

ANÁLISIS DEL CAMBIO ESTRUCTURAL EN EL
MODELO DE REGRESIÓN LINEAL

STRUCTURAL CHANGE ANALYSIS IN LINEAR
REGRESSION MODEL

BLANCA ROSA PÉREZ SALVADOR*
MARÍA GUADALUPE GARCÍA SALAZAR†

Received: 13 Feb 2009; Revised: 15 Nov 2009; Accepted: 6 Apr 2010

Resumen

Suponiendo que las observaciones provienen de una distribución normal se obtuvo la distribución de la estadística de prueba de la razón de verosimilitud cuando existe un cambio en los parámetros del modelo en un tiempo desconocido, también se encontró el estimador de máxima verosimilitud del punto de cambio en el tiempo.

Palabras clave: Cambio estructural, punto de cambio, modelos lineales, estimadores, pruebas de máxima verosimilitud.

*Departamento de Matemáticas, Universidad Autónoma Metropolitana, Unidad Iztapalapa. Av San Rafael Atlixco No.186, Col.Vicentina C.P.09340 Del. Iztapalapa, México D.F., México. E-Mail: psbr@xanum.uam.mx

†Egresada de la maestría en Matemáticas Aplicadas e Industriales; Universidad Autónoma Metropolitana, Unidad Iztapalapa; México D.F., México. E-Mail: mggasa@yahoo.com.mx

Abstract

Assuming that the observations are from normal distribution we obtain de distribution of the maximum likelihood ratio test if there is a change in the parameters at an unknown time and we find the maximum likelihood estimators of the time change too.

Keywords: Structural changes, change point, linear model, maximum likelihood estimators, test.

Mathematics Subject Classification: 62E15, 62F03, 62F10.

1 Introducción

El concepto cambio estructural no tiene una definición única puesto que depende del entorno en el cual se utilice. Antonio Hidalgo (2007) [5] en su tesis doctoral muestra las siguientes definiciones:

- “Un cambio en la estructura productiva.”
- “Un proceso mediante el cual nuevas formas de instituciones surgen a partir de un marco institucional.”
- “Un proceso secuencial por medio del cual distintas estructuras económicas (producción, comercio internacional, utilización de los factores, . . .) de un país subdesarrollado se van transformando hasta que el sector industrial desplaza a la agricultura como centro de gravedad de la actividad económica.”
- “La modificación de un sistema de ecuaciones, en el cual se refleja el equilibrio general de un sistema económico.”
- “Aquella alteración o modificación de los parámetros en un modelo de regresión.”

Aunque el significado de este concepto se entiende de manera diferente por distintos autores; las definiciones anteriores buscan un mismo objetivo el cual es: decidir si un cambio ha ocurrido en el periodo de observación y en caso que este se hubiera dado, se buscará determinar cuando ocurrió. Entonces, surge la pregunta de cómo detectar el punto en el cual se presenta el cambio estructural.

Uno de los métodos que se ha utilizado para probar si ha ocurrido un cambio estructural es conocido como la **Prueba Chow**.

En palabras de Alfonso Nóvales Cinca (1997) [8]: “el test Chow se utiliza cuando el investigador sospecha que el modelo al que responde una parte de la prueba es diferente al que sigue el resto de la muestra.” Es decir,

esta prueba parte del hecho de que *se conoce* en qué momento sucede el cambio estructural, es decir, se supone que se puede dividir una muestra de tamaño n en dos submuestras independientes una de la otra de tamaño n_1 y n_2 respectivamente, $n = n_1 + n_2$, en donde el error de ambas submuestras presenta una distribución normal con media cero y varianza σ^2 .

El modelo de regresión asociado a la existencia de cambio estructural es:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i & i = 1, \dots, n_1 \\ y_i &= \beta_0^* + \beta_1^* x_{1j} + \beta_2^* x_{2i} + \cdots + \beta_p^* x_{pi} + \varepsilon_i & i = n_1 + 1, \dots, n \end{aligned}$$

con $\beta_i \neq \beta_i^*$ para alguna $i = 1, 2, 3, \dots, p$, y $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ por lo que las hipótesis de prueba asociadas a este modelo son

$$\begin{aligned} H_0 &: (\beta_0, \beta_1, \beta_2, \dots, \beta_p) = (\beta_0^*, \beta_1^*, \beta_2^*, \dots, \beta_p^*) \\ H_1 &: (\beta_0, \beta_1, \beta_2, \dots, \beta_p) \neq (\beta_0^*, \beta_1^*, \beta_2^*, \dots, \beta_p^*). \end{aligned}$$

La estadística de la prueba Chow está dada por

$$F_c = \frac{(SCE - (SCE_{1n_1} + SCE_{2n_1})) / (p + 1)}{(SCE_{1n_1} + SCE_{2n_1}) / (n - 2(p + 1))} \sim F_{p+1, n-2p-2}$$

y la región crítica o región de rechazo de H_0 es $F_c > F_{\alpha, p+1, n-2p-2}$.

La característica principal de la prueba Chow es que se conoce el momento en que ocurre el posible cambio, pero qué pasa cuando se sospecha que hubo un cambio, pero **se desconoce** en que momento ocurrió. Esto es, se desconoce cual es el valor de n_1 , inclusive se desconoce si n_1 está en la región de observación, $1 \leq n_1 \leq n$; en este caso las hipótesis de prueba son:

$$H_0 : n_1 \geq n \quad \text{contra} \quad H_1 : n_1 < n.$$

Bajo el supuesto que la hipótesis nula es cierta, el modelo seguido por las observaciones en la región de observación es

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i$$

para toda $i = 1, 2, \dots, n$, esto es, no existe un punto de cambio en la región de observación.

Y bajo el supuesto que la hipótesis nula sea falsa, el modelo es

$$Y_i = \begin{cases} \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i & \text{si } i \leq n_1 \\ \beta_0^* + \beta_1^* x_{1j} + \beta_2^* x_{2i} + \cdots + \beta_p^* x_{pi} + \varepsilon_i & \text{si } i > n_1 \end{cases}$$

con $n_1 < n$ y $\beta_j \neq \beta_j^*$ para algún j entre 0 y p , esto significa que se manifestó un cambio en la región de observación.

Este problema ha sido abordado por diferentes autores, los que han publicado diversas estadísticas de prueba, entre las que se pueden mencionar las siguientes:

Sen y Srivastava (1975) [9] propusieron las siguientes tres estadísticas de prueba para el modelo de localización, $Y = \beta_0 + \varepsilon$, y encontraron su potencia para la hipótesis alternativa mediante el método de Monte Carlo,

$$S = \max_{1 \leq r \leq n-1} \frac{\bar{Y}_{n-r} - \bar{Y}_r}{\sqrt{\left(\frac{1}{r} + \frac{1}{n-r}\right) (\sum_{i=1}^r (Y_i - \bar{Y}_r)^2 + \sum_{i=r+1}^n (Y_i - \bar{Y}_{n-r})^2)}}$$

$$P = \frac{\sum_{i=1}^{n-1} i(Y_{i+1} - \bar{Y})}{\sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{N-1}}} \quad P_1 = \frac{\sum_{i=1}^{n-1} i(Y_{i+1} - \bar{Y})}{\sqrt{\frac{\sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2}{2(N-1)}}$$

donde

$$\bar{Y}_r = \frac{1}{r} \sum_{i=1}^r Y_i \quad y \quad \bar{Y}_{n-r} = \frac{1}{n-r} \sum_{i=r+1}^n Y_i.$$

Beckman y Cook (1979) [3] utilizan el estadístico de prueba, $F = \max(F_{n_1})$ para $p+1 \leq n_1 \leq n-p-1$, para el modelo lineal de localización, $Y = \beta_0 + \beta_1 X + \varepsilon$, donde

$$F_{n_1} = \frac{(SCE - (SCE_{1n_1} + SCE_{2n_1})) / (p+1)}{(SCE_{1n_1} + SCE_{2n_1}) / (n - 2(p+1))}.$$

Worsley (1983) [11], para un modelo de regresión lineal con k variables explicativas, bajo las hipótesis, $H_0 : \delta = 0$ contra $H_1 : \delta \neq 0$, donde $\delta = \beta - \beta^*$, proponen la siguiente estadística,

$$F = \max_k \frac{U_k/p}{(Q - U_k)/(n - 2p)},$$

con

$$U_k = \hat{\delta}'_k \Sigma_k^{-1} \hat{\delta}_k, \quad \hat{\delta}_k = \hat{\beta}_k - \hat{\beta}_k^*, \quad \Sigma_k^{-1} = \text{Varianza bajo } H_0$$

$$Q = \text{Suma de Cuadrados del Error bajo } H_0$$

Horvath y Shao (1993) [6] consideran un modelo de regresión lineal con una variable explicativa. Estos autores utilizan la prueba $H_0 : n_1 \geq n$ contra $H_1 : 1 \leq n_1 < n$ y la estadística de prueba que proponen es,

$$T_n = \max_{d \leq k \leq n-d} \left\{ \frac{1}{\hat{\sigma}^2} (\hat{\beta}_k - \hat{\beta}_k^*) H_k^{-1} (\hat{\beta}_k - \hat{\beta}_k^*) \right\}$$

donde $H_k = (X_k^T X_k)^{-1} + (X_k^{T*} X_k^*)^{-1}$, H_k^{-1} es la varianza total del modelo.

Antoch y Hušcová (2001) [1] consideraron el modelo de localización y propusieron la estadística de prueba,

$$T_n = \max_{1 < k < n} \left\{ \sqrt{\frac{n}{k(n-k)s^2}} \left| \sum_{i=1}^k (\bar{X}_{Ri} - \bar{X}_n) \right| \right\}$$

y rechazan la hipótesis nula si se halla en el 5% más alto de los valores de

$$T_{nl}(R) = \max_{1 < k < n} \left\{ \sqrt{\frac{n}{k(n-k)s^2}} \left| \sum_{i=1}^k (\bar{X}_{Ri} - \bar{X}_n) \right| \right\}$$

donde R indica que el cálculo se hace sobre todas las posibles permutaciones de los datos muestrales, estos autores encontraron que

$$P(\sqrt{2 \log \log n T_{n1}(R)} \leq y + 2 \log \log n - \frac{1}{2} \log \log \log n - \frac{1}{2} \log \pi \mid X_1, \dots, X_n) \\ \rightarrow \exp\{-2 \exp\{-y\}\}, \quad \text{casi seguramente.}$$

Antoch, Gregoire y Jarušcová (2004) [2] generalizaron la propuesta de Antoch y Hušcová [1] al modelo lineal con p variables independientes.

Vito Muggeo (2003) [7] abordó el problema de estimación del punto de cambio estructural; él menciona que la técnica de máxima verosimilitud es inaplicable en el problema por el hecho de que el logaritmo de la función de verosimilitud no es diferenciable en el punto de cambio. Este autor consideró un modelo no lineal para modelar la tendencia en todo el intervalo de observación y propuso estimar a n_1 aplicando una simple técnica de linealización.

En este trabajo se formula la misma prueba de hipótesis pero en regresión múltiple, se obtuvo la estadística de prueba por el método del cociente de verosimilitud y se encontró la distribución exacta de esta estadística de prueba. También se encuentra el estimador de máxima verosimilitud del tiempo en que ocurre el cambio, aún cuando la función de verosimilitud no es derivable en el punto de cambio. Se presentan los resultados de un estudio de simulación con el que se explora el sesgo y la eficiencia de este estimador. El trabajo consta de 5 secciones, en la primera se da una introducción, y se presentan algunos resultados publicados sobre el tema por diferentes autores, en la sección 2 se formula la prueba de hipótesis, se obtiene la estadística de prueba y la región crítica, en la sección 3 se obtiene el estimador del tiempo en que ocurre el cambio y mediante el método de Monte Carlo se exploran sus propiedades de

insesgamiento y eficiencia, en la sección 4 se aplica lo desarrollado en un ejemplo con datos económicos y en la sección 5 se da la discusión y las conclusiones.

2 La prueba de hipótesis

2.1 Formulación del problema

Sea Y_i y $X_{i1}, X_{i2}, \dots, X_{ip}$, $i = 1, 2, 3, \dots, n$, una muestra aleatoria de la variable respuesta Y observada en las variables explicativas X_1, X_2, \dots, X_p , las cuales se relacionan mediante el modelo,

$$Y_i = \begin{cases} \beta_0 + \sum_{j=1}^p X_{ij}\beta_j + \varepsilon_i = X_i^T \beta + \varepsilon_i & \text{si } i \leq n_1 \\ \beta_0^* + \sum_{j=1}^p X_{ij}\beta_j^* + \varepsilon_i = X_i^T \beta^* + \varepsilon_i & \text{si } i > n_1 \end{cases} \quad (1)$$

con $\beta_j \neq \beta_j^*$ al menos para una j , $0 \leq j \leq p$. Donde $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. Al número n_1 lo denominaremos “el punto de cambio estructural”.¹

Las hipótesis a probar son:

$$H_0 : n_1 \geq n \quad \text{contra} \quad H_1 : n_1 < n. \quad (2)$$

Teorema 2.1 *Sea $Y_i, X_{i1}, X_{i2}, \dots, X_{ip}$, $i = 1, 2, 3, \dots, n$ una muestra aleatoria que satisface la relación (1), entonces la región crítica para la prueba (2) obtenida mediante el cociente de verosimilitud está dada por:*

$$\min_m \left\{ \frac{SCE_{1m} + SCE_{2m}}{SCE} \right\} \leq \lambda^* \quad (3)$$

donde SCE es la suma de cuadrados del error de la regresión con toda la muestra; SCE_{1m} es la suma de cuadrados del error con los primeros m datos y SCE_{2m} es la suma de cuadrados del error de la regresión con los restantes datos.

Demostración. El cociente de verosimilitud está dado por

$$\frac{\max L_{H_0}}{\max L_{H_1}} \leq \lambda.$$

¹ En realidad el cambio puede haber ocurrido en cualquier punto del intervalo cerrado $[n_1, n_1 + 1]$, pero para efectos de este trabajo colocaremos el cambio en n_1 .

Para el numerador se tiene que ²

$$\begin{aligned}\max L_{H_0} &= \max_{\beta, \sigma} \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta)} \\ &= \frac{(2\pi)^{-n/2} n^{n/2} e^{-n/2}}{[(Y-X\hat{\beta})^T(Y-X\hat{\beta})]^{n/2}}.\end{aligned}$$

Y para el denominador ³

$$\begin{aligned}\max L_{H_1} &= \\ &= \max_{\beta_1, \beta_2, \sigma, m} \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2}[(Y_1-X_1\beta_1)^T(Y_1-X_1\beta_1)+(Y_2-X_2\beta_2)^T(Y_2-X_2\beta_2)]} \\ &= \max_m \frac{(2\pi)^{-n/2} n^{n/2} e^{-n/2}}{[(Y_1-X_1\hat{\beta}_1)^T(Y_1-X_1\hat{\beta}_1)+(Y_2-X_2\hat{\beta}_2)^T(Y_2-X_2\hat{\beta}_2)]^{n/2}}.\end{aligned}$$

De donde se tiene que

$$\begin{aligned}\left(\frac{\max L_{H_0}}{\max L_{H_1}}\right)^{2/n} &= \\ &= \min_m \frac{(Y_1-X_1\hat{\beta}_1)^T(Y_1-X_1\hat{\beta}_1)+(Y_2-X_2\hat{\beta}_2)^T(Y_2-X_2\hat{\beta}_2)}{(Y-X\hat{\beta})^T(Y-X\hat{\beta})} \\ &\leq \lambda^*.\end{aligned}$$

Esta expresión se puede escribir también como,

$$\min_m \left\{ \frac{SCE_{1m} + SCE_{2m}}{SCE} \right\} \leq \lambda^*. \quad \blacksquare$$

Esta expresión solo tiene sentido cuando $p+1 \leq m \leq n-p-1$ debido a que no se puede realizar una regresión lineal con menos observaciones que parámetros.

Para tener la región de rechazo de tamaño α se debe encontrar la función de densidad de la variable aleatoria $\min_m \left\{ \frac{SCE_{1m} + SCE_{2m}}{SCE} \right\}$, esto se hace con el siguiente teorema.

Teorema 2.2 *Bajo el supuesto que la hipótesis nula es cierta, para la estadística de prueba $\min_m \left\{ \frac{SCE_{1m} + SCE_{2m}}{SCE} \right\}$, existe un vector aleatorio $W \sim N(0, \sigma^2 I_{n-p-1})$, tal que*

$$\min_m \left\{ \frac{SCE_{1m} + SCE_{2m}}{SCE} \right\} = 1 - \max_m \left\{ \frac{W^T Q_m W}{W^T W} \right\}$$

² Y es el vector de las observaciones de la variable respuesta, X es la matriz de variables explicativas.

³ Y_1 y Y_2 son vectores de dimensión m y $n-m$ respectivamente, tales que $Y^T = (Y_1^T, Y_2^T)$ y X_1 y X_2 son las matrices de dimensión $m \times (p+1)$ y $(n-m) \times (p+1)$, tales que $X^T = (X_1^T, X_2^T)$.

con

$$Q_m = P^T \left(\begin{array}{c|c} X_1(X_1^T X_1)^{-1} X_1^T & 0 \\ \hline 0 & X_2(X_2^T X_2)^{-1} X_2^T \end{array} \right) P,$$

$$\text{donde } PP^T = I_n - X(X^T X)^{-1} X^T \quad \text{y} \quad P^T P = I_{n-p-1}.$$

Demostración. Si $Y \sim N(X\beta, \sigma^2 I_n)$, se tiene que $SCE/\sigma^2 = Y^T (I_n - X(X^T X)^{-1} X^T) Y / \sigma^2 \sim \chi_{n-p-1}^2$; de la misma manera se tiene que $SCE_{1m}/\sigma^2 = Y^T (I_{m-p-1} - X_1(X_1^T X_1)^{-1} X_1^T) Y / \sigma^2 \sim \chi_{m-p-1}^2$ y $SCE_{2m}/\sigma^2 = Y^T (I_{n-m-p-1} - X_2(X_2^T X_2)^{-1} X_2^T) Y / \sigma^2 \sim \chi_{n-m-p-1}^2$.

Dado que $(I_n - X(X^T X)^{-1} X^T)$ es una matriz idempotente, simétrica y positiva definida de rango $n - p - 1$, entonces existe una matriz P de tamaño $n \times (n - p - 1)$ tal que

$$PP^T = I_n - X(X^T X)^{-1} X^T \quad \text{y} \quad P^T P = I_{n-p-1}$$

por lo que $SCE = Y^T PP^T Y$, donde $W = P^T Y \sim N(P^T X\beta, \sigma^2 P^T P)$, y dado que

$$0 = X^T (I_n - X(X^T X)^{-1} X^T) P = (X^T P)(P^T X) \Rightarrow P^T X = 0,$$

entonces,

$$SCE = W^T W \quad \text{con} \quad W = P^T Y \sim N(0, \sigma^2 I_{n-p-1}).$$

Ahora, obsérvese que

$$\begin{aligned} SCE_{1m} + SCE_{2m} &= \\ &= Y^T \left(\begin{array}{c|c} I_m - X_1(X_1^T X_1)^{-1} X_1^T & 0 \\ \hline 0 & I_{n-m} - X_2(X_2^T X_2)^{-1} X_2^T \end{array} \right) Y \\ &= W^T Z_m W, \end{aligned}$$

entonces, la matriz Z_m satisface la ecuación

$$Z_m = I_{n-p-1} - P^T \left(\begin{array}{c|c} X_1(X_1^T X_1)^{-1} X_1^T & 0 \\ \hline 0 & X_2(X_2^T X_2)^{-1} X_2^T \end{array} \right) P.$$

Por lo que,

$$SCE_{1m} + SCE_{2m} = W^T W - W^T Q_m W$$

donde

$$Q_m = P^T \left(\begin{array}{c|c} X_1(X_1^T X_1)^{-1} X_1^T & 0 \\ \hline 0 & X_2(X_2^T X_2)^{-1} X_2^T \end{array} \right) P.$$

Con lo que se sigue que

$$\min_m \frac{SCE_{1m} + SCE_{2m}}{SCE} = 1 - \max_m \frac{W^T Q_m W}{W^T W}. \quad \blacksquare$$

Para encontrar el valor de λ^* se requiere tener la función de densidad de $\max_m \frac{W^T Q_m W}{W^T W}$. Para encontrar la función de densidad se utiliza el resultado del siguiente teorema.

Teorema 2.3 Sea el vector $V^T = (V_1, \dots, V_{n-p-2})$ tal que $V_i = W_i / \sqrt{\sum_{i=1}^{n-p-1} W_i^2}$, entonces la función de densidad conjunta de V bajo el supuesto que H_0 es cierta, es igual a

$$f_V(v_1, v_2, \dots, v_{n-p-2}) = \begin{cases} \frac{\Gamma((n-p-1)/2)}{2(\pi)^{(n-p-1)/2} (1-v^T v)^{1/2}}, & \text{si } v^T v < 1 \\ 0, & \text{en otro caso.} \end{cases}$$

Demostración. La función de densidad conjunta del vector $W = P^T Y$ bajo el supuesto que la hipótesis nula es cierta es

$$f_W(w_1, w_2, \dots, w_{n-p-1}) = \frac{e^{-w^T w / 2\sigma^2}}{(2\pi)^{(n-p-1)/2} \sigma^{n-p-1}}.$$

Por definición se tiene que para $j = 1, 2, \dots, n-p-2$

$$v_j^2 \sum_{i=1}^{n-p-1} w_i^2 = w_j^2, \quad (4)$$

al resolver esta ecuación para w_j^2 cuando $j = 1$, se obtiene

$$w_1^2 = \frac{v_1^2}{1-v_1^2} \sum_{i=2}^{n-p-1} w_i^2,$$

con lo que se tiene que

$$\sum_{i=1}^{n-p-1} w_i^2 = w_1^2 + \sum_{i=2}^{n-p-1} w_i^2 = \left(\frac{v_1^2}{1-v_1^2} + 1 \right) \sum_{i=2}^{n-p-1} w_i^2 = \frac{1}{1-v_1^2} \sum_{i=2}^{n-p-1} w_i^2.$$

De esta manera, la ecuación cuando $j = 2$ es igual a

$$\frac{v_2^2}{1-v_1^2} \sum_{i=2}^{n-p-1} w_i^2 = w_2^2,$$

y resolviendo para w_2^2 se obtiene

$$w_2^2 = \frac{v_2^2}{1 - v_1^2 - v_2^2} \sum_{i=3}^{n-p-1} w_i^2,$$

donde

$$\begin{aligned} \sum_{i=1}^{n-p-1} w_i^2 &= \frac{1}{1 - v_1^2} \sum_{i=2}^{n-p-1} w_i^2 = \frac{1}{1 - v_1^2} \left(w_2 + \sum_{i=3}^{n-p-1} w_i^2 \right) \\ &= \frac{1}{1 - v_1^2 - v_2^2} \sum_{i=3}^{n-p-1} w_i^2. \end{aligned}$$

Al seguir el procedimiento hasta $j = n - p - 2$, finalmente se obtiene:

$$\sum_{i=1}^{n-p-1} w_i^2 = \frac{w_{n-p-1}^2}{1 - v_1^2 - \dots - v_{n-p-2}^2} = \frac{w_{n-p-1}^2}{1 - v^T v}.$$

Reemplazando este resultado en la ecuación (2.1) se tiene $w_j^2 = \frac{w_{n-p-1}^2}{1 - v^T v} v_j^2$.

El siguiente paso es encontrar el jacobiano de esta transformación. Las derivadas parciales de la transformación son:

$$\begin{aligned} \frac{\partial w_j}{\partial v_j} &= \frac{|w_{n-p-1}|}{\sqrt{1 - v^T v}} + \frac{|w_{n-p-1}| v_j^2}{(1 - v^T v)^{3/2}} & 1 \leq j \leq n - p - 2 \\ \frac{\partial w_j}{\partial v_i} &= \frac{|w_{n-p-1}| v_j v_i}{(1 - v^T v)^{3/2}} & 1 \leq i \neq j \leq n - p - 2 \end{aligned}$$

y el jacobiano es

$$J = \left(\frac{|w_{n-p-1}|}{\sqrt{1 - v^T v}} \right)^{n-p-2} \left| I + \frac{1}{1 - v^T v} v v^T \right|.$$

El determinante de la matriz $I + \frac{1}{1 - v^T v} v v^T$ puede ser obtenido como el producto de sus valores propios. Los vectores propios de la matriz $I + \frac{1}{1 - v^T v} v v^T$ son v y cualquier vector z ortogonal a v . El valor propio asociado a v es $\frac{1}{1 - v^T v}$, ya que

$$\left(I + \frac{1}{1 - v^T v} v v^T \right) v = v + \frac{v^T v}{1 - v^T v} v = \frac{1}{1 - v^T v} v$$

y el valor propio asociado a z ortogonal a v es uno debido a que

$$\left(I + \frac{1}{1 - v^T v} v v^T \right) z = z,$$

por lo tanto

$$\left| I + \frac{1}{1 - v^T v} v v^T \right| = \frac{1}{1 - v^T v} \quad \text{y} \quad J = \frac{|w_{n-p-1}^{n-p-2}|}{(1 - v^T v)^{(n-p)/2}}.$$

Sustituyendo este término en la función de densidad de W se obtiene

$$\begin{aligned} f_{V, W_{n-p-1}}(v_1, v_2, \dots, v_{n-p-2}, w_{n-p-1}) &= \\ &= \frac{|w_{n-p-1}^{n-p-2}| e^{-\frac{w_{n-p-1}^2}{2\sigma^2(1-v^T v)}}}{(2\pi)^{(n-p-1)/2} \sigma^{n-p-1} (1 - v^T v)^{(n-p)/2}}. \end{aligned}$$

Entonces, la función de densidad del vector V es

$$f_V(v_1, v_2, \dots, v_{n-p-2}) = \frac{\int_{-\infty}^{\infty} |w_{n-p-1}^{n-p-2}| e^{-\frac{w_{n-p-1}^2}{2\sigma^2(1-v^T v)}} dw_{n-p-1}}{(2\pi)^{(n-p-1)/2} \sigma^{n-p-1} (1 - v^T v)^{(n-p)/2}}.$$

Esta integral puede resolverse usando el cambio de variable $v = \frac{w_{n-p-1}^2}{2\sigma^2(1-v^T v)}$,

$$\begin{aligned} f_V(v_1, v_2, \dots, v_{n-p-2}) &= \\ &= \frac{2^{(n-p-3)/2}}{(2\pi)^{(n-p-1)/2} (1 - v^T v)^{1/2}} \int_{-\infty}^{\infty} |v|^{((n-p-1)/2)-1} e^{-v} dv \\ &= \frac{1}{2(\pi)^{(n-p-1)/2} (1 - v^T v)^{1/2}} \int_0^{\infty} |v|^{((n-p-1)/2)-1} e^{-v} dv \\ &= \frac{\Gamma((n-p-1)/2)}{2(\pi)^{(n-p-1)/2} (1 - v^T v)^{1/2}}, \end{aligned}$$

Por tanto,

$$f_V(v_1, v_2, \dots, v_{n-p-2}) = \begin{cases} \frac{\Gamma((n-p-1)/2)}{2(\pi)^{(n-p-1)/2} (1-v^T v)^{1/2}}, & \text{si } v^T v < 1 \\ 0, & \text{en otro caso.} \end{cases} \quad \blacksquare$$

Teorema 2.4 *El valor de λ^* en la expresión (2.1) para el nivel de significancia α satisface la ecuación*

$$\begin{aligned} P \left[\left(\frac{\max L_{H_0}}{\max L_{H_1}} \right)^{2/n} < \lambda^* \right] &= \\ &= 1 - \int \cdots \int_{A_{\lambda^*}} f_V(v_1, v_2, \dots, v_{n-p-2}) dv_1 dv_2 \cdots dv_{n-p-2} \\ &= \alpha \end{aligned}$$

con $A_{\lambda^*} = \{v \in \mathbb{R}^{n-p-2} \mid u^T Q_{p+1} u < 1 - \lambda^*, \dots, u^T Q_{n-p-1} u < 1 - \lambda^*\}$, y

$$f_V(v_1, v_2, \dots, v_{n-p-2}) = \begin{cases} \frac{\Gamma((n-p-1)/2)}{2(\pi)^{(n-p-1)/2}(1-v^T v)^{1/2}}, & \text{si } v^T v < 1 \\ 0, & \text{en otro caso.} \end{cases}$$

Demostración. La región crítica $\max_m \frac{W^T Q_m W}{W^T W} \geq 1 - \lambda^*$ es equivalente a $\max_m u^T Q_m u \geq 1 - \lambda^*$ con $u_i = w_i / \|w\|$ para $i = 1, 2, \dots, n-p-2$ y $u_{n-p-1} = \sqrt{1 - u_1^2 - u_2^2 - \dots - u_{n-p-2}^2}$; entonces

$$P\left(\max_m u^T Q_m u \geq 1 - \lambda^*\right) = 1 - P\left(\max_m u^T Q_m u < 1 - \lambda^*\right)$$

y

$$\begin{aligned} P\left(\max_m u^T Q_m u < 1 - \lambda^*\right) &= \\ &= P\left(u^T Q_{p+1} u < 1 - \lambda^*, \dots, u^T Q_{n-p-1} u < 1 - \lambda^*\right) \\ &= \int \cdots \int_{A_{\lambda^*}} f_V(v_1, v_2, \dots, v_{n-p-2}) dv_1 dv_2 \cdots dv_{n-p-2} \\ &= 1 - \alpha \end{aligned}$$

con $A_{\lambda^*} = \{v \in \mathbb{R}^{n-p-2} \mid u^T Q_{p+1} u < 1 - \lambda^*, \dots, u^T Q_{n-p-1} u < 1 - \lambda^*\}$. ■

Debido a que la región de integración es compleja no existe una solución analítica simple, entonces para tener la región crítica se recurre a efectuar una integración numérica, o bien una integración estocástica usando el método de Monte Carlo.

2.2 Determinación de la región crítica mediante simulación

Una manera de obtener la región crítica es mediante una simulación por el método de Monte Carlo. En la simulación el siguiente algoritmo se ejecutó 5000 veces, y en consecuencia se obtuvieron 5000 valores para la variable R .

1. Mediante un generador de números aleatorios se obtiene el vector aleatorio $W \in \mathbb{R}^{n-p-1}$, $W \sim N(0, I\sigma^2)$, con σ conocida.
2. Con el vector W se calculan los valores de la forma cuadrática $u^T Q_m u$, donde $u = W/\sqrt{W^T W}$ y $m = p+1, p+2, \dots, n-p-1$.
3. Se encuentra el máximo del conjunto $\{u^T Q_{p+1} u, u^T Q_{p+2} u, \dots, u^T Q_{n-p-1} u\}$, $R = \max_{\{m=p+1, \dots, n-p-1\}} \{u^T Q_m u\}$.

Al terminar las 5000 corridas, se ordenan en forma ascendente los 5000 valores de R .

Finalmente se elige el valor $R_{0.05}$ tal que el 5% de los 5000 valores de R están por arriba de él y el 95% está por debajo, esto es:

$$\#\{R \mid R < R_{0.05}\} = 0.95 \times 5000 = 4750$$

al final se hace $R_{0.05} = 1 - \lambda^*$.

2.3 Determinación de la región crítica mediante integración numérica.

Para encontrar la región crítica por integración numérica se debe tener en cuenta que $f_V(v) > 0$ sólo cuando $v^T v < 1$ y por lo tanto al definir la partición en el dominio de f_V se debe considerar esta restricción.

Para eficientar los cálculos computacionales, se calculará el valor α_c tal que

$$P\left(\max_m \frac{w^T Q_m w}{w^T w} \geq 1 - \lambda_c^*\right) = \alpha_c,$$

donde $\lambda_c^* = \min_m \frac{SCE_{1m} + SCE_{2m}}{SCE}$, para la muestra que se tiene, en lugar de calcular el valor de λ^* tal que $P(\max_m w^T Q_m w / w^T w \geq 1 - \lambda^*) = \alpha$.

El algoritmo para encontrar el valor de α_c es el siguiente.

1. Encontrar la partición $P_r = \{x_0, x_1, \dots, x_{2r}\}$ del intervalo $[-1, 1]$ tal que $x_0 = -1$, $x_1 = -1 + 1/r$, $x_2 = -1 + 2/r, \dots, x_{2r} = 1$.
2. Formar los hipercubitos $[x_{i_1}, x_{i_1+1}] \times [x_{i_2}, x_{i_2+1}] \times [x_{i_3}, x_{i_3+1}] \times \dots \times [x_{i_{n-p-2}}, x_{i_{n-p-2}+1}]$, cuyo volumen está dado por $\Delta(v) = 1/r^{n-p-2}$.
3. Obtener el vector c_i igual al centro de cada hipercubito y calcular el término $c_i^T c_i$.
4. Si $c_i^T c_i \leq 1$ calcule $\max(c_i) = \max_m \{u(c_i)^T Q_m u(c_i) < 1 - \lambda^*\}$; donde $u(c_i) = \left(c_{i1}, c_{i2}, \dots, c_{i,n-p-2}, \sqrt{1 - c_i^T c_i}\right)$.
5. Si $\max(c_i) \geq 1 - \lambda_c^*$ entonces se suma a la integral $Int = Int + f_V(c_i)\Delta(v)$.

3 Estimación

3.1 Formulación del problema

El punto de cambio estructural es un parámetro que no necesariamente existe, por lo que es diferente a otros parámetros como la media o la

varianza de una variable aleatoria normal que siempre existen, por lo que es conveniente, una vez que se tiene la estimación, hacer una prueba para determinar si en \hat{m} hay evidencia de que ha ocurrido un cambio.

El método de estimación utilizado es el de máxima verosimilitud aún cuando la función de verosimilitud no es diferenciable en el punto de cambio, comenzamos escribiendo la función de verosimilitud

$$\begin{aligned} L(m) &= L(m; Y, X_1, \dots, X_p) \\ &= \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2}[(Y_1 - X_1\beta_1)^T(Y_1 - X_1\beta_1) + (Y_2 - X_2\beta_2)^T(Y_2 - X_2\beta_2)]}. \end{aligned}$$

El punto \hat{m} donde $L(m)$ alcanza su valor máximo es

$$\hat{m} = \{m \mid SCE_{1m} + SCE_{2m} \leq SCE_{1k} + SCE_{2k}, \quad k = p+1, \dots, n-p-1\}.$$

El rango de valores de este estimador está restringido a $p+1 \leq \hat{m} \leq n-p-1$, debido a dos razones técnicas,

1. Si una submuestra tiene menos de $p+1$ datos, estos serían insuficientes para ajustar un modelo de regresión con p variables explicativas.
2. Si $k < p+1$, el teorema (3.1) afirma que $SCE_{1k} + SCE_{2k} \geq SCE_{1(p+1)} + SCE_{2(p+1)}$ y por lo tanto, k nunca será elegido como estimador de m . Caso similar ocurre cuando $k \geq n-p+1$.

Por esta razón la muestra debe contener suficientes datos antes y después del punto de cambio para que éste pueda ser estimado.

Teorema 3.1 Sea $(Y_1, X_1), \dots, (Y_n, X_n)$ una muestra aleatoria con Y_i la variable de interés y X_i el vector de variables explicativas; sean SCE_{1k} y SCE_{2k} como se definieron anteriormente, entonces, cuando $k \leq p+1$ necesariamente $SCE_{1k} + SCE_{2k} \geq SCE_{1(p+1)} + SCE_{2(p+1)}$.

Demostración. Si $k \leq p+1$ se sigue que $SCE_{1k} = 0$, por lo tanto, $SCE_{1k} + SCE_{2k} = SCE_{2k}$ y entonces, para demostrar el teorema basta ver que $SCE_{2k} \geq SCE_{2(p+1)}$ cuando $k < p+1$.

Considere la matriz de variables explicativas X de $n \times (p+1)$ que se descompone en tres submuestras, X_1 de $k \times (p+1)$, X_2 de $(p+1-k) \times (p+1)$ y X_3 de $(n-p-1) \times (p+1)$ tales que $X^T = (X_1^T, X_2^T, X_3^T)$ de esta manera (con $k < p+1$) se tiene que,

$$SCE_{2(p+1)} = Y_3^T (I - X_3(X_3^T X_3)^{-1} X_3^T) Y_3$$

y

$$\begin{aligned} SCE_{2(k)} &= \\ &= [Y_2^T \ Y_3^T] \left(I - \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} ([X_2^T \ X_3^T] \begin{bmatrix} X_2 \\ X_3 \end{bmatrix})^{-1} [X_2^T \ X_3^T] \right) \begin{bmatrix} Y_2 \\ Y_3 \end{bmatrix} \\ &= [Y_2^T \ Y_3^T] \left(I - \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} (X_2^T X_2 + X_3^T X_3)^{-1} [X_2^T \ X_3^T] \right) \begin{bmatrix} Y_2 \\ Y_3 \end{bmatrix}. \end{aligned}$$

Se puede probar que

$$\begin{aligned} (X_2^T X_2 + X_3^T X_3)^{-1} &= \\ &= (X_3^T X_3)^{-1} - (X_3^T X_3)^{-1} X_2^T (I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1} X_2 (X_3^T X_3)^{-1} \end{aligned}$$

y por lo tanto

$$I - \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} (X_2^T X_2 + X_3^T X_3)^{-1} [X_2^T \ X_3^T] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

donde

$$\begin{aligned} A_{11} &= I - X_2 (X_3^T X_3)^{-1} X_2^T \\ &\quad + X_2 (X_3^T X_3)^{-1} X_2^T (I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1} X_2 (X_3^T X_3)^{-1} X_2^T \\ A_{12} &= -X_2 (X_3^T X_3)^{-1} X_3^T \\ &\quad + X_2 (X_3^T X_3)^{-1} X_2^T (I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1} X_2 (X_3^T X_3)^{-1} X_3^T \\ A_{21} &= -X_3 (X_3^T X_3)^{-1} X_2^T \\ &\quad + X_3 (X_3^T X_3)^{-1} X_2^T (I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1} X_2 (X_3^T X_3)^{-1} X_2^T \\ A_{22} &= I - X_3 (X_3^T X_3)^{-1} X_3^T \\ &\quad + X_3 (X_3^T X_3)^{-1} X_2^T (I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1} X_2 (X_3^T X_3)^{-1} X_3^T \end{aligned}$$

y dado que

$$\begin{aligned} X_2 (X_3^T X_3)^{-1} X_2^T (I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1} &= \\ &= (I + X_2 (X_3^T X_3)^{-1} X_2^T - I) (I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1} \\ &= I - (I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1} \end{aligned}$$

entonces

- $A_{11} = (I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1}$
- $A_{12} = -(I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1} X_2 (X_3^T X_3)^{-1} X_3^T$
- $A_{21} = -X_3 (X_3^T X_3)^{-1} X_2^T (I + X_2 (X_3^T X_3)^{-1} X_2^T)^{-1}$.

Entonces, se tiene que,

$$\begin{aligned}
 SCE_{2k} &= [Y_2^T \quad Y_3^T] \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right) \begin{bmatrix} Y_2 \\ Y_3 \end{bmatrix} \\
 &= Y_2^T (I + X_2(X_3^T X_3)^{-1} X_2^T)^{-1} Y_2 \\
 &\quad - 2Y_2^T (I + X_2(X_3^T X_3)^{-1} X_2^T)^{-1} X_2(X_3^T X_3)^{-1} X_3^T Y_3 \\
 &\quad + Y_3^T (I - X_3(X_3^T X_3)^{-1} X_3^T) Y_3 + Y_3^T X_3(X_3^T X_3)^{-1} X_2^T \times \\
 &\quad \times (I + X_2(X_3^T X_3)^{-1} X_2^T)^{-1} X_2(X_3^T X_3)^{-1} X_3^T Y_3 \\
 &= (Y_2 - X_2(X_3^T X_3)^{-1} X_3^T Y_3)^T (I + X_2(X_3^T X_3)^{-1} X_2^T)^{-1} \times \\
 &\quad \times (Y_2 - X_2(X_3^T X_3)^{-1} X_3^T Y_3) + SCE_{p+1}.
 \end{aligned}$$

Con lo que se demuestra que $SCE_{2k} > SCE_{2(p+1)}$. ■

3.2 Estudios de simulación

El estimador del punto de cambio encontrado tiene todas las propiedades de un estimador máximo verosímil. Para sondear posibles propiedades de insesgamiento y eficiencia de este estimador, se efectuaron una serie de simulaciones. En todos los casos se generaron datos de un modelo de regresión con un punto de cambio en m tal que $p + 1 \leq m \leq n - p - 1$. Se utilizaron 4 posibles valores para σ^2 , $\sigma^2 = 0.01, 0.1, 1$ y 10 . En cada caso se realizaron 1000 corridas, encontrándose el mismo número de estimaciones de \hat{m} , finalmente, se calculó el promedio y la varianza de estos 1000 datos obtenidos, que corresponden a las estimaciones de $E(\hat{m})$ y $V(\hat{m})$ y que se presentan en una tabla.

3.2.1 Estudio de simulación 1

Se genera una muestra de 20 observaciones, de la variable de interés Y y la variable explicativa X relacionadas mediante el modelo,

$$Y_i = \begin{cases} 2 - 1.2X_i + \varepsilon_i, & \text{si } i \leq 10 \\ -1 + 0.7X_i + \varepsilon_i, & \text{si } i > 10 \end{cases} \quad \text{con} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

y $m = 10$.

El resultado de la simulación es,

	σ^2			
	0.01	0.1	1	10
$E(\hat{m})$	10	9.63	9.65	9.82
$V(\hat{m})$	0	0.84	18.41	21.12

Como se puede ver, cuando $\sigma^2 = 0.01$ $\hat{m} = 10$ y no tiene sesgo, conforme la varianza aumenta, se presenta un sesgo que no es mayor a 0.5.

3.2.2 Estudio de simulación 2

En este caso la muestra es de 20 observaciones con el modelo

$$Y_i = \begin{cases} 0.6 + 1.4X_i + \varepsilon_i, & \text{si } i \leq 9 \\ 1.24 + 0.5X_i + \varepsilon_i, & \text{si } i > 9 \end{cases} \quad \text{con} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

el punto de cambio estructural se encuentra en $m = 9$.

El resultado de la simulación es,

	σ^2			
	0.01	0.1	1	10
$E(\hat{m})$	9	9.50	10.44	10.28
$V(\hat{m})$	0	7.75	21.15	22.08

Nuevamente, cuando $\sigma^2 = 0.01$ $\hat{m} = 9$ y no tiene sesgo, conforme la varianza aumenta, se presenta un sesgo.

3.2.3 Estudio de simulación 3

Ahora se considera una muestra de 35 observaciones del modelo

$$Y_i = \begin{cases} 0.076 + 0.042X_i + \varepsilon_i, & \text{si } i \leq 15 \\ -1.659 + 0.086X_i + \varepsilon_i, & \text{si } i > 15 \end{cases} \quad \text{con} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

donde $m = 15$.

Los resultados de la simulación son,

	σ^2			
	0.01	0.1	1	10
$E(\hat{m})$	15.09	14.96	16.75	17.06
$V(\hat{m})$	0.08	7.14	79.74	87.63

3.2.4 Estudio de simulación 4

Sea Y una variable explicativa y X una variable respuesta para una muestra de 20 observaciones relacionadas por el modelo,

$$Y_i = \begin{cases} -1 + 1.5X_i + \varepsilon_i, & \text{si } i \leq 11 \\ 3 + 2X_i + \varepsilon_i, & \text{si } i > 11 \end{cases} \quad \text{con} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

donde $m = 11$

Los resultados de la simulación son,

	σ^2			
	0.01	0.1	1	10
$E(\hat{m})$	11	11	11	10.76
$V(\hat{m})$	0	0	0	1.27

Como se observa, al parecer nuestro estimador es sesgado y el sesgo se carga hacia el lado, con respecto a m , que tiene mas datos, esto es, \hat{m} tiende a colocarse en la muestra dividida en dos submuestras por m del lado de la submuestra con más datos.

4 Ejemplo

En la Tabla 1 y la siguiente gráfica se tiene información sobre el ahorro (Y) y el ingreso (X) del Reino Unido para el período de 1946 — 1963. Se desea saber si ha habido un cambio significativo en la función de ahorro durante el período posterior a la Segunda Guerra Mundial 1946 — 1954 y 1955 — 1963. El diagrama de dispersión correspondiente se muestra en la Figura 1.

Año	Ingreso (X)	Ahorro (Y)
1946	8.8	0.36
1947	9.4	0.21
1948	10.0	0.08
1949	10.6	0.20
1950	11.0	0.10
1951	11.9	0.12
1952	12.7	0.41
1953	13.5	0.50
1954	14.3	0.43
1955	15.5	0.59
1956	16.7	0.90
1957	17.7	0.95
1958	18.6	0.82
1959	19.7	1.04
1960	21.1	1.53
1961	22.8	1.94
1962	23.9	1.75
1963	25.2	1.99

Tabla 1: Datos de ahorro e ingreso, Reino Unido, 1946—1963 (millones de libras). Fuente: [4, p. 259].

Para estos datos se encontró el estimador del punto de cambio y el valor de λ^* para probar si existe un cambio en la región de observación. Los resultados obtenidos son:

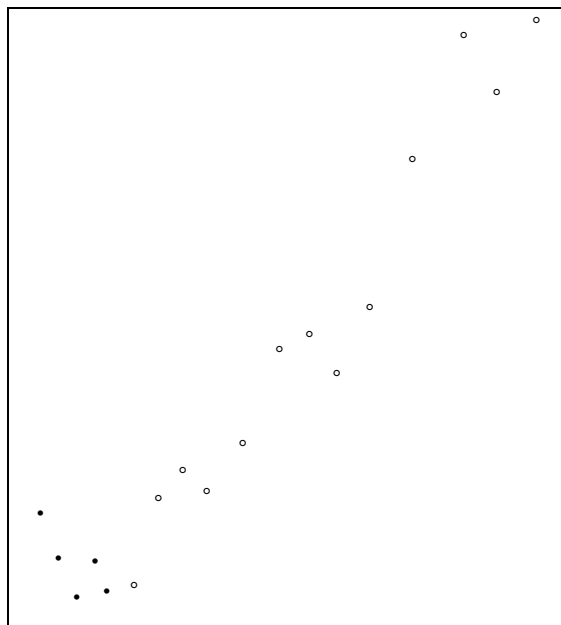


Figura 1: Diagrama de dispersión para los datos de ahorro e ingreso, Reino Unido, 1946—1963 (ver Tabla 1).

\hat{m}	Ingreso (Mill. Lib.)	$\min_m \left\{ \frac{SCE_{1m} + SCE_{2m}}{SCE} \right\}$	$\lambda_{0.05}^*$
5	11	0.4714	0.8300

La estimación del punto de cambio es $\hat{m} = 5$, esto implica que el año donde se presenta el cambio estructural es 1950. En la gráfica de arriba se presentan los primeros 5 datos con circulitos negros, y se ve la tendencia decreciente, el resto de los datos tiene tendencia creciente. Además se observa que

$$0.4714 = \min_m \left\{ \frac{SCE_{1m} + SCE_{2m}}{SCE} \right\} < \lambda^* = 0.83$$

de donde se concluye que, con un nivel de significancia de 0.05 hay suficiente evidencia estadística para afirmar que existe un cambio estructural en el periodo de observación.

5 Conclusiones

En este trabajo se formuló la prueba

$$H_0 : m \geq n \quad \text{contra} \quad H_1 : m < n.$$

con la finalidad de determinar si en el periodo de observación se había manifestado un cambio en los parámetros del modelo. Se encontró la estadística de prueba utilizando el cociente de verosimilitud y se obtuvo analíticamente su función de distribución exacta lo que es una aportación de este trabajo, pues en la literatura se encuentra únicamente la distribución asintótica de la estadística de prueba. De igual manera se obtuvo el estimador de máxima verosimilitud del tiempo en que se produce el cambio y utilizando el método de Monte Carlo se encontró que el estimador encontrado tiene un sesgo pequeño.

Referencias

- [1] Antoch, J.; Hušková, M. (2001) “Permutation tests in change point analysis”, *Statistics and Probability Letters* **53**: 37–46.
- [2] Antoch, J.; Gregoire, G.; Jurušková, D. (2004) “Detection of structural changes in generalized linear models”, *Statistics and Probability Letters* **69**: 315–332.
- [3] Beckman; Cook (1979) “Testing for two-phase Regressions”, *Technometrics* **21**(1).
- [4] Gujarati, D. (1997) *Econometría*. McGraw Hill, México.
- [5] Hidalgo, A.(2007) “El cambio estructural del sistema socioeconómico Costarricense desde una perspectiva compleja y evolutiva (1980–1998)”, <http://www.eumed.net/tesis/alhc/19.htm>
- [6] Horváth, L.; Shao, Q. (1993) “Limit theorems for the union-intersection test”, *Journal of Statistical Planning and Inference* **44**: 133–148.
- [7] Muggeo, V. (2003) “Estimating regression models with unknown break-points”, *Statistics in Medicine* **22**: 3055–3071.
- [8] Nóvalés, A. (1997) *Estadística y Econometría*. McGraw Hill, México.
- [9] Sen, A.; Srivastava, M. (1975) “On tests for detecting change in mean”, *Annals of Statistics* **3**: 98–108.
- [10] Steven, J. (2001) “Inference and estimation in a changepoint regression problem”, *The Statistician* **50**(1): 51–61.
- [11] Worsley. (1983) “Testing for a two-phase multiple regression”, *Technometrics* **25**(1): 35–42.