

Tesis presentada para la obtención del grado de

MAGISTER EN ESTADISTICA APLICADA

Tema: "ESTIMACION DE REGRESION A TROZOS"

por

CHRISTINE ADRIANE ISGRO

Dirigida por

Dr. RAUL PEDRO MENTZ

Comité de Supervisión

Dr. RICARDO ANIBAL LEIVA

Dr. ALDO JOSE VIOLLAZ

UNIVERSIDAD NACIONAL DE TUCUMAN

FACULTAD DE CIENCIAS ECONOMICAS

INSTITUTO DE INVESTIGACIONES ESTADISTICAS

Octubre de 1991

## PROPOSITO

El modelo de regresión a trozos es adecuado cuando la relación entre las variables tiene distintas expresiones funcionales en partes del dominio de definición, como consecuencia del comportamiento de variables que pueden ser o no variables explicativas del modelo. En los últimos años se observa un interés creciente de los estadísticos por este tema dada la gran cantidad de situaciones donde este modelo constituye una descripción adecuada de la relación entre las variables.

En este trabajo se realiza una presentación unificada de los aportes más significativos al tema realizados hasta este momento, que permite superar algunas de las dificultades que presenta su estudio, entre las que se puede citar lo disperso de la bibliografía, las diferencias entre las notaciones utilizadas por los autores y el hecho de que en muchos trabajos se trata un problema particular del modelo de regresión a trozos sin hacer referencia a una teoría general. Todo esto hacía difícil el esfuerzo para integrar y comparar los distintos aportes además de obstaculizar la comprensión de cada problema.

En este trabajo se ha encontrado la expresión de un modelo general que permite replantear los problemas tratados en la bibliografía como casos particulares. Este modelo general incorpora a la variable que gobierna el cambio de régimen al conjunto de variables independientes, con un coeficiente que será estimado cuando pertenezca a las variables explicativas y que se supondrá igual a cero cuando no pertenezca.

Los trabajos se han clasificado de acuerdo al tipo de variable que gobierna el cambio de régimen, que puede ser una variable conocida que pertenece o no al conjunto de variables explicativas del modelo, una combinación lineal de este tipo de variables conocidas (que pertenecen o no al modelo), o, por último, una variable desconocida.

Los problemas estadísticos que presenta este modelo son la

estimación del número de fases de la regresión, de la ubicación del punto de cambio de régimen y de los restantes parámetros de la regresión y la realización de tests de hipótesis sobre los parámetros.

Este trabajo se realizó en el marco de un convenio de colaboración entre el Departamento de Estadística de la Universidad de Sevilla y el Departamento de Estadística de la Universidad de Granada. El autor agradece al Dr. Fernando Arias por su colaboración y al Dr. Juan Carlos Rodríguez por su colaboración en la realización de este trabajo. Este trabajo se realizó en el Instituto de Estadística y Matemáticas de la Facultad de Ciencias Económicas de la Universidad Nacional de Educación.

El autor desea agradecer al Sr. Juan Carlos Rodríguez por su colaboración y al Sr. Fernando Arias por su colaboración en la realización de este trabajo. El autor agradece al Dr. Juan Carlos Rodríguez por su colaboración y al Sr. Fernando Arias por su colaboración en la realización de este trabajo.

Este trabajo se realizó en el marco de un convenio de colaboración entre el Departamento de Estadística de la Universidad de Sevilla y el Departamento de Estadística de la Universidad de Granada. El autor agradece al Dr. Fernando Arias por su colaboración y al Dr. Juan Carlos Rodríguez por su colaboración en la realización de este trabajo.

Este trabajo se realizó en el marco de un convenio de colaboración entre el Departamento de Estadística de la Universidad de Sevilla y el Departamento de Estadística de la Universidad de Granada. El autor agradece al Dr. Fernando Arias por su colaboración y al Dr. Juan Carlos Rodríguez por su colaboración en la realización de este trabajo.

Este trabajo se realizó en el marco de un convenio de colaboración entre el Departamento de Estadística de la Universidad de Sevilla y el Departamento de Estadística de la Universidad de Granada. El autor agradece al Dr. Fernando Arias por su colaboración y al Dr. Juan Carlos Rodríguez por su colaboración en la realización de este trabajo.

Este trabajo se realizó en el marco de un convenio de colaboración entre el Departamento de Estadística de la Universidad de Sevilla y el Departamento de Estadística de la Universidad de Granada. El autor agradece al Dr. Fernando Arias por su colaboración y al Dr. Juan Carlos Rodríguez por su colaboración en la realización de este trabajo.

## AGRADECIMIENTOS

Este proyecto se inició como parte del plan de trabajo de mi Beca Interna de Perfeccionamiento del CONICET bajo la dirección del Dr. Ricardo Aníbal Leiva, que cumplí hasta enero de 1988 en el Centro Regional de Investigaciones Científicas y Tecnológicas-Mendoza (CRICYT-ME) del CONICET, y desde entonces y hasta su finalización en el Instituto de Investigaciones Estadísticas (INIE), en la Facultad de Ciencias Económicas de la Universidad Nacional de Tucumán.

El trabajo se debe en gran medida al estímulo y apoyo permanente del Dr. Ricardo Aníbal Leiva, con quien inicié el tratamiento del tema en Mendoza, cuando era becaria interna del CONICET. El Dr. Leiva con gran paciencia y dedicación me guió y ayudó a superar las dificultades encontradas. Ha sido un privilegio para mí poder trabajar a su lado.

Debo agradecer al Dr. Raúl Pedro Mentz por sus valiosos comentarios y sugerencias sobre una versión preliminar de la tesis que han sido de incalculable valor en su redacción definitiva.

También quiero agradecer a los Ing. Ernesto Gabino Mavrich y María Cristina Herrera que despertaron en mí el interés por la metodología estadística.

Deseo hacer extensivo mi agradecimiento a la CPN María Beatriz Ceballos y a la Lic. María Rosa Santillán, miembros del Instituto de Investigaciones Estadísticas, por el tiempo y la colaboración que me ofrecieron que me ofrecieron para la impresión de este trabajo.

Mi mayor gratitud personal es para mis padres y mi esposo que me brindaron todo su apoyo en la elaboración de esta tesis. A ellos dedico este trabajo.

### III

#### INDICE

I.	PROPOSITO .....	I-II
II.	AGRADECIMIENTOS.....	III
III.	INDICE.....	IV-VI

#### CAPITULO I

INTRODUCCION Y CONTENIDO DEL INFORME.....	1
---	---

#### CAPITULO II

##### MODELO DE REGRESION A TROZOS DE $r$ FASES CON $(r-1)$ PUNTOS DE CAMBIO.

2.1.	Expresión Matricial del Modelo de Regresión.....	6
2.2.	Modelo de Regresión a Trozos.	
2.2.1.	<i>Antecedentes</i> .....	7
2.2.2.	<i>Expresión matricial del modelo de regresión a trozos</i> .....	7

#### CAPITULO III

##### TRABAJOS EN QUE LA VARIABLE QUE GOBIERNA EL CAMBIO DE REGIMEN NO ES UNA VARIABLE EXPLICATIVA DEL MODELO.

3.1.	Estimación de la ubicación del punto de cambio de régimen.	
3.1.1.	<i>Solución propuesta por Quandt (1958-1960)</i> .....	11
3.1.2.	<i>Solución propuesta por Holbert y Broemeling (1977)</i> .....	14
3.1.3.	<i>Solución propuesta por Ferreira (1975)</i> .....	16
3.2.	Tests de hipótesis sobre la existencia de un punto de cambio en el modelo.	
3.2.1.	<i>Solución propuesta por Chow (1960)</i> .....	23
3.2.2.	<i>Solución propuesta por Quandt (1958-1960)</i> .....	28
3.2.3.	<i>Solución propuesta por Broemeling y Chin-Choy (1980)</i> .....	33
3.3.	Tests de hipótesis sobre los restantes parámetros de la regresión.	

CAPITULO V  
CASOS ATIPICOS.

5.1. La naturaleza elige uno de los dos regímenes de una regresión de acuerdo al valor de una función de una variable $Z$ , que puede ser o no una de las variables explicativas de la regresión.	
5.1.1. Solución propuesta por Goldfeld y Quandt (1972).....	125
5.2. La naturaleza elige uno de los dos regimenes de la regresión con probabilidades $\lambda$ y $1-\lambda$ respectivamente.	
5.2.1. Solución propuesta por Quandt (1972).....	128
5.2.2. Solución propuesta por Kiefer (1978).....	131
5.3. La naturaleza elige uno de los dos regímenes con una probabilidad que depende del estado anterior del sistema (proceso de Cadena de Markov).	
5.3.1. Solución propuesta por Goldfeld y Quandt (1972).....	136

CAPITULO VI  
CONCLUSIONES.

6.1. Resumen.....	140
6.2. Temas de estudio para una etapa posterior.....	143
Bibliografía.....	144

## 1. INTRODUCCION Y CONTENIDO DEL INFORME

Una de las metas de la investigación científica es encontrar, describir y analizar relaciones entre variables en el mundo real, y una de las formas de hacer este análisis es hallar funciones que describan dichas relaciones. Puede sostenerse que ésto es una abstracción matemática, ya que no existen relaciones exactas en la naturaleza. Sin embargo el concepto es importante pues aunque la relación entre las variables no sea exacta puede encontrarse a veces alguna función que la describa con gran aproximación. Esta idea origina una cantidad importante de métodos estadísticos: en ellos una variable  $Y$  está relacionada con una o más variables  $X_1, \dots, X_k$ , de manera que se considera la relación funcional

$$Y=f(X_1, \dots, X_k) . \quad (1)$$

Esta relación puede provenir de teorías propias del campo sujeto a análisis, puede ser una manera práctica de resumir una gran cantidad de información sobre un tema, o reconocer otros orígenes. Desde el punto de vista estadístico interesa el caso en que (1) no es exactamente una igualdad sino que existe una componente aleatoria de error (por ejemplo aditiva)

$$Y=f(X_1, \dots, X_k)+\varepsilon; \quad (2)$$

donde  $f(X_1, \dots, X_k)$  es una función de  $X=(X_1, \dots, X_k)'$ , que explicita la relación entre las variables, definida en un dominio  $D$ ; puede ser del tipo paramétrico que es cuando existe un conjunto de parámetros (constantes) que en el caso más sencillo son "portadores" de las variables. En general la forma funcional de  $f(X)$  es conocida pero los parámetros son desconocidos.  $X$  es la variable independiente o predictora e  $Y$  es la variable dependiente o de respuesta. Aunque la variable aleatoria  $\varepsilon$  no es observable algo sobre su distribución es a menudo establecido como parte del modelo.

Una forma funcional ampliamente usada es la lineal

$$Y = \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon \quad (3)$$

pues aún cuando ésta no sea la forma funcional verdadera si el rango de operación es pequeño puede aproximarla adecuadamente.

En el caso hipotético de que no hubiese un término de error en la ecuación, podría obtenerse una solución matemática para los parámetros con sólo tener tantas observaciones como parámetros a estimar, es decir,

$$Y = \beta_1 X_{1i} + \dots + \beta_k X_{ki}, \quad i=1,2,\dots,k. \quad (4)$$

Sin embargo en la práctica el término de error está siempre presente y su valor es desconocido por lo que deben buscarse formas alternativas para "estimar" estadísticamente los parámetros a partir de N observaciones de las variables:  $Y_1, X_{11}, \dots, X_{k1}, \dots, Y_N, X_{1N}, \dots, X_{kN}$ , de manera que tenemos

$$Y = \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i, \quad i=1,2,\dots,N. \quad (5)$$

Las N ecuaciones en (5), donde en general N es mayor que k, constituyen el modelo de regresión múltiple. El término regresión fue usado por primera vez por Francis Galton (1886) en un artículo muy famoso publicado en los Proceedings of the Royal Society.

Una restricción importante de (5) es que el conjunto de parámetros es fijo y tiene validez para todas las observaciones. Esto puede resultar muy restrictivo; si el índice i fuese el tiempo, cambios (suaves) a través del tiempo podrían implicar que si  $\beta_0, \beta_1, \dots, \beta_k$  sirven para ajustar las observaciones  $i=1,2,\dots,\eta_1$  otro conjunto distinto  $\beta'_0, \beta'_1, \dots, \beta'_k$  será necesario para el resto,  $i=\eta_1+1,\dots,N$ . Surge intuitivamente entonces la idea de un modelo de regresión a trozos que es un modelo de regresión que presenta distintas formas funcionales en distintas regiones del dominio de definición, debido a la influencia de variables que pueden pertenecer o no al modelo.

Esto se presenta por ejemplo en:

Economía: donde algunas de las variables macro-económicas sufren modificaciones debidas a los cambios estructurales en la economía ocasionados por cambios repentinos en la política

económica, así como también por desastres naturales (epidemias, inundaciones, guerras, etc.).

Tomografía computada: donde la densidad de los materiales analizados puede tener discontinuidades debidas a tumores o simplemente a la distinta composición del material estudiado.

Procesamiento de imágenes: donde la intensidad de la luz de una imagen cambia de un objeto a otro.

En los últimos años se observa un creciente interés de los estadísticos por resolver los problemas de estimación y tests que presenta este modelo pues a su indudable importancia práctica se suma el interés teórico que su estimación despierta, ya que su tratamiento requiere el desarrollo de nuevas teorías que permitan resolver los complicados problemas que en ella se presentan.

Al abordar el estudio de este modelo se deben enfrentar algunas dificultades. La primera de ellas es lo disperso de la bibliografía; las contribuciones que se han hecho hasta el momento sobre el tema se hallan publicadas en muchas revistas especializadas y tesis doctorales. Además cada una de ellas trata casos particulares del modelo de regresión a trozos, pero ninguna hace referencia a una teoría general. Hasta la notación y planteamiento del modelo es distinto en muchos casos. Esto hace difícil el esfuerzo para integrar y comparar los distintos aportes, además de obstaculizar la comprensión de cada problema. En este trabajo se realiza una presentación unificada de los principales trabajos sobre el tema. Para ello se debió hallar la expresión de un modelo general que permitiera expresar los casos tratados como casos particulares de dicho modelo general.

Esta tesis ha sido dividida en seis capítulos, donde en el capítulo II se presenta la expresión general del modelo de regresión lineal múltiple a trozos de  $r$  fases con  $(r-1)$  puntos de cambio de régimen; en el capítulo III se tratan los trabajos en que la variable que gobierna el cambio de régimen no pertenece al modelo, en el capítulo IV aquellos en que la variable que gobierna el cambio de régimen pertenece al modelo, en capítulo V aquellos en que la variable que gobierna el cambio de régimen es una combinación lineal de variables que pertenecen y no

pertencen al modelo, o es una variable desconocida y por último en el capítulo VI se presentan las conclusiones de este trabajo y se proponen posibles temas de estudio para el futuro. Además cada capítulo ha sido dividido en secciones teniendo en cuenta que el problema tratado sea de estimación o de tests de hipótesis. En el primer caso se deben considerar:(A) estimación del número de fases,(B) estimación de la ubicación de los puntos de cambio de régimen,(C) estimación de los restantes parámetros de la regresión; y en el segundo se deben considerar:(D) tests de hipótesis sobre la existencia de un punto de cambio en el modelo,(E) tests de hipótesis sobre la ubicación de los puntos de cambio de régimen, (F) tests de hipótesis sobre los restantes parámetros de la regresión.

El siguiente cuadro muestra los temas abordados en los capítulos III, IV y V.

Estructura esquemática del contenido del trabajo,  
con indicación de las áreas que han sido  
tratadas en la literatura.

Temas	Capítulo III Variable no pertenece al modelo	Capítulo IV Variable pertenece al modelo	Capítulo V Variable es combinación lineal o es desconocida.
<b>Estimación</b>			
A.Número de fases		*	*
B.Ubicación de los puntos	*	*	
C.Restantes parámetros		*	*
<b>Tests</b>			
D Existencia de puntos de cambio	*		*
E.Ubicación de los puntos		*	
F.Restantes parámetros	*	*	*

En cada sección se han clasificado los trabajos teniendo en

cuenta el número de variables explicativas y el número de fases del modelo.

CON DOS EJEMPLOS DE CÁLCULO

### 3.1. Ejemplo numérico del método de mínimos cuadrados

En este ejemplo se muestra la aplicación del método de mínimos cuadrados al estudio de un fenómeno económico. Para ello se han tomado los datos de producción del sector de maquinaria de España, en unidades de miles de toneladas, durante el período 1950-1960.

Tabla 1. Datos de producción del sector de maquinaria de España (1950-1960)

Tabla 1. Datos de producción del sector de maquinaria de España (1950-1960). Se muestra un extracto de los datos. El vector de variables explicativas es  $X = (X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10})'$  y el vector de dependientes es  $Y = (Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8, Y_9, Y_{10})'$ . El modelo a estimar es  $Y = X\beta + \epsilon$ , donde  $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10})'$  es el vector de parámetros a estimar y  $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6, \epsilon_7, \epsilon_8, \epsilon_9, \epsilon_{10})'$  es el vector de errores. Los datos de la tabla 1 se han utilizado para calcular el vector de mínimos cuadrados  $\hat{\beta}$  y el vector de errores  $\hat{\epsilon}$ . El vector de mínimos cuadrados  $\hat{\beta}$  se calcula mediante el método de mínimos cuadrados ordinarios (MCO) y el vector de errores  $\hat{\epsilon}$  se calcula mediante el método de mínimos cuadrados ordinarios (MCO).

El método de mínimos cuadrados ordinarios (MCO) es el método más utilizado para la estimación de los parámetros de un modelo de regresión lineal. Este método consiste en encontrar el vector de parámetros  $\hat{\beta}$  que minimiza la suma de los cuadrados de los errores  $\hat{\epsilon}$ .

El método de mínimos cuadrados ordinarios (MCO) es el método más utilizado para la estimación de los parámetros de un modelo de regresión lineal. Este método consiste en encontrar el vector de parámetros  $\hat{\beta}$  que minimiza la suma de los cuadrados de los errores  $\hat{\epsilon}$ .

Este método se aplica a los datos de la tabla 1 para estimar los parámetros del modelo de regresión lineal.

2. MODELO DE REGRESION A TROZOS DE  $r$  FASES  
 CON  $(r-1)$  PUNTOS DE CAMBIO

2.1. Expresión Matricial del Modelo de Regresión

En esta sección se recordará la expresión matricial del modelo de regresión múltiple para hacer más sencilla la presentación del modelo de regresión a trozos. En general el modelo puede expresarse:

$$Y = X\beta + \varepsilon, \quad (2.1)$$

donde  $Y = (Y_1, \dots, Y_N)'$  es un vector aleatorio observable,  $X = (x'_1, x'_2, \dots, x'_N)'$  es una matriz de números fijos observables,  $x'_i$  es el vector de observaciones realizadas en cada individuo,  $\beta = (\beta_0, \dots, \beta_k)'$  es un vector de parámetros no observable definido en el espacio paramétrico y  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)'$  es un vector aleatorio, no observable, que en general se supone con distribución normal  $N(0, \sigma^2 I_N)$  donde  $N$  es el tamaño de la muestra. Note que los vectores ó matrices representadas por letras griegas ( $\beta$ ,  $\varepsilon$ , etc.) no están en "negrita"; sin embargo el contexto aclarará en cada caso cómo se usa el símbolo.

La expresión matricial del modelo de regresión múltiple para  $(k+1)$  regresores y  $N$  observaciones es:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} X_{10} & \dots & X_{1k} \\ \vdots & & \vdots \\ X_{N0} & \dots & X_{Nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix} \quad (2.2)$$

con  $\varepsilon$  que se distribuye  $N(0, \sigma^2 I_N)$ .

## 2.2. Modelo de Regresión a Trozos

### 2.2.1. Antecedentes

En una gran cantidad de situaciones existen razones para suponer que la relación cambia en uno o más puntos del dominio de definición por influencia de una o más variables que pueden estar o no entre las variables explicativas de la regresión. El modelo de regresión a trozos postula que la relación entre la variable dependiente  $Y$  y la variable independiente  $X'=(X_0, X_1, \dots, X_k)'$  sigue  $r$  regímenes distintos ( $r=2,3,4,\dots$ ) en el rango de variación de  $X$ , es decir que existen  $(r-1)$  puntos de cambio de régimen.

Los primeros trabajos que hemos considerado en esta tesis pertenecen a Quandt y fueron editados en 1958 y 1960. Desde esa fecha hasta la actualidad la cantidad de publicaciones sobre el tema es considerable; al abordar su estudio se encuentra entre otras las siguientes dificultades: la diversidad de notaciones utilizadas, la falta de fundamentación de los métodos propuestos como soluciones a los problemas estadísticos tratados (sobre todo en los primeros trabajos sobre el tema), el tratamiento separado del caso en que la variable que gobierna el cambio de régimen pertenece o no al modelo, el tratamiento separado de cada caso particular sin hacer referencia a una teoría general, la gran cantidad de problemas de inferencia estadística que se presentan y por último la comparación de las soluciones dadas por distintos autores para resolver problemas de estimación o de tests de hipótesis que presenta el modelo.

### 2.2.2. Expresión matricial del modelo de regresión a trozos

Como ya señalamos el principal objetivo de este trabajo es realizar una presentación unificada del estado actual del

conocimiento sobre el modelo de regresión a trozos que minimice los problemas mencionados en la sección anterior. Para lograrlo fue necesario en primer lugar encontrar una definición general del modelo de regresión a trozos que permita tratar todos los casos particulares citados en la bibliografía como un caso particular del mismo, y de esta forma integrar los aportes de los distintos autores y además unificar y simplificar el estudio del tema.

Una forma sencilla de solucionar este problema es el de incorporar a la variable que gobierna los cambios de régimen entre las variables independientes de la regresión, con un coeficiente que será estimado en el caso de que ella sea una variable que pertenece al modelo y se supondrá igual a cero en el caso que ella no pertenezca. En todo el desarrollo de la tesis se indica con  $X_0$  a esta variable y  $\beta_0$  a su coeficiente. El modelo muestral puede expresarse mediante el siguiente sistema de ecuaciones:

$$\begin{aligned}
 Y_i &= \beta_{01} X_{0i} + \beta_{11} X_{1i} + \dots + \beta_{k1} X_{ki} + \varepsilon_i & \text{si } X_{0i} \leq \gamma_1; i = 1, 2, \dots, n_1 \\
 Y_i &= \beta_{02} X_{0i} + \beta_{12} X_{1i} + \dots + \beta_{k2} X_{ki} + \varepsilon_i & \text{si } \gamma_1 < X_{0i} \leq \gamma_2; i = n_1 + 1, \dots, n_1 + n_2 \\
 &\vdots \\
 Y_i &= \beta_{0r} X_{0i} + \beta_{1r} X_{1i} + \dots + \beta_{kr} X_{ki} + \varepsilon_i & \text{si } X_{0i} \geq \gamma_{r-1}; \\
 & & i = n_1 + \dots + n_{r-1} + 1, \dots, N.
 \end{aligned}$$

En general se supone que los  $\varepsilon_i$  son variables aleatorias independientes idénticamente distribuidas (vaid) normales  $(0, \sigma_i^2)$ , donde la varianza del error es constante para todas las observaciones que pertenecen a la misma fase de la regresión. Si  $X_0$  no está entre las variables explicativas de la regresión entonces  $\beta_{0j} = 0$  para  $j = 1, \dots, r$ . La expresión matricial del modelo es:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_r \end{bmatrix} = \begin{bmatrix} X_1 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & X_r \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_r \end{bmatrix} \quad (2.3)$$

donde  $\varepsilon' = (\varepsilon'_1, \dots, \varepsilon'_r) = (\varepsilon_1, \dots, \varepsilon_N)'$  y se distribuye normal  $(0, \Sigma)$  con

$$\Sigma = \begin{bmatrix} \Sigma_1 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & \Sigma_r \end{bmatrix} \quad (2.4)$$

y  $\Sigma_j = \sigma^2 I_{n_j}$ .  $j$  indica la fase de la regresión y varía entre 1 y  $r$ ,  $r$  es el número de fases del modelo,  $n_j$  es el número de observaciones de la  $j$ -ésima fase de la regresión y  $N$  es el número de observaciones de la muestra. ( $N = \sum_{j=1}^r n_j$ ).  $Y_j(n_j \times 1)$  es el vector de variables dependientes de la  $j$ -ésima fase de la regresión,  $X_j(n_j \times (k+1))$  es la matriz de variables independientes de la  $j$ -ésima fase de la regresión,  $\beta_j((k+1) \times 1)$  es el vector de coeficientes de la  $j$ -ésima fase y  $k$  es el número de variables explicativas en cada fase de la regresión.

A veces existe más de un grupo de observaciones, digamos  $g$  grupos, donde cada una de ellas sigue un régimen de  $r$  fases (por ejemplo observaciones pertenecientes a distintos países, a distintos periodos de tiempo o a distintas unidades económicas); y el modelo se generaliza de la siguiente forma:

### 3. TRABAJOS EN QUE LA VARIABLE QUE GOBIERNA EL CAMBIO DE REGIMEN NO PERTENECE A LAS VARIABLES EXPLICATIVAS DEL MODELO.

#### 3.1. Estimación de la ubicación del punto de cambio de régimen.

##### 3.1.1. Solución propuesta por Quandt (1958-1960)

El primer autor que trata este problema es Quandt en dos trabajos publicados en 1958 y 1960. Estos son los más citados en la bibliografía y los primeros a los cuales nosotros haremos referencia. El considera el caso particular del modelo de regresión lineal con una variable explicativa donde la variable que determina el cambio de régimen es el tiempo, y se sabe de antemano que en el periodo considerado sólo hay un cambio de régimen; pero la solución también es válida para cualquier tipo de variable. Luego el modelo tiene la siguiente expresión:

$$\begin{aligned} Y_i &= \beta_{11} + \beta_{21} X_i + \varepsilon_i && \text{si } i \leq \tau, \quad i=1,2,\dots,\tau \\ Y_i &= \beta_{12} + \beta_{22} X_i + \varepsilon_i && \text{si } i > \tau, \quad i=\tau+1,\dots,N ; \end{aligned}$$

donde  $i$  es la variable ordenadora que determina el cambio de régimen,  $N$  es el tamaño de la muestra y  $\tau$ , que recibe el nombre de switch-point o shift-point, es el valor de la variable ordenadora donde se produce el cambio de régimen. El autor supone que los datos están libres de errores de observación y que los términos de error son independientes entre sí e independientes de la variable explicativa. En forma matricial:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad (3.1)$$

donde la matriz de residuos se distribuye normal  $N(0, V_\tau)$ , con

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} (X'_{1\tau} X_{1\tau})^{-1} X'_{1\tau} Y_{1\tau} \\ (X'_{2\tau} X_{2\tau})^{-1} X'_{2\tau} Y_{2\tau} \end{bmatrix} \quad (3.6)$$

y los estimadores MV de la varianza del error en cada fase de la regresión son respectivamente:

$$\begin{aligned} \tilde{\sigma}_{1\tau}^2 &= \frac{(Y_{1\tau} - X_{1\tau} \hat{\beta}_{1\tau})' (Y_{1\tau} - X_{1\tau} \hat{\beta}_{1\tau})}{\tau} \\ \tilde{\sigma}_{2\tau}^2 &= \frac{(Y_{2\tau} - X_{2\tau} \hat{\beta}_{2\tau})' (Y_{2\tau} - X_{2\tau} \hat{\beta}_{2\tau})}{N - \tau} \end{aligned} \quad (3.7)$$

$\tilde{\sigma}_{1\tau}^2$  y  $\tilde{\sigma}_{2\tau}^2$  son estimadores sesgados por lo que se realiza una corrección para convertirlos en estimadores insesgados,

$$\begin{aligned} \hat{\sigma}_{1\tau}^2 &= \frac{(Y_{1\tau} - X_{1\tau} \hat{\beta}_{1\tau})' (Y_{1\tau} - X_{1\tau} \hat{\beta}_{1\tau})}{\tau - 2} \\ \hat{\sigma}_{2\tau}^2 &= \frac{(Y_{2\tau} - X_{2\tau} \hat{\beta}_{2\tau})' (Y_{2\tau} - X_{2\tau} \hat{\beta}_{2\tau})}{N - \tau - 2} \end{aligned} \quad (3.8)$$

Luego de repetir este proceso para  $\tau = 3, 4, 5, \dots, N-3$  se elige como estimador MV de  $\tau$  a aquel para el cual  $L_\tau$  alcanza su máximo valor.

Como ya dijimos antes Quandt sólo considera el caso del modelo con una variable explicativa, pero la solución propuesta por él es fácilmente extensible al caso general de  $k$  variables explicativas. En este caso el modelo tiene la siguiente expresión:

$$\begin{aligned} Y_i &= \beta_{11} + \beta_{21} X_{2i} + \dots + \beta_{k1} X_{ki} + \varepsilon_i; & i=1, 2, \dots, \tau \\ Y_i &= \beta_{12} + \beta_{22} X_{2i} + \dots + \beta_{k2} X_{ki} + \varepsilon_i; & i=\tau+1, \dots, N. \end{aligned}$$

Su expresión matricial coincidirá con (3.1) y (3.2).

Como en el caso anterior los parámetros que deben estimarse en este modelo son  $\beta_1, \beta_2, \sigma_1^2, \sigma_2^2$  y  $\tau$ . Se calcula el valor de la función de verosimilitud de la muestra para todos los posibles valores de  $\tau$  y se selecciona como estimador MV de  $\tau$  a aquel para el cual  $L_\tau$  alcance su máximo valor. ( $\tau = k+1, \dots, N-(k+1)$ ). Para cada valor de  $\tau$  los estimadores MV del resto de los parámetros se encuentran en la forma usual y serán función del mismo.

### 3.1.2. Solución propuesta por Holbert y Broemeling (1977).

Holbert y Broemeling, en un trabajo publicado en 1977, tratan también el caso del modelo de regresión lineal con una variable explicativa, pero desde un punto de vista bayesiano. Ellos suponen que los  $X_i$  son variables regresoras conocidas, no estocásticas, ordenadas por el valor de la variable que determina el cambio de régimen y que la varianza del error es la misma para ambas fases de la regresión; ésto es  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ .

Las expresiones de las funciones de densidad a priori de los parámetros son:

$$\pi_0(\sigma^2) = \begin{cases} \alpha \frac{1}{\sigma^2} & , 0 < \sigma^2 < \infty \\ 0 & , \text{en cualquier otro punto} \end{cases} ; \quad (3.9)$$

$$\pi_0(\tau) = \begin{cases} \frac{1}{N-3} & \tau = 2, 3, \dots, N-2 \\ 0 & \text{en cualquier otro punto} \end{cases} . \quad y \quad (3.10)$$

$$\pi_0(\beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}) \propto \text{cte} , \quad (3.11)$$

con  $\beta_{11}, \beta_{21}, \beta_{12}$  y  $\beta_{22}$  que varían entre menos infinito e infinito.

La función de verosimilitud es:

$$L(\tau, \sigma^2, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}) = \frac{1}{(2\pi \sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} (Y_\tau - X_\tau \beta)' (Y_\tau - X_\tau \beta)\right\}, \quad (3.12)$$

que si se combina con la función de densidad conjunta a priori de los parámetros se obtiene:

$$\pi_1(\tau, \sigma^2, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}) \propto (\sigma^2)^{-\frac{(N+1)}{2}} \exp\left\{-\frac{1}{2\sigma^2} (Y_\tau - X_\tau \beta)' (Y_\tau - X_\tau \beta)\right\}; \quad (3.13)$$

que es la expresión de la densidad conjunta a posteriori.

Sea  $\hat{\beta}_\tau = (X_\tau' X_\tau)^{-1} X_\tau' Y_\tau$  el estimador MC de  $\beta$  cuando el shift-point es igual a  $\tau$  (el subíndice  $\tau$  indica, como en la sección anterior, que el cambio de régimen se produce en  $\tau$ ) y

$$(X_\tau - X_\tau \beta)' (Y_\tau - X_\tau \beta) = (Y_\tau - X_\tau \hat{\beta}_\tau)' (Y_\tau - X_\tau \hat{\beta}_\tau) + (\beta - \hat{\beta}_\tau)' X_\tau' X_\tau (\beta - \hat{\beta}_\tau)$$

la suma de cuadrados residual cuando el cambio de régimen se produce en  $\tau$ , que si se reemplaza en (3.13)

$$\begin{aligned} \pi_1(\tau, \sigma^2, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}) \propto (\sigma^2)^{-\frac{(N+1)}{2}} \exp\left\{-\frac{1}{2\sigma^2} (Y_\tau - X_\tau \hat{\beta}_\tau)' (Y_\tau - X_\tau \hat{\beta}_\tau) + \right. \\ \left. + (\beta - \hat{\beta}_\tau)' X_\tau' X_\tau (\beta - \hat{\beta}_\tau)\right\} \\ \alpha(\sigma^2)^{-\frac{(N+1)}{2}} \exp\left\{-\frac{1}{2\sigma^2} (R_\tau + (\beta - \hat{\beta}_\tau)' X_\tau' X_\tau (\beta - \hat{\beta}_\tau))\right\} \end{aligned} \quad (3.14)$$

$$\tau = 2, \dots, N-2, \quad 0 \leq \sigma^2 < \infty \text{ y } -\infty < \beta_{jk} < \infty, \quad j=1,2; k=1,2.$$

Integrando respecto a  $\sigma$  y a  $\beta$  se obtiene la función de densidad a posteriori de  $\tau$ :

$$\begin{aligned}
 \pi_1(\tau) &\propto \int_{-\infty}^{\infty} \int_0^{\infty} \sigma^{\left[-\frac{N}{2}+1\right]} \exp\left\{\left(-\frac{1}{2\sigma^2}\right) \left[R_\tau + (\beta - \hat{\beta}_\tau) X'_\tau X_\tau (\beta - \hat{\beta}_\tau)\right]\right\} d\sigma d\beta \\
 &\propto |X'_\tau X_\tau|^{(-1/2)} (R_\tau)^{-\left(\frac{N-4}{2}\right)} \\
 &\propto \left[\tau S_1^2 (N-\tau) S_2^2\right]^{-(1/2)} (R_\tau)^{-\left(\frac{N-4}{2}\right)} ; \tau=2,3,\dots,N-2. \quad (3.15)
 \end{aligned}$$

donde  $S_1^2$  es la  $\text{Var}(X)$  en la primera fase de la regresión y  $S_2^2$  en la segunda.

Holbert y Broemeling aplican este método a los datos de Quandt (1958). La función de densidad a posteriori  $\pi_1(\tau)$  alcanza su máximo valor para  $\tau=12$ , lo cual coincide con el valor del parámetro del modelo a partir del cual se generaron los datos. Los autores ofrecen copia del programa con el cual se hicieron los cálculos sólo con pedirlo. Este método es fácil de aplicar, no necesi<sup>ta</sup> teoría asintótica y además los resultados obtenidos son exactos aun para tamaños de muestra moderado.

### 3.1.3. Solución propuesta por Ferreira (1975)

En los enfoques no bayesianos el problema de establecer intervalos de confianza y tests de hipótesis es muy complicado. Las distribuciones de probabilidad involucradas son difíciles y su cómputo no es simple, por esta razón últimamente se ha dado mucha atención al estudio de los métodos inferenciales asintóticos. Ferreira analiza las dificultades de cómputos en la aplicación de métodos bayesianos y compara el sesgo medio y el error cuadrático medio del estimador bayesiano de  $\tau$  correspondiente a tres funciones de densidad a priori diferentes y del estimador máximo verosímil mediante el método de Monte Carlo. El supone que la matriz regresora  $X$  es no estocástica

donde los valores observados de la variable independiente son valores fijos bajo el control del experimento; y que las densidades a priori de  $\beta_{11}, \beta_{21}, \beta_{12}$  y  $\beta_{22}$  y del  $\log \sigma^2$  son independientes y uniformes en todo el rango de variación de cada uno de los parámetros y que la densidad a priori de  $\tau$  tiene la siguiente forma:

$$\pi_0(\tau) = \begin{cases} 0 & \text{si } \tau = 1, N-1, N \\ \pi_0^*(\tau) & \text{si } 1 < \tau < N-1 \end{cases} \quad (3.16)$$

Ferreira usa tres funciones de densidad a priori  $\pi_0(\tau)$  distintas, con la única restricción, como se deduce de (3.16), de que no haya una observación en el primer régimen y sólo una o ninguna observación en el segundo régimen.

Sea  $\theta = (\tau, \sigma^2, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22})$  el vector de parámetros,

$$L(Y/X, \theta) = \frac{1}{(2\pi \sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}(Y_\tau - X_\tau \beta)'(Y_\tau - X_\tau \beta)\right\} \quad (3.17)$$

la función de verosimilitud conjunta de la muestra y

$$\pi_1(\theta) \propto \frac{L(Y/X, \theta) \pi_0(\tau)}{\sigma^2} ; \quad \tau = 2, 3, \dots, N-2 \quad (3.18)$$

la función de densidad a posteriori de los parámetros. Bajo los supuestos enunciados las funciones de densidad marginal a posteriori de cada uno de los parámetros tendrán las siguientes expresiones:

$$\pi_1(\tau/X, Y) \propto \left[ \tau(N-\tau) S_{x_1}(\tau) S_{x_2}(\tau) \right]^{-\frac{1}{2}} (S^2(\tau))^{-\frac{(N-4)}{2}} \pi_0(\tau), \quad \tau = 2, 3, \dots, N-2. \quad (3.19)$$

$$\pi_1(\beta_{21}/X, Y) \propto \sum_{\tau=2}^{N-2} \left[ 1 + \frac{(\beta_{11} - S_{XY1}(\tau)/S_{X1}(\tau))^2}{S^2/S_{X1}(\tau)} \right]^{-\frac{N-3}{2}} \frac{\pi_0(\tau)}{(S^2(\tau)/S_{X1}(\tau))^{1/2}}, \quad (3.20)$$

$$\pi_1(\beta_{11}/X, Y) \propto \sum_{\tau=2}^{N-2} \left[ 1 + \frac{(\beta_{21} - Y_1(\tau))^2}{S^2(\tau)/\tau} \right]^{-\frac{N-3}{2}} \frac{\pi_0(\tau)}{(S^2(\tau)/\tau)^{1/2}}, \quad (3.21)$$

$$\pi_1(\sigma^2/X, Y) \propto \sum_{\tau=2}^{N-2} \left[ \frac{S^2(\tau)}{\sigma^2} \right]^{\frac{N-4}{2}} \exp\left[-\frac{S^2(\tau)}{2\sigma^2}\right] \pi_0(\tau)/\sigma^2 \quad (3.22)$$

donde  $S_{X1}(\tau)$  es la varianza de la variable independiente en la primera fase de la regresión y  $S_{X2}(\tau)$  en la segunda fase de la regresión,  $S_{Y1}(\tau)$  e  $\bar{Y}_1(\tau)$  son respectivamente la varianza y la media de la variable independiente en la primera fase de la regresión,  $S_{XY1}(\tau)$  es la covarianza de  $(X, Y)$  en la primera fase de la regresión y análogamente, para  $S_{Y2}(\tau)$  y  $S_{XY2}(\tau)$ .  
 $S^2(\tau) = S_{Y1}(\tau) - S_{XY1}(\tau)/S_{X1}(\tau)$ ,  $S^2_2(\tau) = S_{Y2}(\tau) - S_{XY2}(\tau)/S_{X2}(\tau)$  y  
 $S^2_1(\tau) = S^2_{Y1}(\tau) + S^2_2(\tau)$ .

Las esperanzas de estas distribuciones son, respectivamente:

$$E(\beta_{21}/X, Y) = \sum_{\tau=2}^{N-2} (S_{XY1}(\tau)/S_{X1}(\tau)) \pi_0(\tau) / \quad (3.23)$$

$$E(\beta_{11}/X, Y) = \sum_{\tau=2}^{N-2} \bar{Y}_1(\tau) \pi_0(\tau) \quad \gamma \quad (3.24)$$

$$E(\sigma^2/X, Y) = \frac{1}{N-6} \sum_{\tau=2}^{N-2} S^2(\tau) \pi_0(\tau). \quad (3.25)$$

$\frac{S_{XY1}(\tau)}{S_{X1}(\tau)}$  y  $\bar{Y}_1(\tau)$  son los estimadores MV de  $\beta_{21}$  y  $\beta_{11}$  cuando  $t=\tau$ .

(3.23) y (3.24) pueden interpretarse como la media de los

estimadores MV de  $\beta_{21}$  y  $\beta_{11}$  cuando se sabe que hay  $\tau$  puntos en el primer régimen ponderando cada valor mediante la función de densidad a priori.

Pueden hallarse expresiones análogas para  $\pi_1(\beta_{22}/X, Y)$ ,  $\pi_1(\beta_{12}/X, Y)$ ,  $E(\beta_{22}/X, Y)$  y  $E(\beta_{12}/X, Y)$ .

Ferreira realiza un experimento muestral para comparar el comportamiento de tres estimadores bayesianos del punto de cambio de régimen  $\tau$  correspondiente a tres distribuciones a priori diferentes y a una función de pérdida cuadrática. Las funciones usadas por él son las siguientes:  $\pi_{01}(\tau) \propto 1$  (uniforme);  $\pi_{02}(\tau) \propto [\tau(N-\tau)]^{1/2}$  y  $\pi_{03}(\tau) \propto [\tau(N-\tau) S_{X1}(\tau) S_{X2}(\tau)]^{1/2}$  con  $\tau$  que varía entre 2 y  $N-2$ .

Las tres cumplen que  $\pi_0(\tau) = 0$  si  $\tau=1, N-1, N$ . La simulación fue llevada a cabo eligiendo como ecuación del primer régimen la usada por Quandt (1958). Cada repetición consistió en veinte observaciones correspondiente a las abscisas  $X=1, 2, \dots, 20$ . Para  $\tau=5; 10$  y  $15$  se introdujeron cambios en la pendiente entre  $\delta=-1,8$  y  $\delta=1,8$  con incrementos de  $0,1$ . Para cada posición y magnitud del cambio se simularon 100 experimentos y en cada caso se estimaron los parámetros usando las tres funciones a priori de  $\tau$  ya descriptos y los estimadores MV de los parámetros del modelo lo que permitió calcular el sesgo medio y el error cuadrático medio. Se reproducen a continuación las tablas del sesgo y del error cuadrático medio de los cuatro estimadores del punto de cambio de régimen  $\tau$ .

Tabla 1: Sesgo medio de los cuatro estimadores de  $\tau$  para  $\delta=-1,8, \dots, 1,8$  y  $\tau=5, 10$  y  $15$ .

Shift $\delta$	B1	B2	B3	ML
-1.8	-0.44	-0.57	-0.03	-0.23
	-0.67	-0.69	-0.63	-0.77
	-0.52	-0.47	-0.72	-0.63
-1.6	-0.26	-0.46	0.24	-0.06
	-0.36	-0.37	0.33	-0.36
	-0.47	-0.35	-0.78	-0.74

Shift $\delta$	B1	B2	B3	$V_L$
-1.4	0.0 -0.40 -0.01	-0.14 -0.51 -0.01	0.43 -0.44 -0.03	0.32 -0.57 -0.04
-1.2	0.14 -0.44 -0.50	-0.04 -0.45 -0.46	0.70 -0.38 -0.98	0.37 -0.38 -0.65
-1.0	0.36 -0.29 -1.15	0.11 -0.33 -0.95	1.04 -0.21 -1.68	0.89 -0.16 -1.60
-.8	1.68 -0.42 -1.30	1.44 -0.44 -1.12	2.17 -0.38 -1.90	1.84 -0.27 -1.35
-.6	2.13 -0.40 -2.17	1.07 -0.40 -2.03	2.69 -0.34 -2.60	2.57 -0.25 -1.89
-.4	3.08 -0.04 -3.70	4.00 -0.07 -3.63	4.06 0.08 -3.00	3.79 0.43 -3.38
-.2	4.00 0.00 -4.61	5.00 0.0 -4.41	4.00 0.12 -4.71	4.62 0.24 -5.13
0	4.00 -0.15 -6.05	4.94 -0.17 -6.09	4.03 -0.21 -6.07	5.40 -0.24 -4.98
.2	4.60 -0.09 -4.98	4.53 -0.14 -4.99	4.60 -0.02 -4.99	5.12 -0.41 -5.11
.4	5.51 -0.30 -3.46	3.85 -0.29 -3.42	4.25 -0.37 -3.75	4.62 -0.89 -4.07
.6	2.73 -0.25 -2.08	2.62 -0.29 -1.94	3.15 -0.20 -2.01	3.21 -0.07 -2.63
.8	1.32 -0.57 -1.54	1.17 -0.63 -1.33	1.80 -0.52 -2.11	1.76 -0.42 -1.86
1.0	0.44 -0.40 -1.33	0.21 -0.41 -1.19	1.14 -0.32 -1.85	0.52 -0.25 -1.6
1.2	0.28 -0.49 -0.55	0.08 -0.53 -0.45	0.85 -0.44 -0.97	0.23 -0.22 -0.22
1.4	-0.36 -0.55 -0.63	-0.49 -0.57 -0.44	0.21 -0.54 -0.80	0.14 -0.71 -0.47
1.6	-0.34 -0.50 -0.60	-0.53 -0.62 -0.51	0.15 -0.59 -0.81	-0.02 -0.51 -0.82
1.8	-0.43 -0.55 -0.47	-0.54 -0.53 -0.38	0.0 -0.62 -0.73	-0.86 -0.41 -0.71

Tabla 2: Error cuadrático medio de los cuatro estimadores de  $\tau$  para  $\delta = -1,8, \dots, 1,8$  y  $\tau = 5, 10$  y  $15$ .

$\delta$	B1	B2	B3	ML
-1,8	1.10 1.15 1.00	1.21 1.19 0.95	0.91 1.09 1.20	1.81 2.29 1.81
-1,6	1.08 1.02 1.37	1.30 1.03 1.30	1.16 0.89 1.72	2.00 1.70 3.34
-1,4	2.32 1.04 1.55	2.46 1.13 1.41	2.15 0.98 1.97	3.38 2.59 3.88
-1,2	2.42 1.98 1.76	2.38 2.03 1.58	2.76 1.58 2.44	5.19 3.96 3.15
-1,0	2.94 2.17 4.84	3.21 2.41 4.61	3.90 1.75 5.98	8.83 5.34 10.26
-0,8	10.80 3.86 6.50	10.76 4.14 6.46	10.65 2.74 7.12	17.52 7.63 10.37
-0,6	10.91 4.08 13.87	10.83 4.62 14.44	11.85 2.80 13.32	22.75 8.51 17.99
-0,4	23.90 6.08 23.42	25.32 6.71 24.25	21.42 4.08 22.40	32.95 16.95 30.46
-0,2	33.88 9.79 28.29	35.08 10.54 28.37	29.67 5.98 27.85	43.84 19.08 48.19
0	31.80 7.23 33.09	33.12 8.63 34.08	28.55 4.33 30.89	51.61 22.26 43.74
0,2	29.00 9.21 33.52	29.05 11.18 34.77	26.34 5.74 30.73	46.47 19.89 48.55
0,4	23.65 7.28 22.04	24.31 8.13 22.68	23.45 4.57 20.87	41.52 15.76 35.09
0,6	15.89 5.25 9.78	16.54 6.07 10.08	15.35 3.56 11.11	26.15 10.95 17.31
0,8	7.86 4.21 7.42	8.21 4.85 7.17	8.13 3.22 7.97	13.64 8.70 12.76
1,0	4.08 2.18 6.33	4.23 2.19 5.97	4.52 1.62 6.79	7.98 4.14 10.76
1,2	2.64 1.49 2.43	2.50 1.49 2.39	3.49 1.20 2.63	7.29 2.70 4.30

Shift $\delta$	B1	B2	B3	ML
1.4	1.80	1.83	1.63	3.40
	1.35	1.43	1.14	2.68
	1.73	1.60	2.00	3.35
1.6	1.44	1.65	1.41	2.80
	1.26	1.30	1.16	2.35
	1.56	1.51	1.71	3.04
1.8	0.97	1.06	0.64	1.38
	0.99	0.99	0.92	1.43
	1.41	1.32	1.69	3.16

La conclusión más notable que puede extraerse de esta simulación es que el ECM es uniformemente más grande para el estimador MV que para los tres estimadores bayesianos y que el sesgo medio de los cuatro estimadores es comparable. Este experimento muestra la solidez del método sólo respecto de estas tres funciones a priori, pero no brinda un criterio para la elección de la misma, sólo los compara.

### 3.2. Tests de hipótesis sobre la existencia de un punto de cambio en el modelo.

#### 3.2.1. Solución propuesta por Chow (1960)

Chow considera el caso en que dado un conjunto  $n_1$  de observaciones se desea averiguar si  $n_2$  observaciones adicionales pertenecen a la misma regresión que la primera. Es decir si dos conjuntos de observaciones pueden ser tratados como pertenecientes al mismo modelo de regresión; estadísticamente esto es equivalente a testar la hipótesis nula  $\beta_1 = \beta_2$  versus la hipótesis alternativa  $\beta_1 \neq \beta_2$ . Esto permite dar respuesta a preguntas como: ¿permanece la relación estable para dos periodos de tiempo?, o, ¿se mantiene la misma relación para dos grupos de unidades económicas?. Para poder hacer progresos analíticos el autor supone que la varianza del error es la misma para ambas fases de la regresión, esto es  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ , y que el vector de residuos  $\varepsilon$  se distribuye normal  $(0, \sigma^2 I_N)$ . Utiliza el método de los mínimos cuadrados. El define una variable  $d_i = Y_i - x_i' \hat{\beta}_1$  para las observaciones de la segunda muestra,  $i = n_1 + 1, \dots, n_1 + n_2$ , pudiéndose expresar el vector  $d$  de diferencias de  $Y_2$  con los valores de predicción de la primera muestra como:

$$d = Y_2 - X_2 \hat{\beta}_1 = X_2 \beta_2 - X_2 \beta_1 + \varepsilon_2 - X_2 (X_1' X_1)^{-1} X_1' \varepsilon_1 \quad (3.26)$$

cuya esperanza y varianza son respectivamente

$$E(d) = X_2 \beta_2 - X_2 \beta_1 \quad (3.27)$$

$$\text{Var}(d) = (I_{n_2} + X_2 (X_1' X_1)^{-1} X_2') \sigma^2 \quad (3.28)$$

donde  $\hat{\beta}_1$  es el estimador M.C. de  $\beta_1$  y  $\sigma^2$  se estima a partir de las observaciones de la primera muestra, esto es  $\hat{\sigma}^2 = S_1^2$ . Chow analiza por separado dos casos, según que el número de observaciones del segundo grupo sea igual o mayor que uno.

a) Si hay sólo una observación adicional ( $n_2 = 1$ ), bajo la

hipótesis nula  $E(d) = 0$  y el estadístico

$$\frac{d^2}{(1 + X_2(X_1'X_1)^{-1}X_2') S_1^2} \quad (3.29)$$

se distribuye como una F con  $(1, n_1 - k)$  grados de libertad. Si el valor del estadístico es menor que el valor de  $F(1, n_1 - k)$  para un nivel de significancia previamente fijado se rechaza la hipótesis nula y se acepta que la nueva observación no pertenece a la misma regresión.

b) Para el caso en que el número de observaciones adicionales sea mayor que uno Chow propone tres test distintos:

Test 1:

Sea

$$\bar{d} = \frac{1}{n_2} 1'd \quad (3.30)$$

la media de la variable  $d_i$  definida en la sección anterior y

$$\text{Var}(\bar{d}) = \frac{1}{n_2} 1'(I_{n_2} + X_2(X_1'X_1)^{-1}X_2')\sigma^2 1 \quad (3.31)$$

$$E(d) = \frac{1}{n_2} 1'(X_2\beta_2 - X_2\beta_1) \quad (3.32)$$

su varianza y esperanza respectivamente.

Bajo la hipótesis nula el estadístico

$$\frac{\bar{d}^2}{\frac{S_1^2}{n_2} \left[ 1'(I_{n_2} + X_2(X_1'X_1)^{-1}X_2')1 \right]} \quad (3.33)$$

se distribuye como una F con  $(1, n_1 - k)$  grados de libertad. Este test es obviamente débil versus un gran número de hipótesis alternativas, pues  $\bar{d}$  puede anularse en situaciones distintas a la hipótesis nula, razón por la cual Chow mismo no lo considera un

test interesante, sin embargo es útil para construir intervalos de predicción para la media de  $n_2$  observaciones adicionales.

Test 2:

En vez de testar la hipótesis nula  $\beta_1 = \beta_2$  Chow prueba la hipótesis  $E(d) = 0$ . Para ello define la forma cuadrática  $A = d'(\text{Cov } d)d$  donde

$$A = [\beta_2' X_2' - \beta_1' X_2'] [I + X_2 (X_1' X_1)^{-1} X_2']^{-1} [X_2 \beta_2 - X_2 \beta_1] \sigma^{-2} + [\varepsilon_1' \varepsilon_2'] \\ \begin{bmatrix} -X_1 (X_1' X_1)^{-1} X_2' \\ I \end{bmatrix} [I + X_2 (X_1' X_1)^{-1} X_2']^{-1} [-X_2 (X_1' X_1)^{-1} X_1' \quad I] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sigma^{-2} \quad (3.34)$$

Bajo la hipótesis nula el primer término de  $A$  se anula y  $A$  se distribuye como una  $\chi^2_{(n_2)}$ , mientras que bajo la hipótesis alternativa lo hace como una  $\chi^2$  no centrada. Como el estadístico  $\frac{S_1^2 (n_1 - k)}{\hat{\sigma}^2}$  se distribuye como una  $\chi^2_{(n_1 - k)}$  y  $S_1^2$  es independiente de  $\hat{\beta}_1$  y de  $\hat{Y}_2$ . Luego bajo la hipótesis nula el estadístico

$$\frac{d'(\text{Cov } d)^{-1} d \frac{1}{n_2}}{\frac{S_1^2 (n_1 - k)}{\sigma^2 (n_1 - k)}} = \frac{d' [I + X_2 (X_1' X_1)^{-1} X_2']^{-1} d}{S_1^2 n_2} \quad (3.35)$$

se distribuye, bajo la hipótesis nula como una  $F$  con  $(n_2, n_1 - k)$  grados de libertad. La regla de decisión es rechazar la hipótesis nula si el valor del estadístico supera el valor de  $F(n_2, n_1 - k)$  para un nivel de significancia previamente establecido.

Test 3:

En realidad los dos test propuestos son casos especiales de la teoría de hipótesis lineales. Chow propone aplicar la teoría de las hipótesis lineales para probar la hipótesis nula  $\beta_1 = \beta_2$ . Bajo  $H_0$  el modelo puede expresarse de la siguiente forma:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta_0 + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad (3.36)$$

con  $\varepsilon$  que se distribuye  $N(0, \sigma^2 I_N)$ . La suma de cuadrados residual bajo la hipótesis nula ( $SCR_{H_0}$ ) es:

$$SCR_{H_0} = [\varepsilon_1', \varepsilon_2'] \left[ I - \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} (X_1' X_1 + X_2' X_2)^{-1} (X_1', X_2') \right] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

y bajo la hipótesis alternativa ( $\beta_1 \neq \beta_2$ ):

$$SCR_{H_1} = \varepsilon_1' [I_{n_1} - X_1 (X_1' X_1)^{-1} X_1'] \varepsilon_1 + \varepsilon_2' [I_{n_2} - X_2 (X_2' X_2)^{-1} X_2'] \varepsilon_2$$

con lo que la  $SCR_{H_0}$  puede expresarse como:

$$SCR_{H_0} = SCR_{H_1} + (\hat{\beta}_1 - \hat{\beta}_0)' [X_1' - X_1' X_1 (X_2' X_2)^{-1} X_2'] \begin{bmatrix} X_1 \\ -X_2 (X_2' X_2)^{-1} X_1' X_1 \end{bmatrix} (\hat{\beta}_1 - \hat{\beta}_0)$$

$$SCR_{H_0} = SCR_{H_1} + R \quad (3.37)$$

donde  $\hat{\beta}_0$  es el estimador MC de  $\beta$  bajo  $H_0$  y  $R$  es una forma cuadrática e  $\hat{\beta}_1 - \hat{\beta}_0$  de rango  $k$ .  $\hat{\beta}_1 - \hat{\beta}_0$  puede expresarse como:

$$\hat{\beta}_1 - \hat{\beta}_0 = \beta_1 - \beta_0 + \begin{bmatrix} (X_1' X_1)^{-1} X_1' & 0 \\ -[X_1' X_1 + X_2' X_2]^{-1} [X_1' & X_2'] \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad (3.38)$$

Chow considera dos situaciones según que el número de observaciones adicionales ( $n_2$ ) sea mayor o menor que el número de variables explicativas del modelo ( $k$ ).

a) Si  $n_2$  es mayor que  $k$  la  $SCR_{H_0}$  se distribuye como  $\chi^2_{(n_1 - n_2 - k)} \sigma^2$  y  $R$  como  $\chi^2_{(k)} \sigma^2$  y bajo la hipótesis alternativa  $R$  tenderá a aumentar. En este caso el estadístico

$$\frac{(SCR_{H_0} - SCR_{H_1})/k}{SCR_{H_1} / (n_1 + n_2 - 2k)} \quad (3.39)$$

se distribuye como una  $F_{(k, n_1 + n_2 - 2k)}$  g.l. bajo la hipótesis nula. Si el valor del estadístico es mayor que el valor de  $F_{(k, n_1 + n_2 - 2k)}$  para un nivel de significancia previamente establecido se acepta la hipótesis nula y en caso contrario se rechaza.

b) Si  $n_2$  es menor que  $k$  no puede estimarse  $\beta_2$  como en la sección anterior, así que la  $SCR_{H_1} = (Y_1 - X_1 \hat{\beta}_1)' (Y_1 - X_1 \hat{\beta}_1)$  pues la SCR para

la segunda muestra será nula. En este caso  $\frac{SCR_{H_1}}{\sigma^2}$  se distribuye

como  $\chi^2_{(n_1 - k)}$  y  $\frac{SCR_{H_0}}{\sigma^2}$  como  $\chi^2_{(n_1 + n_2 - k)}$ ; con lo que la hipótesis podrá ser testada mediante el estadístico:

$$\frac{(SCR_{H_0} - SCR_{H_1})/n_2}{SCR_{H_1} / (n_1 - k)} \quad (3.40)$$

que se distribuye como una  $F_{(n_2, n_1 - k)}$  bajo la hipótesis nula.

Cuando  $n_2$  es mayor que  $k$  este estadístico sigue siendo válido pero es menos potente que el encontrado en el caso (a). Chow demuestra que el test 2 y su caso particular 1 son un caso especial del test 3 donde se aplicó la teoría de las hipótesis lineales ya que  $A = d' (Covd)^{-1} d = ||X_1 \hat{\beta} - X_1 \hat{\beta}_0||^2 + ||Y_2 - X_2 \hat{\beta}_0||^2 =$

$$SCR_{H_0} - SCR_{H_1} .$$

## 3.2.2. Solución propuesta por Quandt (1958-1960)

Quandt considera el caso en que no se conoce qué observaciones pertenecen a uno u otro régimen, y se desea testar si la regresión sigue un régimen de una o de dos fases. Esto es,  $H_0: \beta_1 = \beta_2$  versus  $H_1: \beta_1 \neq \beta_2$ . Quandt propone para ello dos estadísticos uno para muestras grandes y otro para muestras pequeñas.

El modelo bajo  $H_0$  es (3.36), mientras que bajo  $H_1$  el modelo es (3.1) con matriz de varianza-covarianza (3.2), donde  $\tau$  es el punto de cambio de régimen que debe estimarse a partir de los datos muestrales en la forma ya indicada en la Sección 3.1.1.

## 1) Test para muestras grandes:

El test para muestras grandes propuesto es un test de razón de verosimilitud  $\lambda$  tal que

$$\lambda = \frac{L(\omega)}{L(\Omega)} = \frac{\text{maximo de la funcion de verosimilitud bajo } H_0}{\text{maximo de la funcion de verosimilitud bajo } H_1}$$

$$= \frac{\hat{\sigma}_1^{\tau} \hat{\sigma}_2^{N-\tau}}{\hat{\sigma}^N} \quad (3.41)$$

donde  $\omega$  es el espacio paramétrico bajo  $H_0$ , de dimensión  $N-3=n$  y  $\Omega$  es el espacio paramétrico bajo  $H_1$ , de dimensión  $N-7=m$ . (Si el número de variables explicativas fuese mayor que uno, entonces  $\dim(\omega) = N-(k+1)$  y  $\dim(\Omega) = N-2(k+1)-1$ , siendo  $k$  el número de variables explicativas del modelo).

Bajo ciertas condiciones la distribución  $\chi_{n-m}^2$  es una aproximación aceptable a la distribución de  $-2 \log \lambda$ . En este caso eso no sucede pues  $\tau$  es una variable discreta, pero las distorsiones que produce se hacen menos significativas a medida que el tamaño muestral aumenta. Quandt examina el supuesto de que  $-2 \log \lambda$  se distribuye asintóticamente como  $\chi_{(4)}^2$ , para muestras grandes. Para ello define una variable  $X$  no estocástica con los valores particulares 0,05; 0,10; 0,15; ...; 1 y genera 200 series con 20

observaciones cada una ( $c/X_i$  usado una sola vez); 200 series con 40 observaciones cada una ( $c/X_i$  usado dos veces) y 200 series con 60 observaciones cada una ( $c/X_i$  usado tres veces). Los valores de los residuos  $\varepsilon_i$  fueron generados a partir de la distribución normal con  $\mu=0$  y  $\sigma^2=1$ . Quandt calculó el valor de  $\lambda$  para cada una de las 600 series y se probó la hipótesis de que cada una de las distribuciones empíricas de  $-2\log\lambda$  había sido generada por una distribución  $\chi^2_{(4)}$ . Esta hipótesis fue rechazada en TODOS los casos con un nivel de significancia del 0,00001 con lo cual se concluye que la aproximación  $\chi^2$  falla completamente. Una posible razón teórica es que la derivada de la función de verosimilitud se anula en el máximo, y en este caso sólo se ha definido para valores enteros de  $\tau$ . Además la distribución de  $-2\log\lambda$  no es independiente de la longitud de la serie. Quandt propone tres soluciones para este problema:

a) Hacer un test empírico. Si las tres distribuciones se pudiesen unir se podría derivar una tabla empírica de puntos porcentuales con alto grado de confianza; pero antes debería probarse la hipótesis de que las tres distribuciones empíricas han sido generadas por la misma distribución teórica. Para testar esto último Quandt utiliza tres criterios: el test de contingencia de  $\chi^2$ , el test de Kruskal-Wallis y el test de Kiefer. Para las tres pruebas la probabilidad de mayor discordancia entre las distribuciones fue menor del 1%. Se realizó por lo tanto para cada longitud de serie una tabla empírica que asocia para varios valores de  $k$  la probabilidad estimada de que  $-\log\lambda < k$  e intervalos de confianza del 95% para los mismos.

Se reproduce a continuación la tabla publicada por Quandt de la probabilidad estimada de que  $-\log\lambda < k$  e intervalos de confianza del 95% alrededor de la estimación.

Tabla 3: Puntos porcentuales para  $-\log \lambda$  con intervalos del 95% para  $N=20, 40$  y  $60$ .

K	T=20			T=40			T=60		
	Lower Bound	$\hat{P}(-\log \lambda) < K$	Upper Bound	Lower Bound	$\hat{P}(-\log \lambda) < K$	Upper Bound	Lower Bound	$\hat{P}(-\log \lambda) < K$	Upper Bound
3.0	(.008)	.021	(.052)	(.000)	.000	(.019)	(.000)	.000	(.010)
4.0	(.049)	.080	(.127)	(.004)	.013	(.041)	(.000)	.006	(.030)
5.0	(.057)	.126	(.179)	(.022)	.043	(.081)	(.023)	.043	(.081)
6.0	(.175)	.228	(.291)	(.073)	.109	(.160)	(.083)	.122	(.175)
7.0	(.305)	.370	(.439)	(.160)	.211	(.273)	(.135)	.181	(.243)
8.0	(.380)	.454	(.524)	(.265)	.326	(.394)	(.220)	.278	(.344)
9.0	(.477)	.547	(.615)	(.378)	.445	(.515)	(.323)	.388	(.457)
10.0	(.564)	.633	(.697)	(.520)	.599	(.658)	(.404)	.472	(.542)
11.0	(.602)	.706	(.811)	(.637)	.704	(.763)	(.527)	.607	(.683)
12.0	(.751)	.811	(.860)	(.744)	.805	(.854)	(.654)	.720	(.778)
13.0	(.822)	.875	(.914)	(.785)	.845	(.889)	(.744)	.805	(.854)
14.0	(.864)	.912	(.944)	(.841)	.892	(.928)	(.800)	.856	(.898)
15.0	(.893)	.925	(.957)	(.874)	.920	(.950)	(.829)	.881	(.910)
16.0	(.932)	.967	(.985)	(.894)	.937	(.964)	(.853)	.903	(.937)
17.0	(.958)	.984	(.995)	(.898)	.940	(.965)	(.890)	.935	(.964)
18.0	(.957)	.955	(.995)	(.929)	.965	(.953)	(.914)	.953	(.979)
19.0	(.958)	.986	(.996)	(.947)	.978	(.991)	(.936)	.970	(.980)
20.0	(.959)	.987	(.996)	(.954)	.983	(.994)	(.952)	.982	(.994)
21.0	(.961)	.988	(.997)	(.958)	.986	(.998)	(.956)	.987	(.996)

b) Hacer arbitrariamente  $\tau = N/2$  si  $N$  es par o  $\tau = \frac{(N+1)}{2}$  si  $N$  es impar. De acuerdo a nuestro punto de vista esta solución no es admisible ya que por su arbitrariedad producirá contaminación entre ambos regímenes.

c) Eliminar las observaciones centrales y considerar que pertenecen al primer régimen las primeras  $g$  observaciones y al segundo las últimas  $g$ . La potencia del test dependerá de cuán cerca esté el punto de cambio verdadero del final de la serie. Tampoco consideramos admisible esta solución pues no tiene ninguna conexión con la ubicación del punto de cambio para los datos muestrales que se estén analizando.

Para (b) y (c) la potencia de los tests disminuirá mientras más cerca esté el punto de cambio de régimen a cualquiera de los extremos de la serie y aumentará a medida que sea más brusco el cambio de régimen.

## 2) Test para muestras pequeñas:

El autor propone tres test para muestras pequeñas.

A) El primero de ellos es el siguiente estadístico:

$$\frac{SCR_{H_0} - SCR_{H_1}}{SCR_{H_1}} \quad (3.42)$$

que, según Quandt, bajo la hipótesis nula se distribuye aproximadamente como una  $F(4, N-8)$  grados de libertad; con el supuesto adicional que  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ . Sin embargo los grados de libertad de la  $SCR_{H_0}$  son  $(N-3)$  y de la  $SCR_{H_1}$  son  $(N-7)$ . La discrepancia en los grados de libertad puede deberse a una corrección para mejorar la aproximación a la distribución  $F$  pues el hecho de que  $\tau$  no sea conocido sino que sea una estimación MV del punto de cambio basado en todas las observaciones tenderá a hacer la  $SCR_{H_1}$  más pequeña y aumentará la varianza de la razón pues esto disminuye los g.l del denominador y aumenta los del numerador lo que resulta en un rechazo de  $H_0$  con una probabilidad mayor que la que indica el nivel de significancia del test (3.42). Lamentablemente el autor no realiza ninguna aclaración al respecto. Para solucionar este inconveniente Quandt propone hacer en forma arbitraria  $\tau = N/2$  o  $\tau = (N-1)/2$ . Nosotros rechazamos esta solución por las razones ya expuestas; consideremos que un estadístico más adecuado para testar la hipótesis nula es

$$\frac{(SCR_{H_0} - SCR_{H_1})/4}{SCR_{H_1} / (N-7)} \quad (3.43)$$

que bajo la hipótesis nula se distribuye como una  $F$  con  $(4, N-7)$  g.l., pues dividir el numerados y el denominador por los g.l. mejorará la aproximación a la distribución  $F$ , siempre que se cumpla el supuesto  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ .

Si el número de variables explicativas fuese mayor que uno los grados de libertad serán  $(N-2k-1, N-k)$ , siendo  $k$  el número de variables explicativas del modelo.

B) El segundo es llamado test 1 por Quandt.

El sugiere que una vez encontrados los estimadores MV de

$\tau, \beta_1, \beta_2, \sigma_1^2, \sigma_2^2$  se estime  $Y_1$  usando  $\hat{\beta}_2$  y  $Y_2$  usando  $\hat{\beta}_1$  y se calculen los vectores de residuos  $R_1$  y  $R_2$  tal que  $R_1 = X_1 \hat{\beta}_2 - Y_1$  y  $R_2 = X_2 \hat{\beta}_1 - Y_2$  cuya media y varianza son respectivamente:  $\bar{r}_1 = \frac{1}{\tau} 1' R_1$ ,  $S_1^2 = \hat{\text{Var}}(R_1)$ ,  $\bar{r}_2 = \frac{1}{N-\tau} 1' R_2$  y  $S_2^2 = \hat{\text{Var}}(R_2)$ . Aparentemente los estadísticos

$$\frac{\sqrt{\hat{\tau}} (\bar{r}_1 - E(\bar{r}_1))}{S_1} \quad \text{y} \quad (3.44)$$

$$\frac{\sqrt{N-\hat{\tau}} (\bar{r}_2 - E(\bar{r}_2))}{S_2} \quad (3.45)$$

se distribuyen respectivamente como una  $t_{(\hat{\tau}-1)}$  y  $t_{(N-\hat{\tau}-1)}$ .  $H_0$  será rechazada si cualquiera de los estadísticos es significativo al nivel de significancia previamente establecido.

Este test presenta dos problemas; el primero es que la  $E(\bar{r}_1)$  y  $E(\bar{r}_2)$  pueden anularse no sólo bajo la hipótesis nula sino también

bajo la alternativa cuando 
$$\frac{\hat{\beta}_{12} - \hat{\beta}_{11}}{\hat{\beta}_{22} - \hat{\beta}_{21}} = - \frac{\sum_{i=1}^{\tau} X_{2i}}{\tau} = - \frac{\sum_{i=\tau+1}^N X_{2i}}{N-\tau}$$

Esto tiende a disminuir la potencia del test. Sin embargo esto no es una dificultad seria pues a menos que la distribución de  $X_2$  sea singular esto ocurrirá con probabilidad cero. Una objeción más seria es que la distribución del estadístico será afectada por el hecho de que  $\tau$  ha sido estimado a partir de los datos lo que puede producir que no sea una distribución t de Student.

Nosotros lo rechazamos pues no conocemos su distribución. Para evitar este problema Quandt realiza la misma sugerencia que para los estadísticos anteriores; es decir hacer  $\tau = \frac{N}{2}$  si N es par o  $\tau = \frac{N+1}{2}$  si N es impar. Esto no es para nosotros una solución válida pues uno de los problemas más importantes que presenta el modelo de regresión a trozos es la estimación del punto de cambio y el fijarlo en forma arbitraria no guarda ninguna relación con los datos muestrales que se estén considerando, por lo cual hay un

gran peligro de contaminación.

C) Quandt llama a este estadístico test 2

En el supone que  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ . Bajo la hipótesis nula los estadísticos

$$\frac{(X_1 \hat{\beta}_2 - Y_1)' (X_1 \hat{\beta}_2 - Y_1)}{\sigma^2} \quad \text{y} \quad \frac{(X_2 \hat{\beta}_2 - Y_2)' (X_2 \hat{\beta}_2 - Y_2)}{\sigma^2}$$

se distribuyen respectivamente como una  $\chi^2_{(\hat{\tau}-1)}$  y  $\chi^2_{(N-\hat{\tau}-1)}$ , con lo que la razón

$$\frac{(X_1 \hat{\beta}_2 - Y_1)' (X_1 \hat{\beta}_2 - Y_1)}{(X_2 \hat{\beta}_2 - Y_2)' (X_2 \hat{\beta}_2 - Y_2)} \quad (3.46)$$

bajo la hipótesis nula se distribuye como  $F_{(\hat{\tau}-1, N-\hat{\tau}-1)}$ . Este es el test propuesto por Quandt.

Nosotros consideramos que una mejor aproximación a la distribución F se obtendrá dividiendo el numerador y el denominador por sus grados de libertad, con lo que el test propuesto será:

$$\frac{(X_1 \hat{\beta}_2 - Y_1)' (X_1 \hat{\beta}_2 - Y_1) / (\hat{\tau}-1)}{(X_2 \hat{\beta}_2 - Y_2)' (X_2 \hat{\beta}_2 - Y_2) / (N-\hat{\tau}-1)} \quad (3.47)$$

que bajo la hipótesis nula también se distribuye como una  $F_{(\hat{\tau}-1, N-\hat{\tau}-1)}$ . La potencia de este test está relacionado con la ubicación del punto de cambio de régimen.

### 3.2.3. Solución propuesta por Broemeling y Chin Choy (1981)

Broemeling y Chin Choy proponen un test para detectar si el modelo sigue un régimen de una o de dos fases. Usan técnicas bayesianas donde el test está basado en la función de masa marginal a posteriori de  $\tau$ , que compara la probabilidad a

posteriori de que no ocurra ningún cambio con la probabilidad de que ocurra exactamente un punto de cambio de régimen.

Si  $\tau=N$  la relación es estable en todo el rango de variación de  $X$ , en cambio si  $\tau$  está comprendido entre 1 y  $N-1$   $\beta_1$  es distinto a  $\beta_2$ . El modelo que ellos consideran es (3.1) con el supuesto adicional de que  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ . En forma resumida se lo puede expresar como:

$$Y_\tau = X_\tau \beta + \varepsilon_\tau \quad (3.48)$$

donde el subíndice  $\tau$  indica el punto de cambio de régimen.

La función de verosimilitud de la muestra es:

$$L(\tau, \beta_1, \beta_2, \sigma^2) \propto \frac{1}{(\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} (Y_N - X_N \beta_1)' (Y_N - X_N \beta_1)\right\}; \text{ si } \tau=N$$

$$\frac{1}{(\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} (Y_\tau - X_\tau \beta)' (Y_\tau - X_\tau \beta)\right\}; \text{ si } 1 \leq \tau \leq N-1 \quad (3.49)$$

La función de densidad a priori de  $\tau$  es:

$$\pi_0(\tau) = \begin{cases} q & \text{si } \tau = N \\ \frac{1-q}{N-1} & \text{si } 1 \leq \tau \leq N-1 \end{cases}; \quad (3.50)$$

$q$  es dado y está comprendido entre 0 y 1.

Si  $\tau=N$  la función de densidad a priori de  $(\beta_1, 1/\sigma^2)$  es una gama normal, ésto es:

$$\pi_0(\beta_1, \frac{1}{\sigma^2}) \propto \sigma^{-m} \exp\left\{-\frac{1}{2\sigma^2} [M + (\beta_1 - \beta_{10})' \Lambda_{11} (\beta_1 - \beta_{10})]\right\}, \quad (3.51)$$

la función de densidad condicional de  $\beta_1$  dado  $1/\sigma^2$  es:

$$\pi_0(\beta_1 / \sigma^{-2}) \propto \frac{1}{(\sigma^2)^{K/2}} \exp\left\{-\frac{1}{2\sigma^2} (\beta_1 - \beta_{10})' \Lambda_{11} (\beta_1 - \beta_{10})\right\}, \quad (3.52)$$

luego  $\beta_1$  se distribuye  $N(\beta_{10}, \sigma^2 \Lambda_{11}^{-1})$  y  $\pi_0(1/\sigma^2)$  es una gama con

parámetros  $(b, a)$ ;

$$\pi_0\left(\frac{1}{\sigma^2}\right) \propto (1/\sigma^2)^{a-1} \exp(b/\sigma^2), \text{ con } \frac{1}{\sigma^2} > 0, a > 0, b > 0 \quad (3.53)$$

Si  $\tau$  está comprendido entre 1 y  $N-1$ ,  $\pi_0(\beta, \frac{1}{\sigma^2})$  es una gama normal de la siguiente forma:

$$\pi_0\left(\beta, \frac{1}{\sigma^2}\right) \propto \sigma^{-m} \exp\left\{-\frac{1}{2\sigma^2} [M + (\beta - \beta_0)' \Lambda (\beta - \beta_0)]\right\} \quad (3.54)$$

donde  $M > 0$ ,  $m > 0$  y  $\Lambda$  es una matriz definida positiva.

$\pi_0(\beta/\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{k/2} \exp\left\{-\frac{1}{2\sigma^2} (\beta - \beta_0)' \Lambda (\beta - \beta_0)\right\}$  y  $\pi_0\left(\frac{1}{\sigma^2}\right)$  es (3.53).  $\Lambda$  es la matriz de precisión definida

positiva, particionada como  $\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12} & \Lambda_{22} \end{bmatrix}$  donde  $\Lambda_{11}$ ,  $\Lambda_{12}$  y  $\Lambda_{22}$  son matrices especificadas.

Combinando la función de verosimilitud con las expresiones de las funciones de densidad a priori de los parámetros se obtienen las funciones de densidad a posteriori:

$$\pi_1\left(\beta_1, \frac{1}{\sigma^2}, \tau\right) \propto |\Lambda_{11}|^{1/2} q \left[\frac{1}{\sigma^2}\right]^{\frac{N+k+2a}{2} - 1} \exp\left\{-\frac{1}{2\sigma^2} [D(N) + (\beta_1 - \beta_1^*)' [X_N' X_N + \Lambda_{11}] (\beta_1 - \beta_1^*)]\right\} \quad (3.56)$$

si  $\tau = N$  y

$$\pi_1\left(\beta_1, \frac{1}{\sigma^2}, \tau\right) \propto \frac{|\Lambda|^{1/2} (1-q)}{(N-1)} \left[\frac{1}{\sigma^2}\right]^{\frac{N+2k+2a}{2} - 1} \exp\left\{-\frac{1}{2\sigma^2} [D(\tau) + (\beta - \beta^*(\tau))' [X_\tau' X_\tau + \Lambda] (\beta - \beta^*(\tau))]\right\} \quad (3.57)$$

si  $\tau$  está comprendido entre 1 y  $N-1$ , donde:

$$\beta^*(\tau) = [X_\tau' X_\tau + \Lambda]^{-1} [\Lambda \beta_0 + X_\tau' Y_\tau], \quad 2D(\tau) = 2b + [Y_\tau - X_\tau \beta^*(\tau)]' Y_\tau + [\beta_0 - \beta^*(\tau)]' \Lambda \beta_0, \quad \beta_1^* = [X_N' X_N + \Lambda_{11}]^{-1} [\Lambda_{11} \beta_{10} + X_N' Y_N] \text{ y } 2D(N) = 2b +$$

$$+\{Y'_{N'}Y_{N'} + \beta'_{10} \Lambda_{11} \beta_{10} - \beta_{10}^* [X'_{N'}X_{N'} + \Lambda_{11}] \beta_{10}^*\}.$$

La preocupación principal es encontrar la función de masa a posteriori de  $\tau$ , para ello  $\beta_1$ ,  $\beta_2$  y  $\frac{1}{\sigma^2}$  son tratados como parámetros sin importancia y en consecuencia se integra respecto a ellos obteniéndose la siguiente expresión:

$$\pi_1(\tau) \propto q \left[ \lambda_{11} \right]^{1/2} D(N)^{-(N+2\alpha)/2} \left[ X'_{N'}X_{N'} + \Lambda_{11} \right]^{1/2}, \quad \text{si } \tau=N$$

$$\propto \frac{(1-q)}{(N-1)} \left[ \Lambda \right]^{1/2} D(\tau)^{-(N+2\alpha)/2} \left[ X'(\tau)X(\tau) + \Lambda \right]^{1/2} \quad \text{si } 1 < \tau < N-1. \quad (3.58)$$

Esta función es usada para probar estabilidad, si alcanza su máximo valor para  $\tau=N$  se acepta que la regresión sigue una sola fase, en cambio si alcanza su máximo para cualquier valor comprendido entre 1 y  $N-1$  se considera que la regresión sigue un régimen de dos fases.

Broemeling y Chin Choy proponen además otro procedimiento utilizando el método secuencial de Smith (1975). Este consiste en tomar las primeras  $t$  observaciones  $Y_1, Y_2, \dots, Y_t$  y calcular  $\pi_1(t)$  para  $t=2, 3, \dots, N$ . Consideran que la función de masa a priori tiene la siguiente expresión:

$$\pi_0(\tau) = \begin{cases} q & \text{si } \tau = t \\ \frac{1-q}{t-1} & \text{si } 1 \leq \tau \leq t-1 \end{cases} \quad (3.59)$$

Para cada conjunto de  $t$  observaciones se grafica  $\pi_1(t)$  vs  $t$  con  $t=2, 3, \dots, N$ . Si el gráfico muestra una tendencia descendente se sospecha de la inestabilidad del modelo. Este método es similar al test de razón de verosimilitud de Brown, Durbin y Evans (1975). Los autores comparan los métodos mediante simulación aplicando ambos a los datos de Quandt (1958). Para ello

consideran que la media y la varianza de las funciones de densidad a priori de los parámetros son iguales al valor del parámetro a partir del cual se generaron los datos. Esto es:  $\beta = (2,5; 0,7; 5; 0,5)'$ ,  $\beta_{10} = (2,5; 0,7)'$ ,  $\Lambda = I_4, \Lambda_{11} = I_2, \Lambda_{22} = I_2, \Lambda_{12} = 0_{(2 \times 2)}$  y  $a=b=1$ .

Para cuatro valores de  $q$ : 0,05; 0,5; 0,95 y 0,99 se calcula  $\pi_1(\tau)$ . Esta función es sensible a la elección de  $q$ , pero alcanza su máximo para  $\tau=12$  en los cuatro casos. El método secuencial de Smith se maximiza para  $\tau=11$  para las cuatro elecciones de  $q$ .

Nosotros creemos que esta forma de probar el método no es adecuada pues en las situaciones reales nunca se conoce la función a partir de la cual se generaron las observaciones. Incorporando esta información se favorece la aplicación del mismo pues se incorpora información "verdadera". Una forma más honesta de probarlo es la siguiente: fijar  $q$  y generar un punto de  $\pi_0(\tau)$ : si el punto generado es  $N$  se genera  $\beta_1$  con  $\pi_0(\beta_1/\sigma^{-2})$ , en cambio si el punto generado está comprendido entre 1 y  $N-1$  se genera  $\beta$  con  $\pi_0(\beta/\sigma^{-2})$ . Los parámetros de las distribuciones  $\pi_0(\beta_1/\sigma^{-2})$  y  $\pi_0(\beta/\sigma^{-2})$  se fijan a priori, pero no igual al valor del parámetro a partir de los cuales se generaron los datos. Con estos datos se generan las observaciones  $Y_i$ . Esta es una forma de probar el método no tan tendenciosa como la propuesta por los autores donde en las distribuciones a priori de los parámetros incorpora la información sobre la distribución "verdadera" que generó los datos, la cual en la realidad nunca se halla disponible para el investigador.

Otra crítica que podría hacerse al trabajo es que sólo considera el caso en que haya cero o un punto de cambio, pero no puede generalizarse al caso donde el número de puntos de cambio sea mayor. Una solución puede ser buscar la distribución a priori de  $k$  puntos de cambio.

### 3.3. Tests de hipótesis sobre los restantes parametros de la regresión.

#### 3.3.1. Solución propuesta por Chow(1960).

Chow considera el caso del modelo de regresión de dos fases, donde se conoce de antemano qué observaciones pertenecen a cada una de las fases de la regresión, y se desea averiguar si es idéntica parte de la relación para ambas. Estadísticamente esto es equivalente a testar si subconjuntos de coeficientes de las dos fases son los mismos. Esto permitirá dar respuesta a preguntas como: ¿ha cambiado la elasticidad ingreso de la demanda para un cierto alimento desde la Segunda Guerra, mientras que la elasticidad precio se ha mantenido constante?

Para poder hacer progresos analíticos el autor supone que la variancia del error es la misma para ambas fases de la regresión, esto es  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ , y que el vector de residuos se distribuye  $N(0, \sigma^2 I_N)$ . Chow utiliza el método de los mínimos cuadrados. Para el comentario de este trabajo se han modificado las notaciones del modelo y de los estadísticos propuestos por Chow para seguir la expresión del modelo general ya citado. Esto simplifica notablemente las expresiones.

Sin pérdida de generalidad puede suponerse que los primeros  $m$  coeficientes son los que se mantienen estables, donde obviamente  $m$  debe ser menor que  $k$ . Esto originará la siguiente partición de  $X_1, X_2, \beta_1$  y  $\beta_2$ :  $X_1 = (X_{1m}, X_{1p})$  donde  $X_{1m}$  está formado por las primeras  $m$  columnas de  $X_1$  y  $X_{1p}$  por las últimas  $(k-m)$ ,  $\beta_1 = (\beta_{1m}, \beta_{1p})'$  donde  $\beta_{1m}$  está formado por los primeros  $m$  coeficientes de  $\beta_1$  y  $\beta_{1p}$  por los últimos  $(k-m)$  respectivamente y análogamente con  $X_2$  y  $\beta_2$ .

En términos de esta partición la hipótesis nula a testar es  $H_0: \beta_{1m} = \beta_{2m}$  versus  $H_1: \beta_{1m} \neq \beta_{2m}$ .

El modelo bajo  $H_0$  puede expresarse:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_{1m} & X_{1p} & 0 \\ X_{2m} & 0 & X_{2p} \end{bmatrix} \begin{bmatrix} \beta_{0m} \\ \beta_{1p} \\ \beta_{2p} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}; \quad \varepsilon \sim N(0, \sigma^2 I_N) \quad (3.60)$$

El estimador MC de  $\beta$  bajo  $H_0$  es:

$$\hat{\beta}_{H_0} = \begin{bmatrix} X'_{1m} X_{1m} + X'_{2m} X_{2m} & X'_{1m} X_{1p} & X'_{2m} X_{2p} \\ X'_{1p} X_{1m} & X'_{1p} X_{1p} & 0 \\ X'_{2p} X_{2m} & 0 & X'_{2p} X_{2p} \end{bmatrix}^{-1} \begin{bmatrix} X'_{1m} X_{1p} & X'_{2p} X_{2p} \\ X'_{1p} X_{1p} & 0 \\ 0 & X'_{2p} X_{2p} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \quad (3.61)$$

y la  $SCR_{H_0}$  tendrá  $(n_1 + n_2 - 2k + m)$  g.l. y se distribuirá como una  $\chi^2$  con  $(n_1 + n_2 - 2k + m)$  g.l.

Chow considera dos casos, según el número de observaciones de la segunda muestra sea mayor o menor que el número de parámetros de la regresión.

a) El número de observaciones es mayor que el número de parámetros. En este caso el estimador MC de  $\beta$  bajo  $H_1$  es:

$$\hat{\beta}_{H_1} = \begin{bmatrix} X'_{1m} X_{1m} & 0 & X'_{1m} X_{1p} & 0 \\ 0 & X'_{2m} X_{2m} & 0 & X'_{2m} X_{2p} \\ X'_{1p} X_{1m} & 0 & X'_{1p} X_{1p} & 0 \\ 0 & X'_{2p} X_{2m} & 0 & X'_{2p} X_{2p} \end{bmatrix}^{-1} \begin{bmatrix} X'_{1m} X_{1p} & 0 \\ 0 & X'_{2m} X_{2p} \\ X'_{1p} X_{1p} & 0 \\ 0 & X'_{2p} X_{2p} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \quad (3.62)$$

y la  $SCR_{H_1}$  tendrá  $(n_1 + n_2 - 2k)$  g.l. y se distribuirá como  $\chi^2_{(n_1 + n_2 - 2k)} \sigma^2$ . La  $SCR_{H_0}$  se puede expresar como:

$$SCR_{H_0} = SCR_{H_1} + \left\| \begin{pmatrix} X_{1m} & X_{1p} & 0 & 0 \\ 0 & 0 & X_{2m} & X_{2p} \end{pmatrix} \hat{\beta}_{H_1} - \begin{pmatrix} X_{1m} & X_{1p} & 0 \\ X_{2m} & 0 & X_{2p} \end{pmatrix} \hat{\beta}_{H_0} \right\|^2 \quad (3.63)$$

$$SCR_{H_0} = SCR_{H_1} + q$$

Como el rango de  $q$  es  $m$  el estadístico

$$\frac{(SCR_{H_0} - SCR_{H_1}) / m}{SCR_{H_1} / (n_1 + n_2 - 2k)} \quad (3.64)$$

se distribuye bajo  $H_0$  como una  $F$  con  $(m, n_1 + n_2 - 2k)$  g.l.

b) El número de observaciones es menor que el número de parámetros ( $(k-p) \leq n_2 < k$ ).

La  $SCR_{H_0}$  sólo puede calcularse si  $n_2 \geq (k-p)$ . En este caso se debe modificar el cálculo de la  $SCR_{H_1}$  que es:

$$SCR_{H_1} = (Y_1 - X_{1m} \hat{\beta}_{1m} - X_{1p} \hat{\beta}_{1p})' (Y_1 - X_{1m} \hat{\beta}_{1m} - X_{1p} \hat{\beta}_{1p}) \quad (3.65)$$

que tiene  $(n_1 - k)$  g.l. En este caso el estadístico

$$\frac{(SCR_{H_0} - SCR_{H_1}) / (n_2 - k + m)}{SCR_{H_1} / (n_1 - k)} \quad (3.66)$$

se distribuye, bajo  $H_0$ , como una  $F$  con  $(n_2 - k + m, n_1 - k)$  g.l. La regla del test es rechazar  $H_0$  si el valor del estadístico es mayor que el valor de  $F_{(n_2 - k + m, n_1 - k)}$  para un nivel de significancia previamente establecido.

En todos los tests Chow supone que la varianza del error se mantiene constante tanto bajo  $H_0$  como bajo  $H_1$ . Schmidt y Sickles han investigado el efecto de la violación de este supuesto en el nivel de significancia de los distintos test y han arribado a las siguientes conclusiones:

a) Cuando los tamaños de muestras en ambas fases son iguales ( $n_1 = n_2$ ) hay un aumento pequeño del nivel de significancia sobre el nivel nominal aun cuando haya un gran desvío de la suposición de igual varianza.

b) Cuando los tamaños de muestra son desiguales el nivel de significancia real muestra un desvío mayor del nominal y puede ser mayor o menor que ésta.

En general puede decirse que las consecuencias de la violación de los supuestos del modelo no han sido suficientemente tratados en la bibliografía.

### 3.3.2. Solución propuesta por Tsurumi (1977).

Tsurumi considera el problema del modelo de regresión de dos fases cuando se desea averiguar si un coeficiente de la regresión es idéntico para ambas fases. La solución propuesta es bayesiana.

El deriva un un "highest posterior density interval test" (HPDI test) para examinar cambios en parámetros individuales entre submuestras, en términos de una función de densidad de probabilidad doble t.

La hipótesis nula que se desea testar es:  $H_0: \beta_{1j} = \beta_{2j}$  ó  $\delta_j = 0$  versus la alternativa:  $H_1: \beta_{1j} \neq \beta_{2j}$  ó  $\delta_j \neq 0$ ; donde  $\delta_j = \beta_{2j} - \beta_{1j}$  y  $\beta_{1j}$  y  $\beta_{2j}$  son la j-ésima componente de  $\beta_1$  y  $\beta_2$  respectivamente; (j varía entre 1 y k). La función de verosimilitud de la muestra es:

$$L(\beta_1, \beta_2, \sigma_1, \sigma_2 / \text{datos}) = (2\pi)^{-N/2} \sigma_1^{-n_1} \sigma_2^{-n_2} \exp \left[ -\frac{1}{2\sigma_1^2} (Y_1 - X_1 \beta_1)' (Y_1 - X_1 \beta_1) \right] \exp \left[ -\frac{1}{2} (Y_2 - X_2 \beta_2)' (Y_2 - X_2 \beta_2) \right] \quad (3.67)$$

y las funciones de densidad a priori de  $(\beta_1, \sigma_1)$  y  $(\beta_2, \sigma_2)$  son respectivamente

$$\pi_0(\beta_1, \sigma_1) \propto \frac{1}{\sigma_1} \quad y \quad (3.68)$$

$$\pi_0(\beta_2, \sigma_2) \propto \frac{1}{\sigma_2} \quad ; \quad (3.69)$$

con lo que la función de densidad a posteriori de  $(\beta_1, \beta_2, \sigma_1, \sigma_2)$  es:

$$\pi_1(\beta_1, \beta_2, \sigma_1, \sigma_2) \propto (2\pi)^{-N/2} \sigma_1^{-(n_1+1)} \sigma_2^{-(n_2-1)} \exp\left\{-\frac{1}{2\sigma_1^2}(Y_1 - X_1\beta_1)'\right. \\ \left.(Y_1 - X_1\beta_1)\right\} \exp\left\{-\frac{1}{2\sigma_2^2}(Y_2 - X_2\beta_2)'\right. \\ \left.(Y_2 - X_2\beta_2)\right\} \quad (3.70)$$

Integrando respecto a  $\sigma_1$  y  $\sigma_2$  se obtiene la densidad a posteriori de  $(\beta_1, \beta_2)$

$$\pi_1(\beta_1, \beta_2) \propto \left[ (Y_1 - X_1\hat{\beta}_1)'\right. \\ \left. (Y_1 - X_1\hat{\beta}_1) + (\beta_1 - \hat{\beta}_1)'\right. \\ \left. X_1'X_1(\beta_1 - \hat{\beta}_1) \right]^{-n_1/2} \\ \left[ (Y_2 - X_2\hat{\beta}_2)'\right. \\ \left. (Y_2 - X_2\hat{\beta}_2) + (\beta_2 - \hat{\beta}_2)'\right. \\ \left. X_2'X_2(\beta_2 - \hat{\beta}_2) \right]^{-n_2/2} \quad (3.71)$$

y donde se deduce la densidad a posteriori de  $(\beta_{j1}, \beta_{j2})$

$$\pi_1(\beta_{j1}, \beta_{j2}) \propto \left[ 1 + \frac{(\beta_{j1} - \hat{\beta}_{j1})^2}{\hat{\sigma}_{j1}^2 (n_1 - k)} \right]^{-(n_1 - k + 1)/2} \\ \left[ 1 + \frac{(\beta_{j2} - \hat{\beta}_{j2})^2}{\hat{\sigma}_{j2}^2 (n_2 - k)} \right]^{-(n_2 - k + 1)/2} \quad (3.72)$$

que es el producto de funciones  $t$  univariadas, donde

$$\hat{\sigma}_{ji}^2 = \frac{(Y_i - X_i\hat{\beta}_i)'\right. \\ \left.(Y_i - X_i\hat{\beta}_i)}{(n_i - k)} (X_i'X_i)^{-1}, \text{ con } i=1,2 \text{ es la estimación de}$$

la variancia del error para la primera y segunda fase de la

regresión respectivamente. El estadístico  $\frac{\beta_{ji} - \hat{\beta}_{ji}}{\hat{\sigma}_{ji}}$  se distribuye

como una  $t_{(n_i - k)}$ , con  $i=1,2$  y  $j=1,2,\dots,k$

Para testar la hipótesis nula Tsurumi realiza el siguiente cambio de variable:  $\delta_j = \beta_{2j} - \beta_{1j}$  y  $\eta = \beta_{1j}$  con lo que la función de densidad a posteriori de  $\delta_j$  será:

$$\pi_1(\delta_j/\text{datos}) \propto \left[ 1 + \frac{(\eta - \hat{\eta})^2}{(n_1 - k) \hat{\sigma}_{j1}^2} \right]^{-(n_1 - k + 1)/2} \left[ 1 + \frac{(\delta_j + \eta - \hat{\delta}_j - \hat{\eta})^2}{(n_2 - k) \hat{\sigma}_{j2}^2} \right]^{-2(n_2 - k + 1)/2} d\eta \quad (3.73)$$

La integral puede evaluarse directamente o aproximarse mediante una función t-univariada. Conociendo la función de densidad de  $\delta_j$  pueden encontrarse los IC del  $(1-\alpha)\%$  de confianza con más alta densidad a posteriori para todos los  $j$ ;  $j=1,2,\dots,k$ .

Si el intervalo incluye el cero se acepta  $H_0$  y en caso contrario se rechaza. Una crítica que surge inmediatamente a este método es que sólo pueden hacerse  $k$  pruebas bajo el supuesto de independencia de las muestras lo cual no se cumple en este caso. La única alternativa aceptable cuando se quiere probar si más de dos pares de coeficientes son iguales es hacer una prueba multivariada y determinar si incluye o no el cero. Lograr esto es sumamente complicado por lo que consideramos una solución aceptable la propuesta de Tsurumi.

### 3.3.3. Solución propuesta por Tsurumi y Shaffin (1984)

Estos autores también consideran el problema del modelo de regresión de dos fases donde la varianza del error es distinta para ambas fases con un enfoque bayesiano. Ellos derivan un highest posterior density interval test (HPDI) para examinar cambios en parámetros individuales entre submuestras; lo formulan en términos de una función de densidad de probabilidad  $t$  ponderada, usando como ponderación la función de densidad de

probabilidad a posteriori del radio de los desvíos estandar. Este método tiene la ventaja, respecto del descrito en 3.3.2 de ser más eficiente computacionalmente. Para ello definen las siguientes variables:

$$\omega = \frac{\sigma_1}{\sigma_2}, \quad \delta = \beta_2 - \beta_1, \quad Z = \begin{bmatrix} Y_1 \\ \omega Y_2 \end{bmatrix}, \quad W = [X \ V], \quad X = \begin{bmatrix} X_1 \\ \omega X_2 \end{bmatrix}, \quad V = \begin{bmatrix} 0 \\ \omega X_2 \end{bmatrix}, \quad \theta = \begin{bmatrix} \beta_1 \\ \delta \end{bmatrix}; \quad \text{que}$$

si se reemplazan en la función de verosimilitud conjunta de la muestra:

$$L(\beta_1, \delta, \sigma_1, \sigma_2) = (2\pi)^{-N/2} \sigma_1^{-n_1} \sigma_2^{-n_2} \exp\left\{-\frac{1}{2\sigma_1^2} (Z - W\xi)' (Z - W\xi)\right\} \quad (3.74)$$

las funciones de densidad a priori de los parámetros son (3.68) y (3.70) con lo que la función de densidad a posteriori es:

$$\pi_1(\beta_1, \delta, \sigma_1, \sigma_2 / \text{datos}) \propto \sigma_1^{-(n_1+1)} \sigma_2^{-(n_2+1)} \exp\left\{-\frac{1}{2\sigma_1^2} (Z - W\xi)' (Z - W\xi)\right\} \quad (3.74)$$

Si se realiza ahora un nuevo cambio de variable  $(\beta_1, \delta, \sigma_1, \sigma_2) \Rightarrow (\beta_1, \delta, \sigma_1, \omega)$  donde el jacobiano de la transformación es  $J = \sigma_1 \omega^{-2}$ , con lo que las funciones de densidad a posteriori de los parámetros son:

$$\begin{aligned} \pi_1(\beta_1, \delta, \omega / \text{datos}) &\propto \omega^{n_2-1} [(Z - W\hat{\xi})' (Z - W\hat{\xi}) + (\xi - \hat{\xi})' W' W (\xi - \hat{\xi})]^{-N/2}, \\ \pi_1(\beta_1, \delta, \omega / \text{datos}) &\propto \omega^{n_2-1} [(Z - W\hat{\xi})' (Z - W\hat{\xi}) + (\xi - \hat{\xi})' D (\xi - \hat{\xi})]^{-N/2} \\ &\quad [1 + (\beta_1 - \tilde{\beta}_1)' \frac{X_1' X_1}{\Xi} (\beta_1 - \tilde{\beta}_1)]^{-N/2} \end{aligned} \quad (3.76)$$

donde:  $a = (Z - W\hat{\xi})' (Z - W\hat{\xi}) + (\xi - \hat{\xi})' D (\xi - \hat{\xi})$ ,  $\tilde{\beta}_1 = \hat{\beta}_1 + H(\xi - \hat{\xi})$ ,  
 $H = \omega^2 X_2' X_2 (X_1' X_1)^{-1}$  y  $D = \omega^2 X_2' [I - \omega^2 X_2 (X_1' X_1)^{-1} X_2'] X_2$ . El cambio de variable se realizó para poder integrar respecto a  $\beta_1$  con lo que la función de densidad a posteriori de  $(\delta, \omega)$  es

$$\pi_1(\delta, \omega / \text{datos}) \propto \omega^{n-1} |X_1' X_1|^{-1/2} [(Z - W\hat{\xi})' (Z - W\hat{\xi}) + (\xi - \hat{\xi})' D (\xi - \hat{\xi})]^{-(N-K)/2} \quad (3.77)$$

$$\pi_1(\omega, \delta_j / \text{datos}) \propto \omega^{n-1} [(Z - W\hat{\xi})' (Z - \omega\hat{\xi})]^{-(N-2k+1)/2} |X'X|^{-1/2} |D_{jj}|^{-1/2} \left[ 1 + \frac{(\delta_j - \hat{\delta}_j)^2}{(Z - \omega\hat{\xi})' (Z - \omega\hat{\xi}) d^{jj}} \right]^{(N-2k+1)/2} \quad (3.7)(3.78)$$

Si  $R = (Z - \omega\hat{\xi})' (Z - \omega\hat{\xi})$

$$\pi_1(\omega, \delta_j / \text{datos}) \propto |X'X|^{-1/2} |D| \omega^{n-1} R^{-(N-2k)/2} (R d^{jj})^{-1/2} \left[ 1 + \frac{(\delta_j - \hat{\delta}_j)^2}{R d^{jj}} \right]^{-(N-2k+1)/2} \propto \pi_1(\omega / \text{datos}) \cdot \pi_1(\delta_j / \omega, \text{datos}) \quad (3.79)$$

donde  $d^{jj}$  es el  $j$ -ésimo elemento de la diagonal de  $D^{-1}$ ,  $D_{jj}$  es la matriz  $D$  de la cual se eliminan la  $j$ -ésima fila y la  $j$ -ésima columna y  $|D| = |D_{jj}| (d^{jj})^{-1}$ .

$$\pi_1(\omega / \text{datos}) \propto \omega^{n-1} |X'X|^{-1/2} |D| R^{-(N-2k)/2} \quad (3.80)$$

$$\pi_1(\delta_j / \omega, \text{datos}) \propto (R d^{jj})^{-1/2} \left[ 1 + \frac{(\delta_j - \hat{\delta}_j)^2}{R d^{jj}} \right]^{-(N-2k+1)} \quad (3.81)$$

que es una función  $t$  de Student.

$$\pi_1(\delta_j / \text{datos}) = \int_{\omega}^{\infty} \pi_1(\delta_j / \omega, \text{datos}) \pi_1(\omega / \text{datos}) d\omega \propto \omega_0 |X'X|^{-1/2} |D_{jj}|^{-1/2} \omega^{n-1} R^{-(N-2k+1)/2}$$

$$\left[ 1 + \frac{(\delta_j - \hat{\delta}_j)^2}{R d^{jj}} \right]^{-\frac{(N-2k+1)}{2}} d\omega, \quad (3.1) \quad (3.82)$$

que es una función de densidad  $t$  ponderada con las ponderaciones iguales a la función de densidad marginal a posteriori de  $\omega$ . Para evaluar  $\pi_1(\delta_j/\text{datos})$  se puede usar un método de integración numérica o, si se usa el hecho de que  $\pi_1(\delta_j/\omega, \text{datos})$  es una  $t$  univariada, llevar a cabo un test  $t$  condicionado en la media a posteriori de  $\omega$ . Una vez evaluada  $\pi_1(\delta_j/\text{datos})$  pueden encontrarse los IC del  $(1-\alpha)\%$  de confianza con más alta densidad a posteriori para todos los  $j$ . Si el intervalo incluye el cero se acepta  $H_0$  y en caso contrario se rechaza.

Los autores analizan los datos de una función de producción dada en Maddala ["ECONOMETRICS", 1977, Mc Graw-Hill, New York, pag. 196-199] mediante el test de Chow (1960) y los dos métodos bayesianos propuestos. El test de Chow da el valor 1,31 que es menor que el valor de  $F_{(3;33;0,5)} = 2,90$ , por lo que aparentemente no puede rechazarse la hipótesis de que la relación es estable. Esto puede deberse al hecho de que para realizar esta prueba se supone que la varianza del error es la misma para ambas fases de la regresión, pero el test estandar de homoscedasticidad da como resultado  $(0,04573)^2 / (0,0145)^2 = 9,93$ , cuyo valor de probabilidad es menor del 1%; lo cual sugiere que existen razones para rechazar este supuesto. Como era previsible, no es adecuado aplicar el test de Chow en condiciones de heteroscedasticidad. La función de densidad a posteriori de  $\delta_j$ ;  $\pi_j(\delta_j/\text{datos})$ ,  $j=1,2,3$  basado en la función de densidad doble  $t$  da, como era de esperar, los mismos resultados que la basada en la función de densidad  $t$ -ponderada. La diferencia se encuentra en que el tiempo de CPU requerido para la forma doble- $t$  es más que el doble que para la  $t$ -ponderada. La eficiencia de cómputo de esta última se debe a

que el rango de  $\omega$  es  $(0, \infty)$  mientras que el rango de  $\eta$  es  $(-\infty, \infty)$ . Los autores calculan los intervalos de confianza (IC) del 95% con la mayor densidad a posteriori (HPDI) para  $\pi_1(\delta_j/\text{datos})$ ; la aproximación  $t$  condicionada en la media a posteriori de  $\omega$ , la aproximación de Patil y los IC del 95% basado en el estadístico  $t$  que supone homoscedasticidad. De su análisis se observa que: todos los IC del 95% incluyen el cero, es decir no rechazan  $H_0$  mientras que todos los HPDI del 95% rechazan  $H_0$ , excepto para el coeficiente capital-insumo. Esto ilustra la sensibilidad del estadístico  $t$  al supuesto "falso" de homoscedasticidad y los HPDI aproximados son razonablemente cercanos a los HPDI exactos.

#### 3.3.4. Solución propuesta por Ferreira(1975).

Como ya se discutió en la sección 3.1.3 Ferreira considera el modelo de regresión de dos fases sujeto a los siguientes supuestos:

- 1) la matriz  $X$  es una matriz regresora no estocástica,
- 2) las funciones de densidad a priori  $\pi_0(\beta_{11}), \pi_0(\beta_{21}), \pi_0(\beta_{12}), \pi_0(\beta_{22})$  y  $\pi_0(\log \sigma^2)$  son independientes y uniformes en todo el rango de variación de cada uno de los parámetros,
- 3) la función de densidad a priori de  $\tau$  es igual a cero si  $\tau$  vale 1,  $N-1$  o  $N$ .

El utiliza tres funciones a priori distintas para  $\tau$  (expresión 3.16). Si  $g' = (g_1, g_2, g_3, g_4)$  es un vector fila de constantes y  $\Omega' = (\beta_{21}, \beta_{11}, \beta_{22}, \beta_{12})$  la probabilidad de que el escalar  $g'\Omega$  pertenezca a un intervalo  $(c, d)$  cualquiera es:

$$P[g'\Omega \in (c, d) / X, Y] = \sum_{\tau=2}^{N-2} P_{N-4}^{\tau} [e^{t'(c, \tau)} - e^{t'(d, \tau)}] \pi(\tau / X, Y) \quad (3.83)$$

donde  $t'(u, \tau)$  es una función de la forma  $t'(u, \tau) = (u - g'\hat{\Omega}) / [S^2(\tau) g' C_{\tau} g / (N-4)]^{1/2}$ ,  $C_{\tau}$  es la inversa de la matriz de ecuaciones normales y  $S^2(\tau)$  está definida en (3.21).

Esta expresión permite construir intervalos de confianza para probar hipótesis de interés sobre los parámetros de la regresión. Por ejemplo si  $H_0$  es  $\beta_{21} = \beta_{22}$  entonces  $g'$  será  $(0 \ 1 \ 0 \ -1)$ ; si el intervalo calculado contiene al cero se acepta la hipótesis nula y, en caso contrario se rechaza.

Ferreira también calcula la probabilidad de que la varianza del error pertenezca a un intervalo  $(c,d)$  cualquiera; y es

$$P[\sigma \in (c,d)/X,Y] = \sum_{\tau=2}^{N-2} P[x_{N-4} \in (S(\tau)/d, S(\tau)/c)] \pi(\tau/x,Y) \quad (3.64)$$

Esto permite testar hipótesis sobre la varianza del error.

De las expresiones de las funciones a posteriori marginales de los parámetros del modelo pueden hallarse IC bayesianos y test de hipótesis para los mismos.

#### 4. TRABAJOS EN QUE LA VARIABLE QUE DETERMINA EL CAMBIO DE REGIMEN PERTENECE A LAS VARIABLES EXPLICATIVAS DEL MODELO.

##### 4.1. Estimación del número de fases de la regresión.

##### 4.1.1. Solución propuesta por Sprent (1961).

Sprent estudia el modelo de regresión lineal de dos fases cuando los datos cumplen con los supuestos del modelo de regresión lineal clásico y se conoce el intervalo  $[x_\tau, x_{\tau+1}]$ , en el cual se produce el cambio de régimen; éste es la abscisa del punto de cambio de régimen (change-over point),  $\gamma$ , pertenece al intervalo  $[x_\tau, x_{\tau+1}]$ , con  $\tau$  dado. Si esto último no se cumple Sprent sostiene que puede usarse el método de Quandt (Sección 3.1.1) para estimarlo sin consecuencias serias. La variable  $X$  está ordenada por magnitud.

El autor propone un estadístico para testar la hipótesis de que el modelo sigue una regresión de una fase versus la hipótesis alternativa de que la regresión es de dos fases. Utiliza el método de los mínimos cuadrados. El modelo bajo  $H_0$  es (3.36) mientras que bajo  $H_1$  es (3.1) con el supuesto adicional de que  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ . Las dimensiones de las matrices  $X_1$  y  $X_2$  son respectivamente  $(\tau \times 2)$  y  $X_2$   $((N-\tau) \times 2)$ . Como  $\tau$  es conocido es sencillo estimar por mínimos cuadrados  $\beta_1$  y  $\beta_2$ . Aplicando la teoría de hipótesis lineales se puede construir el estadístico

$$\frac{(SCR_{H_0} - SCR_{H_1}) / 2}{SCR_{H_1} / (N-4)} \quad (4.1)$$

que bajo la  $H_0$  se distribuye como una  $F$  con  $(2, N-4)$  grados de libertad. La regla de decisión es rechazar  $H_0$  si el valor del estadístico es mayor que el valor  $F(2, N-4)$  para un nivel de significancia previamente establecido.

En cuanto a la sugerencia de Sprent de usar el método de Quandt

cuando se desconozca el intervalo en el cual se produce el cambio de régimen tendrá inconvenientes pues modificará la distribución y potencia de los estadísticos sugeridos por él y debería encontrarse la distribución para cada caso; por esa razón nosotros no lo hemos considerado.

#### 4.1.2. Solución propuesta por Mc Gee y Carleton (1970).

Se han ordenado las  $N(k+2)$ -uplas  $(Y_i, X_{oi}, X_{1i}, \dots, X_{ki})$  de acuerdo a la posición de  $X_{oi}$  en el ordenamiento de la variable  $X_o$ . El vector de residuos  $\epsilon$  se distribuye  $N(0, \sigma^2 I_N)$ .

Nota: Este procedimiento también es válido si la variable ordenadora no pertenece al modelo (capítulo 3).

Estos autores trataron el problema sugiriendo un procedimiento que es esencialmente una mezcla de clustering jerárquico y la teoría estándar de regresión; es más pragmático que estadístico pero tiene la ventaja de ser de fácil aplicación en una gran cantidad de problemas y la desventaja que no puede garantizarse que la solución sea óptima en el sentido analítico y el uso del criterio estadístico sólo puede ser justificado por las corridas de Monte Carlo y consideraciones a priori. Ellos consideran el modelo de una variable explicativa, nosotros lo hemos generalizado para  $k \geq 2$  y la única restricción es que el modelo debe tener la misma forma en todas las fases. Se deben ordenar las  $N(k+2)$ -uplas  $(Y_i, X_{oi}, \dots, X_{ki})$  de acuerdo al valor de  $X_o$  en cada observación. Este procedimiento es a la vez más y menos restringido que los problemas usuales de clustering: más en el sentido en que las  $(k+2)$ -uplas están ordenadas lo que elimina varios cluster potenciales y menos pues el número de cluster es desconocido y debe ser averiguado a partir de los datos. En piecewise regression se reemplaza el concepto de "distancia" con el de "bondad del ajuste". Como hay un orden en las observaciones sólo se mide la bondad del ajuste lineal para puntos adyacentes. El procedimiento se puede describir mediante los siguientes pasos:

\*Paso 1: se desean encontrar los  $(k+2)$  puntos adyacentes con el mejor ajuste lineal. Se examinan todos los conjuntos de  $(k+2)$  puntos adyacentes y se computa la bondad del ajuste en cada caso agrupándose aquellos para los cuales esta medida alcance el mínimo valor.

\*Paso 2: se calcula la bondad del ajuste para:

- (a) conjuntos de  $(k+2)$  puntos que todavía no hayan sido agrupados,
- (b) cluster del paso anterior y un punto adyacente a izquierda,
- (c) cluster del paso anterior y un punto adyacente a derecha.

La mejor de estas medidas de ajuste lineal definirá el próximo cluster.

Paso 3: se calculan las medidas de bondad del ajuste para:

- (a) conjuntos de  $(k+2)$  puntos que todavía no hayan sido agrupados,
- (b) cluster/s del paso anterior y un punto adyacente a izquierda,
- (c) cluster/s del paso anterior y un punto adyacente a derecha,
- (d) clusters adyacentes (si existiesen).

La menor de estas medidas de ajuste lineal definirá el próximo cluster.

Se repite este procedimiento hasta que la regla de detención, descrita más abajo, indique suspenderlo.

Para un cluster determinado de tamaño  $n_s$  los estimadores MC de los parámetros se calculan en la forma usual al igual que los residuos estimados; ésto es  $\hat{\beta}_s = (X_s' X_s)^{-1} X_s' Y_s$ , donde  $(Y_s X_s)$  son las matrices de las observaciones que pertenecen al cluster potencial  $s$ ;  $\hat{\epsilon}_i = Y_i - \hat{Y}_i = Y_i - X_i' \hat{\beta}$  es el residuo estimado para una

observación del cluster potencial  $s$  y  $\theta = \frac{\sum_{i=1}^{n_s} \epsilon_i^2}{n_s - (k+1)}$  es la medida de bondad del ajuste.

Si se conoce de antemano el número de fases del modelo el procedimiento se detendrá cuando el número de clusters sea igual al mismo. En cambio si el número de fases es desconocido la regla de detención propuesta por los autores es:

- a) Si el tamaño del cluster es del menor tamaño posible se puede realizar un test de significancia de la relación entre  $Y$  y  $X$  en el mismo.

- b) Si se trata de un cluster de tamaño  $n$  al que se le agregó un punto se puede probar si el punto adicional afecta seriamente la estructura lineal del cluster fijo preexistente.
- c) Si el nuevo cluster es la unión de dos cluster pre-existentes de tamaño  $n_i$  y  $n_j$  se desea probar si los parámetros de ambas regresiones son iguales, es decir si las observaciones de ambos cluster pertenecen a la misma población.

Un estadístico adecuado para (b) y (c) es el propuesto por Chow (1960), descrito en 3.2.1 (expresión 3.39); si supera el valor de  $F(k_i, n_i + n_j - 2(k+1))$  para un nivel de significancia previamente determinado se puede sostener que el punto no debe ser unido al cluster en (b) o que los cluster deben permanecer separados en (c).

Esta regla puede ser modificada de acuerdo al número de puntos aislados que se permitan. Este método indica los intervalos donde se producen discontinuidades pero no la ubicación exacta de los mismos.

Existe una dificultad considerable en desarrollar una teoría formal de muestreo para cualquier tipo de procedimiento jerárquico como el que se ha desarrollado (aún sin haber usado este método la determinación de los puntos de cambio de régimen causa problemas pues la razón de verosimilitud es función de variables continuas y discretas, lo cual puede ser intratable), por ello los autores han recurrido a las corridas de Monte Carlo para indicar la estabilidad de la regresión lineal a trozos. Este es sólo un método de "comprobación".

Este es un problema de enormes proporciones pues aun si se fija el número de fases existe una infinidad de modelos que se pueden ajustar, tal que difieran en el valor de los coeficientes y la distribución de los residuos; por lo que se puede crear una gran variedad de reglas de decisión para determinar qué nivel aceptar como "la solución".

El método presenta cuatro problemas principales:

- 1) Tamaño del cluster: si el tamaño mínimo de cluster que se permite varía ( $n_i > k+2$ ) para tamaños de muestras grandes puede arrivarse a soluciones distintas y si el menor tamaño de cluster

probado es demasiado grande es posible que se pierda mucha información. El único consejo que los autores brindan es calcular el estadístico de Durbin-Watson y un gráfico de residuos en cada trozo y, con experiencia, el investigador puede decidir cuál es el tamaño mínimo de cluster que debe probarse y además detectar si se trata de un modelo polinomial.

2) Intervalo demasiado largo entre las observaciones: puede suceder que este espacio sea cubierto por uno de los cluster estimados, sin ser ésto razonable; el investigador debe examinar cuán compacto es el cluster potencial antes de dejarlo fijo. A los cluster potenciales menos compactos se les asignará menos prioridad.

3) Observaciones extremas (outliers): muchas veces la presencia de este tipo de observaciones impide arriivar fácilmente a las conclusiones y muchos investigadores tienen fuertes convicciones para descartarlos. El procedimiento propuesto permite identificar fácilmente estos puntos pues son los últimos en ser incluidos en los cluster.

4) Observaciones repetidas: si para un valor fijo de  $x'$  se poseen observaciones repetidas de  $Y$  el análisis es una extensión del procedimiento propuesto pero sólo se debe exigir que las observaciones repetidas pertenezcan al mismo cluster.

#### 4.1.3. Solución propuesta por Bellman y Roth (1969).

Estos autores proponen un método para ajustar una curva dados los datos. No corresponde exactamente al modelo de regresión a trozos tratado en la tesis pero lo hemos incluido como un aporte más para solucionar estos problemas. Ellos aplican las ideas de programación dinámica y definen como el mejor ajuste a aquel que minimiza el desvío absoluto entre la curva ajustada y los datos dados.

Dada la función  $g(x)$ , cuyos valores están dados en el intervalo  $[0, a]$  los autores desean aproximar esta curva mediante una secuencia de segmentos, cada uno unido a sus vecinos en los

El test de razón de verosimilitud propuesto es:

$$U = \frac{Z_{\hat{\tau}}(\hat{\gamma})}{S} \quad (4.3)$$

donde  $\hat{\gamma}$  es el estimador MV de la abscisa del punto de cambio de régimen,

$$Z_{\hat{\tau}}(\hat{\gamma}) = \frac{(\tilde{\beta}_{02(\hat{\tau})} - \tilde{\beta}_{01(\hat{\tau})})^2 [C_{\hat{\tau}} - D_{\hat{\tau}}(\tilde{\gamma}_{\hat{\tau}} + \hat{\gamma}) + E_{\hat{\tau}} \tilde{\gamma}_{\hat{\tau}} \hat{\gamma}]^2}{C_{xx,N} (C_{\hat{\tau}} - 2D_{\hat{\tau}} \hat{\gamma} + E_{\hat{\tau}} \hat{\gamma}^2)} \quad (4.4)$$

$\hat{\tau}$  es el estimador del intervalo en el cual se produce el cambio de régimen,  $\tilde{\beta}_{01(\hat{\tau})}$  y  $\tilde{\beta}_{02(\hat{\tau})}$  son los estimadores MC locales de la pendiente de regresión,

$$C_{\hat{\tau}} = C_{xx,\hat{\tau}} + C_{xx,\hat{\tau}}^* + \frac{\hat{\tau} \hat{\tau}^*}{N} (\bar{x}_{\hat{\tau}}^2 C_{xx,\hat{\tau}}^* + x_{\hat{\tau}}^{*2} C_{xx,\hat{\tau}}), \quad (4.5)$$

$$D_{\hat{\tau}} = \frac{\hat{\tau} \hat{\tau}^*}{N} (\bar{x}_{\hat{\tau}} C_{xx,\hat{\tau}}^* + x_{\hat{\tau}}^- C_{xx,\hat{\tau}}), \quad (4.6)$$

$$E_{\hat{\tau}} = \frac{\hat{\tau} \hat{\tau}^*}{N} (C_{xx,\hat{\tau}} + C_{xx,\hat{\tau}}^*), \quad (4.7)$$

$$\hat{\tau}^* = N - \hat{\tau} \quad (4.8)$$

$$\bar{x}_{\hat{\tau}} = \frac{1}{\hat{\tau}} \sum_{i=1}^{\hat{\tau}} x_i \quad (4.9)$$

$$\bar{x}_{\hat{\tau}}^* = \frac{1}{\hat{\tau}^*} \sum_{i=\hat{\tau}+1}^N x_i \quad (4.10)$$

$$C_{xx,\hat{\tau}} = \sum_{i=1}^{\hat{\tau}} (x_i - \bar{x}_{\hat{\tau}})^2 \quad (4.11)$$

$$C_{xx,\hat{\tau}}^* = \sum_{i=\hat{\tau}+1}^N (x_i - \bar{x}_{\hat{\tau}}^*)^2 \quad (4.12)$$

$$y \quad C_{xx,N} = \sum_{i=1}^N (x_i - \bar{x})^2. \quad (4.13)$$

Existe una fuerte evidencia empírica de que  $Z(\hat{\gamma})/\sigma^2$  se distribuye asintóticamente como  $\chi^2_3$  lo que implica que bajo la hipótesis nula

U se distribuye asintóticamente como  $F_{2, N-4}$ . Este resultado está conectado con la no normalidad de  $\hat{\beta}_{02} - \hat{\beta}_{01}$  y  $\hat{\gamma}$  cuando  $\beta_{02} - \beta_{01} = 0$ , pero no ha sido verificado teóricamente. A pesar de ello es un resultado muy útil en la práctica. Encontrar la distribución del estadístico U será muy complicado pues depende de  $Z(\hat{\gamma})$  que ha sido hallado mediante procedimiento iterativos.

#### 4.1.5. Solución propuesta por Schechtman(1983).

Schechtman investiga mediante simulación el comportamiento del test de Hinkley(4.3) para el modelo con restricción de continuidad cuando los términos de error no se distribuyen normalmente. En este caso los supuestos del modelo son:

(i)  $\varepsilon_i \text{ iid } (0, \sigma^2), i=1, 2, \dots, N$

(ii)  $x_\tau \leq \gamma < x_{\tau+1}$

(iii)  $\beta_{11} + \beta_{01}\gamma = \beta_{12} + \beta_{02}\gamma$

donde los parámetros desconocidos son  $\beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}, \gamma, \tau$  y  $\sigma^2$  y la hipótesis de interés es  $H_0: \beta_{01} = \beta_{02}$  vs  $H_1: \beta_{01} \neq \beta_{02}$ .

Para determinar qué consecuencia tiene en el tamaño del test y la potencia del mismo la violación del supuesto de normalidad Schechtman realizó numerosas simulaciones y arribó a las siguientes conclusiones:

(a) Consecuencias en el tamaño del test:

Se generaron quinientas muestras de tamaño sesenta ( $N=60$ ), para cinco distribuciones de los residuos: normal, exponencial, doble exponencial, Cauchy y uniforme  $[-1,7;1,7]$ .

Para cada una de las muestras se calculó el valor de la variable dependiente ( $Y_i$ ) usando la regresión:  $Y_i = 20 + 50i + \varepsilon_i$  con  $i$  que varió entre uno y sesenta. En ella no se produjo cambio en la pendiente y se calculó el número de veces que el estadístico resultó mayor que el valor crítico  $F_{(3,60-4)}$  para  $\alpha = 0,01; 0,05$  y  $0,10$ . Cuando la distribución de los residuos es Cauchy, exponencial y doble exponencial no se mantiene el nivel de

significancia del test, siendo la primera de ellas la que presenta peor comportamiento. Para el caso normal (0,1) los valores críticos de F son muy grandes, con lo que el test de Hinkley es conservativo. Es decir que el test no tienen el nivel de significación nominal ni siquiera cuando se cumple el supuesto de normalidad.

La siguiente tabla, publicada en el trabajo de Schechtman muestra la estimación del nivel real designificación para las distintas distribuciones de los residuos.

Tabla 1:  
Estimación del nivel de significación.

Nominal level $\alpha$	Underlying distributions				
	Cauchy	Exponential	Double exponential	Normal	Uniform
0.01	0.0690	0.0372	0.0210	0.0072	0.0048
0.05	0.0978	0.0812	0.0688	0.0442	0.0240
0.10	0.1234	0.1262	0.1154	0.0940	0.0678

(b) Consecuencias en la potencia del test:

Se hizo un estudio de potencia de Monte Carlo para comparar la potencia del test para distintas distribuciones de los residuos. Como ésta depende de los parámetros abscisa del punto de cambio ( $\gamma$ ) y cambio de pendiente ( $\delta = \beta_{02} - \beta_{01}$ ); para cada una de las siguientes alternativas se generaron cinco mil muestras de tamaño sesenta considerando para cada una las cinco distribuciones mencionadas en (a) con  $X_i = i$ ;  $i = 1, 2, \dots, 60$

- 1)  $(\gamma, \delta) = (30; 0, 2)$  y  $\beta_{01} = 50$
- 2)  $(\gamma, \delta) = (30; 0, 2)$  y  $\beta_{01} = 1$
- 3)  $(\gamma, \delta) = (20; 0, 2)$  y  $\beta_{01} = 1$
- 4)  $(\gamma, \delta) = (30; 0, 1)$  y  $\beta_{01} = 1$

De él se concluyó que la potencia aumenta cuando aumenta el tamaño del cambio de la pendiente y  $\gamma$  se acerca a  $N/2$ . (La función de potencia es, para valores fijos de  $\delta$ , simétrica en el punto de cambio de régimen  $\gamma$  alrededor de  $\gamma = \frac{N}{2} = 30$ ).

La siguiente tabla, publicada en el trabajo de Schechtman, muestra la potencia del test de Hinkley cuando los residuos tienen una de las distribuciones ya citadas, para diversos valores de  $\gamma$  y  $\delta$ .

Tabla 2: Estudio de potencia de Monte Carlo.

$\alpha$	Doubie				
	Exponential	exponential-	Cauchy	Normal	Uniform
a) $(r, \delta) = (30, 0.2), \beta_1 = 50$					
0.01	0.9945	1.000	0.0936	0.9996	1.000
0.05	0.9986	1.000	0.1726	1.000	1.000
0.10	0.9994	1.000	0.2396	1.000	1.000
b) $(r, \delta) = (30, 0.2), \beta_1 = 1$					
0.01	0.9948	1.000	0.0936	0.9996	1.000
0.05	0.9986	1.000	0.1726	1.000	1.000
0.10	0.9994	1.000	0.2396	1.000	1.000
iii) $(r, \delta) = (20, 0.2)$					
0.01	0.9664	0.9998	0.0818	0.9804	0.9888
0.05	0.9918	0.9998	0.1440	0.9960	0.9982
0.10	0.9968	0.9998	0.1954	0.9984	0.9992
iv) $(r, \delta) = (30, 0.1)$					
0.01	0.5566	0.9048	0.0706	0.5360	0.5530
0.05	0.7628	0.9702	0.1148	0.7682	0.7880
0.10	0.8480	0.9870	0.1550	0.8618	0.8766

Resumiendo se puede decir que cuando los residuos no se distribuyen normalmente no se mantiene el nivel de significación del test pero la potencia del mismo es muy alta, lo que hace que siga siendo conveniente la utilización del mismo.

1) Si se conoce de antemano qué observaciones pertenecen a cada régimen y se deriva la función de verosimilitud conjunta de la muestra respecto a  $\beta_1, \beta_2, \sigma_1^2, \sigma_2^2$  se encuentra la expresión de los estimadores MV de los mismos, pero no una expresión explícita de  $\gamma$ . Para ello debe hacerse un cambio de variable 1-1 en (4.15). Se definen dos nuevas variables  $\gamma_j$  y  $\delta_j$  donde  $\gamma_j = \beta_{j1} - \beta_{j2}$  si  $j$  es menor o igual que  $q$  ó  $\gamma_j = \beta_{j1}$  si  $j$  es mayor que  $q$ ; y  $\delta_j = \gamma_j / \gamma_k$ . El cambio de variable que se produce es:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \longrightarrow \begin{bmatrix} \gamma \\ \beta_2 \end{bmatrix} \longrightarrow \begin{bmatrix} \delta \\ \beta_2 \end{bmatrix}$$

con lo que (4.15) se puede expresar como:

$$x' \gamma = \gamma_k x' \delta = \gamma_k (x - x^{(1)}) (x - x^{(2)}) \dots (x - x^{(k)}) = 0 \quad (4.16)$$

donde  $x' = (1 \ x \dots x^k)$  y  $x^{(1)}, x^{(2)}, \dots, x^{(k)}$ , son las raíces del polinomio. Si se introducen estos cambios de variables en la SCR de la primera fase de la regresión ( $SCR_1$ ), ésta queda expresada por:

$$\begin{aligned} SCR_1 &= (Y_1 - X_1 \beta_1)' (Y_1 - X_1 \beta_1) \\ &= (Y_1 - X_1 \beta_1 + X_{11} \beta_2 - X_{11} \beta_2)' (Y_1 - X_1 \beta_1 + X_{11} \beta_2 - X_{11} \beta_2) \\ &= (Y_1 - X_1 \gamma - X_{11} \beta_2)' (Y_1 - X_1 \gamma - X_{11} \beta_2) \\ &= (Y_1 - X_1 \gamma \delta - X_{11} \beta_2)' (Y_1 - X_1 \gamma \delta - X_{11} \beta_2) \\ &= \left[ Y_1 - (X_1 - X^{(1)} 1) (X_1 - X^{(2)} 1) \dots (X_1 - X^{(k)} 1) - X_{11} \beta_2 \right]' \\ &\quad \left[ Y_1 - (X_1 - X^{(1)} 1) (X_1 - X^{(2)} 1) \dots (X_1 - X^{(k)} 1) - X_{11} \beta_2 \right] \end{aligned} \quad (4.17)$$

donde  $1_{(n \times 1)}$  es la matriz suma y  $X_{11}$  está formado por las primeras  $(q+1)$  columnas de  $X_1$ .

Si se realiza un cambio de variable análogo en la SCR de la segunda fase ( $SCR_2$ ), llamando  $X_{12}$  a la matriz formada por las últimas  $(k - (q+1))$  de  $X$ ; la función de verosimilitud conjunta de

la muestra se puede expresar:

$$L = \frac{1}{(2\pi)^{N/2} \sigma_1^{n_1} \sigma_2^{n_2}} \exp\left(-\frac{1}{2\sigma_1^2} SCR_1\right) \exp\left(-\frac{1}{2\sigma_2^2} SCR_2\right) \quad (4.18)$$

De esta manera se encuentran expresados en forma explícita las  $k$  intersecciones de ambos polinomios. En este caso los parámetros a estimar son  $\beta_2, \sigma_1^2, \sigma_2^2$  y las  $k$  intersecciones  $x^{(1)}, \dots, x^{(2)}$ . Como la expresión de los estimadores se obtiene mediante una transformación 1-1 de los estimadores MV de  $\beta_1$  y  $\beta_2$ , poseerán propiedades asintóticas óptimas por la propiedad de invarianza. Lo ideal será encontrar la distribución de  $\hat{\gamma}$ , pero como es una solución de una ecuación polinomial no hay, en general, una fórmula explícita para cualquier  $\hat{\gamma}$  en término de los coeficientes de la ecuación de regresión y esto complica la búsqueda de la distribución. Robinson lo solucionó buscando IC para  $\gamma$  directamente. Estos aunque contienen a  $\gamma$  no son necesariamente simétricos alrededor de él. Para ello define una v.a.U:

$$U = (C_0 \ C_1 \ \dots \ C_k) \begin{pmatrix} \gamma^0 \\ \gamma^1 \\ \vdots \\ \gamma^k \end{pmatrix} = C' \gamma = \gamma' C \quad (4.19)$$

donde  $C = \hat{\beta}_1 - \hat{\beta}_2$  y  $\gamma$  es el shift-point; es decir  $U$  es la distancia entre las dos regresiones estimada en el punto donde se produce el cambio de régimen y se distribuye normal  $(0, \sigma_u^2)$ . Las varianzas de  $\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}$  y  $U$  son respectivamente:

$$\text{Var}(\hat{\beta}_1) = \sigma_1^2 (X_1' X_1)^{-1} \quad (4.20)$$

$$\text{Var}(\hat{\beta}_2) = \sigma_2^2 (X_2' X_2)^{-1} \quad (4.21)$$

$$\text{Var}(\hat{\gamma}) = \begin{pmatrix} (X'_{11} X_{11})^{-1} & \sigma_1^2 \\ (X'_{12} X_{12})^{-1} & \sigma_2^2 \end{pmatrix} + \begin{pmatrix} (X'_2 X_2)^{-1} & \sigma_2^2 \\ & 0 \end{pmatrix} = M \quad (4.22)$$

$$\text{Var}(U) = \sigma_U^2 = r' \text{Var}(\hat{\gamma}) r = (\gamma^0 \dots \gamma^k) M \begin{pmatrix} \gamma^0 \\ \gamma \\ \vdots \\ \gamma^k \end{pmatrix} = \sum_{i=1}^k \sum_{j=1}^k m_{ij} \gamma^i \gamma^j = \sigma_1^2 P(\gamma) + \sigma_2^2 Q(\gamma) \quad (4.23)$$

Como el cociente  $U/\sigma_U$  se distribuye  $N(0,1)$ , la  $SCR_1$  como una  $\chi^2 \sigma_1^2$  con  $(n_1 - k - 1)$  g.l. y la  $SCR_2$  como una  $\chi^2 \sigma_2^2$  con  $(n_2 - q - 1)$  g.l. y  $U$  es independiente de  $SCR_1$  y de la  $SCR_2$ , se puede construir el estadístico  $R$

$$R = \frac{\frac{U}{\sigma_U}}{\sqrt{\frac{SCR_1}{\sigma_1^2} + \frac{SCR_2}{\sigma_2^2}}} \sqrt{(n_1 + n_2 - k - q - 2)} \quad (4.24)$$

que se distribuye como una  $t(\gamma, \sigma_1^2, \sigma_2^2)$ .  $R$  contiene tres parámetros desconocidos  $(\gamma, \sigma_1^2, \sigma_2^2)$ . Si se supone que  $\sigma_1^2 = \sigma_2^2$ , o más generalmente que  $\sigma_1^2 = h^2 \sigma_2^2$ , para alguna constante conocida  $h$  mayor que cero el estadístico será:

$$R = \frac{\sqrt{h^2 P(\gamma) + Q(\gamma)}}{\sqrt{\frac{SCR_1}{h^2} + SCR_2}} \sqrt{n_1 + n_2 - k - q - 2} \quad (4.25)$$

que también tendrá una distribución  $t$  pero sólo involucra un sólo parámetro desconocido  $\gamma$ . Los límites de confianza para niveles de significancia  $\alpha$  son dados por  $-t_{1-\alpha/2} < t < t_{1-\alpha/2}$  ó  $0 \leq t^2 < t_{1-\alpha/2}^2$ .

Sea  $H = \sqrt{\frac{SCR_1/h + SCR_2}{n_1 + n_2 - k - q - 2}}$ . De todo los  $\gamma$  definidos por la ecuación polinomial original, los que satisfacen la restricción

$$0 \leq U^2(\gamma) < \left[ h^2 P_{2k}(\gamma) + Q_{2q}(\gamma) \right] H^2 t_{1-\alpha/2}^2 \quad (4.26)$$

constituyen la región de confianza. Los límites de la región se obtienen resolviendo la ecuación polinomial de grado  $2k$

$$U^2(\gamma) = \left[ h^2 P_{2k}(\gamma) + Q_{2q}(\gamma) \right] H^2 t_{1-\alpha/2}^2 \quad (4.27)$$

Las raíces de la ecuación  $U(\gamma)=0$ , ésto es, las abscisas de los puntos de intersección de los polinomios estudiados, están incluidos en el conjunto de las raíces de la ecuación  $t(\gamma)=0$  que está contenida en la unión desde  $i$  igual a uno hasta  $a$  de los intervalos  $I_i$ . Este polinomio recibe el nombre de IC polinomial. Pares de raíces apropiadamente identificadas con cada extremo definen IC para cada raíz  $x^{(i)}$ . Si todas las  $2k$  raíces son reales la estimación MV de  $\gamma$  está contenida en dicho conjunto. La región definida por  $-t_{1-\alpha/2} < t(\gamma) < t_{1-\alpha/2}$  ha sido reemplazada por un conjunto equivalente de intervalos alrededor de  $\gamma$ .

2) Si no se conoce de antemano qué observaciones pertenecen a cada régimen la función de verosimilitud conjunta de la muestra es función de  $(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, n_1)$ . Robinson realiza la misma sugerencia que Quandt: para cada elección de  $(n_1, n_2)$  estimar  $(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2)$  y calcular el valor de la función de verosimilitud de la muestra; al par  $(n_1, n_2)$  que maximiza dicha expresión se lo considera como el estimador MV de  $n_1$  y las estimaciones de  $\beta_1, \beta_2, \sigma_1^2$ , y  $\sigma_2^2$  obtenidas a partir del mismo son las que se usarán en la estimación de  $\gamma$ . Frecuentemente no es necesario probar todos los pares posibles sino sólo los "factibles". Se debe tener en cuenta que  $n_1$  sea mayor que  $(k+2)$  y que  $n_2$  sea mayor que  $(q+2)$  para no obtener estimaciones degeneradas de  $\sigma_1^2$  y  $\sigma_2^2$ . Al estimar a  $(n_1, n_2)$  en esta forma la teoría de distribución para el caso de  $n_1$  conocido no podrá

aplicarse.

El autor sostiene que para completar la teoría sería necesario calcular niveles de potencia para cualquier IC dado para una elección de  $n_1$  cuando alguna otra alternativa es correcta.  $t(\gamma)$  no tendrá una distribución  $t$  pues depende del valor de  $n_1$  desconocido y los IC calculados serán aproximados y no exactos. Este problema no ha sido resuelto hasta el momento.

Robinson discute algunas de las dificultades prácticas en las cuales los procedimientos propuestos por él no son enteramente adecuados. (Se refiere a las dificultades cuando  $n_1$  es conocido, es decir que se conoce el intervalo  $I=[x_{n_1}, x_{n_1+1}]$  donde se produce el cambio de régimen). Estos inconvenientes son:

- 1) Las dos fases de la regresión tienen más de una intersección en  $I$ .
- 2) Las dos fases de la regresión tienen intersecciones reales cerca de  $I$  pero no en él.
- 3) Las dos fases de la regresión tienen intersecciones reales pero no en  $I$  ni cerca de él.
- 4) Las dos fases de la regresión no tienen intersección real en absoluto.
- 5) Los IC polinomiales no tienen  $2k$  raíces reales y/o las raíces encontradas no pueden ser identificadas con ninguna intersección  $X^{(i)}$ .

Lo primero que él recomienda cuando aparecen estos inconvenientes es controlar los cálculos y los supuestos del modelo. Si ellos son satisfactorios la dificultades pueden hallarse en la muestra y en el hecho de que al estimar  $\gamma$  es forzoso a veces hacer extrapolaciones lo cual produce errores. Las dificultades pueden ser consecuencias no usuales de la teoría para ciertos valores de los parámetros. En este caso propone, para cada caso, las siguientes soluciones:

- 1) Cuando no hay información adicional se puede realizar una elección aleatoria entre todas ellas. (Ya hemos indicado que el método propuesto no permite determinar cuál de las  $k$  raíces es el shift-point).

2) Robinson propone el uso de IC mejorados. El intervalo I puede ser tratado como un intervalo del 100% de confianza para  $\gamma$ . Usando la teoría dada para construir IC se elige un valor de  $\alpha$  suficientemente pequeño de manera que el IC construido (J) contenga al estimador MV  $\gamma$ . Luego se toma la intersección  $G=I \cap J$  como el IC mejorado. El valor de  $\alpha$  puede ser lo suficientemente pequeño como para garantizar que la intersección sea no vacía. El nuevo IC es del  $(1-\alpha)\%$  de confianza debido a la independencia de los intervalos.

$$\begin{aligned} P(\gamma \in G) &= P(\gamma \in I) \cdot P(\gamma \in J) \\ &= (1.0)(1-\alpha) = 1-\alpha \end{aligned}$$

Para obtener una estimación puntual de  $\gamma$  se debe considerar el límite de G a medida que  $(1-\alpha)$  tiende a 1. Para algún valor de  $\alpha$ , G consistirá en un solo punto que es, forzosamente, uno de los extremos de I. Ambos extremos del intervalo pueden obtenerse en un ejemplo simétrico. Ahora la teoría de distribución está implícita en la construcción del intervalo de confianza mejorado. Para estimar la ordenada del punto de cambio de régimen se minimiza la suma ponderada

$$R = \frac{1}{\sigma_1^2} \left[ Y_* - \hat{Y}_1(\hat{\gamma}) \right]^2 + \frac{1}{\sigma_2^2} \left[ Y_* - \hat{Y}_2(\hat{\gamma}) \right]^2 \quad (4.28)$$

donde  $\hat{Y}_1(\hat{\gamma})$  es el valor estimado de la ordenada de  $\hat{\gamma}$  mediante la primera fase de la regresión y  $\hat{Y}_2(\hat{\gamma})$  es el valor estimado de la ordenada de  $\hat{\gamma}$  mediante la segunda fase de la regresión.

$$\hat{Y}_* = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2} \hat{Y}_1(\hat{\gamma}) + \frac{\hat{\sigma}_1^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2} \hat{Y}_2(\hat{\gamma}) \quad (4.29)$$

3) Y 4) Las soluciones para el problema (3) y (4) se da en forma conjunta. Robinson propone definir un procedimiento de estimación que concuerde con el procedimiento MV cuando las

técnicas de MV produzcan una estimación de  $\hat{\gamma}$  útil y cuando esto no suceda de un estimador sensible con propiedades deseables. Es decir que la intención es proceder en el espíritu de la MV. Sea la región R contenida entre las rectas verticales  $X_{n_1}$  y  $X_{n_1+1}$ . Si la estimación  $MV(\hat{\gamma}, \hat{Y}^*(\hat{\gamma}))$  pertenece a R y hubiese sido usado en la estimación no hubiera contribuido a la SCR y por lo tanto no afectará el valor máximo de la verosimilitud. Si el estimador MV no pertenece a R se determinará un punto  $(\tilde{\gamma}, \tilde{Y})$  que sí pertenece, que si hubiese sido incluido en la estimación hubiese reducido el máximo de la función de verosimilitud el mínimo. A esta idea se la llama criterio de quasi-verosimilitud (ql).

$$\tilde{L} = L \frac{1}{\sqrt{2\pi} \sigma_1} \exp\left(-\frac{1}{2\sigma_1^2} (\tilde{Y} - \tilde{\gamma}\beta_1)^2\right) \frac{1}{\sqrt{2\pi} \sigma_2} \exp\left(-\frac{1}{2\sigma_2^2} (\tilde{Y} - \tilde{\gamma}\beta_2)^2\right) \quad (4.30)$$

donde  $\tilde{\gamma}_1 = (1 \ \tilde{\gamma} \ \tilde{\gamma}^2 \ \dots \ \tilde{\gamma}^k)$  y  $\tilde{\gamma}_2 = (1 \ \tilde{\gamma} \ \tilde{\gamma}^2 \ \dots \ \tilde{\gamma}^q)$ .

El único propósito de la ql es obtener un método de estimación para  $(\gamma, Y(\gamma))$  y no mejorar las estimaciones de  $\beta_1, \beta_2, \sigma_1^2$  y  $\sigma_2^2$  pues ya existen estimadores MV de estos parámetros y no pueden ser mejorados. Si se trata a  $\tilde{L}$  como una función de verosimilitud los estimadores de  $\sigma_1^2$  y  $\sigma_2^2$  involucrán al valor desconocido  $(\tilde{\gamma}, \tilde{Y})$ . Esto produce serias dificultades de cálculo pues  $\tilde{Y}^*$  depende de las varianzas. Una alternativa razonable es usar los estimadores MV de  $\sigma_1$  y  $\sigma_2$ , es decir que se las considera como constante conocidas.

Si se deriva la función de Quasi-verosimilitud respecto a cada uno de los parámetros y se iguala a cero se obtendrá la siguiente estimación de  $\tilde{Y}^*$ :

$$\tilde{Y}^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \hat{Y}_1(\tilde{\gamma}) + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \hat{Y}_2(\tilde{\gamma}) \quad (4.31)$$

Es necesario conocer  $\tilde{\gamma}$ . Para ello se debe maximizar  $\tilde{L}$  por ensayo

y error. Esta solución sólo permite obtener una estimación puntual del punto de cambio de régimen pero no se conoce nada sobre sus propiedades ni sobre su distribución. Robinson sostiene que la razón por la cual se prefieren los estimadores  $q_1$  en vez de usar directamente una abscisa en  $R$  donde las dos fases de la regresión ( $\hat{Y}_1$  y  $\hat{Y}_2$ ) estén más cerca es sobre todo por refinamiento.

Otra idea investigada para este problema es la de construir bandas de confianza  $B_1$  y  $B_2$  de igual nivel de confianza alrededor de las dos fases de la regresión y tomar la intersección de ambas como una región de confianza para  $(\gamma, Y)$ . Un estimador puntual puede ser definido como el límite cuando  $B_1 \cap B_2 \rightarrow 0$ . Este método fue usado para regresiones lineales y cuadráticas en las que dio buenos resultados pero en polinomios de mayor grado las bandas de confianza  $B_1$  y  $B_2$  (que son generalizaciones de hipérbolas) pueden no definir una intersección acotada.

5) Cuando el IC polinomial no tiene las  $2k$  raíces reales a veces pueden obtenerse algunos intervalos con un solo límite. En dichos casos es necesario identificar límites de confianza prospectivos con raíces. Esto involucra ensayo y error numérico. No se han realizado estudios adecuados sobre el tema.

Una crítica importante que debe hacerse al trabajo de Robinson es que él sólo considera el caso en que es conocido el intervalo en el cual se produce el cambio de régimen.

Robinson propone como temas de estudio para etapas posteriores los siguientes:

- \*Realizar estudios comparables en regresión múltiple
- \*Estudiar problemas de intersección de regresiones cuando las muestras son funciones continuas de procesos aleatorios.

En todos los casos será necesario más trabajo para obtener la confianza correcta para IC obtenidos de una elección incorrecta de  $n_1$ .

- \*Proponer soluciones para usar la información de que  $\gamma$  pertenece a un intervalo conocido  $I$ . En este trabajo NO se introduce esta información en la formulación de verosimilitud, por lo que se

está en condiciones de obtener teoría sobre las distribuciones ya que el espacio paramétrico de  $\alpha$  y  $\beta$  no se restringe. Blisschke en su trabajo con regresiones lineales fuerza la restricción de que  $\gamma$  pertenezca al intervalo I en las ecuaciones MC, o sea que se restringe el espacio paramétrico de  $\alpha$  y  $\beta$ , pero no obtiene ninguna teoría sobre la distribución por dificultades algebraicas.

#### 4.2.2. Solución propuesta por Hudson(1966).

Hudson analiza el mismo problema que Robinson en la sección anterior pero sujeto a las siguientes restricciones:

- 1) El número de variables explicativas es el mismo en ambas fases de la regresión ( $k=q$ ).
- 2) Ambas fases de la regresión se intersectan en el punto de cambio de régimen:  $\gamma \hat{\beta}_1 = \gamma \hat{\beta}_2$ .
- 3) El punto de cambio de régimen está comprendido entre la última observación de primer grupo y la primera del segundo:

$$x_{n_1} \leq \gamma \leq x_{n_1+1}$$

Utiliza el método de mínimos cuadrados. Hudson define cuatro formas mutuamente exclusivas y exhaustivas en que las dos fases de la regresión pueden unirse de acuerdo a la ubicación del estimador MC de  $\gamma$ .

Sea  $\theta = (\beta_1, \beta_2, \gamma, n_1)$  el vector de parámetros a estimar,  $H_0$  el conjunto de parámetros que satisfacen las restricciones (2) y (3) y  $H_1, H_2, H_3, H_4$  una partición de  $H$ . Para cada  $H_k$  se define  $R_k = \min_{\theta \in H_k} \{R(\theta)\} = R(\hat{\theta}_k)$  que es el menor residuo MC entre todos los  $\theta$  que pertenecen a  $H_k$  y  $R_0 = \min_k \{R_k\} = R(\hat{\theta})$  que es el menor residuo MC para todos los  $\theta$ .

En el algoritmo desarrollado por Hudson se buscan soluciones posibles en cada una de estas cuatro categorías y se elige la que produce el menor residuo. Se restringe la elección de los posibles puntos de cambio de régimen de manera que haya por lo menos  $(k+1)$  puntos en la segunda fase.

El dominio de la función de verosimilitud es el espacio de los valores reales de TODOS los parámetros, tanto si está incluido explícitamente como un argumento de la función de verosimilitud como si no lo está. Por ejemplo si se escribe la función de verosimilitud  $L(\beta_1, \beta_2)$  se entiende que está definida sobre la parte del espacio de  $(\beta_1, \beta_2)$  para la cual la ecuación  $\gamma' \beta_1 = \gamma' \beta_2$  tiene por lo menos una raíz real en el espacio de  $\gamma$ . Esto es distinto a lo planteado por Robinson que define  $L(\beta_1, \beta_2)$  sobre todo el espacio  $(\beta_1, \beta_2)$ , con lo que puede obtenerse una estimación compleja de  $\gamma$ . En el caso tratado por Hudson los estimadores MC son siempre MV si el término de error se distribuye normal y el dominio de la función de verosimilitud se define como se indicó más arriba.

Como ya dijimos el autor define cuatro tipos de unión mutuamente exclusivas y exhaustivas que él llamó unión tipo Uno, Dos, Tres y Cuatro. Se las describe a continuación:

a) Unión de Tipo Cuatro:

Se refiere al caso en que  $\hat{\gamma}$  es menor que el valor más pequeño de la muestra o mayor que el más grande ( $\hat{\gamma} < x_1$  ó  $\hat{\gamma} > x_N$ ). No se la considera en el trabajo pues en realidad reduce el número de fases del modelo, por lo que el estudio completo de un modelo con uno o más puntos de unión tipo cuatro es equivalente al estudio de un modelo con menos submodelos.

b) Unión de tipo uno, dos y tres:

Para definir la unión tipo uno, dos y tres es necesario definir una función  $g(\hat{\gamma})$

$$g(\hat{\gamma}) = \hat{\gamma}' \hat{\beta}_1 - \hat{\gamma}' \hat{\beta}_2 \quad (4.32)$$

cuya derivada primera es:

$$g'(\hat{\gamma}) = \frac{\delta(\hat{\gamma}' \hat{\beta}_1)}{\delta X} - \frac{\delta(\hat{\gamma}' \hat{\beta}_2)}{\delta X} \quad (4.33)$$

Los valores de  $g'(\hat{\gamma})$  y el hecho que  $\hat{\gamma}$  coincida o no con el valor de una observación determinan el tipo de unión como se señala en el siguiente cuadro:

	$\hat{\gamma} \neq X_i$	$\hat{\gamma} = X_i$
$g'(\hat{\gamma}) > 0$	Tipo Uno	Tipo Dos
$g'(\hat{\gamma}) = 0$	Tipo Tres	Tipo Dos

La Unión de tipo tres no se aplica si ambas fases de la regresión son rectas pues  $\hat{\gamma}'\beta_1 = \hat{\gamma}'\beta_2$  y  $g'(\hat{\gamma})=0$  indicarán que se trata de un caso degenerado.

El autor estudia por separado los casos en que el tipo de unión es conocido que cuando es desconocido.

A) Tipo de unión conocido

AI) Unión tipo Uno:

Hudson prueba que si la unión es de Tipo Uno el estimador MC de  $\gamma$  consiste simplemente en los estimadores MC locales.

Se desea minimizar  $R=(Y-X\beta)'(Y-X\beta)$  sujeto a las restricciones:  $\hat{\gamma}'\beta_1 = \hat{\gamma}'\beta_2$  y  $X_{n_1} < \hat{\gamma} < X_{n_1+1}$ . Se supondrá inicialmente que  $n_1$  es conocido. Para poder encontrar los estimadores de  $\beta_1$  y  $\beta_2$  Hudson presenta algunos teoremas que nosotros hemos replanteado para hacerlos consecuentes con el modelo general al cual siempre hacemos referencia.

Teorema 1:

Sea  $Y=f(X)=f_1(X,\beta_1) I_{(\alpha_1,\gamma)}(X) + f_2(X,\beta_2) I_{(\gamma,x_N)}(X)$

con  $f_1(\gamma,\beta_1) = f_2(\gamma,\beta_2)$  donde  $f_i = x'\beta_i$ ,  $i=1,2$ .

Sean  $A=a(X_1, \dots, X_n)$ ,  $B_1=b(X_1, \dots, X_n)$  y  $B_2=b(X_1, \dots, X_n)$  los estimadores MC para este modelo. Si  $\hat{\gamma}=a(X_1, \dots, X_n)$  y  $\hat{\beta}_j=b_j(X_1, \dots, X_n)$ ;  $j=1,2$  (valor que toman las v.a.  $A, B_1$  y  $B_2$  para los datos muestrales  $X_1=x_1, \dots, X_N=x_N$ ) son tales que:

(i)  $X_{n_1} < \hat{\gamma} < X_{n_1+1}$

$$(ii) \quad g'(\hat{\gamma}) = \begin{bmatrix} \frac{\delta f_1}{\delta X} \\ \beta_1 = \hat{\beta}_1 \end{bmatrix}_{X=\hat{\gamma}} - \begin{bmatrix} \frac{\delta f_2}{\delta X} \\ \beta_2 = \hat{\beta}_2 \end{bmatrix}_{X=\hat{\gamma}} \neq 0$$

entonces:  $\hat{\beta}_j = \beta_j^*(n_j)$ ,  $j=1,2$  y  $\hat{\gamma} = \gamma^*(n_1)$  donde los  $\beta_j^*(n_j)$  son los estimadores MC locales de cada una de las fases de la regresión (obtenidos usando separadamente los datos  $x_1, \dots, x_{n_1}$  y  $x_{n_1+1}, \dots, x_n$ ) y  $\gamma^*(n_1)$  es la intersección de  $x_1' \beta_1^*$  y  $x_2' \beta_2^*$ .

Demostración:

$$S = (Y - X\beta)'(Y - X\beta) + 2\lambda((f_1(\gamma, \beta_1) - f_2(\gamma, \beta_2)))$$

Si se deriva  $S$  respecto a cada uno de los parámetros de la expresión y se iguala a cero se obtienen los estimadores MC de los mismos.

$$\frac{1}{2} \frac{\delta S}{\delta \gamma} = \lambda g'(\hat{\gamma}) = 0, \text{ si } \frac{\delta S}{\delta \gamma} \text{ existe en } \hat{\gamma}$$

$$\frac{1}{2} \frac{\delta S}{\delta \lambda} = f_1(\hat{\gamma}, \hat{\beta}_1) - f_2(\hat{\gamma}, \hat{\beta}_2) = 0$$

Por la condición (i) de la hipótesis del teorema la variación de  $S$  con respecto a  $\hat{\gamma}$  depende sólo del segundo término de  $S$  en la vecindad de  $\hat{\gamma}$ . Como  $f_1$  y  $f_2$  son diferenciables respecto al primer argumento se concluye que  $\frac{\delta S}{\delta \gamma}$  existe en  $\gamma = \hat{\gamma}$ . Luego de (ii) y (i)  $\lambda = 0$ ,  $\hat{\beta}_1 = (X_1' X_1)^{-1} X_1' Y_1$  y  $\hat{\beta}_2 = (X_2' X_2)^{-1} X_2' Y_2$ ; que son los estimadores MC locales. Nota: la prueba falla si la unión es de tipo 3 pues  $g'(\hat{\gamma}) \neq 0$  y en general  $\lambda \neq 0$  ó si la unión es de tipo 2 pues  $S$  no es una función diferenciable respecto a  $\gamma$  en el punto  $\hat{\gamma}$  excepto cuando  $f_1(x_i, \beta_1^*) = f_2(x_i, \beta_2^*)$ , es decir que la intersección coincide, por casualidad, con el valor de una observación. En este caso  $S$  es diferenciable.

### Teorema 2:

Sea  $y = f(x) = f_1(x, \beta_1) I_{(x, \gamma)}(x) + f_2(x, \beta_2) I_{(\gamma, x_N)}(x)$  con

$f_1(\gamma, \beta_1) = f_2(\gamma, \beta_2)$  donde  $f_i = x' \beta_i$ ,  $i=1,2$ ; y sean  $A = a(x_1, \dots, x_n)$ ,  $\hat{\beta}_1 = b(x_1, \dots, x_n)$  y  $\hat{\beta}_2 = b(x_1, \dots, x_n)$  los estimadores MC para este modelo. Si  $\gamma = a(x_1, \dots, x_n)$  y  $\hat{\beta}_j = b_j(x_1, \dots, x_n)$ ;  $j=1,2$  son tales que:

(i)  $\hat{\gamma} = x_{n_1}$

(ii)  $f_1[x_{n_1}, \beta_1^*(n_1)] = f_2[x_{n_1}, \beta_2^*(n_1)]$

ésto es:  $\gamma^*(n_1) = x_{n_1}$  entonces  $\hat{\beta}_j = \beta_j^*(n_1)$ , con  $j$  que varía entre uno y dos.

Demostración:

De (ii) la solución no restringida  $(\beta_1^*, \beta_2^*, \gamma^*)$  satisface la restricción (c) del modelo y junto a (i) implica que la solución no restringida es también solución restringida.

Corolario (conversión del teorema 1)

Si no existe  $\gamma^*(n_1)$  tal que  $x_{n_1} < \gamma^*(n_1) < x_{n_1+1}$  no es posible que se cumplan las hipótesis (i) y (ii) del teorema 1.

Demostración:

La demostración se hará por reducción al absurdo.

Supongamos que (i)  $x_{n_1} < \hat{\gamma} < x_{n_1+1}$  y (ii)  $g'(\hat{\gamma}) \neq 0$  fuesen verdaderos.

Entonces por el teorema 1:  $\hat{\gamma} = \gamma^*(n_1)$  y  $x_{n_1} < \gamma^*(n_1) < x_{n_1+1}$  lo cual es absurdo pues contradice la hipótesis. Luego (i) y (ii) no pueden ser verdaderos bajo estas condiciones.

Aplicaciones del teorema 1:

Si se conoce el tipo de unión se pueden usar los resultados del teorema 1 como un medio sencillo para encontrar la solución MC total tanto si  $n_1$  es conocido como si no lo es.

a) Si  $n_1$  es conocido se obtienen los estimadores MC del modelo a partir de los estimadores MC locales  $\beta_1^*(n_1)$  y  $\beta_2^*(n_1)$  usando los datos con abscisa  $(x_1, \dots, x_{n_1})$  y  $(x_{n_1+1}, \dots, x_N)$  respectivamente.

Si se resuelve:  $f_1[\gamma, \beta_1^*(n_1)] = f_2[\gamma, \beta_2^*(n_1)]$  se encuentra que existe por lo menos una solución para  $\gamma = \gamma^*(n_1)$  que pertenece al intervalo  $[x_{n_1}, x_{n_1+1}]$  entonces  $[\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}] = [\beta_1^*(n_1), \beta_2^*(n_1), \gamma^*(n_1)]$ .

b) Si  $n_1$  es desconocido el procedimiento sugerido por Hudson en este caso se denomina Paso 1. Consiste en elegir un valor de  $n_1$  y calcular  $\beta_j^*(n_1)$ ,  $j=1,2$ . Se observa si las curvas tienen por lo menos un punto de unión  $\gamma^*(n_1)$  comprendido entre  $x_{n_1}$  y  $x_{n_1+1}$ ; si esto se cumple se calcula:

$$T(n_1) = SCR_{1n_1}^* + SCR_{2n_2}^* \quad (4.34)$$

donde  $SCR_{1n_1}^*$  y  $SCR_{2n_2}^*$  son la SCR locales de cada una de las regresiones. Si las curvas no se unen  $T(n_1) = \infty$ . Se repite este procedimiento para todos los valores relevantes de  $n_1$  y se elige como valor crítico el valor de  $n_1$  para el cual  $T(n_1)$  se minimiza. La hipótesis de que la unión es de tipo 1 junto al resultado del corolario implica que  $T(n_1) < \infty$  para algún  $n_1$ .

#### A.2) Unión de Tipo Dos:

Si se conoce que la unión es de tipo 2 eso significa que  $\hat{\gamma} = x_{n_1}$  para algún  $n_1$  y  $g'(\hat{\gamma})$  es distinto de cero.

a)  $n_1$  conocido:

Es fácil encontrar el resto de los parámetros pues se trata de un modelo que es lineal en ellos. Se deben encontrar los estimadores MC sujetos a la restricción  $f_1(x_{n_1}, \beta_1) = f_2(x_{n_1}, \beta_2)$  lo cual puede expresarse de la siguiente forma:

$$H\hat{\beta} = 0 \text{ con } H = [1 \ x_{n_1} \ x_{n_1}^2 \ \dots \ x_{n_1}^k \ -1 \ -x_{n_1} \ \dots \ -x_{n_1}^k] \quad (4.35)$$

$$\hat{\beta}_{Rest} = \hat{\beta}^* - (X'X)^{-1}H'[H(X'X)H']^{-1}H'\hat{\beta} \quad (4.36)$$

El punto  $X_{n_1}$  puede ser incluido en la primera o en la segunda fase de la regresión pero no en las dos.

b)  $n_1$  desconocido:

El procedimiento sugerido por Hudson en este caso se denomina Paso 2. Consiste en elegir un valor de  $n_1$  y calcular

$$S(n_1) = SCR_{n_1} = SCR_{1n_1} + SCR_{2n_1} \quad (4.37)$$

Se repite el proceso para todos los valores relevantes de  $n_1$  y se elige como valor crítico a aquel para el cual  $S(n_1)$  es mínimo.

### A.3) Unión de Tipo Tres:

Para este tipo de unión se cumple que:  $X_{n_1} < \hat{\gamma} < X_{n_1+1}$  ;  
 $f_1(\gamma, \hat{\beta}_1) = f_2(\hat{\gamma}, \hat{\beta}_2)$  y  $f'_1(\gamma, \hat{\beta}_1) = f'_2(\hat{\gamma}, \hat{\beta}_2)$ .

Estos casos puede requerir un tratamiento individual para cada caso en particular. Afortunadamente sólo se presentan rara vez y algunos autores sostienen que con probabilidad cero. La recomendación del autor es resolverla mediante aproximaciones sucesivas.

### B) TIPO DE UNIÓN DESCONOCIDO

Este es el caso que normalmente se presenta en la realidad. La clasificación de los tipos de unión es exhaustiva, por lo que la solución final está basada en el tipo de estimación que da la menor SCR. El procedimiento sugerido por Hudson es el siguiente: comenzar buscando soluciones con una unión de tipo uno. Para cada valor relevante de  $n_1$  se registra:  $SCR_{n_1}^*$ ,  $SCR_{n_2}^*$ ,  $\beta_1^*(n_1)$ ,  $\beta_2^*(n_1)$ ,  $\gamma_n$  y  $T(n_1)$ . Luego se comienza a buscar soluciones de tipo 2, y se encuentra que algunos valores posibles para  $\hat{\gamma} = x_{n_1}$  pueden ser desechados de antemano pues el Paso 1 da información valiosa en este sentido.

Se enuncian a continuación algunos teoremas útiles en la ubicación del punto de unión pues permitirán reducir considerablemente el número de cálculos.

#### Teorema 3:

a) Si  $X_{n_1} < \gamma^*(n_1) < X_{n_1+1}$  y  $T(n_1) < \infty$  entonces  $S(n_1) \geq T(n_1)$  y  $S(n_1+1) \geq T(n_1)$ .

b) Si  $R^*(n_1) \geq \text{Min}_j \{T(n_j)\} = T(J)$ ;  $R^*(n_1) = SCR_{n_1}^* + SCR_{n_2}^*$  entonces

$$S(n_1) \geq T(J) \quad \text{y} \quad S(n_1+1) \geq T(J).$$

c) Si: (i)  $S(k)$  ha sido determinado para cierto valor  $k$

$$(ii) R^*(n_1) \geq \underset{\text{esos } k}{\text{Min}} \{S(k)\} = S(k)$$

$$\text{entonces } S(n_1) \geq S(k) \quad \text{y} \quad S(n_1+1) \geq S(k)$$

d) Esta parte del teorema se aplica a submodelos lineales solamente. Si  $\bar{X}_1^*(n_1) < \gamma^*(n_1) < \bar{X}_2^*(n_2)$  entonces  $U(\gamma)$  aumenta estrictamente a medida que  $\gamma$  se aleja de  $\gamma^*(n_1)$  en cualquiera de las dos direcciones mientras que  $\gamma$  se mantenga en el intervalo  $[\bar{X}_1^*(n_1), \bar{X}_2^*(n_2)]$ .

$\bar{X}_1^*(n_1)$  y  $\bar{X}_2^*(n_2)$  son respectivamente la media de la variable  $X$  en cada uno de los grupos.

Demostración:

a) la SCR no restringida es menor o igual que la SCR restringida.

Por lo tanto si se restringe  $\hat{\gamma} = X_{n_1}$  ó  $\hat{\gamma} = X_{n_1+1}$  se encuentra que

$$S(n_1) \geq R^*(n_1) \quad \text{y} \quad S(n_1+1) \geq R^*(n_1); \quad \text{como } T(n_1) < \infty \quad \text{entonces} \\ T(n_1) = R^*(n_1).$$

b) De igual forma que en (a)

$$S(l) \geq R^*(l) \geq R(J) \quad ; \quad l = n_1, n_1+1$$

c) De igual forma que en (a) y en (b)

$$S(l) \geq R^*(l) \geq S(k) \quad ; \quad l = n_1, n_1+1$$

d)  $U(\gamma)$  es la SCR mínima calculada cuando  $\beta_1$  y  $\beta_2$  son estimados minimizando la SCR sujeta a la restricción  $f_1(\gamma, \beta_1) = f_2(\gamma, \beta_2)$ .

La prueba completa se basa en argumentos geométricos dados por Mc Lauren.

Corolario de (a), (b) y (c):

En el paso 2 sólo se debe probar que  $\hat{\gamma} = X_{n_1}$  si para  $m = n_1$  y  $m = n_1 - 1$   $T(m) = \infty$ ,  $R(m) < \text{Min}\{T(j)\}$  y  $R(m) < \text{Min}\{S(k)\}$  para  $S(k)$  previamente determinado.

Corolario de (d) (se aplica sólo a modelos lineales)

$$\text{Si } T(n_1) = \infty, \gamma_i^*(n_1) < X \text{ y } \bar{X}_1^*(n_1) < \gamma^*(n_1) < X_n < X_{n_1+1} < \bar{X}_2^*(n_1)$$

No solamente  $S(n_1) \geq R^*(n_1)$  y  $S(n_1+1) \geq R^*(n_1)$  sino que también

$S(n_1+1) > S(n_1)$ , lo cual implica que no hace falta probar  $\hat{\gamma} = X_{n_1+1}$  bajo estas condiciones.

Similarmente  $\bar{X}_1^*(n_1) < X_{n_1} < X_{n_1+1} < \gamma^*(n_1) < \bar{X}_2^*(n_1)$  implica que  $S(n_1) > S(n_1+1)$  y  $\hat{\gamma} = X_{n_1}$  es imposible.

Cuando se buscan uniones de tipo tres es posible eliminar algunos puntos de antemano.

Teorema 4:

- (a) Si (i)  $X_{n_1} \leq \gamma^*(n_1) \leq X_{n_1+1}$   
 (ii)  $g'(\gamma^*(n_1)) \neq 0$

eso implica que no son posibles (iii)  $X_{n_1} < \hat{\gamma} < X_{n_1+1}$  y (iv)  $g'(\hat{\gamma}) = 0$ .

- (b) Si  $R^*(n_1) \geq \text{Min}_j \{T(j), S(j)\}$

no son posibles (iii)  $X_{n_1} < \hat{\gamma} < X_{n_1+1}$  y (iv)  $g'(\hat{\gamma}) = 0$ .

Demostración:

Las restricciones  $f_1(\hat{\gamma}, \beta_1) = f_2(\hat{\gamma}, \beta_2)$  y  $f_1'(\hat{\gamma}, \beta_1) = f_2'(\hat{\gamma}, \beta_2)$  no pueden reducir la SCR obtenida en los pasos anteriores.

En el Paso 3 sólo se deben buscar uniones de tipo 3 con  $X_{n_1} < \hat{\gamma} < X_{n_1+1}$  para algún  $n_1$  si  $T(n_1) = \infty$  y  $R^*(n_1) < \text{Min}\{T(j), S(j)\}$ .

Simplificando lo que interesa no es que se aplique la restricción de igual pendiente sino la posibilidad de modelar el álgebra con una o dos restricciones.

En particular si  $S^{(2)}$  es un mínimo con respecto a  $\gamma$  cuando  $\gamma = \hat{\gamma}$  para enfatizarlo se escribe  $S^{(2)}(\hat{\gamma})$ . ( $S^{(2)}$ : SCR de la regresión sujeta a las dos restricciones.

Sea  $S^{(1)}(\gamma^*)$  la SCR cuando  $\gamma$  se fija igual a un  $\gamma^*$  y la mínima SCR se obtienen imponiendo solamente la restricción  $f_1(\gamma, \beta_1) = f_2(\gamma, \beta_2)$ .

Teorema 5:

- (i) Si una solución es de tipo 3 entonces  $S^{(1)}(\gamma) = S^{(2)}(\gamma)$ .  
 (ii) Si  $\gamma^*$  es cualquier otro valor de  $\gamma$  entonces  $S^{(1)}(\hat{\gamma}) \geq S^{(2)}(\hat{\gamma})$ .

Demostración:

(i) Como  $S^{(2)}(\hat{\gamma})$  es la solución MC local es menor o igual que la SCR por cualquier otro método esto significa que  $S^{(2)}(\hat{\gamma})$  es menor o igual que  $S^{(1)}(\hat{\gamma})$ . Por otro lado si  $S^{(1)}(\hat{\gamma})$  se obtiene minimizando una SCR sujeta sólo a una de las dos restricciones satisfecha por la solución total  $S^{(1)}(\hat{\gamma}) \leq S^{(2)}(\hat{\gamma})$  entonces  $S^{(1)}(\hat{\gamma}) = S^{(2)}(\hat{\gamma})$ .

(ii) Sea  $\gamma^*$  cualquier otro valor de  $\gamma$ . Como  $S^{(2)}(\hat{\gamma})$  es el valor mínimo de la SCR se tiene que  $S^{(1)}(\gamma^*) \geq S^{(2)}(\hat{\gamma})$ . El efecto es que si se encuentra  $S^j(\gamma)$  para valores tentativos de  $\gamma$  y se localiza el mínimo se obtendrá la misma estimación  $\gamma = \hat{\gamma}$  para  $j=1$  ó  $j=2$ . En la práctica se prefiere  $j=2$  pues  $S^{(2)}(\gamma)$  tiene un mínimo más notable y  $\hat{\gamma}$  es más fácilmente determinable.

Hudson extiende la solución propuesta para el caso en que el modelo conste de más de dos fases, pero con la restricción de que el número de fases sea conocido. Si todos los puntos de unión  $\gamma_j$  de la solución MC total son de tipo 1 entonces todos los submodelos ajustados son soluciones MC locales. En general el tipo de unión será desconocido. En este caso deben probarse todas las combinaciones obtenidas imaginando que cada uno de los puntos de unión puede ser de tipo 1, 2 ó 3 y se elige como solución a la que produce la menor suma de cuadrados. Aunque el número de operaciones es finito puede ser excesivamente grande y es casi imposible de alcanzar en la práctica en algunos casos por lo que el autor propone otro método más general de aproximaciones sucesivas. Por ejemplo en el caso de un modelo de tres fases se puede elegir una grilla de puntos en el plano  $(\gamma_1, \gamma_2)$  y usando un método como el Steepest descent aproximar la solución. Sin embargo la superficie puede tener muchos mínimos locales diferenciables donde cada mínimo local corresponde a un caso donde  $(\gamma_1, \gamma_2)$  son ambos de tipo 1. También es posible tener un segmento de mínimos adyacentes diferenciables si  $(\gamma_1, \gamma_2)$  están en el mismo intervalo  $(X_{n_1}, X_{n_1+1})$  y ambos son uniones de tipo uno. La SCR es continua (excepto cuando un  $\gamma_j = X_{n_1} \text{ ó } X_{n_1}$ ) pero la SCR es

en general no diferenciable si  $\gamma_1$  ó  $\gamma_2$  coincide con una observación; en realidad no es diferenciable parcialmente respecto a  $\gamma_1$  ó  $\gamma_2$  según  $\gamma_1 = X_i$  ó  $\gamma_2 = X_i$  y es parcialmente no diferenciable en ambas direcciones si  $(\gamma_1, \gamma_2) = (X_i, X_j)$ , con  $X_i \leq X_j$ . Mínimos locales no diferenciables son posibles si ambos puntos de unión son de tipo 2.

De todo ésto es claro que un programa de computadora tendrá dificultades en la ubicación del mínimo por lo que se prefiere el algoritmo finito desarrollado en este trabajo.

#### 4.2.3. Solución propuesta por Hinkley (1969).

Hinkley estudia el problema del modelo de regresión de dos fases donde se desea estimar y hacer inferencia sobre el punto de cambio de régimen ( $\gamma$ ); bajo el supuesto de que ambas fases se intersecan en  $\gamma$  y se desconoce el intervalo en el cual se produce el cambio de régimen. En un trabajo publicado en 1969 encuentra la distribución asintótica del estimador MV de la intersección que es una buena aproximación para muestras finitas. (Es el primer trabajo del que tengamos referencia donde se logra ésto).

Como ambas fases se intersecan en  $\gamma$  es  $\beta_{01}\gamma + \beta_{11} = \beta_{02}\gamma + \beta_{12}$ ; por lo tanto  $\gamma = (\beta_{11} - \beta_{12}) / (\beta_{02} - \beta_{01})$  y está comprendido entre  $x_\tau$  y  $x_{\tau+1}$ , es decir que el modelo es suave, ocurre un cambio en la pendiente pero no hay discontinuidad matemática en la función de regresión. El parámetro que debe estimarse es  $\theta' = (\beta_{01}, \beta_{11}, \beta_{02}, \beta_{12}, \tau, \sigma^2)$  el cual deriva una distribución asintótica para  $\gamma$  que da un buen ajuste a la distribución muestral para tamaños de muestras moderados.

La función de verosimilitud de la muestra es:

$$L(Y; \theta) = \left[ \frac{1}{2\pi\sigma^2} \right]^{(1/2)N} \exp \left[ -\frac{1}{2\sigma^2} (Y_\tau - X_\tau \beta_\tau)' (Y_\tau - X_\tau \beta_\tau) \right] \quad (4.38)$$

donde  $\tau$  varía entre 2 y  $N-2$ ,  $(\beta_{12} - \beta_{11}) + \gamma(\beta_{02} - \beta_{01}) = 0$  y  $\gamma$  está

comprendido entre  $x_t$  y  $x_{t+1}$ .

Sea  $S_0^2$  la SCR si la regresión siguiese un régimen de una fase. La función  $Z_t^2(r) = S_0^2 - S_t^2(r)$  es equivalente a  $L_t(r)$ .  $\hat{\gamma}$  es el valor de  $r$  que maximiza  $Z_t^2(r)$  para todos los valores de  $r$  que pertenecen a  $[x_t, x_{t+1}]$ , con  $t$  que varía entre 2 y  $N-2$ .

$$Z_t^2(r) = \frac{(\tilde{\beta}_{02(t)} - \tilde{\beta}_{01(t)})^2 [C_t - D_t(\tilde{\gamma}_t + r) + E_t \tilde{\gamma}_t r]^2}{C_{xx,N} (C_N - 2D_t r + E_t r^2)} \quad ; t=2, \dots, N-2 \quad (4.42)$$

donde  $C_t, D_t, E_t, t^*, \bar{x}_t, \bar{x}_t^*, C_{xx,t}, C_{xx,t}^*$  y  $C_{xx,N}^*$  están dados por las expresiones (4.5) a (4.13).

La aparición de  $\tilde{\beta}_{02} - \tilde{\beta}_{01}$  en el numerador de  $Z_t^2(r)$  sugiere que la maximización de  $L_t(r)$  está asociada con maximizar la diferencia  $\tilde{\beta}_{02(t)} - \tilde{\beta}_{01(t)}$ ; luego el estimador MV de  $\tilde{\beta}_{02} - \tilde{\beta}_{01}$  será sesgado.

Como  $L_t(r)$  corresponde a la verosimilitud de  $L(r)$  sólo para los valores de  $r$  que pertenecen al intervalo  $[x_t, x_{t+1})$ ,  $\hat{\gamma}$  puede coincidir con  $\tilde{\gamma}_t$  para algún  $t$ , en cuyo caso  $\tilde{\gamma}_t$  es mayor o igual que  $x_t$  y menor que  $x_{t+1}$  ó puede coincidir con algún valor de  $x_t$ , para algún  $t$ . Por lo tanto para determinar  $\hat{\gamma}$  se deben comparar los valores de  $L_t(\tilde{\gamma}_t)$  para todo  $t$  tal que  $\tilde{\gamma}_t$  pertenezca a  $[x_t, x_{t+1})$  y de  $L_t(x_t)$  y  $L_t(x_{t+1})$  para todo  $t$  en que  $\tilde{\gamma}_t$  no pertenezca a dicho intervalo. Se logra reducir el número de cálculos usando los siguientes resultados:

1)  $L(r)$  es continua ya que se verifican las siguientes igualdades para todos los valores de  $t$ . (No hemos considerado necesario desarrollar aquí la demostración)

$$\begin{aligned} \text{a) } (\tilde{\beta}_{02(t)} - \tilde{\beta}_{01(t)})^2 [C_t - D_t(\tilde{\gamma}_t + x_t) + E_t \tilde{\gamma}_t x_t] &= \\ &= (\tilde{\beta}_{02(t-1)} - \tilde{\beta}_{01(t-1)})^2 [C_{t-1} - D_{t-1}(\tilde{\gamma}_{t-1} + x_t) + E_{t-1} \tilde{\gamma}_{t-1} x_t] \end{aligned} \quad (4.43)$$

$$\text{b) } Q_t = C_t - 2D_t x_t + E_t x_t^2 = C_{t-1} - 2D_{t-1} x_t + E_{t-1} x_t^2 = Q_{t-1} \quad (4.44)$$

Luego  $Z_{t-1}^2(x_t) = Z_t^2(x_t)$  que junto a la continuidad de  $Z_t^2(r)$  da el resultado deseado. Las derivadas de  $L(r)$  no existen en  $r=x_t$ .

2) Bajo ciertas condiciones en  $\tilde{\gamma}_t$ ,  $L_t(r)$  es convexa para los valores de  $r$  que pertenecen al intervalo  $[x_t, x_{t+1}]$ . (Las condiciones son que  $D_t - E_t x_{t+1} \leq 0$  y  $D_t - E_t x_t \leq 0$ ).

Sea

$$\delta_t = \frac{C_t - D_t \gamma_t}{D_t - E_t \gamma_t}, \quad t=2, \dots, N-2 \quad (4.45)$$

el valor de  $r$  para el cual  $Z_t^2(r)$  alcanza su mínimo valor, cero. Si se considera a  $\delta_t$  como una función de la variable  $\tilde{\gamma}_t$  es fácil mostrar que si

$$\frac{C_t - D_t x_{t+1}}{D_t - E_t x_{t+1}} \leq \tilde{\gamma}_t \leq \frac{C_t - D_t x_t}{D_t - E_t x_t} \implies \delta_t < x_t \text{ ó } \delta_t > x_{t+1}$$

Esto es,  $L_t(r)$  disminuye a mediada que  $|r - \tilde{\gamma}_t|$  aumenta para  $x_t \leq r \leq x_{t+1}$ . Estos dos resultados permiten dar una versión levemente mejorada del procedimiento de Hudson (1966) para estimar  $\hat{\gamma}$  pues busca  $L(r)$  con un número mínimo de cálculos. Se prefiere usar  $Z_t^2(r)$  en vez de  $S_t^2(r)$  no sólo por simplicidad sino porque el máximo total  $Z_t^2(\hat{\gamma})$  es el estadístico de razón de verosimilitud para testar igualdad de pendientes en ambas fases de la regresión (Sección 4.1.4). Por otro lado es obvio que hay más información en cada  $\tilde{\gamma}_t$  de la que se ha usado hasta ahora en la búsqueda de estimador MV  $\hat{\gamma}$ .

Los sucesivos valores  $\tilde{\gamma}_t, \tilde{\gamma}_{t+1}, \dots$  son dependientes unos de otros de manera tal que si  $\hat{\gamma} = \gamma_t$  eso implica que  $\tilde{\gamma}_{t+1}$  estará cercano a  $\hat{\gamma}$ , y estará más cerca mientras más grandes sean  $t$  y  $N-t$ . Para ilustrar esto Hinkley genera una muestra de cincuenta observaciones usando desvíos pseudonormales con valores de los parámetros  $\beta_{01} = 0,0$ ,  $\beta_{02} = 0,4$ ,  $\gamma = 25,5$ ,  $\sigma = 1$  y  $x_t = t$ ;  $t=1,2,\dots,50$ . La verosimilitud  $L(\gamma)$  aunque no es convexa es muy suave y  $\hat{\gamma} = 27,95$ .

En el gráfico de  $x_t$  versus  $\tilde{\gamma}_t$  se observa un comportamiento interesante. Aparentemente siempre que  $\hat{\gamma}$  es mayor que  $x_{t+1}$ ,  $\tilde{\gamma}_{t+1}$  es mayor que  $x_{t+1}$ ; y, similarmente cuando  $\hat{\gamma}$  es menor que  $x_t$ ,  $\tilde{\gamma}_t$  es menor que  $x_t$ . El examen de otros ejemplos muestra que este comportamiento es más consistente a medida que  $N$  aumenta. Esto está relacionado con el hecho de que las fluctuaciones de  $\tilde{\gamma}_t$  alrededor de  $\hat{\gamma}$  son del orden  $O(N^{-1/2})$  cuando  $t$  no está cerca a 1 ó  $N$ . Esta conexión entre  $\hat{\gamma}$  y los  $\tilde{\gamma}_t$  puede ser usado para construir un procedimiento que converja en  $\hat{\gamma}$  comenzando con valores tentativos de  $t$  ya que  $\tilde{\gamma}_t$  es una buena aproximación de  $\hat{\gamma}$ . Un examen más sofisticado de la conexión entre  $\hat{\gamma}$  y los  $\tilde{\gamma}$  lleva a la distribución asintótica de  $\hat{\gamma}$ . Hinkley obtiene dicha distribución.

Feder y Sylvester (1968) demostraron que  $\hat{\gamma}$  se distribuye asintóticamente normal; pero un estudio empírico indica que ésta es una aproximación inadecuada para muestras pequeñas. Sin embargo la normalidad asintótica de  $\hat{\gamma}$  indica un grado de suavidad asintótica poseído por  $L(r)$  útil en la búsqueda de un resultado asintótico alternativo para la distribución de  $\hat{\gamma}$ .

Es difícil trabajar con  $\hat{\gamma}$  pues no se tiene una definición explícita del mismo; pero su normalidad asintótica indica que  $\hat{\gamma} = x_t$  con probabilidad cero, por lo menos asintóticamente. Se ha mostrado que la siguiente aseveración se cumple con una alta probabilidad:

Teorema:

$[\tilde{\gamma}_t > x_t \Rightarrow \tilde{\gamma}_{t-1} > x_t]$  si y sólo si  $[\tilde{\gamma}_t > x_t \Rightarrow \hat{\gamma} > x_t]$  con  $t < t'$ .

A) H)  $\tilde{\gamma}_t > x_t \Rightarrow \tilde{\gamma}_{t-1} > x_t$

T)  $\tilde{\gamma}_t > x_t \Rightarrow \hat{\gamma} > x_t$

D) Supongamos que  $\tilde{\gamma}_t > x_t$  entonces  $\tilde{\gamma}_{t-1} > x_t$

$$\tilde{\gamma}_t > x_t \Rightarrow \tilde{\gamma}_{t-1} > x_t$$

$$\tilde{\gamma}_{t-1} > x_{t-1} \Rightarrow \tilde{\gamma}_{t-2} > x_{t-1}$$

$\vdots$   
 $\tilde{\gamma}_3 > x_3 \Rightarrow \tilde{\gamma}_2 > x_3$   
 Esto implica que  $\hat{\gamma} \neq \tilde{\gamma}_1$  y  $\hat{\gamma} > x_1$ .

- B) H)  $\tilde{\gamma}_t > x_t \Rightarrow \hat{\gamma} > x_t$   
 T)  $\tilde{\gamma}_t > x_t \Rightarrow \tilde{\gamma}_{t-1} > x_t$   
 D)  $\tilde{\gamma}_t > x_t \Rightarrow \hat{\gamma} > x_t$   
 $\tilde{\gamma}_{t-1} > x_{t-1} \Rightarrow \hat{\gamma} > x_{t-1}$   
 $\vdots$   
 $\tilde{\gamma}_2 > x_2 \Rightarrow \hat{\gamma} > x_2$

Si  $x_{t-1} < \tilde{\gamma}_{t-1} < x_t$  esto contradice la hipótesis pues  $\hat{\gamma}$  puede ser igual a  $\tilde{\gamma}_{t-1}$  para algún  $t < t'$ , luego  $\tilde{\gamma}_{t-1} > x_t$ . Sin pérdida de generalidad se supone que  $\beta_{02} > \beta_{01}$  y que los  $x_t$  están igualmente espaciados. Usando las propiedades (4.43) y (4.44) de continuidad de  $L(r)$  se tiene:

$$\begin{aligned}
 (\tilde{\beta}_{02(t-1)} - \tilde{\beta}_{01(t-1)}) \{Q_{t-1} + (D_{t-1} - E_{t-1} x_t)(x_t - \tilde{\gamma}_{t-1})\} = \\
 = (\beta_{02(t)} - \beta_{01(t)}) \{Q_{t-1} + (D_t E_t x_t)(x_t - \tilde{\gamma}_t)\}, t=3, 4, \dots, N-2 \quad (4.46)
 \end{aligned}$$

Para  $x_t$  igualmente espaciados  $D_{t-1} - E_{t-1} x_t$  y  $D_t - E_t x_t$  son ambos menores que cero si  $t$  es menor que  $\frac{1}{2}(N-1)$  y ambos son mayores que cero si  $t$  es mayor que  $\frac{1}{2}(N+3)$  y  $Q_t$  es mayor que cero para todo  $t$ .

Por ejemplo si  $t$  es menor que  $\frac{1}{2}(N-1)$  para todo  $\epsilon$  mayor que cero la expresión (4.46) será:  $\{P(\tilde{\gamma}_{t-1} > x_t / \tilde{\gamma}_t = x_t + \epsilon) \cong P(\tilde{\gamma}_{t-1} > x_t / \tilde{\gamma}_t = x_t + \epsilon)$

De (4.46):

$$\begin{aligned}
 x_t - \tilde{\gamma}_{t-1} = \frac{\Delta \tilde{\beta}_{0t}}{\Delta \tilde{\beta}_{0, t-1}} \frac{\{Q_{t-1} + (D_t - E_t x_t)(x_t - \tilde{\gamma}_t)\}}{D_{t-1} - E_{t-1} x_t} - \frac{Q_{t-1}}{D_{t-1} - E_{t-1} x_t} < 0 \\
 \Rightarrow x_t - \tilde{\gamma}_{t-1} \text{ menor que cero dado que } x_t - \tilde{\gamma}_t = -\epsilon \Leftrightarrow
 \end{aligned}$$

$$\begin{aligned} & \Leftrightarrow \frac{\Delta \tilde{\beta}_{0t}}{\Delta \tilde{\beta}_{0,t-1}} \cdot \frac{\{Q_{t-1} + (E_t x_t - D_t) \varepsilon\}}{D_{t-1} - E_{t-1} x_t} < \frac{Q_{t-1}}{D_{t-1} - E_{t-1} x_t} \Leftrightarrow \\ & \Leftrightarrow \frac{\Delta \tilde{\beta}_{0t}}{\Delta \tilde{\beta}_{0,t-1}} \{a_{t-1} + (E_t x_t - D_t) \varepsilon\} > a_{t-1} \Leftrightarrow \\ & \Leftrightarrow \frac{\Delta \tilde{\beta}_{0t-1}}{\Delta \tilde{\beta}_{0t}} < \frac{Q_{t-1} + (E_t x_t - D_t) \varepsilon}{Q_{t-1}} = 1 + \frac{(E_t x_t - D_t) \varepsilon}{Q_{t-1}} \\ P(\tilde{\gamma}_{t-1} > x_t / \tilde{\gamma}_t = x_t + \varepsilon) &= P \left\{ \frac{\Delta \tilde{\beta}_{0t-1}}{\Delta \tilde{\beta}_{0t}} < 1 + \frac{E_t x_t - D_t}{Q_{t-1}} \varepsilon \right\} \cong \\ & \cong P \left\{ \frac{\Delta \tilde{\beta}_{0t-1}}{\Delta \tilde{\beta}_{0t}} < 1 + \frac{3[(N-t)-t]}{2t(N-t)} \varepsilon \right\} \\ & \cong P \left\{ \Delta \tilde{\beta}_{0t-1} - \Delta \tilde{\beta}_{0t} \left\{ 1 + \frac{3[(N-t)-t]}{2t(N-t)} \varepsilon \right\} < 0 \right\} \end{aligned}$$

La media y la varianza de  $\tilde{\beta}_{t-1} - \tilde{\beta}_t \left\{ 1 + \frac{3[(N-t)-t]}{2(N-t)t} \varepsilon \right\}$  son del orden  $O(N^{-1})$  y  $O(N^{-5})$  respectivamente, con lo que  $P(\tilde{\gamma}_{t-1} > x_t / \tilde{\gamma}_t = x_t + \varepsilon) \cong \Phi[-O(T^{3/2})]$  para  $t < \frac{1}{2}(N-1)$  y para todo  $\varepsilon > 0$ .

Como la distribución de  $\tilde{\gamma}_t$  es continua y acotada en  $x_t$  entonces:

$$P(\tilde{\gamma}_{t-1} > x_t / \tilde{\gamma}_t > x_t) \cong \Phi[-O(N^{3/2})], \quad t < \frac{1}{2}(N-1) \quad (4.47)$$

Para aplicar esta expresión a la distribución asintótica de  $\hat{\gamma}$ , debe notarse que por definición:

$$\begin{aligned} P(\hat{\gamma} > x_t) &\geq P(\tilde{\gamma}_t > x_t, \tilde{\gamma}_{t-1} > x_t, \dots, \tilde{\gamma}_2 > x_t) = \\ &= P(\tilde{\gamma}_t > x_t) P(\tilde{\gamma}_{t-1} > x_t / \tilde{\gamma}_t > x_t) \dots P(\tilde{\gamma}_2 > x_t / \tilde{\gamma}_2 > x_t \dots \tilde{\gamma}_t > x_t) \end{aligned} \quad (4.48)$$

Esta desigualdad es estrictamente necesaria debido a la posibilidad de que  $\hat{\gamma} > x_t$  cuando  $x_s \leq \tilde{\gamma}_s < x_{s+1}$  para algún  $s < t$ . Sin embargo la suavidad asintótica de  $L(\gamma)$  indica que se puede ignorar esta posibilidad cuando se deriva la distribución

$$F_1(x) = \frac{x - x_t}{x_{l+1} - x_t} F_1(x_{l+1}) + \frac{x_{l+1} - x}{x_{l+1} - x_t} F_1(x_t), \quad x_t \leq x \leq x_{l+1} \quad (4.53)$$

(4.53) se reduce a (4.52) cuando  $x=x_t$  ó  $x=x_{l+1}$ . El cálculo de las probabilidades en (4.52) se simplifica usando una aproximación a la distribución de  $\tilde{\gamma}_l$ .

$$P(\tilde{\gamma}_l < x) \cong \Phi[-\beta(\gamma + \theta_l - (1 - \theta_l))/\Sigma_l(x)] \quad (4.54)$$

donde: 
$$\beta = \frac{\beta_{02} - \beta_{01}}{\sigma}$$

$$\theta_l = \begin{cases} \frac{1}{N-t} \left\{ \sum_{i=l+1}^{\tau} (x_i - \gamma) - (N-t) \bar{x}_l^* \theta_l \right\} & \text{si } t < \tau \\ = 0 & \text{si } t = \tau \\ \frac{1}{t} \left\{ \sum_{i=\tau+1}^{\tau} (x_i - \gamma) - t \bar{x}_l \theta_l \right\} & \text{si } t > \tau \end{cases}$$

$$\theta_l = \begin{cases} \sum_{i=l+1}^{\tau} (x_i - \bar{x}_l^*) (x_i - \gamma) / C_{xx,t}^* & \text{si } t < \tau \\ = 0 & \text{si } t = \tau \\ \sum_{i=\tau+1}^{\tau} (x_i - \bar{x}_l) (x_i - \gamma) / C_{xx,t} & \text{si } t > \tau \end{cases}$$

$$\Sigma_l^2(x) = \frac{1}{t} + \frac{1}{N-t} + \frac{(\bar{x}_l - x)^2}{C_{xx,t}} + \frac{(\bar{x}_l^* - x)^2}{C_{xx,t}^*}$$

Una alternativa de  $F_1$  puede ser derivada del hecho de que la suavidad asintótica de  $L(\gamma)$  implica la equivalencia asintótica de  $\tilde{\gamma}_\tau, \tilde{\gamma}_{\tau+1}, \dots$

Una forma de ver esto es notar que la diferencia entre las derivadas a izquierda y a derecha de la función de verosimilitud conjunta de la muestra con respecto a  $\gamma$  en  $\gamma=x_t$  es del orden  $O(N^{-1})$  relativo a la magnitud de la derivada sin dar argumento más formal se deduce:  $P(\hat{\gamma} < x) \cong F_2(x) = P(\tilde{\gamma}_\tau < x)$   
 Esto incidentalmente lleva a la normalidad asintótica de  $\hat{\gamma}$

$$P(\hat{\gamma} < x) = \Phi\left\{\frac{\beta(x-\gamma)}{\sqrt{(\tau(N-\tau)/4N)}}\right\} \quad (4.55)$$

Debe enfatizarse el hecho de que la dicotomía de  $F_1(x_i)$  depende del hecho de que los  $x_i$  en este caso están igualmente espaciados. Para cada configuración de los  $x_i$  se debe hallar  $F_1(x)$  y  $F_2(x)$ . Un análisis completo de los datos incluye también el cálculo de  $\hat{\beta} = (\hat{\beta}_{02} - \hat{\beta}_{01})/\sigma$ . El estimador MV de  $\beta$  es  $(\hat{\beta}_{02} - \hat{\beta}_{01})/\sigma$  donde  $\hat{\beta}_{01}$  y  $\hat{\beta}_{02}$  son estimadores MC cuando las dos rectas se restringe a intersectarse en  $x = \hat{\gamma}$ .  $\hat{\beta}$  se distribuye asintóticamente  $N(\beta, V^2)$ , con  $V^2 = \frac{1}{C_{xx,\tau}} + \frac{1}{C_{xx,\tau}^*}$ . Se espera un sesgo positivo en  $\hat{\beta}$  para muestras finitas pues la maximización de  $L(\gamma)$  está asociado con maximizar la diferencia entre las pendientes de ambas regresiones.

Para estudiar la distribución de  $\hat{\gamma}$  empíricamente con la idea de comparar las aproximaciones  $F_1(\cdot)$  y  $F_2(\cdot)$  Hinkley generó observaciones variando  $\beta, \gamma$  y  $N$ , tomando en cada caso 500 muestras. El examen de los resultados mostró que para ciertas combinaciones de  $\beta, \gamma$  y  $N$  el problema de estimación estaba mal definido, esto es la  $\text{Var}(\hat{\gamma})$  es grande y hay un sesgo significativo en  $\hat{\gamma}$  y  $\hat{\beta}$  y además la  $\text{Var}(\hat{\beta})$  resultó ser mucho mayor que su valor asintótico. Basándose en esta y otras simulaciones se puede considerar que los casos mal definidos son aquellos para los cuales  $\beta_\tau(N-\tau)/N$  es menor que cinco aproximadamente. En los casos bien definidos la comparación entre  $F_1(\cdot)$  y la distribución empírica de  $\hat{\gamma}$  indica que la aproximación es buena.  $F_2(\cdot)$  no es una aproximación tan buena, aunque mejora considerablemente cuando  $N$  aumenta. Gráficos normales de la distribución empírica de  $\hat{\beta}$  revelan un ajuste excelente a la distribución normal asintótica en los casos bien definidos, pero con un pequeño sesgo positivo. En la práctica la magnitud del sesgo es despreciable en comparación con el error estándar. En ninguna de las comparaciones la posición de  $\gamma$  parece tener demasiado efecto, salvo en determinar si un caso está mal definido o no. La distribución empírica de  $\hat{\gamma}$  para  $25 \leq N \leq 100$  es uniforme en el rango  $2 \leq \gamma \leq N-1$ , con una leve tendencia a aumentar en el punto final. Esta distribución asintótica alternativamente

de  $\hat{\gamma}$  hallada por Hinkley, aunque es una buena aproximación para tamaños de muestras moderados, es complicada de usar, por lo que él continúa trabajando y encuentra la distribución del estimador MV de  $\gamma$  y de su intervalo de confianza.

Considera por separado el caso de que  $\beta_{02}$  sea conocido o desconocido.

1)  $\beta_{02}$  desconocido:

Los estimadores MV de  $\beta_{01}$  y  $\beta_{02}$  van a satisfacer la ecuación matricial:

$$A_{\tau}(\hat{\gamma}, \hat{\gamma}) \begin{bmatrix} \hat{\beta}_{01} \\ \hat{\beta}_{02} \end{bmatrix} = A_{\tau}(\tilde{\gamma}_{\tau}, \hat{\gamma}) \begin{bmatrix} \tilde{\beta}_{01, \tau} \\ \tilde{\beta}_{02, \tau} \end{bmatrix} \quad (4.56)$$

donde:

$$A_{\tau}(r, z) = \begin{bmatrix} NC_{xx, \tau} + tt^*(\bar{x}_{\tau} - r)(\bar{x}_{\tau} - z) & -tt^*(\bar{x}_{\tau}^* - r)(\bar{x}_{\tau} - z) \\ -tt^*(\bar{x}_{\tau} - r)(\bar{x}_{\tau}^* - z) & NC_{xx, \tau}^* + tt^*(\bar{x}_{\tau}^* - r)(\bar{x}_{\tau}^* - z) \end{bmatrix}$$

y  $\hat{\gamma}$  está comprendido entre  $x_{\tau}^*$  y  $x_{\tau+1}^*$  por definición.

Si  $\hat{\gamma} = \tilde{\gamma}_S$  entonces  $(\hat{\beta}_{01}, \hat{\beta}_{02}) = (\beta_{01}, \beta_{02})$

$$A_{\tau}(\hat{\gamma}, \hat{\gamma}) = \begin{bmatrix} NC_{xx, \tau} + tt^*(\bar{x}_{\tau} - \hat{\gamma})(\bar{x}_{\tau} - \hat{\gamma}) & -tt^*(\bar{x}_{\tau} - \hat{\gamma})(\bar{x}_{\tau} - \hat{\gamma}) \\ -tt^*(\bar{x}_{\tau} - \hat{\gamma})(\bar{x}_{\tau}^* - \hat{\gamma}) & NC_{xx, \tau}^* + tt^*(\bar{x}_{\tau}^* - \hat{\gamma})(\bar{x}_{\tau}^* - \hat{\gamma}) \end{bmatrix}$$

$$A_{\tau}(\tilde{\gamma}_{\tau}, \hat{\gamma}) = \begin{bmatrix} NC_{xx, \tau} + tt^*(\bar{x}_{\tau} - \tilde{\gamma}_{\tau})(\bar{x}_{\tau} - \hat{\gamma}) & -tt^*(\bar{x}_{\tau}^* - \tilde{\gamma}_{\tau})(\bar{x}_{\tau} - \hat{\gamma}) \\ -tt^*(\bar{x}_{\tau} - \tilde{\gamma}_{\tau})(\bar{x}_{\tau}^* - \hat{\gamma}) & NC_{xx, \tau}^* + tt^*(\bar{x}_{\tau}^* - \tilde{\gamma}_{\tau})(\bar{x}_{\tau}^* - \hat{\gamma}) \end{bmatrix}$$

$$\text{Si } \hat{\gamma} = \tilde{\gamma}_S \text{ entonces } \begin{bmatrix} \hat{\beta}_{01} \\ \hat{\beta}_{02} \end{bmatrix} = \begin{bmatrix} \tilde{\beta}_{01S} \\ \tilde{\beta}_{02S} \end{bmatrix}$$

Sea S la SCR alrededor de la regresión ajustada entonces

$S = \sum_{i=1}^N (y_i - \bar{y})^2 - \beta_{00}^2 C_{xx, N} - Z(\hat{\gamma})$  tiene  $(N-4)$  g.l. (debe estimarse  $\beta_{01}, \beta_{02}, \sigma^2$  y  $\gamma$ ).  $\beta_{00}$  es el estimador MC de  $\beta_0$  si la regresión

siguiere un régimen de una fase. La distribución asintótica normal de los estimadores MV es fácilmente derivable (Sylvester). La consistencia implica que  $[E(Y/X=\hat{\gamma}), \hat{\beta}_{01}, \hat{\beta}_{02}, \hat{\gamma}]$  se distribuye asintóticamente como  $(\hat{\theta}_\tau, \hat{\beta}_{01\tau}, \hat{\beta}_{02\tau}, \hat{\gamma}_\tau)$ , excepto en casos patológicos. El cómputo de la matriz de información de Fisher condicionada en  $\tau$  da la siguiente expresión para las varianzas asintóticas.

$$\text{Var}(\hat{\theta}) = \sigma^2 \left\{ \frac{\tau \beta_{01}^2 + \tau^* \beta_{02}^2}{\tau \tau^* (\beta_{01} - \beta_{02})^2} + \frac{\beta_{01}^2 (\bar{x}_\tau - \gamma)^2}{C_{xx, \tau}} + \frac{\beta_{02}^2 (\bar{x}_\tau^* - \gamma)^2}{C_{xx, \tau}^*} \right\} \quad (4.57)$$

$$\text{Var}(\hat{\beta}_{01}) = \frac{\sigma^2}{C_{xx, \tau}} \quad (4.58)$$

$$\text{Var}(\hat{\beta}_{02}) = \frac{\sigma^2}{C_{xx, \tau}^*} \quad (4.59)$$

$$\text{Var}(\hat{\gamma}) = \sigma^2 \left\{ \frac{1}{\tau} + \frac{1}{\tau^*} + \frac{(\bar{x}_\tau - \gamma)^2}{C_{xx, \tau}} + \frac{(\bar{x}_\tau^* - \gamma)^2}{C_{xx, \tau}^*} \right\} / (\beta_{01} - \beta_{02})^2 \quad (4.60)$$

De la normalidad asintótica se deduce que la SCR  $S$  se distribuye asintóticamente como  $\sigma^2 \chi_{N-4}^2$ .

2)  $\beta_{02}$  conocido e igual a cero:

Sea  $S_t^2(r)$  la SCR del modelo sujeto a la restricción  $\beta_{02} = 0$ ,  $\tau = t$  y  $\gamma = r$ .

$$S_t^2(r) = \frac{\sum_{i=1}^N (y_i - \bar{y}_N)^2 - \tilde{\beta}_{01t}}{C_t' - 2D_t' r + \varepsilon_t' r^2} \frac{\{C_t' - D_t' (\tilde{\gamma}_t' + r) + \varepsilon_t' \tilde{\gamma}_t' r\}^2}{C_t' - 2D_t' r + \varepsilon_t' r^2} \quad (4.61)$$

$\tilde{\gamma}_t'$  es el estimador MC no restringido de  $\gamma$  condicionado a que  $\tau = t$  y

$$\beta_{02} = 0. \quad C_t' = C_{xx, t} + t t^* \frac{\bar{x}_t^2}{N}, \quad D_t' = t t^* \frac{\bar{x}_t}{N} \quad \text{y} \quad \varepsilon_t' = \frac{t t^*}{N}.$$

El estimador MV  $\hat{\gamma}$  maximiza la función continua  $Z'(r)$  definida como el término final de (4.61) para los valores de  $r$  que están comprendidos entre  $x_t$  y  $x_{t+1}$ ; con  $t$  que varía entre 2 y  $N-2$ . El procedimiento para calcular  $\hat{\gamma}$  es el mismo que se desarrolló cuando  $\beta_{02}$  es desconocido, pero ahora usando la función  $Z'(r)$  en vez de  $Z(r)$ .

$S^2(r) = \sum_{i=1}^N (y_i - \bar{Y}_N)^2 - Z'(\hat{\gamma})$  tiene  $N-3$  g.l. y se distribuye como  $\sigma^2 \chi_{N-3}^2$ .

Las varianzas de la distribución normal asintótica de  $(\hat{\theta}, \hat{\beta}_{01}, \hat{\gamma})$  son:

$$\text{Var}(\hat{\theta}) = \frac{\sigma^2}{\tau^*} \quad (4.62)$$

$$\text{Var}(\hat{\beta}_{01}) = \frac{\sigma^2}{C_{xx, \tau}} \quad (4.63)$$

$$\text{y } \text{Var}(\hat{\gamma}) = \sigma^2 \left\{ \frac{1}{\tau} + \frac{1}{\tau^*} + \frac{(\bar{x}_t - \gamma)^2}{C_{xx, \tau}} \right\} / \beta_{01}^2. \quad (4.64)$$

Existen dos problemas para usar la distribución normal asintótica de los estimadores directamente, en primer lugar las varianzas dependen de parámetros desconocidos, particularmente la varianza de  $\hat{\gamma}$  depende fuertemente de  $\beta_{02} - \beta_{01}$  y además la distribución normal asintótica puede ser una pobre aproximación para muestras pequeñas, en especial en el caso de  $\hat{\gamma}$ . Esto llevó a Hinkley a construir intervalos de confianza simultáneos para  $\gamma$  y  $E(Y_i / X_i = \gamma)$ . Para tamaños de muestras moderados las distribuciones de  $\hat{\beta}_{01}, \hat{\beta}_{02}$  y  $\hat{\gamma}$  son muy bien aproximadas por sus respectivas distribuciones normales asintóticas.

Como la la distribución  $\chi^2$  es aparentemente una aproximación razonable de la distribución del estadístico de razón de verosimilitud para muestras pequeñas, él considera la construcción de IC para  $\gamma$  usando los estadísticos de razón de verosimilitud apropiados. Nuevamente el argumento usado por el autor es empírico. Como ya hizo en la otras situaciones trata por separado el caso de  $\beta_{02}$  sea conocido o desconocido.

#### 1) $\beta_{02}$ desconocido:

Por la construcción de  $Z(r)$  (4.42)[Hinkley,1969] el test de

razón de verosimilitud para  $H_0^*: \gamma = \gamma_0$  con tamaño asintótico  $\delta$  es aceptar  $H_0^*$  si:

$$\frac{Z(\hat{\gamma}) - Z(\gamma_0)}{S} (N-4) \leq F_{1, N-4} (1-\delta) \quad (4.65)$$

$\delta$  es una aproximación del tamaño del test para muestras pequeñas. No puede determinarse el tamaño exacto del test pues las distribuciones no son exactas. La región de  $1-\delta$  de confianza para  $\gamma$  es el conjunto de  $\gamma_0$  que satisface (4.65). Como  $Z(r)$  no es necesariamente una función monótona decreciente de  $Z(\hat{\gamma})$ , la región no es necesariamente un único intervalo. En la práctica parece razonable tomar como una aproximación de la región de confianza el intervalo más pequeño que contenga las soluciones de la misma, a menos que la función de verosimilitud tenga dos picos distintos. Al hacer esto se hace el test conservativo. No se conoce la probabilidad del intervalo, sino sólo una cota inferior  $1-\delta$ . El error de ésta aproximación depende de  $\left| \frac{(\beta_{02} - \beta_{01})}{\sigma} \right|$  y tiende a cero cuando  $\tau$  y  $N-\tau$  aumentan. De manera que ahora el intervalo de confianza aproximado para  $\gamma$  es  $(\gamma_1, \gamma_2)$  con  $\gamma_1$  y  $\gamma_2$  la raíz más pequeña y más grande respectivamente de:

$$Z(r) = Z(\hat{\gamma}) - S F_{1, N-4} (1-\delta) / (N-4) \quad (4.66)$$

Por la definición de  $Z(r)$  para cada  $t$  la ecuación (4.66) es cuadrática para los  $r$  comprendidos entre  $x_t$  y  $x_{t+1}$ . Sustitución directa en  $Z(r)$  da

$$L_t r^2 - Z M_t r + N_t = 0 \quad (4.67)$$

donde  $L_t = \{Z(\hat{\gamma}) - SF_{1, N-4} (1-\delta) / (N-4)\} E_t - D_t - E_t \tilde{\gamma}_t)^2 (\tilde{\beta}_{02t} - \tilde{\beta}_{01t})^2 / C_{XX, N}$ ,

$M_t = \{Z(\hat{\gamma}) - SF_{1, N-4} (1-\delta) / (N-4)\} D_t - (D_t - E_t \tilde{\gamma}_t) (C_t - D_t \tilde{\gamma}_t) (\tilde{\beta}_{02t} - \tilde{\beta}_{01t})^2 /$

$/ C_{XX, N}$  y  $N_t = \{Z(\hat{\gamma}) - SF_{1, N-4} (1-\delta) / (N-4)\} C_t - (C_t - D_t \tilde{\gamma}_t)^2 (\tilde{\beta}_{02t} - \tilde{\beta}_{01t})^2 C_{XX, N}$ .

superior del intervalo de una cola se obtiene en una forma similar.

2)  $\beta_{02}$  conocido e igual a cero:

La región de confianza para muestras grandes para  $\gamma$  es el conjunto de valores de  $r$  que satisfacen:

$$(N-3)\{Z^*(\hat{\gamma})-Z^*(r)\}/S^* \leq F_{1,N-3}(1-\delta) \quad (4.69)$$

con nivel de confianza asintótico  $1-\delta$ .

Adoptando el mismo procedimiento para aproximar el intervalo que en la sección anterior se reemplaza (4.67) por:

$$L_1^* r^2 - 2M_1^* r + N_1^* = 0 \quad (4.70)$$

donde:  $L_1^* = \{Z^*(\hat{\gamma}) - S^* F_{1,N-3}(1-\delta)/(N-3)\} E_1^* - \tilde{\beta}_{02}^2 (D_1^* - E_1^* \hat{\gamma}_1^*)^2$ ,

$M_1^* = \{Z^*(\hat{\gamma}) - S^* F_{1,N-3}(1-\delta)/(N-3)\} D_1^* - \tilde{\beta}_{02}^2 (C_1^* - D_1^* \hat{\gamma}_1^*)(D_1^* - E_1^* \hat{\gamma}_1^*)$  y

$N_1^* = \{Z^*(\hat{\gamma}) - S^* F_{1,N-3}(1-\delta)/(N-3)\} C_1^* - \tilde{\beta}_{02}^2 (C_1^* - D_1^* \hat{\gamma}_1^*)^2$ .

La derivación de  $\gamma_1$  y  $\gamma_2$  se hace en la misma forma que en la sección anterior. Como en el caso de  $\beta_{02}$  desconocido la aproximación normal a la distribución de  $\hat{\theta}$  y  $\beta_{01}$  es adecuada para tamaños de muestra moderados.

Como ya indicamos antes, Hinkley (1971) encuentra una región de confianza conjunta para  $\gamma$  y  $\theta = E(Y_i/X=\gamma)$ . Una de dichas regiones es un rectángulo formado por dos intervalos de confianza individuales, pero esto tiene dos defectos familiares: hay infinitas regiones de ese tipo para ese nivel de confianza y  $\hat{\theta}$  y  $\hat{\gamma}$  están correlacionados, de manera que la región será conservativa. La alternativa es usar regiones de verosimilitud. Como en los casos anteriores considera por separado el caso en que  $\beta_{02}$  sea conocido o desconocido.

1)  $\beta_{02}$  conocido y cero:

La SCR del modelo condicionado a que  $\beta_{02} = 0$ ,  $\tau = t$ ,  $\gamma = r$  y  $\theta = w$  es:

$$S_t^*(r, w) = \sum_{i=1}^N (y_i - \bar{y}_N)^2 - \frac{\left\{ \sum_{i=1}^t (y_i - w) (x_i - r) \right\}^2}{\sum_{i=1}^t (x_i - r)^2} \quad (4.71)$$

$$Z_t^*(r, w) = -N(\bar{y}_N - w)^2 + \frac{\left\{ \beta_{01t} C_{xx,t} + t(\bar{x}_t - r)(\bar{y}_t - w) \right\}^2}{C_{xx,t} + t(\bar{x}_t - r)^2} \quad (4.72)$$

r esta comprendido entre  $x_t$  y  $x_{t+1}$  y t varía entre 2 y N-2.  $Z^*(r) = \max_w Z^*(r, w)$ . En particular  $Z^*(\hat{\gamma}) = Z^*(\hat{\gamma}, \hat{\theta})$ . El estadístico  $Z^*(\hat{\gamma}) - Z^*(\hat{\gamma}, \hat{\theta}) \sim \sigma^2 \chi^2_2$  por la normalidad asintótica de los estimadores MV de manera que la región de confianza  $1-\delta$  de  $(\gamma, \theta)$  para muestras grandes es el conjunto de  $(r, w)$  que satisfacen:

$$Z^*(r, w) \geq Z^*(\hat{\gamma}) - 2S^*F_{2, N-3}(1-\delta)/(N-3) \quad (4.73)$$

Se puede resolver (4.73) para muchos valores de w, o para muchos valores de r. En cualquiera de los dos casos se obtiene un conjunto discreto de puntos en el límite de la región de confianza. Para valores dados de w una solución aproximada se obtiene en una forma similar a la sección anterior. Para valores dados de r (4.73) define una única función cuadrática para todo w, de manera que puede ser resuelta sin dificultad.

Se puede ver fácilmente que los puntos límites para un r dado son las soluciones reales de:

$$G_t^*(r)w^2 - 2H_t^*(r)w + K_t^*(r) = 0 \quad (4.74)$$

$$\text{con } G_t^*(r) = N\{C_{xx,t} + t(\bar{x}_t - r)^2/N\}$$

$$H_t^*(r) = N\{C_{xx,t}(\bar{x}_t - \beta_{01t}(\bar{x}_t - r)/N) - t(\bar{x}_t - r)^2\bar{x}_t/N\}$$

$$K'_t(r) = \{C_{xx,t} + t(\bar{x}_t - r)^2\} \{Z(\hat{\gamma}) - 2SF_{2,N-3}(1-\delta)/(N-3) + N\bar{x}_t^2\} - (\beta_{01}^2 C_{xx,t} + t(\bar{x}_t - r)\bar{x}_t)^2$$

Toda la grilla debe ser cubierta pues pueden ocurrir regiones disjuntas, lo que refleja la no convexidad de la verosimilitud.

## 2) $\beta_{02}$ desconocido:

Cuando  $\beta_{02}$  es desconocido la SCR correspondiente a  $Z'(r,w)$  puede escribirse:

$$Z'(r,w) = -N(\bar{x}_t - w)^2 - \beta_{02}^2 C_{xx,t} + \beta_{01} C_{xx,t} + t(\bar{x}_t - r)(\bar{x}_t - w)^2 / \{C_{xx,t} + t(\bar{x}_t - r)^2\} + t\{\beta_{02}^* C_{xx,t}^* + t^*(\bar{x}_t^* - r)(\bar{x}_t^* - w)^2\} / \{C_{xx,t}^* + t^*(\bar{x}_t^* - r)^2\} \quad (4.75)$$

$$(\bar{x}_t \leq r < \bar{x}_{t+1} \quad ; \quad t=2, \dots, N-2).$$

La región de confianza para muestras grandes es ahora el conjunto de puntos  $(r,w)$  que satisfacen:

$$Z(r,w) \geq Z(\hat{\gamma}) - 2SF_{1,N-4}(1-\delta)/(N-4) \quad (4.76)$$

Las soluciones para un  $w$  dado requieren la solución en una serie de ecuaciones cuadráticas en  $r$ .

Las propiedades de los procedimientos de inferencia discutidos por el autor se basan en la teoría de muestras grandes y pueden no ser aproximaciones adecuadas en muestras pequeñas. Por esta razón Hinkley considera importante investigar las propiedades para éste caso empíricamente. Se generaron 500 muestras de un modelo de regresión de dos fases. En cada caso se ajustaron los dos modelos considerando por separado los casos:  $\beta_{02}$  conocido y  $\beta_{02}$  desconocido. Comparando la media empírica de  $\hat{\beta}_{02}$ ,  $\hat{\gamma}$  y  $\hat{\sigma}^2$  con los valores verdaderos de estos parámetros se observó que hay un sesgo pequeño pero positivo en  $\hat{\gamma}$  cuando  $\beta_{02}$  es conocido, y además la  $\text{Var}(\hat{\gamma})$  es considerablemente mayor que la de la distribución normal que la aproxima. Por otro lado el tamaño empírico del test de razón de verosimilitud conjunto de  $(\gamma, \theta)$  usado en el procedimiento de la región de confianza y el test de razón de verosimilitud de  $H_0: \beta_{02} = 0$  estuvieron cercanos a los valores teóricos.

Hinkley grafica la distribución empírica de  $\hat{\gamma}$  versus los cuartiles normales para  $\beta_{02}$  conocido, siendo evidente la no normalidad de la distribución. La distribución de Hinkley para muestras grandes también presenta un mal comportamiento en este caso. La distribución empírica de  $\hat{\beta}_1$  tiene una buena concordancia con la aproximación normal, con una varianza apenas aumentada.  $\hat{\theta}$  también se comporta razonablemente bien cuando  $\beta_{02}$  es conocido, pero el efecto de sacar esta restricción es drástico, pues en este último caso la  $\text{Var}(\hat{\theta})$  empírica es considerablemente normal para la distribución de  $\hat{\theta}$  es malo para muestras pequeñas. El estimador insesgado de la varianza de  $\hat{\sigma}^2$  presenta un comportamiento muy bueno.

El tamaño pequeño de muestras también tiene influencia en la distribución empírica de  $Z(\hat{\gamma}) - Z(\gamma)$ , pues produce colas un poco más largas que la distribución  $\chi_{(1)}^2$ , que la aproxima, pero debe enfatizarse el hecho de que en muestras un poco mayores que las consideradas por Hinkley en esta simulación la distribución asintótica de  $\hat{\theta}$  y  $Z(\hat{\gamma}) - Z(\gamma)$  son muy buenas aproximaciones, mientras que la distribución de  $\hat{\gamma}$  converge muy lentamente a la normalidad.

Como ya se indico antes Hinkley supone que las observaciones son independientes, se distribuyen normalmente y tienen varianza constante, generalmente desconocida.

#### 4.2.4. Solución propuesta por Williams (1970).

Williams estudia el modelo de regresión de tres fases con una variable explicativa donde se desean estimar los dos puntos de cambio de régimen,  $\gamma_1$  y  $\gamma_2$ . El modelo se puede expresar mediante el siguiente sistema de ecuaciones:

$$\begin{aligned}
 Y_i &= \beta_{01} X_i + \beta_{11} + \varepsilon_i && \text{si } x_i \leq \gamma_1 \\
 Y_i &= \beta_{02} X_i + \beta_{12} + \varepsilon_i && \text{si } \gamma_1 < x_i \leq \gamma_2 \\
 Y_i &= \beta_{03} X_i + \beta_{13} + \varepsilon_i && \text{si } x_i > \gamma_2
 \end{aligned}$$

donde los residuos  $\varepsilon_i$  son independientes, idénticamente distribuidos  $N(0, \sigma^2)$ ; y en forma matricial:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} \quad (4.77)$$

donde  $\varepsilon$  se distribuye  $N(0, \sigma^2 I_N)$ . Esto último no se cumple muchas veces en la práctica, sin embargo el autor ha seguido la práctica usual de suponerlo válido.

La estimación de  $\gamma_1$  y  $\gamma_2$  es complicada pues la función de verosimilitud no es diferenciable respecto a los mismos. En cualquiera de los intervalos  $[x_i, x_{i+1}]$ ,  $[x_j, x_{j+1}]$ , con  $i$  menor que  $j$ , puede existir un máximo local (o supremo) de la función de verosimilitud tal que  $x_i \leq \gamma_1 \leq x_{i+1}$  y  $x_j \leq \gamma_2 \leq x_{j+1}$ .

Los procedimientos iterativos estándares para obtener estimadores MV no son apropiados y es necesario evaluar el supremo local para cada par de intervalos. Williams extiende el procedimiento sugerido por Hudson (Sección 4.2.2) al caso en que haya dos puntos de cambio de régimen. Para ello define un rectángulo en el plano  $(T_1, T_2)$  con vértices  $(x_i, x_j)$ ,  $(x_{i+1}, x_j)$ ,  $(x_i, x_{j+1})$ ,  $(x_{i+1}, x_{j+1})$  definido por  $x_1 < x_i \leq T_1 < x_{i+1} < x_j \leq T_2 \leq x_{j+1} \leq x_N$  de manera que el número de observaciones en los intervalos  $[x_i, x_{i+1}]$ ,  $[x_{i+1}, x_j]$  y  $[x_{j+1}, x_N]$  será respectivamente  $n_1, n_2$  y  $n_3$ . Para cada una de los pares posibles  $(i, j)$  con  $i$  menor que  $j$  la expresión del estimador MC de  $\beta$  es:

$$\hat{\beta}_{(i,j)} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} (X_1' X_1)^{-1} X_1' Y_1 \\ (X_2' X_2)^{-1} X_2' Y_2 \\ (X_3' X_3)^{-1} X_3' Y_3 \end{bmatrix} = (X'_{(i,j)} X_{(i,j)})^{-1} X'_{(i,j)} Y_{(i,j)} \quad (4.78)$$

Sean  $(\gamma_{1(i,j)}^*, \gamma_{2(i,j)}^*)$  las intersecciones de la primera y la segunda y de la segunda y tercera fase de la regresión estimada respectivamente y  $SCR_{i,j}^*$  la SCR de este modelo que para este punto de intersección alcanza su mínimo valor. Si se restringe a las fases de la regresión a que se intersequen en un punto determinado  $(\gamma_1', \gamma_2')$ , con  $\gamma_1'$  que pertenezca al intervalo  $[x_i, x_{i+1}]$  y  $\gamma_2'$  que pertenezca al intervalo  $[x_j, x_{j+1}]$ , con  $i$  menor que  $j$ , el estimador de los parámetros del modelo restringido será:

$$\hat{\beta}_{R(i,j)} = \hat{\beta}_{(i,j)} + (X'_{(i,j)} X_{(i,j)})^{-1} [H(X'_{(i,j)} X_{(i,j)})^{-1} H']^{-1} (-H; \hat{\beta}_{(i,j)}) \quad (4.79)$$

y la SCR

$$SCR_{R(i,j)} = SCR_{(i,j)}^* + (H; \hat{\beta}_{(i,j)}) [H(X'_{(i,j)} X_{(i,j)})^{-1} H']^{-1} (H; \hat{\beta}_{(i,j)}) \quad (4.80)$$

donde  $H = \begin{bmatrix} 1 & \gamma_1' & -1 & -\gamma_1' & 0 & 0 \\ 0 & 0 & 1 & \gamma_2' & -1 & -\gamma_2' \end{bmatrix}$ .

Si  $X_i \leq \gamma_1^* \leq X_{i+1}$  y  $X_j \leq \gamma_2^* \leq X_{j+1}$  entonces la  $SCR_{R(i,j)}$  es un mínimo igual a  $SCR_{(i,j)}^*$ . Si esto no se cumple la  $SCR_{R(i,j)}$  alcanza su mínimo valor en algún punto de los límites del rectángulo con vértices  $(X_i, X_{i+j})$ ;  $(X_i, X_{i+j+1})$ ,  $(X_{i+1}, X_{i+j})$ ,  $(X_{i+1}, X_{i+j+1})$  indicado en la figura.

Williams supone que la función es cóncava en ese rectángulo pero es incapaz de determinar en qué punto del límite alcanza su mínimo, pero para tamaños muestrales suficientemente grandes (mayores que 70) considera que es suficiente calcular la SCR para los cuatro vértices del rectángulo y elegir como estimador de  $(\gamma_1, \gamma_2)_{(i,j)}$  a aquel que la minimiza. El método propuesto por Williams consiste en repetir este procedimiento para todos los

valores posibles de  $(i,j)$  y elegir como estimador de  $(\gamma_1, \gamma_2)$  a aquel que minimiza la  $SCR_{R(i,j)}$  (es necesario que el número de observaciones en cada fase permita estimar los parámetros).

Es obvio que este procedimiento insumirá un tiempo considerable de cálculo; ésto es una restricción importante en el número de simulaciones que se pueden llevar a cabo cuando se quiere discriminar entre modelos alternativos.

#### 4.2.5 Solución propuesta por Gurnow (1973).

Gurnow estudia, al igual que Williams, el caso de las curvas estímulo-respuesta, pero realizando los siguientes supuestos:

- El animal o planta sólo comienza a responder al estímulo cuando éste excede un cierto nivel  $\gamma_1$ , es decir que la respuesta es nula siempre que el valor del estímulo sea menor o igual que  $\gamma_1$ .
- La respuesta será lineal hasta que ésta alcance un cierto valor  $Y^*$  a partir del cual la respuesta permanece constante, aunque el valor del estímulo aumente.

Con lo que la curva estímulo-respuesta individual es:

$$\begin{aligned}
 E(Y_i/X_i) &= 0 && \text{si } X_i < \gamma_1 \\
 &= \beta(X_i - \gamma) && \text{si } \gamma_1 < X_i < \gamma_1 + \frac{Y^*}{\beta} \\
 &= Y^* && \text{si } X_i > \gamma_1 + \frac{Y^*}{\beta}
 \end{aligned} \tag{4.81}$$

que gráficamente se puede observar en la siguiente figura:

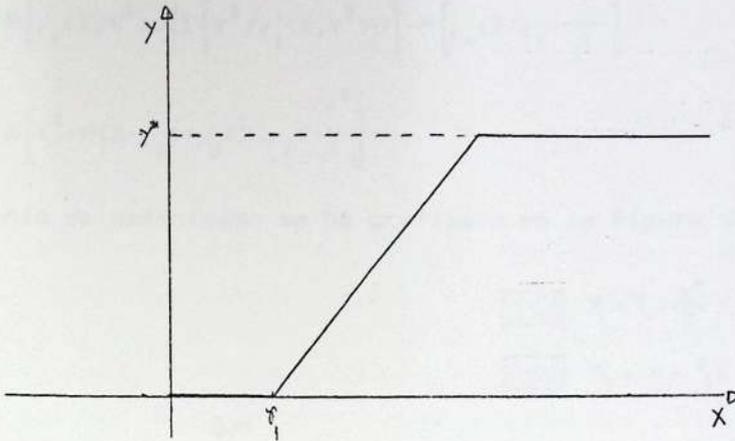


Figura N°1.

Las variaciones individuales en los valores de  $\gamma_1$  y  $Y^*$  producen una curva suave para la población. Para poder hacer progresos analíticos en la estimación del modelo Gurnow se ve obligado a hacer dos supuestos adicionales:

- c)  $\beta$  es igual para todos los individuos de la población  
 d)  $(\gamma_1, Y^*)$  se distribuye como una normal bivariada ( $\gamma_1$  e  $Y^*$  en este caso indican variables, y no parámetros, como en el modelo planteado por Williams, pues se refieren a los valores para cada individuo de la población).

El objetivo del trabajo es estimar los parámetros de la curva estímulo-respuesta para la población. Se elijen las escalas y el origen de los ejes de manera que  $E(\gamma_1)=0$ ,  $E(Y^*)=\mu$ ,  $\text{Var}(\gamma_1)=1$ ,  $\text{Cov}(\gamma_1, Y^*)=\rho$  y  $\text{Var}(Y^*)=1$ .

Para un nivel dado del estímulo,  $X$ , la respuesta esperada será:

$$E(Y/X) = 0 \cdot P\left[X < \gamma_1 \text{ ó } Y^* < 0\right] + \beta P\left[\gamma_1 < X < \gamma_1 + \frac{Y^*}{\beta}\right] E\left[X - \gamma_1 / \gamma_1 < X < \gamma_1 + \frac{Y^*}{\beta}\right] + \\ + P\left[\gamma_1 < \gamma_1 + \frac{Y^*}{\beta} < X\right] E\left[Y^* / \gamma_1 < \gamma_1 + \frac{Y^*}{\beta} < X\right]$$

$$\begin{aligned}
 &= P\left[\gamma_1 < X, Y^* > 0\right] E\left[Y^* / \gamma_1 < X, Y^* > 0\right] - P\left[\gamma_1 < X < \gamma_1 + \frac{Y^*}{\beta}\right] \\
 &E\left[Y^* - \beta(X - \gamma_1) / \gamma_1 < X < \gamma_1 + \frac{Y^*}{\beta}\right] \quad (4.82)
 \end{aligned}$$

El dominio de definición se ha graficado en la Figura N°2.

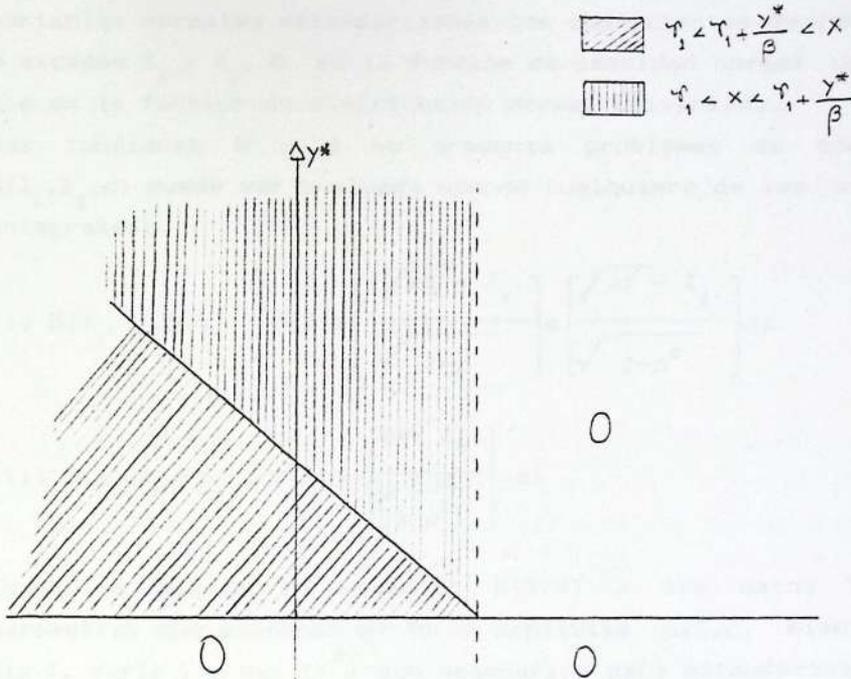


Figura N°2.

Para describir (4.82) en términos de funciones conocidas se usa la expresión de la distribución conjunta de dos variables estandarizadas que se distribuyen como una normal bivariada, con coeficiente de correlación  $\rho$ . En este caso  $(\gamma_1, Y^*)$  se distribuyen como una normal bivariada con parámetros  $E(Y^*) = \mu$ ,  $\text{Var}(Y^*) = 1$ ,  $\text{Cor}(\gamma_1, Y^*) = \rho$ ,  $E(\gamma_1) = 0$ ,  $\text{Var}(\gamma_1) = 1$ ; con lo que la  $E(Y/X)$  tiene la siguiente expresión:

$$\begin{aligned}
 E(Y/X) = & -\mu B\left[-X, -\mu, -\rho\right] - (\mu - \beta X) B\left[-X, \frac{\beta X - \mu}{\theta}, -\frac{\beta + \rho}{\theta}\right] + \theta(\mu) \phi\left[\frac{X + \rho\mu}{\sqrt{1 - \mu^2}}\right] - \\
 & - \theta \phi\left[\frac{X - \mu}{\theta}\right] \phi\left[\frac{-(1 + \beta\rho)X + (\beta + \rho)\mu}{\theta \sqrt{1 - \rho^2}}\right] + \beta \theta(X) \phi\left[\frac{\beta X + \mu}{\sqrt{1 - \rho^2}}\right] \quad (4.83)
 \end{aligned}$$

donde:  $\theta = +\sqrt{1 + \beta^2 + 2\beta\rho}$ ,  $B(Z_1, Z_2, \rho)$  es la probabilidad de que dos variables normales estandarizadas con coeficientes de correlación  $\rho$  excedan  $Z_1$  y  $Z_2$ ,  $\theta$  es la función de densidad normal univariada y  $\phi$  es la función de distribución normal univariada.

Las funciones  $\theta$  y  $\phi$  no presenta problemas de cálculos y  $B(Z_1, Z_2, \rho)$  puede ser evaluada usando cualquiera de las siguientes integrales:

$$(i) B(Z_1, Z_2, \rho) = \theta(X) \phi\left[\frac{\sqrt{\rho X} - Z_1}{\sqrt{1 - \rho}}\right] \phi\left[\frac{\sqrt{\rho X} - Z_2}{\sqrt{1 - \rho^2}}\right] dX \quad (4.84)$$

$$(ii) B(Z_1, Z_2, \rho) = \theta(X) \phi\left[\frac{\rho X - Z_2}{\sqrt{1 - \rho^2}}\right] dX \quad (4.85)$$

Cuando se ajusta la ecuación  $E(Y/X)$  a los datos hay tres parámetros que aparecen en forma explícita  $\mu, \beta, \rho$ , mientras que  $E(\gamma_1)$ ,  $\text{Var}(\gamma_1)$  y  $\text{Var}(Y^*)$  son necesarios para estandarizar estos valores.  $\mu, \beta$  y  $E(\gamma_1)$  determinan la forma general de la curva y  $\rho^*$ ,  $\text{Var}(\gamma_1)$  y  $\text{Var}(Y^*)$  la curvatura de la misma fuera de la región en la que es aproximadamente lineal.

La expresión de la  $\text{Var}(Y/X)$  puede ser hallada, pero será mucho más engorrosa que la de  $E(Y/X)$ ; por esta razón Gurnow rechaza el uso de determinaciones repetidas de  $Y$  para cada valor de  $X$  para obtener una estimación de la misma. En algunas situaciones se dispone de estimaciones externas de  $\rho$ ,  $\text{Var}(\gamma_1)$  y  $\text{Var}(Y^*)$ .

La  $\text{Var}(Y^*)$  puede ser estimada de la  $\text{Var}(Y)$  cuando se conoce que los correspondientes valores del estímulo  $X$  son suficientemente grande como para no limitar la respuesta.

En el ejemplo presentado por Gurnow se ajusta la  $E(Y/X)$  mediante el método de MC no ponderados y el procedimiento iterativo usado es el método simplex de Nelder y Mead.

#### 4.2.6. Solución propuesta por Gallant y Fuller (1973).

Estos autores propusieron una solución para el modelo polinomial con puntos de unión tipo 3 según la denominación de Hudson (Sec.4.2.2) problema al que este último dedicó muy poca atención. En especial ellos consideraron el modelo polinomial de  $r$  fase sujeto a las siguientes restricciones:

- 1) Los residuos  $\epsilon_i$  son independientes y se distribuyen idénticamente  $(0, \sigma^2)$ .
- 2) El modelo de la regresión en cada fase es un polinomio de grado  $K$ , ésto es:

$$g_k(x, \beta_j) = \sum_{i=0}^k \beta_{ij} x^i, \quad k=1, 2, \dots, r$$

donde algunos de los coeficientes pueden ser nulos y no necesariamente todas las fases del polinomio son del mismo grado.

- 3) La función es continua en cada punto de cambio de régimen, ésto es:  $g_j(\gamma_j, \beta_j) = g_{j+1}(\gamma_{j+1}, \beta_{j+1})$ ,  $j=1, 2, \dots, r-1$
- 4) La función es derivable en todo punto de cambio de régimen, ésto es:  $\frac{\delta}{\delta x} g_j(\gamma_j, \beta_j) = \frac{\delta}{\delta x} g_{j+1}(\gamma_{j+1}, \beta_{j+1})$ ,  $j=1, 2, \dots, r-1$ .

Como  $\gamma' = (\gamma_1, \gamma_2, \dots, \gamma_{r-1})$  es un vector desconocido, el modelo es no lineal en los parámetros y la estimación por MC es difícil de obtener.

En este trabajo Gallant y Fuller muestran cómo pueden reparametrizarse modelos de regresión polinomial que cumplan las restricciones (1), (2), (3) y (4) para poder aplicar el ajuste de Gauss-Newton modificado; pues en esta forma a veces es posible satisfacer las condiciones de Gallant [Gallant, "Statistical Inference for Nonlinear Regression Models", Ph.D dissertation, Iowa State University, 1971] de manera que los estimadores MC

sean fuertemente consistentes y se distribuyan asintóticamente normal. Además el modelo reparametrizado permite, en cierta forma, la aplicación de los resultados de Gallant para test de hipótesis de modelos no lineales.

No es posible explicar en forma general cómo se realiza la reparametrización así que los autores muestran mediante un ejemplo, el de la forma cuadrático-cuadrático, cómo deben incorporarse las condiciones (3) y (4) en la expresión del modelo para lograrlo.

Sean  $g_1(x, \beta_1) = \beta_{01} + \beta_{11}x + \beta_{21}x^2$  y  $g_2(x, \beta_2) = \beta_{02} + \beta_{12}x + \beta_{22}x^2$ . Por las condiciones (3) y (4):  $\beta_{01} + \beta_{11}\gamma + \beta_{21}\gamma^2 = \beta_{02} + \beta_{12}\gamma + \beta_{22}\gamma^2$  y  $\beta_{11} + 2\beta_{21}\gamma = \beta_{12} + 2\beta_{22}\gamma$ . Despejando  $\beta_{01}$  y  $\beta_{11}$  y reemplazando en  $g_1(x)$ :  $g_1(x) = g_2(x) + (\beta_{21} - \beta_{22})(\gamma - x)^2$ . Esto sugiere la reparametrización:  $\theta_1 = \beta_{02}$ ,  $\theta_2 = \beta_{12}$ ,  $\theta_3 = \beta_{22}$ ,  $\theta_4 = \beta_{21} - \beta_{22}$  y  $\theta_5 = \gamma$ ; de manera que los submodelos puedan escribirse conjuntamente como un único modelo de una ecuación diferencial:

$$f(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 (e_5 - x)^2 I_+(e_5 - x)$$

$$\text{donde } I_+(Z) = 1 \text{ si } Z \geq 0 \\ = 0 \text{ si } Z < 0.$$

El vector de derivadas parciales es:

$$\nabla' f(x, \theta) = (1, x, x^2, (e_5 - x)^2 I_+(e_5 - x), 2\theta_4 (e_5 - x) I_+(e_5 - x))$$

y la matriz  $F(\theta)_{n \times 5} = |\nabla' f(x_i, \theta)|$  cuyas filas son los  $\nabla' f(x_i, \theta)$  tendrá rango cinco si  $\theta_4$  es distinto de cero, o si hay cinco puntos  $x_{t_1}, x_{t_2}, \dots, x_{t_5}$  en  $\{x_i\}_{i=1}^n$  tal que:  $x_1 < x_{t_1} < x_{t_2} < \theta_5 < x_{t_3} < x_{t_4} < x_{t_5} < x_N$  ó

$$x_1 < x_{t_1} < x_{t_2} < x_{t_3} < \theta_5 < x_{t_4} < x_{t_5} < x_N.$$

En la tesis doctoral ya citada se dan las condiciones en la secuencia  $\{x_i\}$  para el modelo cuadrático-cuadrático de manera que los estimadores MC de  $\theta$  sean consistentes y se distribuyen asintóticamente normal. Este es el primer trabajo que tengamos referencia donde las condiciones para que los estimadores posean propiedades óptimas se dan sobre los valores observados de la

variable independiente.

El objeto de haber reparametrizado el modelo es que ahora se puede aplicar el método modificado de Gauss-Newton que es una forma eficiente de encontrar el mínimo de:

$$Q(\theta) = \sum_{i=1}^n \{y_i - f(x_i, \theta)\}^2 \quad (4.86)$$

cuando  $f$  es un "grafted polynomial model" en la forma reparametrizada. Esto ha sido demostrado por Gallant y Fuller.

Hartley ["The Modified Gauss-Newton Method for the fitting on Non-Linear Regression Functions by least squares". *Technometrics*, 3 (May 1961), 269-80] requiere que las derivadas parciales segundas de  $f(x, \theta)$  en  $\theta$  existan y sean continuas en su prueba de que las iteraciones converjan. Afortunadamente esta condición no es necesaria para obtener convergencia y los autores replantean el teorema de Hartley como sigue: (la demostración se haya en la tesis doctoral ya citada).

Teorema:

Dado un modelo de regresión  $Y_i = f(x_i, \theta) + \varepsilon_i$  y los pares de observaciones  $(y_i, x_i)$  ( $i=1, 2, \dots, N$ ). Si existe un subconjunto  $S$  convexo y acotado de  $R^p$  y un  $\theta_0$  interior a  $S$  tal que:

- 1-  $\nabla f(x_i, \theta)$  existe y es continua sobre  $S$  para  $i=1, 2, \dots, N$
- 2-  $\theta \in S \Rightarrow \text{rango} F(\theta) = p$
- 3-  $Q(\theta_0) < \inf \{Q(\theta) : \theta \text{ es un punto del límite de } S\}$ .
- 4- No existe  $\theta', \theta''$  en  $S$  tal que:  $\nabla Q(\theta') = \nabla Q(\theta'') = 0$   
y  $Q(\theta') = Q(\theta'')$

La secuencia  $\{\theta_k\}_{k=1}^{\infty}$  se construye de la siguiente forma:

$$0- \text{Calcular } D'_0 = [F'(\theta_0) F(\theta_0)]^{-1} F'(\theta_0) [Y - f(\theta_0)]$$

Encontrar  $\lambda_0$  que minimiza  $Q(\theta_0 + \lambda D'_0)$  sobre  $\Lambda_0 = \{\lambda : 0 \leq \lambda \leq 1, \theta_0 + \lambda D'_0 \in S\}$

$$1- \text{Establecer } \theta_1 = \theta_0 + \lambda_0 D'_0$$

$$\text{Calcular } D'_1 = [F'(\theta_1) F(\theta_1)]^{-1} F'(\theta_1) [Y - f(\theta_1)]$$

Encontrar  $\lambda_1$  que minimiza  $Q(\theta_1 + \lambda D'_1)$  sobre  $\Lambda_1 = \{\lambda : 0 \leq \lambda \leq 1, \theta_1 + \lambda D'_1 \in S\}$

$$2- \text{Establecer } \theta_2 = \theta_1 + \lambda_1 D'_1$$

Entonces para la secuencia  $\{\theta_k\}_{k=1}^{\infty}$  así construída se cumple que:

- 1-  $\theta_k$  es un punto interior de S para  $k=1,2,\dots$ .
- 2- La secuencia  $\{\theta_k\}$  converge a un límite  $\theta^*$  que es interior a S.
- 3-  $\nabla Q(\theta^*) = 0$ .

Nota: en la práctica no es necesario encontrar cada valor de  $\lambda_k$  exactamente sino sólo encontrar  $\hat{\lambda}_k \in \Lambda_k$  tal que  $Q(\theta_k + \hat{\lambda}_k D_k) \leq Q(\theta_k)$  y  $\lambda_k \neq 0$ . Si dicho valor no puede ser hallado y  $d_k$  no está suficientemente cercano a cero el método falla. Si se elige otro valor inicial  $\theta_0$  se puede obtener convergencia.

El punto inicial  $\theta_0$  se obtiene graficando los pares  $(x_i, y_i)$  y eligiendo puntos de unión por inspección y los restantes parámetros pueden ser estimados por métodos de regresión múltiple estándares.

En resumen podemos decir que la propuesta de Gallant y Fuller es que reparametrizando el modelo puede aplicarse el teorema de Hartley que asegura convergencia. Si además los  $x_i$  cumplen las condiciones de Gallant los estimadores MC serán consistentes y asintóticamente normales. La reparametrización bajo estas condiciones sólo asegura convergencia pero no permite saber nada sobre las propiedades de los estimadores obtenidos.

El aporte es válido sólo para los casos particulares que él trata, por esta razón el método es explicado mediante ejemplos pues debe darse un tratamiento particular a cada modelo.

## 4.3. Estimación de los parámetros de la regresión.

## 4.3.1. Solución propuesta por Bacon y Watts (1971).

Estos autores sostienen que aunque la forma en que tradicionalmente se ha planteado el modelo de regresión a trozos es la más obvia no es adecuada en muchos casos pues no es una parametrización sensible para detectar cambios suaves en la pendiente.

Ellos proponen un modelo empírico más general que incluye una curva de transición (trn) con un parámetro que describe la transición del modelo en el punto de cambio de régimen que es adecuado tanto cuando hay una transición suave como cuando hay un cambio abrupto de un régimen a otro.

Aunque la forma en que Bacon y Watts han planteado el modelo no es consecuente con el modelo general al cual siempre hacemos referencia lo hemos incorporado en la ordenación pues constituye un enfoque interesante al problema.

La expresión del modelo es:

$$Y_i = \alpha_0 + \alpha_1(x_i - \gamma) + \alpha_2(x_i - \gamma) \operatorname{trn} \left[ \frac{(x_i - \gamma)}{\delta} \right] + \varepsilon_i \quad (4.86)$$

donde  $\gamma$  es el change-over value (abscisa del punto donde se produce el cambio de régimen) y  $\delta$  es el parámetro de transición. Si  $\delta = 0$  la transición será abrupta mientras que si  $\delta$  es mayor que uno la transición será suave. No es necesario hacer ningún supuesto sobre el valor de  $\delta$ .

La función de transición debe satisfacer las siguientes condiciones:

$$(i) \lim_{s \rightarrow \infty} \operatorname{trn} \left[ \frac{|s|}{\delta} \right] = 1$$

$$(ii) \operatorname{trn}(0) = 0$$

$$(iii) \lim_{\delta \rightarrow 0} \operatorname{trn} \left( \frac{S}{\delta} \right) = \operatorname{sgn}(s) \quad (4.87)$$

$$(iv) \lim_{S \rightarrow \infty} S \operatorname{trn} \left( \frac{S}{\delta} \right) = S$$

Existen muchas funciones de transición que pueden ser usadas, por ejemplo la función de distribución de cualquier función de densidad simétrica, o la tangente hiperbólica.

En realidad la forma de la función de transición es de importancia secundaria pues el efecto de la variabilidad de los datos puede encubrir cualquier diferencia introducida por diferentes funciones de transición; además aparece en el modelo multiplicada por  $(x-\gamma)$  y el comportamiento general de la cantidad

$(x-\gamma) \operatorname{trn} \left\{ \frac{(x-\gamma)}{\delta} \right\}$  es apenas afectado por la forma particular de la función. Para ilustrar ésto el trabajo muestra los valores de

$(x-\gamma) \operatorname{trn} \left\{ \frac{(x-\gamma)}{\delta} \right\}$  para tres funciones distintas, observándose que las diferencias son extremadamente pequeñas. Aunque en el trabajo los autores emplearon una forma particular, el procedimiento de estimación es aplicable a cualquier caso. Ellos hicieron  $\operatorname{trn} \left( \frac{S}{\delta} \right) = \tanh \left( \frac{S}{\delta} \right)$  con lo que el modelo es:

$$Y_i = \alpha_0 + \alpha_1 (x_i - \gamma) + \alpha_2 (x_i - \gamma) \tanh \left[ \frac{(x_i - \gamma)}{\delta} \right] + \varepsilon_i \quad (4.88)$$

y en forma matricial:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 - \gamma & (x_1 - \gamma) \tanh \left[ \frac{x_1 - \gamma}{\delta} \right] \\ \vdots & \vdots & \vdots \\ 1 & x_N - \gamma & (x_N - \gamma) \tanh \left[ \frac{x_N - \gamma}{\delta} \right] \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

donde  $\varepsilon$  se distribuye  $N(0, \sigma^2 I_N)$ .

Es interesante notar que para cualquier función de transición el radio de curvatura  $R$  para el modelo en el punto  $\gamma$  es  $R = \frac{\delta}{2|\alpha_2|} (1 + \alpha_1^2)^{3/2}$ ; con lo que  $\delta$  puede ser llamado el parámetro del radio de curvatura.

Si se grafica  $(x_i - \gamma)$  versus  $(x_i - \gamma) \tanh\left(\frac{x_i - \gamma}{\delta}\right)$  se observa que valores pequeños de  $\delta$  producen sólo desvíos pequeños de la recta. Bacon y Watts aplicaron un enfoque bayesiano para tratar este problema.

Sea  $\theta = (\alpha_0, \alpha_1, \alpha_2, \gamma, \delta, \sigma)$  el vector de parámetros a estimar;

$$L(\theta/X, Y) = \frac{1}{(\gamma \sqrt{2\pi} \sigma)^N} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \alpha_0 - (x_i - \gamma) [\alpha_1 + \alpha_2 \tanh(\frac{x_i - \gamma}{\delta})])^2\right\}$$

$$= \frac{1}{(\gamma \sqrt{2\pi} \sigma)^N} \exp\left\{-\frac{1}{2\sigma^2} S(\theta)\right\} \quad (4.89)$$

la función de verosimilitud conjunta de la muestra y  $\pi(\theta) \propto \frac{1}{\sigma}$  la función de densidad a priori de  $\theta$ , con lo que la función de densidad a posteriori de  $\theta$  es:

$$\pi_1(\theta/X, Y) \propto \left(\frac{1}{\sigma}\right)^{N+1} \exp\left\{-\frac{1}{2\sigma^2} S(\theta)\right\} \quad (4.90)$$

Para valores dados de  $\gamma$  y  $\delta$  se puede expresar como:

$$S(\alpha/\gamma, \delta) = S(\hat{\alpha}/\gamma, \delta) + (\alpha - \hat{\alpha})' X'_{(\gamma, \delta)} X_{(\gamma, \delta)} (\alpha - \hat{\alpha}) \quad (4.91)$$

donde  $X_{(\gamma, \delta)}$  es la matriz regresora para valores dados de  $(\gamma, \delta)$  y  $\hat{\alpha}$  es el estimador mínimo cuadrado de  $\alpha$ . Reemplazando (4.91) en (4.90):

$$\pi_1(\theta/X, Y) \propto \left(\frac{1}{\sigma}\right)^{N+1} \exp\left\{-\frac{1}{2\sigma^2} [S(\hat{\alpha}/\gamma, \delta) + (\alpha - \hat{\alpha})' X'_{(\gamma, \delta)} X_{(\gamma, \delta)} (\alpha - \hat{\alpha})]\right\} \quad (4.92)$$

que si se integra respecto a  $\sigma$  y  $\alpha$  se obtiene la función a

posteriori de  $(\gamma, \delta)$

$$\pi_1(\gamma, \delta) \propto \frac{1}{|X'(\gamma, \delta) X(\gamma, \delta)|^{1/2}} \left[ S(\hat{\alpha}/\gamma, \delta) \right]^{-\frac{1}{2}(N-3)} \quad (4.93)$$

$\pi_1(\gamma, \delta)$  es proporcional a la SCR condicional elevada a la potencia  $\frac{1}{2}(N-3)$ . La reducción del exponente de  $N$  a  $(N-3)$  refleja el uso del estimador MC  $\hat{\alpha}$ . El determinante  $|X(\gamma, \delta) X(\gamma, \delta)|^{1/2}$  es relativamente constante sobre la región de interés de  $\gamma$  y  $\delta$ , y por lo tanto juega un rol secundario en la determinación del comportamiento a posteriori.

Debe maximizarse  $\pi_1(\gamma, \delta)$ , por algún método de maximización no lineal, para determinar qué par  $(\gamma, \delta)$  hace que ésta alcance su máximo valor. Una vez estimados  $(\gamma, \delta)$  pueden estimarse los restantes parámetros del modelo.

#### 4.4. Test de hipótesis sobre la ubicación del punto de cambio de régimen.

##### 4.4.1. Solución propuesta por Sprent (1961)

Sprent estudia el modelo de regresión lineal de dos fases con una variable explicativa, donde se conoce de antemano el intervalo en el cual se produce el cambio de régimen  $[x_{\tau}, x_{\tau+1}]$  y los residuos cumplen con los supuestos del modelo de regresión lineal clásico; con lo que el método de mínimos cuadrados es apropiado para estimar los parámetros del modelo.

El propone un estadístico para testar la hipótesis nula de que el punto de cambio de régimen coincide con un valor dado ( $H_0: \gamma = \gamma_0$  vs  $H_1: \gamma \neq \gamma_0$ ). Está basado en la comparación del ajuste mediante técnicas estándares de regresión directas no restringidas en cada fase con aquel obtenido si se restringen las rectas a que se intersequen en  $\gamma_0$ . La restricción puede expresarse como  $\beta_{11} + \beta_{01}\gamma_0 = \beta_{12} + \beta_{02}\gamma_0$ . Para estimar los parámetros del modelo restringido usa el método de los multiplicadores de Lagrange. Para testar  $H_0$  sugiere realizar un test F con una varianza del error apropiada pero no lo construye.

Nosotros hemos decidido seguir un procedimiento más sencillo usando la teoría de modelos lineales que permitirá realizar esto último con mayor rapidez y facilidad.

La hipótesis nula  $H_0: \gamma = \gamma_0$  puede replantearse como  $H_0: H\beta = 0$  versus  $H_1: H\beta \neq 0$ , con  $H = [-1 \ -\gamma_0 \ 1 \ \gamma_0]$ .

La diferencia entre la SCR bajo  $H_0$  y  $H_1$  es:

$$SCR_{H_0} - SCR_{H_1} = (H\hat{\beta}_{H_1})' [H(X'X)^{-1}H']^{-1} (H\hat{\beta}_{H_1}) \quad (4.94)$$

que tiene un grado de libertad.  $\hat{\beta}_{H_1}$  es el estimador de los parámetros sin restringir. Como por los supuestos del modelo tanto bajo  $H_0$  como bajo  $H_1$  la varianza del error es la misma se puede construir el siguiente estadístico:

$$\frac{SCR_{H_0} - SCR_{H_1}}{SCR_{H_1} / (N-4)} \quad (4.95)$$

que bajo la hipótesis nula se distribuye como una  $F_{(1, N-4)}$ . Si su valor es menor que el valor de  $F_{(1-\alpha)}(1, N-4)$  para un nivel de significancia previamente determinado se acepta la hipótesis nula y en caso contrario se rechaza.

#### 4.4.2. Solución propuesta por Hinkley (1969-1970)

Ya hemos tratado en otras secciones (4.1.4 y 4.2.3) los trabajos de Hinkley. Él estudia el modelo de regresión lineal de dos fases. En 4.2.3 se presenta la distribución que él encuentra del estimador MV del punto de cambio de régimen  $\gamma$ . Si se desean encontrar intervalos de confianza para  $\gamma$ ,  $\theta$ ,  $\beta_{01}$ ,  $\beta_{22}$  existen dos obstáculos para usar la distribución asintótica normal de las estimaciones directamente: la varianza depende de parámetros desconocidos, en especial la  $\text{Var}(\hat{\gamma})$  depende fuertemente de  $\beta_{22} - \beta_{21}$  y la distribución normal asintótica puede ser una pobre aproximación. Afortunadamente la distribución asintótica  $\chi^2$  del estadístico de razón de verosimilitud parece ser una aproximación razonable para muestras pequeñas.

Hinkley (1969) presenta el estadístico de razón de verosimilitud para testar la hipótesis nula  $\gamma = \gamma_0$

$$\Lambda = -2 \log [L(\gamma_0) / L(\hat{\gamma})] \quad (4.96)$$

que se distribuye como una  $\chi^2_{(1)}$  bajo la hipótesis nula. Esto se deduce de la normalidad asintótica del estimador M.V.  $L$  es la función de verosimilitud evaluada en  $\gamma_0$  y  $\hat{\gamma}$  respectivamente, y  $\hat{\gamma}$  es el estimador M.V. de  $\gamma$ . Por la definición de  $Z_{\tau}^2(\gamma)$ , expresión (4.42), (4.96) puede expresarse de la siguiente forma:

$$\Lambda = \frac{1}{\sigma^2} \left[ Z_{\tau}^2(\hat{\gamma}) - Z_{\tau}^2(\gamma_0) \right] \quad (4.97)$$

donde  $\gamma_0$  está comprendido entre  $x_{\tau_0}$  y  $x_{\tau_0+1}$  y  $\hat{\gamma}$  entre  $x_{\tau}^{\wedge}$  y  $x_{\tau+1}^{\wedge}$ . Cuando  $\sigma^2$  es desconocido puede reemplazarse por la SCR estimada, en cuyo caso la distribución de  $\Lambda$  se aproxima mejor mediante una distribución F y la regla del test en este caso es aceptar  $H_0$  si

$$\frac{Z_{\tau}^2(\hat{\gamma}) - Z_{\tau_0}^2(\gamma_0)}{\hat{\sigma}^2} (N-4) \leq F_{1, N-4}(1-\delta) \quad (4.98)$$

donde  $\delta$  es el tamaño asintótico del test y es una aproximación para muestras pequeñas.  $Z(r)$  no es necesariamente una función monótona decreciente de  $Z(\hat{\gamma})$ , por lo que la región no es necesariamente un único intervalo. Los argumentos que él esgrime son empíricos.

Hinkley examina la precisión de la distribución mediante el método de Monte Carlo. Si el tamaño de muestra es pequeño la distribución empírica de  $Z(\hat{\gamma}) - Z(\gamma)$  se ve afectada y produce colas un poco más largas que la  $\chi_{(1)}^2$ . También examinó la distribución empírica de  $[Z(\hat{\gamma}) - Z(\gamma)] / \hat{\sigma}^2$ , calculó las potencias empíricas de los tests de dos colas de tamaño 0,90; 0,95 y 0,99 para la hipótesis nula  $\gamma = \gamma_0$  y  $\beta_{02}$  desconocido para cinco valores distintos de  $\gamma_0$ , incluyendo el verdadero y observó que existe gran coincidencia entre el tamaño teórico y el empírico.

Nota: debe recordarse que en todo el desarrollo del trabajo Hinkley supone que las observaciones son independientes e idénticamente distribuidas normales  $(0, \sigma^2)$ .

## 4.5. Tests de Hipótesis sobre los parámetros de la Regresión.

## 4.5.1. Soluciones propuestas por Sprent (1961).

Sprent considera el caso en que existen  $g$  grupos de observaciones (por ej. datos de distintas unidades económicas de distintos países o distintas regiones) y el modelo para cada uno de ellos es una regresión de dos fases con una variable explicativa y se desean testar distintas hipótesis sobre los parámetros de los distintos grupos.

El autor supone que los datos de cada grupo cumplen con los supuestos del modelo de regresión lineal clásico y además que se conoce, para cada grupo, el intervalo en el cual se produce el cambio de régimen. (El problema cuando el número de grupos es igual a uno se discutió en la Sección 4.4.1). Si esta última condición no se verifica Sprent sostiene que puede usarse el método de Quandt para estimarlo sin consecuencias serias en la utilización de los test propuestos. Nosotros creemos que esa sugerencia no es válida pues en el modelo tratado por Quandt la variable que gobierna el cambio de régimen; además, suponiendo que este inconveniente pudiese subsanarse habría que averiguar qué propiedades y distribuciones tendrán los estimadores M.C. propuestos por él.

El modelo en este caso es (2.6) con  $r$  igual a dos y  $\beta' = (\beta_{011} \beta_{111} \beta_{021} \beta_{121} \dots \beta_{01g} \beta_{11g} \beta_{02g} \beta_{12g})$ .

Se detallan a continuación las hipótesis citadas por Sprent:

$$A) H_0: \beta_{01j} = \beta_{01} \text{ y } \beta_{02j} = \beta_{02}; j=1,2,\dots,g.$$

(Se desea testar si las rectas de cada fase, para todos los grupos son paralelas).

En realidad Sprent postula esta hipótesis pero no propone ningún estadístico. Nosotros lo hemos desarrollado utilizando la teoría de las hipótesis lineales. En este caso las hipótesis son  $H_0: H\beta=0$  vs  $H_1: H\beta \neq 0$ ; con

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & & & & & \\ \vdots & & & & & & & & & & & & \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & & & & & \\ \vdots & & & & & & & & & & & & \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & -1 & 0 \end{pmatrix}.$$

La diferencia entre la SCR bajo  $H_0$  y bajo  $H_1$  es:

$$SCR_{H_0} - SCR_{H_1} = (H\beta_{H_1})' \left[ H(X'X)^{-1}H' \right]^{-1} (H\beta_{H_1}) \quad (4.99)$$

donde  $SCR_{H_0}$  tiene  $(N-2g-2)g.l.$  y  $SCR_{H_1}$   $(N-4g)g.l.$  Si la varianza del error es la misma en ambos casos puede construirse el siguiente estadístico de prueba.

$$\frac{(SCR_{H_0} - SCR_{H_1}) / (2g-2)}{SCR_{H_1} / (N-4g)} \quad (4.100)$$

que bajo  $H_0$  se distribuye como una F con  $(2g-2, N-4g)g.l.$

$$B) H_0: \beta_{01j} = \beta_{01} \text{ y } \beta_{11j} = \beta_{11} \quad ; \quad j=1,2,\dots,g$$

(Se desea testar si las rectas dentro de una fase coinciden. En este caso, sin pérdida de generalidad, se ha supuesto que la coincidencia se produce en la primera fase).

Sprent plantea esta hipótesis pero no propone ningún estadístico para testarlo. Nosotros lo hemos desarrollado utilizando la teoría de las hipótesis lineales.

La hipótesis nula también puede expresarse como  $H\beta=0$  y la alternativa  $H\beta \neq 0$ , con

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ \cdot & & & & & & & & & & & & & & & & & \\ \cdot & & & & & & & & & & & & & & & & & \\ \cdot & & & & & & & & & & & & & & & & & \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ \cdot & & & & & & & & & & & & & & & & & \\ \cdot & & & & & & & & & & & & & & & & & \\ \cdot & & & & & & & & & & & & & & & & & \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & -1 & 0 & 0 \end{bmatrix}$$

La diferencia entre la  $SCR_{H_0}$  y la  $SCR_{H_1}$  es (4.99) donde  $SCR_{H_0}$  tiene  $(N-2g-2)g.l.$  y la  $SCR_{H_1}$   $(N-4g)g.l.$  Si la varianza del error es la misma en ambos casos se puede construir el siguiente estadístico:

$$\frac{(SCR_{H_0} - SCR_{H_1}) / (2g-2)}{SCR_{H_1} / (N-4g)} \tag{4.101}$$

que bajo la hipótesis nula se distribuye como una F con  $(2g-2, N-4g)g.l.$

C) Test para determinar si los puntos de cambio de régimen están sobre una recta conocida (de pendiente conocida o desconocida).

Como en los casos anteriores Sprent sólo plantea esta hipótesis pero no propone ningún estadístico. Nosotros hemos propuesto uno siguiendo un procedimiento similar al aplicado en (A) y (B).

Si la recta tiene pendiente infinita  $X=\gamma_0$  bajo  $H_0$  entonces  $\beta_{11j} + \beta_{01j} \gamma_0 = \beta_{12j} + \beta_{02j} \gamma_0 \Rightarrow (\beta_{12j} - \beta_{11j}) + \gamma_0 (\beta_{02j} - \beta_{01j}) = 0$  donde  $j$  varía entre 1 y  $g$ . Las hipótesis a testar puede expresarse como  $H_0: H\beta=0$  vs  $H_1: H\beta \neq 0$ , donde

$$H = \begin{bmatrix} -\gamma_0 & -1 & \gamma_0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\gamma_0 & -1 & \gamma_0 & 1 & 0 & \dots & \dots & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & \dots & -\gamma_0 & -1 & \gamma_0 & 1 \end{bmatrix}_{(g \times 4g)}$$

$SCR_{H_0} - SCR_{H_1}$  es igual a (4.9B) y  $\hat{\beta}_{H_0} = \hat{\beta}_{H_1} - (X'X)^{-1}H'[H(X'X)^{-1}H']^{-1}H\hat{\beta}_{H_1}$ .

El estadístico de prueba es:

$$\frac{(SCR_{H_0} - SCR_{H_1}) / g}{SCR_{H_1} / (N-4g)} \quad (4.102)$$

que bajo la hipótesis nula se distribuye como una F con  $(g, N-4g)g.l.$

D) Test para determinar si los punto de cambio de régimen de todas las regresiones yacen sobre la recta  $X = \delta$ , con  $\delta$  desconocido, o sobre la recta  $Y = \alpha_0 + \alpha_1 x$  con  $\alpha_0$  desconocido. Para cada grupo se conoce el intervalo donde se produce el cambio de régimen. Para hacer más sencilla la prueba de esta hipótesis Sprent supone que se ha aceptado antes que cada una de las regresiones sigue un régimen de dos fases y que las rectas de cada fase son paralelas entre sí.

$$H_0: \beta_{01j} = \beta_{01} \text{ , } \beta_{02j} = \beta_{02} \text{ y } (\beta_{12j} - \beta_{11j}) + \delta(\beta_{02} - \beta_{01}) = 0 \text{ ; } j=1,2,\dots,g.$$

Sprent propone encontrar los estimadores MC sujetos a las  $g$  restricciones mediante el método de los multiplicadores de Lagrange. Nosotros hemos desarrollado la solución propuesta por Sprent en forma matricial para hacerlo consecuente con el modelo general al cual siempre hacemos referencia. En este caso  $H_0$  puede expresarse como  $H_0: H\beta = r$ , con

$$H = \begin{bmatrix} 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & . & . & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & . & . & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & . & . & 0 & -1 & 0 & 1 \end{bmatrix} \quad \text{y } r = \gamma \mathbf{1}_{(g \times 1)}$$

$g \times 4$

Los estimadores MC de  $\beta$  sujetos a la restricción se pueden calcular derivando la siguiente expresión:

$$\psi = (Y - X\beta_{H_0})' (Y - X\beta_{H_0}) - 2\lambda' (H\beta_{H_0} - \gamma \mathbf{1}), \quad (4.103)$$

( $\lambda'$  es el vector columna de  $g$  multiplicadores de Lagrange). En este caso la diferencia en la SCR bajo la hipótesis nula y la alternativa es:

$$SCR_{H_0} - SCR_{H_1} = (\hat{\gamma} \mathbf{1} - H\hat{\beta}_{H_1})' [H(X'X)^{-1}H']^{-1} (\hat{\gamma} \mathbf{1} - H\hat{\beta}_{H_1}) \quad (4.104)$$

y el estadístico para testar  $H_0$  es:

$$\frac{(SCR_{H_0} - SCR_{H_1}) / (g-1)}{SCR_{H_1} / (N-2g-2)} \quad (4.105)$$

que bajo la hipótesis nula se distribuye como una  $F$  con  $(g-1, N-2g-2)$  grados de libertad.

#### 4.5.2. Solución propuesta por Hinkley (1971)

Hinkley se refiere en su trabajo al modelo de regresión de dos fases con una variable explicativa (Sección 4.1.4) y propone dos test asintóticamente equivalentes para testar la hipótesis

nula de que la pendiente en la segunda fase es nula ( $\beta_{02}=0$ )

a) El primero es:

$$\hat{\beta}_{02} \sqrt{(N-4)C_{xx,\tau}^* / S^2} \quad (4.106)$$

que bajo la hipótesis nula se distribuye asintóticamente como una  $t_{(N-4)}$  si  $\beta_{01}$  es distinto de cero.  $\hat{\beta}_{02}$  es el estimador MC de  $\beta_{02}$ ,  $S^2$  es la suma de cuadrados residual usada para estimar  $\sigma^2$ ; y  $C_{xx,\tau}^*$  y  $\bar{X}_\tau^*$  están dados respectivamente por (4.12) y (4.9).

Habría que determinar si la distribución del estadístico es una  $t$  aun cuando el punto de cambio de régimen sea estimado a partir de los datos.

b) El segundo es un test de razón de verosimilitud

$$V = (S' - S)(N-4) / S \quad (4.107)$$

que bajo la hipótesis nula se distribuye asintóticamente como una  $F$  con  $(1, N-4)$  grados de libertad.  $S$  es la SCR alrededor del modelo ajustado de dos fases y  $S'$  es la SCR bajo la hipótesis nula. Tiene la ventaja de no involucrar al parámetro sin sentido  $\tau$ .

El argumento del autor para sostener cuáles son las distribuciones de los estadísticos propuestos son las distribuciones asintóticas de los estimadores que él derivó en forma empírica.

## 4.6. Discriminación entre modelos alternativos.

## 4.6.1. Solución propuesta por Williams (1970).

Williams desarrolla un estadístico para discriminar entre el modelo exponencial y el modelo de regresión lineal de tres fases con una variable explicativa. El compara estos dos modelos para la síntesis de enzimas durante el ciclo de la célula. Usa el criterio de razón de verosimilitud.

El modelo exponencial es:

$$Y_i = \alpha_0 + \alpha_1 \exp(\alpha_2 X_i) + \varepsilon_i, \quad i=1,2,\dots,N; \quad (4.108)$$

y en forma matricial:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & \varepsilon^{\alpha_2} X_1 \\ 1 & \varepsilon^{\alpha_2} X_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & \varepsilon^{\alpha_2} X_N \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_N \end{pmatrix}$$

que resumiendo es  $Y = X_e \alpha + \varepsilon$ , con  $\varepsilon$  que se distribuye  $N(0, \sigma_g^2 I_N)$ . El parámetro que debe estimarse es  $\psi = (\alpha_0, \alpha_1, \alpha_2)$  y para ello se deben aplicar métodos no lineales.

El modelo lineal de tres fases es (4.77) con  $\varepsilon$  que se distribuye  $N(0, \sigma_f^2 I_N)$ . En este caso el parámetro a estimar es  $\theta = (\beta, \gamma_1, \gamma_2)$ . El método para hacerlo ya se indicó en la Sección 4.2.4.

Los dos modelos son distintos en el sentido de que ninguno de ellos es un caso especial del otro. Cox usa como criterio de discriminación la razón de verosimilitud que se reduce a la razón de la SCR. Si se considera como la hipótesis nula a testar que el modelo es exponencial versus la hipótesis alternativa de que el modelo es lineal de tres fases el test de razón de verosimilitud tendrá la siguiente expresión:

$$\lambda = \frac{L(\text{exp})}{L(\text{lineal})} = \frac{\left(\frac{1}{2\pi\hat{\sigma}_g}\right)^{N/2} \exp\left[-\frac{1}{2\hat{\sigma}_g^2} \sum_{i=1}^N (Y_i - \hat{\alpha}_0 - \hat{\alpha}_1 - e^{\hat{\alpha}_2 X_{i1}})^2\right]}{\left(\frac{1}{2\pi\hat{\sigma}_f}\right)^{N/2} \exp\left[-\frac{1}{2\hat{\sigma}_f^2} (Y-X\beta)'(Y-X\beta)\right]}$$

$$= \frac{\left(\frac{1}{2\pi\hat{\sigma}_g}\right)^{N/2} \exp\left(-\frac{N}{2}\right)}{\left(\frac{1}{2\pi\hat{\sigma}_f}\right)^{N/2} \exp\left(-\frac{N}{2}\right)}$$

$$\lambda^{2/N} = \frac{\hat{\sigma}_f^2 \text{ SCR alrededor del modelo lineal}}{\hat{\sigma}_g^2 \text{ SCR alrededor del modelo exponencial}} \quad (4.109)$$

Para usar el criterio de razón de verosimilitud es necesario conocer la distribución de  $\lambda$  que depende no sólo de la forma del modelo verdadero sino también del valor de los parámetros desconocidos. Si existiese un conjunto mínimo de estadísticos suficientes de los parámetros un test exacto usaría la distribución condicional de  $\lambda$  dados los valores observados de los estadísticos suficientes. Cox indica que cuando éstos no existen la distribución de  $\lambda$  puede ser determinada condicionada en los estimadores MV pero esta opción será intratable tanto analíticamente como por simulación. El procedimiento consiste en encontrar la distribución de  $\lambda$  suponiendo que los parámetros toman valores iguales a aquellos estimadores de MV obtenidos utilizando los datos. Esto se justifica por la suficiencia asintótica de los estimadores MV. Para cada modelo Cox encuentra la distribución de  $\lambda$  si ese modelo se supone verdadero con parámetro igual a las estimaciones MV obtenidas a partir de los datos, desarrolla expresiones asintóticas para la media y

varianza de  $\lambda$  y establece que se distribuye asintóticamente normal. Estos resultados suponen que la función de verosimilitud es diferenciable respecto a TODOS los parámetros del modelo y por lo tanto no es válido para el modelo de regresión de tres fases, por lo que Williams sugiere usar simulación para indicar el rango de la distribución de  $\lambda$ .

Sean  $\hat{\theta}$  y  $\hat{\psi}$  las estimaciones MV de los parámetros de la regresión bajo ambos modelos y  $\hat{\sigma}_f^2$  y  $\hat{\sigma}_g^2$  las estimaciones de la varianza del error. Suponiendo que el modelo segmentado es el correcto se simula un conjunto de observaciones  $Y_i = f(\hat{\theta}, x) + \varepsilon_i$  donde los  $\varepsilon_i$  son un conjunto de residuos pseudo-normales con varianza  $\hat{\sigma}_f^2$  y para estos datos simulados se ajustan ambos modelos y se calcula  $\lambda$ . Se realiza este procedimiento diez veces usando diferentes conjuntos de residuos obteniéndose  $\lambda_{fk}$ ;  $k=1,2,\dots,10$ . A la distribución de  $\lambda$  cuando el modelo segmentado es correcto se la llama  $\Lambda_f$ . Se repite la misma operación pero ahora suponiendo que el modelo exponencial es el correcto obteniéndose  $\lambda_{gk}$ ,  $k=1,2,\dots,10$ . A la distribución de  $\lambda$  cuando el modelo exponencial se supone correcto se la llama  $\Lambda_g$ .

Con lo que se han efectuado diez observaciones de la distribución  $\Lambda_f$  y diez de la distribución  $\Lambda_g$ . El valor de  $\lambda$  calculado a partir de los datos,  $\lambda_o$ , debe ser asignado a una de las dos distribuciones.

Sean  $m_f, m_g, s_f$  y  $s_g$  la media y desviación de  $\Lambda_f$  y  $\Lambda_g$  con  $m_f$  menor que  $m_g$ ;  $d_f = \max \{m_f + 2s_f, \max \lambda_{fk}\}$  y  $d_g = \min \{m_g - 2s_g, \min \lambda_{gk}\}$ .

La regla de decisión dada por el autor es la siguiente:

- Si  $\lambda_o < d_f$  y  $\lambda_o < d_g$  entonces el modelo verdadero es el segmentado,
- si  $\lambda_o > d_f$  y  $\lambda_o > d_g$  entonces el modelo verdadero es el exponencial,
- si  $\lambda_o > d_f$  y  $\lambda_o < d_g$  entonces ningún modelo es verdadero y
- si  $\lambda_o < d_f$  y  $\lambda_o > d_g$  entonces no es posible realizar ninguna discriminación entre los modelos.

No puede calcularse la probabilidad de errores de clasificación pues no se conocen las distribuciones de  $\Lambda_f$  y  $\Lambda_g$ . A veces el mejor ajuste del modelo segmentado puede deberse a que tiene más

grados de libertad que el exponencial.

Williams considera, en todos los casos, que los residuos son independientes y homoscedásticos. Sin embargo en muchos casos la varianza del error es mayor para valores mayores de la esperanza de  $Y$  y los residuos están correlacionados; por lo que habrá que determinar si las distribuciones  $\Lambda_g$  y  $\Lambda_f$  son críticamente dependientes de los mismos, pues es éste caso no serían válidas sus sugerencias. El realizó una investigación limitada sobre los efectos de la violación de este supuesto. Eligió un conjunto de parámetros representativo  $\beta$  y usando un conjunto de setenta desvíos pseudo-normales con varianza uno ( $u_i$ ) construyó tres conjuntos de desvíos:

(a) Control:  $\varepsilon_i = k_1 u_i$ ;  $k_1 = 6$

(b) Heteroscedásticos:  $\varepsilon_i = k_2 x_i' \beta u_i$

(c) Correlacionados:  $\varepsilon_i = 0.0333\varepsilon_{i-1} + 0.943k_3 u_i$ ,  $i > 1$ ,  $\varepsilon_1 = k_3 u_1$

$k_2$  y  $k_3$  se eligieron de manera que la SCR del modelo segmentado fuese el mismo para los tres conjuntos de residuos. Para cada conjunto se ajustaron ambos modelos y se calculó el valor  $\lambda$ . Se generaron 40 conjuntos de residuos y se repitió el mismo procedimiento con todos ellos.

La distribución de  $\Lambda_f$  fue similar para los tres tipos de error y la discriminación hecha entre los modelos de regresión resultó válida aunque las suposiciones de independencia y errores homoscedásticos no lo fuesen.

#### 4.6.2 Solución propuesta por Gallant y Fuller(1973)

Los autores proponen un test para discriminar entre modelos alternativos en el caso en que ambos cumplan condiciones de continuidad y diferenciabilidad que permitan ser reparametrizado y aplicar el ajuste de Gauss-Newton modificado. Esto ha sido tratado en la Sección 4.2.6, no en forma general sino mediante ejemplos. Lo mismo sucede con el test propuesto que sólo puede explicarse a través de ejemplos pues cada caso requiere un

tratamiento distinto.

Por ejemplo si la hipótesis a testar es que el modelo es cuadrático-lineal [ $f_1(x, \theta)$ ] versus a que el modelo sea cuadrático-cuadrático-lineal [ $f_2(x, \theta)$ ] donde:

$$f_1(x, \theta) = \theta_1 + \theta_2 x + \theta_3 (\theta_4 - x)^2 I_+(\theta_4 - x)$$

$$f_2(x, \theta) = \theta_1 + \theta_2 x + \theta_3 (\theta_4 - x)^2 I_+(\theta_4 - x) + \theta_5 (\theta_6 - x)^2 I_+(\theta_6 - x)$$

Las hipótesis pueden plantearse como:  $H_0: \theta_5 = 0; \theta_4 = \theta_6$  vs  $H_1: \theta_5 \neq 0; \theta_4 \neq \theta_6$  para  $f_2(x, \theta)$ ; donde los residuos  $\varepsilon_i$  son independientes idénticamente distribuidos  $N(0, \sigma^2)$ .

El test construido por analogía a la teoría de modelos lineales, es

$$t(Y) = \frac{(\frac{1}{2}) \sum_{i=1}^N [(Y_i - f_1(x_i, \hat{\theta}))^2 - (Y_i - f_2(x_i, \hat{\theta}))^2]}{[1/(N-6)] \sum_{i=1}^N [Y_i - f_2(x_i, \hat{\theta})]^2} \quad (4.110)$$

que bajo la hipótesis nula se distribuye como una  $F(2, N-6)$ ; la regla del test es rechazar  $H_0$  si  $t(Y) > F(\alpha, 2, N-6)$  para un nivel de significancia previamente establecido. ( $\hat{\theta}$  es el estimador de  $\theta$  bajo  $H_0$  y  $\hat{\hat{\theta}}$  lo es bajo la hipótesis alternativa).

interpretada como una medida de la discriminación entre ambos regímenes.

5.2 La naturaleza elige uno de los dos regímenes de la regresión con probabilidad  $\lambda$  y  $1-\lambda$  respectivamente.

5.2.1. Solución propuesta por Quandt (1972).

El modelo considerado es:

$$Y_i = x_i' \beta_1 + \varepsilon_{1i} \quad ; \quad \varepsilon_{1i} \sim N(0, \sigma_1^2) \quad (5.4)$$

$$Y_i = x_i' \beta_2 + \varepsilon_{2i} \quad ; \quad \varepsilon_{2i} \sim N(0, \sigma_2^2) \quad (5.5)$$

Se supone que la naturaleza elige a (5.4) con probabilidad  $\lambda$  y a (5.5) con probabilidad  $(1-\lambda)$ , es decir que la probabilidad de que la naturaleza elija uno y otro régimen se mantiene constante para cada observación. Es obvio que  $\lambda$  está comprendido entre cero y uno. Se desconoce qué variables gobiernan el cambio de régimen por lo tanto  $X_0$  no está presente en el modelo. Resumiendo:  $P[E(Y_i) = x_i' \beta_1] = \lambda$  y  $P[E(Y_i) = x_i' \beta_2] = 1-\lambda$ .

Debe estimarse  $\lambda$  y los parámetros de la regresión  $\beta_1, \beta_2, \sigma_1^2$ , y  $\sigma_2^2$ . Para ello Quandt aplica el método de máxima verosimilitud.

Sea  $f(Y_i/x_i)$  la función de densidad condicional de  $Y_i$  dado  $x_i'$

$$f(Y_i/x_i) = \frac{\lambda}{\sqrt{2\pi} \sigma_1} \exp\left\{-\frac{1}{2\sigma_1^2} (Y_i - x_i' \beta_1)^2\right\} + \frac{1-\lambda}{\sqrt{2\pi} \sigma_2} \exp\left\{-\frac{1}{2\sigma_2^2} (Y_i - x_i' \beta_2)^2\right\}, \quad (5.6)$$

con  $\lambda$  comprendido entre 0 y 1;

$$f_y(Y) = \prod_{i=1}^N \left[ \frac{\lambda}{\sqrt{2\pi} \sigma_1} \exp\left\{-\frac{1}{2\sigma_1^2} (Y_i - x_i' \beta_1)^2\right\} + \frac{1-\lambda}{\sqrt{2\pi} \sigma_2} \exp\left\{-\frac{1}{2\sigma_2^2} (Y_i - x_i' \beta_2)^2\right\} \right], \quad (5.7)$$

la función de verosimilitud conjunta de la muestra y

$$L = \sum_{i=1}^N \log \left[ \frac{\lambda}{\sqrt{2\pi} \sigma_1} \exp\left\{-\frac{1}{2\sigma_1^2} (Y_i - x_i' \beta_1)^2\right\} + \frac{1-\lambda}{\sqrt{2\pi} \sigma_2} \exp\left\{-\frac{1}{2\sigma_2^2} (Y_i - x_i' \beta_2)^2\right\} \right], \quad (5.8)$$

el log de la función de verosimilitud. (5.8) debería maximizarse respecto a  $\beta_1, \beta_2, \sigma_1^2, \sigma_2^2$  y  $\lambda$ , para encontrar los estimadores MV. Esto involucra un problema de maximización no lineal. Para lograrlo Quandt aplica el algoritmo de Powell. Para probar el método generó observaciones de los siguientes regímenes:

$$Y_{1i} = 1.0 + 10x_i + \varepsilon_{1i} \quad \varepsilon_{1i} \text{ iid } N(0, \sigma_1^2) \quad (5.9)$$

$$Y_{2i} = 0.5 + 1.5x_i + \varepsilon_{2i} \quad \varepsilon_{2i} \text{ iid } N(0, \sigma_2^2) \quad (5.11)$$

con los  $x_i$  uniformes e idénticos en todas las muestras. Se generaron cinco conjuntos de observaciones variando en cada caso  $N, \sigma_1^2, \sigma_2^2$  y  $\lambda$ . Se detallan a continuación los detalles de cada simulación:

Caso	N	Rango de los valores de x	$\sigma_1^2$	$\sigma_2^2$	$\lambda$	número de repeticiones
1	60	10 a 20	2.0	2.5	0.5	30
2	120	10 a 20	2.0	2.5	0.5	30
3	60	10 a 20	2.0	2.5	0.75	30
4	60	10 a 20	2.0	25.0	0.5	30
5	60	0 a 40	2.0	2.5	0.5	30

Sea  $Y_{2i} - Y_{1i} = 0.5(x_i - 1.0) + \varepsilon_{2i} - \varepsilon_{1i}$ . Para los primeros tres casos existe una alta probabilidad de que  $Y_{2i} - Y_{1i}$  sea suficientemente grande de manera que muchos métodos podrían separar adecuadamente ambas regresiones. Esto no es cierto para lo últimos dos casos pues los puntos de ambos regímenes se superponen y la separación de los mismo no es obvia.

El algoritmo de maximización no convergió el 22% de las veces (en particular el 27%, 23%, 3% y 53% respectivamente para el caso

1, 2, 3, 4 y 5). Esto puede deberse a que se excedió un límite arbitrario en el número de iteraciones permitidas o porque en el punto donde se supone se halla el máximo la matriz de derivadas parciales segundas no es definida negativa. Cualquiera de estas razones sugiere que la función de verosimilitud tiene una región de considerable chatura. Sin embargo la sugerencia de que la indeterminación aparece por la superposición de los puntos de ambos regímenes no es sostenible pues los dos casos con una superposición considerables se comportan radicalmente diferentes a este respecto.

Quandt compara el sesgo medio en la estimación de cada uno de los parámetros del modelo, el error cuadrático y la razón del valor medio de la varianza asintótica de las repeticiones exitosas al error cuadrático medio en cada experimento para cada uno de los cinco casos con los resultados obtenidos si se hubiesen aplicado a los mismos datos el método de Goldfeld y Quandt (5.1.1). La inspección de estos resultados revela que los dos métodos son difícilmente comparables pero aparentemente los resultados favorecerían el método de Goldfeld y Quandt(5.1.1) aunque no en forma notoria en términos del sesgo y del ECM y a este nuevo método propuesto por Quandt respecto a la razón de la media de la varianza asintótica al ECM que converge más rápidamente a uno(para el otro método la convergencia se produce para muestras de tamaño a cientoveiente en cambio para éste se produce para casi todos los casos). Esta razón tiende a uno a medida que el tamaño de la muestra aumenta indefinidamente.

Este método tiene la ventaja de que no se necesita conocimiento previo sobre la variable que determina el cambio de régimen y la desventaja de que no permite identificar observaciones individuales con uno de ambos regímenes sino solamente calcular la probabilidad de que ocurra uno u otro régimen durante el período muestral.

## 5.2.2. Solución propuesta por Kiefer(1978).

Kiefer (1978) trata el mismo modelo que Quandt. Propone un método que provee estimadores eficientes de los parámetros y de los errores estandar asintóticos, lo cual no había podido lograrse hasta ese momento. Demuestra que aunque la función de verosimilitud no es acotada en algunos puntos del límite del espacio paramétrico las ecuaciones de verosimilitud tienen una raíz que es consistente y se distribuye asintóticamente normal. Esto clarifica la inconsistencia entre el descubrimiento empírico de Quandt de que la función de verosimilitud tiene un máximo interno que brinda estimadores y la creencia generalizada de que las técnicas de verosimilitud son inaplicables al problema sin serias restricciones. Kiefer considera en particular el modelo tratado por Quandt(1972) discutido en la Sección 5.2.1 cuyas expresiones son (5.4) y (5.5) con función de densidad de cada observación (5.6) y función de verosimilitud conjunta de la muestra (5.8).

Si  $\hat{\beta}_1$  es elegido de manera que  $y_i = x_i' \hat{\beta}_1$  para algún  $i$ , entonces a medida que  $\hat{\sigma}_1$  tiende a cero (5.6) aumenta sin límite; además como el segundo término de  $f_i$  es distinto de cero para las restantes observaciones, (5.8) es una función no acotada. Esta característica de la función de verosimilitud ha llevado al desarrollo de técnicas distintas a las MV para problemas en que la restricción  $\sigma_1 = k\sigma_2$  no puede ser impuesta a priori, a pesar del descubrimiento de Quandt de que la función de verosimilitud a menudo tiene un máximo interior que brinda estimadores con propiedades muestrales razonables.

Una técnica popular y consistente en el caso de las medias y de las varianzas es el método de los momentos. La extensión de la misma al modelo de regresión no es inmediata ni eficiente para este problema, y además no brinda estimadores de los errores estandar de los coeficientes.

Como ya es sabido el método de MV postula que los estimadores de los parámetros se encuentran como una solución de la ecuación

$\frac{\delta \ln L}{\delta \theta} = 0$ , donde  $L$  es la función de verosimilitud de la muestra.

Cramer muestra bajo qué condiciones existe una única raíz consistente; su prueba es extendida al caso multivariado por Chanda, y en forma más correcta por Tarone y Gruenhage. Kiefer extiende estas condiciones, en los siguientes teoremas al modelo de regresión a trozos expresado mediante las ecuaciones (5.4) y (5.5).

Teorema 1:

Sea  $\theta = (\lambda, \beta_1, \beta_2, \sigma_1, \sigma_2)$  y sea el espacio paramétrico dado por:  $0 \leq \lambda \leq 1$ ,  $-\infty < \beta_{ij} < \infty$  y  $0 \leq \sigma_i < \infty$  con  $i=1,2$  y  $j=1,2,\dots,k$ . [ $\beta'_i = (\beta_{i1}, \dots, \beta_{ik})$ ]. Sea  $\theta_0$  el valor verdadero del parámetro contenido en alguna región cerrada  $\bar{\Omega}$  que no contiene a  $\lambda=0$ ,  $\lambda=1$ ,  $\sigma_1=0$  ó  $\sigma_2=0$ . Dicha región siempre existe pues, por hipótesis, éstos no son los valores verdaderos de los parámetros. Si la v.a.  $Y_i$  se distribuye de acuerdo a la función de densidad (5.6) entonces para  $N$  suficientemente grande existe una única raíz consistente  $\theta_N$  de las ecuaciones de verosimilitud y  $(\sqrt{N}(\theta_N - \theta_0))$  se distribuye asintóticamente normal con media cero y varianza  $I(\theta_0)^{-1}$ ; donde  $I(\theta_0)$  es la matriz de información de Fisher.

[Nota: al tener esta matriz de información como matriz de covarianza el estimado es eficiente]

Demostración:

La prueba consiste en la verificación de las tres condiciones del lema de Chanda.

\*Condición 1:

Para casi todos los  $Y_i$  y para todo  $\theta \in \bar{\Omega}$ ,  $\frac{\delta \ln f}{\delta \theta_r}$ ,  $\frac{\delta^2 \ln f}{\delta \theta_r \delta \theta_s}$  y  $\frac{\delta^3 \ln f}{\delta \theta_r \delta \theta_s \delta \theta_t}$  existe para todo  $r,s,t$ . ( $\theta_r, \theta_s$  y  $\theta_t$  simbolizan cualquier elemento de vector de parámetros  $\theta$ ).

\*Condición 2:

Para casi todos los  $Y_i$  y para todo  $\theta \in \bar{\Omega}$ ,  $\left| \frac{\delta f}{\delta \theta_r} \right| < F_r(Y)$ ,

$$\left| \frac{\delta^2 f}{\delta \theta_r \delta \theta_s} \right| < \langle F_{rs}(Y) \rangle \text{ y } \left| \frac{\delta^3 \ln f}{\delta \theta_r \delta \theta_s \delta \theta_t} \right| < \langle H_{rst}(Y) \rangle; \text{ donde } H \text{ es tal que}$$

$H_{rst}(Y) \int_{-\infty}^{\infty} f dY \leq M$  y  $F_r(Y)$  y  $F_{rs}(Y)$  son acotadas para todo  $r, s, t$ .

\*Condición 3:

Para todo  $\theta \in \bar{\Omega}$  la matriz  $I(\theta) = \int_{-\infty}^{\infty} \left[ \frac{\delta \ln f}{\delta \theta} \right] \left[ \frac{\delta \ln f}{\delta \theta} \right]' f dy$  es

definida positiva.

La verificación de éstas se halla en el apéndice del trabajo de Kiefer. El teorema no indica cuál raíz, en el caso que haya más de una, es consistente. Para iniciar el proceso de maximización se necesita un estimador inicial consistente.

Este punto inicial puede ser obtenido por el método de la función generatriz de momentos (fgm) sugerido por Quandt ["Estimating Mixtures of Normal Distributions and Switching Regression", Research Memorandum N° 190, Econometric Research Program, Princeton, 1975], que está basado en la relación:

$$E(e^{tY_i}) = \lambda \exp\left\{ t x_i' \beta_1 + t^2 \frac{\sigma_1^2}{2} \right\} + (1-\lambda) \exp\left\{ t x_i' \beta_2 + t^2 \frac{\sigma_2^2}{2} \right\} \quad (5.11)$$

$$= m(t, x, \theta)$$

Para cada  $i$  y para un  $t$  dado las v.a.  $e^{tY_i}$  son independientes cada uno con media  $m(t, x_i, \theta)$ . Por la ley fuerte de los grandes números la cantidad  $\left(\frac{1}{N}\right) \sum_{i=1}^N e^{tY_i}$  converge fuertemente a  $\left(\frac{1}{N}\right) \sum_{i=1}^N m(t, x_i, \theta)$ . La técnica de estimación sugerida es elegir un conjunto de valores para  $t$ , por ejemplo  $(t_1, t_2, \dots, t_k)$ , de algún intervalo alrededor de cero y minimizar

$$S_N = \sum_{j=1}^k \left[ \frac{1}{N} \sum_{t=1}^N e^{z_j^t y_t} - \frac{1}{N} \sum_{t=1}^N m(t, x_t, \theta) \right]^2 \quad (5.12)$$

La consistencia de los estimadores obtenidos minimizando  $S_N$  puede ser mostrada como un resultado del siguiente lema de Rockafellar.

Lema: Sea  $h$  una función convexa propia cerrada que alcanza su ínfimo en un único punto  $x$ . Si  $x_1, x_2, \dots$  es una secuencia de vectores tal que  $h(x_1), h(x_2), \dots$  converge al  $\inf h$ , entonces  $x_1, x_2, \dots$  converge a  $x$ .

Teorema 2:

El estimador  $\hat{\theta}_N$  obtenido minimizando  $S_N$  es consistente para  $\theta_0$ .

Demostración:

Si se expresa a  $S_N$  como  $S_N = \sum_{j=1}^p [z_j - h_j(\theta)]^2$ , la derivada segunda es:

$$\frac{\delta^2 S_N}{\delta \theta \delta \theta'} = -2 \sum_{j=1}^p [z_j - h_j(\theta)] \frac{\delta^2 h_j}{\delta \theta \delta \theta'} + 2 \sum_{j=1}^k \left( \frac{\delta h_j}{\delta \theta} \right) \left( \frac{\delta h_j}{\delta \theta'} \right) \quad (5.13)$$

Por la ley fuerte de los grandes números y por ser acotada  $\frac{\delta^2 h}{\delta \theta \delta \theta'}$ , el primer término de (5.13) converge fuertemente a cero en  $\theta = \theta_0$ ; esto es existe un  $N$  tal que para todo  $N^* > N$ ,  $S_{N^*}$  es una función convexa en la vecindad de  $\theta_0$  (bajo el supuesto de que  $p \geq 2k+3$ ). Como  $S_N$  converge a cero (ley fuerte de los grandes números) y satisface las condiciones del lema,  $\hat{\theta}_N$  converge en  $\theta_0$ . Para el número de parámetros que generalmente se encuentran en la práctica el problema de minimización es solucionable por los programas de optimización no lineal comunes. Los estimadores calculados con la técnica de la función generatriz de momentos pueden ser usados como puntos iniciales de un Newton step hacia el máximo local del logaritmo de la función de verosimilitud.

El procedimiento iterativo de Newton es definido por:

$$\theta^{i+1} = \theta^i - \left( \frac{\delta^2 \ln L(\theta^i)}{\delta \theta \delta \theta'} \right)^{-1} \left( \frac{\delta \ln L(\theta^i)}{\delta \theta} \right). \quad (5.14)$$

Un paso provee estimadores asintóticamente eficientes (en el sentido que siguen siendo consistentes y su matriz de Var-Cov es la de información de Fisher). El opuesto de la inversa de la matriz de derivadas segundas del  $\ln L$  es un estimador consistente de la matriz asintótica de Var-Cov de los parámetros estimados.

En resumen podemos decir que Kiefer fundamenta el método de Quandt y sugiere el uso de los estimadores de la función de generatriz de momentos como puntos iniciales del método iterativo. El autor no determina si así se logrará la convergencia del método el 100% de las veces, ni si la iteración provee mejores estimadores en el caso de muestras pequeñas.

5.3. La naturaleza elige uno de los dos regímenes con una probabilidad que depende del estado anterior del sistema.

(proceso de Cadena de Markov).

5.3.1. Solución propuesta por Goldfeld y Quandt (1972)

En este trabajo Goldfeld y Quandt se refieren al modelo de regresión de dos fases, ya descrito anteriormente en (5.4) y (5.5), pero con la particularidad de que el sistema puede pasar de un régimen a otro cada vez que se genera una nueva observación con una probabilidad que dependerá del estado anterior del sistema. Esta interpretación convierte al proceso de cambio de régimen en una cadena de Markov. En este caso además de estimarse los parámetros del modelo  $\beta_1, \beta_2, \sigma_1^2$  y  $\sigma_2^2$  se debe estimar una matriz de transición del sistema  $\pi$ .

Sea  $\lambda'_0 = (\lambda_{10}, \lambda_{20})$  la probabilidad de que el sistema esté inicialmente en uno de los dos estados con  $\lambda_{20} = 1 - \lambda_{10}$  y  $\lambda'_i$  la probabilidad de que el sistema esté en el estado uno ó dos en la  $i$ -ésima observación, con  $\lambda'_i = \lambda'_{i-1} T = \lambda'_0 T^i = (\lambda_{1i} \lambda_{2i})$ ; con lo que la densidad condicional de  $Y_i$  dado  $x'_i$  es:

$$f(y_i/x_i) = \lambda_{1i} f_1(y_i/x_i) + (1 - \lambda_{1i}) f_2(y_i/x_i)$$

$$\begin{aligned} &= \frac{\lambda_{1i}}{\sqrt{2\pi} \sigma_1} \exp\left[-\frac{1}{2\sigma_1^2} (y_i - x'_i \beta_1)^2\right] + \frac{1 - \lambda_{1i}}{\sqrt{2\pi} \sigma_2} \exp\left[-\frac{1}{2\sigma_2^2} (y_i - x'_i \beta_2)^2\right] \\ &= \lambda'_{1i} f_i \end{aligned} \quad (5.15)$$

donde  $f'_i = (f_1(Y_i/x_i), f_2(Y_i/x_i))$ . El logaritmo de la función de verosimilitud conjunta de la muestra es:

$$L = \sum_{i=1}^N \log(\lambda'_{1i} \cdot f_i) \quad (5.16)$$

Sea

$$T = \begin{pmatrix} \tau_1 & 1-\tau_1 \\ 1-\tau_2 & \tau_2 \end{pmatrix} = \begin{pmatrix} t'_{11} & t'_{12} \\ t'_{21} & t'_{22} \end{pmatrix} \quad (5.17)$$

la matriz de probabilidades de transición para las condiciones iniciales. ( $t'_{sj}$  es la probabilidad de que el sistema cambie del estado  $s$  al estado  $j$ ,  $s, j=1,2$ ) y

$$T^i = \begin{pmatrix} t'_{sj} \end{pmatrix} = T^i = \begin{pmatrix} t_{11}^i & 1-t_{11}^i \\ t_{21}^i & 1-t_{21}^i \end{pmatrix} \quad (5.18)$$

la matriz de transición en la  $i$ -ésima observación, con  $i$  que varía entre 1 y  $n$ , donde  $t_{11}^i = \tau_1 t_{11}^{i-1} + (1-\tau_1) t_{21}^{i-1}$  y  $t_{21}^i = (1-\tau_2) t_{11}^{i-1} + \tau_2 t_{21}^{i-1}$ .

La solución general de este sistema en diferencias es:

$$\begin{pmatrix} t_{11}^i \\ t_{21}^i \end{pmatrix} = \frac{(\tau_1 + \tau_2 - 1)^i}{\tau_1 + \tau_2 - 2} \begin{pmatrix} \tau_1 - 1 \\ 1 - \tau_2 \end{pmatrix} - \frac{1}{\tau_1 + \tau_2 - 2} \begin{pmatrix} 1 - \tau_2 \\ 1 - \tau_2 \end{pmatrix}$$

$$\lambda'_i = (\lambda_{11} \lambda_{21}) = (\lambda_{10} t_{11}^i + \lambda_{20} t_{21}^i ; \lambda_{10} t_{12}^i + \lambda_{20} t_{22}^i) \quad (5.19)$$

Si se reemplaza (5.19) en (5.16)

$$L = \sum_{i=1}^N \log (\lambda'_i f_i) \quad (5.20)$$

que es función de  $\beta_1, \beta_2, \sigma_2^2, \lambda_{10}, \tau_1$  y  $\tau_2$ . Derivado  $L$  parcialmente respecto a cada uno de los parámetros se pueden encontrar los estimadores MV de los mismos. Este método recibe el nombre de método  $\tau$ .

Goldfeld y Quandt hacen referencia a un caso especial en que la función de transición es función de una variable  $Z$  que no pertenece al modelo. En este caso la matriz de transición es:

$$T_i = \begin{pmatrix} \tau_1(z_i) & 1-\tau_1(z_i) \\ 1-\tau_2(z_i) & \tau_2(z_i) \end{pmatrix} \quad (5.21)$$

donde:  $\lambda'_i = \lambda'_{0j=1} \prod T_j$ .

Si por ejemplo se supone, por razones teóricas, que valores grandes de  $Z$  están asociados con probabilidades altas de pasar al primer estado la función se puede definir de la siguiente forma:

$$\tau_1(z_i) = \frac{1}{\sqrt{2\pi} \sigma_1} \int_{-\infty}^{z_i} \exp\left[-\frac{1}{2} \left(\frac{\xi - z_{01}}{\sigma_1}\right)^2\right] d\xi \quad (5.22)$$

$$\tau_2(z_i) = \frac{1}{\sqrt{2\pi} \sigma_2} \int_{z_i}^{\infty} \exp\left[-\frac{1}{2} \left(\frac{\xi - z_{02}}{\sigma_2}\right)^2\right] d\xi \quad (5.23)$$

La función de verosimilitud puede ser derivada y maximizada respecto a  $\beta_1, \beta_2, \sigma_1^2, \sigma_2^2, z_{01}, z_{02}, \lambda_{10}$ . Este método se conoce como Método de  $\tau(z)$ .

Ninguno de los métodos citados permite que la elección de un régimen dependa del registro histórico en el sentido de que la selección de los regímenes no está influenciado por el número de veces que un determinado régimen ha prevalecido ni por el patrón temporal de la elección de regímenes. Existen muchas aplicaciones económicas donde dicha dependencia inercial puede ser deseable. Una de las formas de lograr ésto es definiendo  $\lambda_i$  como  $\lambda_i = \rho(1 - \delta_{i-1}) + (1 - \rho)\lambda_{i-1}$ .  $\delta_{i-1}$  es igual a cero si prevaleció el primer régimen en el estado anterior e igual a uno si el que prevaleció fue el segundo régimen. Si  $\rho=0$  este método coincide

con el método  $\lambda$ , y si  $\rho=1$  el sistema permanece indefinidamente en el régimen que estaba en el período inicial. Para valores intermedios, mientras mayor sea el valor de  $\rho$  más tenderá  $\lambda_i$  a reflejar el estado del sistema en el período anterior. Los  $\lambda_i$  pueden formar  $2^{i-1}$  valores distintos, cada uno expresable en términos de  $\lambda_0, \delta_0$  y  $\rho$ .

Otro modelo que citan los autores es uno que permite cambios entre estados pero donde haya persistencia de los mismos a través del tiempo. Por ejemplo en algún período el sistema obedece al primer régimen, en otros al segundo y en otros obedece a un híbrido de transición entre ambos regímenes. Si se reemplazan los siguientes valores en la matriz diagonal D descrita en 5.1.1

$$\gamma(z_i) = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{z_i} \exp \left\{ -\frac{1}{2} \left[ \frac{\xi - z_0}{\sigma} \right]^2 \right\} d\xi \quad (5.24)$$

$d_i = \rho d_{i-1} + (1-\rho)\xi(z_i)$ , donde  $\rho$  está comprendido entre cero y uno y la condición inicial es  $d_1 = \gamma(z_1)$ .  $d_i$  puede expresarse como:

$$d_i = \rho^{i-1} \gamma(z_1) + (1-\rho) \sum_{j=2}^i \rho^{i-j} \gamma(z_j) \quad (5.25)$$

y la función de versosimilitud se expresa como una función de los parámetros desconocidos  $\delta, Z_0$  y  $\sigma$ . En el caso extremo en que  $\rho=0$  no existe persistencia temporal en el proceso de cambio y el grado de hibridización sólo depende de  $Z$ , mientras que en todos los otros casos depende del valor de  $\rho$ .

## 6. CONCLUSIONES.

### 6.1. Resumen.

En los últimos tiempos ha crecido el interés por el modelo de regresión a trozos, dado que existen muchas situaciones reales donde la relación entre las variables no es estable sino que adopta distintas formas analíticas en distintas regiones del dominio. En estos casos no es posible aplicar los métodos de estimación y de tests de hipótesis del modelo de regresión, pues existen ahora nuevos parámetros a estimar que son los puntos donde la relación cambia de forma. A éstos se los denomina puntos de cambio de régimen o "change-values". El cambio de régimen puede estar determinado por una variable conocida, por una combinación lineal de variables conocidas o por una variable desconocida.

Por tratarse de un tema cuyo estudio se ha iniciado alrededor de los años 60 y cuya aplicación práctica se difundió más ampliamente en la última década, los aportes realizados se encuentran publicados en revistas especializadas y tesis doctorales. Por esta razón un usuario que tenga un problema concreto que resolver, deberá hacer una búsqueda bibliográfica trabajosa y lenta, con el inconveniente adicional de que cada uno de los artículos trata un caso particular, sin hacer referencia a un modelo general, lo que hace difícil la integración y comparación de los distintos aportes. Inclusive la notación utilizada es distinta en cada caso.

El objetivo de este trabajo es elaborar un resumen del estado actual del conocimiento sobre el modelo de regresión a trozos, en forma tal que cualquier persona pueda usar rápida y fácilmente los resultados que se han presentado en la bibliografía.

Para lograr esta presentación unificada se debió proponer una definición general del modelo de regresión a trozos que permita considerar a todos los casos tratados en la literatura como casos particulares de él. Al unificar la notación de los

trabajos estudiados pudimos elaborar un paquete de resultados de fácil acceso para los usuarios del modelo de regresión a trozos. Para ello se agregó una variable explicativa al modelo que es la que rige los cambios de régimen, cuyo coeficiente es cero si no pertenece al modelo y deberá ser estimado en el caso que sí pertenezca (Capítulo II).

En segundo lugar los artículos estudiados se ordenaron en tres grandes grupos de acuerdo al siguiente criterio:

1) que la variable que gobierna el cambio de régimen sea una variable conocida que no pertenezca al modelo (Capítulo III);

2) que la variable que gobierna el cambio de régimen sea una variable conocida que pertenezca al modelo (Capítulo IV);

3) que la variable que gobierna el cambio de régimen sea una combinación lineal de variables conocidas que pertenezcan o no al modelo, o sea una variable desconocida, que llamaremos "Casos Atípicos" (Capítulo V).

Cada capítulo se ha dividido en las siguientes secciones, de acuerdo al tema tratado:

- a) Estimación del punto de cambio de régimen.
  - b) Estimación de los parámetros de la regresión.
  - c) Estimación del número de fases de la regresión.
  - d) Tests de hipótesis sobre la ubicación del punto de cambio de régimen.
  - e) Tests de hipótesis sobre los parámetros de la regresión.
  - f) Tests de hipótesis sobre el número de fases de la regresión.
- (No todos los temas han sido tratados en todos los capítulos).

De esta manera creemos que hemos logrado simplificar notablemente la búsqueda de soluciones para los problemas particulares que pueda tener cualquier lector.

Finalmente el hecho de que el punto de cambio de régimen sea un parámetro desconocido hace complicado el encontrar la distribución exacta o asintótica de los estimadores de los parámetros y de los estadísticos para testar distintas hipótesis. Sin embargo la mayor parte de los autores consultados extienden los resultados clásicos del modelo de regresión para el caso de regresión a trozos o proponen estimadores afirmando que poseen

cierta distribución, sin bases teóricas o experimentales. Esto nos llevó a considerar detenidamente cada resultado y realizar para cada estimador o test propuesto las críticas que consideramos pertinentes sobre las limitaciones en su uso.

## 6.2. Temas de estudio para una etapa posterior.

Del estudio de los trabajos sobre el modelo de regresión a trozos citados en la Bibliografía se observa que existen muchos problemas de interés que todavía no han sido abordados. El estudio detallado y riguroso de los temas que se enuncian a continuación será un aporte fundamental para extender el uso de este modelo a las situaciones donde describe adecuadamente la relación entre las variables.

- 1) Extender las propuestas al caso no lineal de  $r$  fases (donde el número de fases sea conocido), pues en casi todos los trabajos se hace referencia al modelo lineal de dos fases y sólo en algunos al de tres. Lograr ésto será un aporte importante pues permitirá considerar el caso general de  $r$  fases, donde la relación no sea necesariamente lineal en cada una de ellas.
- 2) Verificar para los distintos tests propuestos el nivel de significación y la potencia para distintas hipótesis alternativas. La distribución de los tests en casi todos los estudios no ha sido derivada teórica ni empíricamente.
- 3) Uno de los problemas más importantes que presenta este modelo es la estimación del punto de cambio de régimen ("change-value"). Muchos autores suponen conocido el intervalo en el cual se produce el cambio de régimen, lo cual no ocurre siempre en la práctica. Un aporte valioso será averiguar las propiedades de los estimadores del modelo cuando este intervalo no se conozca.
- 4) Por otro lado para los casos particulares en que se conoce el intervalo en el que se produce el cambio de régimen, no se ha podido incorporar esta información a la función de verosimilitud para encontrar estimadores máximo verosímiles. La propuesta es usar estimadores mínimos cuadráticos, pero en este caso no se ha logrado conocer la distribución por inconvenientes algebraicos.
- 5) Por último creemos que un aporte importante será generalizar las soluciones al caso de  $r$  fases, con  $r$  desconocido.

## BIBLIOGRAFIA

- BACON, D.W. and WATTS, D.G., "Estimating the transition between two intersecting straight lines", Biometrika, Vol. 58, No.3 (1971) págs 525-534.
- BELLMAN, R. y ROTH, R., "Curve fitting by segmented straight lines", JASA, N 64(1969), págs. 1079-1084.
- BROEMELING, L.D. and CHIN CHOY, J.H., "Detecting structural change in linear models. Commun. Statisti-Theor. Meth., Vol. A10(24), (1981), págs. 2551-2561.
- BROWN, R.L. DURBIN, J. and EVANS, J.M., "Techniques for testing the constancy of regression relationships over time", Journal of the Royal Statistical Society, Vol. 37, N°2(1975), págs. 149.
- CHOW, G.E. "Tests of equality between sets of coefficients in two linear regressions", Econometrica, N 28,3(1960), págs. 591-605.
- FARLEY, J.U. and HINICH, M.J., "A test for a shifting slope coefficient in a linear model", JASA, Vol. 5, N° 331(1970), págs.1320-9.
- FEDER, P.I., "On asymptotic distribution theory in segmented regression problems-identified case. Ann. Statist., Vol. 3, N° 1,(1975), págs. 49-83.
- FEDER, P.I., "The log likelihood ratio in segmented regression", The Annals of Statistics, Vol. 3, No. 1,(1975), págs. 84-97.
- FERREIRA, P.E., "A Bayesian analysis of switching regression model:known number of regimes", JASA, Vol. 70, (1975), págs. 370-4.
- GALLANT, A.R. and FULLER, W.A., "Fitting segmented polynomial regression models whose join points have to be estimated", JASA N° 341(1973), págs. 144-7.
- GOLDFELD, S.M. and QUANDT, R.E., "A Markov model for switching

- reegressions. Journal of Econometrics, Vol. 1(1973), págs.3-16.
- GOLDELF, S.M. and QUANDT, R.E., "Nonlinear Methods in Econometrics", Alemania, North-Holland, 1972.
- GRAYBILL, F.A., "Theory and Application of the Linear Model", Massachusetts, E.E.U.U., Duxbury Press, 1976.
- GURNOW, R.N., "A smooth population response curve based on an abrupt threshold and plateau model for individuals", Biometrics, N 29 (1973), págs. 1-10.
- HAWKINS, D.M., "Testing a Sequence of observations for a shift in location", Vol. 72, N° 357(1977), págs. 180-186.
- HINKLEY, D.V., "On the ratio of two correlated normal random variables", Biometrika, Vol. 56, N 3(1969), págs. 635-9.
- HINKLEY, D.V., "Inference about the intersection in two-phase regression", Biometrika, Vol. 56, N 3(1969), págs. 495-504.
- HINKLEY, D.V., "Inference about the change-point in a sequence of random variables", Biometrika, Vol.57, No 1(1970), págs.1-16.
- HINKLEY, D.V., "Inference in two-phase regression", JASA, Vol.66, No. 336(1971), págs. 736-743.
- HOLBERT, D. and BROEMELING, L., "Bayesian inferences related to shifting sequences and two-phase regression", Commun.-Statist.-Theor. Meth., Vol. A6(3)(1977), págs. 265-275.
- HUDSON, D.J., "Fitting segmented curves whose join points have to be estimated", JASA, N 61(1966), págs. 1097-1129.129.
- JOHNSTON, J., "Econometric Method", 3a.Ed., E.E.U.U., McGraw-Hill, 1984.
- KIEFER, N.M. "Discrete parameter variation: Efficient estimation of a switching regression model", Econometrica, Vol. 46, N° 2(1978), págs. 427-34.
- Mc GEE, V. and CARLETON, W.T., "Piecewise regression", JASA, Vol.

- 65, N° 331(1970), págs. 1109-1124.124.
- MOEN, D.H. and BROEMELING, L.D., "Testing for a change in the regression matrix of a multivariate linear model", Commun. Statist.-Theor. Metho., Vol. 13(12)(1984), págs. 1521-1531.
- MOOD, A.M., GRAYBILL, F.A. and BOES, D.C., "Introduction to the theory of Statistics", 3d. Ed. Nueva York, McGraw-Hill, 1974.
- PAGE, E.S., "A test for a change in a parameter occurring at an unknown point", Biometrika, N 42(1955), págs. 523-7.
- PAGE, E.S., "On problems in which a change in a parameter occurs at an unknown point", Biometrika, N 44(1957), págs. 248-52.
- QUANDT, R.E., "The estimation of the parameters of a linear regression system obeying two separate regimes, JASA, No. 53 (1958), págs. 873-880.
- QUANDT, R.E., "Tests of the hypothesis that a linear regression system obeys two separate regimes, JASA, N 55(1960), págs. 324-330.
- QUANDT, R.E., "A new approach to estimating switching regressions", JASA, Vol. 67, N° 338(1972), págs. 306-310.
- ROBINSON, D.E., "Estimates for the points of intesection of two polynomial regressions", JASA N 59(1964), págs. 214-24.
- SCHECHTMAN, E., "Inference in two-phase regression: a simulation study with nonnormal observations", J. Statist. Comput. Simul., Vol. 17 (1983), págs. 223-229.
- SCHWEDER, T., "Some optimal methods to detect structural shift or outliers in regression", JASA, Vol. 71(1976), págs. 491-501.
- SPRENT, P., "Some hypothesis concerning two phase regression lines", Biometrics, N 17(1961), págs. 634-45.
- TSURUMI, H. and SHEFLIN, N., "Bayesian tests of a parameter shift under heteroscedasticity: weighted-T vs double-T approaches"

Commun. Statist-Theor. Meth., Vol. 13(8)(1984), págs. 1003-1013.

TSURUMI, H., "A bayesian and maximum likelihood analysis of a gradual swithching regression in a simultaneous variables", Journal of Econometrics, Vol. 19. págs.165-182.

WILLIAMS, D.A., "Discrimination between regression models to determine the pattern of enzyme synthesis in synchronous cell cultures", Biometrics, (March 1970), págs. 23-32.