results obtained in the same study from the BMDP computer program (17), which being a maximum likelihood procedure is comparable to the three FORMS. The estimated variances appear in Table 2.

In our study of the homogeneity of variances, we will use λ_1 defined in Equation (11) [and in Eq. (15) for $k = 2$ groups] and also λ_3 defined in Equation (20).

Table 2. Estimated Variances of Four Estimation Procedures for the Error Variance in the First Order Moving Average Model

Method	$\alpha = 0.30$
FORM1	0.0049318
FORM2	0.0049336
FORM3	0.0049334
BMDP	0.0049863

We recall that they coincide for $k = 2$.

$$\lambda_1^{2/n} = \lambda_{11}\lambda_{12} \tag{23}$$

Each one of these test statistics is the product of two factors, one involving the sample variances and the other the sample correlations. Let us denote where λ_{11} is the first factor and λ_{12} the second factor in the right-hand side of Equation (11), and

$$\lambda_3^{2/n} = \lambda_{31}\lambda_{32} \tag{24}$$

with similar definitions. Note that $\lambda_{11} = \lambda_{31}$.

Table 3 presents an analysis of the test of homogeneity of variances for all pairs of estimation procedures. It shows λ_{11}, λ_{12}, $\lambda_1^{2/n}$, and $\phi_1 = -2\log\lambda_1$, which is asymptotically distributed as chi squared with one degree of freedom.

From the asymptotic chi-squared distribution we have that $\chi^2(1;0.05) = 3.84$, whereas $\chi^2(1;0.01) = 6.63$. When we compare the three FORMs in the first three rows of Table 3, we find no significant differences at the 5% level. When we compare the FORMs with BMDP, higher values of the test statistic are obtained in all cases, all leading to the rejection of the hypotheses of homogeneity of the variances at the indicated levels.

It is clear that the factors involving the correlations determine the rejection of the hypotheses of homogeneity when the FORMs are compared with BMDP. Also note that neither the variances in Table 2 nor the factor λ_{11} show indications of the

Table 3. Numerical Values Associated with the Likelihood Ratio Test Statistic Defined in Equation (15) for all Pairs of Estimation Procedures

Methods	λ_{11}	λ_{12}	$\lambda_1^{2/n}$	ϕ_1
FORM1, FORM2	1.	0.9493249	0.9493249	2.26
FORM1, FORM3	1.	0.9605563	0.9605563	1.75
FORM2, FORM3	1.	0.9991810	0.9991810	0.04
FORM1, BMDP	0.9999698	0.7971065	0.7971065	9.85
FORM2, BMDP	0.9999717	0.7978803	0.7978803	9.81
FORM3, BMDP	0.9999716	0.7975602	0.7975602	9.82

Table 4. Numerical Values of the Factors in λ_j and ϕ_j, $j = 1, 3$, for All Sets of 3 or 4 Estimation Procedures

	a) Testing Homogeneity of Variances by Equation (11)			b) Testing Homogeneity and Correlations by Equation (20)		
Methods	λ_{11}	λ_{12}	ϕ_1	λ_{31}	λ_{32}	ϕ_3
F1,F2,F3	1.	0.92309	8.00	1.	0.57343	55.61
F1,F2,BMDP	0.99996	0.63427	45.53	0.99996	0.00807	481.90
F1,F3,BMDP	0.99996	0.63412	45.55	0.99996	0.00943	466.36
F2,F3,BMDP	0.99996	0.63436	45.52	0.99996	0.00324	573.06
F1,F2,F3,BMDP	0.99995	0.50431	68.46	0.99995	0.00006	978.82

heterogeneity. Finally, note that if no account is taken of the correlation structure, the variances in Table 2 when compared with F-tests with 99 and 99 degrees of freedom, will be declared equal in pairs at any reasonable level of significance.

With the test statistic Equation (11) we can now test the homogeneity of variances for groups of three or four estimation procedures. Let $\phi_j = -2\log\lambda_j$ for $j = 1, 3$, be the transformations that are approximately distributed as χ^2. The numerical results are given in Table 4.

We first analyze Part (a) of Table 4, where we test homogeneity of variances in the intraclass correlation model by Equation (11). We have $\chi^2(2;0.05) = 5.99$, $\chi^2(2;0.01) = 9.21$, $\chi^2(3;0.05) = 7.81$, and $\chi^2(3;0.01) = 11.34$.

To compare FORM1, FORM2, and FORM3 we consider ϕ_1. Homogeneity of variances is accepted at 1% level. When the results for BMDP are included in groups with those for the FORMs, higher values of the sample test statistic ϕ_1 are obtained, leading to rejection of the hypotheses of homogeneity. Ref. (17) has a computer program for parameter estimation in ARMA models based upon ideas coming from (18), which are different from those used to derive the FORMS. For a comparison of these two approaches, see (19).

A different test situation is to test H_e against H_g, in which not only variances but also correlations are allowed to be different. As discussed in point 7 of Section 3.3, this can be approached by using the test statistic Equation (20). The numerical results are given in Part (b) of Table 4.

The asymptotic distribution associated with Equation (20) is considered in (15, Sec. 8.6.2). A simple approximation, which is adequate in our case, is to take ϕ_3 as approximately χ^2 with $f = k(k+1)/2 - 2$ degrees of freedom. For 100 observations we have the results given in Table 5.

The interpretation of the results stemming from Table 4 is similar to that of Table 3. We recall that small values of λ_j lead to rejection of the correponding null hypotheses. For the test (11) of H_e of ellipticity against H_{ic} (under which

Table 5. Significance Points of χ^2 for Selected Degrees of Freedom

Number of Groups, k	Degrees of Freedom, f	$\chi^2(f, 0.05)$	$\chi^2(f, 0.01)$
2	1	3.48	6.63
3	4	9.49	13.28
4	8	15.51	20.09

Σ is given by (1)), we consider the first two columns of numbers in Part (a) of Table 4. We notice that the main contribution towards rejecting H_e comes from λ_{12}, the factor involving the sample correlations. The factor defined by the sample variances either equals 1 or is very close to 1. We conclude that in those cases where rejection is indicated by the test statistic, this is due mostly to the effect of the factor depending on the correlations, over and above the effect of the differing variances as captured by the other factor.

The same kind of analysis holds when we compare H_e against H_g, under which the correlations are unrestricted: we find that λ_{32} is numerically different from λ_{31}.

In summary, λ_1 tests $\sigma_1^2 = \cdots = \sigma_k^2$ under a model of intraclass correlation, in which ρ is treated as a nuisance parameter entering in both the null and the alternative hypotheses. On the other hand, λ_3 tests $\sigma_1^2 = \cdots = \sigma_k^2$ and $\rho_{12} = \cdots = \rho_{k-1,k}$ (i.e., H_e against H_g), in which both the variances and the correlations vary freely. In our numerical example, the test statistics ϕ_1 and ϕ_3 differ considerably, by a factor of approximately 10.

5. Concluding remarks

Testing the homogeneity of k variances ($k \geq 2$) is frequently encountered in the following format: There are k populations $N(\mu_j, \sigma_j^2)$, $j = 1, \ldots, k$, and the null hypothesis $\sigma_1^2 = \cdots = \sigma_k^2$ is to be tested, with the μ_j nuisance parameters; the test is being conducted with k independent samples each of size n. This can be posed as a multivariate problem, in which a sample of n independent k-dimensional random vectors is assumed to be generated by a $N_k(\mu, \mathbf{D})$ population, and H_0 is that $\mathbf{D} = \sigma^2\mathbf{I}$ for some unspecified σ, where again μ is a nuisance parameter. The likelihood ratio test statistic Equation (16) is a power of the ratio of the geometric to the arithmetic means of the $\hat{\sigma}_j^2$, and rejects H_0 if and only if the ratio is smaller than 1. A well-known procedure is Bartlett's test, [Eqs. (17) and (18)], which is a modification of the likelihood ratio procedure.

A different problem arises when there is correlation among the univariate samples, which corresponds to assuming a nondiagonal covariance matrix in the

multivariate setup. One case that we considered in detail is when there is "intraclass correlation," under which the population is $N_k(\mu, \mathbf{D}^{1/2}\mathbf{P}\mathbf{D}^{1/2})$, and the null hypothesis of homogeneity of variances is that the distribution is $N_k(\mu, \sigma^2\mathbf{P})$. This is what we labelled testing the null hypothesis H_e against H_{ic}, and now μ and ρ are nuisance parameters. The likelihood ratio test statistic we found is Equation (11), and it is composed of two factors, one is identical to the test statistic in the likelihood ratio test of homogeneity of variances in the case of independence, and the other is a function of the sample correlation coefficient computed under H_e and under H_{ic}. The numerical example that we presented in Section 4 illustrates the role of each factor of the test statistic in an empirical application, in particular that of the factor involving the correlations.

Testing homogeneity of variances is considered in the literature of multivariate analysis in two other forms that are different from that of intraclass correlation. 1) A test presented in (15) leading to Equation (20), considers that the data is generated by a normal $N_k(\mu, \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2})$ population, and the null hypothesis is that the covariance matrix is $\sigma^2\mathbf{P}$: we are testing that $\sigma_1^2 = \cdots = \sigma_k^2$ and $\rho_{12} = \cdots = \rho_{k-1,k}$ with μ a nuisance parameter. In our notation the null hypothesis H_e is tested against H_g. In the numerical example of Section 4 this test statistic takes values larger than those corresponding to the test of H_e against H_{ic}, as was to be expected because there are more conditions to be satisfied on the parameter space. 2) The "sphericity test" also considers data generated by a $N_k(\mu, \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2})$ population, but the null hypothesis is that the covariance matrix is spherical, $\sigma^2\mathbf{I}$. Hence, we are testing that $\sigma_1^2 = \cdots = \sigma_k^2$ and $\rho_{ij} = 0$ for all $i \neq j$, with only μ being a nuisance parameter. In our notation, the null hypothesis H_s is tested against H_g and the test statistic appears in Equation (22).

6. Proof of theorem 1.

The likelihood function is

$$L(\mu, \Sigma) \propto |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{n}(\mathbf{x}_\alpha - \mu)'\Sigma^{-1}(\mathbf{x}_\alpha - \mu)\right\}. \tag{25}$$

Taking logarithms, and using that the expected values are estimated by the sample means, we can write using the notation introduced in Equations (8) and (9)

$$\log L = \text{constant} - \frac{n}{2}\left\{\sum_{i=1}^{k}\log\sigma_i^2 + \log[(1-\rho)^{k-1}\{1+(k-1)\rho\}]\right\} - \frac{n}{2(1-\rho)[1+(k-1)\rho]} \times \left\{[1+(k-2)\rho]\sum_{i=1}^{k}\frac{\hat{\sigma}_i^2}{\sigma_i^2} - \rho\sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\frac{\hat{\sigma}_{ij}}{\sigma_i\sigma_j}\right\}. \tag{26}$$

Then

$$\frac{\partial \log L}{\partial \sigma_i^2} = -\frac{n}{2\sigma_i^2} - \frac{n}{2(1-\rho)[1+(k-1)\rho]} \times \left\{ [1+(k-2)\rho]\left(-\frac{\hat{\sigma}_i^2}{\sigma_i^4}\right) - \rho \sum_{\substack{j=1 \\ j\neq i}}^{k} \frac{\hat{\sigma}_{ij}}{\sigma_i^3 \sigma_j} \right\}$$

$$= -\frac{n}{2\sigma_i^2}\left\{1 - \frac{1}{(1-\rho)[1+(k-1)\rho]} \times \left[[1+(k-2)\rho]\frac{\hat{\sigma}_i^2}{\sigma_i^2} - \rho \sum_{\substack{j=1 \\ j\neq i}}^{k} \frac{\hat{\sigma}_{ij}}{\sigma_i \sigma_j} \right]\right\} = 0, \tag{27}$$

which implies that

$$[1+(k-2)\rho]\frac{\hat{\sigma}_i^2}{\sigma_i^2} - \rho \sum_{\substack{j=1 \\ j\neq 1}}^{k} \frac{\hat{\sigma}_i^2}{\sigma_i^2} = (1-\rho)[1+(k-1)\rho], \quad i = 1, 2, \ldots, k, \tag{28}$$

and hence that

$$[1+(k-2)\rho]\sum_{i=1}^{k} \frac{\hat{\sigma}_i^2}{\sigma_i^2} - \rho \sum_{\substack{i=1 \\ i\neq j}}^{k} \sum_{j=1}^{k} \frac{\hat{\sigma}_{ij}}{\sigma_i \sigma_j} = k(1-\rho)[1+(k-1)\rho]. \tag{29}$$

Next,

$$\frac{\partial \log L}{\partial \rho} = \frac{nk(k-1)\rho(1-\rho)^{k-2}}{2(1-\rho)^{k-1}[1+(k-1)\rho]} + \frac{n[(k-2)-2\rho(k-1)]}{2(1-\rho)^2[1+(k-1)\rho]^2} \times \left\{ [1+(k-1)\rho]\sum_{i=1}^{k} \frac{\hat{\sigma}_i^2}{\sigma_i^2} - \rho \sum_{\substack{i=1 \\ i\neq j}}^{k} \sum_{j=1}^{k} \frac{\hat{\sigma}_{ij}}{\sigma_i \sigma_j} \right\}$$

$$- \frac{n}{2(1-\rho)[1+(k-1)\rho]}\left\{ (k-2)\sum_{i=1}^{k} \frac{\hat{\sigma}_i^2}{\sigma_i^2} - \sum_{\substack{i=1 \\ i\neq j}}^{k} \sum_{j=1}^{k} \frac{\hat{\sigma}_{ij}}{\sigma_i \sigma_j} \right\}$$

$$= -\frac{n}{2(1-\rho)[1+(k-1)\rho]} - k(k-1)\rho - \frac{(k-2)-2\rho(k-1)}{(1-\rho)[1+(k-1)\rho]} k(1-\rho)[1+(k-1)\rho]$$

$$+(k-2)\sum_{i=1}^{k} \frac{\hat{\sigma}_i^2}{\sigma_i^2} - \sum_{\substack{i=1 \\ i\neq j}}^{k} \sum_{j=1}^{k} \frac{\hat{\sigma}_{ij}}{\sigma_i \sigma_j}$$

$$= 0, \tag{30}$$

where we used Equation (29). We then have

$$(k-2)\sum_{i=1}^{k} \frac{\hat{\sigma}_i^2}{\sigma_i^2} - \sum_{\substack{i=1 \\ i\neq j}}^{k} \sum_{j=1}^{k} \frac{\hat{\sigma}_{ij}}{\sigma_i \sigma_j} = -k(k-1)\rho + k(k-2). \tag{31}$$

Combining Equations (29) and (31),

$$\sum_{\substack{i=1 \\ i\neq j}}^{k} \sum_{j=1}^{k} \frac{\hat{\sigma}_{ij}}{\sigma_i \sigma_j} = \rho k(k-1), \tag{32}$$

from which ρ in Equation (8) follows. Substitution in Equation (29) completes the proof for the estimators of the variances in Equation (8).

Under H_0 the maximum likelihood estimator of the μ_j are again the sample means. We have to find the maximum likelihood estimators of σ^2 and ρ in

$$\log L_0 = \text{constant} - \frac{n}{2}\{\log \sigma^{2k} + \log(1-\rho)^{k-1}[1+(k-1)\rho]\} - \frac{n}{2(1-\rho)[1+(k-1)\rho]\sigma^2} \times \left\{ [1+(k-2)\rho]\sum_{i=1}^{k} \hat{\sigma}_i^2 - \rho \sum_{\substack{i=1 \\ i\neq j}}^{k} \sum_{j=1}^{k} \hat{\sigma}_{ij} \right\}. \tag{33}$$

By a similar algebra than in the previous case, we can show that $\partial \log L_0/\partial \sigma^2 = 0$ leads to an expression that corresponds to Equation (29), namely

$$[1+(k-2)\rho]\sum_{i=1}^{k} \hat{\sigma}_i^2 - \rho \sum_{\substack{i=1 \\ i\neq j}}^{k} \sum_{j=1}^{k} \hat{\sigma}_{ij} = k\sigma^2(1-\rho)[1+(k-1)\rho], \tag{34}$$

and that $\partial \log L_0/\partial \rho = 0$ leads to an expression that corresponds to Equation (31), namely,

$$(k-2)\sum_{i=1}^{k} \hat{\sigma}_i^2 - \sum_{\substack{i=1 \\ i\neq j}}^{k} \sum_{j=1}^{k} \hat{\sigma}_{ij} = \sigma^2[k(k-2) - k(k-1)\rho]. \tag{35}$$

Operating as in the previous case, the results in Equation (10) are proved.

Acknowledgements

I thank professors T. W. Anderson and V. Nair, and an Associate Editor, for helpful comments on this material. I also thank Eng. Carlos Martinez for doing the computations given in Section 4.

References

1. Anderson, T.W.; Mentz, R.P.; Jarma, N.M.; Martinez, C.I. Simulations of Iterative Procedures for Maximum Likelihood Estimation in MA(1) Models. Commun. in Statis.-Simul. and Comput. **1996**, *25* (4), 851–65.
2. Olkin, I. Correlations Revisited (with Discussion). In *Improving Experimental Design & Statistical Analysis*; Stanley, J.C. Ed.; Rand McNally: Chicago, 1967; Chap. 3, 102–56.
3. De Lury, D.B. Note on Correlations. Ann. Math. Stat **1933**, *9*, 149–51.
4. Kendall, M.G. *The Advanced Theory of Statistics*; Charles Griffin: London, 1947; Vol I.
5. Glasser, R.E. Bartlett's Test of Homogeneity of Variances *Encyclopedia of Statistics*; John Wiley & Sons, Inc.: New York, 1982.
6. Koch, G.G. Intraclass correlation coefficient, *Encyclopedia of Statistics*; John Wiley & Sons, Inc.: New York, 1983.
7. Graybill, F.A. *Matrices with Statistical Applications*, 2nd Edn.; Wadsworth & Brooks/Cole: California, 1969.
8. Searle, S.R.; Casella, G.; McCulloch, C.E. *Variance Components*; John Wiley & Sons, Inc.: New York, 1992.
9. Donner, A. A Review of Inference Procedures for the Intraclass Correlation Coefficient in the One-Way Random Effects Model. Intern. Stat. Review, **1986**, *54* (1), 67–82.
10. Anderson, T.W. Estimation of Covariance Matrices which are Linear Combinations or whose Inverses are Linear Combinations of given Matrices. In *Essays in Probability and Statistics*; U. of North Carolina Press: Chapel Hill, 1969; 1–24.
11. Khuri, A.I.; Mathew, T.; Sinha, B.K. *Statistical Tests for Mixed Linear Models*; John Wiley & Sons, Inc.: New York, 1998.
12. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*, 2nd Edn.; John Wiley & Sons, Inc.: New York, 1984.
13. Bartlett, M.S. Properties of Sufficiency and Statistical Tests. Proc. of the Royal Soc. of London A, **1937**, *160*, 268–82.
14. Stigler, S.M. Francis Galton's Account of the Invention of Correlation. Stat. Sci. **1989**, *4* (2), 73–86.
15. Siotani, M.; Hayakawa, T.; Fujikoshi, Y. *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*; American Sciences Press, Inc.: Columbus, Ohio, 1985.
16. Anderson, T.W.; Mentz, R.P. Iterative Procedures for Exact Maximum Likelihood Estimation in the Gaussian First-Order Moving Average Model. In *Statistics and Probability: A Raghu Raj Bahadur Festchrift*; Ghosh, J.K., Mitra, S.K., Parthasarathy, K.R., Prakasa Rao, B.L.S. Eds.; Wiley Eastern Ltd., New Delhi, 1993; 19–35.
17. BMDP. *BMDP Statistical Software*, Los Angeles, California, 1990.
18. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. *Time Series Analysis: Forecasting and Control*, 3rd Edn.; Prentice Hall: New Jersey, 1994.
19. Anderson, T.W.; Mentz, R.P. Evaluation of Quadratic Forms and Traces for Iterative Estimation in First-Order Moving Average Models. Commun. Statist.-Theory Meth., **1993**, *22* (4), 931–63.

Received September, 1999

Revised September, 2000

Bias correction for estimators of the residual variance in the ARMA(1,1) model

Raúl P. Mentz
University of Tucumán, Argentina

Pedro A. Morettin and Clélia M.C. Toloi
University of São Paulo, Brazil

Abstract

We consider the ARMA(1,1) model and deal with the estimation of the residual variance. General results are known for the maximum likelihood(ML) estimators under normality, both for known and unknown mean, in which case the asymptotic biases depend on the number of parameters (including the mean) and on the true residual variance, but not on the values of the remaining parameters. For moment and least squares estimators the situation is different: the asymptotic biases depend on the values of the parameters, besides the true variance. Some simulation results are also presented.

Key words

Autoregressive moving average models, bias, least squares, maximum likelihood, method of moments, residual variance, time series.

1. Introduction

1.1. The Model

We consider the stationary autoregressive moving average model of order (1,1), denoted ARMA(1,1), defined by

$$(X_t - \mu) + \beta(X_{t-1} - \mu) = \varepsilon_t + \alpha\varepsilon_{t-1}, \quad t = 0, \pm 1, \ldots, \tag{1}$$

where ε_t is a white noise sequence with zero mean and finite variance $\sigma^2 > 0$, μ, α, β are real parameters and $|\beta| < 1$. The model is then invertible into an infinite moving-average and if $|\alpha| < 1$ it is invertible into an infinite autoregression.

The autocovariance sequence is $\gamma_s = E(X_t - \mu)(X_{t+|s|} - \mu), s = 0, \pm 1, \ldots$, following the equations

$$\begin{aligned} \gamma_0 &= \frac{1 + \alpha^2 - 2\alpha\beta}{1 - \beta^2}\sigma^2, \\ \gamma_1 &= \frac{(1 - \alpha\beta)(\alpha - \beta)}{(1 - \beta^2)}\sigma^2, \\ \gamma_s &= -\beta\gamma_{s-1}, \qquad s \geq 2. \end{aligned} \tag{2}$$

The autocorrelation sequence is $\rho_s = \gamma_s/\gamma_0, s = 0, \pm 1, \ldots$ and from (2) is given by

$$\rho_s = \begin{cases} 1, & \text{if } s = 0 \\ \frac{(1-\alpha\beta)(\alpha-\beta)}{1+\alpha^2-2\alpha\beta}(-\beta)^{s-1}, & \text{if } s \geq 1. \end{cases} \tag{3}$$

The autocovariance and autocorrelation sequences are two of the basic tools in the time domain analysis of the model (1). The autocovariance satisfies the inversion formula

$$\gamma_s = \int_{-\pi}^{\pi} e^{i\lambda s} f(\lambda) d\lambda,$$

where

$$f(\lambda) = \frac{\sigma^2}{2\pi}\frac{|1 + \alpha e^{i\lambda}|^2}{|1 + \beta e^{i\lambda}|^2}, \quad -\pi \leq \lambda \leq \pi \tag{4}$$

is the spectral density of the process.

1.2. Objectives of the Paper

Model (1) is an important tool both in theoretical and empirical time series analysis. In this paper we consider the estimation of the parameter σ^2. This is important because estimates of the residual variance enter, for example,

into confidence sets for parameters, in the estimation of the spectrum, in the expression of the estimated error of prediction and in the determination of the order of the model using some information criterium (as the AIC or BIC).

For purposes of inference we consider a sample $X_1 \ldots\ldots X_T$ from Eq. (1). Estimation procedures frequently used are the methods of moments (MM). least squares(LS) and maximum likelihood(ML) under normality. Frequency domain arguments have also been considered in the literature.

The main purposes of this paper are: (1) to derive approximate expresions for the biases of MM and LS estimators: (2) to show through simulations the behavior of the theoretical results. in particular to show that the asymptotic formulas for the bias and variance of the ML estimator of σ^2 apply also to the LS estimator(LSE). at least for values of the parameters far from the limits of the regions of stationarity and invertibility.

To deal with (1) above we assume that the ϵ_t are independent identically distributed with mean 0 and variance σ^2; to deal with (2) we add that they are normal. It follows that our simulations are made with pseudonormal random numbers. Conditions for the asymptotic theory of MLE under which LSE have the same properties are given by Brockwell and Davis (1991 Ch. 8 and 10) or Fuller (1996 Ch. 8). Relevant questions as whether the LSE are consistent under more general conditions than the MLE or whether LSE are more robust than MLE to deviations from normality will not be addressed here.

The objects of the inference will be the autocovariances and autocorrelations introduced in (2) and (3), respectively. and the parameters α, β and σ^2. For the first two sequences, large sample expectations, variances, covariances and

distributions are available, for several standard definitions of the sample quantities. This point will be briefly considered in Section 1.3.

Moment estimators are important, since they can be used as good starting values for iterative procedures, as maximum likelihood. Least squares estimators(LSE) are frequently considered in theory and practice of time series and in many computer programs available. In our case we used SPLUS and SCA statistical software. These programs consider the estimation of parameters like α, β and μ in Eq. (1) by minimizing a properly defined sum of squares. ML will be only briefly considered in Section 4, since results are available in the literature.

Tanaka (1984) suggests a technique for obtaining the Edgeworth type asymptotic expansions associated with ML estimates in ARMA models. He obtains biases up to order $1/T$ for AR(1),AR(2), MA(1),MA(2) and ARMA(1,1) models with and without constant terms. Biases for the residual variance are also considered.

Cordeiro and Klein (1994) present a general procedure to obtain the biases of ML estimates in ARMA models. It turns out that the formula is difficult to obtain for models other than the lower order ones, but numerically it is easy to be implemented.

De Gooijer and Pukkila (1994) present a technique for obtaining expressions for the approximate expectations of estimates in ARMA models. They derive first and second order approximations, based on Taylor series expansions of the log-likelihood in terms of the expected values of the sample covariances or in terms of the expected values of the periodogram ordinates.

For pure moving average and autoregressive models see Mentz, Morettin and Toloi (1997, 1998, 1999).

1.3. Asymptotic Results for Sample Autocovariances and Autocorrelations

We consider estimating γ_s by

$$c_s = \frac{1}{T}\sum_{t=1}^{T-s}(X_t - \bar{X})(X_{t+s} - \bar{X}) = c_{-s}, \quad s = 0, \ldots, T-1, \tag{5}$$

where $\bar{X}$ is the sample mean $\bar{X} = \sum_{t=1}^{T} X_t/T$.

Other estimators are considered in the literature, for example by changing in (5) the denominator, the value to be subtracted from the X's or the range of the sums; see, for example, Anderson (1971) or Fuller (1996). We use (5) because for $T > 1$ the autocovariance matrix with elements $c_{|s-t|}$ is positive definite, a fact that we shall use below.

With these estimators of the autocovariances, we form estimators of the autocorrelations ρ_s,

$$r_s = c_s/c_0, \quad s = 0, 1, \ldots, T-1. \tag{6}$$

For LSE a different autocovariance estimator will be defined in Section 3.

Large sample moments of these estimators are available in the literature. They are derived under the assumption that the process follows a general linear model, of which (1) is a special case. Some useful results for the ARMA model are, Fuller (1996), for $h \geq q \geq 0$,

$$E(c_h - \gamma_h) = -\frac{h}{T}\gamma_h - \frac{T-h}{T}\mathrm{Var}(\bar{X}) + O(T^{-2}), \tag{7}$$

$$\mathrm{Cov}(c_h, c_q) = \frac{T-h}{T^2}\sum_p(\gamma_p\gamma_{p-h+q} + \gamma_{p+q}\gamma_{p-h}) + O(T^{-2}), \tag{8}$$

$$E(r_h) = \frac{T-h}{T}\rho_h - \frac{1}{\gamma_0}(1-\rho_h)\mathrm{Var}(\bar{X}) + \frac{1}{\gamma_0^2}[\rho_h\mathrm{Var}(c_0) - \mathrm{Cov}(c_h, c_0)] + O(T^{-2}), \tag{9}$$

$$\mathrm{Cov}(r_h, r_q) = \frac{1}{T}\sum_p(\rho_p\rho_{p-h+q} + \rho_{p+q}\rho_{p-h} - 2\rho_q\rho_p\rho_{p-h} - 2\rho_h\rho_p\rho_{p-q} + 2\rho_h\rho_q\rho_p^2) + O(T^{-2}). \tag{10}$$

For the ARMA(1,1) model we have, using (7),

$$E(c_0 - \gamma_0) = -\mathrm{Var}(\bar{X}) + O(T^{-2}) \approx \frac{-\sigma^2}{T}\frac{(1+\alpha)^2}{(1+\beta)^2}. \tag{11}$$

2. Estimation by the method of moments

2.1. Estimation of the Residual Variance

Box, Jenkins and Reinsel (1994 p. 206) give a general procedure for obtaining initial estimates of parameters for an ARMA(p,q) model, which can be viewed as moment type estimates. A related reference is Brockwell and Davis (1991 p.250).

From (2) we find that an estimate for σ^2 is given by

$$\hat{\sigma}^2_{MM} = \frac{1-\hat{\beta}^2}{1+\hat{\alpha}^2-2\hat{\alpha}\hat{\beta}} c_0. \tag{12}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the moment estimators of α and β, respectively, defined implicitly by the sample analog of (2).

In deriving the asymptotic bias of (12), the asymptotic biases, variances and covariances of $\hat{\alpha}$ and $\hat{\beta}$ will be needed. We proceed as follows: from (3) we obtain

$$\rho_1 = \frac{(1-\alpha\beta)(\alpha-\beta)}{1+\alpha^2-2\alpha\beta}, \quad \rho_2 = -\beta\rho_1, \tag{13}$$

from which we get $\alpha = f(\rho_1, \rho_2)$ and $\beta = g(\rho_1, \rho_2)$.

2.2. Asymptotic Theory for the Residual Variance Estimator

Asymptotic biases of the moment estimators $\hat{\alpha}$ and $\hat{\beta}$ are obtained as follows. Let

$$\dot{\alpha}_i = \frac{\partial f}{\partial \rho_i}, \quad \ddot{\alpha}_{ij} = \frac{\partial^2 f}{\partial \rho_i \partial \rho_j}, \quad \dot{\beta}_i = \frac{\partial g}{\partial \rho_i}, \quad \ddot{\beta}_{ij} = \frac{\partial^2 g}{\partial \rho_i \partial \rho_j}, \tag{14}$$

for $i, j = 1, 2$. Here, $\partial f/\partial \rho_i$ is a shorthand notation for $\partial f(r_1, r_2)/\partial r_i|_\theta$, where θ is the vector of parameters. The same holds for the other derivatives and this simplification will be used consistently in the paper.

Then, using a Taylor expansion that retains terms contributing up to second order and taking expected values we obtain

$$E(\hat{\alpha}-\alpha) = \sum_{i=1}^{2} \dot{\alpha}_i E(r_i-\rho_i) + \frac{1}{2}\sum_{i=1}^{2} \ddot{\alpha}_{ii} E(r_i-\rho_i)^2 + \ddot{\alpha}_{12} E(r_1-\rho_1)(r_2-\rho_2) + o(T^{-1}), \tag{15}$$

$$E(\hat{\beta}-\beta) = \sum_{i=1}^{2} \dot{\beta}_i E(r_i-\rho_i) + \frac{1}{2}\sum_{i=1}^{2} \ddot{\beta}_{ii} E(r_i-\rho_i)^2 + \ddot{\beta}_{12} E(r_1-\rho_1)(r_2-\rho_2) + o(T^{-1}). \tag{16}$$

Using the expansions for $\hat{\alpha}-\alpha$ and $\hat{\beta}-\beta$ and taking expectations, we obtain

$$E(\hat{\alpha}-\alpha)(\hat{\beta}-\beta)=\sum_{i=1}^{2}\dot{\alpha}_i\dot{\beta}_i E(r_i-\rho_i)^2+(\dot{\alpha}_2\dot{\beta}_1+\dot{\alpha}_1\dot{\beta}_2)E(r_1-\rho_1)(r_2-\rho_2)+o(T^{-1}), \tag{17}$$

$$E(\hat{\alpha}-\alpha)^2=\sum_{i=1}^{2}\dot{\alpha}_i^2 E(r_i-\rho_i)^2+2\dot{\alpha}_1\dot{\alpha}_2 E(r_1-\rho_1)(r_2-\rho_2)+o(T^{-1}). \tag{18}$$

$$E(\hat{\beta}-\beta)^2=\sum_{i=1}^{2}\dot{\beta}_i^2 E(r_i-\rho_i)^2+2\dot{\beta}_1\dot{\beta}_2 E(r_1-\rho_1)(r_2-\rho_2)+o(T^{-1}). \tag{19}$$

We now consider finding the asymptotic bias of $\hat{\sigma}^2_{MM}$. Using a Taylor expansion up to second order and taking expectations, we obtain

$$\begin{aligned}E(\hat{\sigma}^2_{MM}-\sigma^2)=&\frac{\partial\sigma^2}{\partial\gamma_0}E(c_0-\gamma_0)+\frac{\partial\sigma^2}{\partial\alpha}E(\hat{\alpha}-\alpha)\\&+\frac{\partial\sigma^2}{\partial\beta}E(\hat{\beta}-\beta)+\frac{1}{2}\frac{\partial^2\sigma^2}{\partial\alpha^2}E(\hat{\alpha}-\alpha)^2+\frac{1}{2}\frac{\partial^2\sigma^2}{\partial\beta^2}E(\hat{\beta}-\beta)^2\\&+\frac{\partial^2\sigma^2}{\partial\gamma_0\partial\alpha}E(c_0-\gamma_0)(\hat{\alpha}-\alpha)+\frac{\partial^2\sigma^2}{\partial\gamma_0\partial\beta}E(c_0-\gamma_0)(\hat{\beta}-\beta)\\&+\frac{\partial^2\sigma^2}{\partial\alpha\partial\beta}E(\hat{\alpha}-\alpha)(\hat{\beta}-\beta)+o(T^{-1}).\end{aligned} \tag{20}$$

In the Appendix we show that (20) leads to

$$E(\hat{\sigma}^2_{MM}-\sigma^2)=\frac{-\sigma^2}{T}\frac{M(\alpha,\beta)}{(1-\alpha^2)(\beta^2-1)(\alpha\beta-1)^2}+o(T^{-1}), \tag{21}$$

with

$$\begin{aligned}M(\alpha,\beta)=&-3+5\alpha^2-2\alpha^3-17\alpha^4+4\alpha^5+14\alpha^6-2\alpha^7-6\alpha^8-\alpha^{10}\\&+2\alpha\beta+2\alpha^2\beta+2\alpha^3\beta-2\alpha^4\beta+12\alpha^5\beta-2\alpha^6\beta-6\alpha^7\beta\\&+2\alpha^8\beta+6\alpha^9\beta-2\alpha^2\beta^2+3\alpha^4\beta^2-11\alpha^6\beta^2+4\alpha^8\beta^2+\alpha^{10}\beta^2\\&-10\alpha\beta^3-2\alpha^2\beta^3+6\alpha^3\beta^3+2\alpha^4\beta^3-4\alpha^5\beta^3+2\alpha^6\beta^3-2\alpha^7\beta^3\\&-2\alpha^8\beta^3-6\alpha^9\beta^3+5\alpha^2\beta^4+2\alpha^3\beta^4-6\alpha^4\beta^4-4\alpha^5\beta^4\\&+5\alpha^6\beta^4+2\alpha^7\beta^4+4\alpha^8\beta^4.\end{aligned}$$

In the special case of $\alpha=0$, (21) reduces to

$$\frac{-\sigma^2}{T}\frac{3-5\beta^2}{1-\beta^2}+o(T^{-1}), \tag{22}$$

which can be compared with the bias of the residual variance estimator for the AR(1) model, given by Mentz *et al.* (1998) as

$$\frac{-\sigma^2}{T}\frac{2-4\beta^2}{1-\beta^2}+o(T^{-1}). \tag{23}$$

The difference between (22) and (23) arises from the use of derivatives. For example, a term equal to α appearing in the analysis of the ARMA(1,1) model, has partial derivative equal to 1 with respect to α; but, if α is set equal to 0 the derivative becomes 0. Expressions (22) and (23) are compared in Figure 1.

In the special case of $\beta = 0$, (21) reduces to

$$\frac{-\sigma^2}{T}\frac{3-5\alpha^2+2\alpha^3+17\alpha^4-4\alpha^5-14\alpha^6+2\alpha^7+6\alpha^8+\alpha^{10}}{(1-\alpha^2)^3}+o(T^{-1}). \tag{24}$$

which can be compared with the bias obtained for the MA(1) model by Mentz *et al.* (1997) as

$$\frac{-\sigma^2}{T}\frac{2-6\alpha^2-2\alpha^3+15\alpha^4+4\alpha^5-4\alpha^6-2\alpha^7+\alpha^8}{(1-\alpha^2)^3}+o(T^{-1}). \tag{25}$$

Expressions (24) and (25) are compared in Figure 2.

In the special case of $\alpha = \beta$, (21) reduces to

$$\frac{-\sigma^2}{T}\frac{3-3\alpha^2+9\alpha^4-\alpha^6}{(1-\alpha^2)^3}+o(T^{-1}), \tag{26}$$

which can be compared with $\frac{-\sigma^2}{T}$. Expression (26) is shown in Figure 3 for $|\alpha| \leq 0.6$.

3. Estimation by least squares

One procedure that is often used in practice is to minimize the sum

$$S(\alpha,\beta)=\sum_{j=1}^{T}(X_j-\hat{X}_j)^2/v_{j-1} \tag{27}$$

with respect to α and β, where $E(X_{n+1}-\hat{X}_{n+1})^2 = v_n\sigma^2$ is the mean square error of prediction and the one-step ahead predictors $\hat{X}_j$ of X_j can be computed recursively through the Innovations Algorithm (see Brockwell and Davis (1991)), namely

$$v_0=\frac{1+2\theta\phi+\theta^2}{1-\phi^2}, \quad v_{j+1}=1+\theta^2-\frac{\theta^2}{v_j}, \quad j=0,1,2,\ldots.$$

$$\hat{X}_1=0, \quad \hat{X}_{j+1}=\phi X_j+\frac{\theta}{v_{j-1}}(X_j-\hat{X}_j), \quad j=1,2,\ldots.$$

The estimators obtained in this way will be referred to as the least squares estimators(LSE) $\hat{\alpha}_{LS}$ and $\hat{\beta}_{LS}$, of α and β, respectively. For the minimization of $S(\alpha,\beta)$ it is necessary to restrict α and β to $|\alpha|<1$ and $|\beta|<1$.

To develop the theory we use the idea of Durbin (1959), approximating the ε_t in (1) by a "long autoregression"

$$\varepsilon_t^* = \sum_{s=0}^{B_T} \delta_s^* (X_{t-s} - \mu), \tag{28}$$

for some set of constants δ_s^* with $\delta_0^* = 1$, and where $B_T \to \infty$ as $T \to \infty$ but in such a way that $B_T/T \to 0$ as $T \to \infty$. In fact, we use (28) together with the substitution of μ by $\bar{X}$.

The corresponding LSE of the residual variance will be taken as

$$\hat{\sigma}_{LS}^2 = \frac{1}{T-3} \sum_{t=1}^{T} \hat{\epsilon}_t^2, \tag{29}$$

taking into account the estimation of the parameters α and β.

Defining $u_t = \varepsilon_t + \alpha\varepsilon_{t-1}$ we have

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\alpha)^j u_{t-j} = \sum_{j=0}^{\infty} (-\alpha)^j [X_{t-j} - \mu + \beta(X_{t-j-1} - \mu)]$$

$$= (X_t - \mu) + (\beta - \alpha) \sum_{j=1}^{\infty} (-\alpha)^{j-1} (X_{t-j} - \mu)$$

that can be approximated by

$$\varepsilon_t^* = (X_t - \mu) + (\beta - \alpha) \sum_{j=1}^{B_T} (-\alpha)^{j-1} (X_{t-j} - \mu). \tag{30}$$

Thus the LSE of σ^2 using the long AR approximation is

$$\hat{\sigma}_{LS}^2 = \frac{1}{T - B_T - 3} \sum_{t=B_T+1}^{T} [\hat{\varepsilon}_t^*]^2$$

$$= \frac{1}{T - B_T - 3} \sum_{t=B_T+1}^{T} [(X_t - \bar{X}) + (\hat{\beta} - \hat{\alpha}) \sum_{j=1}^{B_T} (-\hat{\alpha})^{j-1} (X_{t-j} - \bar{X})]^2, \tag{31}$$

which is approximately equal to

$$\hat{\sigma}_{LS}^2 = c_{00}^* + (\hat{\beta} - \hat{\alpha})^2 \sum_{j=1}^{B_T} \sum_{k=1}^{B_T} (-\hat{\alpha})^{j+k-2} c_{jk}^* + 2(\hat{\beta} - \hat{\alpha}) \sum_{j=1}^{B_T} (-\hat{\alpha})^{j-1} c_{j0}^*, \tag{32}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the least squares estimators of α and β, respectively and

$$c_{jk}^* = \frac{1}{T - B_T - 3} \sum_{t=B_T+1}^{T} (X_{t-j} - \overline{X})(X_{t-k} - \overline{X}), \tag{33}$$

for $j, k = 0, 1, \ldots, B_T$.

These sample autocovariances are different from (5) in that they all have $(T - B_T)$ summands and they do not form Toeplitz matrices, since for example c^*_{11} and c^*_{22} are different. It is not difficult to verify that

$$E(c^*_{jk} - \gamma_{j-k}) = \frac{-2\pi f(0)}{T} + o(T^{-1}) = -\frac{\sigma^2}{T}\frac{(1+\alpha)^2}{(1+\beta)^2} + o(T^{-1}). \tag{34}$$

This result shows that all the c^*_{jk} have the same asymptotic biases and this has to be used instead of (7), valid for moment estimators. Using the fact that $\frac{\partial}{\partial 3}\sum_{t=B_T+1}^{T}[\hat{\varepsilon}^*_t]^2 = 0$, we can rewrite (32) as

$$\hat{\sigma}^2_{LS} \simeq c^*_{00} + (\hat{\beta} - \hat{\alpha})\sum_{j=1}^{B_T}(-\hat{\alpha})^{j-1}c^*_{j0}. \tag{35}$$

Using (2) we have that

$$\sigma^2 = \gamma_0 + (\beta - \alpha)\sum_{j=1}^{\infty}(-\alpha)^{j-1}\gamma_j \tag{36}$$

and taking $\hat{\sigma}^2_{LS}$ as a function of $\hat{\alpha}, \hat{\beta}$ and c^*_{j0}, $j = 1, \ldots, B_T$, we derive a Taylor expansion in terms up to second order and obtain the asymptotic bias as

$$E(\hat{\sigma}^2_{LS} - \sigma^2) = -\frac{\sigma^2}{T}\frac{(1+\alpha)}{(1+\beta)} + \sum_{i=1}^{6} A_i + o(T^{-1}), \tag{37}$$

where

$$A_1 = -\sigma^2\frac{(\alpha - \beta)}{1 - \alpha\beta}E(\hat{\alpha} - \alpha)\,,$$

$$A_2 = \sigma^2\frac{(\alpha - \beta)}{1 - \beta^2}E(\hat{\beta} - \beta)\,,$$

$$A_3 = \sigma^2\frac{\beta(\alpha - \beta)}{(1 - \beta^2)(1 - \alpha\beta)}E(\hat{\alpha} - \alpha)(\hat{\beta} - \beta)\,,$$

$$A_4 = -\sigma^2\frac{\beta(\alpha - \beta)}{(1 - \alpha\beta)^2}E(\hat{\alpha} - \alpha)^2\,,$$

$$A_5 = \sum_{j=1}^{B_T}(-\alpha)^{j-2}[\alpha j - \beta(j-1)]E(\hat{\alpha} - \alpha)(c^*_{j0} - \gamma_j)\,,$$

$$A_6 = \sum_{j=1}^{B_T}(-\alpha)^{j-1}E(\hat{\beta} - \beta)(c^*_{j0} - \gamma_j).$$

See the Appendix for details.

The first term in the right-hand side of (37) is the sum of the contributions coming from terms involving the (asymptotic) biases of the covariance estimators, that is, approximations to $E(c^*_{j0} - \gamma_j), j = 0, 1, \ldots$. The second term includes two parts: (1) The contributions from the biases, variances and covariances of $\hat{\alpha}$ and $\hat{\beta}$, namely approximations to $E(\hat{\alpha} - \alpha)$, $E(\hat{\beta} - \beta)$, $E(\hat{\alpha} - \alpha)^2$, $E(\hat{\beta} - \beta)^2$ and $E(\hat{\alpha} - \alpha)(\hat{\beta} - \beta)$; these can be taken as equivalent (asymptotically) to those

of the maximum likelihood estimators, as given, for example, in Brockwell and Davis (1991 Section 8.8). (2) The contributions from the asymptotic covariances between the c_{j0}^{*}'s and $\hat{\alpha}$ and $\hat{\beta}$; for these there are no closed-form approximations available, and we may resort to some numerical procedure, such as the bootstrap, Efron and Tibshirani (1993).

4. Estimation by maximun likelihood

Here we assume that the ε_t are normally distributed with zero mean and variance σ^2. Then the likelihood of the observations $\mathbf{X} = (X_1, \ldots, X_T)'$ is the function of μ, σ, α and β given by

$$L = (2\pi)^{-T/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}(\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu)\}. \tag{38}$$

where Σ is the covariance matrix of $\mathbf{X}$.

A convenient way to treat the problem is by using the prediction error decomposition of the likelihood. This avoids the direct calculation of the determinant and inverse of the covariance function of $\mathbf{X}$. We can also write

$$L = (2\pi\sigma^2)^{-T/2}|\mathbf{P}|^{-1/2}\exp\{-\frac{1}{2\sigma^2}(\mathbf{X}-\mu)'\mathbf{P}^{-1}(\mathbf{X}-\mu)\}, \tag{39}$$

where $\Sigma = \sigma^2\mathbf{P}$. Maximum likelihood estimators of the parameters are obtained by maximizing (39) over the parametric space. No explicit formulas for

these estimators are known, not even in the case $p = q = 1$. For the moving average case see Cryer and Ledolter (1981) and Godolphin and De Gooijer (1982). However, asymptotic expressions for the biases of the estimators are known: Tanaka (1984) and Cordeiro and Klein (1994) give for the ARMA(1,1) model,

$$E(\hat{\alpha}-\alpha) = \frac{A_0}{T} + O(T^{-1}), \tag{40}$$

$$E(\hat{\beta}-\beta) = \frac{B_0}{T} + O(T^{-1}), \tag{41}$$

$$E(\hat{\sigma}_{ML}^2) = \begin{cases} -\frac{3\sigma^2}{T} + O(T^{-2}), & \text{if } \mu \text{ unknown;} \\ -\frac{2\sigma^2}{T} + O(T^{-2}), & \text{if } \mu = 0, \end{cases} \tag{42}$$

where A_0 and B_0 are complicated expressions of α and β.

The MLE of the residual variance is then

$$\hat{\sigma}_{ML}^2 = \frac{1}{T}(\mathbf{X}-\hat{\mu})'\hat{\mathbf{P}}^{-1}(\mathbf{X}-\hat{\mu}), \tag{43}$$

or

$$\hat{\sigma}_{ML}^2 = \frac{S(\hat{\alpha},\hat{\beta})}{T}, \tag{44}$$

where $S(\alpha, \beta)$ is given by (27) and $\hat{\alpha}, \hat{\beta}$ are the values of α, β which minimize the reduced likelihood

$$\ell(\alpha, \beta) = \frac{\ell n S(\alpha, \beta)}{T} + \frac{1}{T}\sum_{j=1}^{T} \ell n(v_{j-1}). \tag{45}$$

In our simulations the program SCA gives the MLE of the variance.

In the case of invertible models ($|\alpha| < 1$), and if the determinant of the covariance matrix appearing in the Gaussian likelihood function is asymptotically negligible compared with the sum of squares in its exponent, then the minimization of the sum of squares is equivalent to the maximization of the likelihood and LS and ML estimators have the same asymptotic properties, Brockwell and Davis (1991 Section 8.7).

5. Simulations

To verify empirically the conclusions of the theoretical results presented so far, we report in this section the findings of a simulation study. Twelve sets of parameters are considered for model (1), with $\mu = 0$ and pseudorandom $a_t \sim$ iid N(0, 1). Sample sizes $T = 100, 200, 400$ and $1,000$ replications for each case were taken. For the computations we have used the SPLUS and SCA packages.

For moment estimators, the relation $\hat{\beta} = -r_2/r_1$ yields, for some samples, values greater than one in absolute value, which violates the assumption of

stationarity. For these cases we computed a restricted estimator, denoted $\hat{\sigma}^2_{MM}$, with the results of the first $1,000$ replications having $|\beta| < 1$ reported. The generation of the samples and the computations were done with SPLUS.

For MLE and LSE the simulations were carried out with the SCA package, "option exact likelihood" and "conditional", respectively. Conditional LSE within SCA, denoted $\hat{\sigma}^2_{LS}$, are asymptotically equivalent to our estimators.

Tables 5.1 and 5.2 report the findings. The first shows the models that have at least one parameter close to the non-stationarity or non-invertibility boundaries; the second brings the models for which both parameters are far from these limits.

In each cell we present the estimated bias (EST. BIAS), the asymptotic bias (ASYM. BIAS) given by the theoretical formulas (for LS we used the same formula as for ML), the studentized bias (STU. BIAS) computed as (EST. BIAS -ASYM. BIAS)/(ST. ERROR) and finally the standard error (ST. ERROR) of the estimated bias, computed as $s/\sqrt{1,000}$, where

$$s^2 = \sum_{i=1}^{1000} (\hat{b}_i - \overline{b})^2/1000.$$

In this formula, $\hat{b}_i = \hat{\sigma}^2_i - 1$, $\overline{b} = \sum_{i=1}^{1000} \hat{b}_i/1000$ and $\hat{\sigma}^2_i$ is the variance estimated for each method. For example, for MM estimators, $\hat{\sigma}^2_i = \hat{\sigma}^2_{MM,i}$, $i = 1, \ldots, 1000$.

The results of this analysis can be summarized as follows:

i. For the models with parameter values far from the boudaries (Table 5.2), namely, values with absolute values not greater than 0.5, most of the estimated biases agree with the asymptotic theory, for all values of T ($|\text{STU. BIAS}|$) ≤ 2.0. Some exceptions occur for the method of moments, when $T \leq 200$ and α near 0.5. For MLE and LSE the constancy of the biases over changes in parameter values, and their fit by $-3/T$ is well supported by our results for all T, and further $|STU.BIAS| < 2$. This kind of behavior holds for MM only when $T = 400$ or $\alpha < 0.3$.

ii. For the model with $\alpha = -0.90$ and $\beta = 0.95$ (Table 5.1), that is, values near the boundaries, the results are bad ($|\text{STU. BIAS}| > 2.0$) for all methods and all sample sizes. Moreover, STU.BIAS tends to be smaller for MLE.

iii. For the models with only one parameter near the corresponding boundary, the behaviors were different:

a. For $\alpha = -0.8$ (Table 5.1) the estimated bias of MM does not agree with the asymptotic theory, even for $T = 400$. For $|\alpha| = 0.9$, that is, very near the noninvertibility limits, all methods behave badly, for all sample sizes. Sample sizes with $T > 400$ are probably necessary to get better results.

b. The previous behavior does not occur when only $|\beta|$ is near one. The first three models in Table 5.1. indicate that, for this situation, all methods work well, except LS, for $\alpha = -0.5$, $\beta = 0.9$ and $T = 100$.

iv. The asymptotic bias for the variance estimators are always negative, for LS and ML methods. The estimated biases in general follow this pattern, with a single exception for LS (see Table 5.1).

v. Moment estimators, as expected, are more sensitive to values of α near one, in absolute value, showing disagreement with the asymptotic theory for $|\alpha| \geq 0.8$. Moreover, the asymptotic values for MM estimators are, in most situations, bigger than those for LS and ML estimators. For some combinations of α and β values (see Figure 4), the biases are positive.

vi. We also notice that the numerical values of estimated standard errors, for a given sample size, tend to be similar across methods, a trend that does not hold for estimated or asymptotic biases.

6. Concluding Remarks

In this paper we considered the estimation of the residual variance in the ARMA(1,1) model. This variance is a nuisance parameter and its estimation is important because estimators enter into prediction errors, confidence intervals, tests of hypotheses, spectral estimators and other inferencial procedures.

In spite of the indicated usefulness, not many results are available about properties of estimators of the residual variance in ARMA models, except for maximum likelihood estimators under normality.

We considered estimation by three standard methods, namely moments, least squares and maximum likelihood under normality. In the analytical part of our work we concentrated in the study of the asymptotic biases of the estimators by using Taylor-type expansions and asymptotic results in the literature for means, variances, autocovariances and autocorrelations of linear processes. Small sample behavior of the estimators in terms of biases were studied through simulations.

Method of Moments. For the MM we obtained a closed-form expression for the asymptotic bias in the estimator defined by (12), namely (21), which is a function of α and β (besides σ^2 and T). This expression is studied graphically in Figure 5 for $-0.9 \leq \beta \leq 0.9$ and selected values $|\alpha| = 0.9$, 0.6 and 0.3. The graphs are of $T\ (ASYM.BIAS)/\sigma^2$, and they are computed with the corresponding value for the ML estimator, namely -3, for all values of the parameters. The three graphs show that:

i. For $|\alpha| = 0.9$ and $-0.9 \leq \beta \leq 0.9$ the value of T/σ_a^2. ASYM.BIAS is large and negative, indicating an underestimation of σ^2;

ii. For $|\alpha| = 0.6$ and $-0.7 \leq \beta \leq 0.7$, the above quantity assumes values which are smaller (around -9), indicating also an underestimation of σ^2, but with a smaller bias;

iii. For $|\alpha| = 0.3$ and $-0.65 \leq \beta \leq 0.65$, T/σ^2. ASYM.BIAS assumes negative values, but very small and often smaller than the corresponding ML ones.

The estimated bias of the variance by the method of moments will be always positive when β approaches the nonstationarity limit, $|\beta| \geq 0.70$, and if α is away from the noninvertibility region, $|\alpha| \leq 0.70$. But this asymptotic bias will never be positive if both parameters are negative.

For values of $|\alpha|$ not very close to one, $|\alpha| \leq 0.6$, we have reasonable values for the asymptotic biases, when compared with the ML estimators biases, for $-0.65 \leq \beta \leq 0.65$ (not too close to the nonstationarity region). Figure 6 confirms the above conclusions. When α approaches the noninvertibility limits, large sample sizes are necessary to give estimates compatible with the asymptotic theory. In our simulations we showed that the behavior of the MM estimator is comparable to LS and ML in terms of our proposed figures of merit, provided both parameters have small values, and $T = 400$ (Table 5.2). For $|\alpha| > 0.8$ (Table 5.1) samples larger than $T = 400$ are necessary for all estimation methods work properly. This important and not well known fact holds independently of the values of β.

Least Squares. Our analysis for the LS estimator is less complete. We obtained an asymptotic closed-form representation of the estimator, namely (35) and an asymptotic closed-form expression for the bias, namely (37), for which we lack analytical approximations for the covariances between the sample covariances and the estimators of the parameters. We performed some simulations with LS estimators available in standard computer programs, and found that their behavior is reasonable.

The simulations confirm expected results, that ML and LS estimators perform better than MM or that better fits are obtained for large sample sizes. The fit of the simulated results is much better for ML and LS than for MM procedures.

In conclusion, correction for biases when using ML estimators is simple, since the correction does not depend on the values of the model parameters. For LS and MM estimators the correction will include a more complex function of all parameters, which, in practice will have to be estimated. We should also expect considerable biases near the limits of the admissibility regions for MM estimators.

Acknowledgements

This work was done with partial support of Fundación Antorchas, Argentina. The authors acknowledge partial support from CONICET, Argentina (RPM) and CNPq, Brazil (PAM and CMCT).

Appendix

A1. Proof of (21)

Expression (12) gives a formula of $\hat{\sigma}^2$ as a function of $c_0, \hat{\alpha}$ and $\hat{\beta}$, and is the sample analog of σ^2 defined in (2). We now consider (20): $E(c_0 - \gamma_0)$ is given by (11) and the expectations $E(\hat{\alpha} - \alpha)^r(\hat{\beta} - \beta)^s$, for $0 \leq r + s \leq 2$ are given by (15)-(19).

The necessary inputs to apply formulas (15)-(19) are:

(a) Calculate the first and second-order derivatives of α and β with respect to ρ_k; this can be done by (13) where $\rho_1 = F(\alpha, \beta), \rho_2 = G(\alpha, \beta)$, so $\alpha = f(\rho_1, \rho_2), \beta = g(\rho_1, \rho_2)$.

Denoting the various derivatives with subindices, the first and second-order derivatives of the inverse functions f and g are determined by the matrix relation $\mathbf{AB} = \mathbf{I}_6$ where

$$\mathbf{A} = \begin{pmatrix} F_1 & F_2 & F_{11} & F_{12} & F_{21} & F_{22} \\ G_1 & G_2 & G_{11} & G_{12} & G_{21} & G_{22} \\ 0 & 0 & F_1^2 & F_1F_2 & F_2F_1 & F_2^2 \\ 0 & 0 & F_1G_1 & F_1G_2 & F_2G_1 & F_2G_2 \\ 0 & 0 & G_1F_1 & G_1F_2 & G_2F_1 & G_2F_2 \\ 0 & 0 & G_1^2 & G_1G_2 & G_2G_1 & G_2^2 \end{pmatrix}$$

and

$$\mathbf{B} = \begin{pmatrix} f_1 & f_2 & f_{11} & f_{12} & f_{21} & f_{22} \\ g_1 & g_2 & g_{11} & g_{12} & g_{21} & g_{22} \\ 0 & 0 & f_1^2 & f_1f_2 & f_2f_1 & f_2^2 \\ 0 & 0 & f_1g_1 & f_1g_2 & f_2g_1 & f_2g_2 \\ 0 & 0 & g_1f_1 & g_1f_2 & g_2f_1 & g_2f_2 \\ 0 & 0 & g_1^2 & g_1g_2 & g_2g_1 & g_2^2 \end{pmatrix}$$

The results can be checked by repeated application of the chain rule.

(b) From (9) and (10) evaluate $E(r_j - \rho_j), E(r_j - \rho_j)^2, j = 1, 2$ and $E(r_1 - \rho_1)(r_2 - \rho_2)$.

In (20) it remains to evaluate $E(c_0 - \gamma_0)(\hat{\alpha} - \alpha)$ and $E(c_0 - \gamma_0)(\hat{\beta} - \beta)$. From (13) we can expand $(\hat{\alpha} - \alpha)$ in a Taylor expansion up to order $O(T^{-1})$ and multiplying by $(c_0 - \gamma_0)$ we have that, to this same order,

$$(c_0-\gamma_0)(\hat{\alpha}-\alpha)\simeq(c_0-\gamma_0)(r_1-\rho_1)\dot{\alpha}_1+(c_0-\gamma_0)(r_2-\rho_2)\dot{\alpha}_2.$$

Also,

$$r_j-\rho_j\simeq-(c_0-\gamma_0)\gamma_j/\gamma_0^2+(c_0-\gamma_0)^2\gamma_j/\gamma_0^3+\gamma_0^{-1}(c_j-\gamma_j)-\gamma_0^{-2}(c_0-\gamma_0)(c_j-\gamma_j)$$

and finally

$$E(c_0-\gamma_0)(\hat{\alpha}-\alpha)=[\frac{-\gamma_1}{\gamma_0^2}E(c_0-\gamma_0)^2+\frac{1}{\gamma_0}E(c_0-\gamma_0)(c_1-\gamma_1)]\dot{\alpha}_1$$
$$+[\frac{-\gamma_2}{\gamma_0^2}E(c_0-\gamma_0)^2+\frac{1}{\gamma_0}E(c_0-\gamma_0)(c_2-\gamma_2)]\dot{\alpha}_2+o(T^{-1}), \qquad (A.1)$$

which can be evaluated using (8).

The same procedure is used to find an entirely similar expression for $E(c_0-\gamma_0)(\hat{\beta}-\beta)$, replacing in (A.1) $\dot{\alpha}_i$ by $\dot{\beta}_i, i=1,2$.

Finally, substituting all these expressions in (20) leads to (21).

A.2. Proof of (37)

By (35) and (36) we have that

$$\hat{\sigma}^2_{LS}-\sigma^2=c^*_{00}-\gamma_0+(\hat{\beta}-\hat{\alpha})\sum_{j=1}^{B_T}(-\hat{\alpha})^{j-1}c^*_{j0}-(\beta-\alpha)\sum_{j=1}^{\infty}(-\alpha)^{j-1}\gamma_j$$

Hence, an expansion of up to second order terms leads to

$$E(\hat{\sigma}^2_{LS}-\sigma^2)=E(c^*_{00}-\gamma_0)+BE(\hat{\alpha}-\alpha)+CE(\hat{\beta}-\beta)$$

$$+\sum_{j=1}^{B_T}D_jE(c^*_{j0}-\gamma_j)+\sum_{j=1}^{B_T}M_jE(c^*_{j0}-\gamma_j)(\hat{\alpha}-\alpha)+\sum_{j=1}^{B_T}F_jE(\hat{\beta}-\beta)(c^*_{j0}-\gamma_j)$$
$$+GE(\hat{\alpha}-\alpha)(\hat{\beta}-\beta)+\frac{I}{2}E(\hat{\alpha}-\alpha)^2+o(T^{-1}). \qquad (A.2)$$

where the needed derivatives are

$$B=\frac{\partial\sigma^2}{\partial\alpha},C=\frac{\partial\sigma^2}{\partial\beta},D_j=\frac{\partial\sigma^2}{\partial\gamma_j},$$

$$M_j=\frac{\partial^2\sigma^2}{\partial\alpha\partial\gamma_j},\ F_j=\frac{\partial^2\sigma^2}{\partial\beta\partial\gamma_j},\ G=\frac{\partial^2\sigma^2}{\partial\alpha\partial\beta},\ I=\frac{\partial^2\sigma^2}{\partial\alpha^2}.$$

Replacing $E(c^*_{jk}-\gamma_{j-k}), j,k=0,\ldots,B_T$, we obtain (37).

Note: The computations were done with the Mathematica program (Wolfram, 1988).

References

ANDERSON, T.W. (1971). *The Statistical Analysis of Time Series.* New York: John Wiley and Sons.

BOX, G.E.P. JENKINS, G.M. and REINSEL, G. (1994). *Time Series Analysis: Forecasting and Control.* Third Edition. New Jersey: Prentice-Hall.

BROCKWELL, P.J. and DAVIS, R.A. (1991). *Time Series: Theory and Methods.* New York: Springer Verlag.

CORDEIRO, G.M. and KLEIN, R. (1994). Bias Correction of Maximum Likelihood Estimates for ARMA Models. *Probability and Statistics Letters.* :169-176.

CRYER, J.D. and LEDOLTER, J. (1981). Small Sample Properties of the Maximum Likelihood Estimator in the First Order Moving Average Model. *Biometrika.* **68**: 691-694.

DE GOOIJER, J.D. and PUKKILA, T. (1994). On the Expectation of Estimators for General ARMA Processes. *Statistica*, anno LIV. (1): 39-50.

DURBIN, J. (1959). Efficient Estimation of Parameters in Moving Average Models. *Biometrika.* **46**: 306-316.

EFRON, B. and TIBSHIRANI, R. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

FULLER, W.A. (1996). *Introduction to Statistical Time Series.* Second Edition. New York: John Wiley and Sons.

GODOLPHIN, E.J. and DE GOOIJER, J.G. (1982). On the Maximum Likelihood Estimation of the Parameters of a Gaussian Moving Average Process. *Biometrika.* **69**: 443-451.

MathSoft, Inc. (1998). The SPlus 4.5. Seattle: Washington.

MENTZ, R.P., MORETTIN, P.A. and TOLOI, C.M.C. (1997). Residual Variance Estimation in Moving Average Models. *Communications in Statistics - Theory and Methods.* **26**: 1905-1923.

MENTZ, R.P., MORETTIN, P.A. and TOLOI, C.M.C. (1998). On Residual Variance Estimation in Autoregressive Models. *Journal of Time Series Analysis.* **19**: 187-208.

MENTZ, R.P., MORETTIN, P.A. and TOLOI, C.M.C. (1999). On Least Squares Estimation of the Residual Variance in the First Order Moving Average Model. *Computational Statistics and Data Analysis.* **29**: 485-499.

Scientific Computing Associates. (1993). The SCA Statistical System (Release V.2). Lisle, Illinois.

TANAKA, K. (1984). An Asymptotic Expansion Associated with the Maximum Likelihood Estimators in ARMA Models. *Journal of the Royal Statistical Society.* Series B. 46: 58-67.

WOLFRAM, S. (1988). *Mathematica: A System for Doing Mathematics by Computer.* Reading: Addison-Wesley.

TABLE 5.1

Results of a simulation study. Estimated biases of residual variance (with standard error), asymptotic biases and studentized bias for moment, least squares and maximum likelihood estimators in the ARMA(1,1) model.

T		$\alpha = 0.4, \beta = -0.7$			$\alpha = 0.4, \beta = 0.95$			$\alpha = -0.5, \beta = 0.9$		
		MM	LS	ML	MM	LS	ML	MM	LS	ML
100	Est. Bias	0.0072	-0.0305	-0.0311	0.0391	-0.0324	-0.0328	0.2292	-0.0207	-0.0340
	(St. Error)	(0.0052)	(0.0044)	(0.0044)	(0.0053)	(0.0043)	(0.0043)	(0.0093)	(0.0046)	(0.0044)
	Asym. Bias	0.0051	-0.0300	-0.0300	0.0345	-0.0300	-0.0300	0.2122	-0.0300	-0.0300
	Stu. Bias	0.397	-0.108	-0.243	0.869	-0.543	-0.647	1.824	2.041	-0.900
200	Est. Bias	0.0069	-0.0122	-0.0168	0.0230	-0.0177	-0.0179	0.1119	-0.0126	-0.0140
	(St. Error)	(0.0035)	(0.0032)	(0.0031)	(0.0036)	(0.0031)	(0.0031)	(0.0057)	(0.0032)	(0.0031)
	Asym. Bias	0.0026	-0.0150	-0.0150	0.01726	-0.0150	-0.0150	0.1061	-0.0150	-0.0150
	Stu. Bias	1.244	0. 857	-0.589	1.625	-0.896	-0.957	1.033	0.731	0.309
400	Est. Bias	0.0031	-0.0053	-0.0086	0.0118	-0.0081	-0.0082	0.0543	-0.0081	-0.0079
	(St. Error)	(0.0024)	(0.0022)	(0.0022)	(0.0024)	(0.0022)	(0.0022)	(0.0034)	(0.0022)	(0.0022)
	Asym. Bias	0.0013	-0.0075	-0.0075	0.0086	-0.0075	-0.0075	0.0530	-0.0075	-0.0075
	Stu. Bias	0.749	1.009	-0.484	1.329	-0.285	-0.330	0.369	-0.252	-0.176

TABLE 5.1 (continued)

Results of a simulation study. Estimated biases of residual variance (with standard error), asymptotic biases and studentized bias for moment, least squares and maximum likelihood estimators in the ARMA(1,1) model.

T		$\alpha = -0.8, \beta = -0.6$			$\alpha = 0.9, \beta = 0.3$			$\alpha = -0.9, \beta = 0.95$		
		MM	LS	ML	MM	LS	ML	MM	LS	ML
100	Est. Bias	0.0127	-0.0185	-0.0333	0.0984	0.0583	-0.0684	1.1605	-0.1194	-0.0427
	(St. Error)	(0.0048)	(0.0045)	(0.0045)	(0.0056)	(0.0049)	(0.0046)	(0.0261)	(0.0026)	(0.0056)
	Asym. Bias	-0.9531	-0.0300	-0.0300	-8.3520	-0.0300	-0.0300	-4.3495	-0.0038	-0.0038
	Stu. Bias	22.356	2.543	-0.729	1,508.1	-5.815	-8.284	210.79	-34.797	-2.256
200	Est. Bias	0.0137	-0.0089	-0.0161	0.0926	0.0331	-0.0527	0.6881	- 0.0746	-0.0313
	(St. Error)	(0.0035)	(0.0032)	(0.0032)	(0.0041)	(0.0032)	(0.0037)	(0.0152)	(0.0025)	(0.0037)
	Asym. Bias	-0.4766	-0.0150	-0.0150	-4.1760	-0.0150	-0.0150	-2.1748	-0.0150	-0.0150
	Stu. Bias	140.89	1.911	-0.342	1,051.4	15.037	-10.281	187.98	-23.836	-4.379
400	Est. Bias	0.0122	-0.0048	-0.0102	0.0888	0.0208	-0.0370	0.4280	-0.0097	-0.0236
	(St. Error)	(0.0024)	(0.0022)	(0.0023)	(0.0031)	(0.0023)	(0.0029)	(0.0093)	(0.0044)	(0.0022)
	Asym. Bias	-0.2383	-0.0075	-0.0075	-2.0880	-0.0075	-0.0075	-1.0874	-0.0075	-0.0075
	Stu. Bias	104.80	1.239	-1.176	711.36	12.344	-10.398	163.47	-0.501	-8.107

TABLE 5.2

Results of a simulation study. Estimated biases of residual variance (with standard error), asymptotic biases and studentized bias for moment, least squares and maximum likelihood estimators in the ARMA (1,1) model.

T		$\alpha = 0.05, \beta = 0.4$			$\alpha = 0.1, \beta = 0.4$			$\alpha = 0.3, \beta = 0.1$		
		MM	LS	ML	MM	LS	ML	MM	LS	ML
100	Est. Bias	-0.0204	-0.0332	-0.0336	-0.0198	-0.0413	-0.0301	-0.0204	-0.0345	-0.0360
	(St. Error)	(0.0045)	(0.0044)	(0.0044)	(0.0045)	(0.0043)	(0.0044)	(0.0045)	(0.0044)	(0.0044)
	Asymp. Bias	-0.0272	-0.0300	-0.0300	-0.0284	-0.0300	-0.0300	-0.0369	-0.0300	-0.0300
	Stu. Bias	1.526	-0.724	-0.837	1.913	-2.639	-0.030	3.665	-1.034	-1.370
200	Est. Bias	-0.0073	-0.0181	-0.0182	-0.0116	-0.0205	-0.0163	-0.0124	-0.0186	-0.0193
	(St. Error)	(0.0032)	(0.0031)	(0.0031)	(0.0032)	(0.0030)	(0.0031)	(0.0033)	(0.0031)	(0.0030)
	Asym. Bias	-0.0136	-0.0150	-0.0150	-0.0142	-0.0150	-0.0150	-0.0185	-0.0150	-0.0150
	Stu. Bias	1.950	-1.010	-1.003	0.809	-1.829	-0.416	1.847	-1.177	-1.404
400	Est. Bias	-0.0046	-00820	-0.0082	-0.0068	-0.0075	-0.0066	-0.0072	-0.0082	-0.0083
	(St. Error)	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0023)	(0.0022)	(0.0022)	(0.0022)
	Asym. Bias	-0.0068	-0.0075	-0.0075	-0.0071	-0.0075	-0.0075	-0.0092	-0.0075	-0.0075
	Stu. Bias	0.987	-0.311	-0.320	0.113	0.000	0.403	0.923	-0.306	0.374

TABLE 5.2 (continued)

Results of a simulation study. Estimated biases of residual variance (with standard error), asymptotic biases and studentized bias for moment, least squares and maximum likelihood estimators in the ARMA(1,1) model.

T		$\alpha = 0.4, \beta = 0.05$			$\alpha = 0.4, \beta = -0.2$			$\alpha = 0.5, \beta = -0.3$		
		MM	LS	MM	MM	LS	MM	MM	LS	MM
100	Est. Bias	-0.0192	-0.0338	-0.0363	-0.0221	-0.0331	-0.0352	-0.0192	-0.0330	-0.0372
	(St. Error)	(0.0045)	(0.0044)	(0.0044)	(0.0047)	(0.0044)	(0.0044)	(0.0049)	(0.0044)	(0.0045)
	Asym. Bias	-0.0457	-0.0300	-0.0300	-0.0409	-0.0300	-0.0300	-0.0551	-0.0300	-0.0300
	Stu. Bias	5.850	-0.874	-1.431	4.000	-0.716	-1.173	7.359	-0.696	-1.584
200	Est. Bias	-0.0160	-0.0182	-0.0189	-0.0155	-0.0181	-0.0184	-0.0173	-0.0181	-0.0188
	(St. Error)	(0.0033)	(0.0031)	(0.0031)	(0.0034)	(0.0030)	(0.0030)	(0.0035)	(0.0030)	(0.0031)
	Asym. Bias	-0.0229	-0.0015	-0.0150	-0.0204	-0.0150	-0.0150	-0.0275	-0.0150	-0.0150
	Stu. Bias	2.123	-1.036	-1.282	1.465	-1.033	-1.112	2.897	-1.010	-1.249
400	Est. Bias	-0.0104	-0.0081	-0.0083	-0.0094	-0.0081	-0.0082	-0.0133	-0.0081	-0.0083
	(St. Error)	(0.0023)	(0.0022)	(0.0022)	(0.0023)	(0.0022)	(0.0022)	(0.0026)	(0.0022)	(0.0022)
	Asym. Bias	-0.0114	-0.0075	-0.0075	-0.0102	-0.0075	-0.0075	-0.0138	-0.0075	-0.0075
	Stu. Bias	0.439	-0.288	-0.342	0.351	-0.284	-0.329	0.204	-0.257	-0.351

Figure 1

(T/σ^2) *Asym. Bias for MM Estimators.*

Figure 2

(T/σ^2) *Asym. Bias for MM Estimators.*

Figure 3

(T/σ^2) *Asym. Bias for MM Estimators in the special case of* $\alpha = \beta$.

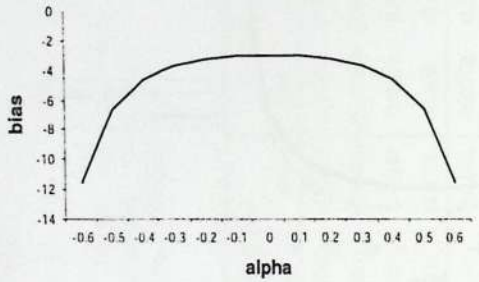


Figure 4

(T/σ^2) *Asym. Bias for MM Estimators.*

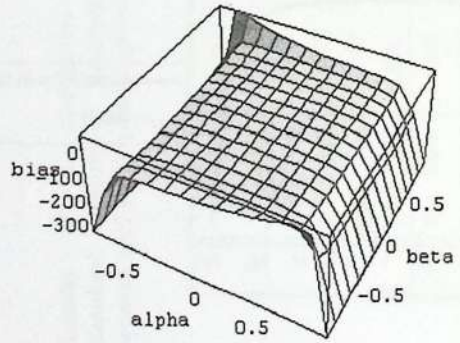
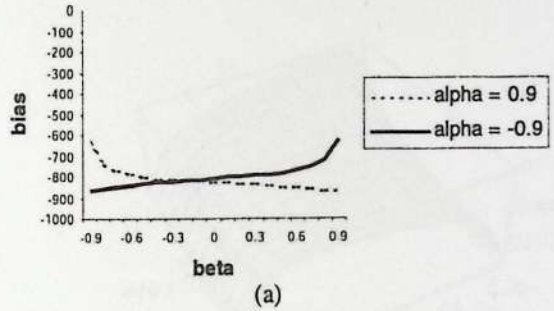
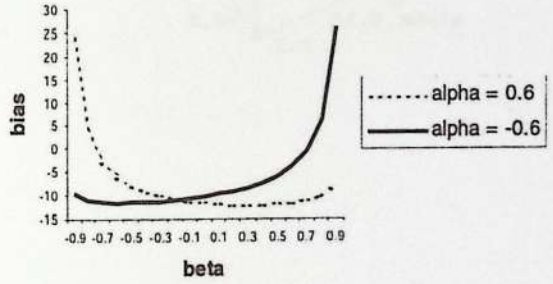


Figure 5

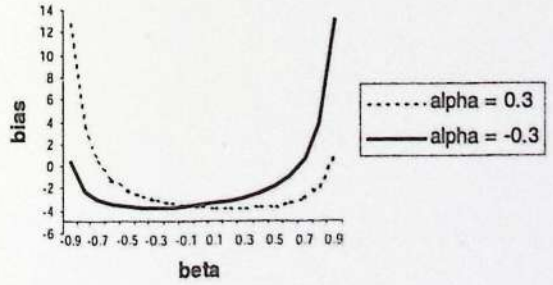
(T/σ^2) *Asym. Bias for MM and ML Estimators for some values of* α *and* $-0.9 \leq \beta \leq 0.9$.



(a)



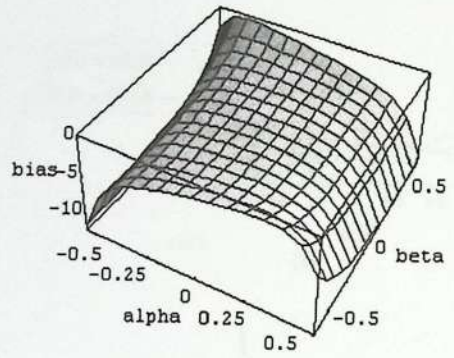
(b)



(c)

Figure 6

(T/σ^2) *Asym. Bias for MM Estimators for* $|\alpha| \leq 0.6$ *and* $|\beta| \leq 0.65$.



Robust estimation in time series

Raúl P. Mentz and Carlos I. Martínez

Facultad de Ciencias Económicas
Universidad Nacional de Tucumán and CONICET, Argentina

Abstract

The main purpose of this work is to study empirically by means of simulations, the robustness of a set of proposals to estimate the parameters in the MA(1) time series model. The non-normal populations are mixtures of normal distributions, defined by $g(x) = pN(0,k) + (1-p)N(0,1)$, where the proportion of contamination most frequently used is $p = 0.10$ and k is the variance of the distribution used in the contamination; α is taken to be 0.90, which is close to the region of non-invertibility. Key results are that the estimation procedures used in the study provide good results in terms of biases in the estimation of the parameters, and that the biases are not changed when contaminated errors (mixtures) are considered. The estimation of the variance of the contaminated errors is also studied through simulations.

Key Words: Maximum likelihood estimation, contaminated errors, robustness, estimation of error variance, biases.

AMS subject classification: 62M10, 62F35

1. Introduction

Many estimation procedures used in time series analysis are deduced under restrictive assumptions. For example to estimate the parameters in ARMA models, frequently it is assumed that errors form a white noise process: they are independent (or at least non-correlated), with constant expected value (usually taken to be 0), and constant, finite variance. Under these assumptions, or similar ones, they are used computationally and their basic theoretical properties are studied; this applies, for example, to the method of moments or similar procedures (Burg's algorithm, for example), or some version of the least squares procedure.

*Correspondence to: Raúl P. Mentz, Casilla de Correo 209, (4000) Tucumán, Argentina, Fax 54 (381) 4364-105, phone 54 (381)4364-093. E-mail: rmentz@herrera.unt.edu.ar

Received: April 2001; Accepted: April 2002

With the additional assumption of normality, the method of maximum likelihood is frequently considered. Experiences in the use and analysis of this procedure are in general positive: asymptotic results are known covering various types of investigations, and studies conducted through simulations have frequently shown that the asymptotic results can be used with moderate and even small samples. In general, simulation studies use values generated by normal distributions.

The assumption of normality is often too demanding for applications. Several approaches have been used in the literature to deal with departures from this assumption in the case of ARMA models. We shall briefly review some of them.

One possibility is to apply methods to deal with outliers. The presence of outliers is an indication of the normality of the series. Fox (1972) discussed the idea of additive (AO, Type I) and innovations (IO, Type II) outliers. The former are the effect of external or exogenous causes and the latter of internal or endogenous causes. Chang et al. (1988) used the technique of intervention analysis (Box and Tiao (1975)) to deal with both kinds of outliers. Peña (1990) dealt with measuring the influence of outliers (Cook and Weisberg (1982)); in this approach observations are deleted and the effect of such deletions on the estimates is measured: a high influence means that the deletion strongly affects the estimates.

The treatment of outliers can also be approached through the use of robust statistical procedures. An exposition is Martin (1980) who dealt with autoregressive models. The basic idea is to replace the common weighting schemes (quadratic or absolute value), by more elaborate ones. A collection of these procedures is in Andrews et al. (1972), as will be discussed below.

Another approach is to use for the error term some non-normal distribution. Tikku et al. (2000) considered autoregressive models with errors terms modelled by a symmetric family of distributions, namely Student's t. These are known to have heavy tails for small degrees of freedom. These authors derive a modified maximum likelihood estimation procedure and show that it has good properties in estimation and testing problems.

An alternative related to the previous paragraph, is to use mixtures of normal distributions as models for the error term. In the frequently-quoted study Andrews et al. (1972), the behavior of a collection of as many as 68 point estimators of location was studied, but not in the context of

time series. These authors used simulation and considered a wide variety of distributions. In particular, they used mixtures of normal distributions with equal expected values and differing variances, which are known in the literature as robustness models (Lindsay (1995), Section 1.3.12, Aitkin and Tunnicliffe (1980)).

In the present study we consider the first order moving average model with errors generated by robustness models. Through simulations we compare an iterative estimation procedure presented in Anderson and Mentz (1993b) with methods available in five well-known computer programs. Our main objective is to evaluate whether the use of mixture in the error term, affects the outcome of the procedures by introducing biases in the estimation of the moving average parameter.

2. The first-order moving average model and mixtures of two normal densities

The MA(1) (first-order moving average) model assumes that an observable time series y_t is generated by

$$y_t = u_t + \alpha u_{t-1}, \tag{2.1}$$

where the u_t are Gaussian white noise, that is, independent $N(0, \sigma^2)$ random variables. Instead of the parameters (α,σ^2), often the pair $(\sigma_0,\ \rho)$ is considered, where $\sigma_0 = \sigma^2(1+\alpha^2) = E(y_t^2)$ is the variance of y_t, and $\rho = \alpha/(1+\alpha^2) = E(y_t y_{t-1})/E(y_t^2)$ is the first-order autocorrelation coefficient. The relation between these parameters is $\rho = \alpha/(1+\alpha^2)$ and $\alpha = \left\{1-(1-4\rho^2)^{1/2}\right\}/(2\rho)$. The invertibility regions are $|\alpha| < 1$ and $|\rho| < 1/\{2\cos[\pi/(T+1)]\}$ (Anderson and Takemura (1986)). In general r is used to designate an estimator of ρ. The results are given in terms of ρ and only by exception in terms of α.

Instead of normal errors we consider the mixture of two normal densities, with the same expected values and different variances: $N(\theta, \sigma^2)$ y $N(\theta, k\sigma^2)$, where $k > 0$. Following Lindsay (1995), we write the mixture as $pN(\theta, k\sigma^2) + (1-p)N(\theta, \sigma^2)$, where $0 < p < 1$. Without loss of generality we take $\theta = 0$ y $\sigma = 1$.

The simulation procedure consists in generating pseudorandom independent numbers, uniformly distributed on $[0, 1]$, and to consider this interval divided by p. The experiment is interpreted as the selection of the

$N(0,k)$ density with relative expected proportion p, and density $N(0,1)$ with relative expected proportion $1-p$. When p is small, this is a model for the generation of a proportion p of outliers.

Hence, the model density can be written as

$$g(x;p,k) = p\frac{1}{\sqrt{2\pi k}}\exp\left\{-\frac{x^2}{2k}\right\} + (1-p)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}, \quad -\infty < x < \infty \tag{2.2}$$

where p and k are the model parameters.

The central moments of this density are obtained by integrating over x, so that they "may be expressed as weighted sums (using the same weights) of the expectations calculated using fixed numbers" (Andrews et al. (1972), Section 4D3). Hence, the density 2.1 has 0 expected value and variance

$$Var(X) = E(X^2) = pk + (1-p) \tag{2.3}$$

depending on the sampling fraction p and the scale factor k.

Given that the selection of the uniform random variables is done independently, it follows that the random variables with the mixture distribution are also independent. In the time series terminology, if X_t denotes the stochastic process generated in the indicated way, it constitutes a white noise process: independent random variables with 0 expected value and constant variance $pk + (1-p)$.

Following Andrews et al. (1972) (Section 5.4, Table 5-2), we take $k = 1, 9$ and 100: the first value gives random variables which are independent and identically distributed $N(0,1)$, and the other two correspond to standard deviations of 3 and 10, respectively.

In Figure 1 three densities are compared, namely: the standard normal $N(0,1)$, the mixture $0.5N(0,1) + 0.5N(0.9)$ (which has variance equal to 5, according to (2.3)), and the $N(0,5)$ density. The three densities are symmetric with respect to 0, but while the first and third ones are normal, the mixture is clearly non-normal, has "heavy tails", that is to say, it assigns high densities to values far from its 0 expected value.

Figures 2 and 3 compare de $N(0,1)$ density with mixtures having $k = 9$ and $k = 100$ respectively.

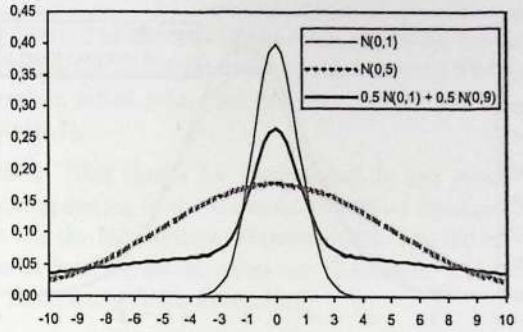


Figure 1: Comparison of the densities $N(0,1)$ and $N(0,5)$ with the mixture $0.5\ N(0,1) + 0.5N(0,9)$.

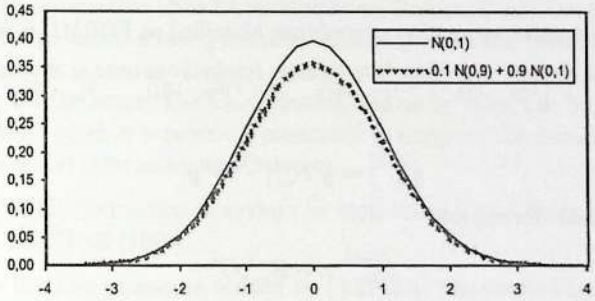


Figure 2: Comparison of the mixtures used in Table 1. $N(0,1)$ and mixture $0.1\ N(0,9) + 0.9\ N(0,1)$

3. Estimation procedures in time series analysis

In terms of the parameters of the MA(1) model, the likelihood function of a vector of observations $y = (y_1, ..., y_T)'$ is

$$L(\sigma_0, \rho) = (2\pi\sigma_0)^{-T/2}\,|R|^{-1/2}\exp\left\{-y'R^{-1}y/(2\sigma_0)\right\} \tag{3.1}$$

where R is the $T \times T$ autocorrelation matrix: $R = I + \rho G$, and G has 1's in its two diagonals adjacent to its main diagonal and 0's elsewhere.

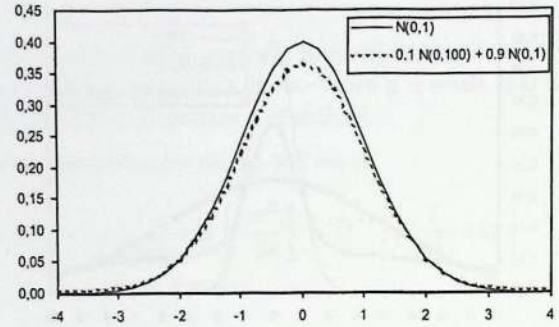


Figure 3: Comparison of the mixtures used in Table 1. $N(0,1)$ *and mixture* $0.1\ N(0,100) + 0.9\ N(0,1)$

One iterative estimation procedures, identified as FORM1, is written

$$FORM1: \left\{t_{22}^{(i-1)}q_{10}^{(i-1)} - t_{21}^{(i-1)}q_{11}^{(i-1)}\right\} = r_i t_{20}^{(i-1)}q_{11}^{(i-1)} - t_{21}^{(i-1)}q_{10}^{(i-1)}, \quad (3.2)$$

where

$$q_{jk}^{(i-1)} = \boldsymbol{y}'\boldsymbol{R}_{i-1}^{-(j+1)}\boldsymbol{G}^k\boldsymbol{y}, \quad (3.3)$$

are quadratic forms, and

$$t_{jk}^{(i-1)} = tr\boldsymbol{R}_{i-1}^{-1}\boldsymbol{G}^k \quad (3.4)$$

are traces. The computational details and other analyses of this proposal can be found in Anderson and Mentz (1993a).

This iterative procedure is derived from the normal likelihood function when the method of scoring is used in the expansion of the log-likelihood. Other procedures are derived in Anderson and Mentz (1993b) by expanding the log-likelihood by the Newton-Raphson method, and also by using the two expansions with the concentrated likelihood functions. These procedures are mathematically equivalent, but they lead to different estimating equations, and hence often to different values of the estimates.

The estimation procedure defined by (3.2), (3.3), and (3.4) was compared in a simulation study with the following five:

1. *BMDPCls.* (Cls stands for conditional least squares) This is a preliminary estimation procedure, a variant of the method of moments, available

in BMDP (1990). The estimate of the main parameter minimizes the sum of squares appearing in the exponent of the Gaussian likelihood function, assuming certain initial values for the error terms. See Box and Jenkins (1970)(Chapter 7).

2. *BMDPBak.* (Bak stands for backcasting) In the previous procedure, the term corresponding to the Gaussian likelihood function is omitted. In the present one the full expression is considered. The initial values are not assumed to be fixed values, but they are "forecasted": this originates the expression backcasting (back-forecasting).

3. *S-PLUS.* Estimation procedure available in the package with this name. The procedure minimizes a likelihood function conditional on a set of initial values (Venables and Ripley (1997), Section 15.2). The program includes a set of alternative computations for the case of missing observations.

4. *ITSM.* (Interactive Time Series Modelling, Brockwell and Davis (1991)). This package includes two types of estimation procedures, preliminary and final, and this in turn can be least squares or Gaussian maximum likelihood; we only use the latter. The maximization is done by using the "innovations algorithm", which is a recursive procedure to compute the one-step ahead predictors and their mean square errors.

5. *MINITAB.* Estimation procedure for time series that follow an ARIMA model. MINITAB (1996).

The iterative procedure defined by (3.2)-(3.4) was studied by means of simulations in Anderson et al. (1996). In this paper the procedures mentioned above, were compared with the preliminary estimator of ρ given by the first-order sample autocorrelation r, and with the preliminary estimation procedure BMDPCls. Considering only pseudorandom normal numbers, the study by simulations detected that (3.4) operates quite satisfactorily for $T = 100$ when $\alpha = 0.30$, a value not in the non-invertibility region given by $|\alpha| > 1$; in this case, (3.4) substantially improves the simple estimator r, and BMDPCls provides results of a quality comparable with (3.4). However, when $\alpha = 0.90, T = 100$ was insufficient to justify the use of the asymptotic approximations known for the estimators and their standard errors. $T = 250$ was used and the fit to the asymptotic theory improved considerably. In the analysis the use of standard errors coming from the asymptotic theory was emphasized, as a means of evaluating the statistical significance of the simulation results; $n = 100$ replicates were

done in each case.

The present paper has a structure similar to that in the cited publication, but here non-normal errors in model (2.1), generated by mixtures of normals as defined in (2.2) are used.

4. Description and analysis of the results

4.1. Simulation design

We consider samples sizes $T = 100$ or 250 and repetitions $n = 100$ or 500. We take the pair $(T, n) = (100, 500)$ as an example.

To generate the MA(1) observations with errors defined by (2.1), for each repetition the following steps produce the desired result:

1. Generate independent pseudorandom $N(0,1)$ numbers, by using Wolfram (1991). These are denoted $u_1, u_2, \ldots, u_{100}$.
2. Generate a value $\hat{p}$ from the uniform distribution on $(0,1)$.
3. Transform by defining $z_i^c = \sqrt{k}u_i$, if $u_i \leq \hat{p}$, $z_i^c = u_i$, if $u_i > \hat{p}$.
4. Model: $y_i = z_i^c + \alpha z_{i-1}^c$, $i = 1, 2, ..., 100$.
5. Use the y_i to estimate the parameters of the MA(1) model.
6. Repeat $n = 500$ times.

To approximate the idea of independence, between each set of T observations, 50 were discarded. With a given set of y's, all the estimation procedures were applied; this design tries to control variability, at least in part: if the numbers were changed from the calculations for one method to the next one, differences due to these numbers will be added to the simulation process itself.

Calculations were done for three values of k, the variance of the contaminating distribution $N(0,k)$. We choose $k = 1, 9$ and 100, as was indicated in Section 2.

In the following subsection we describe the structure of Table 1, which contains many of the basic findings of our simulation study.

4.2. Structure of table 1

Considering $r_i, i = 1, 2, ..., n$, the values coming from the n repetitions, Table 1 contains in its columns: (1)$m = m(r) = \sum_{i=1}^{n} r_i/n$, the estimates average, whose estimated standard error is $s/\sqrt{n}$, where s is in column (4); (2) $bi(r) = m - \rho$, the average bias, whose standard error is also $s/\sqrt{n}$, and is included in this column; (4)$s = s(r) = \left\{\sum_{i=1}^{n}(r_i - m)^2/(n-1)\right\}^{1/2}$, the standard deviation of the r_i, whose estimated asymptotic standard error is $s/\{2(n-1)\}^{1/2}$; (5) $as_m(r) = \left\{T^{-1}\left[1-\alpha^2\right]^3/\left[1+\alpha^2\right]^4\right\}^{1/2}$, the estimated asymptotic standard deviation of r; (6) $rs_\rho(r) = s/as_\rho(r)$, the empirical standard deviation of the r_i divided into its asymptotic standard deviation, whose estimated standard error is the standard error of r divided by the value of (5); (7) $mse(r) = s^2(r) + bi^2(r)$, the estimated mean squared error. The average estimate of α $m(\alpha)$, is also reported.

In columns (2) and (3), dividing the average estimates by their estimated standard errors, we obtain asymptotic tests of the null hypotheses that the true values are 0. For example, in the case of FORM1 and $k = 1, bi(r)/\{s/\sqrt{n}\} = -0.00093/0.00040 = -2.325$ indicates that for this sample of $n = 100$ replications, the bias is not significantly different from 0 at the 1% level. In column (6) we are interested in knowing if the sample quantities differ significantly from 1; for example, for BMDPBak and $k = 1, (1.01029 - 1)/0.00432 = 2.381$, which means that the hypothesis that the ratio does not differ from 1 is not rejected: the estimated standard deviation does not differ significantly from its asymptotic value.

4.3. Numerical results

Numerical results coming from simulation experiments are summarized in Table 1, which follows the structure described in Subsection 4.2.

The table has three parts, for $k = 1, 9$ and 100, respectively. In each part the following elements remain constant: (1) The MA(1) parameters, in particular $\alpha = 0.90$, to which $\rho = 0.49724$ corresponds; (2) Sample size $T = 100$; (3) The number of repetitions, $n = 100$, and (4) The contamination proportion $p = 0.10$ which is adequate for a robustness study, and is frequently used as such in the literature.

Results are given as functions of ρ and its estimators (denoted in general

Part a. $k = 1$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,89739	FORM1	0,49631	-0,00093	-0,00187	0,00396	0,00264	1,56644	0,00303
			0,00040	0,00080	0,00028		0,00670	
0,91832	BMDP Bak	0,49749	0,00025	0,00050	0,00255	0,00183	1,01029	0,00280
			0,00026	0,00051	0,00018		0,00432	
0,88043	BMDP Cls	0,49451	-0,00273	-0,00548	0,00628	0,00338	2,48488	0,00356
			0,00063	0,00126	0,00045		0,01063	
0,90426	S-PLUS	0,49671	-0,00053	-0,00106	0,00292	0,00236	1,15699	0,00240
			0,00029	0,00059	0,00021		0,00495	
0,90410	ITSM	0,49671	-0,00053	-0,00107	0,00294	0,00236	1,16131	0,00240
			0,00029	0,00059	0,00021		0,00497	
0,91321	MINITAB	0,49727	0,00003	0,00006	0,00273	0,00201	1,07969	0,00276
			0,00027	0,00055	0,00019		0,00462	

Part b. $k = 9$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,90829	FORM1	0,49705	-0,00019	-0,00038	0,00262	0,00220	1,03480	0,00243
			0,00026	0,00053	0,00019		0,00443	
0,92470	BMDP Bak	0,49780	0,00056	0,00113	0,00228	0,00160	0,90324	0,00285
			0,00023	0,00046	0,00016		0,00387	
0,88587	BMDP Cls	0,49492	-0,00232	-0,00467	0,00540	0,00313	2,13731	0,00308
			0,00054	0,00109	0,00038		0,00915	
0,91278	S-PLUS	0,49713	-0,00011	-0,00022	0,00266	0,00203	1,05136	0,00255
			0,00027	0,00053	0,00019		0,00450	
0,91228	ITSM	0,49712	-0,00012	-0,00024	0,00267	0,00205	1,05428	0,00254
			0,00027	0,00054	0,00019		0,00451	
0,92043	MINITAB	0,49758	0,00034	0,00069	0,00256	0,00175	1,01198	0,00290
			0,00026	0,00051	0,00018		0,00433	

Part c. $k = 100$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,91111	FORM1	0,49708	-0,00016	-0,00032	0,00289	0,00209	1,14333	0,00273
			0,00029	0,00058	0,00021		0,00489	
0,93017	BMDP Bak	0,49790	0,00066	0,00133	0,00238	0,00142	0,94122	0,00304
			0,00024	0,00048	0,00017		0,00403	
0,89318	BMDP Cls	0,49394	-0,00330	-0,00665	0,01617	0,00281	6,39648	0,01287
			0,00162	0,00325	0,00115		0,02737	
0,91593	S-PLUS	0,49724	0,00000	0,00001	0,00284	0,00191	1,12333	0,00284
			0,00028	0,00057	0,00020		0,00481	
0,91562	ITSM	0,49723	-0,00001	-0,00001	0,00285	0,00192	1,12844	0,00285
			0,00029	0,00057	0,00020		0,00483	
0,92601	MINITAB	0,49772	0,00048	0,00097	0,00268	0,00156	1,05984	0,00316
			0,00027	0,00054	0,00019		0,00454	

Table 1: Simulation results with the MA(1) model with parameter $\alpha = 0.90$, $(\rho = 0.49724)$, $T = 100$, *contamination of* $p = 0.10$, *and 3 values of* k

by r). Estimation results in terms of α are also presented to facilitate understanding.

Each part has 6 rows corresponding to the estimation procedures. The contents of each part was described in Subsection 4.2.

The main observations coming from the analysis of this information, are the following:

1.*Bias in the Estimation of* ρ. Column (1) contains the estimates aver-

age of ρ, column (2) the bias in the estimation of this parameter and (3) the relative bias. In column (1) we observe that the 5 procedures, except for BMDPCls, show non-significant biases, since these biases, divided by the corresponding standard errors (printed in parenthesis) are (approximately) less than 2. BMDPCls shows a significant bias, which is in agreement with its condition of being a preliminary estimation procedure.

These remarks are valid even when errors contaminated with $N(0, 9)$ and $N(0, 100)$ are used.

It is interesting to note the proximity of the results obtained by use of the programs S-PLUS and ITSM. This is understandable, since maximum likelihood estimation procedures to be used with ARMA models are widely available in the current literature.

In conclusion, the 5 programs provide good results in terms of biases in the estimation of ρ, and these biases do not change when we consider contaminated errors (mixtures) in the MA(1) model.

2. *Estimating the Variance of* ρ*'s Estimators.* Column (4) contains the estimates of $\mathrm{Var}(r)$ for the given methods. These values can be compared with those coming from the asymptotic theory, as given in column (5). There exists a clear similarity between these two values, except in the case of BMDPCls (which was already discussed) and for FORM1 when $k = 1$.

The standard error of $s(r)$ in column (4) is computed, as indicated in Subsection 4.1, by means of the asymptotic expression $s/\sqrt{2(n-1)}$. For $k = 100$ and FORM1, this ratio is 0.00021, which is exactly the value in the table. In general, except for BMDPCls, the asymptotic approximations work very satisfactorily, with or without contamination.

To complete this section, we now compare the results obtained in Anderson et al. (1996) with those in the present paper. Table 2 summarizes these results.

We observe that the estimates coming from FORM1 are, in general, smaller than those of the other methods. This is due, in part at least, to the fact that this procedure, being an exact maximum likelihood, forces the estimates to be less than 1 in absolute value. This, in turn, comes from the fact that the likelihood function takes the same values in a given value of α and in its reciprocal. See, for example, Anderson and Mentz (1980).

Results in Table 2 show that the pairs of columns are quite similar,

MA(1), $\alpha = 0,90$ ($\rho = 0.49724$), $T = 100, n = 100, k = 1$

Figures of Merit	FORM 1		BMDP Bak	
	This Paper	Anderson	This Paper	Anderson
M(r)=m	0.49631	0.49682	0.49749	0.49750
Bi(r)	-0.00093	-0.00042	0.00025	0.00026
	(0.00040)	(0.0041)	(0.00026)	
Rbi(r)	-0.00187	-0.00084	0.00050	0.00052
	(0.00080)	(0.00082)	(0.00051)	
S(r)	0.00396	0.00406	0.00255	0.00367
	(0.00028)	(0.00029)	(0.00018)	
Rsa(r)	0.00264	0.00282	0.00183	0.00234
Mse(r)	1.56644	1.605	1.01029	1.452
	(0.00670)	(0.115)	(0.0432)	

Table 2: Comparison between simulation results in the present paper and those in Anderson et al. (1996)

except for the estimated mean square error for procedure BMDPBak. A consequence of this similarity, is that in both works, biases and relative biases in the estimation of ρ are not significantly different from 0. The differences among estimated means square errors are due to the differences among the estimates of variances, since the contributions of the squared biases are small.

4.4. Other numerical results

Table 5 is similar to Table 1, except for the following: (1) Only 4 programs were analyzed, since BMDPCls was excluded (its behavior was inferior to that of BMDPBak), and ITSM (it was similar to S-PLUS, which is easier to use computationally); (2) $\alpha = 0.90$ ($\rho = 0.49724$) were used again, sample size $T = 100$ and three values of k (1,9 and 100); (3) $n = 500$ repetitions were done and the contamination percentage was $p = 0.50$: this value is not associated with the usual ideas in robustness studies, but it was used as an extreme value to analyze its effects on the results.

A comparison of the results in Tables 1 and 5 is made in Table 7. The main observations stemming from this table are the following.

1. Estimates coming from FORM1 are in general smaller than those of the other methods. See Subsection 4.3.

$k = 1$

N°	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	1.0442	1.0081	1.0014
					(0.1711)	(0.1596)	(0.1573)
2	500	100	0.90	0.10	1.0367	0.9964	
					(0.1508)	(0.1384)	
3	500	250	0.90	0.10	1.0139	0.9952	
					(0.0926)	(0.0895)	
4	100	100	0.40	0.10	0.9895	09864	
					(0.1337)	(0.1336)	
				0.20	0.9946	0.9914	
					(0.1447)	(0.1436)	
				0.30	1.0085	1.0055	
					(0.1421)	(0.1415)	
				0.40	1.0051	0.9999	
					(0.1441)	(0.1422)	
				0.50	0.9875	0.9841	
					(0.1466)	(0.1461)	
				0.75	0.9634	0.9609	
					(0.1287)	(0.1286)	
5	100	100	0.40	0.05	1.0200	1.0163	
					(0.1486)	(0.1484)	
				0.10	1.0203	1.0161	
					(0.1484)	(0.1483)	
				0.20	0.9774	0.9743	
					(0.1342)	(0.1335)	
				0.30	1.0055	1.0015	
					(0.1308)	(0.1307)	
				0.40	1.0239	1.0180	
					(0.1482)	(0.1460)	
				0.50	0.9683	0.9648	
					(0.1467)	(0.1463)	
				0.75	1.0157	1.0107	
					(0.1582)	(0.1570)	

$k = 9$

N°	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	1.9086	2.0092	1.8302
					(0.5693)	(1.7823)	(0.5488)
2	500	100	0.90	0.10	1.8544	1.7844	
					(0.5038)	(0.4838)	
3	500	250	0.90	0.10	1.8267	1.7964	
					(0.3037)	(0.2970)	
4	100	100	0.40	0.10	1.8295	1.8249	
					(0.4508)	(0.4500)	
				0.20	2.5811	2.5711	
					(0.6883)	(0.6898)	
				0.30	3.4589	3.4450	
					(0.7483)	(0.7468)	
				0.40	4.9258	4.9052	
					(1.0339)	(1.0298)	
				0.50	5.0561	5.0444	
					(0.9219)	(0.9170)	
				0.75	6.9365	6.8488	
					(1.1903)	(1.1017)	

Continued

Table 3: Simulation results with the MA(1) model, estimation of the variance of the error term defined as a mixture: average and standard deviation.

Table 3. Continued

N°	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
5	100	100	0.40	0.05	1.4578	1.4535	
					(0.4370)	(0.4377)	
				0.10	1.8581	1.8523	
					(0.5011)	(0.5009)	
				0.20	2.5462	2.5383	
					(0.6137)	(0.6130)	
				0.30	3.4686	3.4549	
					(0.7846)	(0.7794)	
				0.40	4.3917	4.3681	
					(0.9300)	(0.9200)	
				0.50	4.8003	4.7832	
					(1.0838)	(1.0831)	
				0.75	6.9926	6.9545	
					(1.2268)	(1.2180)	

$k = 100$

N°	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	11.6665	11.2929	11.2742
					(6.2579)	(6.1470)	(6.0732)
2	500	100	0.90	0.10	11.1083	10.7577	
					(5.4224)	(5.3152)	
3	500	250	0.90	0.10	11.0699	10.9078	
					(3.2999)	(3.2530)	
4	100	100	0.40	0.05	6.4369	6.4230	
					(4.6777)	(4.6791)	
				0.10	11.4051	11.3766	
					(5.4712)	(5.4687)	
				0.20	20.4162	20.3548	
					(6.8408)	(6.8412)	
				0.30	31.4441	31.3293	
					(8.9559)	(8.8902)	
				0.40	42.6938	42.4690	
					(10.4204)	(10.3216)	
				0.50	48.4238	48.2631	
					(12.2090)	(12.2041)	
				0.75	74.9865	74.5799	
					(13.6208)	(13.5185)	

$k = 1$

N°	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	1.0442	1.0081	1.0014
2	500	100	0.90	0.10	1.0367	0.9964	
3	500	250	0.90	0.10	1.0139	0.9952	
4	100	100	0.40	0.10	0.9895	09864	
				0.20	0.9946	0.9914	
				0.30	1.0085	1.0055	
				0.40	1.0051	0.9999	
				0.50	0.9875	0.9841	
				0.75	0.9634	0.9609	
5	100	100	0.40	0.05	1.0200	1.0163	
				0.10	1.0203	1.0161	
				0.20	0.9774	0.9743	
				0.30	1.0055	1.0015	
				0.40	1.0239	1.0180	
				0.50	0.9683	0.9648	
				0.75	1.0157	1.0107	

$k = 9$

N°	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	1.9086	2.0092	1.8302
2	500	100	0.90	0.10	1.8544	1.7844	
3	500	250	0.90	0.10	1.8267	1.7964	
4	100	100	0.40	0.10	1.8295	1.8295	
				0.20	2.5811	2.5811	
				0.30	3.4589	3.4450	
				0.40	4.9258	4.9052	
				0.50	5.0561	5.0444	
				0.75	6.9365	6.8488	
5	100	100	0.40	0.05	1.4578	1.4535	
				0.10	1.8581	1.8523	
				0.20	2.5462	2.5383	
				0.30	3.4686	3.4549	
				0.40	4.3917	4.3681	
				0.50	4.8003	4.7832	
				0.75	6.9926	6.9545	

$k = 100$

N°	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	11.6665	11.2929	11.2742
2	500	100	0.90	0.10	11.1083	10.7577	
3	500	250	0.90	0.10	11.0699	10.9078	
44	100	100	0.40	0.05	6.4369	6.4230	
				0.10	11.4051	11.3766	
				0.20	20.4162	20.3548	
				0.30	31.4441	31.3293	
				0.40	42.6938	42.4690	
				0.50	48.4238	48.2631	
				0.75	74.9865	74.5799	

Table 4: Simulation results with the MA(1) model, estimation of the variance of the error term defined as a mixture: summary of average estimates.

Part a. $k = 1$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,90230	FORM1	0,49641	-0,00083	-0,00166	0,00457	0,00109	4,03951	0,00374
			0,00020	0,00041	0,00014		0,00770	
0,91847	BMDP Bak	0,49722	-0,00002	-0,00003	0,00411	0,00081	3,63660	0,00409
			0,00018	0,00037	0,00013		0,00693	
0,90767	S-PLUS	0,49648	-0,00076	-0,00153	0,00472	0,00099	4,17675	0,00396
			0,00021	0,00042	0,00015		0,00796	
0,91657	MINITAB	0,49709	-0,00015	-0,00029	0,00432	0,00084	3,81777	0,00417
			0,00019	0,00039	0,00014		0,00325	

Part b. $k = 9$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,90505	FORM1	0,49661	-0,00063	-0,00126	0,00426	0,00104	3,76911	0,00364
			0,00019	0,00038	0,00013		0,00718	
0,92158	BMDP Bak	0,49740	0,00016	0,00033	0,00387	0,00076	3,42288	0,00403
			0,00017	0,00035	0,00012		0,00652	
0,90920	S-PLUS	0,49670	-0,00054	-0,00108	0,00426	0,00097	3,77224	0,00373
			0,00019	0,00038	0,00013		0,00719	
0,91869	MINITAB	0,49725	0,00001	0,00002	0,00412	0,00081	3,64629	0,00413
			0,00018	0,00037	0,00013		0,00311	

Part c. $k = 100$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,90501	FORM1	0,49669	-0,00055	-0,00111	0,00378	0,00104	3,34737	0,00323
			0,00017	0,00034	0,00012		0,00638	
0,92423	BMDP Bak	0,49753	0,00029	0,00057	0,00335	0,00072	2,96039	0,00363
			0,00015	0,00030	0,00011		0,00564	
0,91084	S-PLUS	0,49686	-0,00038	-0,00076	0,00370	0,00094	3,27483	0,00332
			0,00017	0,00033	0,00012		0,00624	
0,92080	MINITAB	0,49739	0,00015	0,00030	0,00354	0,00078	3,13142	0,00369
			0,00016	0,00032	0,00011		0,00267	

Table 5: Simulation results with the MA(1) model with parameter $\alpha = 0.90$ ($\rho = 0.49724$) $T = 100, n = 500$, *contamination of* $p = 0.50$ *and 3values of* k

2. For $p = 0.10$, relative biases are non-significant at the 5% level, in all cases of Table 7 (which includes the two columns headed this paper in Table 1), except FORM1 when $k = 1$.

3. Estimates of $S(r)$ range from 0.00228 to 0.00292 for all procedures and values of k, except that FORM1 has the value 0.00396 for $k = 1$.

4. As a consequence of these observations, the estimated mean square error rme(r) takes values similar for the various procedures, except FORM1 as indicated.

5. Final comments, summary and conclusions

The main objective of this study is to perform an empirical analysis by means of simulations, the robustness of different proposals to estimate the parameters in the MA(1) time series model. The non-normal populations

that we consider are mixtures of normal densities, defined in general by $g(x) = pN(0,k) + (1-p)N(0,1)$, where the contamination proportion is taken to be $p = 0, 10$ or 0.50, the variance of the contaminating normal distribution $k = 1, 9$ or 100 (Andrews et al. (1972)), and the estimation procedures are Cls (conditional least squares) and Bak ("backcasting") of BMDP (BMDP (1990)), SPLUS (Venables and Ripley (1997)), ITSM (Brockwell and Davis (1991)) and MINITAB (MINITAB (1996)).

The other components of the simulations program are: the MA(1) model parameter is taken to be $\alpha = 0.90$, except in Tables 3 and 4 where $\alpha = 0.40 : 0.90$ is a value close to the region of non-invertibility ($|\alpha| < 1$ for the MA(1) to be invertible into an infinite autoregression), sample size is $T = 100$, while some partial experiments were done with $T = 250$, and the number of repetitions is $n = 100$ or 500.

The results obtained in the estimation of α or ρ are presented with the format designed in Anderson et al. (1996), which is a simulation study using only FORM1 and BMDPBak.

The empirical results (Table 1) show that the programs (except for BMDPCls, as expected) give satisfactory results in terms of biases in the estimation of ρ, and that these biases do not change when contaminated errors (mixtures) are used in the MA(1) model. These results are compared, partially at least, with those in Anderson et al. (1996), having found a great deal of similarity among them, also as expected (Table 2); this similarity exists even when the sources of generation of the pseudorandom numbers are changed.

Since the error variance changes when we consider different mixtures, and is given by $\mathrm{Var}(u_t) = pk + (1-p)$, Tables 3 and 4 show the results of some experiments done with BMDPBak. This procedure is selected because it provides complete results, and it improves the estimation by BMDPCls. The following values are taken: $\alpha = 0.40, T = 100$ or $250, n = 100$ or 500 and $p = 0.05, 0.10, 0.20, 0.30, 0.40, 0.50$ and 0.75. A summary of these results, compared with the corresponding theoretical or expected values, constitutes Table 6.

The good agreement of the empirical results with the underlying theory is observed.

Finally, in the experiment done by changing p from 0.10 to 0.50 (and n from 100 to 500), good stability of the results is observed.

Contamination (p)	$k=1$ $\text{Var}(u_t)$	$k=1$ Estimation BMDPBak	$k=9$ $\text{Var}(u_t)$	$k=9$ Estimation BMDPBak	$k=100$ $\text{Var}(u_t)$	$k=100$ Estimation BMDPBak
0.05	1.00	1.0163	1.40	1.4535	5.90	6.4230
0.10	1.00	1.0161	1.80	1.8523	10.90	11.3766
0.20	1.00	0.9743	2.60	2.5383	20.80	20.3548
0.30	1.00	1.0015	3.40	3.4549	30.70	31.3293
0.40	1.00	1.0180	4.20	4.3681	40.60	42.4690
0.50	1.00	0.9648	5.00	4.7832	50.50	48.2631
0.75	1.00	1.0107	7.00	6.9545	75.25	74.5799

Table 6: Comparison between simulation with the MA(1) model, in estimating the variance of the error term defined by a mixture: comparison between empirical averages and theoretical Values

References

AITKIN, M. and TUNNICLIFFE, W. G. (1980). Mixtures models, outliers and the EM algorithm. *Technometrics*, 22:325–331.

ANDERSON, T. W. and MENTZ, R. P. (1980). On the structure of the likelihood function of autoregressive and moving average models. *Journal of Time Series Analysis*, 1:83–94.

ANDERSON, T. W. and MENTZ, R. P. (1993a). Evaluation of quadratic forms and traces for iterative estimation in first-order moving average models. *Communications in Statistics: Theory and Methods*, 22:931–963.

ANDERSON, T. W. and MENTZ, R. P. (1993b). *Iterative Procedures for Exact Maximun Likelihood Estimation in the First-order Gaussian Moving Average Model.* A Raghu Raj Bahadur Fetchrift. Wiley Eastern Limited.

ANDERSON, T. W., MENTZ, R. P., JARMA, N. M., and MARTÍNEZ, C. I. (1996). Simulations of iterative procedures for maximum likelihood estimation in MA(1) models. *Communications in Statistics: Simulation and Computation*, 25:851–865.

ANDERSON, T. W. and TAKEMURA, A. (1986). Why do noninverrible estimated moving averages occur? *Journal of Time Series Analysis*, 7:235–254.

Part a. $k = 1$

Item	Form1		BMDBPBak		S-PLUS		MINITAB	
	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$
m(r) = m	0,49631	0,49641	0,49749	0,49722	0,49671	0,49648	0,49727	0,49709
bi(r)	-0,00093	-0,00083	0,00025	-0,00002	-0,00053	-0,00076	0,00003	-0,00015
	-0,0004	-0,0002	-0,00026	-0,00018	-0,00029	-0,00021	-0,00027	-0,00019
rbi(r)	-0,00187	-0,00166	0,00050	-0,00003	-0,00106	-0,00153	0,00006	-0,00029
	-0,0008	-0,00041	-0,00051	-0,00037	-0,00059	-0,00042	-0,00055	-0,00039
s(r)	0,00396	0,00457	0,00255	0,00411	0,00292	0,00472	0,00273	0,00432
	-0,00028	-0,00014	-0,00018	-0,00013	-0,00021	-0,00015	-0,00019	-0,00014
asm(r)	0,00264	0,00109	0,00183	0,00081	0,00236	0,00099	0,00201	0,00084
rsa(r)	1,56644	4,03951	1,01029	3,63660	1,15699	4,17675	1,07969	3,81777
	-0,0067	-0,0077	-0,00432	-0,00693	-0,00495	-0,00796	-0,00462	-0,00325
mse(r)	0,00303	0,00374	0,00280	0,00409	0,00240	0,00396	0,00276	0,00417

Part b. $k = 9$

Item	Form1		BMDBPBak		S-PLUS		MINITAB	
	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$
m(r) = m	0,49705	0,49661	0,49780	0,49740	0,49713	0,49670	0,49758	0,49725
bi(r)	-0,00019	-0,00063	0,00056	0,00016	-0,00011	-0,00054	0,00034	0,00001
	-0,00026	-0,00019	-0,00023	-0,00017	-0,00027	-0,00019	-0,00026	-0,00018
rbi(r)	-0,00038	-0,00126	0,00113	0,00033	-0,00022	-0,00108	0,00069	0,00002
	-0,00053	-0,00038	-0,00046	-0,00035	-0,00053	-0,00038	-0,00051	-0,00037
s(r)	0,00262	0,00426	0,00228	0,00387	0,00266	0,00426	0,00256	0,00412
	-0,00019	-0,00013	-0,00016	-0,00012	-0,00019	-0,00013	-0,00018	-0,00013
asm(r)	0,00220	0,00104	0,00160	0,00076	0,00203	0,00097	0,00175	0,00081
rsa(r)	1,03480	3,76911	0,90324	3,42288	1,05136	3,77224	1,01198	3,64629
	-0,00443	-0,00718	-0,00287	-0,00652	-0,0045	-0,00719	-0,00433	-0,00311
mse(r)	0,00243	0,00364	0,00285	0,00403	0,00255	0,00373	0,00290	0,00413

Part c. $k = 100$

Item	Form1		BMDBPBak		S-PLUS		MINITAB	
	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$	$p = 0.10$ $n = 100$	$p = 0.50$ $n = 500$
m(r) = m	0,49708	0,49669	0,49790	0,49753	0,49724	0,49686	0,49772	0,49739
bi(r)	-0,00016	-0,00055	0,00066	0,00029	0,00000	-0,00038	0,00048	0,00015
	-0,00029	-0,00017	-0,00024	-0,00015	-0,00028	-0,00017	-0,00027	-0,00016
rbi(r)	-0,00032	-0,00111	0,00113	0,00057	0,00001	-0,00076	0,00097	0,00030
	-0,00058	-0,00034	-0,00048	-0,0003	-0,00057	-0,00033	-0,00054	-0,00032
s(r)	0,00289	0,00378	0,00238	0,00335	0,00284	0,00370	0,00268	0,00354
	-0,00021	-0,00012	-0,00017	-0,00011	-0,0002	-0,00012	-0,00019	-0,00011
asm(r)	0,00209	0,00104	0,00142	0,00072	0,00191	0,00094	0,00156	0,00078
rsa(r)	1,14333	3,34737	0,94122	2,96039	1,12333	3,27483	1,05984	3,13142
	-0,00489	-0,00638	-0,00403	-0,00564	-0,00481	-0,00624	-0,00454	-0,00267
mse(r)	0,00273	0,00323	0,00304	0,00363	0,00284	0,00332	0,00316	0,00369

Table 7: Comparison of simulation results with the MA(1) model with contaminations of p = 0.10 and 0.50, n = 100 and 500. Parameter α = 0.90 (ρ = 0.49724), T = 100, and 3 values of k

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H., and TUKEY, J. W. (1972). *Robust estimates of location: survey and advances*. Princeton University Press, Princeton.

BMDP (1990). *BMDP Statistical Software*. Los Angeles, California.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis Forecasting and Control*. Holden-Day, San Francisco.

BOX, G. E. P. and TIAO, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70:70–79.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *ITSM: An Interactive Time Series Modelling Package for the PC*. Springer-Verlag, New York.

CHANG, I., TIAO, G. C., and CHAN, C. (1988). Estimation of Time Series parameters in the presence of outliers. *Technometrics*, 30:193–204.

COOK, R. D. and WEISBERG, S. (1982). *Residuals and influence in regression*. Chapman and Hall, London.

FOX, A. J. (1972). Outliers in Time Series. *Journal of the Royal Statistical Society*, B 43:350–363.

LINDSAY, B. G. (1995). Mixture models: theory, geometry and applications. Regional conference series in probability and statistics, Institute of Mathematical Statistics.

MARTIN, R. D. (1980). Robust estimation of autoregressive models. Directions in time series, Institute of Mathematical Statistics, Hayward, California.

MINITAB (1996). *Guía del Usuario de MINITAB en Español, Versión 2 para Windows*. Minitab Inc., Los Angeles, California, US.

PEÑA, D. (1990). Influential observations in Time Series. *Journal of Business and Economics Statistics*, 8:235–241.

TIKKU, M. L., WONG, W. K., VAUGHAN, D. C., and BIAN, G. (2000). Time Series models in non-normal situations: symmetric innovations. *Journal of Time Series Analysis*, 21:571–596.

VENABLES, W. N. and RIPLEY, B. D. (1997). *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York, 2nd ed.

WOLFRAM, S. (1991). *Mathematica*. Addison-Wesley Publishing Co., Reading, Massachusetts, US, 2nd ed.

On confidence bands for time series problems in the time and frequency domains

Raúl P. Mentz, Aldo J. Viollaz and Carlos I. Martinez

Universidad Nacional de Tucumán and CONICET

Abstract: The construction of (asymptotic) simultaneous confidence bands for some time series problems is studied, typically for the sample autocorrelogram and windowed spectral density estimate. The following approaches are explored: (1) To use the close-form results available in the literature; (2) To use the asymptotic independence of the sample quantities to derive new procedures; (3) To resort to inequalities. As expected, the bands turn out being wider than those frequently encountered in the literature, based on point-by-point confidence intervals. Numerical values of the necessary constants are given for selected values of the joint confidence coefficient and various numbers of sample quantities. The use of confidence sets of non-uniform width is also briefly explored. Monte Carlo simulations are presented, for problems in the time and in the frequency domains.

Key words: Asymptotic independence; autocorrelogram; Bonferroni inequality; simultaneous confidence bands; windowed spectral density estimate.

1. Introduction

Given an observed time series $y_1, \ldots, y_T$, two useful data-analytic techniques are to compute, plot and interpret the sample autocorrelogram in the time domain, or a (windowed) spectral density estimate in the frequency domain. These are just two instances among several sample quantities often considered: in the time domain, the sample partial, inverse, and partial inverse autocorrelograms for univariate series, and the cross correlogram for bivariate series; in the frequency domain, the sample coherence and phase for bivariate time series.

From an empirical point of view, and in agreement with what the theoretical sampling properties indicate, it is often the case that the sample autocorrelogram and sample spectral density tend to exhibit fluctuations that must be accurately interpreted. The question then arises as how to set "control lines" or "confidence bands" on these sample functions.

Setting control lines in the sample autocorrelogram has been defined to mean that (asymptotic) point-by-point confidence intervals are plotted for the whole set of sample estimates. In a simple case, straight control lines at $\pm c\hat{\sigma}$ have been suggested by Box and Jenkins (1976, page 185), and used by many authors and practitioners. Here c is a constant (often taken to be 1 or 2) and $\hat{\sigma}^2$ is an estimate of the residual variance of an underlying linear model.

Similarly, control lines for spectral density estimates have been often recommended and used in the literature: see, for example, Granger and Hatanaka (1984, page 66). The technique used here is also to derive an (asymptotic) confidence interval for one ordinate and to use it for the whole sample function.

The approach of control lines has the advantage of its simplicity, but has the obvious shortcoming of using a point-by-point inference tool to make inferences about the whole set of values under consideration. This means that control of the confidence coefficient is lost. To avoid this difficulty a joint or simultaneous confidence approach is need, and that will be discussed below.

In this paper and in terms of frequency analysis, we consider estimation of the spectral density function by means of estimators related to the sample spectral density or *periodogram*. We consider smoothing the periodogram by using adequate "windows" (that is, systems of weights) defined in the time or in the frequency domain.

An alternative approach to estimate the spectral density of a stationary time series, is to use an *autoregressive estimator*, defined as follows: the given series is approximated by a finite-order autoregression, denoted by AR($\hat{p}$), whose order is suitably estimated. Then all parameters are estimated, and the resulting values are "plugged-in" the formula of the spectral density of the AR($\hat{p}$) model. The resulting estimator possesses some good properties, and is, in general, quite smooth, in comparison with the sample periodogram.

This approach is closely related to what is called the *maximum entropy spectral estimator*. On these topics, see, among others, Akaike (1969), Parzen (1974), Priestley (1981), Beamish and Priestley (1981), Newton and Pagano (1984), Koslov and Jones (1985), Sakai and Sakaguchi (1990), Hrafnkelsson and Newton (2000).

In Section 2 we review point-by-point confidence intervals; in Section 3 we consider simultaneous confidence sets; in Section 4 we present some simple or elementary approaches to solve or approximate the simultaneous inference problem. Sections 5 and 6 contain simulation results, and Section 7 is of discussion and conclusions.

2. Basic definitions and point-by-point confidence bands

According to one standard definition, the sample autocorrelogram is the set of sample quantities

$$r_s = \frac{\sum_{t=1}^{T-s}(y_t - \overline{y})(y_{t+s} - \overline{y})}{\sum_{t=1}^{T}(y_t - \overline{y})^2}, \quad s = 1, 2, \ldots, m, \tag{2.1}$$

where m is usually taken to be considerably smaller than T. Other definitions of the sample autocorrelations are considered in Mentz (1983a) in connection with their roles in the sample autocorrelogram.

The sampling properties of (2.1) are usually studied in terms of asymptotic results, when $T \to \infty$ while m remains fixed. If the underlying stochastic process is stationary and linear ($y_t = \mu + \sum_{j=-\infty}^{\infty} \gamma_j u_{t-j}$, where the innovations u_t are independent, identically distributed, with 0 expected value and finite fourth-order moment, and $\sum|\gamma_j| < \infty$) it can be shown that $r_1, r_2, \ldots, r_j$ is asymptotically normal around the true parameter $\rho_1, \rho_2, \ldots, \rho_m$, with variances and covariances given by

$$\tau_{gh} = \frac{1}{T}\sum_{r=-\infty}^{\infty}(\rho_{r+g}\rho_{r+h} + \rho_{r-g}\rho_{r+h} - 2\rho_h\rho_r\rho_{r+g} - 2\rho_g\rho_r\rho_{r+h} + 2\rho_g\rho_h\rho_r^2), \quad g, h = 1, 2, \ldots, m. \tag{2.2}$$

See, for example, Anderson (1971). Whenever (2.2) can be evaluated numerically (i.e., whenever the ρ_j can be expressed as functions of some set of parameters) we have a way to solve problems of inference, in the sense of approximating the needed distributions by their asymptotic normal limits. In particular, a confidence interval for ρ_s is then

$$r_s \pm c\hat{\tau}_{ss}^{1/2}, \quad s = 1, 2, \ldots, m, \tag{2.3}$$

where c is chosen from standard normal tables to give the desired confidence coefficient.

One standard definition of the (windowed) spectral density estimator is

$$\hat{f}(v_j) = \frac{1}{2\pi}\sum_{s=-m}^{m} k\left(\frac{s}{m}\right)\cos(v_j s)c_s, \quad j = 0, 1, \ldots, l, \tag{2.4}$$

where $c_s = T^{-1}\sum_{t=1}^{T-s}(y_t - \overline{y})(y_{t+s} - \overline{y}) = c_{-s}$, and k is a "kernel function": it is normalized, bounded, symmetric about 0, and sufficiently smooth (see, for example, Anderson, 1971, Section 9.4). Note that in (2.1) $r_s = c_s/c_0$.

The sampling properties of (2.4) are usually studied in terms of asymptotic results, when $T \to \infty$ and $m = m_T \to \infty$ in such a way that $m_T^d/T \to 0$ for some suitable d. If the underlying stochastic process is stationary and linear (see above), it can be shown that $\hat{f}(v_0), \hat{f}(v_1), \ldots, \hat{f}(v_l)$ is asymptotically normal around the spectral ordinates $f(v_0), f(v_1), \ldots, f(v_l)$ with variances

$$\tau^2(v_j) = \frac{m}{T}2f^2(0)\int_{-1}^{1}k^2(x)dx, \quad v_j = 0, \tag{2.5}$$

$$= \frac{m}{T}2f^2(\pi)\int_{-1}^{1}k^2(x)dx, \quad v_j = \pm\pi, \tag{2.6}$$

$$= \frac{m}{T}f^2(v_j)\int_{-1}^{1}k^2(x)dx, \quad v_j \neq 0, \pm\pi, \tag{2.7}$$

and covariances equal to 0.

Since these integrals are known for the standard windows proposed in the literature, we have a way to construct confidence intervals at each frequency. One such form is

$$\frac{\hat{f}(v_j)}{1+\tau c\sqrt{m/T}} \leq f(v_j) \leq \frac{\hat{f}(v_j)}{1-\tau c\sqrt{m/T}}, \quad v_j \neq 0, \pm\pi, \tag{2.8}$$

where $\tau^2 = \int_{-1}^{1} k^2(x)dx$, and c is chosen from standard normal tables. It can also be shown that to the transformation $\ln f(v_j)$ confidence intervals can be set by

$$\ln \hat{f}(v_j) - \tau c\sqrt{\frac{m}{T}} \leq \ln f(v_j) \leq \ln \hat{f}(v_j) + \tau c\sqrt{\frac{m}{T}}. \tag{2.9}$$

Two important differences between the two cases that we considered in this section are the following: (a) Asymptotic theories are developed in such a way that the final results are comparable, but while in (2.1) m is treated as fixed when $T \to \infty$, in (2.4) $m = m_T \to \infty$ with T; (b) The asymptotic covariances for (2.4) are 0, while (2.2) is not necessarily equal to 0 when $g \neq h$.

In fact, it is well known that sample spectral density or periodogram obtained from (2.4) by setting $k \cong 1$ and $m = T - 1$ is such that for independent $y_1, \ldots, y_T$ normal the periodogram ordinates at different frequencies are independent for a finite sample size T.

3. Simultaneous inference

The comments in the last part of the preceding section about the asymptotic uncorrelatedness of spectral density estimators at different frequencies contribute to explain why detailed (asymptotic) results for simultaneous inference are available only for the frequency domain. An early contribution is by Walker (1967).

Woodroofe and Van Ness (1967) proved that under certain conditions on the underlying process and the kernel function, the asymptotic distribution of the (normalized) maximum of the windowed spectral density estimator's ordinates can be found in an explicit form useful for statistical inference. Their main result can be written as

$$\lim_{T\to\infty} P\left[\max_{0\leq s\leq m} \frac{\left|\hat{f}\left(\frac{\pi s}{m}\right) - f\left(\frac{\pi s}{m}\right)\right|}{f\left(\frac{\pi s}{m}\right)} \leq ax + b\right] = \exp\left(-e^{-x}\right), \tag{3.1}$$

where $a = [2\ln(2m)]^{-1/2}$, $b = a^{-1} - a[\ln\ln(2m) + \ln 2\pi]/2$.

Some remarks about this result follow.

(a) *Assumptions of the theorem.* It is assumed that the underlying process is linear and the innovations have finite eighth-order moment. Some other regularity conditions are set on the coefficients of the linear representation, on the spectral density, and on the kernel function.

(b) *Nature of the result.* Expression (3.1) shows that a key consequence of this approach is that the limiting extreme-value distribution function $\exp[-\exp(-x)]$ is used to determine the constant in the confidence set, instead of the standard normal distribution function of our Section 2. Tables of significance points of this extreme-value distribution are given by Owen (1962).

(c) *Asymptotic order of the result.* Woodroofe and Van Ness (page 1558) indicate that "the difference between the maximum deviation and the deviation at a single frequency point ... manifests itself in the factor $(\ln m)^{-1/2}$. Thus in practice a confidence band for all frequencies is $O(\ln m)^{1/2}$ times that for a finite set". This observation has also been recorded by Priestley (1981, page 486).

It should be noted that the comment refers to the "asymptotic order" or the result, and should not be interpreted to mean that, for example, in (2.8) c should be replaced by $c(\ln m)^{1/2}$. In effect, even when the order of magnitude is correct in an asymptotic sense, the constant for a finite set of observed values may have to be altered. See Appendix 1.

(d) *Usefulness of the result.* Referring to this result, Hannan (1970), page 294, wrote: "These results are important in principle but surrounded by some doubt in practice, of a greater magnitude than that accorded to results of earlier sections, because of their asymptotic nature. It is known that such extreme value formulas are relevant only in enormously large samples, when the largest of a series of independent and identically distributed random variables is under consideration. Here further approximations are involved and for the relevance of the formulas it is evidently m as much as T whose magnitude is of importance. One conjectures that the formulas are the roughest of approximations only."

(e) *Other problems in the frequency domain.* Hannan (1970), page 294, notes that the approach of Woodroofe and Van Ness can be used to derive simultaneous probability statements analogous to (3.1) for sample coherence and phase in multiple time series analysis.

All these results and remarks correspond to estimation in the frequency domain. It is important to discuss what can be said about a close-form asymptotic result similar to (3.1) for the sample autocorrelogram and other quantities in the time domain.

One can conjecture that the extreme-value distribution function $\exp(-e^{-x})$ used in (3.1) should be also relevant for the sample autocorrelogram, under suitable conditions on the underlying process, and after suitable normalization. From general results in Cramer and Leadbeter (1962), for example, it follows that the normalized maximum of a stationary Gaussian stochastic process follows the distribution attained in (3.1). However, the sample autocorrelations can be regarded only as an *asymptotically* Gaussian stochastic process; this will then require an explicit treatment, that we have been unable to trace in the literature.

4. Other approaches to simultaneous inference

4.1. Introduction

The solution studied in Section 3 has two main shortcomings or difficulties: (a) Its derivation is quite complicated, and it will not be easy to extend the approach to other cases of interest, in particular, to those in the time domain; (b) The sample sizes implied by the result may be too high for some fields of application. These and other reasons indicate that it pays to study further the problem, with simpler tools, and to derive exact or approximate results that may be useful.

In general, we want to study what can be said about joint probabilities that in the case of the sample autocorrelogram have the form

$$P(\sqrt{T}|r_1| \le c, \ldots, \sqrt{T}|r_m| \le c|H) = P\left(\max_{1\le s\le m} \sqrt{T}|r_s| \le c|H\right), \qquad (4.1)$$

where H is some suitable hypothesis on the underlying process.

4.2. Using the asymptotic independence

Mentz (1983b) evaluated (4.1) when H is that the underlying process is white noise. Then if the probability in (4.1) is set equal to γ, c is defined by

$$\Phi(c) - \Phi(-c) = \sqrt[m]{\gamma}, \qquad (4.2)$$

where Φ is the standard normal cumulative distribution function.

This approach was applied to the joint inference based on $r_1, \ldots, r_m$ to test the null hypothesis that the process is white noise. It uses the fact that the set of sample autocorrelations is asymptotically normal, as indicated in Section 2. However, in view of (2.2) the independence of the r_j has to be assumed.

The approach can be extended to the frequency domain as follows: the set

$$\sqrt{\frac{T}{m}} \frac{\left[\hat{f}(v_j) - f(v_j)\right]}{\tau f(v_j)}, \quad v_j \ne 0, \pm\pi,$$

is asymptotically unit normal, that is, multivariate normal with means 0, variance 1, and covariances 0. Hence, the constants defined by (4.2) are also those needed for simultaneous confidence intervals for $f(v_j)$. These will then be of the forms given in (2.8) or (2.9).

Note that the main difference between the results in the time and frequency domains is that in the former the asymptotic independence of the sample quantities has to be assumed, while in the latter is given as part of the asymptotic distribution.

4.3. Some useful inequalities

In Section 1 we discussed the fact that the accurate interpretation of the fluctuations exhibited by a set of sample quantities, requires the evaluation of the probabilities of some events in the joint distribution of the sample quantities. This is

often difficult due to two main reasons: (1) In general, the joint distributions are not known for finite sample sizes; (2) Even that the asymptotic joint distributions often turn out being multivariate normal, the evaluation of probabilities for the events of interest is complicated.

In this section we present three inequalities useful to construct confidence bands for a set of parameters. The inequalities will be presented in general form, for a random vector $X = (X_1, \ldots, X_m)'$; then in Section 4.4 they will be used to derive confidence bands for the problems discussed so far.

Theorem 1 *Let $X = (X_1, \ldots, X_m)'$ be distributed as multivariate normal $N(0, \Sigma)$. Then for all $a_i > 0$, $i = 1, \ldots, m$,*

$$P(|X_1| \le a_1, \ldots, |X_m| \le a_m) \ge \prod_{i=1}^{m} P(|X_i| \le a_i). \qquad (4.3)$$

The inequality is strict if Σ is positive definite and at least one pair (X_i, X_j) has non-null correlation.

This theorem was proved by Dunn (1958) for $m \le 3$, and by Katri (1967) and Sidák (1967) for arbitrary m. The hypothesis of normality can be relaxed as follows.

Corollary 1 *Let $\mathbf{X} = (X_1, \ldots, X_m)'$ be a random vector. Assume that there exist Borel measurable and monotone functions $g_i : \Re \to \Re$, $i = 1, \ldots, m$, which are symmetric in the sense that $g_i(x) = -g_i(-x)$ for all real x, and such that $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ the vector has a normal distribution, where $Y_i = g_i(x_i)$, $i = 1, \ldots, m$. Then the conclusion of Theorem 1 holds.*

Theorem 2 *Let $\mathbf{X} = (X_1, \ldots, X_m)'$ be distributed as $N(\mathbf{0}, \sigma^2\mathbf{B})$, where $\mathbf{B}$ is a positive semidefinite matrix, and let S be independent of $\mathbf{X}$ such that nS^2/σ^2 has a $\chi^2(n)$ distribution. Let $T_i = X_i/S$, $i = 1, \ldots, m$. Then*

$$P(|T_1| \le a_1, \ldots, |T_m| \le a_m) \ge \prod_{i=1}^{m} P(|T_i| \le a_i), \qquad (4.4)$$

for all $a_i > 0$, $i = 1, \ldots, m$.

The proof of this theorem can be found in Sidák (1967) and Tong (1980), page 37.

The inequalities for probabilities of rectangles of the type (4.3) are basic for constructing confidence bands for a set of parameters. In this sense, Theorem 1 and Corollary 1 are useful. They are limited, however, because they require the assumption of a normal (or closely related) joint distribution.

Finally, we present Bonferroni's inequality which does not require any assumption about the distribution of the random variables involved. In spite of its generality, Bonferroni's inequality if quite sharp and hence very useful.

Theorem 3 *Let $\mathbf{X} = (X_1, \ldots, X_m)'$ be a random vector and let $A_1, \ldots, A_m$ be Borel measurable sets of the real line. Then*

$$1 - \frac{Q_1^2}{Q_1 + 2Q_2} \geq P\left(\bigcap_{i=1}^{m}(X_1 \in A_i)\right) \geq 1 - Q_1, \tag{4.5}$$

where

$$Q_1 = \sum_{i=1}^{m} P(X_i \notin A_i), \tag{4.6}$$

$$Q_2 = \sum_{i=1}^{m}\sum_{j=1}^{i-1} P(X_i \notin A_i,\ X_j \notin A_j). \tag{4.7}$$

For a proof of this theorem see Chung and Erdos (1952).

4.4. Evaluation and comparison of simultaneous confidence bands

4.4.1. Introduction

We now evaluate the constants needed to set confidence bands in the sample autocorrelogram and a windowed spectral density estimate. We shall continue with the general notation introduced in Section 4.3. In fact, it should be clear that the constants to be evaluated to use in (2.3) and (2.8), correspond to the following set up.

Let $\mathbf{X} = (X_1, \ldots, X_m)'$ be a random vector with which we want to construct a confidence band for the set $\mu_1, \ldots, \mu_m$ of expected values. We assume that the corresponding variances $\sigma_1^2, \ldots, \sigma_m^2$ are known, and that the confidence band is of the form

$$P\left(\bigcap_{i=1}^{m}|X_i - \mu_i| \leq c\sigma_i\right) \geq \gamma. \tag{4.8}$$

Without loss of generality we assume that $\sigma_i = 1$, $i = 1, \ldots, m$.

When dealing with the autocorrelogram, we have that in general ρ_i is not the expected value of r_i, and that the variance of r_i is not known since it depends on the unknown ρ_i, see (2.3). In this case the constants to be determined below still apply in an asymptotic sense, and the variances will have to be estimated on the basis of the sample observations (as indicated in (2.3); this will be another source of approximation.

In the case of the spectral density, forms (2.8) and (2.9) are such that the component $\tau\sqrt{m/T}$ is a constant for a given window.

Using the inequalities introduced in Section 4.3 we now evaluate the constants to be used in the bands.

From (4.3) we have that

$$P\left(\bigcap_{i=1}^{m}|X_i - \mu_i| \leq c\right) \geq \prod_{i=1}^{m} P(|X_i - \mu_i| \leq c) = [\Phi(c) - \Phi(-c)]^m = \gamma, \tag{4.9}$$

where we assumed that the joint distribution is normal. Solving (4.9) for c, that we may denote as $c_m(\gamma)$ to emphasize its dependence on m and γ, we have that

$$c_m(\gamma) = \Phi^{-1}\left(\frac{1 + \sqrt[m]{\gamma}}{2}\right). \tag{4.10}$$

This set of values was presented in Section 4.2 for the case of asymptotic independence.

Using Bonferroni's inequality we have:

$$\begin{aligned} P\left(\bigcap_{i=1}^{m}|X_i - \mu_i| \leq b\right) &\geq 1 - \sum_{i=1}^{m} P(|X_i - \mu_i| > b) \\ &= -mP(|X_1 - \mu_1| > b) \\ &= 1 - 2m[1 - \Phi(b)] = \gamma, \end{aligned} \tag{4.11}$$

and from this we deduce the constant

$$b_m(\gamma) = \Phi^{-1}\left(1 - \frac{1-\gamma}{2m}\right). \tag{4.12}$$

In the case of the frequency domain we can also use the asymptotic distribution of the maximum. This can be written as

$$\lim_{T\to\infty} P\left(\max_{1\leq i\leq m}|X_i - \mu_i| \leq z\right) = \exp\left(-e^{-z}\right) = \gamma, \tag{4.13}$$

where $z = ax + b$, and a and b are given below (3.1). Explicitly, the constants to be used here are

$$z_m(\gamma) = b - a\ln(-\ln\gamma). \tag{4.14}$$

Note that even when γ is the confidence coefficient of the joint procedure, the evaluation in the marginal normal distribution of (4.13) is done as if the level were $1 - (1-\gamma)/2m$; a similar argument holds for (4.11). This means that the normal density is used far apart in the tails, and hence that the use of the resulting values should be done with care.

4.4.2. Numerical results

We now evaluate and compare numerically $c_m(\gamma)$, $b_m(\gamma)$ and $z_m(\gamma)$ introduced in Section 4.4.1. We recall that $z_m(\gamma)$ is only justified in the case of the spectral density function estimator.

Table 1 presents values of the indicated expressions for selected values of γ and m. To facilitate the writing γ is often omitted.

Detailed tables are appended to the present work; see Tables 3 and 4.

Table 1 *Constants to be used in (asymptotic) joint confidence bands derived under different assumptions*

m	Confidence level γ								
	0,90			0,95			0,99		
	c_m	b_m	z_m	c_m	b_m	z_m	c_m	b_m	z_m
1	1.645	1.645	-	1.960	1.960	-	2.576	2.576	-
2	1.949	1.960	2.367	2.236	2.241	2.799	2.806	2.807	3.778
5	2.311	2.326	2.572	2.569	2.576	2.908	3.089	3.090	3.667
10	2.560	2.576	2.768	2.800	2.807	3.062	3.289	3.291	3.728
20	2.791	2.807	2.966	3.016	3.023	3.231	3.480	3.481	3.831
50	3.075	3.090	3.222	3.283	3.291	3.459	3.718	3.719	3.996
100	3.276	3.291	3.408	3.474	3.481	3.629	3.889	3.891	4.130
200	3.467	3.481	3.588	3.656	3.662	3.796	4.055	4.056	4.266
500	3.706	3.719	3.815	3.885	3.891	4.009	4.265	4.265	4.447
1000	3.878	3.891	3.980	4.050	4.056	4.165	4.417	4.417	4.583
2000	4.043	4.056	4.140	4.209	4.214	4.317	4.565	4.565	4.717
5000	4.253	4.265	4.343	4.412	4.417	4.511	4.751	4.751	4.891

Table 2 *Relative differences of the constants to be used in (asymptotic) joint confidence bands derived under different assumptions*

m	Confidence level γ					
	0.90		0.95		0,99	
	$\frac{c_m - z_m}{z_m} 100$	$\frac{b_m - z_m}{z_m} 100$	$\frac{c_m - z_m}{z_m} 100$	$\frac{b_m - z_m}{z_m} 100$	$\frac{c_m - z_m}{z_m} 100$	$\frac{b_m - z_m}{z_m} 100$
2	-17.7	-17.2	-20.1	-19.9	-25.7	-25.7
5	-10.2	-9.6	-11.7	-11.4	-15.8	-15.7
10	-7.5	-6.9	-8.6	-8.3	-11.8	-11.7
20	-5.9	-5.4	-6.7	-6.4	-9.2	-9.1
50	-4.6	-4.1	-5.1	-4.9	-7.0	-6.9
100	-3.9	-3.5	-4.3	-4.1	-5.8	-5.8
200	-3.4	-3.0	-3.7	-3.5	-5.0	-4.9
500	-2.9	-2.5	-3.1	-3.0	-4.1	-4.1
1000	-2.6	-2.3	-2.8	-2.6	-3.6	-3.6
2000	-2.4	-2.0	-2.6	-2.4	-3.3	-3.2
5000	-2.3	-1.8	-2.4	-2.1	-3.1	-2.9

About b_m and c_m note that we consider the variance as known (cf. Section 4.4.1); this is reasonable in our case since we have in mind large sample sizes for practical applications. For b_m, when the variance is unknown and replaced by an estimate, Miller (1966) presents a table of percentage points of the corresponding (Student) distribution; our constant correspond to Miller's table for a number of degrees of freedom tending to ∞. However, even for small numbers degrees of freedom the approximation is very good.

Table 3 *Constants $c_m(\gamma)$ to be used in (asymptotic) joint confidence band*

m \ γ	0.6	0.7	0.8	0.9	0.95	0.99	0.995	0.999
1	0.84162	1.03643	1.28155	1.64485	1.95996	2.57583	2.80703	3.29053
2	1.21228	1.39393	1.61842	1.94882	2.23648	2.80623	3.02296	3.48069
3	1.41671	1.58884	1.80113	2.11405	2.38774	2.93416	3.14349	3.58783
4	1.55524	1.72068	1.92477	2.22627	2.49092	3.02220	3.22668	3.66216
5	1.65898	1.81940	2.01746	2.31066	2.56876	3.08904	3.28996	3.71892
6	1.74140	1.89787	2.09123	2.37800	2.63104	3.14276	3.34090	3.76472
7	1.80948	1.96273	2.15229	2.43386	2.68280	3.18757	3.38345	3.80306
8	1.86732	2.01787	2.20424	2.48148	2.72701	3.22596	3.41993	3.83600
9	1.91747	2.06571	2.24937	2.52292	2.76553	3.25950	3.45183	3.86484
10	1.96168	2.10791	2.28921	2.55955	2.79963	3.28926	3.48015	3.89048
20	2.23805	2.37234	2.53969	2.79102	3.01599	3.47948	3.66165	4.05552
30	2.38924	2.51751	2.67781	2.91951	3.13675	3.58665	3.76422	4.14930
40	2.49237	2.61675	2.77248	3.00791	3.22009	3.66101	3.83551	4.21469
50	2.57018	2.69174	2.84416	3.07500	3.28348	3.71777	3.89000	4.26478
60	2.63243	2.75180	2.90164	3.12893	3.33450	3.76359	3.93401	4.30531
70	2.68417	2.80176	2.94952	3.17391	3.37712	3.80194	3.97089	4.33931
80	2.72836	2.84447	2.99048	3.21244	3.41366	3.83489	4.00258	4.36857
90	2.76687	2.88171	3.02622	3.24610	3.44561	3.86374	4.03035	4.39423
100	2.80095	2.91469	3.05790	3.27596	3.47398	3.88939	4.05505	4.41707
200	3.01724	3.12438	3.25975	3.46682	3.65575	4.05446	4.21424	4.56468
300	3.13796	3.24170	3.37300	3.57432	3.75846	4.14826	4.30487	4.64903
400	3.22127	3.32277	3.45139	3.64890	3.82984	4.21367	4.36813	4.70803
500	3.28464	3.38449	3.51114	3.70583	3.88440	4.26377	4.41663	4.75332
600	3.33565	3.43421	3.55931	3.75178	3.92848	4.30431	4.45590	4.79004
700	3.37825	3.47576	3.59958	3.79024	3.96540	4.33832	4.48885	4.82088
800	3.41478	3.51141	3.63416	3.82327	3.99713	4.36758	4.51722	4.84744
900	3.44673	3.54258	3.66441	3.85220	4.02494	4.39325	4.54211	4.87077
1000	3.47508	3.57027	3.69129	3.87792	4.04966	4.41609	4.56426	4.89154

Table 4 *Constants $z_m(\gamma)$ to be used in (asymptotic) joint confidence band*

m \ γ	0.6	0.7	0.8	0.9	0.95	0.99	0.995	0.999
2	1.41856	1.63429	1.91595	2.36663	2.79893	3.77782	4.19561	5.16338
3	1.60839	1.79814	2.04590	2.44232	2.82257	3.68360	4.05109	4.90235
4	1.73862	1.91476	2.14474	2.51271	2.86569	3.66494	4.00607	4.79625
5	1.83644	2.00383	2.22238	2.57207	2.90751	3.66705	3.99122	4.74214
6	1.91426	2.07539	2.28578	2.62239	2.94529	3.67644	3.98849	4.71133
7	1.97861	2.13496	2.33911	2.66575	2.97907	3.68854	3.99135	4.69276
8	2.03331	2.18585	2.38502	2.70370	3.00938	3.70156	3.99698	4.68129
9	2.08077	2.23017	2.42524	2.73736	3.03675	3.71468	4.00402	4.67424
10	2.12263	2.26938	2.46099	2.76757	3.06164	3.72754	4.01175	4.67009
20	2.38490	2.51715	2.68982	2.96610	3.23111	3.83120	4.08731	4.68058
30	2.52890	2.65443	2.81833	3.08057	3.33212	3.90171	4.14482	4.70795
40	2.62736	2.74870	2.90713	3.16061	3.40376	3.95435	4.18934	4.73367
50	2.70179	2.82015	2.97469	3.22196	3.45915	3.99623	4.22545	4.75643
60	2.76141	2.87750	3.02907	3.27158	3.50421	4.03096	4.25578	4.77655
70	2.81103	2.92529	3.07447	3.31318	3.54215	4.06062	4.28190	4.79448
80	2.85344	2.96619	3.11340	3.34894	3.57488	4.08648	4.30484	4.81063
90	2.89043	3.00189	3.14742	3.38028	3.60364	4.10941	4.32527	4.82530
100	2.92320	3.03354	3.17762	3.40815	3.62928	4.12999	4.34370	4.83873
200	3.13163	3.23539	3.37088	3.58766	3.79561	4.26647	4.46743	4.93295
300	3.24832	3.34874	3.47986	3.68966	3.89091	4.34661	4.54110	4.99162
400	3.32899	3.42723	3.55550	3.76074	3.95761	4.40339	4.59365	5.03437
500	3.39043	3.48707	3.61326	3.81515	4.00881	4.44734	4.63450	5.06804
600	3.43993	3.53532	3.65987	3.85915	4.05031	4.48316	4.66789	5.09583
700	3.48131	3.57568	3.69889	3.89604	4.08515	4.51337	4.69613	5.11949
800	3.51680	3.61032	3.73241	3.92777	4.11516	4.53949	4.72059	5.14009
900	3.54786	3.64063	3.76177	3.95558	4.14150	4.56248	4.74215	5.15835
1000	3.57544	3.66757	3.78786	3.98033	4.16495	4.58300	4.76142	5.17472

For z_m in Tables 1 and 4, we note that for example for $\gamma = 0.99$, z_m first decreases and then increases. This behavior is contrary to expectation, and we interpret that for the corresponding values of γ the asymptotic results should not be used for very small m.

Table 2 presents values of the relative differences between c_m, b_m and z_m, respectively.

Tables 1 and 2 show that for m as low as 10 or 20, the differences among the values are small, in the order of 10% or less, c_m and b_m are always less than z_m, and c_m is always less than b_m; see Section 4.4.3 below.

The first line of Table 1 contains the standard normal deviates. For $m = 50$ the constants bear to those in the first line the following relations: they are about twice for $\gamma = 0.90$. It follows that joint confidence bands are considerably wider than point-by-point confidence bands, for the usual values of γ, and for the frequently encountered values of m.

4.4.3. Analytic and asymptotic results

Since for $0 < \gamma < 1$ it holds that $\sqrt[m]{\gamma} < 1 - (1-\gamma)/m$, we have that

$$\frac{1 + \sqrt[m]{\gamma}}{2} < 1 - \frac{1-\gamma}{2m}, \tag{4.15}$$

and since Φ^{-1} is monotone increasing,

$$c_m(\gamma) < b_m(\gamma) \tag{4.16}$$

for all $m > 1$ and γ, $0 < \gamma < 1$.

As m increases, it is interesting to study the behavior of the constants. In Appendix 1 we derive asymptotic expressions for them, and show that

$$\lim_{m\to\infty} \frac{c_m}{z_m} = \lim_{m\to\infty} \frac{b_m}{z_m} = \lim_{m\to\infty} \frac{c_m}{b_m} = 1. \tag{4.17}$$

Moreover, these as well as the various constants themselves, are asymptotically independent of the confidence coefficient γ.

The interpretation of these asymptotic results should be done with care, due to the very large values of m (and hence of the sample size T) involved. We recall the comment by Hannan that we reproduced in Section 3. This fact motivates our inclusion of values as large as $m = 5000$ in Tables 1 and 2.

5. A simulation study: frequency domain, estimation by confidence bands

To study empirically the performance of some confidence bands for the spectral density function, we carry out a Monte Carlo experiment. We take three autoregressive models considered by Beamish and Priestley (1981) and Newton and

Pagano (1984), namely,

Model I $y_t - 0.4y_{t-1} - 0.45y_{t-2} = u_t$,
Model II $y_t + 1.7y_{t-1} + 2.4y_{t-2} + 1.634y_{t-3} + 0.872y_{t-4} + 0.168y_{t-5} = u_t$,
Model III $y_t - 2.7607y_{t-1} + 3.8106y_{t-2} - 2.6535y_{t-3} + 0.9238y_{t-4} = u_t$,

where u_t is a Gaussian white noise with unit variance. These models are used because they can be classified as being easy, moderately difficult and very difficult to fit, respectively, and also because they were studied by the indicated authors. Models will be identified as I, II and III. For each model we simulate N=1000 replicates of trajectories of length $T = 100$, 200 or 400, and for each of them we construct joint confidence bands.

Instead of the spectral density estimator defined in (2.4) by using the lag window, we use the average spectral density estimator defined by

$$\hat{f}(v_j) = (2\pi)^{-1} \sum_{|k| \le l_n} W_n(k) I_n(v_{j+k}),$$

where I_n is the periodogram, $l_n = l$ is a sequence of integers, and W_n is a sequence of weight functions. We take $l = 2$ and $W(-2) = W(2) = 1/8$, $W(-1) = W(0) = W(1) = 1/4$. Then, the asymptotic variance of $\hat{f}(v_j)$ for v_j unequal to 0 or π is given by

$$\beta^2 = \sum_{|k| \le 2} W^2(k) = \frac{7}{32}.$$

This choice corresponds to modified Daniel's window $w(x) = \text{sine}(\pi x)/(\pi x)$ for $-1 \le x \le 1$, and $m = T/5$.

Five different definitions of the confidence bands are considered, where logarithms are decimal, namely:

(i) $\log \hat{f}(v_j) + \log v_j + \log \chi^2_{(1-\gamma)/2}(\nu)$, $\log \hat{f}(v_j) + \log v_j - \log \chi^2_{\gamma/2}(\nu)$, where $\nu = 2\beta^{-2} = 9.14$.

(ii) Confidence band for $\log \hat{f}(v_j)$ defined in the S-PLUS program, which is based on χ^2 with 8.29 degrees of freedom.

(iii) $\log \hat{f}(v_j) \pm z_{(\gamma-1)/2}\beta$, based on a normal approximation to the distribution of $\log \hat{f}(v_j)$.

(iv) $\log \hat{f}(v_j) + \beta^{-2}/2 \pm z_{(\gamma-1)/2}\beta$, based on a normal approximation to the distribution of $\log \hat{f}(v_j)$ corrected for (asymptotic) bias.

(v) $\log \hat{f}(v_j) - \log(1+\beta)$, $\log \hat{f}(v_j) - \log(1-\beta)$. This is based on taking logarithms in (2.8).

Bands are computed with a joint confidence level of 0.95. They will be identified by (i) to (v).

Table 5 shows the results when the calculations are done at $T/10$ frequency points, which is the number used in the design. We also considered shown the results when all $T/2$-2 frequencies, $4\pi \le \nu \le 2\pi(T/2 - 1)/T$ are used in the calculations; this, however, is too demanding, and only the approach in Table 5 is shown.

Table 5 *Frequency distributions of points out of the 0.95 confidence band and percentage of coverage (0 or 1 points) ($v_j = 2\pi j 10/T$, $j = 3, 8, \ldots, T/10 - 2$)*

Sample Size	Model and Method	Points out of the band					Percentage of 0 or 1	Average
		0	1	2	3	4 or more		
$T = 100$	I (S-PLUS)	876	123	1		0	99.90%	0.125
	I (df = 9.14)	795	191	14		0	98.60%	0.219
	I (Normal)	662	282	54	1	1	94.40%	0.397
	I (Normal corr.)	860	135	5		0	99.50%	0.145
$T = 200$	I (S-PLUS)	922	71	7		0	99.30%	0.085
	I (df = 9.14)	836	149	14	1	0	98.50%	0.18
	I (Normal)	609	311	72	7	1	92.00%	0.48
	I (Normal corr.)	865	125	10		0	99.00%	0.145
$T = 400$	I (S-PLUS)	936	63	1		0	99.90%	0.065
	I (df = 9.14)	861	124	14	0	1	98.50%	0.156
	I (Normal)	571	314	96	14	5	88.50%	0.569
	I (Normal corr.)	864	123	12	1	0	98.70%	0.15
$T = 100$	II (S-PLUS)	276	548	142	32	2	82.40%	0.936
	II (df = 9.14)	290	526	160	23	1	81.60%	0.919
	II (Normal)	365	504	115	16	0	86.90%	0.782
	II (Normal corr.)	433	502	60	5	0	93.50%	0.637
$T = 200$	II (S-PLUS)	582	394	24		0	97.60%	0.442
	II (df = 9.14)	567	383	47	3	0	95.00%	0.486
	II (Normal)	569	336	82	12	1	90.50%	0.541
	II (Normal corr.)	698	280	21	1	0	97.80%	0.325
$T = 400$	II (S-PLUS)	820	168	12		0	98.80%	0.192
	II (df = 9.14)	772	211	14	3	0	98.30%	0.248
	II (Normal)	579	325	88	8	0	90.40%	0.525
	II (Normal corr.)	844	145	9	2	0	98.90%	0.169
$T = 100$	III (S-PLUS)	13	122	291	331	243	13.50%	2.74
	III (df = 9.14)	20	144	336	316	184	16.40%	2.552
	III (Normal)	60	226	412	229	73	28.60%	2.039
	III (Normal corr.)	64	236	418	217	65	30.00%	1.991
$T = 200$	III (S-PLUS)	482	382	116	19	1	86.40%	0.675
	III (df = 9.14)	444	403	123	28	2	84.70%	0.741
	III (Normal)	416	417	139	23	5	83.30%	0.785
	III (Normal corr.)	644	300	50	5	1	94.40%	0.419
$T = 400$	III (S-PLUS)	886	108	6		0	99.40%	0.12
	III (df = 9.14)	842	145	13		0	98.70%	0.171
	III (Normal)	621	298	63	16	2	91.90%	0.48
	III (Normal corr.)	880	114	6		0	99.40%	0.126

Table 5 shows the frequencies of coverage of the bands in $N = 1000$ replications. Method (iv) based on the normal approximation corrected for bias, and method (ii) based on the χ^2 approximation available in S- PLUS, give the best results. Method (v) produced poor results and it was discarded. Method (iii) based on the normal approximation without correction for bias is clearly inferior.

Even when the frequencies of strict coverage are not too close to the 95% theoretical level, we can accept the performance of the confidence bands as providing a correct frequency of coverage, if we are willing to allow up to one point out of the band. Figures 1, 2 and 3 show that methods (i), (ii) and (iv) give almost the same confidence bands. Figures 4 to 8 show bands for five replicates calculated with method (iv) for each model and selected sample sizes. In the case of Model III, Figures 6, 7 and 8, it is clear the effect of increasing sample size, in that the bands tend to show less bias and are more concentrated among them.

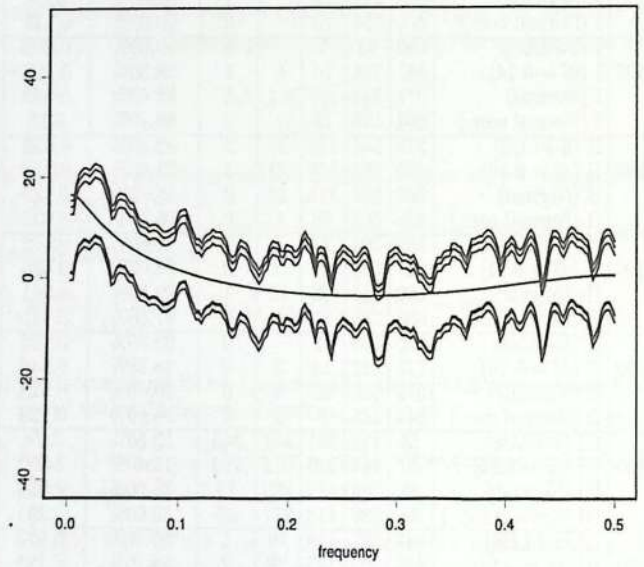


Figure 1 *0.95 confidence bands for the spectral density, AR(2) model I, $T = 400$, methods (i) to (iv)*

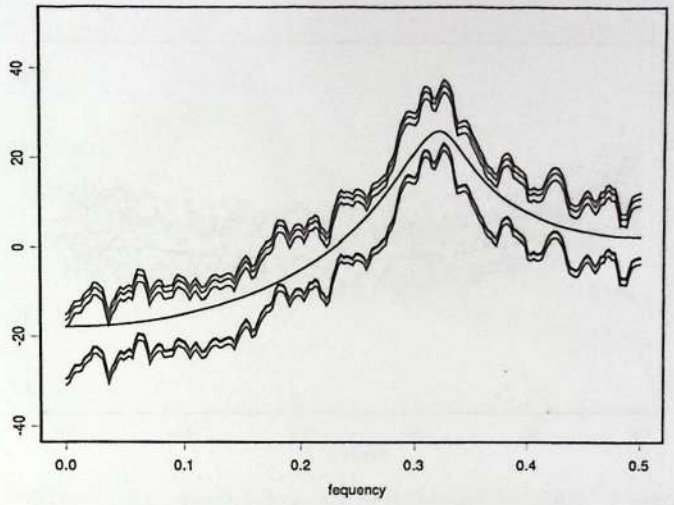


Figure 2 *0.95 confidence bands for the spectral density, AR(5) model II, $T = 400$, methods (i) to (iv)*

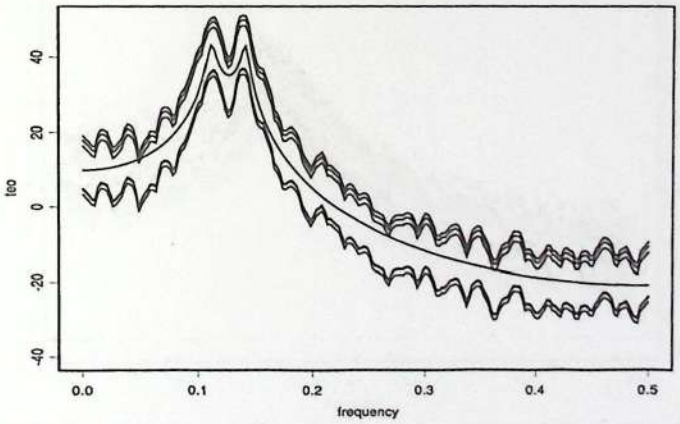


Figure 3 *0.95 confidence bands for the spectral density, AR(4) model III, $T = 400$, methods (i) to (iv)*

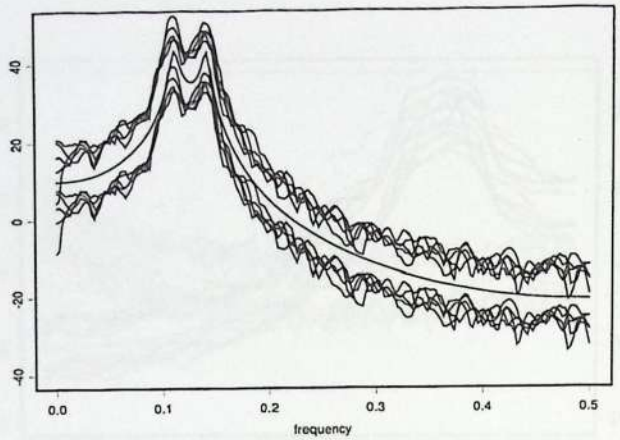


Figure 8 *0.95 confidence bands for the spectral density, AR(4) model III, $T = 400$, method (iv), 5 replicates*

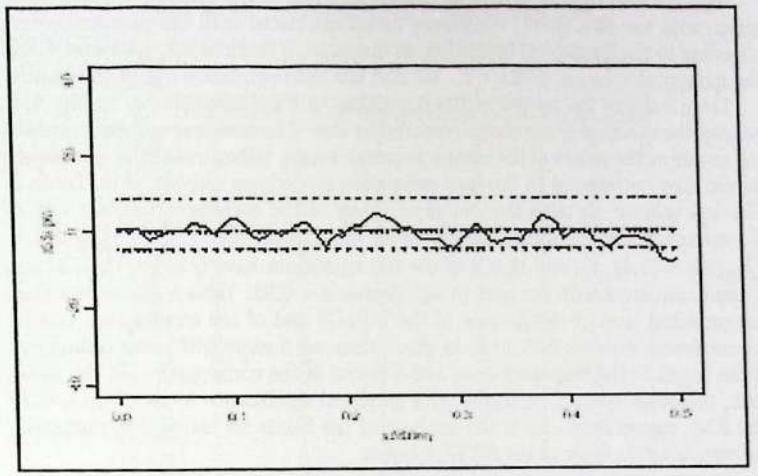
6. A simulation study: testing hypotheses

In this section we explore the use of confidence bands to test the hypothesis that a given time series is white noise. The procedures in the frequency and time domains are compared with the use of the modified Ljung-Box-Pierce portmanteau statistic (Box, Jenkins and Reinsel, Section 8.2.2), defined by

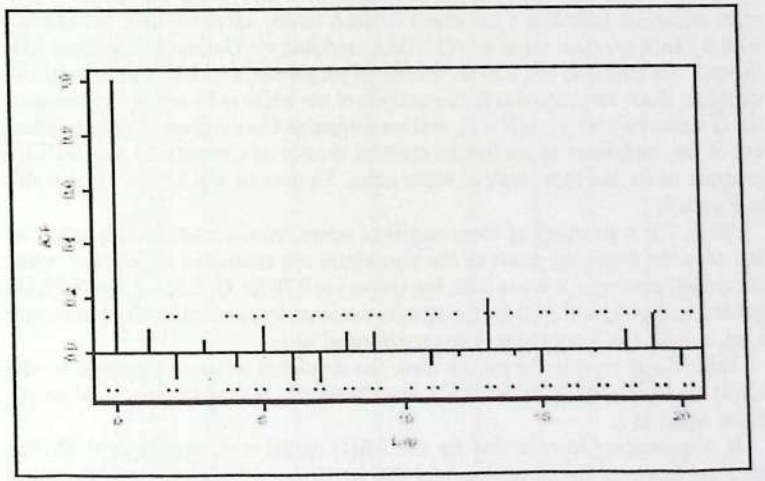
$$\tilde{Q} = T(T+2)\sum_{t=1}^{K} \frac{r_t^2}{T-t}\,, \tag{6.1}$$

where the r_j are the sample autocorrelations of the given series, and K is a suitably chosen time lag so that little is lost by omitting r_t's with $t > K$ (Box et al., op. cit.)

Figure 9 shows the kind of situation that we are studying. Represented are the logarithms of the spectral quantities, the center straight line corresponding to $\log[2\pi(\sigma^2/2\pi)] = \log(\sigma^2) = \log 1 = 0$, where $\sigma^2/2\pi$ is the spectral density of white noise. The solid line is $\log[2\pi\hat{f}(\nu)]$, where $\hat{f}$ comes from the S-PLUS program, and the broken lines are the 95% (joint) confidence bands for white noise compute by S-PLUS. The spectral densities are computed at $T/10$ frequency points; at none of these the estimated spectral density falls out of the bands. However, the plot is made at $T/2$ frequency points, and at some of these the solid line lies out of the bands.



(a) Spectral density series



(b) Correlogram

Figure 9 *An example of a simulated white series $T = 200$, tested for H_0: white noise. Level of significance 0.05*

Figure 9 also shows the sample autocorrelogram of the generated white noise series, with the 95% (joint) confidence bands calculated with the constant corresponding to the Bonferroni inequality, as discussed in Sections 4.3, 4.4.1 and 4.4.2; the numerical values is $3.023/\sqrt{T}$. We find one autocorrelation out of the bands.

Table 6 shows the results of 100 repetitions of these calculations. In the first column the values of $\tilde{Q}$ are shown, ordered by size. The next four columns contain the counts of the points of the sample spectral density falling out of the confidence bands; they correspond to the four estimation procedures considered in Table 5. The last column contains the counts of values of the correlogram falling out of the corresponding confidence bands. Since the critical value of $\tilde{Q}$ is $\chi^2_{0.05}(K) = \chi^2_{0.05}(20) = 31.41$, we find that 8 of the 100 repetitions have $\tilde{Q}$ larger than 31.41, a result consistent with the level of significance $\alpha = 0.05$. Table 5 also shows that the empirical level of significance of the S-PLUS and of the correlogram bands, are consistent with the 0.05 level: in effect, there are 3 series with point falling out of the bands in the frequency case, and 5 points in the correlogram. At the same time, the other spectral estimates give empirical significance levels of 0.11, 0.30 and 0.10, respectively. As in the analysis of the bands for estimation purposes, the results are in favor of the S-PLUS bands.

We next apply the indicated procedure to gain some indication of the discriminatory power of the procedures. For this purpose, we test whether simulated AR(1) series, lead to rejection of the white noise hypothesis. We generate by Monte Carlo, 100 AR(1) series of lengths $T = 200$ using the S-PLUS program arima.sim (simulate a univariate ARIMA series), estimate their parameters with S-PLUS program arima.mle (ARIMA modeling via Gaussian maximum likelihood). We used 0.30 and 0.60 for the model parameter, and 1 for the innovations variance. Then we proceed as in the analysis of the white noise series: we compare the $\tilde{Q}$ statistics with $\chi^2_{0.05}(K-1)$, and we determine the numbers of points falling out of the confidence bands for the spectral density as compute by the S-PLUS program under the hypothesis of white noise. Figures 10 and 11 are comparable to Figure 9.

Table 7 is a summary of these empirical power calculations. In this table we find that the empirical power of the procedures are estimated as follows: when the model parameter is set at 0.30, the values are 0.76 for $\tilde{Q}$, 0.34 for the S-PLUS spectral estimate, and 0.90 for the sample autocorrelogram; when the parameter is set at 0.60, the 3 estimates of power are equal to 1.

Table 7 also reports the results when the simulated series corresponds to the AR(2) model introduced in Section 5, and the finding is that the empirical power is also equal to 1.

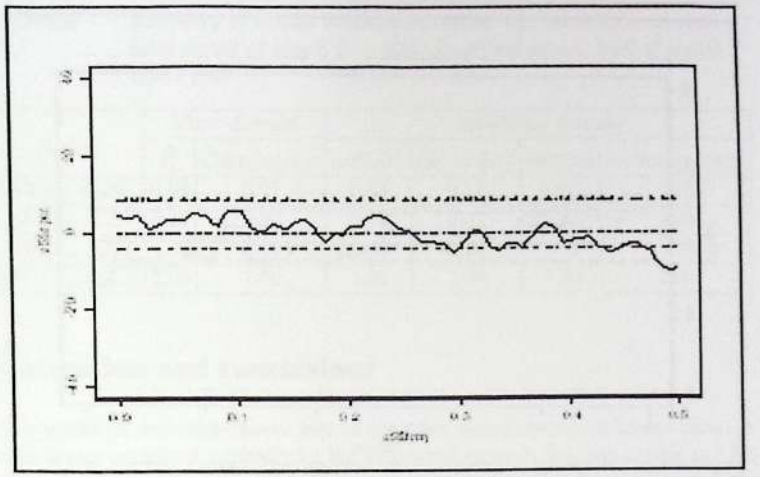
It is interesting to note that for the AR(1) model with parameter 0.30, the empirical power of the correlogram for this particular set of series is 0.90, which represents an improvement over the value of 0.76 corresponding to the use of the $\tilde{Q}$ statistic. More detailed studies should be conducted to investigate the power of the various procedures considered in the present study, but this is beyond the scope of the present paper.

Table 6 *100 simulated white noise series, $T = 200$, tested for H_0: white noise by six procedures, level of significance 0.05*

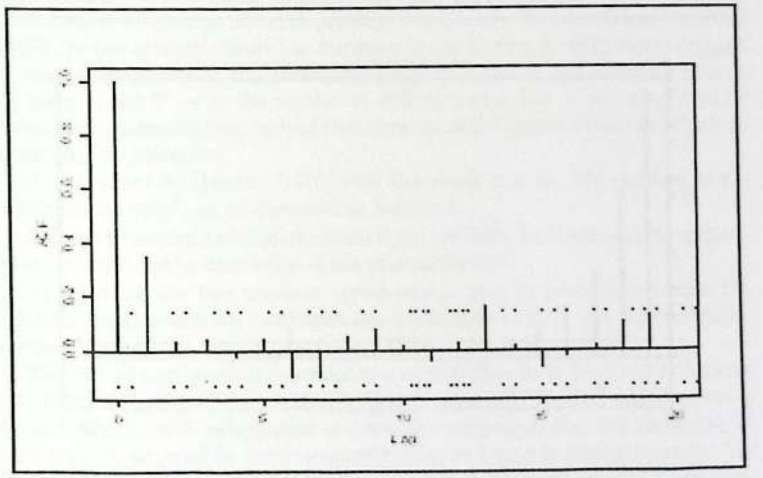
Time domain	Frequency domain				Time domain
$\tilde{Q}$	S-PLUS	DF = 9.14	Normal	Normal correc.	Correlogram
9.090	0	1	1	1	0
11.096	0	0	0	0	0
11.526	0	0	0	0	0
11.954	0	0	0	0	0
12.244	0	0	0	0	0
12.325	0	0	0	0	0
12.387	0	0	1	0	0
12.399	0	0	0	0	0
12.604	0	0	0	0	0
12.626	0	0	0	0	0
13.683	0	0	0	0	0
13.732	0	0	0	0	0
13.774	0	0	0	0	0
13.994	0	0	0	0	0
14.165	0	0	1	0	0
14.282	1	1	1	1	0
14.418	0	0	0	0	0
14.483	0	0	0	0	0
14.516	0	0	0	0	0
14.595	0	0	0	0	0
14.666	0	0	0	0	0
14.786	0	0	0	0	0
14.880	0	0	0	0	0
15.013	0	0	0	0	0
15.424	0	0	0	0	0
15.443	0	0	0	0	0
15.445	0	0	0	0	0
15.617	0	0	0	0	0
15.632	0	0	0	0	0
15.858	0	0	0	0	0
16.243	0	0	1	0	0
16.244	0	0	0	0	0
16.618	0	0	0	0	0
16.770	0	0	0	0	0
16.787	0	0	0	0	0
16.805	0	0	0	0	0
16.912	0	0	0	0	0
17.069	0	0	0	0	0
17.130	0	0	0	0	0
17.206	0	0	0	0	0
17.400	0	0	1	0	0
18.075	0	0	1	0	0
18.284	0	0	0	0	0
18.495	0	0	0	0	0
18.608	0	1	2	1	0
18.751	0	0	0	0	1
18.767	0	0	0	0	0
18.777	0	0	0	0	0
18.810	0	0	0	0	0
18.916	0	1	1	1	0
19.258	0	0	0	0	0
19.289	0	0	1	0	0

Table 6 *Continuation*

Time domain	Frequency domain				Time domain
$\tilde{Q}$	S-PLUS	DF = 9.14	Normal	Normal correc.	Correlogram
19.429	0	0	0	0	0
19.595	0	0	0	0	0
19.615	0	0	0	0	0
19.868	0	0	0	0	0
19.877	0	0	1	0	0
20.203	0	0	0	0	0
20.217	0	0	1	0	0
20.506	0	0	0	0	0
20.663	0	0	0	0	0
21.396	0	0	0	0	1
21.409	0	0	0	0	0
21.563	0	0	0	0	0
22.013	0	0	0	0	0
22.195	0	0	1	0	0
22.197	0	1	2	1	0
22.368	0	0	0	0	0
22.529	0	0	0	0	0
22.733	0	0	0	0	0
23.192	0	0	0	0	0
23.303	0	0	0	0	0
23.655	0	0	1	0	0
23.694	0	0	0	0	0
23.761	0	0	1	0	0
23.765	0	0	1	0	0
24.134	0	0	0	0	0
24.475	0	0	1	0	0
25.379	0	1	1	0	0
25.560	0	0	0	0	0
26.289	1	1	1	1	0
26.426	0	1	2	1	0
26.887	0	0	0	0	0
26.896	0	0	0	0	0
27.258	0	0	0	0	0
27.625	0	0	1	0	0
28.476	0	0	1	0	0
29.354	0	0	0	0	0
30.051	0	0	1	0	0
30.419	0	0	0	0	0
31.109	1	1	2	1	0
31.318	0	0	1	0	0
31.483	0	0	0	0	0
31.793	0	1	1	1	0
32.372	0	0	0	0	1
32.792	0	0	1	0	0
33.748	0	0	1	0	1
34.513	0	1	1	1	0
42.475	0	0	0	0	0
50.619	0	0	0	0	2
Number of rejections of H_0	0	3	11	30	10	5
Levels	0.00	0.03	0.11	0.30	0.10	0.05

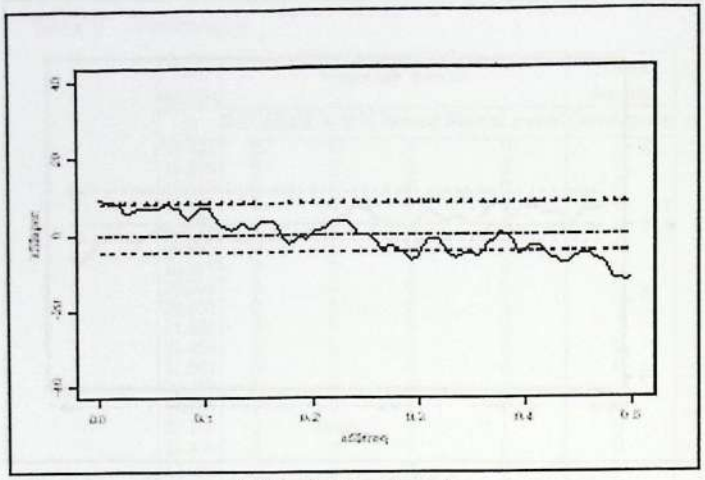


(a) Spectral density series

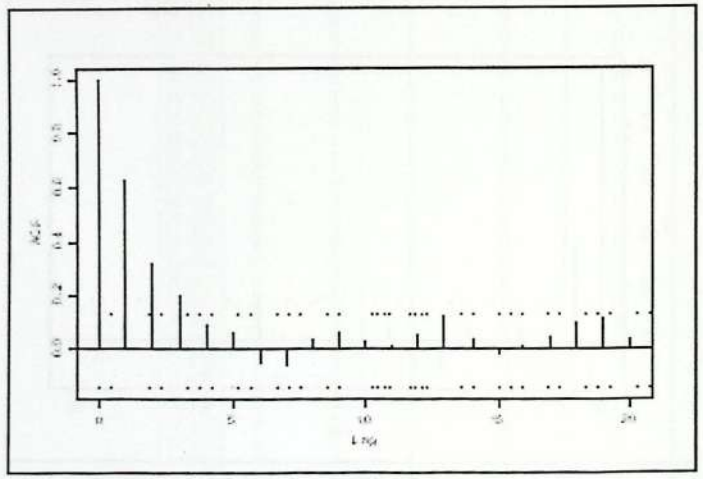


(b) Correlogram

Figure 10 *An example of a simulated AR(1) series with parameters 0.30, $T = 200$, tested for H_0: white noise. Level of significance 0.05*



(a) Spectral density series



(b) Correlogram

Figure 11 *An example of a simulated AR(1) series with parameters 0.60, $T = 200$, tested for H_0: white noise. Level of significance 0.05*

Table 7 *Summary of results obtained in testing H_0: white noise in simulated series of length $T = 200$. Empirical power, level of significance 0.05*

Series	Time domain		Frequency domain			
	$\tilde{Q}$	Correlogram	S-PLUS	DF = 9.14	Normal	Normal correct.
White noise	0,08	0.05	0.03	0.11	0.30	0.10
AR(1), 0.30	0.76	0.90	0.34	0.37	0.58	0.25
AR(1), 0.60	1.00	1.00	1.00	1.00	1.00	1.00
AR(2)[Sect.5]	1.00	1.00	1.00	1.00	1.00	1.00

7. Discussion and conclusions

The question was raised about how to construct simultaneous confidence bands in time series problems, typically for the autocorrelogram in the time domain and for the spectral density function in the frequency domain. Simultaneous procedures are such that control of the confidence coefficient is given to the researcher, who then knows what probability level (either exact or approximate) is associated with the simultaneous inferencial statement he makes on the basis of his observations.

A formal solution to one such problem was given by Woodroofe and Van Ness (1967) for the spectral density, as discussed in our Section 3. With some adequate general assumptions on the underlying stochastic process, and assuming that as the sample size $T \to \infty$ the number of ordinates $m = m_T \to \infty$, an asymptotic probability statement was derived that uses one of the extreme value distributions known in the literature.

It was noted by Hannan (1970) that this result may be "the roughest of approximations only", as we discussed in Section 3.

A close-form result of similar nature is not available for the time domain quantities, at least to the knowledge of the present writers.

At least for the two previous arguments, it pays to investigate further the problem. In this note we considered two possibilities: (1) To use the asymptotic independence of the sample quantities; (2) To resort to inequalities.

The use of asymptotic independence is explored briefly in Section 4.2. It turns out that the asymptotic joint distribution of windowed spectral estimators at a set of frequencies is independent multivariate normal, so that the evaluation of constants to be used in simultaneous confidence bands is straightforward. We denote these constants by $c_m(\gamma)$. These same constants can be used in the time domain for the autocorrelations, but here the asymptotic independence has to be assumed; then we interpret that the confidence band is directed to a comparison against a white noise hypothesis.

The constants $c_m(\gamma)$ evaluated under the asymptotic normality and independence can also be justified on the basis of asymptotic joint normality only, as follows: For a multivariate normal density there is an inequality (4.3) that leads to the use of the $c_m(\gamma)$ in the construction of conservative simultaneous confidence

bands for m quantities, with probability level γ.

The need to use the asymptotic joint normality can in turn be relaxed: Bonferroni's inequality is used in Section 4.4.1 to derive (conservative) constants $b_m(\gamma)$ that use the asymptotic *marginal* normal distribution of each sample quantity under consideration.

In terms of assumptions we see that the three approaches (for the frequency domain) can be interpreted as follows: (a) Bonferroni's inequality, and hence the use of the associated constants $b_m(\gamma)$, requires only that the marginal distribution of each sample quantity be approximated by its normal limit; (b) The inequality in (4.3), and hence the constant $c_m(\gamma)$, require that the joint distribution of the sample statistics be approximated by its normal limit; (c) The approach of Woodroofe and Van Ness and the corresponding constants, denoted $z_m(\gamma)$ by us, requires not only that the sample size $T \to \infty$ but that m, the sample quantities under consideration, satisfies $m = m_T \to \infty$, in such a way that $g(m_T)/T \to \infty$ for some suitable function g.

We conclude that the use of inequalities in the way we we did in Section 4.4. has several advantages: (1) It requires less assumptions on the set up of the problem; (2) It is simpler, since for example the proof of the results in Woodroofe and Van Ness is quite complicated; (3) It provides more flexibility, in particular, can be used to derive bands of non-uniform width.

About this last point, the results we presented so far are for bands of uniform width. This is not what is always recommended in the literature. The BMDP-81 package, for example, produces a "control line" for the sample autocorrelogram of the form given in (2.3), where the approximation

$$\hat{\tau}_{ss} \cong \frac{1}{T}\sum_{j=0}^{s-1} r_j^2 \tag{7.1}$$

is used, as suggested originally by Bartlett (1946). This then produces a band that in general has increasing width, and that in the computer appears with nondecreasing width, due to rounding.

The idea of nonuniform significance levels was also considered in multiple decision problems by Anderson (1971). In our case this will lead to allowing m to be considerably large, but still small compared with T, and letting the width increase with the index of the sample statistics. This appears as a reasonable procedure, and will make less important the exact determination of m in the definition of the sample autorrelogram.

The numerical values of $z_m(\gamma)$, $c_m(\gamma)$ and $b_m(\gamma)$ derived for the result of Woodroofe and Van Ness, the inequality in (4.3), and Bonferroni's inequality, respectively, were evaluated and compared. They differ for very small values of m, for which in fact $z_m(\gamma)$ is not recommended. For about 20 or 30 sample quantities, which may be a useful number for practical use, $c_m(\gamma)$ and $b_m(\gamma)$ differ only slightly, and their differences with $z_m(\gamma)$ are also small.

For standard values of γ (0.90, 0.95, 0.99) and for m in the order of 30, the constants are roughly in the order of 1.5 times those of the univariate confidence

intervals. Hence, simultaneous confidence bands will be wider that those formed with point-by-point constants.

To see how well this kind of procedure works for finite sample sizes (T), Monte Carlo simulation experiments were conducted. In terms of estimation, four procedures are compared for $T = 100$, 200 and 400 with 1000 replications. The results are interesting: defining as satisfactory coverage of the true spectral density by 0.95 confidence bands, the occurrence up to one point out of the band, three of the procedures work reasonable well. As expected, they are wider than point-by-point bands of the same probability level. In terms of using some of the results to test the hypothesis of white noise, the joint confidence bands for the correlogram showed to perform better than using the $\bar{Q}$ statistic, and also that the bands in the frequency domain. The study of power presented in this paper is, however, quite limited, and further studies are required to evaluate the corresponding empirical measures.

Appendix

In this Appendix we prove some assertions made in Section 4.4.3. For $m \to \infty$ we derive asymptotic expansions for b_m, c_m and z_m, show that they are independent of the confidence coefficient γ, and that they are asymptotically equal.

The following inequality holds for $x > 0$ (Pickands, 1969),

$$\psi(x)\left(1 - \frac{1}{x^2}\right) \le 1 - \Phi(x) \le \psi(x), \tag{A.1}$$

where $\Phi(x)$ is the standard normal distribution function, and

$$\psi(x) = \frac{1}{x}\phi(x) = \frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}. \tag{A.2}$$

Lemma 1 *Let $0 < \beta < 1$, and x_1, x_2 and x_3 be respectively the roots of the equations*

$$\psi(x_1)\left(1 - \frac{1}{x_1^2}\right) = \beta, \tag{A.3}$$

$$1 - \Phi(x_2) = \beta, \tag{A.4}$$

$$\psi(x_3) = \beta. \tag{A.5}$$

Then $x_1 \ge x_2 \ge x_3$.

To find asymptotic expansions for b_m and c_m we find asymptotic expansions for $x_1 = x_1(m)$ and $x_3 = x_3(m)$ by setting $\beta = (1-\gamma)/2m$ and $\beta = (1 - \sqrt[m]{\gamma})/2$, respectively.

Lemma 2 *Let b_m be the root of the equation*

$$1 - \Phi(b_m) = \frac{1-\gamma}{2m},$$

from (A.4) and (4.12). Then $b_m^2 \sim 2\ln(2m)$ *as* $m \to \infty$, *where the notation* $\sim$ *means asymptotically equal.*

Proof. Equation (A.5) is now

$$\frac{1}{x_3\sqrt{2\pi}}e^{-\frac{1}{2}x_3^2} = \frac{1-\gamma}{2m};$$

it follows that

$$x_3^2 = -2\ln x_3 - \ln 2\pi - 2\ln(1-\gamma) + 2\ln(2m). \qquad \text{(A.6)}$$

Equation (A.3) is now

$$\frac{1}{x_1\sqrt{2\pi}}e^{-\frac{1}{2}x_1^2}\left(1-\frac{1}{x_1^2}\right) = \frac{1-\gamma}{2m};$$

it follows that

$$x_1^2 = -2\ln x_1 - \ln 2\pi - 2\ln(1-\gamma) + 2\ln(2m) + \ln\left(1-\frac{1}{x_1^2}\right). \qquad \text{(A.7)}$$

From (A.6) and (A.7) it follows that both x_1^2 and x_3^2 are asymptotically equal to $2\ln(2m)$. From Lemma 1 it follows that $b_m^2 \sim 2\ln(2m)$.

Lemma 3 *Let* c_m *be the root of the equation*

$$1 - \Phi(c_m) = \frac{1-\sqrt[m]{\gamma}}{2},$$

from (A.4) and (4.10). Then $c_m^2 \sim 2\ln(2m)$ *as* $m \to \infty$.

Proof. Equation (A.5) is now

$$\frac{1}{\sqrt{2\pi}x_3}e^{-\frac{1}{2}x_3^2} = \frac{1-\sqrt[m]{\gamma}}{2} = \frac{1-\gamma}{2m} + 0\left[\left(\frac{1-\gamma}{m}\right)^2\right]$$

and therefore

$$e^{-\frac{1}{2}x_3^2} = \sqrt{2\pi}x_3\left(\frac{1-\gamma}{2m} + 0\left[\left(\frac{1-\gamma}{m}\right)^2\right]\right),$$

$$x_3^2 = -2\ln x_3 - \ln 2\pi(1-\gamma) + 2\ln(2m) + 0\left(\frac{1-\gamma}{m}\right), \qquad \text{(A.8)}$$

Equation (A.3) is now

$$\frac{1}{\sqrt{2\pi}x_1}e^{-\frac{1}{2}x_1^2}\left(1-\frac{1}{x_1^2}\right) = \frac{1-\sqrt[m]{\gamma}}{2},$$

and therefore

$$x_1^2 = -2\ln x_1 - \ln 2\pi - 2\ln(1-\gamma) + 2\ln(2m) + 0\left(\frac{1-\gamma}{m}\right) + \ln\left(1-\frac{1}{x_1^2}\right). \qquad \text{(A.9)}$$

Hence, from (A.8), (A.9) and Lemma 1 it follows that $c_m^2 \sim 2\ln(2m)$ as $m \to \infty$.

Lemma 4 *Let* z_m *be defined by (4.14). Then* $z_m^2 \sim 2\ln(2m)$ *as* $m \to \infty$.

Proof. From the definitions of a and b following (3.1) we have that

$$z_m = \sqrt{2\ln(2m)} - \frac{1}{2\sqrt{2\ln(2m)}}[\ln\ln(2m) + \ln 2\pi + 2\ln(-\ln\gamma)]. \qquad \text{(A.10)}$$

and the conclusion of the lemma follows.

By using Taylor's theorem, the following asymptotic expansion can be derived,

$$z_m^2 \sim 2\ln(2m) - \ln\ln(2m) - \ln 2\pi - 2\ln(-\ln\gamma), \qquad \text{(A.11)}$$

which compares directly with those for x_1^2 and x_3^2 derived in the proofs of lemmas 2 and 3.

The results in (4.17) follow immediately from Lemmas 2, 3 and 4. Further c_m, b_m and z_m are asymptotically (as $m \to \infty$) independent of the confidence coefficient γ.

Acknowledgements

The computations of Tables 1 to 4 were kindly done for us by Eng. Ricardo E. Gonzalez. Parts of this work were presented in meetings and workshops, where they received useful comments from Professors William Kruskal (U. of Chicago), Gunnar Kulldorff (U. of Umea), Peter Schonfeld (U. of Bonn), and members of the institutes headed by the last two. We also thank useful comments by members of our Institute of Statistics, University of Tucumán.

(Received August, 2002. Accepted December, 2003.)

References

Akaike, H. (1969). Power spectrum estimation through autoregressive model fitting. *Ann. Inst. Statist. Math.*, **21**, 407–419.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* John Wiley and Sons.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. *J. R. Stat. Soc. (suppl)*, **7**, 211.

Beamish, N. and Priestley, M. B. (1981). A study of AR and window spectral estimation. *Appl. Statist.*, **30**.

Box, G. E. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Revised Edition, Holden-Day.

Cramer, H. and Leadbetter, M. (1967). *Stationary and Related Stochastic Processes.* Hohn Wiley and Sons.

Chung, K. L. and Erdos, P. (1952). On the applications of the Borel-Cantelli lemma. *Trans. Amer. Math. Soc.*, **72**, 178–186.

Das Gupta, S., Eaton, M. L., Olkin, I., Perlman, M. K., Savage, L. K. J. and Sobel, M. (1972). Inequalities on the probability content of convex regions for elliptically contoured distributions. *Proc. Sixth Berkely Symp. Math. Statisti. Probab., 2*, University of California Press.

Dunn, O. (1958). Estimation of the means of dependent variables. *Ann. Math. Stat.*, **29**, 1095–1111.

Granger, W. and Hatanaka, M. (1964). *Spectral Analysis of Economic Time Series.* Princenton University Press.

Hannan, E. (1970). *Multiple Time Series.* John Wiley and Sons.

Hrafnkelsson, B. and Newton, H. J. (2000). Asymptotic simultaneous confidence bands for vector autoregressive spectra. *Biometrika*, **87**, 173–182.

Katri, C. G. (1967). On certain inequalities for normal distributions and their applications to simultaneous confidence bounds. *Ann. Math. Stat.*, **38**, 1853–1867.

Koslov, J. W. and Jones, R. H. (1985). A unified approach to confidence bounds for the autoregressive spectral estimator. *J. of Time Series Anal.*, **6**, 141–151.

Mentz, R. P. (1983a). On the use of the sample autocorrelogram. Presented at the *First Catalán International Symposium on Statistics*, Barcelona, September 1983.

Mentz, R. P. (1983b) Plotting control lines in the sample autocorrelogram. *Contributed Papers, 44th. Session of the I.S.I.*, **2**, 709–712.

Miller, R. (1966). *Simultaneous Statistical Inference.* McGraw-Hill Book Co.

Newton, H. J. and Pagano, M. (1984). Simultaneous confidence bands for autoregressive spectra. *Biometrika*, **71**, 1, 197–202.

Owen, D. (1962). *Handbook of Statistical Tables.* Addison Wesley Publishing Co.

Parzen, E. (1974), Some recent advances in time series modelling. *I.E.E.E. Trans. Automatic Control*, **AC-19**, 723–729.

Priestley, M. (1981). *Spectral Analysis and Time Series.* Vols. 1 and 2. London: Academic Press.

Pickands III, J. (1969). Upcrossing probabilities for stationary Gaussian Processes. *Trans. Amer. Math. Soc.*, **145**, 51–73.

Sakai, H. and Sakaguchi, F. (1990). Simultaneous confidence bands for the spectral estimate of two-channel autoregressive processes. *J. of Time Series Anal.*, **11**, 49–56.

Sidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Stat. Assoc.*, **62**, 626–633.

Tomasek, L. (1987). Asymptotic simultaneous confidence bands for autoregressive spectral density. *J. of Time Series Anal.*, **8**, 469–477.

Tong, G. L. (1980). *Probability Inequalities in Multivariate Distributions.* Academic Press.

Walker, A. M. (1965). Some asymptotic results for the periodogram of a stationaty time series. *J. Austral. Math. Soc.*, **5**, 107–128.

Woodroofe, M. and Van Ness, J. (1967). The maximum deviation of sample spectral densities. *Ann. Math. Stat.*, **36** 1558–1569.

Raúl Pedro Mentz, Aldo J. Viollaz and Carlos I. Martinez

Universidad Nacional de Tucumán and CONICET

Casilla de correo 209

4000 - San Miguel de Tucumán, República Argentina

E-mail: rmentz@herrera.unt.edu.ar

Correlación, simetría y variabilidad

Raúl P. Mentz

Introducción

En muchos problemas de estadística aplicada, las observaciones numéricas disponibles para el análisis forman pares, un par de números asociado con cada individuo. Por ejemplo, podemos tener los pesos de niños antes y después de un tratamiento, calificaciones escolares al comienzo y al final de un experimento educativo, altura y peso de atletas, cantidades compradas y precios unitarios pagados por una persona en varias compras o por diferentes personas, etc. Una técnica útil del análisis de datos para conjuntos bidimensionales de la forma $\{(x_1, y_1),...,(x_n, y_n)\}$, es construir un gráfico de "puntos" $P_i = (x_i, y_i)$ en un sistema común de coordenadas cartesianas ortogonales, con sus correspondientes ejes x o de las abscisas e y de las ordenadas. Esta representación gráfica del conjunto de puntos o pares será llamado un *diagrama de dispersión*.

El diagrama de dispersión es adecuado para estudiar problemas de *correlación*. Más adelante definiremos el concepto de correlación para un diagrama de dispersión, pero previamente trataremos de desarrollar un sentido intuitivo para este concepto y para r, la medida de correlación. Una alternativa es leer la Sección II antes de la I.

1. Ejemplos
Significado de correlación

Para ilustrar el significado de la correlación, se presenta a menudo un argumento gráfico como el siguiente. En el Gráfico 1 se muestran cinco diagramas de dispersión, cada uno con 100 puntos o pares. Cuando algunos pares coinciden se usa una marca mas gruesa. Los pares en la parte (a) del gráfico son aproximadamente no correlacionados, r es aproximadamente igual a cero. A medida que descendemos en el gráfico, la correlación aumenta, y en la parte (e) tenemos el valor extremo de r=1, los pares están todos en una recta, en este caso una que pasa por el origen del sistema de coordenadas.

El lector puede suponer que en todos los casos, a medida que los puntos se concentran alrededor de una recta, la correlación aumenta y se aproxima a 1. En primer lugar, si la concentración ocurre a lo largo de una recta de pendiente negativa [por ejemplo una que pasa por el origen y por el punto (x, y) = (-1, 1)], no se aproxima a 1 sino a –1. Más importante aún es que si la concentración ocurr

con relación a una recta paralela al eje x (o coincidente con él), r no se aproxima a 1 o –1 sino que lo hace a 0.

Por lo tanto el signo y el grado de la correlación (y el valor de r) son diferentes según el diagrama de dispersión se concentre con relación a una recta o a otra, y debemos diferenciar las rectas con pendientes no nulas de aquellas con pendiente nula: pendiente 0 de la recta de concentración significa que r = 0. Si denotamos por b la pendiente de la "recta de concentración" (que se definirá con precisión en la Sección II), tenemos que en la parte (a) del Gráfico 1, r = b = 0, en la parte (e) r = + 1 con b > 0, mientras que en las partes intermedias de (b) a (d), r y b son ambos positivos y 0 < r < 1.

Detalle Técnico. La parte (a) del diagrama de dispersión del Gráfico 1 se construyó con pares de números seudo-aleatorios, distribuidos uniformemente entre 0 y 1. Para cada par (x, y), su proyección a la línea y = x es ((x + y)/2, (x + y)/2), y ellos forman los pares de la parte (e). Los puntos de las rectas que unen ambos pares están dados por (ax + (1-a)(x + y)/2, ay + (1-a)(x + y)/2) para a entre 0 y 1. Por lo tanto las partes (a) hasta (e) corresponden a los valores a = 1, 0.75, 0.5, 0.25 y 0, respectivamente. A estos diagramas de dispersión corresponden los valores r = 0.03, 0.31, 0.62, 0.89 y 1 señalados en el gráfico.

Diagramas de dispersión simétricos

Un enfoque que se utiliza a menudo para ilustrar el sentido de la falta de correlación (r = 0) es recurrir a los diagramas simétricos. Los diagramas en el Gráfico 2 tienen r = 0.

Cualquiera de estos diagramas puede alterarse (haciéndolo más disperso) en una o ambas direcciones, y todavía corresponderá a r = 0. Por ejemplo, los diagramas de dispersión en el Gráfico 3 se obtuvieron del diagrama a la izquierda del Gráfico 2, y todos tienen r=0.

Sin embargo, la presencia de simetría no es fácil de detectar. Algunos de los diagramas de dispersión en el Gráfico 4 tienen r=0 mientras que otros tienen r<0 ó r>0. Por lo tanto se justifica analizar el tema de la simetría con más detención. Lo haremos después de introducir algo de teoría y una notación.

2. Notación

Para un conjunto de n pares de números, $\{(x_1, y_1), ..., (x_n, y_n)\}$, definimos el *coeficiente de correlación de Pearson* por

$$(1) \qquad r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Donde $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \quad \bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j$.

Es útil representar a las sumas con la letra S. Entonces,

$$(2) \qquad r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{S_{xy}}{S_x S_y},$$

donde $S_x = \sqrt{s_{xx}}$, $S_y = \sqrt{s_{yy}}$. Si a un conjunto de n pares ajustamos una recta por el llamado *método de los cuadrados mínimos*, la recta tendrá pendiente

$$(3) \qquad b = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

y ordenada al origen

$$(4) \qquad a = \bar{y} - b\bar{x}.$$

Mantenemos como supuesto que $S_{xx} \neq 0$ y $S_{yy} \neq 0$, lo que significa que los casos en que todas las x son iguales, o todas las y iguales, o ambos, están descartados. Bajo estas condiciones, está claro que r y b sólo pueden ser iguales a 0 en un caso, esto es cuando $S_{xy}=0$.

Nótese que S_{xx} y S_{yy} (o bien S_x y S_y que están, en valor absoluto, en las mismas unidades de las variables) miden la variabilidad, la *variabilidad marginal*, presente en las respectivas variables, mientras que S_{xy} mide la *variabilidad conjunta* de los

pares (x,y) y tiene como unidades el producto de las dos variables consideradas. Utilizando divisores adecuados, constituyen medidas estadísticas utilizadas frecuentemente: $S_x^2=S_{xx}/(n-1)$ y $S_y^2=S_{yy}/(n-1)$ son las *varianzas muestrales*, sus raíces cuadradas (no-negativas) con las *desviaciones estándares muestrales* y $S_{xy}=S_{xy}/(n-1)$ [o S_{xy}/n] es la *covarianza muestral*.

Propiedades de invariancia

Es importante analizar cuestiones de *invariancia*. Primero consideremos el efecto de *traslaciones*, que convierten a x en x+p, a y en y+q, o a ambos simultáneamente. Estas traslaciones afectan de la misma manera a los promedios, esto es, la media de los valores trasladados x+p es $\bar{x}+p$, y la de los valores trasladados y+q es $\bar{y}+q$. Por lo tanto, los *desvíos* $x_i-\bar{x}$ e $y_i-\bar{y}$ no son afectados por las traslaciones. Dado que r y b fueron definidos en (1) y (3) en términos de sumas de estos desvíos, concluimos que *r y b son invariantes a las traslaciones*.

Sean ahora c y d constantes positivas. Dado que

$$(5)\quad \sum_{i=1}^{n}(cx_i-c\bar{x})^2 = c^2\sum_{i=1}^{n}(x_i-\bar{x})^2 = c^2 S_{xx},$$

$$\sum_{i=1}^{n}(dy_i-d\bar{y})^2 = d^2\sum_{i=1}^{n}(y_i-\bar{y})^2 = d^2 S_{yy},$$

$$\sum_{i=1}^{n}(cx_i-c\bar{x})(dy_i-d\bar{y}) = cd\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = cdS_{xy},$$

se deduce que el coeficiente de correlación de los pares (cx_i, dy_i) es $cdS_{xy}/(cS_x dS_y)=r$, mientras que la pendiente de la recta computada por el método de cuadrados mínimos se toma el valor $cdS_{xy}/(c^2S_{xx})=(d/c)b$.

La conclusión es que r es *invariante a las transformaciones (positivas) de escala*, mientras que b resulta afectado de la manera indicada. Una manera interesante de enfatizar estos resultados es notar que en el análisis de correlación para diagramas de dispersión como los presentados en la Sección I, no es necesario referir las observaciones a un sistema de coordenadas, pues las traslaciones y los cambios (positivos) de escala no tienen efecto sobre el coeficiente de correlación. Cuando se

estudian las rectas por cuadrados mínimos, se deben considerar o trazar los ejes coordenados, para recordar que las elecciones de escala son importantes.

Las restricciones c>0 y d>0 son importantes: si una de estas constantes es positiva y la otra negativa, los signos de r y b cambian, mientras que si ambas son positivas o ambas negativas, r y b no cambian de valor ni de signo.

3. Cómputos

Para analizar con más detalle los ejemplos de la Sección I, presentamos a continuación ejemplos numéricos. El diagrama a la izquierda del Gráfico 4 (o del Gráfico 5) puede considerarse generado por los pares (z_i, y_i) de la Tabla 1.

Tabla 1. Ejemplo numérico con 5 puntos bidimensionales.

1 Abscisas z_i	2 Ordenadas Y_i	3 Abscisas Centradas $x_i=z_i-\bar{z}$	4 Productos x_iy_i	5 Cuadrados x^2_i
1	1	-2.5	-2.5	6.25
2	-1	-1.5	1.5	2.25
3	1	0	0	0
4	-1	1.5	-1.5	2.25
5	1	2.5	2.5	6.25
Suma = 15	1	0	0	17.00

Nótese que la última fila contiene las sumas de las columnas.

Teniendo en cuenta la invariancia a las traslaciones, operamos con $x_i = z_i - \bar{z}$ en vez de hacerlo con z_i. Como la suma de los x es igual a 0, también lo es su promedio. Dado que los y son muy simples operamos sin modificarlos, siendo su promedio 1/5. Tenemos que

$$(6)\qquad S_{xy} = \sum_{i=1}^{5}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{5}(x_i-\bar{x})y_i = \sum_{i=1}^{5}x_iy_i = 0,$$

la segunda igualdad se debe a que $\sum(x_i-\bar{x})\bar{y} = \bar{y}\sum(x_i-\bar{x}) = 0$, y la

tercera a que $\bar{x} = 0$, lo que conduce a que $S_{xy}=\sum x_i y_i$. También obtenemos $S_{xx}=17$. En conclusión,

$$(7) \qquad r=\frac{S_{xy}}{S_x S_y}=0, \quad b=\frac{S_{xy}}{S_{xx}}=0,$$

como se dijo. Sin embargo consideremos ahora el diagrama a la derecha del Gráfico 5 (o el segundo contando desde la izquierda en el Gráfico 4); los datos aparecen en la Tabla 2.

Tabla 2. Ejemplo numérico con 6 puntos bidimensionales

1 Abscisas z_i	2 Ordenadas y_i	3 Abscisas Centradas $x_i=z_i-\bar{z}$	4 Productos $x_i y_i$	5 Cuadrados x^2_i
1	1	-2.5	-2.5	6.25
2	-1	-1.5	1.5	2.25
3	1	-0.5	-0.5	0.25
4	-1	0.5	-0.5	0.25
5	1	1.5	1.5	2.25
6	-1	2.5	-2.5	6.25
Suma 21	0	0	-3.0	17.50

Ahora calculamos

$$(8) \qquad r=\frac{S_{xy}}{S_x S_y}=\frac{-3}{\sqrt{17.50 x 6}}=-0.29, \qquad b=\frac{S_{xy}}{S_{xx}}=-0.17,$$

mientras que $a = \bar{y} - b\bar{x} = 0$. Diagramas correspondientes a las tablas 1 y 2 forman el Gráfico 5.

estudian las rectas por cuadrados mínimos, se deben considerar o trazar los ejes coordenados, para recordar que las elecciones de escala son importantes.

Las restricciones c>0 y d>0 son importantes: si una de estas constantes es positiva y la otra negativa, los signos de r y b cambian, mientras que si ambas son positivas o ambas negativas, r y b no cambian de valor ni de signo.

3. Cómputos

Para analizar con más detalle los ejemplos de la Sección I, presentamos a continuación ejemplos numéricos. El diagrama a la izquierda del Gráfico 4 (o del Gráfico 5) puede considerarse generado por los pares (z_i, y_i) de la Tabla 1.

Tabla 1. Ejemplo numérico con 5 puntos bidimensionales.

1 Abscisas z_i	2 Ordenadas Y_i	3 Abscisas Centradas $x_i=z_i-\bar{z}$	4 Productos $x_i y_i$	5 Cuadrados x^2_i
1	1	-2.5	-2.5	6.25
2	-1	-1.5	1.5	2.25
3	1	0	0	0
4	-1	1.5	-1.5	2.25
5	1	2.5	2.5	6.25
Suma = 15	1	0	0	17.00

Nótese que la última fila contiene las sumas de las columnas.

Teniendo en cuenta la invariancia a las traslaciones, operamos con $x_i = z_i - \bar{z}$ en vez de hacerlo con z_i. Como la suma de los x es igual a 0, también lo es su promedio. Dado que los y son muy simples operamos sin modificarlos, siendo su promedio 1/5. Tenemos que

$$(6) \qquad S_{xy}=\sum_{i=1}^{5}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{5}(x_i-\bar{x})y_i=\sum_{i=1}^{5}x_i y_i=0,$$

la segunda igualdad se debe a que $\sum(x_i-\bar{x})\bar{y}=\bar{y}\sum(x_i-\bar{x})=0$, y la

tercera a que $\bar{x} = 0$, lo que conduce a que $S_{xy}=\sum x_i y_i$. También obtenemos $S_{xx}=17$. En conclusión,

$$(7) \qquad r=\frac{S_{xy}}{S_x S_y}=0, \quad b=\frac{S_{xy}}{S_{xx}}=0,$$

como se dijo. Sin embargo consideremos ahora el diagrama a la derecha del Gráfico 5 (o el segundo contando desde la izquierda en el Gráfico 4); los datos aparecen en la Tabla 2.

Tabla 2. Ejemplo numérico con 6 puntos bidimensionales

1 Abscisas z_i	2 Ordenadas y_i	3 Abscisas Centradas $x_i=z_i-\bar{z}$	4 Productos $x_i y_i$	5 Cuadrados x^2_i
1	1	-2.5	-2.5	6.25
2	-1	-1.5	1.5	2.25
3	1	-0.5	-0.5	0.25
4	-1	0.5	-0.5	0.25
5	1	1.5	1.5	2.25
6	-1	2.5	-2.5	6.25
Suma 21	0	0	-3.0	17.50

Ahora calculamos

$$(8) \qquad r=\frac{S_{xy}}{S_x S_y}=\frac{-3}{\sqrt{17.50 x 6}}=-0.29, \quad b=\frac{S_{xy}}{S_{xx}}=-0.17,$$

mientras que $a = \bar{y} - b\bar{x} = 0$. Diagramas correspondientes a las tablas 1 y 2 forman el Gráfico 5.

A medida que aumenta la cantidad de pares de datos del tipo que estamos analizando, disminuyen b y r. Por ejemplo, podemos controlar que para conjuntos de pares como los de las tablas 1 y 2, se cumple lo siguiente:

n=4 pares, b=-0.4, r=-0.44,
n=6 pares, b=-0.17, r=-0.29 (como se vio en (8))
n=50 pares, b=-0.0024, r=-0.00346.

4. Diagramas de dispersión simétricos

Definición Operativa. Un diagrama de dispersión con n puntos o pares bidimensionales (x_i , y_i) (donde $n\geq 1$) se dice que es *simétrico* si: (a) contiene n_1 puntos $(0\leq n_1\leq n)$ para los cuales $(x_i-\bar{x})(y_i-\bar{y})=0$; y (b) los restantes $2n_2$ puntos (donde $0\leq n_2\leq n/2$) aparecen en pares, (x_j, y_j) y (x_k, y_k) en los que $(x_j-\bar{x})(y_j-\bar{y})=-(x_k-\bar{x})(y_k-\bar{y})\neq 0$.

Los n_1 puntos cuyos productos cruzados son iguales a 0 pertenecen a las líneas $x = \bar{x}$ o $y = \bar{y}$, o a ambas si ellos coinciden con el par de promedios $(\bar{x}, \bar{y})$. Estas líneas son paralelas a los ejes coordenados, o coinciden con ellos.

Una consecuencia de la definición operativa es la siguiente:

Proposición. Un diagrama de dispersión simétrico tiene $S_{xy}=0$, y por lo tanto $r=b=0$.

Esto proviene directamente de la definición de S_{xy} dada, por ejemplo, en (6).

Proposición. Existen diagramas de dispersión que no satisfacen la condición de simetría de la definición operativa y sin embargo tienen $S_{xy}=0$.

Ejemplo. Para n=3, el conjunto de puntos {(-2, 2), (-1, -1), (3, 1)} tiene esta propiedad.

Generación de diagramas de dispersión simétricos

En esta sección analizamos cómo generar diagramas de dispersión simétricos. De

acuerdo con nuestras observaciones sobre invariancia, en los siguientes diagramas omitimos los ejes coordenados. Considere los diagramas siguientes:

(9)

```
X    X X    X X    X X X    X X X X
X     X     X X      X
X     X
X
```

Ellos cumplen con las condiciones de la definición operativa y son por lo tanto diagramas de dispersión simétricos. Corresponden a las *particiones* de 4 con los enteros 1, 2, 3 y 4,

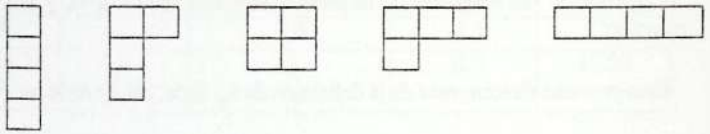
(10) 1+1+1+1 2+1+1 2+2 3+1 4

Euler demostró que el número p(n) de maneras de representar un entero positivo n como la suma de enteros positivos (sin considerar el orden) es igual al coeficiente de q^n en la expansión en serie de potencias del producto infinito

$$(11) \qquad \prod_{k=1}^{\infty} \frac{1}{1-q^k} = 1+q+2q^2+3q^3+5q^4+7q^5+\ldots,$$

(Bressoud and Propp, 1999). Estos autores usan lo que llaman *diagramas de Young* para representar las particiones. En nuestro caso ellos son,

(12)



De los diagramas de dispersión simétricos presentados en (10), podemos generar otros por permutación, con lo que obtenemos lo siguiente:

(13)

```
X    X X    X     X     X X   X X X   X     X X X X
X     X     X X   X     X X     X     X X X
X      X      X    X X
X
```

1+1+1+1 2+1+1 1+2+1 1+1+2 2+2 3+1 1+3 4

Proposición. La cantidad s(n) de diagramas de dispersión simétricos generados por las permutaciones de las particiones es igual a 2^{n-1}.

Esto es válido por ser s(n) la cantidad de soluciones de las ecuaciones

$$(14) \qquad x_1+x_2+\ldots+x_k = n$$

para k=1,2,…,n, cuando las soluciones son enteros de 1 a n. Para cada k la cantidad de soluciones es $\binom{n-1}{k-1}$ (Niven, 1965, Capítulo 5). Por lo tanto

$$(15) \qquad s(n) = \sum_{k=1}^{n} \binom{n-1}{k-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} = 2^{n-1}.$$

Observamos que algunos (pero no todos) los diagramas de dispersión de (13) pueden escribirse con orientaciones distintas, y todavía retener la propiedad de simetría. En efecto, tenemos lo siguiente:

```
X X        X        X X X
 X         X          X
 X       X X

 X          X         X
  X X     X X        X X
X            X       X

2+1+1     1+1+2     3+1
```

Sin embargo no podemos hacer algo semejante con 1+2+1 o con 2+2.

5. Conclusiones

Hemos analizado algunas propiedades de la correlación. El elemento básico de nuestro enfoque es el *diagrama de dispersión* de un conjunto de pares o puntos $P_i=(x_i, y_i)$ en el espacio bidimensional.

La correlación se mide con relación a rectas: si los pares están en una curva, aún siendo la relación (no lineal) aparente, su correlación (lineal) puede ser baja, incluso igual a 0, cuando utilizamos el coeficiente de correlación de Pearson definido en (1). Por lo tanto la *correlación es lineal* aún cuando omitimos el calificativo.

Cuando medimos el signo y el grado de la correlación con el coeficiente r de Pearson, como se hace habitualmente en la práctica, las traslaciones y los cambios de escala son irrelevantes, excepto que si multiplicamos a los x con c y a los y con d, y se satisface que cd<0, ocurre un cambio de signos. La conclusión es que los diagramas de dispersión, cuando se analiza la correlación, pueden dibujarse sin los ejes coordenados. Desde este punto de vista la correlación no está relacionada con la variabilidad marginal de x o de y por separado, sino que sólo está relacionada con la variabilidad conjunta medida por S_{xy}, la suma de productos de las desviaciones con relación a los promedios. El coeficiente r es S_{xy} estandarizado por el producto de las desviaciones estándares, de manera que sus valores posibles están entre –1 y +1.

En consecuencia hemos analizado un conjunto de conexiones entre la correlación y su medida r, con la variabilidad que presenta el diagrama de dispersión. Hemos enfatizado que ciertas aseveraciones con respecto a r deben relacionarse con el valor de la pendiente b.

A continuación relacionamos correlación con simetría. Encontramos que los diagramas de dispersión simétricos (según una "definición operativa") tienen r=0, pero que la recíproca no es cierta, podemos tener r=0 en diagramas de dispersión no simétricos. También analizamos ejemplos de diagramas de dispersión que se alejan levemente de la simetría.

Finalmente desarrollamos un procedimiento simple para generar diagramas de dispersión simétricos. Este procedimiento revela una relación interesante con la teoría matemática de los números (enteros): los diagramas de dispersión simétricos están relacionados con particiones de enteros como sumas de enteros positivos, y a soluciones en enteros de ciertas ecuaciones. Ilustramos el procedimiento sistemático para generar 2^{n-1} diagramas de dispersión simétricos de n puntos. Dos de ellos, correspondientes a la partición de un entero n como 1+1+...+1 o bien como n, tienen $S_{xx}=0$ y $S_{yy}=0$, respectivamente, de manera que r no está definido para ellos; todos los restantes tienen $S_{xy}= r =0$.

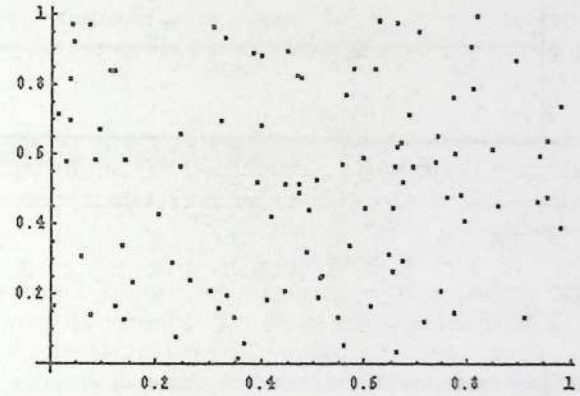
Bibliografía

Bressoud, David and James Propp (1999), How the alternating sign matrix Conjecture was solved, *Notices of the American Mathematical Society*, Vol. 46, No. 6, pp. 637-646.

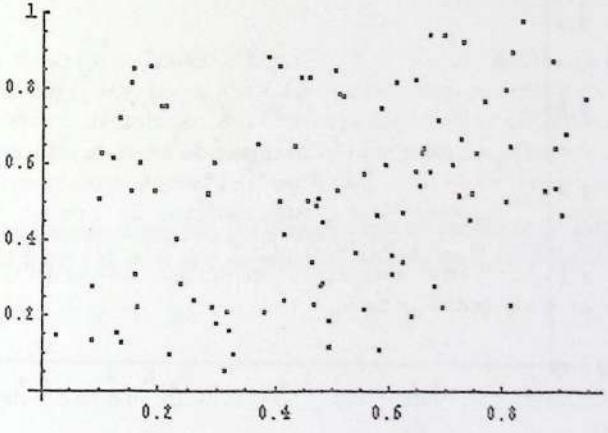
Niven, Ivan (1965), *Mathematics of Choice*. New York: Random House, the L. W. Singer Company.

Grafico 1. Los diagramas de dispersión tienen 100 puntos o pares. El diagrama a) tiene r aproximadamente igual a 0, el diagrama e) tiene r exactamente igual a 1. Los puntos del diagrama a) se proyectan a la recta y = x; b), c), y d) representan pasos sucesivos de la aproximación de cada par (x,y) a su proyección ((x+y)/2,(x+y)/2).

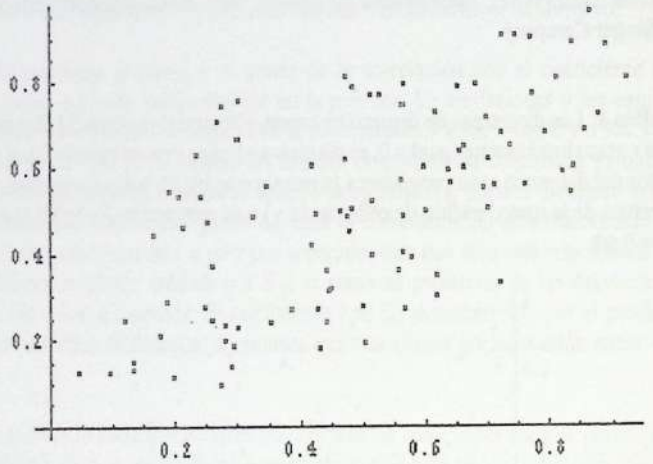
a) r = 0,03



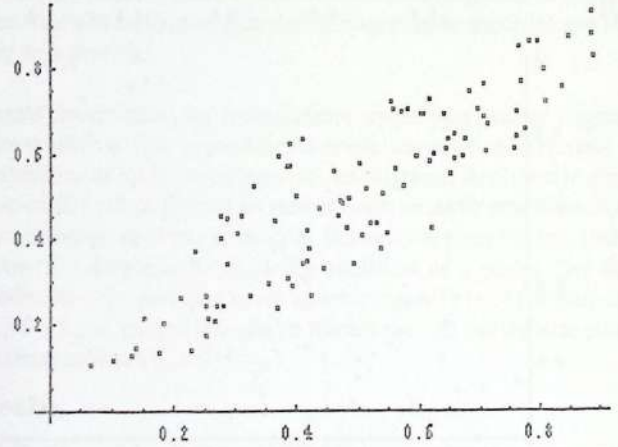
b) r = 0,31



c) r = 0,62



d) r = 0,89



e) r = 1,00

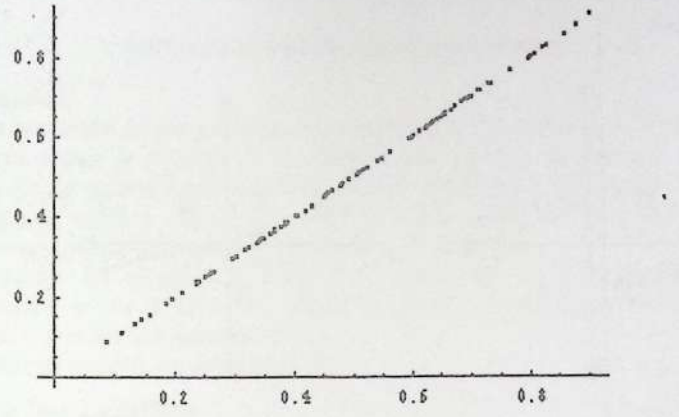


Gráfico 2. Diagramas de dispersión que tienen r=0.

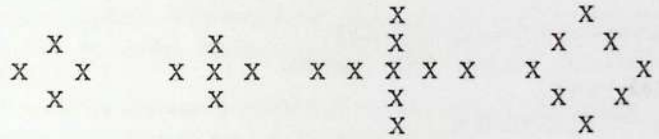


Gráfico 3. Diagramas de dispersión con r=0 deducidos del primer diagrama del Gráfico 1.

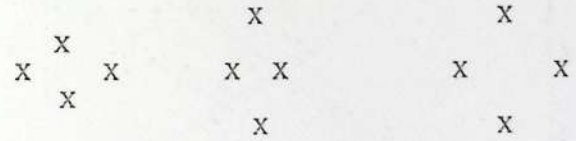


Gráfico 4. Diagramas de dispersión que muestran que un cambio de un solo punt puede afectar el signo y el valor de r, en particular, que sea igual a 0 o que no lo se

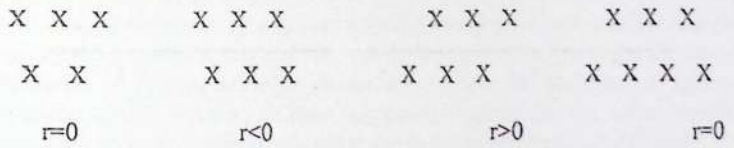
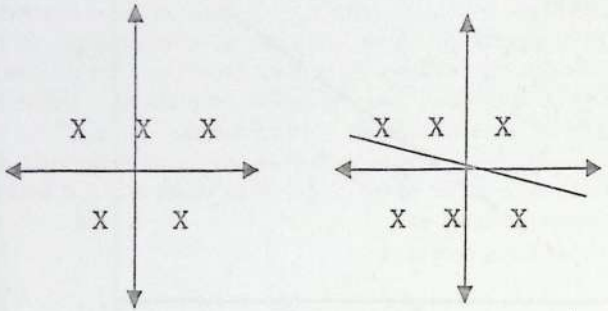


Gráfico 5. Izquierda: Pares de la Tabla 1, r=b=0, la recta por cuadrados mínimos es el eje de las abscisas. Derecha: Pares de la Tabla 2, r=-0.29, b=-0.17 con la recta calculada.



Universidad Nacional deTucumán y CONICET.
Casilla de Correo 209. Tucumán (4000) – Argentina.
Fax 54 (0381) 436-4105.
Email: rmentz@herrera.unt.edu.ar

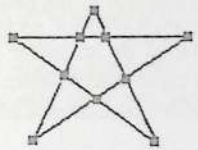
El rey y los castillos

Gamondi Rubén Mario y Cassoli Darío

Introducción:
Con la resolución de este problema se pretende lograr dos objetivos, en el primero se busca captar la atención de los alumnos para resolver un problema sin la utilización de conceptos matemáticos y con mucha utilización de la idea intuitiva y el segundo objetivo consiste en crear el ámbito apropiado para aplicar la herramienta matemática conocida como inducción matemática.
Este trabajo fue propuesto a alumnos de primer año de un profesorado en matemática, y con el correr del tiempo se generó un intercambio que llevó a plantear y resolver una generalización.
El problema original fue extraído de [1].

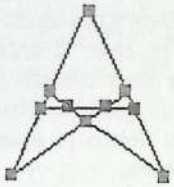
El rey y los castillos

"Érase una vez, en tiempos antiguos, un poderoso rey que tenía ideas excéntricas en materia de arquitectura militar. Sostenía que había gran fuerza y economía en las formas simétricas, y siempre citaba el ejemplo de las abejas, que construyen sus panales en celdas hexagonales, para demostrar que la naturaleza lo respaldaba.
Decidió construir diez nuevos castillos en su país, todos conectados por murallas fortificadas, que debían formar cinco líneas con cuatro castillos en cada línea. El real arquitecto presentó su plano preliminar en la forma

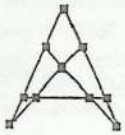


Pero el monarca señaló que era posible atacar cada castillo desde el exterior, y ordenó modificar el plano para que la mayor cantidad posible de castillos quedaran a salvo de un ataque externo, y sólo se pudiera llegar a ellos cruzando las murallas fortificadas. El arquitecto replicó que le parecía imposible disponerlos de tal modo que aún un solo castillo – el edificio que el rey se proponía usar como residencia- se pudiera proteger de ese modo, pero Su Majestad se apresuró a explicarle como hacerlo. ¿Cómo se pueden construir los diez castillos y fortificaciones para satisfacer del mejor modo los requerimientos del rey?.

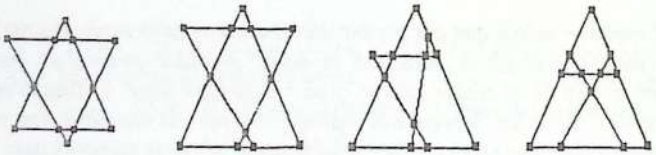
Ante esta situacion se conviene en definir *punto interior* o *castillo interior*, como aquel punto que para poder llegar a él desde el exterior es necesario pasar por lo menos por una línea, y *punto exterior o castillo exterior* aquel que se accede desde el exterior sin pasar por ninguna línea.
Planteado el problema surge la siguiente respuesta por parte del alumnado, la cual coincide con [1].



El docente interviene nuevamente y propone que se analice la posibilidad de defender un castillo. Aparece la siguiente situación:

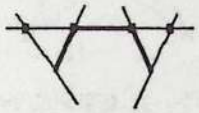


Agotado el problema, el docente propone modificar el enunciado identificando cada castillo con un punto y cada muralla con una línea, donde cada línea tiene únicamente 4 puntos y cada punto, surge de la intersección de sólo dos líneas, este enunciado se analiza para el caso de 6 líneas y 12 puntos, luego se propone a los alumnos que muestren la posibilidad de encontrar cero, uno, dos y tres puntos interiores respectivamente o tanto como sean posibles.
Para cada uno de los casos los gráficos encontrados son:



Surge por parte de los alumnos la observación de la imposibilidad de encontrar una cantidad de puntos interiores que superen a 3, en el caso de 12 puntos y 6 líneas.

Se les propone como actividad siguiente la posibilidad de realizar el mismo problema pero con la variante de 14 puntos y 7 líneas.
Luego de deducido el caso de $n = 7$ se pretende generalizar el problema, el cual consiste en: **trazar un diagrama con solamente puntos exteriores, que contenga $2n$ puntos y n líneas, con la condición que cada línea pase por sólo 4 puntos.**
La respuesta, pronto aparece con la intervención del docente al considerar un polígono (se puede pensar regular) de n lados y prolongar sus lados, se tiene que cada lado finaliza en 2 puntos y la prolongación del mismo hacia ambos sentidos se corta en dos puntos con las respectivas prolongaciones de los 2 lados vecinos.
La gráfica siguiente muestra tres lados vecinos y sus prolongaciones en un polígono de n lados.



Con todas las situaciones propuestas y que quizás han sido resueltas se da una respuesta a la siguiente pregunta:

Si se tienen 2n puntos y n líneas de manera que cada línea contenga 4 puntos, que cada punto pertenezca a la intersección de únicamente dos líneas; y además con las n líneas se cubran los 2n puntos. ¿Cuál es la mayor cantidad de puntos que pueden ser interiores?.

Al hacer el análisis del caso $n = 5$, hemos visto que hay sólo dos puntos interiores.
De la situación $n = 6$, se observa que el máximo posible es de 3 puntos interiores, del caso $n = 7$ resulta un máximo de 5 puntos interiores, luego para $n = 8$ es posible encontrar a lo sumo 7 puntos interiores y para $n = 9$ son 9 los puntos interiores, esta sucesión no es otra cosa que la secuencia de números impares, excepto en el primer caso.
A partir de todo este trabajo manual surge la siguiente conjetura:

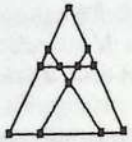
Conjetura: ***La máxima cantidad de puntos interiores que resultan de trazar n líneas que pasen por 2n puntos de manera que cada línea contenga 4 puntos y cada uno de ellos pertenezca a la intersección de dos líneas es:***

$$2 \cdot n - 9$$

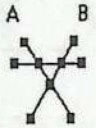
Para poder afirmar o contradecir esta suposición se debería ver cuáles son las herramientas matemáticas disponibles para resolver esta situación, una de ellas consiste en utilizar el método de inducción matemática, la idea inicial es realizar la

solucion en conjunto con el alumnado, analizando uno por uno los distintos valores de n.

En el caso $n = 6$, se ha encontrado el siguiente diagrama:

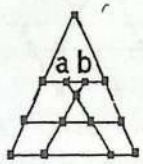


Que no es otra cosa que un triángulo (exterior) en el cual se le ha insertado el siguiente diseño:

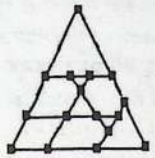


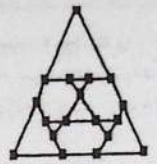
El próximo paso radica en la posibilidad de poder pasar del caso $n = 6$ al caso $n = 7$, para ello es necesario trazar un nuevo segmento y que aparezcan dos nuevos puntos en el esquema, nos bastamos de la siguiente estrategia. En la figura anterior identificamos los puntos A y B:

Luego se traza una línea de manera tal que los puntos A y B no sean los puntos extremos del segmento que formarán con otros dos puntos, de esta manera se obtiene el siguiente esquema:

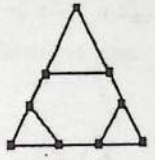


Si se repite el proceso en cada una de las dos puntas restantes del diseño insertado en el triángulo exterior, se tiene la manera de pasar a $n = 8$ y $n = 9$, obteniéndose respectivamente:

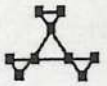




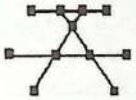
Se puede observar que para $n = 9$ se ha insertado en la figura siguiente

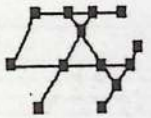


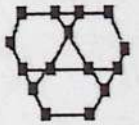
el diseño:



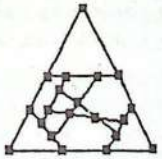
Repasando lo realizado hasta el momento, los diseños que permitieron pasar de 6 hasta 9, haciéndolo por 7 y 8 fueron:

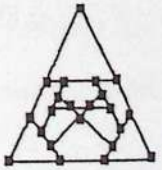


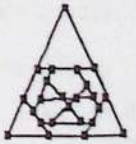




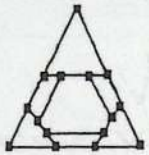
A continuación se muestra los casos de $n = 10, 11$ y 12.



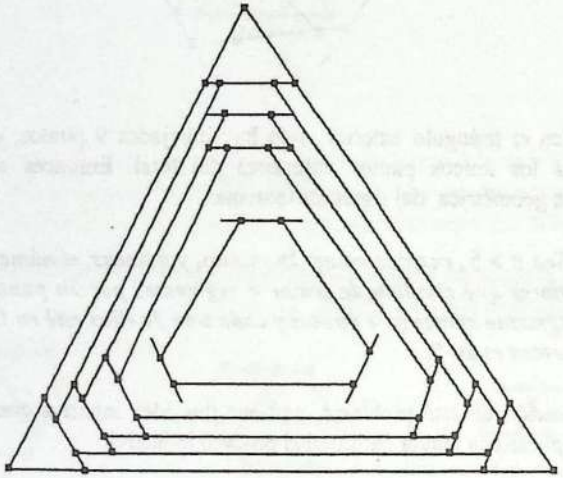




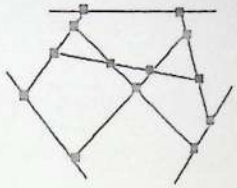
Se nota que los tres últimos casos, tienen en común la figura "exterior" siguiente:

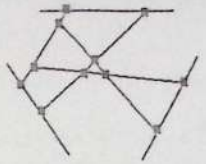


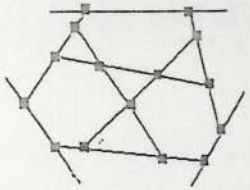
Y los diseños interiores inscriptos son los mismos que los que permitieron el paso de 6 hasta llegar a 9.
Mediante la repetición de este proceso, es posible pasar a diagramas exteriores del tipo:



Que luego con alguno de los diseños siguientes antes elaborados se puede construir el caso que se requiera:







Puesto que en el triángulo exterior, sólo hay dibujados 9 puntos, estos serán en consecuencia los únicos puntos exteriores del total. Entonces se tiene una demostración geométrica del siguiente teorema:

Teorema: ***Sea $n > 5$, consideremos 2n puntos, y n líneas, el número máximo de puntos interiores que resultan de trazar n segmentos por 2n puntos de manera que cada segmento contenga 4 puntos y cada uno de ellos esté en la intersección de dos segmentos es 2n-9.***

Con la resolución de este problema, se tiene una idea intuitiva que creemos que permitirá explicar con mayor facilidad el proceso inductivo.

Bibliografía

[1]-. *El Acertijo del Mandarín y otras diversiones matemáticas*. Henry Dudeney. Juegos & co./Zugarto ediciones. 1992.

[2]-. *Para pensar Mejor.* Guzmán M. Editorial Labor . Barcelona 1991.

[3]-. *Matemáticas, discreta y combinatoria.*Grimaldi, Ralph P. Addison - Wesley Iberoamericana. 1989.

[4]-. *Matemáticas Discretas.* Ross, Kennet A. Wright, Charles. Prentice Hall, Hispanoamericana 1990.

[5]-. *Handbook of Applicable Mathematics. Combinatorics and Geometry.* Jhon Wiley & Sons. 1961.

{*} NUCOMPA – Facultad de Ciencias Exactas. UNICEN.
Tandil (7000). Prov. de Buenos Aires.
rgamondi@exa.unicen.edu.ar

{**} Inst. Sup. Formación Docente N° 27 – Bolívar.

Problemas para resolver

1) Estoy pensando en un número menor que 100. Si el número es múltiplo de 4 entonces está entre 50 y 59 inclusive.

Si no es múltiplo de 6, entonces está entre 20 y 29 inclusive. Si no es múltiplo de 7, entonces está entre 70 y 79 inclusive. Qué número es? (nota: hay tres soluciones posibles, encuentre las tres).

2) Pruebe :

a) $2^n - 1$ es divisible por 3 sí y sólo si n es par.

b) $2^n - 1$ es divisible por 7 sí y sólo si n es múltiplo de 3.

c) $2^n - 1$ es divisible por 15 sí y sólo si n es múltiplo de 4.

d) Enuncie y pruebe la fórmula general.

3) Suponga que se tienen números de la forma $ab0a$, $ee2$, $bb5$, $ccdgcd5d$, donde a, b, c, d, e, g son dígitos entre 0 y 1. Se sabe que:

$$(ab0a) \times (ee2) \times (bb5) = ccdgcd5d$$

y se sabe además que $g = 5 + c$. Halle a y b.

4) En la ciudad de Metrópolis hay alrededor de 20 millones de personas, de las cuales 14.971.389 son adultos. No todos los adultos hablan el mismo idioma, pero se da la situación que, en cualquier grupo de tres metropolitanos adultos al menos dos hablan el mismo idioma. Cada metropolitano adulto habla entre 1 y 20 idiomas. Pruebe que hay al menos un idioma que es hablado por al menos 374.285 adultos.

5) Halle todos los primos p tales que $2p - 1$ y $2p + 1$ también sean primos.

Normas para la presentación de artículos

Los artículos que se presenten a esta revista no deben haber sido publicados o estar siendo considerados para su publicación en otra revista.

Cada trabajo deberá ser enviado en forma definitiva con las indicaciones necesarias para su impresión. No se envían pruebas de imprenta a los autores.

Cada artículo debe presentarse por duplicado mecanografiado a doble espacio. Los diagramas deben dibujarse perfectamente. Los símbolos manuscritos deben ser claramente legibles. Salvo en la primera página, deben evitarse en lo posible notas al pie.

La recepción de cada trabajo se comunicará a vuelta de correo y en su oportunidad la aceptación del mismo para su publicación.

Dirección de e-mail:
revm@mate.uncor.edu

Los trabajos deben enviarse a:
Revista de Educación Matemática
Facultad de Matemática, Astronomía y Física
Medina Allende y Haya de la Torre
Ciudad Universitaria - 5000 - Córdoba

Nota sobre la asignación de fechas a los picos y valles de una serie de lluvias mensuales

Nora M. Jarma y Raúl P. Mentz

Facultad de Ciencias Económicas
Universidad Nacional de Tucumán
CONICET

Resumen

En este trabajo se estudia el efecto que tiene sobre la asignación de fechas a los ciclos de corto plazo de una serie de lluvias mensuales, la selección del tratamiento destinado a estimar el componente de tendencia. Se analizan tres maneras de estimar la tendencia, dos que provienen directamente del programa X-12-ARIMA de ajuste estacional, y otro que surge de recomendaciones aparecidas recientemente en la literatura. Los resultados muestran la importancia de seleccionar bien el método de estimación de la tendencia y coinciden con lo que se encontró en otro estudio al analizar series de indicadores económicos mensuales y trimestrales, que tienen (en general) menos variabilidad que las series de lluvias.

Palabras Claves

Tendencia, estacionalidad, X-12-ARIMA, ciclos, picos, valles, suavizado.

1. Introducción

En un trabajo reciente (Jarma y Mentz, 2000) se consideró el tema de la asignación de fechas a los picos y valles de un conjunto de indicadores económicos argentinos. Se puso especial atención a la selección de opciones en el programa X-12-ARIMA y al suavizado de las series ajustadas por estacionalidad.

El propósito de esta nota es aplicar procedimientos estadísticos similares a una serie de lluvia. La serie con datos hasta 1995 fue analizada en Mentz *et al* (2000) y se encontró que presentaba mucha variabilidad, mas que las series económicas citadas.

2. Datos

La serie considerada es la de los totales mensuales de lluvia caída en la localidad de La Cocha, Provincia de Tucumán. Los datos fueron captados por un productor agropecuario particular. Se dispone de información desde enero de 1982 hasta diciembre de 2002. En Mentz *et al* (2000) la serie fue estudiada hasta diciembre de 1995.

3. Métodos

3.1. Ajuste Estacional

El ajuste estacional se hizo con el programa X-12-ARIMA (Release Version 0.2.10 del año 2002). Este programa es una versión moderna (Findley *et al*, 1998) del método X-11 desarrollado por el Bureau of the Census de los Estados Unidos (Shiskin et al, 1967). Las opciones básicas en nuestro caso son que se usó la versión aditiva del modelo estacional, y se requirió la selección automática de un modelo ARIMA de los disponibles en el programa.

El modelo que se considera para los datos observados, y_t, es aditivo en los componentes inobservables, y puede escribirse como $y_t = t_t + C_t + S_t + I_t$, donde los componentes son, respectivamente la tendencia, los ciclos, el componente estacional y el irregular. Sin embargo los programas de la familia X-11 no distinguen entre tendencia y ciclo, de manera que el modelo efectivamente utilizado es $y_t = C_t + S_t + I_t$, donde ahora C es el «componente tendencia-ciclo». Este modelo se aplica a los datos observados en los meses t = 1,2,...,T; además se dispone de extrapolaciones obtenidas al ajustar a la serie un modelo ARIMA estacional, lo que permite extender la serie por un año mas.

El resultado del procesamiento con el programa X-12-ARIMA es un conjunto de tablas y gráficos. Un primer resumen lo constituye la Tabla 1. Luego tenemos las siguientes series: (1) La serie observada, yt que se presenta en el Gráfico 1; (2) La serie de estimaciones estacionales, Tabla D10 del programa, corresponde a nuestro Gráfico 2. (3) La serie corregida o ajustada por estacionalidad, Tabla D11 del programa, corresponde a nuestro Gráfico 3. (4) La serie corregida por estacionalidad y suavizada,

Tabla D12 del programa, corresponde a nuestro Gráfico 4. finalmente (5) El Gráfico 5 contiene la versión designada como D12* mas abajo, que es una versión mas suavizada aún, de la serie corregida por estacionalidad. En el análisis se usa también la Tabla E2 (ver abajo), que no se presenta por razones de simplicidad.

En los gráficos la escala horizontal está expresada en meses y la vertical en milímetros, que corresponden a un litro de agua por metro cuadrado de superficie.

El resultado del análisis cíclico se muestra en algunos de los gráficos (3, 4 y 5) y en las tablas 2, 3, 4 y 5. Esto será explicado mas abajo.

A pesar de este suavizado se sostiene (Dagum, 1996) que es aconsejable profundizar la reducción del efecto del componente irregular. Este es el procedimiento utilizado en esta nota, lo identificamos como D12* y opera de la siguiente manera: (1) Además de las tablas D11 y D12, el programa X-12-ARIMA emite la Tabla E2, "Serie Ajustada por Estacionalidad Modificada", en la que los valores extremos ("outliers" en inglés) han sido reemplazados por otros valores generados en el programa. (2) La serie de la Tabla E2 es extendida por medio de un modelo ARIMA estacional, que en nuestro caso es el modelo (0,1,1)(0,0,1)12 recomendado por la autora. (3) Usando las opciones provistas por el programa, la serie no recibe ajuste por estacionalidad y se extrapola usando el modelo ARIMA. (4) A la serie resultante se le aplica un promedio móvil de Henderson de 23 meses. (5) En la serie resultante se reemplaza nuevamente los valores extremos, y, finalmente, (6) El resultado deseado se lee en la Tabla D12.

En la Tabla 1 damos algunos de los principales estadísticos descriptivos de la bondad del ajuste. Para comparar presentamos los del período 1982-1995 con los del actual, 1982-2002. Algunas observaciones son: (1) El estadístico F de la "estacionalidad estable" describe cuan importante es el componente estacional de la serie; el F de la "estacionalidad móvil" describe si debe asignarse importancia a los cambios en la estacionalidad durante el lapso considerado. Los valores observados en el período 1982-2002 son, respectivamente, 26.82 y 1.42. (2) Sobre la base de estos dos estadísticos F se califica al ajuste estacional como "estable al 0.1%". (3) MDC es el "mes de dominancia cíclica", definido como la cantidad de meses de variación del componente tendencia-ciclo que hacen falta para compensar la variabilidad del componente irregular; el valor 12 es el mas desfavorable. El "cociente I/C" relaciona la variabilidad del irregular estimado con la del componente tendencia-ciclo estimado. (4) Los estadísticos M son medidas de la bondad del ajuste y son 11 en total. En la serie 1982-2002 hay cuatro que tienen valores desfavorables. M1 mide la contribución del componente irregular a la varianza total; M2 tiene un significado similar, solo que la varianza total se calcula cuando la serie originaria ha sido convertida en estacionaria. M3 mide la contribución del componente irregular sobre la base de la razón I/C mediante la expresión M3 = (I/C - 1)/2. M5 está basado en MCD. M5 = (MCD' - 0.5)/5, donde MCD' ha sido previamente corregido en base a una interpolación lineal. Es importante destacar que el estadístico M7 no ha resultado en la lista de indicadores desfavorables. M7 es un test combinado de la presencia de estacionalidad identificable. Utilizando los valores de estos estadísticos se computa el de Q, una medida de resumen, que es un promedio ponderado de los estadísticos M y que en nuestro caso conduce a calificar el ajuste como rechazado. Debe advertirse que el uso de los estadísticos M y Q con el modelo aditivo no es del todo aconsejable.

TABLA 1
Algunas Medidas de la Bondad del Ajuste Estacional,
Modelo Aditivo, Programa X-12-ARIMA
Serie de Lluvias Mensuales en La Cocha, Tucumán, en dos Periodos

Años	Estacionalidad Estable (F)	Estacionalidad Móvil (F)	Ajustes	MDC	Cociente I/C	Cantidad de M>1	Detalle	Q	Ajustes
1982-1995	13.18	0.98	Estable al 0.1%	12	4.34	4	M1.M2. M3.M5	1.31	Rechazado
1982-2002	26.82	1.42	Estable al 0.1%	12	4.61	4	M1.M2 M3.M5	1.31	Rechazado

Algunas observaciones son: (1) Aumentaron los dos estadisticos F, es decir hay un componente estacional mas importante y tiene mayor variabilidad a través de los años, confirmando lo que muestra los gráficos que se verán mas adelante. (2) Hay pocos cambios en las otras medidas.

3.2. Análisis Cíclico

Las series identificadas como D11, D12 y D12*, que representan a los valores corregidos por estacionalidad con distintos grados de suavizado, ingresan al programa de asignación de fechas a los picos y valles. Esto lo ejecuta un programa desarrollado originariamente en el National Bureau of Economic Research de los Estados Unidos y tuvimos acceso a él a través del Center for Internationoal Business Research de la Universidad de Columbia (Moore, 1989).

El resultado de la aplicación de este método se presenta en forma tabular y gráfica.

4. Análisis de los resultados

4.1. Presentación Gráfica

La información disponible para el período 1982-2002 se presenta inicialmente en 5 gráficos, como sigue:

Gráfico 1. Serie originaria. Obsérvese la presencia de ceros, en algunos casos por lapsos de varios meses. Estos determina que no podamos usar directamente el ajuste multiplicativo. Las opciones del programa son el modelo aditivo y uno llamado "pseudo-aditivo"; los ensayos que realizamos con este último no fueron satisfactorios.

Gráfico 1
Lluvias en La Cocha, Tucumán, Argentina. Serie originaria
Enero 1982-Diciembre 2002.

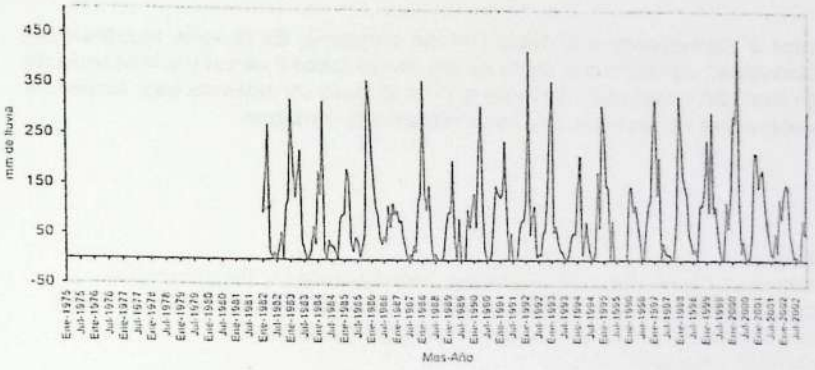


Gráfico 2. Corresponde a la Tabla D10 de la salida del programa X-12-ARIMA. Observamos cambios suaves en la amplitud del componente estacional estimado. Nótese el cambio a partir de 1992. En el estudio previo (Mentz *et al*, 2000) no se observó este tipo de movimiento.

Gráfico 2
Lluvias en La Cocha, Tucumán, Argentina. Factores Estacionales,
Tabla D10, Programa X-12-ARIMA, Método de Ajuste Aditivo.
Enero 1982-Diciembre 2002.

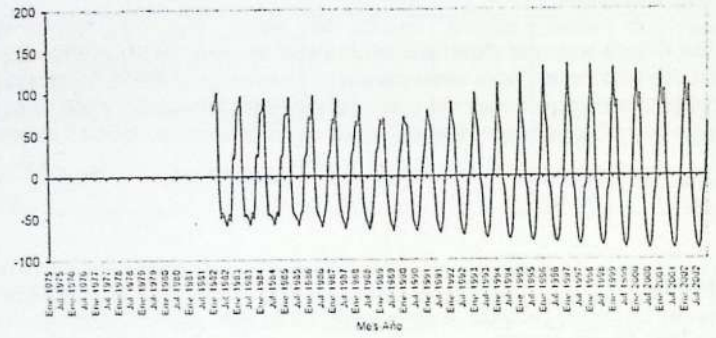
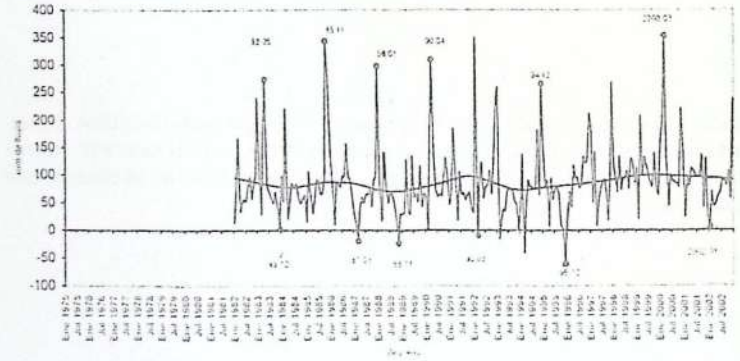


Gráfico 3. Corresponde a la Tabla D11 del programa. Es la serie ajustada por estacionalidad, con las fechas de los puntos de giro (picos y valles) y la tendencia de largo plazo del crecimiento. Notamos que en el lapso considerado esta tendencia presenta pocas oscilaciones, y es aproximadamente constante.

Gráfico 3
Lluvias en La Cocha, Tucumán, Argentina. Serie Ajustada por Estacionalidad,
Tabla D11, Programa X-12-ARIMA, Método de Ajuste Aditivo, Tendencia de Largo Plazo y Puntos de Giro Determinados por el Programa del NBER.
Enero 1982-Diciembre 2002.



4. Análisis de los resultados

4.1. Presentación Gráfica

La información disponible para el período 1982-2002 se presenta inicialmente en 5 gráficos, como sigue.

Gráfico 1. Serie originaria. Obsérvese la presencia de ceros, en algunos casos por lapsos de varios meses. Estos determina que no podamos usar directamente el ajuste multiplicativo. Las opciones del programa son el modelo aditivo y uno llamado "pseudo-aditivo"; los ensayos que realizamos con este último no fueron satisfactorios.

Gráfico 1
Lluvias en La Cocha, Tucumán, Argentina. Serie originaria
Enero 1982-Diciembre 2002.

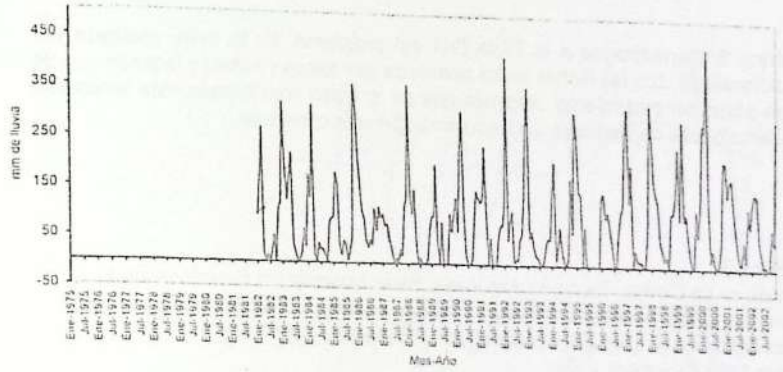


Gráfico 2. Corresponde a la Tabla D10 de la salida del programa X-12-ARIMA. Observamos cambios suaves en la amplitud del componente estacional estimado. Nótese el cambio a partir de 1992. En el estudio previo (Mentz *et al*, 2000) no se observó este tipo de movimiento.

Gráfico 2
Lluvias en La Cocha, Tucumán, Argentina. Factores Estacionales,
Tabla D10, Programa X-12-ARIMA, Método de Ajuste Aditivo.
Enero 1982-Diciembre 2002.

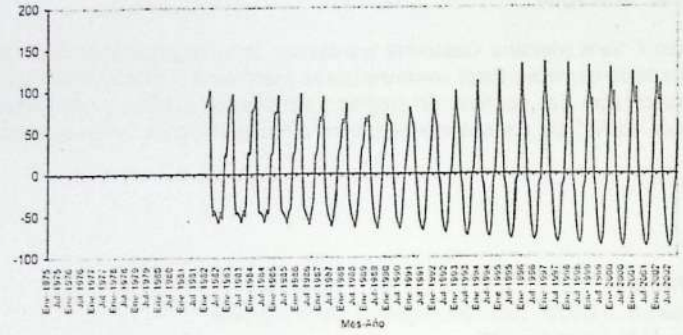
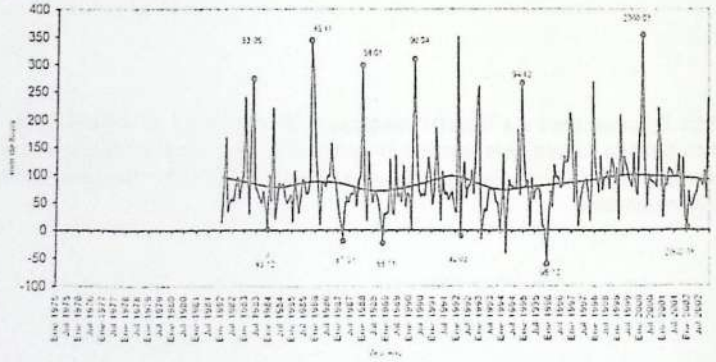


Gráfico 3. Corresponde a la Tabla D11 del programa. Es la serie ajustada por estacionalidad, con las fechas de los puntos de giro (picos y valles) y la tendencia de largo plazo del crecimiento. Notamos que en el lapso considerado esta tendencia presenta pocas oscilaciones, y es aproximadamente constante.

Gráfico 3
Lluvias en La Cocha, Tucumán, Argentina. Serie Ajustada por Estacionalidad,
Tabla D11, Programa X-12-ARIMA, Método de Ajuste Aditivo, Tendencia de Largo
Plazo y Puntos de Giro Determinados por el Programa del NBER.
Enero 1982-Diciembre 2002.



Tendencia de largo plazo. Los gráficos 3, 4 y 5 incluyen una tendencia de largo plazo calculada por los programas del National Bureau of Economic Research. Se trata de un promedio móvil de 75 meses (Moore, 1980). Se llegó a este promedio móvil por un procedimiento de análisis empírico basado en muchas series observadas y se encontró que una alta proporción de los casos el ajuste logrado y la extrapolación practicada dieron resultados razonables.

Gráfico 4. Corresponde a la Tabla D12, componente de tendencia-ciclo estimado, con los puntos de giro y la tendencia de largo plazo del crecimiento. En comparación con el Gráfico 3, esta serie está mas suavizada. Las fecha de los puntos de giro no coinciden exactamente en todos los casos. La tendencia del crecimiento tiene un movimiento a partir, aproximadamente, de fines de 1994, y al final del período tiene una tendencia decreciente.

Gráfico 4
Lluvias en La Cocha, Tucumán, Argentina. Serie Tendencia-Ciclo, Tabla D12,
Programa X-12-ARIMA, Método de Ajuste Aditivo, Tendencia de Largo Plazo y
Puntos de Giro Determinados por el Programa del NBER.
Enero 1982-Diciembre 2002.

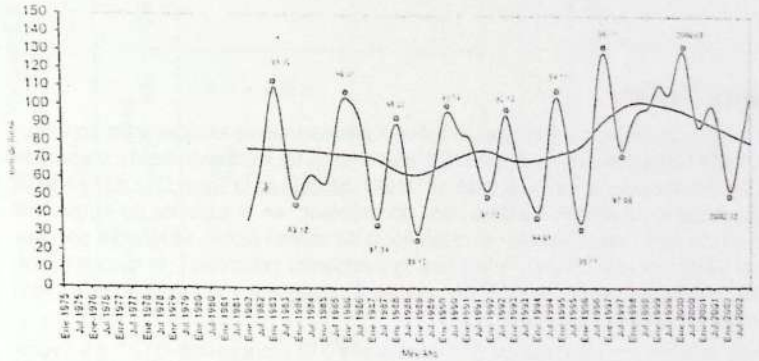
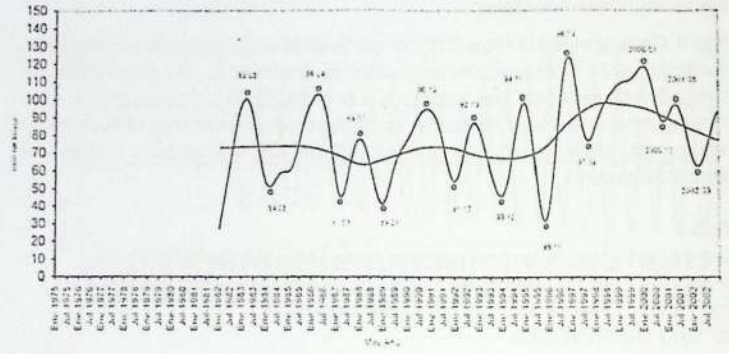


Gráfico 5. Corresponde a lo que hemos identificado como D12*, es decir, una serie mas suavizada que la de la Tabla D12. Hay algunos cambios en las fechas de picos y valles, se observa nuevamente un movimiento a partir de fines de 1994, y al final del período la tendencia de largo plazo señala una disminución importante.

Gráfico 5

Lluvias en La Cocha, Tucumán, Argentina. Serie Tendencia-Ciclo, Tabla D12*, Programa X-12-ARIMA. Tendencia de Largo Plazo y Puntos de Giro Determinados por el Programa NBER. Enero 1982-Diciembre 2002.



4.2. Análisis Tabular

Una manera de resumir la información sobre el procesamiento estacional lo constituye la Tabla 1 presentada en la Sección 3.1 precedente. La correspondiente presentación del fechado constituye en un caso, la Tabla 2, basada en la Tabla D11 de la salida del programa X-12-ARIMA. La tabla tiene dos cuerpos, en el superior se analiza la sucesión de pico - valles - picos y en la inferior la de valles - picos - valles. Se dan las fechas asignadas por los programas (las tres primeras columnas), la duración en meses, y estadísticos de resumen: las medias (aritméticas) y las medianas de las duraciones, y el desvío estándar.

De esta tabla hay una para la salida D11, una para la D12 y otra para la D12*; a su vez, hay un juego de estas tres tablas para el llamado "procesamiento clásico" de los ciclos y otro juego para el "procesamiento de crecimiento".

La determinación de los puntos de giro en los *ciclos clásicos* se realiza en los niveles de la serie observada de lluvias, mientras que en el *análisis de crecimiento* los puntos de giro se determinan con respecto a la tendencia de largo plazo de la serie.

TABLA 2
Lluvias en La Cocha, Tucumán, Argentina. Análisis de los Puntos de Giro, Programa NBER.
Serie Ajustada por Estacionalidad, Tabla D11, Programa X-12-ARIMA
Método de Ajuste Aditivo. Enero 1982-Diciembre 2002

(a) Duracion del Ciclo de Lluvias, *Análisis Pico-Valle-Pico*

PICOS AÑO	PICOS MES	VALLES AÑO	VALLES MES	PICOS AÑO	PICOS MES	DURACION EN MESES DISMINUCION LLUVIAS	DURACION EN MESES AUMENTO LLUVIAS	DURACION EN MESES CICLO TOTAL
1983	5	1983	12	1985	11	7	23	30
1985	11	1987	3	1988	1	16	10	26
1988	1	1988	11	1990	4	10	17	27
1990	4	1992	3	1993	1	23	10	33
1993	1	1994	3	1994	12	14	9	23
1994	12	1995	12	2000	3	12	51	63
2000	3	2002	1			22		
MEDIAS						14.9	20.0	33.7
MEDIANAS						14.0	13.5	28.5
DESVIO ESTANDAR						6.0	16.1	14.9

(b) Duracion del Ciclo de Lluvias, *Análisis Valle-Pico-Valle*

VALLES AÑOS	VALLES MESES	PICOS AÑOS	PICOS MESES	VALLES AÑOS	VALLES MESES	DURACION EN MESES AUMENTO LLUVIAS	DURACION EN MESES DISMINUCION LLUVIAS	DURACION EN MESES CICLO TOTAL
1983	12	1985	11	1987	3	23	16	39
1987	3	1988	1	1988	11	10	10	20
1988	11	1990	4	1992	3	17	23	40
1992	3	1993	1	1994	3	10	14	24
1994	3	1994	12	1995	12	9	12	21
1995	12	2000	3	2002	1	51	22	73
2002	1							
MEDIAS						20.0	16.2	36.2
MEDIANAS						13.5	15.0	31.5
DESVIO ESTANDAR						16.1	5.3	20.1

En el análisis clásico las variaciones pico - valle - pico se interpretan como contracciones y expansiones en el régimen de lluvias, mientras que en el análisis de crecimiento estos conceptos se sustituyen, respectivamente, por los de desaceleración y aceleración en el régimen de lluvias.

Algunos datos aparecen en la Tabla 3.

TABLA 3
Medidas de Posición y Variabilidad del Fechado de los Ciclos
Serie de Lluvias Mensuales en La Cocha, Tucumán, 1982-2002

Tipo	Medias Aritméticas			Medianas			Desvíos Estándares		
	Dismin	Aumen	Total	Dismin	Aumen	Total	Dismin	Aumen	Total
Análisis Pico – Valle- Pico									
D11 C	14.9	20.0	33.7	14.0	13.5	28.5	6.0	16.1	14.8
D12 C	14.5	16.1	29.1	13.5	12.0	26.0	5.6	8.2	5.8
D12* C	13.8	16.9	28.7	11.5	12.0	26.0	5.5	7.5	6.8
D11 Cr	15.0	26.8	40.4	14.0	23.0	30.0	6.5	15.9	17.7
D12 Cr	14.3	16.3	29.3	13.5	12.0	26.0	5.1	8.5	6.1
D12*Cr	16.9	14.0	30.9	12.0	11.0	24.0	7.5	5.9	12.1
Análisis Valle – Pico – Valle									
D11 C	20.0	16.2	36.2	13.5	15.0	31.5	16.1	5.3	20.1
D12 C	16.1	15.1	31.3	12.0	15.0	25.0	8.2	5.7	13.0
D12*C	11.7	15.4	27.3	12.0	11.0	24.5	1.7	8.0	8.2
D11 Cr	26.8	16.6	43.4	23.0	16.0	40.0	15.9	5.8	19.1
D12 Cr	16.3	14.9	31.1	12.0	15.0	25.0	8.5	5.1	12.7
D12*Cr	15.4	11.6	27.0	11.0	11.5	23.5	8.0	1.8	7.7

Clave. (1) Las primeras seis filas de datos corresponden al análisis pico- valle - pico y las restantes al análisis valle - pico - valle. (2) Como medidas de posición se presentan las medias aritméticas y las medianas de las duraciones de los ciclos estimados, separando las fases de disminución de la lluvia de las de aumento. (3) Como medida de variabilidad se presenta el desvío estándar.

Las tablas 4 y 5 resumen la asignación de fechas a los picos y valles de la serie de lluvias en el periodo 1982-2002, para el análisis clásico y de crecimiento, respectivamente. Las columnas (1), (2) y (3) contienen las fechas asignadas. Las columnas (4), (5) y (6) las diferencias en meses entre ellas. Se advierte una cierta coincidencia (diferencias hasta -2 y +2 meses). Analizamos este tema mas adelante.

TABLA 4
Lluvias en La Cocha, Tucumán, Argentina, Fechas de los Pico y Valles
Establecidos para las Tablas D11, D12 y D12*, Programa NBER, Análisis Clásico
Enero 1982-Diciembre 2002

PICOS			Diferencias en meses		
1	2	3	4	5	6
D11 (1)	D12 (2)	D12* (3)	(2) vs. (1)	(3) vs. (1)	(3) vs. (2)
1983/05	1983/02	1983/03	-3	-2	1
1985/11	1986/01	1986/04	2	5	3
1988/01	1988/02	1988/01	1	0	-1
1990/04	1990/04	1990/10	0	6	6
1993/01	1992/10	1992/10	-3	-3	0
1994/12	1994/11	1994/12	-1	0	1
	1996/11	1996/11	-	-	0
2000/03	2000/02	1999/12	-1	-3	-2
				-	-

VALLES			Diferencias en meses		
D11 (1)	D12 (2)	D12* (3)	(2) vs. (1)	(3) vs. (1)	(3) vs. (2)
1983/12	1983/12	1984/03	0	3	3
1987/03	1987/04	1987/03	1	0	-1
1988/11	1988/12	1988/12	1	1	0
1992/03	1991/12	1991/12	-3	-3	0
1994/03	1994/01	1993/12	-2	-3	-1
1995/12	1995/11	1995/11	-1	-1	0
	1997/08	1997/09	-	-	1
2002/01	2002/03	2002/03	2	2	0

3. Duración de los Ciclos. (Tabla 3) Las medias aritméticas varían entre 27.0 y 43.4 meses y las medianas entre 23.5 y 40.0 meses, es decir, aproximadamente, entre 2 y 3 ½ años. Esto está de acuerdo con la observación meteorológica tradicional de que muchas series de este tipo tienen una "oscilación cuasi-bienal" (ver Burroughs, 1992, citado en Mentz et al op. cit.). Observamos que las medianas conducen a estimaciones mas cortas de los ciclos.

4. Efecto del Suavizado en la Duración (Estimada) de los Ciclos. El efecto final del suavizado es reducir la duración estimada.

5. Variabilidad. El efecto del suavizado es importante, sobre todo cuando utilizamos el procedimiento D12*. En algunos casos el desvío estándar estimado con D12* es menos de la mitad del correspondiente a D12. Vale decir que el efecto del suavizado sobre la variabilidad es importante.

6. Fechado de los Picos y Valles. Comparando el fechado bajo D12 con el de D12*, observamos lo siguiente para los meses mas recientes, a partir de 1997: (1) El primer valle está fechado, respectivamente, en 97/08 y 97/09, una diferencia de un mes. (2) El primer pico está fechado en 2000/03 y 2000/01, una diferencia de dos meses. (3) El último valle está fechado en 2002/02 y 2002/03, una diferencia de un mes. (4) El método basado en D12* asigna fechas a un ciclo intermedio, con valle en 2000/10 y pico en 2000/10. Al haber suavizado la tendencia, este movimiento resulta significativo y el programa lo registra.

7. Comparando el Gráfico 4 con el correspondiente al trabajo Mentz *et al.* (2000), Gráfico 5, notamos los siguiente: (1) Las fechas hasta 1991 coinciden; (2) El pico fechado 92/09 pasó a ser 92/10 en el presente trabajo, y el pico 94/12 pasó a ser 94/11; (3) Al final de 1995 la serie mostraba un claro movimiento descendente, en el presente trabajo esto se concretó al asignar fecha 95/11 al valle correspondiente.

8. Conclusión. La principal conclusión referida al suavizado es que el procedimiento D12* es efectivo en lograr un mayor suavizado, al punto de señalar la presencia de un ciclo que de otra manera queda sin identificación explícita.

Agradecimientos: Los autores agradecen al Lic. Antonio Benito por el suministro de los datos y el interés en el trabajo, y al referee por una cuidadosa lectura.

Referencias

DAGUM, E. B. (1996) "A new method to reduce unwanted ripples and revisions in trend-cycle estimates from X-11-ARIMA". *Survey Methodology*, Vol. **22**, No. 1, 77-83.

FINDLEY, D. F., B. C. MONSELL, W. R. BELL, M. C. OTTO AND B-C. Chen (1998) "New capabilities and methods of the X-12-ARIMA seasonal adjustment program, with discussion". *Journal of Business and Economic Statistics*, Vol. **16**, No. 2,127-176.

JARMA, N. M. AND R. P. MENTZ (2000) "Smoothing seasonal time series for cyclical analysis". *Estadística*, Vol. **52**, No. 158-159, 125-159.

MENTZ, R. P., M. A. D'URSO, N. M. JARMA AND G. B. MENTZ (2000) "Cyclical components of local rainfall data". *International Journal of Climatology*, Vol. **20**, 211-228.

MOORE, G. H. (1980) "A long-run look at the business cycle, in Joint Economic Committee", *The business cycle and public policy*, 1929-1980, 96th. Congress, 2nd. Session. Washington, D.C.: Joint Committee Print.

MOORE, G. H. (1989) "The development and use of international economic indicators", in *Statistical Methods for Cyclical and Seasonal Analyses*. R.P. Mentz et al (editors) **63**.84. Panamá, IASI.

SHISKIN, J., A. H. YOUNG and J. C. (1967) The X-11 variant of the census method II seasonal adjustment program. US Bureau of the Census, Washington.

Bootstrap estimation of standard errors in seasonal analysis

R. P. Mentz, N. M. Jarma and C.I. Martínez

Abstract

For the X-12-ARIMA seasonal adjustment procedure, standard errors of year-ahead seasonal estimates are computed by means of the bootstrap method. Direct estimates arise when the original estimated irregular is resampled, AR estimates arise when, before resampling, the correlation structure of an AR model is estimated and incorporated to the bootstrap irregulars. The calculations are done with four series covering various situations. Results are reasonably stable for two of the series with good seasonal fits, and less stable for the other two where fits are less satisfactory. The AR estimates are in general larger than the direct ones, as expected. The bootstrap procedures are suitable to be used as an optional calculation in the seasonal analysis.

Key words: Seasonal analysis, X-12-ARIMA, year-ahead seasonal, standard errors, bootstrap, autoregressive model.

* Endereço para correspondência: Universidad Nacional de Tucumán and CONICET - Casilla de Correo 209, (4000)Tucumán – Argentina, Fax 54-381-4364105, e-mail: mentz@herrera.unt.edu.ar.

1. Introduction

We consider the estimation of the standard errors of certain estimators arising in seasonal analysis. Among seasonal adjustment procedures, of singular importance is the family of X-11 procedures, originated in the USA Bureau of the Census in the early 1960's. It is well known that this procedure is statistically complex, being based upon ideas of smoothing by means of moving average filters, of iteration of sets of operations to which a series is subjected, and other techniques. Pfeffermann (1994) developed formulas for the standard errors of some estimates, by assuming a linear approximation to the seasonal adjustment procedure (Wallis, 1974). His analysis can take into account the sampling error of the survey estimates when relevant. We shall not pursue this interesting line of work in our paper. The bootstrap procedure appears as a reasonable alternative to obtain numerical estimations of some of the desired standard errors.

We use the bootstrap procedure to obtain estimations of standard errors of so-called year-ahead seasonal estimates. These are provided as part of the procedure, and consist of 12 monthly constants for monthly data, and 4 quarterly constants for quarterly data. The importance of these estimates is clear, since they are projections to be used in short term forecasting of the seasonal series.

We designed our study in terms of 4 observed series of different lengths, two arising in economic areas and two in meteorological contexts. The economic series are quarterly Gross National Product (GNP) and monthly Index of Industrial Production (IIP), both of Argentina; these are fairly regular series for which the seasonal procedures provide good fits. The meteorological series are two of those considered in Mentz et al (1999), namely monthly rainfall series for two locations in the State of Tucuman, Argentina, Las Cejas and La Cocha; in relation to the two economic series, these have less satisfactory seasonal fits.

The analysis is presented in two parts. In section 2 the bootstrap procedure is applied directly to the estimated irregular components of the series, which are the residuals of the seasonal fits. In section 3 the nature of these irregular components is analyzed in closer detail, and an alternative approach is used in which an autocorrelation structure is dealt explicitly.

2. Direct estimation of standar errors

2.1. The X-12 ARIMA seasonal adjustment program

The X-12-ARIMA seasonal adjustment program (Findley et al 1998) is a recent variant of the family of X-11 programs introduced in 1965 (Shiskin et al 1967). The mention of ARIMA in the name is due to the fact that for each given series a set of seasonal ARIMA models is automatically considered, so that if one of them provides a good fit (and is the best fit among the models available in the program), then one year of extrapolated data generated by that model is added at the end or beginning and end of the series, previous to the operation of the seasonal adjustment.

The program assumes a multiplicative model, $X_t = C_t S_t I_t$, where X is the observable time series, C is a trend cycle component, S is a seasonal (intraanual) component, and I is an irregular component. For many series a multiplicative model available in the program is preferred; in the case of positive components it can be reduced to the additive case by taking logarithms.

A complex computational procedure is used to produce final estimates $\hat{C}$ and $\hat{S}$. A basic device used in the program is a set of moving average filters, which are symmetric in the case of observations in the middle of the series, and become asymmetric for observations near the ends of the series. Wallis (1974) argued that the combined operation of these filters can be taken (approximately) as represented by a long moving average. The use of ARIMA models for extrapolation will increase the number of months at which symmetric filters can be used.

The program includes a series of options that the user can consider, unless he chooses to use the automatic selections. Besides allowing for additive or multiplicative models and monthly or quarterly data, the program includes a treatment of extreme observations, the possibility of accounting for trading day variation, iterations of the procedures to produce preliminary estimates of the trend-cycle and seasonal components, use of a set of summary (diagnostic) measures, and others.

We concentrate in the results provided for the program, for monthly data, of 12 estimates of the seasonal component that can be used for prediction, and that are labeled *one year ahead seasonal factors*. The objective of this paper is to compute standard errors of these estimates, via the bootstrap procedure.

2.2. The Boostrap procedure

Given an observed time series $x_1,\ldots,x_T$ with monthly observations, where $T=12n$ for simplicity, we assume that this series has been generated by the *unknown probability model* $P=P(C,S,F)$, where F is the distribution function of the irregular component I. We assume that the indicated components generate the series by means of the additive model

$$X_t = C_t + S_t + I_t, \qquad t=1,2,\ldots,T. \tag{1}$$

or by a multiplicative model postulating a product of components.

In our presentation of the bootstrap ideas we follow Efron and Tibshirani (1993), in particular their Chapter 8. The fact that it can be assumed that a model holds for a series which is to be processed by one of the X-11 programs has been frequently considered explicitly in the literature; see, in particular, Cleveland and Tiao (1976).

For the given observed time series, the X-12-ARIMA program produces (final) estimates $\hat{C}$ (Table D12), $\hat{S}$ (Table D10), $\hat{I}$ (Table D13), and one year ahead seasonal factors (Table D10A), that we denote by $\hat{S}_{aj}$, $j=1, 2,\ldots,12$. The estimated $\hat{S}_{aj}$ are standardized so that their mean is 100 for the multiplicative and 0 for the additive model. For further use we note that we operate with centered irregulars, in either the additive or multiplicative situations.

Our objective is to use the available information on C, S and I to estimate the standard error of the $\hat{S}_{aj}$. Formally we consider an *estimated probability model* $\hat{P}=P\left(\hat{C},\hat{S},\hat{F}\right)$ corresponding to P defined above. Here $\hat{F}$ is a discrete distribution obtained from the estimated irregular terms: they define an empirical distribution

function having jumps of size $1/T$ at the estimated values $\hat{I}_1,\ldots,\hat{I}_T$. In the context of this model we let

$$se_F(\hat{S}_{aj}), \qquad j=1,2,\ldots,12, \tag{2}$$

denote the standard errors of the statistics $\hat{S}_{aj}$.

The estimates of the $se_F(\hat{S}_{aj})$ that we propose to compute are denoted by $\hat{se}_B(\hat{S}_{aj})$ (or by $\hat{se}_B$ when there is no risk of confusion), and are calculated by means of the following steps: (a) Let $\hat{I}_1^*,\ldots,\hat{I}_B^*$ be B bootstrap samples of size T from $\hat{F}$. That is to say, each one is a sample of size T selected with replacement from among $\hat{I}_1,\ldots,\hat{I}_T$; (b) With a given set $\hat{I}_b^*$, we associate a bootstrap time series

$$\hat{X}_{tb}^* = \hat{C}_t + \hat{S}_t + \hat{I}_{tb} \qquad t=1,\ldots,T \text{ and } b=1,\ldots,B, \tag{3}$$

where $\hat{C}$ and $\hat{S}$ are the estimates coming from processing the data $x_1,\ldots,x_T$ with the X-12-ARIMA program; (c) With each bootstrap series, we use the X-12-ARIMA program to process it and compute $\hat{S}_{aj}^*(b)$, the corresponding one year ahead seasonals; and (d) We then define the mean of the bootstrap values by

$$\hat{S}_{aj}^*(.) = \frac{1}{B}\sum_{b=1}^{B}\hat{S}_{aj}^*(b) \tag{4}$$

and we have that

$$\hat{se}_B(\hat{S}_{aj}) = \left\{\frac{1}{B-1}\sum_{b=1}^{B}\left[\hat{S}_{aj}^*(b)-\hat{S}_{aj}^*(.)\right]^2\right\}^{1/2}, \quad j=1,2,\ldots,12 \tag{5}$$

are our bootstrap estimates of the standard errors of the $\hat{S}_{aj}$.

Efron and Tibshirani (1993, pages 57-58) also consider robust estimates of standard errors, defined by

$$\widetilde{se}_{B,\alpha}(\hat{S}_{ij}) = \frac{\hat{S}^{*(\alpha)}_{ij} - \hat{S}^{*(1-\alpha)}_{ij}}{2z^{(\alpha)}}, \quad j = 1,2,...12, \tag{6}$$

where $\hat{S}^{+(\alpha)}$ is the 100α% quantile of the bootstrap replications, and $z^{(\alpha)}$ is the 100α% percentile of a standard normal distribution.

2.3. Series used in our study

Since our final results are numerical estimates of the desired standard errors, we payed attention to the design of our calculations, in the sense that a collection of series was used in an attempt to cover several different scenarios. Some of the dichotomies that we considered are economic *versus* other types of series, monthly *versus* quarterly observations, multiplicative *versus* additive models, satisfactory *versus* poor seasonal fits.

The series we analyze are all from Argentine sources: (1) Gross National Product - GNP - at market prices, quarterly observations, 1980-1996, multiplicative model; (2) Index of Industrial Production – IIP, monthly observations, 1980-1996, multiplicative model; (3) Rainfall in Las Cejas, Tucuman, monthly observations, 1975-1989, additive model; and (4) Rainfall in La Cocha, Tucuman, monthly observations, 1982-1997, additive model. The economic series GNP and IIP have very good seasonal fits in terms of X-11-ARIMA, considerable better than the rainfall series.

For further information about the series see the Appendix, where we also report on some of the main results obtained in processing these series with the X-11-ARIMA program.

2.4. Results of the bootstrap procedure

In this section we report on the results obtained in using the bootstrap procedure in the manner indicated in Section 2.2. We used B = 200 in all cases, a number that can be considered satisfactory for our problem (cf Efron and Tibshirani, op.cit.). For some of the examples we compared the results obtained with B = 100, 200 and 300, and noted a great deal of stability.

The present Section 2.4 contains the analysis of what we call the *direct method*, to be distinguished from the autoregressive (AR) procedure developed in Section 3; the two procedures are compared in Section 4. For each series processed by the direct method we present a box plot of the bootstrap replicates (Figures 1, 3, 6 and 8) and a table summarizing in numerical form the main results (Tables 1, 2, 4 and 5); to facilitate the comparison of the two indicated procedures, each one of them contains the corresponding results for the direct and AR procedures. For the direct method we also present histograms, to facilitate the analysis of the symmetry and approximate normality of the empirical distributions (Figures 2, 4, 7 and 9), and the results of a comparison with a very simple regression estimation of the seasonal components and their standard errors (Tables 3 and 6 and Figures 5 and 10).

2.4.1. Economic series

2.4.1.1. Gross National Product - GNP

Figure 1 contains the box plots of the 200 bootstrap replicates of each of the one year ahead seasonal factors for quarters I, II, III and IV. Since the model is multiplicative, the seasonal factors cluster around 1; however, the traditional form of the output in these seasonal programs is to write them as percentages. Hence, our figures are expressed in terms of percentages.

Each box extends from the first to the third quartile, and also shows the median of the replicates. The whiskers have lengths equal to the 1.5 times the difference between the quartiles, and the stars correspond to replicates falling outside these limits. The starred replicates fell within 3 times the quartile differences, and hence can be qualified as distant (but not very distant). As we will see in comparison with the other cases, this is a rather regular diagram.

The same information is presented in Figure 2 by means of histograms. The four histograms have the same scales and the same numbers of classes to facilitate the comparisons. In general the histograms show that the distributions of the bootstrap replicates of the one year ahead seasonal factors do not depart strongly from symmetry

On the basis of the bootstrap estimates, Table 1 presents the analysis corresponding to formulas (4), (5) and (6) of Section 2.2. The first two lines of the table compare the $\hat{S}_{ej}$ coming from the X-12-ARIMA program with the averages $\hat{S}^*_{ej}(.)$ of the bootstrap replicates: we find a great deal of coincidence. The third line contains the bootstrap estimates of the standard errors $s\hat{e}_{200}$ defined in (5), and the next three lines contain, respectively, the robust estimates of the standard errors, $s\tilde{e}_{200,\alpha}$ calculated for α=0,84, 0.90 and 0.95 (we take $z^{(0.84)} = 1$). The differences are small, confirming the idea that the bootstrap distributions are fairly regular, that the influence of *outliers* is small.

2.4.1.2. Index of Industrial Production - IIP

Figures 3 and 4, and Table 2 contain the information corresponding to the IIP, in correspondence with Figures 1 and 2 and Table 1 for the GNP. The box plots are quite regular, with essentially all bootstrap replicates lying within 1.5 times the difference between quartiles. The 12 histograms in Figure 4 do not contradict the idea of approximate normality of these bootstrap distributions. The first two lines of Table 2 show a very good agreement in terms of means, while the differences between the bootstrap estimates of standard errors and the three robust estimates are small.

2.4.1.3. Analysis of the results

1. The two economic time series that we studied, quarterly GNP and monthly IIP, are very well suited to be treated by the X-12-ARIMA program in its multiplicative form. In the Appendix we argue that, in terms of some of the main figures of merit commonly used, the signals coming from the program are all in the indicated direction.

2. The bootstrap procedure works very well, at least to be judged by the close agreement between the one year ahead seasonal factors produced by the X-12-ARIMA program and the mean of the corresponding bootstrap replicates (Lines 1 and 2 of

Tables 1 and 2, respectively). We recall that seasonal factors are expressed (arbitrarily) as percentages.

3. The bootstrap replicates show variability, but by considering the box plots (Figures 1 and 3, respectively) we find that no large *outliers* are present. This is also confirmed by the relation between the two types of estimates of standard errors considered below.

4. The distributions of the bootstrap replicates show reasonable similarities with normal distributions, as seen in the histograms contained in Figures 2 and 4, respectively.

5. Tables 1 and 2 exhibit the numerical values of the bootstrap estimates of the standard errors of the one year ahead seasonal factors. We recall that the seasonal factors are expressed as percentages, so that the standard errors are also multiplied by 100. They are computed according to formula (5), and then according to formula (6) for α = 0.84, 0.90 and 0.95, which are robust estimators. In agreement with the preceding comments, the differences among these various estimates are in general small.

6. In the case of the GNP series, referring to the estimates coming from formula (5), we find that there are differences among the quarters: the smallest standard error is for the first quarter, the fourth quarter is larger, and the second and third quarters have similar standard errors and they are numerically the largest.

Bootstrap estimates of standard errors tend to move in consonance with the size of the corresponding statistic (or of the average over *bootstrap* replications), without being proportional. In fact they represent 0.41%, 0.46%, 0.47% and 0.44%, respectively.

7. In the case of the IIP series, referring to the estimates coming from formula (5), we find a good deal of coincidence among the estimates of standard errors: 7 of the 12 estimates can be rounded to 1.2.

8. While comparison of the $\hat{S}_{aj}$ with the averages $\hat{S}^{*}_{aj}(.)$ is useful to evaluate the performance of the procedure, we do not have a similar benchmark to compare the numerical values of the bootstrap estimates of the standard errors. In view of the very regular structure of the seasonal patterns in our series, we performed the following calculations. To the series adjusted for trend (i.e. the original series divided into the final estimate of the trend cycle component), we fitted regression models without constant, having linear trend terms and dummy seasonal variables. The fits of these models are compared with the X-12-ARIMA estimates in Figure 5, and we see that the approximations are quite reasonable. The vertical scales in Figure 5 are not in percentages.

The advantage of these approximations is that for the regression models we have estimates of the standard errors obtained as part of the inferential part of the regression procedure. The numerical values of the one year ahead seasonal factors and their standard errors are compared in Table 3. The estimates of standard errors coming from the regressions are quite similar among themselves, which is a consequence of the approach (The seasonal dummies are orthogonal). We find that the agreement in the case of GNP is quite good, and better than for the IIP. What is important from our point of view is that the orders of magnitude of the standard errors give reassurance to the bootstrap estimates.

9. We conclude that for the two economic series that we considered, one with quarterly and one with monthly observations, which are acknowledged to be very suitable for the seasonal treatment programmed in X-12-ARIMA under multiplicative models, the bootstrap procedure works very well to produce numerical estimates of the standard errors of the one year ahead seasonal factors which are obtained as a regular part of the output of the program.

2.4.2. Rainfall series

2.4.2.1. Las Cejas

Figure 6 contains the box plots of the 200 bootstrap replicates of each of the one year ahead seasonal factors for the 12 months. The model is now additive, and the seasonal factors are expressed as percentages and centered at 0. The variability and presence of *outliers* is more notorious than in the economic series. In particular we find a great deal of variability (including *outliers* in some cases) in January, March, August, and November. The corresponding histograms in Figure 7 do not show noticeable departures from (approximate) normality.

The main results of the bootstrap procedure are in Table 4. There is some level of agreement between the estimated seasonal factors $\hat{S}_{aj}$ and the averages of the bootstrap replicates $\hat{S}^{*}_{aj}(.)$. In comparing the bootstrap estimates of standard errors corresponding to formula (5) appearing in the third line of the table, with the robust estimates in the fourth, fifth and sixth lines, there are some differences, with the first estimates being larger than the robust versions, as expected. In fact, the differences are larger for most of the months mentioned in the preceding paragraph as exhibiting large variabilities.

2.4.2.2. La Cocha

Figure 8 contains the box plots of the 200 bootstrap replicates of each of the one year ahead seasonal factors for the 12 months. In comparison with the box plots for the two economic series and the rainfall series for Las Cejas, the present one shows considerably more variability, and a great deal of outlying values. Note in particular the small *outliers* in the otherwise rainy month of January, opposite to the large *outliers* in the otherwise dry month of July.

The indicated nature of the bootstrap replicates is also apparent in the histograms of Figure 9: their central parts do not contradict the idea of approximate normality, but the presence of very large or very small values is clear.

These remarks help interpret the bootstrap results in Table 5. There are differences between the original estimates of the seasonal factors and the averages of the bootstrap replicates (Lines 1 and 2 of the Table 5). What is more important for our analysis, there exist large differences between the usual bootstrap estimates of the standard errors computed according to (5), and the robust estimates computed according to (6). For the months of January and July mentioned above, the differences are the largest, the usual estimates are larger than 30, while the robust estimates are around 20. One consequence of this fact, is that the usual estimates range from 20.5 for May to 32.1 for January, while the robust estimates show less variability: They range from 18.5 for October to 24.8 for November when $\alpha = 0.95$, from 18.1 for May to 22.7 for February when $\alpha = 0.90$, and from 16.3 for May to 21.8 for April when $\alpha = 0.84$.

2.4.2.3. Analysis of the results

1. The two monthly rainfall series that we studied, Las Cejas and La Cocha (Province of Tucuman, Argentina) are not particularly well suited for treatment with the X-12-ARIMA program. They are processed under the additive model because they have zero values, corresponding to months without rain. In the Appendix we argue that among the figures of merit commonly used, some indicate that a reasonable fit can be expected, while others reject the quality of the fit.

2. The bootstrap procedure also finds difficulties. For example, there is disagreement between the one year ahead seasonal factors produced by the X-12-ARIMA program, and the mean of the corresponding bootstrap replicates (Lines 1 and 2 of Tables 4 and 5, respectively).

3. The bootstrap replicates show variability, including the presence of large *outliers* (box plots in Figures 6 and 8). This is particularly important in the case of La Cocha, and is also confirmed by the relation between the two types of estimates of standard errors considered below.

4. The distributions of the bootstrap replicates show reasonable similarities with normal distributions in their central portions, but the presence of *outliers* is notorious in the histograms contained in Figures 7 and 9.

5. Tables 4 and 5 exhibit the numerical values of the bootstrap estimates of the standard errors of the one year ahead seasonal factors. They are computed according to formula (5) in the third line, and to formula (6) with $\alpha = 0.84$, 0.90 and 0.95 in the fourth, fifth and sixth lines. The differences among these various estimates are large. It makes an important difference whether we adopt the usual or any of the robust estimates of the standard errors, in particular in the case of La Cocha.

In a similar situation, Efron and Tibshirani (1993, page 69) comment: " The large values of $\hat{se}_{200}$ for [some months] are seen to be caused by a few extreme values of [the bootstrap replicates]. The approximate confidence interval $\theta \in \hat{\theta} \pm Z^{(1-\alpha)}\hat{se}$ will be more accurate with $\hat{se}$ equaling $\tilde{se}_{200,\alpha}$ rather than $\hat{se}_{200}$, at least for moderate values of α like .843."

6. Taking the values in the fourth line of Tables 4 and 5 as the estimates of the standard errors of the one year ahead seasonal factors, we find that they range from 9.8 in April to 13.5 in February for Las Cejas, and from 16.3 in May to 21.8 in April for La Cocha.

7. The regression approximation proposed for the economic series can also be tried here (the model has now coefficients restricted to sum to 0). The results are in Figure 10 and in Table 6. This information is still of some use, in terms of orders of magnitude.

8. We conclude that for the two monthly rainfall series that we considered, which are acknowledged to be hardly suitable for the seasonal treatment programmed by the X-12-ARIMA under additive models, the bootstrap procedure finds some difficulties, and the analysis points out to the convenience of using robust estimates of the standard

$$\Phi_{\hat{p}}(\overline{B}) = 1 - \hat{\phi}_1\overline{B} - \ldots - \hat{\phi}_{\hat{p}}\overline{B}^{\hat{p}}, \tag{8}$$

and $\hat{\varepsilon}_t$ is white noise. We now invert (7) into

$$\hat{I}_t = \Phi_{\hat{p}}^{-1}(\overline{B})\hat{\varepsilon}_t \cong \Pi_c(\overline{B})\hat{\varepsilon}_t, \tag{9}$$

where

$$\Pi_c(\overline{B}) = 1 + \hat{\pi}_1\overline{B} + \ldots + \hat{\pi}_c\overline{B}^c, \tag{10}$$

is an approximation of finite order c to the infinite moving average defined by the inverse operator in (9). The order c is chosen so that $\hat{\pi}_s$ is small for $s > c$.

We now generate *bootstrap* samples of the $\hat{\varepsilon}_t$ that we designate by $\hat{\varepsilon}^*_{tb}$, $b = 1,2,\ldots,B$. With these we generate the bootstrap irregulars,

$$\hat{I}^{**}_{tb} = \Pi_c(\overline{B})\hat{\varepsilon}^*_{tb}, \quad t = 1,2,\ldots,T, \quad b = 1,2,\ldots,B, \tag{11}$$

and finally the bootstrap series

$$\hat{X}^{**}_{tb} = \hat{C}_t + \hat{S}_t + \hat{I}^{**}_{tb}, \quad t = 1,2,\ldots,T, \quad b = 1,2,\ldots,B, \tag{12}$$

We use two asterisks to distinguish from (3) of the direct method.

To the series in (12) we apply the X-12-ARIMA procedure to obtain the desired estimates of the year-ahead seasonals, and in turn their bootstrap estimate of standard errors.

We see that if the autocorrelation structure in the estimate irregular for a given series is important, the AR bootstrap series $\hat{X}^{**}_t$ will differ from the $\hat{X}^*_t$ considered in Section 2, and hence the estimates of the year-ahead seasonals may also differ, and also their estimated standard errors. The magnitude of this effect will be studied by comparing the results obtained by application of the two procedures.

3.3. Results of the empirical study

As indicated in Section 3.2, we start the empirical study by fitting to each time series an AR model of an order estimated from the same data. We used the programs in Brockwell and Davis (1991b), which make preliminar estimations of the parameters by using Burg's or Yule and Walker's procedures, then estimate the order by the AICC procedure, and finally produce the final estimates of the AR($\hat{p}$) model parameters. We found the orders of 6 trimesters for GNP, 7 months for the IIP, 12 months for rainfall in Las Cejas, and 4 months for rainfall in La Cocha. We explored this last series in more detail, and found that similar results were obtained by using order 12 also for this rainfall series. In inverting the AR operator, we considered the way the coefficients decreased as the lag increased, and came out with the uniform suggestion of using lags up to c = 50 in all 4 series. Hence, the irregular in (11) is generated by sums truncated at c = 50, while we use three different values of the autoregressive order.

With these irregulars we generate the bootstrap time series defined in (12), obtain the estimates of the year-ahead seasonals, and their standard errors.

3.3.1. Bootstrap study of the economic series

The results of the bootstrap study for the quarterly GNP and monthly IIP are presented in tables and graphs with the same formats as those discussed in Section 2 for the direct method. The only exception is that we do not have tables like Tables 3 and 6, and figures like Figures 5 and 10, which contain regression analyses of the seasonal patterns.

Tables are 1 for GNP and 2 for IIP, and figures containing box plots are 1 for GNP and 3 for IIP. For the sake of simplicity we do not present the histograms corresponding to the bootstrap results under the AR procedure, since they lead to the same kind of conclusion that in the case of the direct method.

In view of the indicated circumstance we do not describe in detail these tables and figures, and postpone further discussion of these results until Section 4 in which they will be compared with those coming from the direct method.

3.3.2. Bootstrap study of the rainfall series

The numerical results of the bootstrap study appear in Tables 4 for Las Cejas and 5 for La Cocha, and box plots of the bootstrap replicates appear in Figures 6 for Las Cejas and 8 for La Cocha. The same comments of the preceding section apply to this case.

4. Comparing the direct and AR approaches

The comparison of the direct and AR approaches is facilitated by the approach followed in the presentation of the results in many of the tables and figures: in Tables 1, 2, 4 and 5 which summarize the main numerical results, and in Figures 1, 3, 6 and 8 containing the box plots of the bootstrap replicates, the results of the two approaches are presented in the same page.

In Tables 1, 2, 4 and 5 the 4 estimated standard errors are summarized by their medians and standard deviations: medians are used to compare positions, standard deviations to compare variabilities. Variability is related to the appearence of the corresponding box plot diagram, including the possible effect of *outliers*. Position and variability as measured in the tables, are then exhibited graphically in Figures 11 and 12.

GNP is the simplest series to analyze, partly because there are only 4 quarters to consider, and also because the series corresponds to a highly aggregated concept. Box plots for the AR(6) procedure show less variability than for the direct method (Figure1), standard errors $s\hat{e}_{200}$ and medians of the 4 estimates of standard errors are higher for the AR(6) than for the direct method (Table 1), standard deviations of the estimates of se for the AR procedure are equal to or larger than those of the direct method in 3 of the 4 quarters. Hence, the graphical summary in Figure 11 can be interpreted as showing that, compared with the direct method, the AR(6) procedure produces larger estimates of $se\left(\hat{S}_{aj}\right)$, its bootstrap replicates tend to exhibit less variability, and as a consequence its robust estimates $s\tilde{e}_{200,\alpha}$ differ less among themselves and from $s\hat{e}_{200}$.

For monthly IIP, $s\hat{e}_{200}$ under AR(7) is higher than under the direct approach in 11 of the 12 months (the exception is January), and medians are also higher in 11 months

(the exception is June). Hence, the conclusion is that AR estimates of standard errors tend to be higher than those of the direct procedure, which is confirmed by Figure 11. In terms of variability of the 4 estimates of $se\left(\hat{S}_{aj}\right)$, and hence of the effect of the variability of the bootstrap replicates, standard deviations under AR are smaller than those of the direct method only in 6 of the 12 months.

As expected, the results for the rainfall series are less conclusive. We recall that these series have more variability that the economic ones, and that the X-12-ARIMA procedure found more difficulties in its fit to the data. For Las Cejas $s\hat{e}_{200}$ is larger under the AR(12) procedure in 8 of the 12 months, and the median of the 4 estimates of se is higher in 9 months, so that AR estimates of standard errors tend to be higher than those of the direct procedure, which is confirmed by Figure 12. Standard deviations of the 4 estimates of se are smaller for the AR procedure in 8 of the 12 months.

For monthly rainfall in La Cocha, $s\hat{e}_{200}$ is smaller under the AR(12) procedure in the 12 months, while the median of the 4 estimates of se is larger than for the direct approach only in 3 of the 12 months. This should be related to the shape of the box plot in Figure 8, and is also evident by observing Figure 12 and comparing La Cocha with the other 3 series considered in this study. Another consequence of these facts is that the standard deviations of the 4 estimates are smaller for the AR procedure in 9 of the 12 months.

5. Conclusions

This study deals with some estimation problems generated in applying the X-12-ARIMA procedure to time series of monthly or quarterly observations. The X-12-ARIMA procedure assumes that the series can be decomposed multiplicatively or additively into certain components; it also has available an option known as pseudo-additive. In the simplest case, these are the trend-cycle, seasonal and irregular components. A more detailed analysis of the given series may lead to consider other components: in a related resampling study, Findley et al (1990) considered in addition trading day, holiday and

outlier components, and treated them with a module of X-12-ARIMA called Reg-ARIMA (Findley et al 1989).

For any of these models, the program X-12-ARIMA produces estimates of the components and of other related quantities. Since one of the main objectives of seasonal analysis is forecasting, we consider the one year ahead seasonal factors estimated by the program. These are extrapolations of the seasonal component, which are recommended to be used in the analysis of the first year following the period of observation.

For these one year ahead seasonal factors, which consist of 12 estimates for monthly data and 4 estimates for quarterly data, we concentrate in estimating their corresponding standard errors. These are the key quantities needed to appraise the variability of the estimates. However, due to the complicated nature of the X-12-ARIMA procedure, no closed-form either exact or approximate expressions are available for these standard errors.

It then follows that the bootstrap procedure is *a priori* adequate to be used for standard error estimation. In this paper we study two procedures, namely: (a) the direct method, in which resampling is performed with the estimated irregular component estimated in the empirical decomposition, as is usually done in bootstrap studies dealing with (independent) cross-section data; (b) the AR($\hat{p}$) procedure, in which the autocorrelation structure of the estimated residual is studied, and if it is found of importance, incorporated into the resampling procedure.

The differences between the two approaches are found to be important. The explicit consideration of the AR structure leads in general to more concentrated values of the bootstrap replicates, less need to consider robust estimates of the standard errors, and larger values of the estimated standard errors.

The procedures were applied to 4 different time series, two arising in the economic field (Argentine's quarterly GNP and monthly IIP), and two monthly rainfall series. The economic series are analyzed by means of multiplicative models, and the rainfall series by means of additive models because they have zero observed values in some months. The indications coming, when the X-12-ARIMA procedure is applied to these series, are that for the economic series the fit is quite good, while for the rainfall

series has considerably more difficulties. This can be regarded as a strength of our analysis, since it is possible to check the expectation that the bootstrap results will tend to be weaker for the rainfall series.

In view of the preceding considerations, in all cases the automatic (or defaut) options available in the X-12-ARIMA program are used. As indicated in the first paragraph of this section, a more detailed or specific model could be developed for each series. However, the rationale in the application of the bootstrap procedure will not change, since all estimated components will enter unchanged in the formation of the bootstrap replicates of the series, only the treatment of the irregular being subject to analysis.

Two main remarks can be offered in conclusion. One is concerned with the numerical value of the estimated standard errors. In terms of order of magnitude, we can use the summary measures available in Tables 1, 2, 4 and 5. Two possible selections are the medians (over periods) of the $\hat{se}_{200}$ estimates, and the medians (over periods) of the four estimates of se. These results are as follows:

Box 1 – Medians of the four estimates of se and $\hat{se}_{200}$, AR procedure

Series	Number of observations	Estimated AR order	Median of $\hat{se}_{200}$, AR procedure	Median of medians of estimates of se, AR procedure
GNP	68	6	0.518	0.508
IIP	204	7	1.260	1.196
Las Cejas	180	12	13.786	13.073
La Cocha	192	12	21.173	19.644

As for overall consistency, these median values can be compared with the simple regression estimates in Tables 3 and 6, the values above being slightly larger.

The second remark is that a bootstrap procedure similar to that developed here, can be incorporated in the empirical application of programs like X-12-ARIMA, for the purpose of obtaining numerical estimates of the standard errors of interest. In this way they will complement the list of numerical estimates produced by this kind of programs.

Appendix. More information about the series used in this study

In Section 2.3 we introduced the four series studied in this paper. The two economic series (quarterly GNP and monthly IIP) are two standard measures used in economic description and analysis. The two monthly rainfall series are included in a study emphasizing cyclical analysis (Mentz et al, 1999). Brief descriptions follow.

GNP. It is a frequently-used indicator of the global behavior of Argentine's economy, computed and published quarterly. It is a weighted aggregate of economic series representing added value in the production of goods and services. It is often defined as the value (at market prices) of the production done in the economic space of a country, with the participation of residing and non-residing factors of production.

IIP. This weighted index is based upon series of the manufacturing sector of the economy, is produced by a private institution, Fundación FIEL, and comprises 61 products; often 100 is regarded as a desirable number of products. The design is of a longitudinal or panel type. Most of the information is gathered from associations and chambers of producers, which simplifies the procedure and reduces the costs. The selection of products is done according to two basic ideas: (a) that the available series provide good coverages for the selected products; (b) that the series are readily and regularly available.

Rainfall Series. These series are formed by monthly totals of rainfall in the indicated places. Those for Las Cejas were provided by a state-supported agricultural experimental station, while those for La Cocha were collected by a private farmer.

In our applications of the X-12-ARIMA procedure, in general the default or automatic options were used, when available; see our comments in Section 5. All series were processed in their original available scales, without introduction transformations; the possibility of transforming the rainfall series was considered in Mentz et al (1999), and found that they did not imply improvements in the application of the procedures.

Appendix Table 1 contains information about the results of applying the X-12-ARIMA procedure to the four series. The two economic series have figures of merit showing that the fit is quite good. The F statistics for stable seasonality (month effect in a one-way ANOVA) are large, and the F statistics for moving seasonality (year effect

in a two-way ANOVA) are small, so that the seasonal components are judged present at 0.1%.

An important measure is the quarter or month of cyclical dominance, which is defined as the number of periods needed to make equal to 1, the ratio of the estimated irregular component to the estimated trend-cycle component. One quarter for GNP and 3 months for IIP are indication of good fits.

Under M statistics analysis the table reports the results of computing 11 sample quantities computed by the X-12-ARIMA program, meaning different characteristics of the seasonal fit. Each M statistic ranges from 0 to 3, and has acceptance values from 0 to 1. The Q statistic is an overall measure defined as a weighted average of the 11 M statistics. For the two economic series we find that no M statistics is larger than 1, and the Q statistics indicated that the seasonal fit is considered as accepted. A reference for these ideas is Lothian et al (1978).

Another way to study the quality of the seasonal fit is by means of the sliding spans analysis (Findley et al, 1990b). In the table we find that for the two economic series, the empirical values are either 0 or very small, and hence that the seasonal fit is deemed likely which is the most satisfactory category.

These analyses are also performed for the two rainfall series. The results for the F statistics, even when they differ in magnitude from those of the economic series, still lead to the qualification of seasonal present at 0.1%. The interpretation of the M and Q statistics are also not too negative: the series for Las Cejas is conditionally accepted, and that for La Cocha is rejected, in agreement with what we saw in other parts of this study. The month of cyclical dominance is 12 for both series, which is the worse possible value. Finally, the sliding spans analysis leads to the qualification of unlikely seasonal fits; we should point out that there are some reservations about the use of these methods in the case of an additive model (Findley et al, 1990).

Table 1 - Argentine's quarterly GNP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: bootstrap means and standard error estimates

a) Direct method

	Quarters				
	I	II	III	IV	Medians
1. One year ahead seasonal from X-12-ARIMA	94.746	103.512	101.232	100.578	100.905
2. Means over bootstrap replicates	94.748	103.430	101.160	100.650	100.905
3. Standard errors se_{200}	0.390	0.476	0.474	0.442	0.458
4. Standard errors $se_{200,\ 0.84}$	0.360	0.455	0.400	0.385	0.392
5. Standard errors $se_{200,\ 0.90}$	0.371	0.488	0.421	0.406	0.413
6. Standard errors $se_{200,0.95}$	0.386	0.474	0.498	0.426	0.450
Medians of the estimates of se	0.378	0.475	0.448	0.416	0.432
Standard deviations of the estimates of se	0.012	0.012	0.039	0.021	0.027

b) AR(6) procedure

	Quarters				
	I	II	III	IV	Medians
1. One year ahead seasonal from X-12-ARIMA	94.746	103.512	101.232	100.578	100.905
2. Means over bootstrap replicates	94.672	103.489	101.201	100.627	100.914
3. Standard errors se_{200}	0.490	0.588	0.525	0.510	0.518
4. Standard errors $se_{200,\ 0.84}$	0.440	0.570	0.550	0.475	0.512
5. Standard errors $se_{200,\ 0.90}$	0.480	0.554	0.527	0.480	0.503
6. Standard errors $se_{200,0.95}$	0.517	0.568	0.489	0.480	0.503
Medians of the estimates of se	0.485	0.569	0.526	0.480	0.508
Standard deviations of the estimates of se	0.028	0.012	0.022	0.014	0.006

Table 2 - Argentine's monthly IIP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: bootstrap means and standard error estimates

a) Direct method

	Months												
	1	2	3	4	5	6	7	8	9	10	11	12	Medians
1. One year ahead seasonal from X-12-ARIMA	92.320	80.520	102.930	99.260	104.320	100.770	102.960	105.570	102.710	108.000	104.380	98.270	103.545
2. Means over bootstrap replicates	92.279	80.740	102.119	99.583	103.633	101.285	103.724	105.047	103.314	105.627	103.782	98.783	103.548
3. Standard errors se_{200}	1.188	1.003	1.246	1.214	1.227	1.208	1.132	1.143	1.111	1.172	1.204	1.112	1.142
4. Standard errors $se_{200,\ 0.84}$	1.110	0.820	1.275	1.120	1.245	1.035	1.130	1.075	1.005	1.025	1.115	1.090	1.058
5. Standard errors $se_{200,\ 0.90}$	1.076	0.870	1.283	1.100	1.174	1.069	1.115	1.057	1.085	1.135	1.174	1.108	1.121
6. Standard errors $se_{200,0.95}$	1.137	0.891	1.240	1.225	1.188	1.380	1.030	1.079	1.185	1.201	1.240	1.094	1.193
Medians of the estimates of se	1.123	0.880	1.260	1.167	1.208	1.138	1.123	1.077	1.088	1.153	1.189	1.101	1.132
Standard deviations of the estimates of se	0.041	0.067	0.018	0.055	0.029	0.136	0.042	0.033	0.066	0.067	0.046	0.009	0.049

b) AR(7) procedure

	Months												
	1	2	3	4	5	6	7	8	9	10	11	12	Medians
1. One year ahead seasonal from X-12-ARIMA	92.320	80.520	102.930	99.260	104.320	100.770	102.960	105.570	102.710	108.000	104.380	98.270	103.545
2. Means over bootstrap replicates	92.397	80.823	102.036	99.531	103.555	101.356	103.672	104.907	103.546	105.473	103.743	98.865	103.644
3. Standard errors se_{200}	1.094	1.043	1.373	1.332	1.405	1.261	1.193	1.265	1.379	1.235	1.284	1.183	1.260
4. Standard errors $se_{200,\ 0.84}$	1.075	0.975	1.285	1.260	1.285	1.080	1.150	1.235	1.205	1.175	1.160	1.055	1.168
5. Standard errors $se_{200,\ 0.90}$	1.174	0.991	1.377	1.279	1.307	1.045	1.119	1.287	1.154	1.186	1.229	1.115	1.170
6. Standard errors $se_{200,0.95}$	1.201	0.979	1.459	1.356	1.398	1.140	1.195	1.204	1.307	1.237	1.207	1.191	1.222
Medians of the estimates of se	1.134	0.985	1.375	1.306	1.352	1.110	1.172	1.250	1.256	1.210	1.218	1.149	1.196
Standard deviations of the estimates of se	0.053	0.027	0.092	0.039	0.054	0.082	0.032	0.031	0.087	0.028	0.045	0.055	0.038

Table 3 - Argentine's economic series, means $S^*_{aj}(.)$ and standard errors $se_{200}(S_{aj})$, X-12-ARIMA, direct method, compared with constant regression seasonals and their standard errors

a) Quarterly GNP, 1980-1996

Quarter	Seasonal factors		Standard errors	
	Means over bootstrap replicates	Regression	Bootstrap se_{200} in (5)	Regression
I	94.748	94.008	0.390	0.477
II	103.430	101.969	0.476	0.483
III	101.160	102.081	0.474	0.489
IV	100.650	102.030	0.442	0.495

b) Monthly IIP, 1980-1996

Month	Seasonal factors		Standard errors	
	Means over bootstrap replicates	Regression	Bootstrap se_{200} in (5)	Regression
1	92.279	93.514	1.188	0.956
2	80.740	80.161	1.003	0.958
3	102.119	100.007	1.246	0.960
4	99.583	99.131	1.214	0.961
5	103.633	101.980	1.227	0.963
6	101.285	99.199	1.208	0.965
7	103.724	102.873	1.132	0.967
8	105.047	105.108	1.143	0.969
9	103.314	105.013	1.111	0.971
10	105.627	107.171	1.172	0.973
11	103.782	105.010	1.204	0.975
12	98.783	101.059	1.112	0.977

Table 4 - Las Cejas monthly rainfall, 1975-1999, one year ahead seasonal factors from X-12-ARIMA: bootstrap means and standard error estimates

a) Direct method

	Months												
	1	2	3	4	5	6	7	8	9	10	11	12	Medians
1. One year ahead seasonal from X-12-ARIMA	57.990	72.990	46.790	-10.440	-52.720	-43.900	-53.680	-51.490	-36.230	-26.400	16.810	78.960	-4.795
2. Means over bootstrap replicates	67.883	82.036	40.589	-9.432	-54.120	-46.056	-55.534	-53.787	-40.894	-25.933	22.173	71.632	-1.880
3. Standard errors SE_{200}	16.134	14.349	14.845	11.560	11.775	12.767	12.618	13.370	12.047	12.698	14.387	11.991	12.371
4. Standard errors $SE_{200,\ 0.94}$	13.435	13.510	12.015	9.830	9.895	11.270	10.625	11.620	10.690	11.245	11.295	10.170	11.068
5. Standard errors $SE_{200,\ 0.50}$	14.150	13.920	11.732	10.445	10.230	11.849	11.927	11.927	11.693	10.917	11.268	11.899	11.460
6. Standard errors $SE_{200,\ 0.06}$	14.854	14.550	12.599	10.970	10.266	12.574	12.845	12.216	11.991	11.084	13.44	11.739	11.885
Medians of the estimates of SE	14.502	14.134	12.307	10.707	10.258	12.212	12.272	12.071	11.842	11.170	12.369	11.819	11.672
Standard deviations of the estimates of SE	0.966	0.401	1.223	0.639	0.725	0.601	0.865	0.662	0.492	0.707	1.359	0.744	0.481

b) AR(12) procedure

	Months												
	1	2	3	4	5	6	7	8	9	10	11	12	Medians
1. One year ahead seasonal from X-12-ARIMA	57.990	72.990	46.790	-10.440	-52.720	-43.900	-53.680	-51.490	-36.230	-26.400	16.810	78.960	-4.795
2. Means over bootstrap replicates	69.762	82.372	41.718	-10.110	-54.010	-47.328	-53.208	-52.782	-40.749	-30.408	22.742	70.138	-3.833
3. Standard errors SE_{200}	14.600	14.395	14.342	13.499	13.582	14.057	12.342	13.140	13.715	12.700	16.276	13.857	13.786
4. Standard errors $SE_{200,\ 0.94}$	13.920	12.670	12.400	13.305	11.880	11.985	12.115	11.260	11.600	12.20	13.910	12.855	12.636
5. Standard errors $SE_{200,\ 0.50}$	13.935	12.574	12.664	13.522	13.112	12.820	11.669	12.367	11.603	12.356	14.423	13.986	13.171
6. Standard errors $SE_{200,\ 0.06}$	13.872	12.769	14.213	13.550	13.438	14.307	11.246	12.207	12.653	12.334	15.699	13.298	12.976
Medians of the estimates of SE	13.928	12.719	13.436	13.510	13.275	13.438	11.902	12.207	12.128	12.368	15.061	13.578	13.073
Standard deviations of the estimates of SE	0.366	0.760	0.878	0.096	0.670	0.948	0.419	0.669	0.878	0.148	0.950	0.453	0.418

Table 5 – La Cocha monthly rainfall, 1982-1997, one year ahead seasonal factors from X-12-ARIMA: bootstrap means and standard error estimates

a) Direct method

	Months												
	1	2	3	4	5	6	7	8	9	10	11	12	Mediana
1. One year ahead seasonal from X-12-ARIMA	151.597	60.295	50.552	-42.848	-9.776	-55.899	-71.133	-71.530	-55.469	-25.306	-17.846	91.006	-21.576
2. Means over bootstrap replicates	134.472	67.066	31.899	-32.199	-11.240	-52.615	-67.510	-68.620	-53.334	-12.433	-9.510	76.116	-10.972
3. Standard errors se_{200}	32.104	27.556	24.408	24.069	20.543	25.546	30.646	28.751	23.212	19.994	26.275	25.152	24.182
4. Standard errors $se_{200,\,0.84}$	19.655	21.260	18.760	21.785	16.330	17.750	18.970	20.515	20.145	18.245	20.690	19.550	19.848
5. Standard errors $se_{200,\,0.80}$	20.265	22.695	19.628	22.430	18.105	18.175	19.446	20.913	19.368	18.927	21.537	19.083	19.226
6. Standard errors $se_{200,\,0.95}$	21.277	22.787	21.036	22.781	18.997	20.976	19.921	21.921	20.711	18.541	24.845	20.043	20.377
Medians of the estimates of se	20.771	22.741	20.331	22.605	18.551	19.575	19.684	21.417	20.428	18.734	23.191	19.796	20.112
Standard deviations of the estimates of se	5.101	2.377	2.151	0.833	1.524	3.107	4.862	3.346	1.440	0.662	2.300	2.445	1.934

b) AR(12) procedure

	Months												
	1	2	3	4	5	6	7	8	9	10	11	12	Medians
1. One year ahead seasonal from X-12-ARIMA	151.597	60.295	50.552	-42.848	-9.776	-55.899	-71.133	-71.530	-55.469	-25.306	-17.846	91.006	-21.576
2. Means over bootstrap replicates	129.928	70.261	34.261	-29.135	-13.013	-54.149	-70.308	-71.341	-54.833	-8.816	-7.696	76.239	-6.256
3. Standard errors se_{200}	25.561	24.883	24.198	22.013	18.947	20.610	19.637	20.973	19.504	20.683	21.663	21.845	21.173
4. Standard errors $se_{200,\,0.84}$	19.005	19.150	19.240	19.225	19.030	16.670	16.880	18.425	17.585	18.865	21.065	20.355	19.610
5. Standard errors $se_{200,\,0.80}$	19.532	20.320	20.476	19.282	17.964	18.986	17.880	20.289	19.115	18.397	20.242	20.480	19.678
6. Standard errors $se_{200,\,0.95}$	20.562	20.258	25.322	21.395	18.350	18.559	18.857	20.694	18.915	18.866	21.100	19.891	19.403
Medians of the estimates of se	20.047	20.289	22.337	20.339	18.648	18.773	18.418	20.591	19.015	18.866	21.083	20.417	19.644
Standard deviations of the estimates of se	2.599	2.203	2.521	1.245	0.439	1.403	0.991	1.028	0.730	0.876	0.507	0.728	0.704

Table 6 – Rainfall series, means $S^{*}_{ej}(.)$ and standard errors $se_{200}(S_{ej})$, X-12-ARIMA, direct method, compared with constant regression seasonals and their standard errors

a) Las Cejas, monthly, 1975-1989

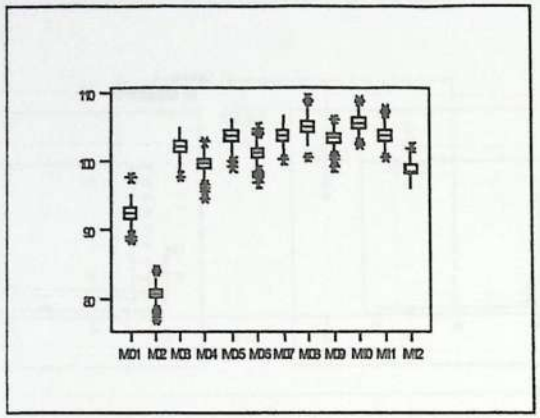
Month	Seasonal factors		Standard errors	
	Means over bootstrap replicates	Regression	Bootstrap se_{200} in (5)	Regression
1	67.893	102.503	13.435	12.206
2	82.036	89.794	13.510	12.233
3	40.569	84.610	12.015	12.260
4	-9.432	-6.486	9.830	12.286
5	-54.120	-55.483	9.895	12.314
6	-46.056	-57.368	11.270	12.341
7	-55.534	-59.863	10.625	12.369
8	-53.787	-58.907	11.620	12.397
9	-40.894	-40.234	10.890	12.425
10	-25.933	-20.502	11.245	12.453
11	22.173	20.104	11.295	12.482
12	71.632	96.549	10.170	12.510

b) Las Cocha, monthly, 1982-1997

Month	Seasonal factors		Standard errors	
	Means over bootstrap replicates	Regression	Bootstrap se_{200} in (5)	Regression
1	134.472	123.468	19.655	17.451
2	67.066	83.284	21.260	17.487
3	31.899	81.662	18.760	17.522
4	-32.199	8.766	21.785	17.559
5	-11.240	-15.371	16.330	17.595
6	-52.615	-42.735	17.750	17.632
7	-67.510	-58.069	18.970	17.669
8	-68.620	-63.266	20.515	17.706
9	-53.334	-45.931	20.145	17.744
10	-12.433	-20.593	18.245	17.782
11	-9.510	29.982	20.690	17.820
12	76.116	80.064	19.550	17.858

Figure 3 – Argentine's monthly IIP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: boxplots of bootstrap replicates

a) Direct method



b) AR(7) procedure

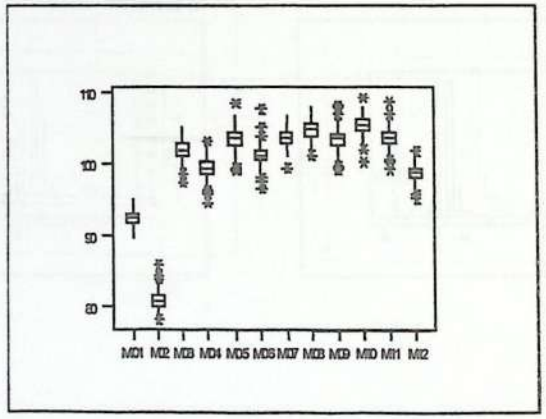
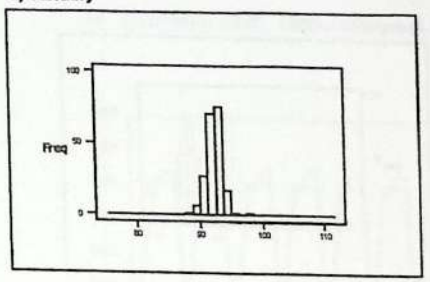
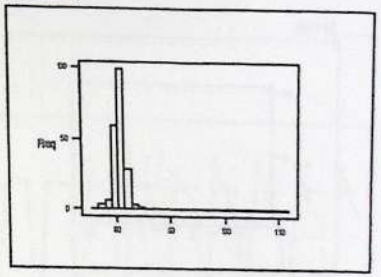


Figure 4 – Argentine's monthly IIP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (continued)

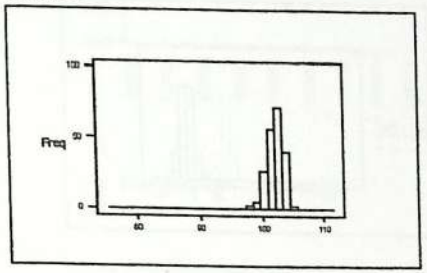
a) January



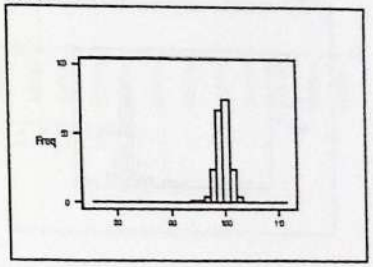
b)February



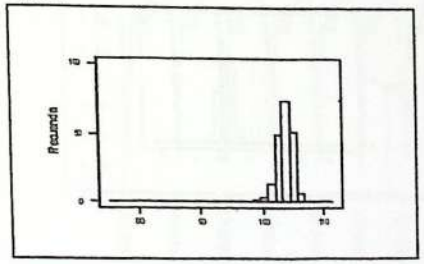
c) March



d) April



e) May



f) June

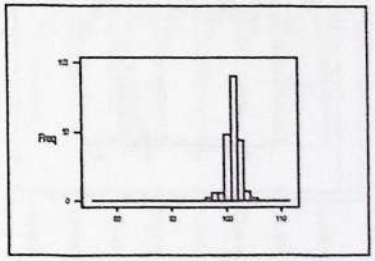


Figure 4 – Argentine's monthly IIP, 1980-1996, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (end)

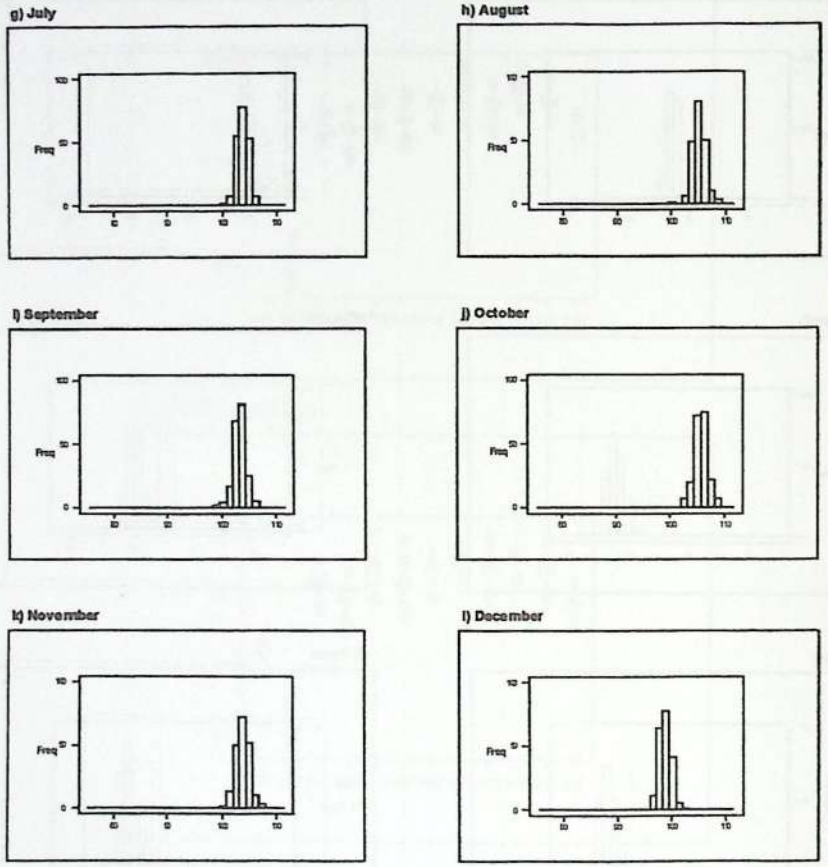
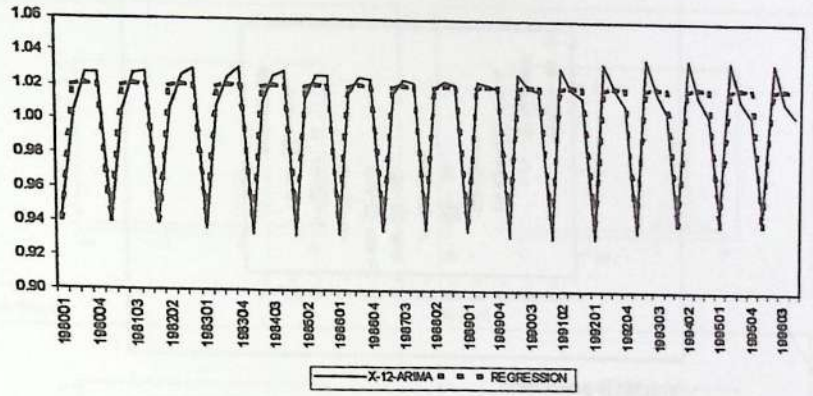


Figure 5 – Argentine's economic series, seasonal factors from X-12-ARIMA compared with constant regression seasonal

a) Quarterly GNP, 1980-1996



b) Monthly IIP, 1980-1996

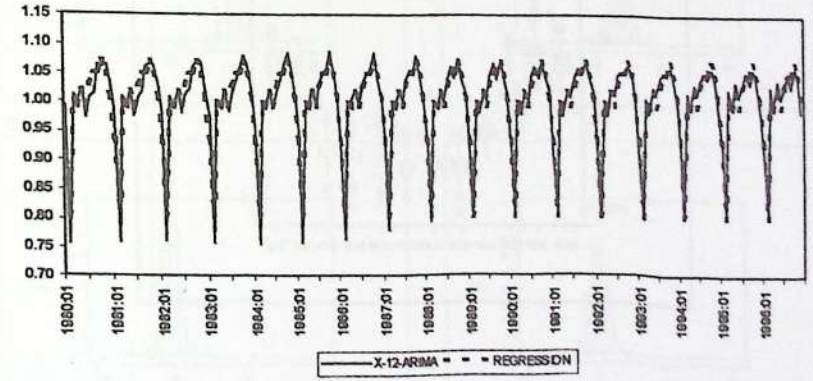
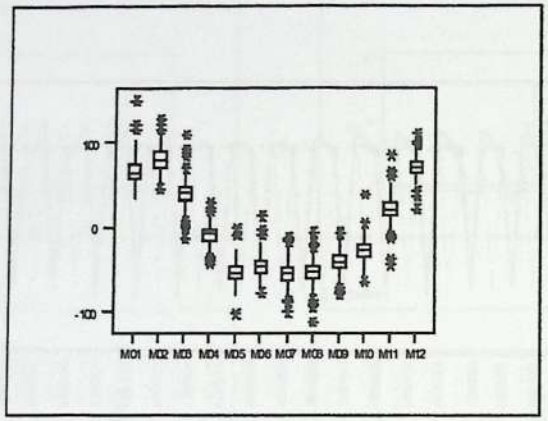


Figure 6 – Las Cejas monthly rainfall, 1975-1989, one year ahead seasonal factors from X-12-ARIMA: boxplots of bootstrap replicates

a) Direct method



b) AR(12) procedure

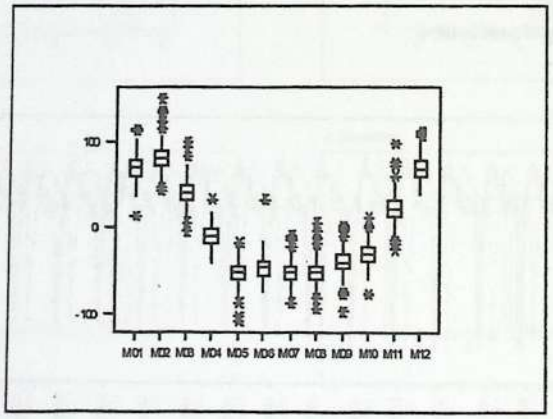


Figure 7 – Las Cejas monthly rainfall, 1975-1989, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (continued)

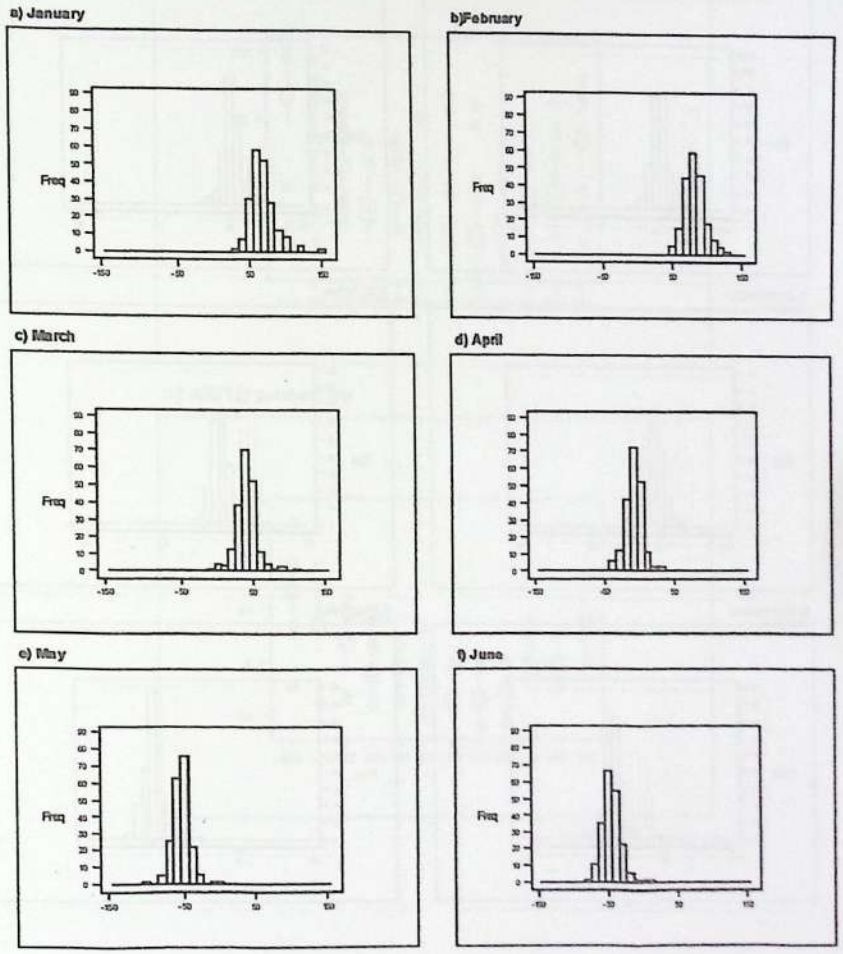


Figure 7 – Las Cejas monthly rainfall, 1975-1989, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (end)

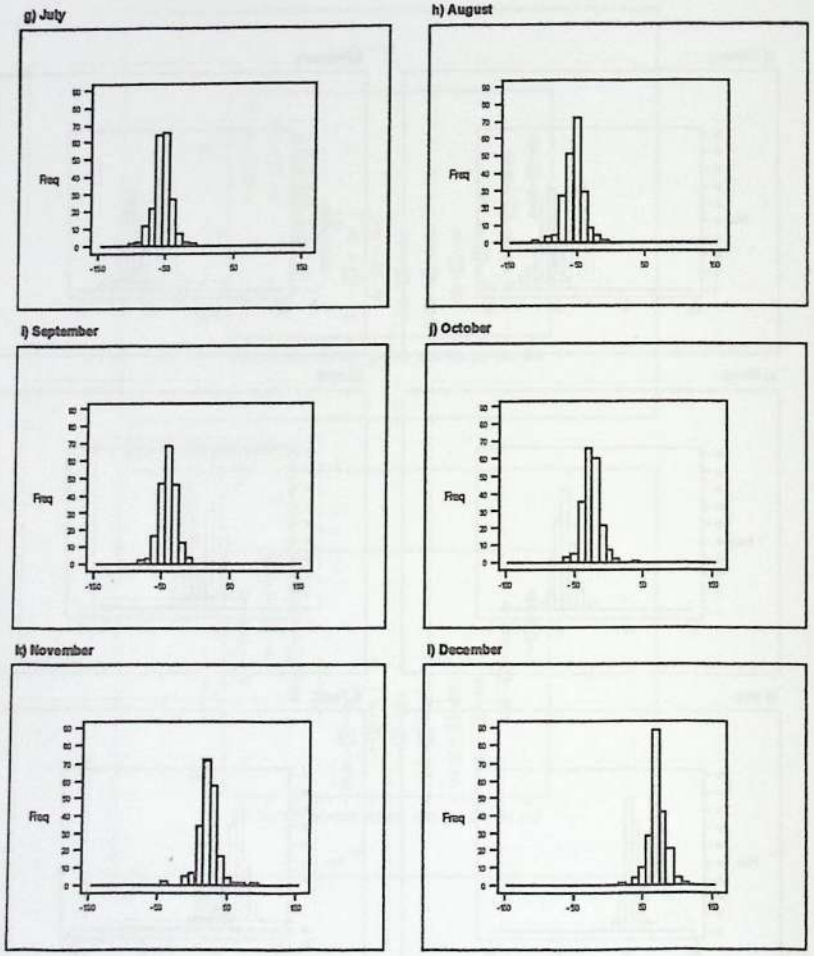


Figure 8 – La Cocha rainfall, 1982-1997, one year ahead seasonal factors from X-12-ARIMA: boxplots of bootstrap replicates

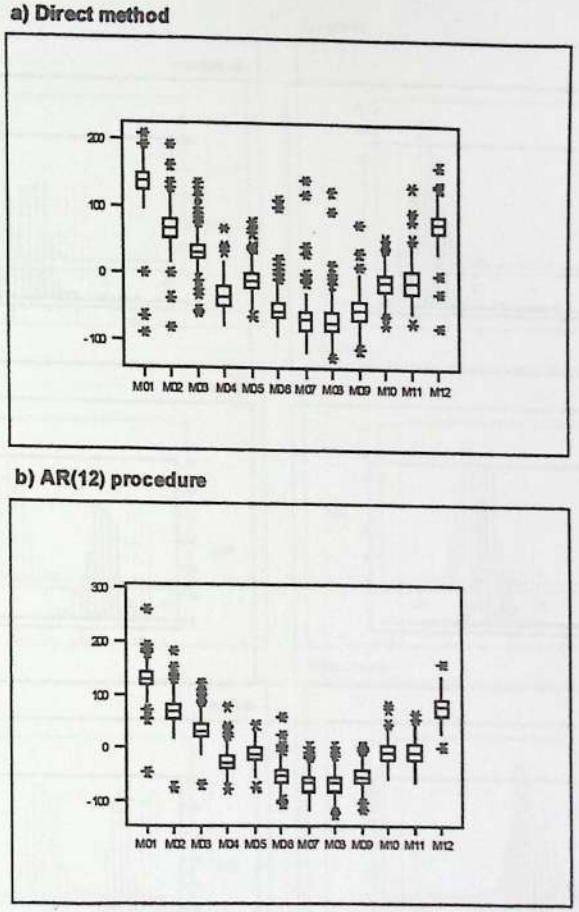


Figure 9 – La Cocha rainfall, 1982-1997, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (continued)

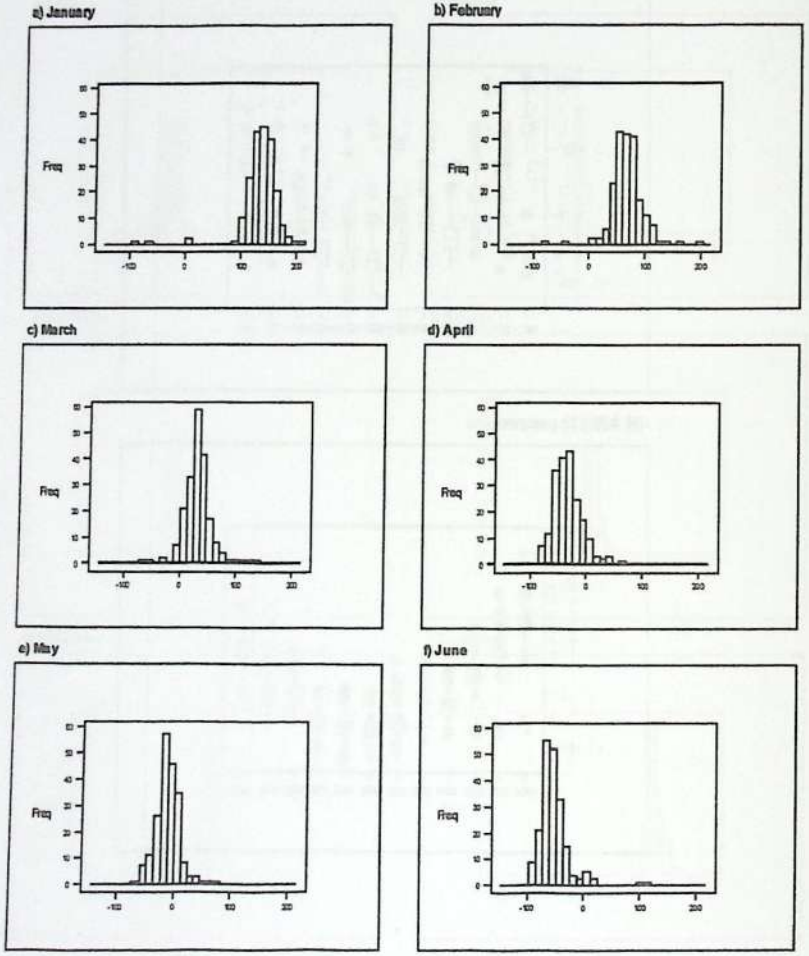


Figure 9 – La Cocha rainfall, 1982-1997, one year ahead seasonal factors from X-12-ARIMA: histograms of bootstrap replicates by direct method, common 32 class intervals (end)

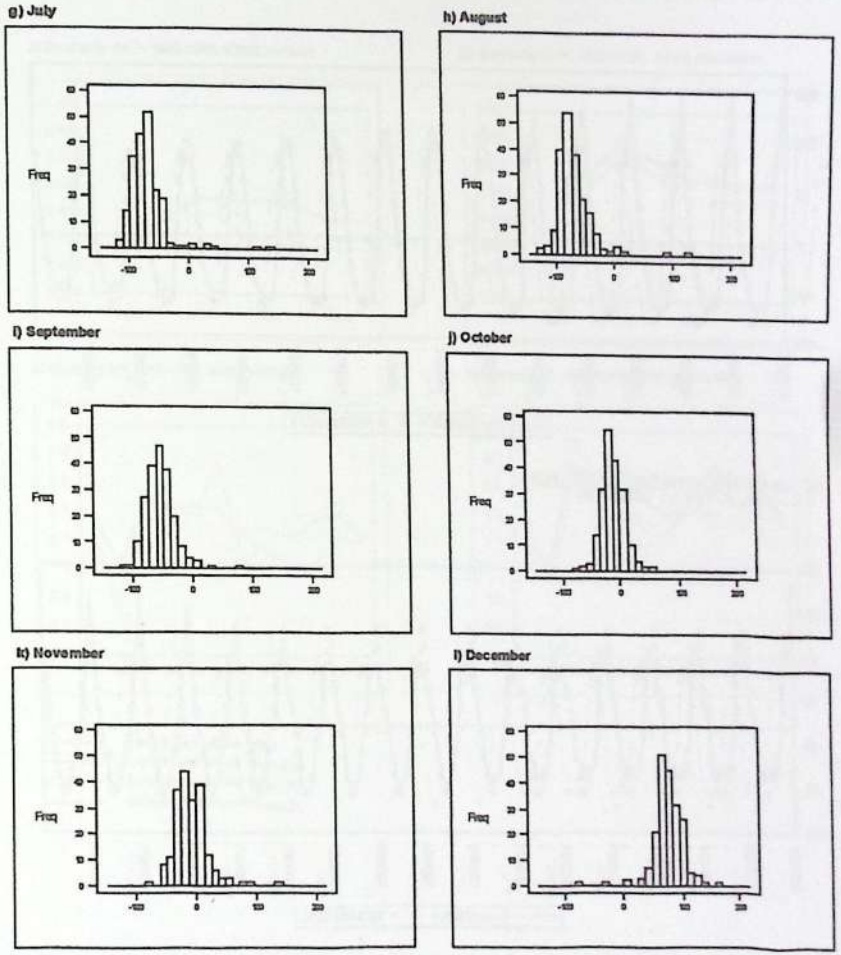
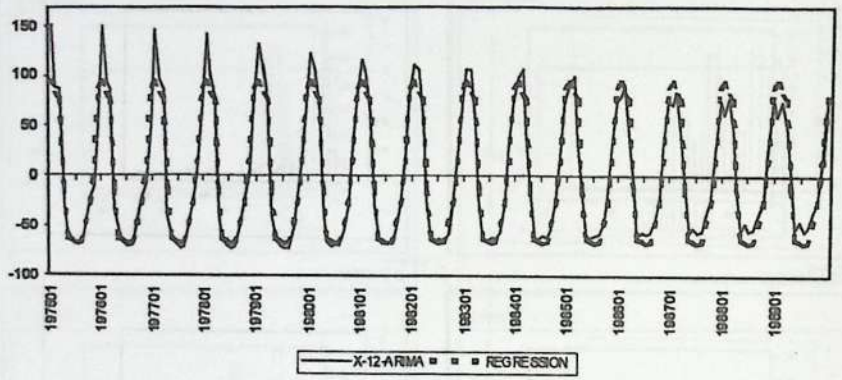


Figure 10 – Rainfall series, seasonal factors from X-12-ARIMA compared with constant regression seasonal

a) Las Cejas, monthly, 1975-1989



b) La Cocha, monthly, 1981-1997

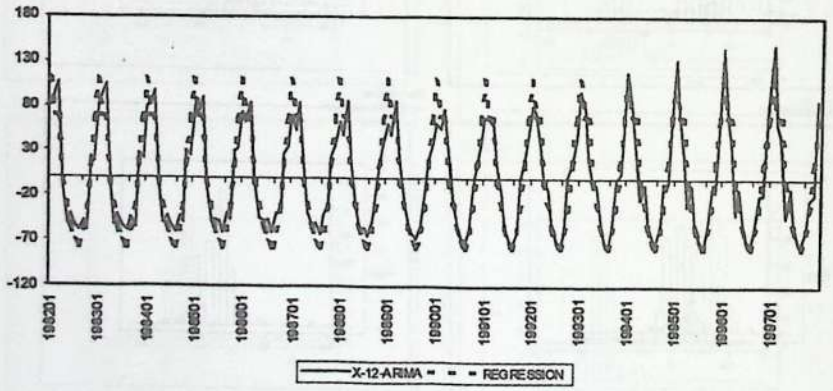


Figure 11 – Argentine's economic series, one year ahead seasonal factors from X-12-ARIMA: graphical comparison of four standard error estimates

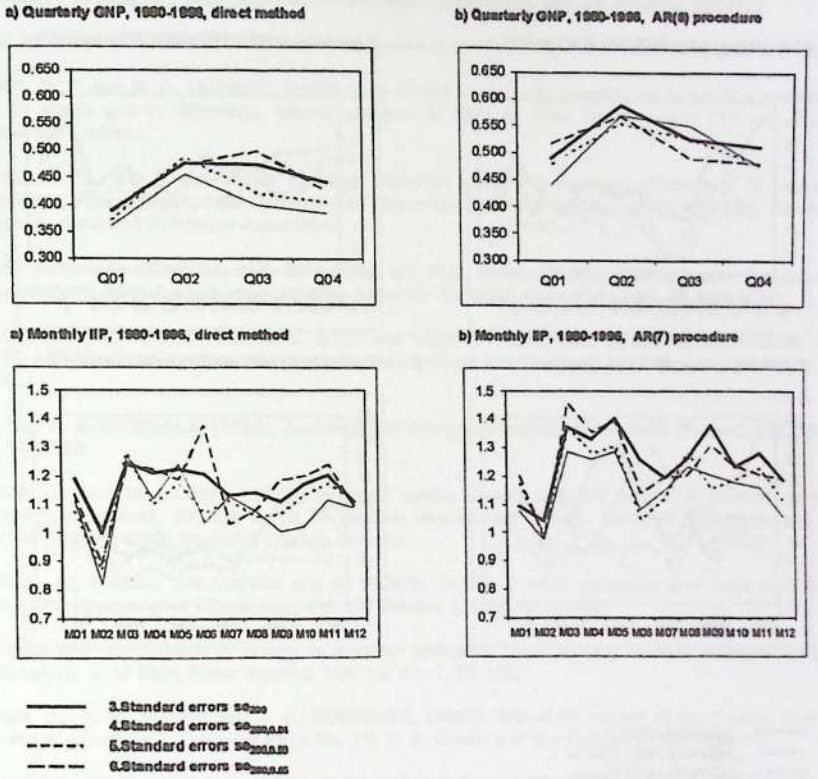
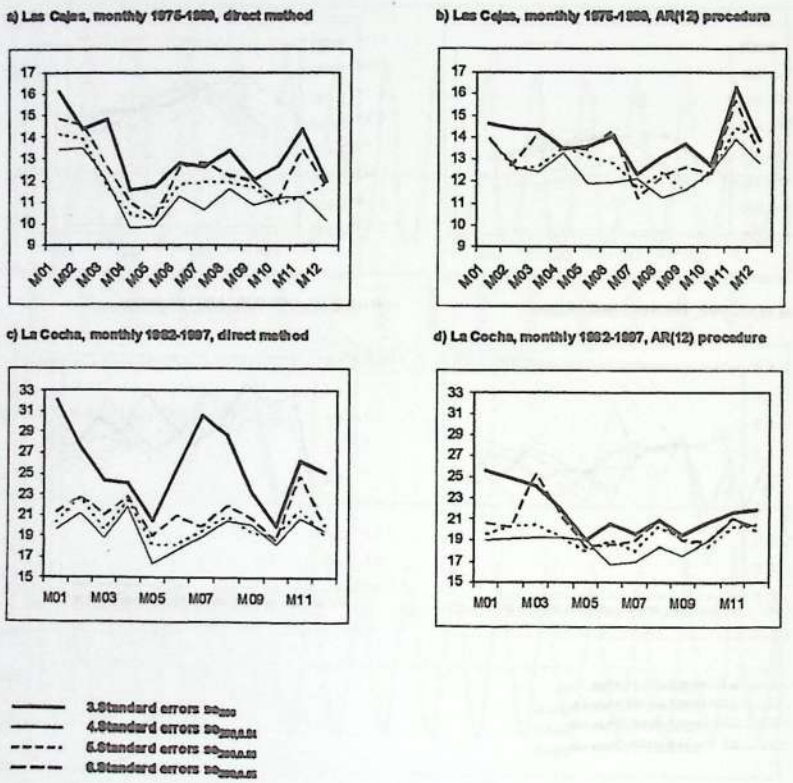


Figure 12 – Argentine's rainfall series, one year ahead seasonal factors from X-12-ARIMA: graphical comparison of four standard error estimates



Bibliography

BROCKWELL, P.J. and R.A. DAVIS, (1991a). *Time Series: Theory and Methods, 2nd Edition*, New York: Springer Verlag.

BROCKWELL, P.J. and R. A. DAVIS, (1991b). ITSM: An Interactive *Time Series Modelling Package for the PC*, New York: Springer Verlag.

CLEVELAND, W. P. and G. C. TIAO, (1976). Decomposition of seasonal time series: a model for the Census X-11 program, *J. of the American Statistical Association*, Vol. 71, No 355, 581-587.

EFRON, B. and F. TIBSHIRANI, (1993). *An Introduction to the Bootstrap*, New York: Chapman and Hall.

FINDLEY, D. F. and B. C. MONSELL, (1989). Reg-ARIMA model preprocessing for seasonal adjustment. In A. C. Singh and P. Whitridge, editors, *Analysis of Data in Time Series*, pages 117-123, Ottawa; Statistics Canada.

FINDLEY, D. F. and B. MONSELL, (1990a). Standard errors for seasonal adjustment: A resampling approach, *Proceedings of the Business and Economics Statistics Section*, pages 455-460, Alexandria, Virginia, American Statistical Association.

FINDLEY D.F., B.C. MONSELL, H.B. SHULMAN, and M.G. PUGH, (1990b). Sliding spans diagnostics for seasonal and related adjustment, *J. of the American Statistical Association*, Vol. 85, 345-355

FINDLEY, D. F., B., C. MONSELL, M. C. OTTO and B.CHEN, (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program, *J. of Business and Economic Statistics*, Vol. 16, No. 2, 127-176.

LI, H. and G. S. MADDALA, (1996). Bootstrapping time series models, *Econometric Reviews*, Vol. 15, No. 2, 115-158

LOTHIAN, J. and M. MORRY, (1978). A set of quality control statistics for the X11-ARIMA seasonal adjustment method, *Working paper* 78-10-005, Methodology Branch, Seasonal Adjustment and Time Series Analysis Staff, Statistics Canada, Ottawa.

MENTZ R, M. D'URSO , N. JARMA and G. MENTZ, (2000). Cyclical components of local rainfall data, *International Journal of Climatology*, Vol. 20, Number 2, pages 211-228.

PFEFFERMANN, D., (1994). A general method for estimating the variances of X-11 seasonally adjusted estimators, *J. of Time Series Analysis*, Vol. 15, No. 1, 85-116.

SHISKIN, J., A. H. YOUNG and J. C. MUSGRAVE, (1967). The X-11 variant of the Census Method II seasonal adjustment. Technical Paper No. 15, U. S. Bureau a of the Census, Washington.

U.S. BUREAU OF THE CENSUS, (1999). *X-12-ARIMA Reference Manual*, Final Version 0.2.5 Washington D.C.: U.S. Bureau of the Census.

Wallis, K. F., (1974). Seasonal adjustment and relations between variables, *J. of the American Statistical Association*, Vol. 69, No. 345, 18-31.

Acknowledgments

The authors thank Marcela D'Urso for help in the initial phase of this study and in the preparation of the Appendix. They thank the comments by Professor A.J. Viollaz, and a conversation held with Professor Lutz Kilian at the University of Michigan. Reports by two referees were very useful to improve the contents and presentation.

Note on the autoregressive spectral estimator

Raúl P. Mentz and Carlos I. Martínez
Universidad Nacional de Tucumán and CONICET

Abstract: An estimator of the spectral density of a stationary time series is obtained by fitting to the observations an autoregressive model (often including the estimation of its order), and computing with sample values the spectrum of the indicated model. In the present note we consider the calculation of simultaneous confidence bands, according to Newton and Pagano (1984). The procedure is illustrated by means of Monte Carlo simulations, for series generated by autoregressive models or orders up to 5.

Key words: Asymptotic properties, confidence bands, estimation of autoregressive order, Monte Carlo.

1. Introduction

Given a time series (or stochastic process) y_t, $t = 0, \pm 1, \pm 2, \ldots$ with the property of being stationary, we define its *autocovariance function* as the set of expected values

$$\sigma_{st} = E(y_t - \mu)(y_s - \mu), \; s, t = 0, \pm 1, \pm 2, \ldots \tag{1.1}$$

where μ is the expected value of the process.

These expectations are assumed to exist, in the sense that the defining series or integrals converge, and be positive semidefinite, in the sense that a matrix of arbitrary size formed with them should possess this property.

We also define its *autocorrelation function*

$$\rho_s = \frac{\sigma_s}{\sigma_0}, \; s = 0, \pm 1, \pm 2, \ldots \tag{1.2}$$

It is often convenient to consider transforming these parametric functions by defining the *spectral density function* of the process, namely

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sigma_h \cos(\lambda h) = \frac{\sigma_0}{2\pi} \sum_{h=-\infty}^{\infty} \rho_h \cos(\lambda h), \; -\pi \leq \lambda \leq \pi. \tag{1.3}$$

This will exist provided the series converges. If the series of absolute values of the covariances converges, then (1.3) converges uniformly; see, for example, Anderson (1971).

An important example of time series is that corresponding to an *autoregressive model of order p*, that we identify by AR(p), with coefficients $\alpha_1, \alpha_2, \ldots, \alpha_p$ and variance σ^2; it satisfies the (stochastic) finite difference equation

$$\sum_{j=0}^{p} \alpha_j (y_{t-j} - \mu) = \varepsilon_t, \; t = 0, \pm 1, \pm 2, \ldots, \tag{1.4}$$

where we define $\alpha_0 = 1$, and the ε_t constitute a stochastic process of uncorrelated (or independent) random variables with 0 expected value and constant variance σ^2. The α_j satisfy the condition that the associated polynomial equation, $\sum \alpha_j z^j = 0$ has all of its roots, real or complex, larger than 1 in absolute value. Under the stated condition y_t can be "inverted", i.e., expressed as an infinite linear combination of the ε_t's, current and past, and further, y_t is independent of ε_t for $s > t$, i.e., independent of future values of the ε_t's.

Under these conditions, it can shown that the spectral density function of the AR(p) process is

$$f_{\mathrm{AR}}(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{\left| \sum_{j=0}^{p} \alpha_j e^{i\lambda j} \right|^2}, \; -\pi \le \lambda \le \pi, \tag{1.5}$$

so that if, for example, $p = 1$, defining $\alpha = \alpha_1$ we find that the spectral density function of the AR(1) model is

$$f_{\mathrm{AR}(1)}(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{1 + \alpha^2 + 2\alpha \cos \lambda}, \; -\pi \le \lambda \le \pi. \tag{1.6}$$

The condition on the roots now implies that α is less than 1 in absolute value. The *autoregressive estimator* of the spectral density function of a given (stationary) time series is defined as follows: (1) The series is approximated by an AR($\hat{p}$) model, whose order $\hat{p}$ is suitably chosen; (2) The $\hat{p} + 1$ parameters of this model are estimated; (3) With the estimates we form a function similar to (5), and we take this function (of frequency λ) as the sample estimator of the spectral density.

This procedure is related to that of Maximum Entropy. Priestley (1981, page 604) states: The method of maximum entropy is based on choosing a spectral estimate $\hat{h}(\omega)$ which is such that the entropy $E = \int_{-\pi}^{\pi} \log\{\hat{h}(\omega)\} d\omega$ is maximized subject to the constraints (in our notation) $\int_{-\pi}^{\pi} \hat{h}(\omega) e^{i\omega r} d\omega = c_r$, $r = 0, \pm 1, \ldots, \pm k$. In Akaike (1977) it is shown that the resulting form of $\hat{h}(\omega)$ is exactly the same as for the AR spectral estimate, where the c_r satisfy the Yule-Walker equations.

Newton and Pagano (1984) considered the problem of computing simultaneous confidence bands for the spectral density function. The problem arises because the estimator consists of a set of values (in general it is computed at frequencies $\lambda_j = 2\pi j/T$, $j = 0, 1, 2, \ldots, [T/2]$, where T is the sample size), each one having sampling variability.

In this note we analyze through simulated examples, the behavior of the proposed confidence band. We want to compare results with those in Newton and

Pagano (1984) and with those in Mentz et al (2003). In the former no logarithmic transformation is introduced, so that we follow this option. Logarithms are used in the second reference.

Newton and Pagano (1984) approached the problem by considering that the reciprocal of the spectral density function of an AR(p) model, has the form of the spectral density function of a moving average model of the same order, with the same coefficients and with the variance of the error term equal to $4\pi^2/\sigma^2$; we denote this model by MA(p).

2. The AR(1) model

With sample values $y_1, \ldots, y_T$, parameters μ, α and σ^2 are estimated by one of the known methods of parametric estimation: Yule-Walker, least squares, what is known as *Burg's algorithm*, or maximum likelihood under the assumption that the errors in the models have a normal distribution. We can also estimate the autocovariances of the process, which means that the output of the estimation stage is, in the case of an AR(1) model, the following:

$$\hat{\mu}, \hat{\alpha}, \hat{\sigma}^2, c_j = \frac{1}{T} \sum_{j=1}^{T-j} (y_t - \overline{y})(y_{t+j} - \overline{y}), \; j = 0, 1, 2. \tag{2.1}$$

With these sample values we define a set of autocovariances corresponding to an MA(p) model namely,

$$\hat{\gamma}_0 = \frac{4\pi^2}{\hat{\sigma}^2}(1 + \hat{\alpha}^2), \;\; \hat{\gamma}_1 = \frac{4\pi^2}{\hat{\sigma}^2}\hat{\alpha}, \tag{2.2}$$

and the empirical spectral density function of the MA(1) model is

$$\hat{h}(\omega) = \frac{1}{\hat{f}(\omega)} = \frac{1}{2\pi}\{\hat{\gamma}_0 + 2\hat{\gamma}_1 \cos(\omega)\}. \tag{2.3}$$

We next define two matrices as follows:

$$\mathbf{C}(\theta) = \mathrm{diag}(\sigma^2 \mathbf{\Gamma}^{-1}, 2\sigma^{-4}), \tag{2.4}$$

where $\mathbf{\Gamma}$ is the matrix of variances and covariances of the process; in the special case of an AR(1) model this is σ_0, and is estimated by the sample value c_0. The other matrix is

$$\mathbf{B} = \begin{pmatrix} \frac{4\pi^2}{\sigma^2} 2\alpha & \sigma^2 \gamma_0 \\ \frac{4\pi^2}{\sigma^2} & \sigma^2 \gamma_1 \end{pmatrix}. \tag{2.5}$$

These two matrices lead to estimates of $\mathbf{C}$ and $\mathbf{B}$ by substituting sample values for α and σ^2 and further estimate γ using definitions (2.2). Defining

$$\hat{\mathbf{D}} = \hat{\mathbf{B}}\hat{\mathbf{C}}\hat{\mathbf{B}} \tag{2.6}$$

the confidence band for the spectral density function is

$$\frac{1}{\hat{h}(\omega)+\hat{s}(\omega)} \le f(\omega) \le \frac{1}{\hat{h}(\omega)-\hat{s}(\omega)}, \tag{2.7}$$

where

$$\hat{s}^2(\omega) = \frac{1}{T}\chi^2_{\delta,p+1}\mathbf{x}'(\omega)\hat{\mathbf{D}}(\omega)\mathbf{x}(\omega), \tag{2.8}$$

$\chi^2_{\delta,p+1}$ is the percentile of the chi-square distribution with $p+1=2$ degrees of freedom, and we introduce the two-dimensional vector $\mathbf{x} = (1/\pi)(1/2\cos(\omega))'$.

If it happens that $\hat{h}(\omega)-\hat{s}(\omega) \le 0$, the upper bound of the confidence band is taken to be infinite.

Under these conditions, the (simultaneous) confidence bands have confidence coefficient approximately equal to δ. The statistical argument is based on projections proposed by Scheffé (1959, page 407), as cited by Newton and Pagano (1984).

3. The AR(5) model

As a more general example we consider the AR(5) model, one that is studied by Newton and Pagano (1984). Observations y_t, $t = 1, 2, \ldots, T$ come from the model

$$y_t + \sum_{j=1}^{5} \alpha_j y_{t-j} = \varepsilon_t, \tag{3.1}$$

where the ε_t are independent with expected value 0 and constant variance $\sigma^2 > 0$. The output of the computer program includes the estimates $\hat{\alpha}_j$, $\hat{\sigma}^2$ and $c_j = (1/T)\sum_{t=1}^{T-j}(y_t-\overline{y})(y_{t+j}-\overline{y})$ where $\hat{\mu} = \overline{y} = \sum_{t=1}^{T} y_t/T$.

The MA(5) used to approximate the autoregression has autocovariances estimated by

$$\hat{\gamma}_j = \frac{4\pi^2}{\hat{\sigma}^2}\sum_{s=0}^{5-j}\hat{\alpha}_s\hat{\alpha}_{s+j}, \quad |j| = 0, 1, \ldots, 5, \tag{3.2}$$

and by 0 for all other values of j. The spectral density of the MA(5)model is estimated by

$$\hat{h}(\omega) = \frac{1}{2\pi}\left\{\hat{\gamma}_0 + 2\sum_{j=1}^{5}\hat{\gamma}_j\cos(j\omega)\right\}, \quad -\pi \le \omega \le \pi. \tag{3.3}$$

Defining matrices

$$\hat{\Gamma} = \|c_{|i-j|}\|, \tag{3.4}$$

$$\hat{\mathbf{C}} = \operatorname{diag}(\hat{\sigma}^2, \hat{\Gamma}^{-1}, 2\hat{\sigma}^{-4}), \tag{3.5}$$

and defining c as

$$c = \frac{4\pi^2}{\hat{\sigma}^2} \tag{3.6}$$

we form the following matrix of order $(p+1)\times(p+1) = 6\times 6$:

$$\hat{\mathbf{B}} = \frac{1}{c}\begin{pmatrix} 2\hat{\alpha}_1 & 2\hat{\alpha}_2 & 2\hat{\alpha}_3 & 2\hat{\alpha}_4 & 2\hat{\alpha}_5 & c\hat{\gamma}_0\hat{\sigma}^2 \\ 1+\hat{\alpha}_2 & \hat{\alpha}_1+\hat{\alpha}_3 & \hat{\alpha}_2+\hat{\alpha}_4 & \hat{\alpha}_3+\hat{\alpha}_5 & \hat{\alpha}_4 & c\hat{\gamma}_1\hat{\sigma}^2 \\ \hat{\alpha}_3 & 1+\hat{\alpha}_4 & \hat{\alpha}_1+\hat{\alpha}_5 & \hat{\alpha}_2 & \hat{\alpha}_3 & c\hat{\gamma}_2\hat{\sigma}^2 \\ \hat{\alpha}_4 & \hat{\alpha}_5 & 1 & \hat{\alpha}_1 & \hat{\alpha}_2 & c\hat{\gamma}_3\hat{\sigma}^2 \\ \hat{\alpha}_5 & 0 & 0 & 1 & \hat{\alpha}_1 & c\hat{\gamma}_4\hat{\sigma}^2 \\ 0 & 0 & 0 & 0 & 1 & c\hat{\gamma}_5\hat{\sigma}^2 \end{pmatrix}. \tag{3.7}$$

This matrix contains those of smaller order. For example, in the case of the AR(4) model, we omit the last row and the fifth column, and set $\hat{\alpha}_5 = 0$, since $\alpha_5 = 0$.

With these matrices we form

$$\hat{\mathbf{D}} = \hat{\mathbf{B}}\hat{\mathbf{C}}\hat{\mathbf{B}} \tag{3.8}$$

and

$$\hat{s}^2(\omega) = \frac{1}{T}\chi_{\delta,6}\,\mathbf{x}'\hat{\mathbf{D}}\mathbf{x} \tag{3.9}$$

where the vector $\mathbf{x}$ is defined by

$$\mathbf{x} = \frac{1}{\pi}\left(\frac{1}{2}\ \cos\omega\ \cos 2\omega\ \cos 3\omega\ \cos 4\omega\ \cos 5\omega\right)'. \tag{3.10}$$

Finally, the confidence bands for the spectral density can be written as

$$\frac{1}{\hat{h}(\omega)+\hat{s}(\omega)} \le f(\omega) \le \frac{1}{\hat{h}(\omega)-\hat{s}(\omega)}. \tag{3.11}$$

This band has, approximately, level δ. Again, this is the key result in Newton and Pagano's main theorem.

As indicated, Newton and Pagano's approach is to set a band to $h(\omega) = 1/f(\omega)$. The band is $\hat{h}(\omega)-s(\omega) \le h(\omega) \le \hat{h}(\omega)+s(\omega)$, and has no problem even if the left hand side is negative for some frequency. When deriving the band for $f(\omega)$ they write: "It could happen that $\hat{h}(\omega)-s(\omega) < 0$ for some ω in which case infinity for the upper limit which does not diminish the probability contents of the band." One could solve this problem by letting this band to be $\max\{\hat{h}(\omega)-s(\omega), d\}$ where d is some constant, for example, $d = 0.001$. This procedure could alter the probability contents of the band or, to say it in other words, the level δ will be only approximately met. In our rather extensive simulations with the given models, we found no cases of negative estimates.

4. Numerical examples

In this section we present and analyze some numerical examples. Observations were generated by Monte Carlo simulations, taking some values for the parameters and taking the error variance equal to 1. This choice coincides with that of Newton and Pagano, which means that the variance is treated as a scaling factor, the main interest being in estimating the spectral density. In all cases $\delta = 0.05$, which means that the confidence bands have (approximately) simultaneous 95% confidence level.

We use the S-PLUS package. The functions we use are: "arima.sim" to simulate the errors; "ar.burg" to fit AR models using Burg's algorithm; "arima.mle" to fit AR models via maximum likelihood estimation under the Gaussian distribution.

4.1. Models

Observations were generated by four models, namely:

I. AR(1) with $\alpha_1 = -0.60$.

II. AR(2) with $\alpha_1 = -0.40$ and $\alpha_2 = -0.45$.

III. AR(4) with $\alpha_1 = -2.760$, $\alpha_2 = 3.8106$, $\alpha_3 = -2.6536$ and $\alpha_4 = 0.9238$.

IV. AR(5) with $\alpha_1 = 1.7$, $\alpha_2 = 2.4$, $\alpha_3 = 1.634$, $\alpha_4 = 0.872$ and $\alpha_5 = 0.168$.

By using models II, III and IV we follow Beamish and Priestley (1981), Newton and Pagano (1984) and others. The argument is that they are "easy, moderately easy and very difficult to fit, respectively". These models have roots (1.11, 2.00), (1.12, 1.12, 1.19, 1.19, 3.33), and (1.02, 1.02, 1.02, 1.02), respectively, so that the roots of model IV are all close to 1 in absolute value. Model I with parameter 0.60 is expected to behave like model II.

4.2. Dealing with ρ and parameter estimation

4.2.1. Dealing with ρ

The autoregressive model has known order p, since this is the model generating the simulated observations. In the computer program used to estimate the parameters, this value of p can be forced, which implies that exactly as many α_j's as is the value of p are estimated. In our case, parameter estimation is done by maximum likelihood using the normal distribution; if observations have only approximately a normal distribution, asymptotic theory shows that use of this procedure provides a valid approximation; see, for example, Brockwell and Davis (1987, Section 10.8).

Observations generated by a model with known p can be used to estimate the order $\hat{p}$ of the AR model that provides the best fit to the data. In the S-PLUS program, this is done by *Burg's algorithm* for parameter estimation, and the AIC

procedure to estimate the order. In Newton and Pagano, Parzen's CAT procedure is used together with Burg's algorithm. Our choice of AIC come from the program we use, namely, S-PLUS. One of our objectives is to explore the behavior of the computer programs in this package.

Table 1 shows the results of applying the indicated procedure to 1000 Monte Carlo replications. Values concentrate for $\hat{p} = p$, but we also observe an important number of cases in which $\hat{p} > p$. In Table 1a) the estimated $\hat{p}$ is restricted to be ≤ 5 while in Table 1b) it is restricted to be ≤ 10.

Table 1a *Frequency distributions of autoregressive orders chosen by S-PLUS using Burg's estimation algorithm, in 1000 replications ($p \leq 5$).*

Model	True order p	Sample size T	Chosen order				
			1	2	3	4	5
I	1	100	733	113	68	46	40
		200	744	121	61	44	30
		400	722	146	55	38	39
II	2	100	1	745	134	69	51
		200	0	756	133	53	58
		400	0	766	109	58	57
III	4	100	0	0	0	819	181
		200	0	0	0	836	164
		400	0	0	0	850	150
IV	5	100	0	0	0	428	572
		200	0	0	0	207	793
		400	0	0	0	38	962

Table 1b *Frequency distributions of autoregressive orders chosen by S-PLUS using Burg's estimation algorithm, in 1000 replications ($p \leq 10$).*

Model	True order p	Sample size T	Chosen order									
			1	2	3	4	5	6	7	8	9	10
I	1	100	699	108	63	37	30	18	11	11	11	12
		200	702	115	57	36	21	21	20	12	7	9
		400	691	138	49	34	30	19	14	10	8	7
II	2	100	1	694	118	57	29	32	25	13	17	14
		200	0	719	118	49	38	26	18	15	8	9
		400	0	730	101	53	42	23	14	18	11	8
III	4	100	0	0	0	665	135	58	46	44	24	28
		200	0	0	0	712	123	63	34	36	16	16
		400	0	0	0	742	107	54	33	25	18	21
IV	5	100	0	0	0	365	399	105	52	38	26	25
		200	0	0	0	171	569	113	49	41	32	25
		400	0	0	0	33	744	98	46	38	20	21

Shibata (1976) derived the explicit asymptotic distribution of the estimated order $\hat{p}$. If this order is selected by means of the AIC criterion in the range from 0 to, say, P, the probability distribution of $\hat{p}$ has values from the true order to P. The distribution does not depend on the α_j, and these are estimated by maximum likelihood. In a simulation study for the AR(1) model with parameter $\alpha = 0.80$, he showed that the frequency of selecting the order $\hat{p} = 1$ is roughly 0.70, and our results in Table 1 agree with this.

Since the models proposed as examples have $p \leq 5$, the restriction that $\hat{p} \leq 10$ is reasonable: larger values of this estimate will introduce further estimates of α parameters and raise sample variability. Note that in Table 1, $p = 1$ and $\alpha = 0.60$, $\hat{\alpha}_1$ is close to 0.60 and $\hat{\alpha}_j$, $j = 2, 3, 4$ and 5 are small. Similar behavior is observed in the case of $p = 2$, $\hat{p} \leq 5$. Further discussion of the effect of estimating the AR order is presented below.

For models I and II ($p = 1$ and 2), changing from the restriction $\hat{p} \leq 5$ to $\hat{p} \leq 10$ alters the selection of order in about 5% of the cases: for model I selecting $\hat{p} = p$ reduces from 733 to 699 cases (out of 1000): in the remaining cases some value $\hat{p} > p$ is selected. For models III and IV ($p = 4$ and 5) the effect is larger: for $p = 4$, selecting $\hat{p} = p = 4$ reduces in about 15% of the cases in favor of choosing values of $\hat{p} > 4$; for $p = 5$, selecting $\hat{p} = p = 5$ reduces in about a 30% of the cases in favor of choosing values of $\hat{p} \neq 5$ (for some series the selected order is 4).

4.2.2. Parameter estimation

Table 2 shows the average results in the estimation of parameters. Sample sizes were $T = 100, 200$ and 400, and 1000 replications (repetitions) were done in each case. Table 2 is done under the restriction $\hat{p} \leq 10$. Another table under the restriction $\hat{p} \leq 5$ is available from the authors.

Table 2 *Average estimates in 1000 replications*

(a) Model I, $p = 1$, $\boxed{\alpha_1 = 0.60}$ MLE

T	α_1	$\mathrm{Var}(\alpha_1)^{1/2}$	σ^2
100	0.5895	0.0805	0.9818
200	0.5944	0.0568	0.9932
400	0.5971	0.0401	0.9958

(b) Model I, $p = 1$, $\boxed{\alpha_1 = 0.60}$ p_{est} Burg's algorithm

T	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	σ^2
100	0.5907	-0.0084	0.0091	-0.0494	-0.0178	0.0296	-0.0117	-0.0545	0.0299	-0.0136	0.9589
200	0.5964	-0.0048	-0.0082	-0.0224	0.0220	-0.0118	-0.0062	-0.0460	0.0422	0.0286	0.9818
400	0.5977	-0.0002	-0.0088	-0.0013	0.0168	-0.0292	-0.0144	0.0038	0.0498	0.0792	0.9899

Table 2 *(continued)*

(c) Model II, $p = 2$, $\boxed{\alpha_1 = 0.40}$ $\boxed{\alpha_2 = 0.45}$ MLE

T	α_1	α_2	$\mathrm{Var}(\alpha_1)^{1/2}$	σ^2
100	0.3966	0.4255	0.0909	0.9707
200	0.3978	0.4369	0.0637	0.9884
400	0.3986	0.4446	0.0448	0.9934

(d) Model II, $p = 2$, $\boxed{\alpha_1 = 0.40}$ $\boxed{\alpha_2 = 0.45}$ p_{est} Burg's algorithm

T	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	σ^2
100	0.3986	0.4287	0.0050	-0.0364	-0.0052	0.0105	-0.0152	-0.0260	0.0695	-0.0465	0.9499
200	0.3991	0.4395	-0.0015	-0.0254	0.0058	-0.0095	-0.0334	-0.0100	0.0445	0.0623	0.9776
400	0.3985	0.4445	-0.0010	-0.0004	0.0071	-0.0042	0.0030	-0.0102	0.0256	-0.0278	0.9881

(e) Model III, $p = 4$, $\boxed{\alpha_1 = 2.760}$ $\boxed{\alpha_2 = -3.8106}$ $\boxed{\alpha_3 = 2.6535}$ $\boxed{\alpha_4 = -0.9238}$ MLE

T	α_1	α_2	α_3	α_4	$\mathrm{Var}(\alpha_1)^{1/2}$	σ^2
100	2.7370	-3.7478	2.5892	-0.8944	0.0442	0.9493
200	2.7481	-3.7774	2.6190	-0.9077	0.0295	0.9786
400	2.7539	-3.7935	2.6356	-0.9155	0.0201	0.9880

(f) Model III, $p = 4$, $\boxed{\alpha_1 = 2.760}$ $\boxed{\alpha_2 = -3.8106}$ $\boxed{\alpha_3 = 2.6535}$ $\boxed{\alpha_4 = -0.9238}$ p_{est} Burg's algorithm

T	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	σ^2
100	2.7397	-3.7579	2.6074	-0.9192	0.0802	-0.1246	0.1192	-0.0932	0.0549	-0.0338	0.9335
200	2.7484	-3.7806	2.6278	-0.9231	0.0617	-0.0990	0.1112	-0.0949	0.0753	-0.0455	0.9711
400	2.7540	-3.7962	2.6420	-0.9237	0.0272	-0.0517	0.0963	-0.1323	0.1048	-0.0483	0.9839

(g) Model IV, $p = 5$, $\boxed{\alpha_1 = -1.7}$ $\boxed{\alpha_2 = -2.4}$ $\boxed{\alpha_3 = -1.634}$ $\boxed{\alpha_4 = -0.872}$ $\boxed{\alpha_5 = -0.168}$ MLE

T	α_1	α_2	α_3	α_4	α_5	$\mathrm{Var}(\alpha_1)^{1/2}$	σ^2
100	-1.6802	-2.3508	-1.5788	-0.8355	-0.1558	0.1008	0.9432
200	-1.6906	-2.3763	-1.6061	-0.8537	-0.1606	0.0705	0.9724
400	-1.6957	-2.3879	-1.6208	-0.8626	-0.1650	0.0496	0.9848

Table 2 *(continued)*

(h) Model IV, $p = 5$, $\alpha_1 = -1.7$ $\alpha_2 = -2.4$ $\alpha_3 = -1.634$ $\alpha_4 = -0.872$ $\alpha_5 = -0.168$
p_{est} Burg's algorithm

T	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	σ^2
100	-1.6680	-2.3278	-1.5388	-0.8103	-0.2139	-0.0014	0.0319	0.0255	0.0272	-0.0161	0.9308
200	-1.6844	-2.3641	-1.5839	-0.8383	-0.1833	-0.0208	-0.0400	-0.0562	-0.0392	-0.0296	0.9666
400	-1.6948	-2.3860	-1.6174	-0.8604	-0.1696	-0.0076	-0.0146	-0.0230	-0.0203	-0.0103	0.9818

For model AR(1) with parameters $\alpha_1 = -0.60$ and $\sigma^2 = 1$, the table contains the averages of the estimates, and of the corresponding standard deviation. These values must be compared with $\alpha_1 = -0.60$, $\mathrm{Var}(\hat{\alpha}_1) \approx (1-\alpha^2)/T$, with numerical values 0.08, 0.056 and 0.04, respectively.

Next we find the averages corresponding to estimates of the parameters obtained when the order value $\hat{p}$ is used, and estimation is done by using Burg's algorithm. The estimation of α_1 has improved to some extent, while the averages for the other parameters are (comparatively) small. We recall that estimates are based upon the number of series indicated in Tables 1 and 2, except that α_1 is always estimated with the 1000 series.

Similar considerations correspond to the other three models. In the case of model AR(2), use of the approximation $\mathrm{Var}(\hat{\alpha}) \approx (1-\alpha_j^2)/T$ for the variance appears valid. This variance corresponds to the asymptotically normal distribution of the maximum likelihood estimator, see, for example, Brockwell and Davis (1987, page 252).

In general, changing the restriction on the chosen order has very little effect on the estimation of parameters for the chosen parameter values.

4.3. Confidence bands for the spectral density

4.3.1. Graphical illustration

In this section we show some examples of use of the techniques developed for the construction of simultaneous confidence bands. In all cases the logarithm of the spectral estimates is presented. The abscissas, which are usually given by $T/2+1$ frequency points from 0 to π, are presented in a transformed scale from 0 to 50, and labeled "Index" in the graphs. These graphs were also made by an S-PLUS function.

Figure 1. Observations were generated by simulation with the AR(1) model with parameter $\alpha = -0.60$. The graph covers only the frequencies from 0 to π (0 to 50), since the remaining values, from $-\pi$ to 0 can be omitted due to symmetry. The central dotted line is the theoretical spectral density of the AR(1) model used to generate the data, and hence corresponds to (1.6) with $\sigma^2 = 1$. The solid line is the spectral density coming from the MA(1) approximation, namely $\hat{h}^{-1}(\lambda)$ as defined in (2.3). Broken lines are the simultaneous 95% confidence bands, defined

by (2.7), one above and one below. They were computed with a sample of size $T = 100$.

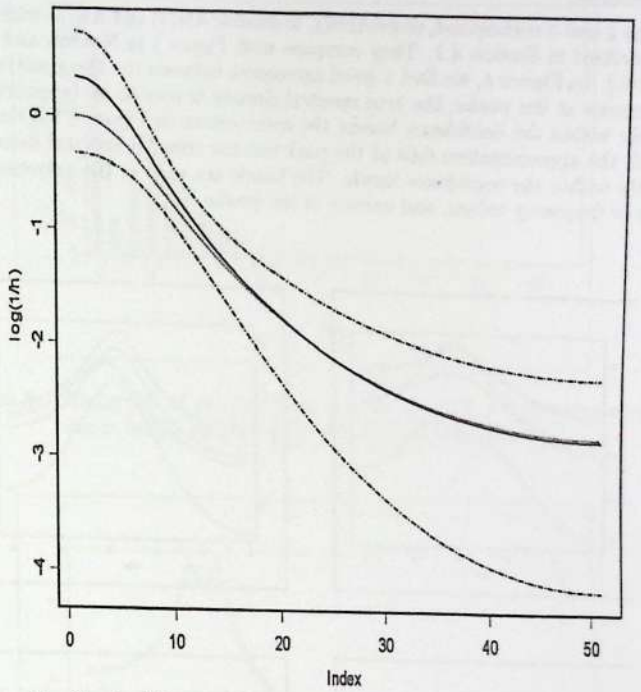


Figure 1 *Spectral density function of AR(1) model with parameter −0.60, MA(1) approximation, and 95% simultaneous confidence bands.*

Figures 2 and 3. Observations generated by simulation with the AR(2) model with parameters −0.40 and −0.45. This model was considered by Newton and Pagano. We observe the following: (1) There are 4 examples out of the 1000 replicates, (2) The theoretical spectral density and its MA(1) approximation in some case are parallel and in other have one or more crossings, (3) The theoretical spectral density in some cases falls entirely inside the confidence bands, in which case we say that there are 0 points outside the bands, and in other there one or more points outside the bands, (4) Since the bands are computed at frequencies $\lambda_j = 2\pi j/T$ for $j = 0, 1, 2, \ldots, T/2$ if T is even, we compare the number of points falling out of the bands for $j \geq 0$ with $T/2+1$, which for $T = 100$ equals 51. (5) The bar graph in Figure 3 is an example of the distribution of points falling outside the 95% confidence bands, for model AR(2) with parameters values

-0.40 and -0.45, and with 1000 replications. We find a bar of size approximately equal to 300 at 0, and then a distribution reaching 24. The average is 2.25 (Table 3); its expected number is $0.05 \times 51 = 2.55$. The shape is typical of a mixture of distributions.

Figures 4 and 5 correspond, respectively, to models AR(4) and AR(5) with parameters defined in Section 4.1. They compare with Figure 1 in Newton and Pagano (op. cit.). In Figure 4, we find a good agreement between the two spectral densities, except at the peaks; the true spectral density is seen to lie (approximately) entirely within the confidence bands; the same occurs in Figure 5 for the AR(5) model: the approximation fails at the peak but the true AR spectral density falls entirely within the confidence bands. The bands are wide at the extremes of the range of frequency values, and narrow at the peaks.

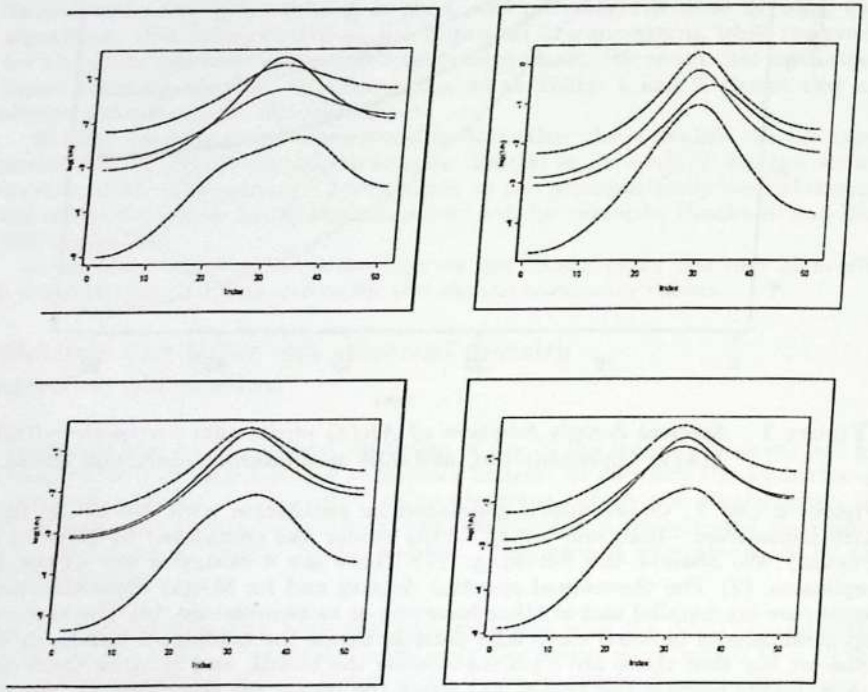


Figure 2 *Spectral density function of AR(2) model with parameter* -0.40 *and* -0.45*, MA(2) approximation, and 95% simultaneous confidence bands.*

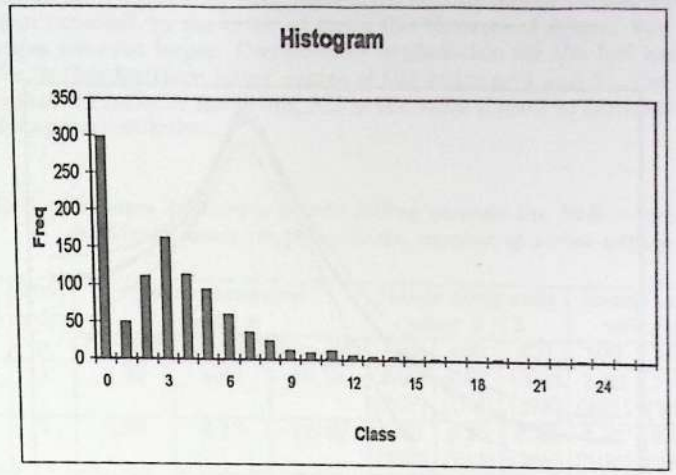


Figure 3 *Histogram of points falling outside the 95% simultaneous confidence bands, AR(2) model, AR spectral estimations.*

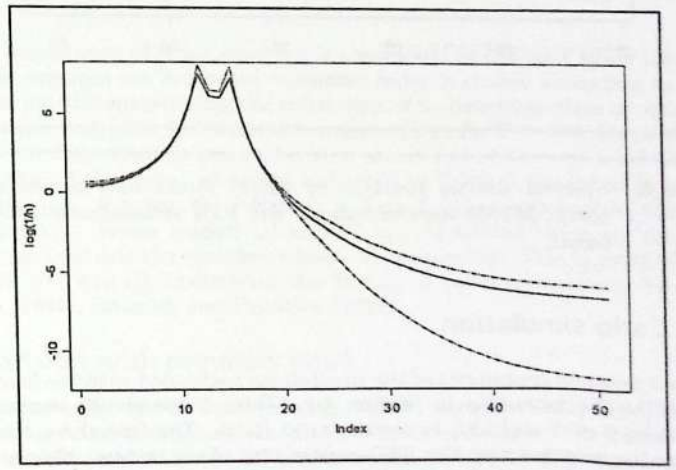


Figure 4 *Spectral density function of AR(4) model with parameter (see text), MA(4) approximation, and 95% simultaneous confidence bands.*

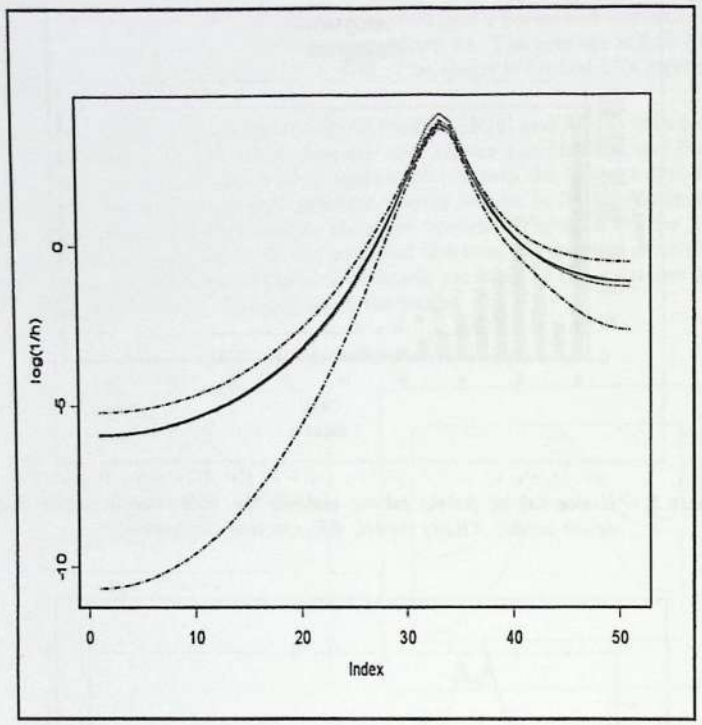


Figure 5 *Spectral density function of AR(5) model with parameter (see text), MA(5) approximation, and 95% simultaneous confidence bands.*

4.3.2. Monte Carlo simulation

Table 3 presents summaries of the simulations performed with the four models mentioned at the beginning of Section 4.1. Table 3 has results corresponding to restricting $\hat{p} \leq 5$ and also to restricting $\hat{p} \leq 10$. The first three columns of results are the averages over the 1000 replications, of the number of points of the theoretical spectral density function, that occurred outside the 95% (simultaneous) confidence band. The expected numbers are 2.55 for $T = 100$, 5.05 for $T = 200$ and 10.05 for $T = 400$. We find that for models I and II the observed (average)

values are in good agreement with these figures. For model IV the averages are larger than expected, in the order of twice the theoretical figure. For model III the averages are even larger. One possible explanation for the left hand part of the tables, is that for these larger values of the order p (4 and 5), the limitation imposed that the order of the fitting AR is the same p used to generate the data, results being too restrictive.

Table 3 *Averages of sample points falling outside the 95% simultaneous confidence bands (in parenthesis, number of series with $\hat{p} = p$).*

Model	True order p	Bands computed using p			Bands computed using $\hat{p} \leq 5$			Bands computed using $\hat{p} \leq 10$		
		$T = 100$	$T = 200$	$T = 400$	100	200	400	100	200	400
I	1	2.25	5.46	18.13	3.05 (267)	6.54 (256)	10.03 (278)	7.52 (301)	7.15 (298)	9.82 (309)
II	2	3.09	6.17	14.49	5.49 (255)	6.13 (244)	9.42 (224)	7.37 (306)	7.04 (281)	9.37 (270)
III	4	14.73	30.57	65.21	17.87 (181)	36.16 (164)	74.57 (150)	18.16 (335)	32.32 (288)	64.42 (258)
IV	5	8.99	17.73	38.01	12.47 (428)	27.95 (207)	64.82 (38)	11.87 (597)	20.29 (427)	32.09 (249)

The importance of order selection is considered in the next three columns. In them the averages are computed when the order is chosen according to the AIC criterion, but still maintaining the restriction of its being less than or equal to 5 or to 10, respectively. We find a sizeable improvement for $T = 400$, for models I and II. Here the improvement occurs because about 1/4 of the series led to changes in the value of the order, which, as indicated in Table 1, consisted in increasing the value of p. Further, for orders $p = 1$ or 2, the restriction on the order is not important. When models III and IV are considered there are more points (on average) outside the confidence band than expected. This is particularly true of model III (AR(4)), confirming the finding of previous authors, Newton and Pagano (1984), Beamish and Priestley (1981).

4.3.2. Comparision with previous work

In a paper by the authors (Mentz, Viollaz and Martínez, 2003), standard, windowed spectral estimators were studied by Monte Carlo. They were: (1) Confidence bands available in the S-PLUS package, in which the χ^2 approximation is used, with 8.29 degrees of freedom; (2) Similar bands with 9.14 degrees of freedom; (3) Confidence bands using the normal approximation, and (4) Similar bands with a correction for asymptotic bias. A summary of results is Table 4.

Table 4 *Averages of sample points falling outside the 95% simultaneous confidence bands.*

Model	True AR order	Sample size T	This study		Previous study			
			p	$\hat{p}$	(1)	(2)	(3)	(4)
II	2	100	3.09	7.37	0.716	1.171	1.962	0.84
		200	6.17	7.04	0.471	0.971	2.262	0.739
		400	14.49	9.37	0.306	0.773	2.797	0.763
III	4	100	14.73	18.16	14.401	13.954	11.669	11.081
		200	30.57	32.32	4.052	3.953	3.579	2.293
		400	65.21	20.29	0.693	0.945	2.512	0.657
IV	5	100	8.99	11.87	3.141	3.014	2.66	1.987
		200	17.73	20.29	1.451	1.679	2.423	1.2
		400	38.01	32.09	0.631	0.934	2.62	0.753

It follows that (except for Model III with $T = 100$), the bands based on windowed spectral estimates, leave, on average, less points outside the bands than the AR spectral estimates. One main point is that while the windowed estimates are based on (95%) point-by-point or marginal analysis, the bands for the AR spectral estimators have 95% joint confidence coefficient, and we expect the former to leave less points outside of them.

5. Summary and conclusions

In this paper we analyze the behavior of the AR spectral estimator in simulated time series. There exist a set of models frequently studied in the literature, namely, models AR(p) with $p = 2$, 4 and 5. There were considered by the present authors in a previous paper dealing with standard windowed spectral estimates, that is those coming from smoothing the periodogram (Mentz, Viollaz and Martínez, 2003). They are confidence intervals computed separately at each frequency with a common confidence coefficient (e.g. 95%). Hence, the level of confidence for the whole set of frequencies is considerably low. In the indicated reference point-by-point inference is compared with joint inference based on extreme value distributions (Woodroofe and Van Ness, 1967).

The bands for the AR spectral estimator developed by Newton and Pagano (op. cit.) are based on projections proposed by Scheffé (1959), and hence have a joint confidence level (e.g. 95%). This is an asymptotic approximation.

The behavior of the AR bands is studied in some detail: (1) The role of selecting the order of the approximating AR model is studied, and found that for the models considered, restricting it to be less than or equal to 10 improves over restricting it to be less than or equal to 5, in particular, for the more complicated models AR(4) and AR(5). However, it affects little the estimation of parameters. (2) Graphical examples are provided to illustrate the results. (3) Tabled results show average numbers of sample points falling outside the bands.

In summary, the AR spectral simultaneous confidence bands are seen to behave reasonably well for the models considered.

As pointed out by a referee, bootstrap estimation of the standard error of the confidence bands could be attempted, but we do not include this work in the present paper.

There has been a good deal of concern about the merits of the AR spectral estimator. See, for example, Priestley (1981, p. 612) who wrote: "It is seen that the AR estimate is highly irregular, with quite marked oscillations over the whole frequency range". Harvey (1994, p. 205) commented on the importance of selecting an adequate (estimated) AR order. See also Brockwell and Davis (1987, Section 10.6), Koopmans (1974, Section 9.4).

Acknowledgements

The authors thank two detailed referee reports and also comments by the editor.

(Received March, 2004. Accepted March, 2006.)

References

Akaike, H. (1977). An entropy maximisation principle. In *Proc. Symp. On Applied Statistics* (Ed. P.Krishnaiah). North-Holland, Amsterdam.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* New York: John Wiley.

Beamish, N. and M. B. Priestley (1981). A study of autoregressive and window spectral estimation. *Appl. Statist.*, **30**, 41–58.

Brockwell, P. J. and R. A. Davis (1987). *Time Series: Theory and Methods.* New York: Springer-Verlag.

Harvey, A. C. (1994). *Time Series Models.* Second Edition. Cambridge, Massachusetts: The MIT Press.

Koopmans, L. H. (1974). *The Spectral Analysis of Time Series.* New York: Academic Press.

Mentz, R. P., A. J. Viollaz and C. I. Martínez (2003). Confidence bands for time series problems in the time and frequency domains. *Brazilian Journal of Probability and Statistics*, **17**, 179–211.

Newton, H. J. and M. Pagano (1984). Simultaneous confidence bands for autoregressive spectra. *Biometrika*, **71**, 197–202.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Vol. 1: Univariate Series. London: Academic Press.

Scheffé, H. (1959). *The Analysis of Variance.* New York: John Wiley.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, **63**, 117–126.

Woodroofe, M. and J. Van Ness (1967). The maximum deviation of spectral densities. *Ann. Math. Stat.*, **36**, 1558–1569.

Raúl Pedro Mentz and Carlos I. Martinez

Instituto de Investigaciones Estadísticas

Facultad de Ciencias Económicas

Universidad Nacional de Tucumán and CONICET

Casilla de correo 209

4000 - San Miguel de Tucumán, Argentina

E-mail: rmentz@herrera.unt.edu.ar and cimartine@herrera.unt.edu.ar