

TESIS QUE PARA OBTENER EL GRADO DE MAESTRO EN ESTADISTICA
E INVESTIGACION DE OPERACIONES PRESENTA EL ACTUARIO
MANUEL MENDOZA RAMIREZ.

TESIS CON
FALTA DE ORIGEN



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

Prefacio	iv
CAPÍTULO I	
Introducción	
CAPÍTULO II	
Sucesiones de variables aleatorias con puntos de cambio.	
II.1 Resultados del enfoque Frecuentista	
II.2 Resultados del enfoque Bayesiano.	
II.3 Discusión.	
CAPÍTULO III	
Modelos de Regresión con puntos de cambio.	
III.1 Resultados del enfoque Frecuentista	
III.2 Resultados del enfoque Bayesiano	
III.3 Discusión	
CAPÍTULO IV	
Otras contribuciones	
CAPÍTULO V	
Consideraciones finales	
BIBLIOGRAFÍA.	

PREFACIO

El objetivo fundamental de este trabajo es presentar un panorama general del estado que guarda la investigación sobre modelos con puntos de cambio en Estadística. A pesar de que no hace mas de treinta años que este tópico empezó a ser motivo de publicaciones, la inesperada abundancia de contribuciones por una parte y la inaccesibilidad de algunos títulos por otra, me han impedido producir una revisión exhaustiva en un sentido estricto. Sin embargo, de entre los trabajos disponibles he incluido aquellos que en mi opinión, representan las contribuciones mas notables y toda referencia de que he tenido conocimiento se encuentra en la bibliografía.

El contenido esta organizado básicamente en cuatro capítulos. El primero, es muy breve y consiste en una presentación del problema, precedido de algunos comentarios de naturaleza general sobre las características del proceso del modelado estadístico. El segundo capítulo se refiere a los resultados relacionados con el caso de sucesiones de variables. En el capítulo tres se tiene un esquema similar para modelos de regresión. El capítulo final consigna de manera mucho más breve algunos resultados y apli

caciones recientes, así como algunos trabajos que consideran el problema de puntos de cambio para otro tipo de modelos o desde otra perspectiva.

En cada uno de los capítulos 2 y 3 la información se divide en dos secciones: Los resultados clásicos y los correspondientes Bayesianos. Esto se debe, entre otras razones, a que desde mi punto de vista la materia objeto de este trabajo constituye un campo particularmente propicio para ilustrar las características de estos dos enfoques de la Estadística. De cualquier forma, en cada sección las contribuciones incluidas son presentadas y discutidas en estricto orden cronológico de publicación. Espero que esto permita apreciar la evolución en el tratamiento que el problema ha recibido, a lo largo del tiempo.

Como complemento a los breves comentarios con que finaliza el trabajo, cada capítulo incluye una sección de discusión que junto con la tónica en la presentación de las contribuciones, pretende satisfacer el objetivo mencionado originalmente.

Quiero hacer patente mi profundo agradecimiento al M.en C. Joaquín Díaz S., por su cuidadosa revisión de la versión preli-

minar y su constante y decidido apoyo en el transcurso de la prolongada elaboración de este trabajo. Asimismo quiero hacer extensivo mi reconocimiento a los Dres. Alfredo Bustos, Raúl Cavajal, José Luis Farah, Raúl Hudlet, Ignacio Méndez, Federico O'Reilly y Pedro Vit así como al M. en C. Guillermo Baz, profesores míos todos ellos, con quienes me considero en deuda.

Finalmente, quiero dejar constancia del trabajo de mecanografía que fué realizado por Ma. Analia Zaragoza, secretaria del Laboratorio de Estadística de la Facultad de Ciencias, en donde he tenido la inapreciable oportunidad de desempeñarme profesionalmente.

CAPITULO I
Introducción.

Uno de los aspectos más importantes en el trabajo científico y en particular, en la Inferencia Estadística, es la construcción de modelos para describir el comportamiento de los fenómenos bajo estudio. El papel que juegan los modelos en el proceso de adquisición de conocimiento es básico, puesto que constituyen abstracciones, representaciones aproximadas de la realidad, mediante las cuales se pueden analizar la naturaleza del fenómeno y las relaciones que guardan sus componentes, aisladas de otros factores que en la realidad pueden interferir con el estudio.

Es importante tomar en cuenta que todo modelo es una aproximación y por tanto en un sentido estricto, no resulta una réplica exacta de la realidad. Esta consideración resulta particularmente valiosa en el proceso de selección de un modelo. Recurriendo a los conceptos de Box [10], es claro que el mejor modelo no necesariamente es el más complicado, que puede resultar tan difícil de analizar como la realidad misma, si no el que proporcione una descripción simple y razonablemente aproximada del fenómeno en cuestión. De hecho, Box comenta que la elección del modelo apro

piado se reduce a establecer un compromiso entre la simplicidad y la flexibilidad que se pretenda. Esto es, el modelo debe ser simple, de manera que produzca una interpretación clara de los resultados pero a la vez, flexible en el sentido de que pueda ser modificado sin dificultad para aproximar satisfactoriamente una variedad de fenómenos.

En el ámbito de la Inferencia Estadística se hace uso intenso de modelos probabilísticos y estadísticos para describir fenómenos de las más diversas disciplinas: Medicina, Economía, Psicología, Biología, Geografía, etc. Usualmente, esos modelos son estáticos en el sentido de que se considera que no cambian a lo largo del proceso de estudio. Sin embargo, en algunas ocasiones esta suposición no parece razonable porque las condiciones experimentales o algunos factores exógenos se modifican e inducen cambios en el modelo. En otros casos, ocurre que la propia naturaleza del fenómeno limita las posibilidades de un modelo a un rango limitado. En cualquiera de estas situaciones, resulta necesario utilizar modelos que describan ese comportamiento de manera conveniente. Una alternativa es proponer modelos con una

estructura mas compleja pero también existe la posibilidad de utilizar, por segmentos, diferentes modelos que usualmente poseen el mismo grado de dificultad que el originalmente propuesto. Este tipo de modelo recibe el nombre de modelo con puntos de cambio, en donde los puntos de cambio se refieren a los puntos de transición entre los diferentes segmentos. Existen, en la literatura estadística, algunos estudios en donde se efectúan comparaciones de este tipo de alternativas. Un ejemplo es el trabajo de Atkinson [3].

De cualquier forma, en los últimos años ha cobrado importancia, el problema de ajustar modelos con puntos de cambio. La utilidad de este tipo de esfuerzos ha sido establecida en Econometría por un número creciente de autores entre los que destacan Alf y Giacotto [1], Cameron y Nash [17], Hsu [50], Khan [56] y Poirier [70]. En Ciencias Biológicas, por otra parte las contribuciones de Carter y Blight [18], Green y Fekete [41], Knapp et al. [57], Ohki [67], Pool y Borchgrevick [71] y Needham [66] entre otros, ilustran aplicaciones de este tipo de modelos.

Una situación de un cambio en el proceso de observación de un fenómeno puede identificarse de manera general cuando el modelo que describe el comportamiento de las variables que se registran se modifica a partir de algún punto generando, de esta manera, dos conjuntos de observaciones: Z_1, Z_2, \dots, Z_T cuyo comportamiento es descrito por un primer modelo M_1 y $Z_{T+1}, Z_{T+2}, \dots, Z_T$ que se describe con otro modelo M_2 , en donde M_2 es diferente de M_1 . Las diferencias entre los modelos pueden ser de diversa naturaleza, desde cambios en el valor de algunos de los parámetros que los definen hasta el más complicado caso de cambios estructurales.

En cualquier forma, el problema más interesante y de mayor aplicabilidad, aparece cuando se desconoce si efectivamente el cambio tuvo lugar y en ese caso, como lo mencionan Chin Choy y Broemcling [23], existen básicamente dos cuestiones por resolver: Determinar si es razonable suponer que efectivamente existe el cambio y en caso afirmativo, realizar inferencias sobre el punto de cambio, que se considera un parámetro adicional, así como sobre los demás parámetros involucrados en los modelos.

Naturalmente, el problema de modelos con puntos de cambio puede complicarse considerando más de un cambio o imponiendo restric

ciones para que los cambios sean graduales de un modelo a otro. El interés y la relevancia de las diferentes posibilidades, están claramente relacionados con la naturaleza del fenómeno bajo estudio y del objetivo que se persiga: predicción de observaciones futuras, descripción de la relación entre las variables de interés, etc.

En este trabajo se examina una variedad de casos que son tratados con diversos procedimientos y que probablemente permitan, ante una situación particular, elegir el modelo y el tratamiento más adecuado.

CAPITULO II

Sucesiones de variables
aleatorias con puntos de cambio.

II.1 RESULTADOS DEL ENFOQUE FRECUENTISTA

En relación al problema de puntos de cambio en sucesiones de variables aleatorias, se han publicado diversos trabajos con el enfoque frecuentista. De hecho, la mayor parte de las primeras contribuciones en la literatura estadística sobre este tema tratan el problema básicamente desde esa perspectiva.

Entre los primeros autores que abordan el t6pico, se encuentra Page, [68] y [69], quien se plantea el problema de probar la hip6tesis nula de que las observaciones independientes X_1, X_2, \dots, X_T provienen de una misma poblaci6n con funci6n de distribuci6n $F(x; \theta_1)$ contra la alternativa de que las primeras r de ellas ($0 < r < T$) tienen una distribuci6n $F(x; \theta_1)$ pero X_{r+1}, \dots, X_T siguen una distribuci6n $F(x; \theta_2)$ en donde, el 6nico par6metro desconocido es, precisamente, r . En el primero de sus trabajos, [68], el autor sugiere el empleo de una regla de decisi6n basada en sumas acumuladas de las observaciones. La variaci6n a esa propuesta que es discutida y examinada posteriormente con mayor detalle por Page, resume en forma apropiada la contribuci6n de este autor en el tema de este trabajo.

Page desarrolla una solución al problema [69], considerando que en realidad se enfrenta a la discriminación de $T+1$ hipótesis H_0, H_1, \dots, H_T , donde H_i es la hipótesis de que las primeras i observaciones proceden de $F(x; \theta_1)$ y el resto de $F(x; \theta_2)$. Es interesante notar que desde este punto de vista existen dos hipótesis, H_0 y H_T , que implican la inexistencia de cambios puesto que bajo H_0 todas las observaciones son idénticamente distribuidas de acuerdo a $F(x; \theta_2)$ mientras que bajo H_T lo son según $F(x; \theta_1)$. En cualquier forma, Page siguiendo el criterio de discriminación de Rao [77], emplea la siguiente regla de decisión. Se prefiere H_i sobre el resto de las hipótesis si

$$L_i(X_1, \dots, X_T) \geq L_j(X_1, \dots, X_T) \quad \forall j \neq i$$

en donde $L_i(X_1, \dots, X_T)$ es la verosimilitud asociada a la hipótesis H_i . Page demuestra que en el caso particular en que $F(x; \theta)$ es una distribución Normal de varianza conocida y constante y considerando $\theta_1 = \mu$, la media de la distribución y $\theta_2 = \mu + \delta$ en donde, tanto μ como $\delta > 0$ son valores especificados, el criterio

propuesto es equivalente a la siguiente regla:

Se prefiere H_i sobre el resto de las hipótesis si:

$$S_i < S_j \quad \forall j \neq i$$

en donde

$$S_0 \equiv 0, \quad S_k = \sum_{m=1}^k (x_m - \mu - \delta/2)$$

de tal manera que basta con calcular las sumas acumuladas S_i ; $i=0,1,\dots,T$ y elegir la hipótesis que corresponda con el punto mínimo de la sucesión. El procedimiento es también empleado para el caso en que el cambio se produce en la varianza (de σ_1^2 a σ_2^2 con $\sigma_1^2 > \sigma_2^2$) encontrándose un resultado análogo, en donde la diferencia radica en que la suma acumulada por calcular está dada por:

$$S_k = \sum_{m=1}^k \left\{ (x_m - \mu)^2 + 2 \left[\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \right]^{-1} \log \frac{\sigma_1}{\sigma_2} \right\}$$

y de nuevo se elige la hipótesis asociada al valor más pequeño de S_i . En el caso más general, que incluye a los anterior-

res, se deduce que para cualquier distribución con función de densidad del tipo:

$$f(x;\theta) = \exp \{A(\theta) B(x) + C(\theta) + D(x)\}$$

que se conoce como función de la forma Koopman, y empleando el procedimiento indicado, el resultado se puede recuperar.

Esto es, la hipótesis H_i se prefiere si $S_i < S_j \forall j \neq i$ con

$$S_k = \sum_{n=1}^k \left\{ B(x_n) + \frac{\Delta C(\theta)}{\Delta A(\theta)} \right\}; \quad k = 1, 2, \dots, T, \quad S_0 \equiv 0$$

en donde $\Delta C(\theta) = C(\theta_2) - C(\theta_1)$, $\Delta A(\theta) = A(\theta_2) - A(\theta_1)$ siempre que $\Delta A(\theta) > 0$.

En el mismo artículo se menciona que la técnica desarrollada puede considerarse de una sola cola, puesto que la forma de la decisión depende de que $\Delta A(\theta) > 0$, si bien el procedimiento para el caso $\Delta A(\theta) < 0$ resulta simétrico. La prueba de dos colas es también desarrollada en términos análogos. Adicionalmente, se considera el caso en que de antemano se asigne a la hipótesis H_T una probabilidad de ser cierta, C veces mayor que al res

to, en cuyo caso de manera similar al problema de clasificación de observaciones, (Anderson [2]pp.130), el criterio sufre una leve modificación en el sentido de que cuando la hipótesis de interés es H_T , ésta se prefiere aunque no corresponda al valor mínimo de S_x siempre y cuando S_T no exceda el valor $\log c / \Delta A(\theta)$. Finalmente, se puntualiza que desde el punto de vista frecuentista las propiedades del criterio propuesto no pueden ser evaluadas sin dificultad, pero se propone una posible estimación de la función potencia, considerando a la sucesión $\{S_i\}$ como un proceso aleatorio. A través de esta aproximación y empleando un razonamiento frecuentista, Page ilustra la forma de determinar una región de rechazo aproximada de H_T para un nivel de significancia preestablecido. En resumen, se presenta una técnica basada en sumas acumulativas, para tratar el problema, que se aplica en el caso en que los modelos involucrados sean totalmente especificados y permite llevar a cabo, una prueba aproximada de la hipótesis de que no existe cambio alguno. En el caso de que esta hipótesis se rechace y aunque el autor no hace referencia a ello, se puede ob

tener como un producto asociado el estimador de máxima verosimilitud del punto de cambio. Sin embargo, la efectividad del procedimiento no es establecida claramente en los trabajos de Page.

D.V.Hinkley es otro autor que ha publicado diversos artículos en el tema de este trabajo. En particular dos de ellos, [46] y [47], están muy ligados al problema que se ha discutido. En el primero de ellos, se considera el mismo esquema de una sucesión X_1, X_2, \dots, X_T de variables aleatorias independientes y continuas tales que la función de densidad de X_i esta dada por:

$$f(X_i, \theta) = \begin{cases} f(X_i, \theta_0) ; & i=1, 2, \dots, \tau \\ f(X_i, \theta_1) ; & i=\tau+1, \tau+2, \dots, T \end{cases}$$

en donde tanto θ_0 como θ_1 son valores conocidos y $\theta_0 \neq \theta_1$. Así pues, el único parámetro desconocido es τ . Partiendo del supuesto de que necesariamente se verificó un cambio, el problema de interés para este autor es la estimación de τ .

Para este fin procede de acuerdo al método de máxima verosimilitud como sigue. La función de verosimilitud está definida por:

$$L(t) = \prod_{i=1}^t f(X_i; \theta_0) \prod_{i=t+1}^T f(X_i; \theta_1); \quad t=1, 2, \dots, T-1$$

de tal manera que el logaritmo de esta función se expresa de la siguiente manera:

$$\ell(t) = \ln L(t) = \sum_{i=1}^t \ln f(x_i; \theta_0) + \sum_{i=t+1}^T \ln f(x_i; \theta_1); \quad t=1, 2, \dots, T-1$$

y el estimador de máxima verosimilitud de τ , llámese $\hat{\tau}$, es el valor de t que maximiza $L(t)$, o equivalentemente $\ell(t)$.

En virtud de que $\ell(t)$ es una función discreta de t , los métodos usuales de optimización mediante diferenciación no se aplican para encontrar el valor máximo. Esta situación adicionada al hecho de que no se supone una forma específica para $f(x;\theta)$ conduce a que no exista una forma explícita para $\hat{\tau}$. Asimismo, no se satisfacen las condiciones indispensables para garantizar que su distribución asintótica sea Normal como ocurre usualmen-

te con los estimadores de máxima verosimilitud, Kendall y Stuart [54]. Bajo esta perspectiva, Hinkley se concentra en el problema de establecer la distribución de \hat{r} .

El procedimiento que emplea consiste en observar que para que \hat{r} sea igual a un valor t^* es necesario y suficiente que $\ell(t^*) > \ell(t)$ para $t=1, 2, \dots, T-1$. Por su parte $\ell(t)$ es una función de la muestra, que en general, puede escribirse como

$$\begin{aligned} \ell(t) &= \sum_{i=1}^t \ell \text{nf}(x_i; \theta_0) + \sum_{i=t+1}^T \ell \text{nf}(x_i; \theta_1) \\ &= \sum_{i=1}^t \ell \text{nf}(x_i; \theta_0) - \sum_{i=1}^t \ell \text{nf}(x_i; \theta_1) + \sum_{i=1}^t \ell \text{nf}(x_i; \theta_1) + \sum_{i=t+1}^T \ell \text{nf}(x_i; \theta_1) \\ &= \sum_{i=1}^t \{ \ell \text{nf}(x_i; \theta_0) - \ell \text{nf}(x_i; \theta_1) \} + \sum_{i=1}^T \ell \text{nf}(x_i; \theta_1) \\ &= \sum_{i=1}^t u_i + \sum_{i=1}^T \ell \text{nf}(x_i; \theta_1) \end{aligned}$$

Así que el comportamiento de $\ell(t)$ puede describirse a través de la suma de variables aleatorias independientes e idénticamente distribuidas $\sum_{i=1}^t u_i$.

Básicamente, el problema de establecer la distribución de \hat{r} es equivalente al de describir el comportamiento del valor máximo que alcanza el proceso estocástico $\{S_k\}_{k=1, \dots, T-1}$ en donde $S_k = \sum_{i=1}^k u_i$. En esas condiciones y empleando algunos resultados de procesos estocásticos, Hinkley obtiene una aproximación asintótica, cuando r y $T-r$ tienden a infinito, para el cálculo de que $\hat{r} = t$ con $t=1, 2, \dots, T-1$.

Las características sobresalientes que se obtienen mediante este procedimiento son las siguientes: a) la distribución de \hat{r} depende de los valores de θ_0 y θ_1 . Más aún, en el caso específico de sucesiones de variables Normales, se tiene que depende de estos parámetros a través de la función $\Delta = |\theta_0 - \theta_1|$. b) El cálculo de las probabilidades se realiza mediante aproximaciones numéricas a la expresión asintótica que involucra series infinitas. c) \hat{r} no es un estimador consistente en el sentido de que los momentos de su distribución no necesariamente convergen a cero a medida que T tiende a infinito. d) La distribución de \hat{r} para el caso en que θ_0 o θ_1 es desconocido, es la misma siempre que estos valores sean reemplaza-

dos por estimadores de máxima verosimilitud consistentes.

Como un medio adicional, Hinkley muestra que si el interés radica en efectuar pruebas de hipótesis del tipo:

$$H_0: \tau = \tau^* \text{ vs. } H_1: \tau \neq \tau^*$$

el método de cociente de verosimilitudes puede efectuarse con una estadística de la forma:

$$\begin{aligned} \Lambda &= \ell(\hat{\tau}) - \ell(\tau^*) \\ &= \sum_{j=1}^{\hat{\tau}} u_j - \sum_{j=1}^{\tau^*} u_j \end{aligned}$$

en donde H_0 se rechaza si $\Lambda \geq C$.

De esta manera, una vez establecida la distribución de $\hat{\tau}$, es relativamente sencillo calcular la correspondiente a Λ bajo la hipótesis H_0 , siempre que θ_0 y θ_1 sean conocidos. Sin embargo, en el caso de que θ_0 o θ_1 sea desconocido, la distribución de Λ no resulta totalmente especificada y las aproximaciones no son completamente satisfactorias.

En un trabajo posterior [41], el mismo autor trata el problema de efectuar inferencias sobre el punto τ en una sucesión de

variables discretas, en donde el modelo sufre un cambio. Específicamente, considera el caso de variables Bernoulli que cambian de un modelo con probabilidad de éxito θ_0 , a otro con probabilidad θ_1 . La distribución asintótica del estimador del punto de cambio se deriva utilizando un razonamiento análogo al empleado en el artículo previo. Igualmente, se obtiene la distribución asintótica de la estadística correspondiente al cociente de verosimilitudes para probar hipótesis sobre τ . La principal contribución de este trabajo es que extiende los resultados obtenidos para variables continuas al caso de variables Bernoulli. El autor indica que el procedimiento podría ser, en principio, generalizando a cualquier distribución discreta, pero reconoce que el tratamiento de las caminatas aleatorias asociadas se complica notablemente salvo en el caso de variables dicotómicas. Una observación interesante es que en este trabajo Hinkley afirma que debido a la poco usual naturaleza del parámetro τ , el enfoque Bayesiano, para la inferencia sobre el punto de cambio, parece tener poco que ofrecer. La fortuna de esta opinión podrá ser juzgada con la informa-

ción contenida en este trabajo. Particularmente la siguiente sección, donde se presentan algunos procedimientos de origen mas reciente, basados en el enfoque Bayesiano, puede resultar de interés.

11.2 RESULTADOS DEL ENFOQUE BAYESIANO

Los resultados que serán presentados en esta sección se basan en argumentos Bayesianos. Por esta razón es conveniente recordar que el enfoque Bayesiano parte de una interpretación subjetiva de la probabilidad en base a la cual se amplía el espectro de fenómenos que pueden ser descritos mediante modelos probabilísticos. En general, se considera la probabilidad como una medida de la incertidumbre, que bajo condiciones específicas, está asociada a la ocurrencia de un evento. Algunos autores, de manera equivalente, se refieren a la probabilidad como el grado de creencia que se posee con respecto a la ocurrencia de un evento en condiciones determinadas. Esta interpretación, aunada a la formulación de un conjunto de principios de coherencia, considerados esenciales en los procesos de toma de decisiones que incluyen como caso particular a los procesos de Inferencia Estadística, conduce a un tratamiento muy específico que apunta, de acuerdo a Bernardo [6], hacia una teoría general unificada para la solución de los problemas que se plantean en la Estadística Inferencial, evitando las soluciones y técnicas particulares que en ocasiones conducen a inconsistencias o contradicciones.

Una amplia discusión, tanto de los fundamentos como del tratamiento de este enfoque no es el propósito de este texto pero puede encontrarse en Bernardo [6], De Finetti [26],

De Groot [19], Lindley [54] y [60], Ramsey [46] y Savage [82].

En relación al problema de puntos de cambio, entre los primeros artículos que utilizan el argumento Bayesiano, se encuentran los trabajos de Chernoff y Zacks [19] y Kander y Zacks [63], en donde fundamentalmente se parte del problema de estimar la media actual de una población de la cual se tienen observaciones sucesivas, bajo el supuesto de que posiblemente tuvieron lugar cambios a lo largo del proceso de observación en el valor del parámetro. Como un producto asociado se obtiene un criterio, supuestamente Bayesiano, para probar la hipótesis de que no hay cambios contra la alternativa de que un cambio (y solo uno) ha tenido lugar en la sucesión (la misma situación que considera Page). En el primer artículo, Chernoff y Zacks [19], se plantea la situación como sigue:

Sea X_1, X_2, \dots, X_n una sucesión de variables aleatorias normales independientes con medias respectivas $\mu_1, \mu_2, \dots, \mu_n$ y varianza común igual a 1. Cada media μ_i es igual a la anterior salvo en el caso en que haya tenido lugar un cambio. El objetivo fundamental que se propone, es la estimación de μ_n . Con tal propósito, y con la intención de emplear el enfoque Bayesiano, los autores reformulan la situación en los

siguientes términos:

- i) Se asigna a los puntos de cambio una distribución a priori.
- ii) μ_n tiene una distribución a priori Normal de media cero y varianza τ^2 .
- iii) Las magnitudes de los cambios, en caso de que estos tengan lugar, siguen una distribución a priori Normal de media cero y varianza σ^2 .
- iv) Todas las distribuciones a priori son mutuamente in dependientes.

Bajo estas condiciones establecen la siguiente relación entre las medias de las observaciones:

$$\mu_i = \mu_{i+1} + J_i Z_i \quad (i=1, 2, \dots, n-1)$$

en donde J_i toma el valor 1 si un cambio ha tenido lugar entre las observaciones i e $i+1$ y toma el valor cero en caso contrario. Por su parte, Z_i representa la magnitud del cambio para el cual como ya se indicó se supone una distribución a priori Normal de media cero y varianza σ^2 .

Considerando que:

$$X_i = \mu_i + \epsilon_i \quad i=1,2,\dots,n$$

en donde ϵ_i se distribuye Normal de media cero y varianza 1, Chernoff y Zacks mediante la aplicación de la recursión entre las medias obtienen que:

$$X_i = \begin{cases} \mu_n + \epsilon_i + \sum_{k=i}^{n-1} J_k Z_k & ; i=1,2,\dots,n-1 \\ \mu_n + \epsilon_n & ; i=n \end{cases}$$

de tal forma que la distribución condicional de X_i dado el valor de μ_n y el vector $\underline{J} = (J_1, J_2, \dots, J_{n-1})'$ de cambios, es Normal con los siguientes parámetros:

$$E(X_i) = \mu_n; \quad i=1,2,\dots,n$$

$$\text{Var}(X_i) = \begin{cases} 1 + \sigma^2 \sum_{k=i}^{n-1} J_k & ; i=1,2,\dots,n-1 \\ 1 & ; i=n \end{cases}$$

y además:

$$\text{Cov}(X_i, X_j) = \begin{cases} \sigma^2 \sum_{k=i}^{n-1} J_k & ; i < j < n \\ 0 & ; i=n \text{ ó } j=n \end{cases}$$

Esto es, se obtiene la distribución de las observaciones condicional al valor de μ_n y al vector de cambios J pero marginalizada respecto a la distribución a priori de la magnitud de los cambios.

A partir de esta distribución, en el trabajo mencionado se obtiene la distribución a posteriori de μ_n , el parámetro de interés, como una mezcla de distribuciones Normales, cada una correspondiente a un diferente valor del vector J. Suponiendo que una función de pérdida cuadrática fuese adecuada, el estimador de Bayes resulta en ese caso como es usual, la media de la distribución a posteriori. Una versión más sencilla de este estimador se obtiene tomando el límite cuando $\tau^2 \rightarrow \infty$, es decir cuando la varianza a priori de μ_n tiende a infinito y por tanto se emplea una distribución a priori impropia.

Naturalmente, una simplificación adicional muy importante se obtiene al imponer la restricción de que se verifique a lo más un cambio (como en el caso de Page). Bajo estas condiciones la expresión para el estimador es como sigue:

$$\hat{\mu}_n = D^{-1} \sum_{m=0}^{n-1} \frac{P_m}{[n + \sigma^2 m (n-m)]^{1/2}} \cdot \exp \left[\frac{1}{2} \frac{\sigma^2 m^2 (n-m)^2 (\bar{X}_n - \bar{X}_{n-m})^2}{n^2 + \sigma^2 m (n-m) n} \right] \frac{n \bar{X}_n + \sigma^2 m (n-m) \bar{X}_{n-m}}{n + \sigma^2 m (n-m)}$$

en donde:

$$D = \sum_{m=0}^{n-1} \frac{P_m}{[n + \sigma^2 m (n-m)]^{1/2}} \cdot \exp \left[\frac{1}{2} \frac{\sigma^2 m^2 (n-m)^2 (\bar{X}_m - \bar{X}_{n-m})^2}{n^2 + \sigma^2 m (n-m) n} \right]$$

y además:

$$P_0 = P(\underline{J} = \underline{0})$$

$$P_m = P\{J_n = 1, J_m = 0 \quad m' \neq m\} \quad n > 1$$

$$\bar{X}_m = \frac{1}{m} \sum_{i=1}^m X_i$$

$$\bar{X}_{n-m} = \frac{1}{n-m} \sum_{i=m+1}^n X_i$$

Adicionalmente, considerando la distribución de las observaciones condicional al valor de \underline{J} y μ_n marginalizada sobre la distribución de \underline{Z} , se aplica el enfoque frecuentista para calcular el estimador lineal insesgado de varianza mínima de μ_n . Finalmente, los autores proponen un procedimiento específicamente adaptado para ser aplicado en el tipo de proble-

ma que se plantean, esto es, la estimación del valor de la media de la última observación registrada, que consiste en detectar el último cambio ocurrido en la sucesión y estimar el valor de μ_n con las observaciones posteriores a tal punto de cambio, empleando el estimador Bayesiano correspondiente al caso de un cambio como máximo. Para detectar el último punto de cambio, Chernoff y Zacks consideran un proceso secuencial en cuya r -ésima etapa se consideran las últimas $r+1$ observaciones y bajo el supuesto de que a lo más ocurre un cambio se calcula la distribución a posteriori del posible punto de cambio y se decide al respecto a través del valor con mayor probabilidad posterior. En caso de que la decisión sea de no cambio se pasa a la siguiente etapa, mientras que de lo contrario se ubica ese punto de cambio que se considera el último de la sucesión y se procede a la estimación de $\hat{\mu}_n$ según se ha descrito.

El trabajo referido incluye algunas simulaciones en que se comparan, a través del Error Cuadrático Medio y bajo diferentes condiciones, el estimador lineal insesgado de varianza mínima, tres variaciones del estimador Bayesiano para el caso de a lo más un cambio y dos del obtenido mediante el procedimiento secuencial, reportando resultados que por lo demás eran de esperarse, esto es, que ninguno de los estimadores puede considerarse uniforme-

mente superior al resto.

Probablemente la contribución del artículo de Chernoff y Zacks que resulta más relevante para este trabajo es el desarrollo de un procedimiento de prueba para la hipótesis de no cambio contra la alternativa de un cambio en la media de la sucesión de variables aleatorias Normales.

La media común μ_m de las observaciones que anteceden al punto de cambio se considera conocida mientras que el valor al que cambia, μ_{m+1} , se supone tal que $\Delta = \mu_{m+1} - \mu_m$ es una variable aleatoria no negativa, con distribución a priori semi normal, esto es:

$$h_{\Delta}(\delta) = \frac{1}{\sigma} \left(\frac{2}{\pi} \right)^{1/2} \exp \left[- \frac{1}{2\sigma^2} \delta^2 \right]; \delta > 0$$

Adicionalmente, se asigna, a priori, una probabilidad de 1/2 a la hipótesis nula de no cambio mientras que el restante 1/2 se distribuye uniformemente entre los diferentes casos en que se descompone la hipótesis alternativa. Esto es, si H_i es la hipótesis de que el cambio ocurrió inmediatamente después de la i -ésima observación, se tiene que a priori:

$$P(H_i) = \frac{1}{2(n-1)} \quad i = 1, 2, \dots, n-1$$

$$P(H_n) = \frac{1}{2}$$

en donde se tiene que las hipótesis H_1, H_2, \dots, H_n inducen una partición del espacio de los posibles estados de la naturaleza, del tal manera que el complemento de H_n (\bar{H}_n) es la unión de las H_i 's para $i=1, 2, \dots, n-1$. Bajo estas condiciones, el criterio de prueba consiste en comparar las probabilidades a posteriori de H_n y de \bar{H}_n a partir de las siguientes expresiones (para el caso $\mu_n = 0$):

$$P^*(H_n) = \frac{1}{2} f(X_1, \dots, X_n | H_n) \\ = \frac{1}{2} (2\pi)^{-\frac{n}{2}} \exp \left[-\frac{1}{2} \sum_{j=1}^n X_j^2 \right]$$

$$P^*(H_i) = \frac{1}{2(n-1)} f(X_1, \dots, X_n | H_i); i=1, 2, \dots, n-1$$

$$= \frac{1}{2(n-1)} \int_0^{\infty} f(X_1, \dots, X_n | H_i, \Delta = \delta) h(\delta) d\delta$$

$$= \frac{1}{2(n-1)} \int_0^{\infty} (2\pi)^{-\frac{n}{2}} \exp \left[-\frac{1}{2} \left\{ \sum_{j=1}^i X_j^2 + \sum_{j=i+1}^n (X_j - \delta)^2 \right\} \right] h(\delta) d\delta$$

$$= \frac{1}{\sigma} \frac{(2\pi)^{-\frac{n}{2}}}{(n-1)} \exp \left[-\frac{1}{2} \sum_{j=1}^n X_j^2 \right] \exp \left[-\frac{1}{2} \frac{S_n^* - 1}{(n-1) \frac{1}{\sigma^2}} \right] \phi \left(\frac{S_n^* - 1}{(n-1) \frac{1}{\sigma^2}} \right)^{1/2} (n-1) \frac{1}{\sigma^2}^{-1/2}$$

en donde

$$S_{n-1}^* = \sum_{j=i+1}^n X_j$$

y $\Phi(z)$ es la función de distribución Normal (0,1) evaluada en el valor z . A partir de esas relaciones se tiene que:

$$\begin{aligned} P^*(\bar{H}_n) &\propto \sum_{i=1}^{n-1} P^*(H_i) \\ &= \frac{1}{\sigma} \frac{(2\pi)^{-n/2}}{(n-1)} \exp \left[-\frac{1}{2} \sum_{j=1}^n X_j^2 \right] \sum_{i=1}^{n-1} \exp \left[\frac{1}{2} \frac{S_{n-1}^{*2}}{(n-i+1)\sigma^2} \right] \\ &\quad \cdot \Phi \left(\frac{S_{n-1}^*}{(n-i+1)\sigma^2} \right)^{1/2} (n-i+1)^{-1/2} \end{aligned}$$

De tal forma que el procedimiento Bayesiano, considerando una pérdida igual en cada una de las dos posibles decisiones erróneas, consiste en considerar los momios a posteriori:

$$\eta(\bar{H}_n, H_n) = \frac{P^*(\bar{H}_n)}{P^*(H_n)}$$

y aceptar H_n si $\eta(\bar{H}_n, H_n)$ es menor que uno, rechazando esta hipótesis en caso contrario. Sin embargo en este punto los autores mezclan el tratamiento Bayesiano con el frecuentista y manipulan los momios como si se tratase de un cociente

de verosimilitudes en el enfoque clásico y proponen una región de rechazo de la forma:

$$\eta(\bar{H}_n, H_n) > C$$

y concentran sus esfuerzos en encontrar la distribución de $\eta(\bar{H}_n, H_n)$ o de una función monótona aproximada del tal cantidad bajo H_n con el objetivo de fijar C de acuerdo a un nivel de significancia preestablecido.

Es fácil verificar que:

$$\eta(\bar{H}_n, H_n) = \frac{2}{\sigma(n-1)} \sum_{i=1}^{n-1} \exp \left[\frac{1}{2} \cdot \frac{S_{n-1}^*}{(n-1+1)} \right] \phi \left(\frac{S_{n-1}^*}{(n-1+1/\sigma^2)^{1/2}} \right) (n-1 + \frac{1}{\sigma^2})^{-1/2}$$

Los autores consideran una aproximación de esta expresión para el caso en que $\sigma^2 \rightarrow \infty$, esto es, para cuando el factor de escala en la distribución a priori de Δ tiende a cero en cuyo caso $(n-1 + \frac{1}{\sigma^2})^{-1/2} \approx \sigma$, $\exp \left[\frac{1}{2} \cdot \frac{S_{n-1}^*}{(n-1+1)} \right] \approx 1$ y si se desarrolla $\phi \left(\frac{S_{n-1}^*}{(n-1+1/\sigma^2)^{1/2}} \right)^{1/2}$ alrededor de cero se obtiene que:

$$\eta(\bar{H}_n, H_n) \approx \frac{2}{\sigma(n-1)} \sum_{i=1}^{n-1} \sigma \left\{ \frac{1}{2} + \frac{\sigma}{(2\pi)^{1/2}} S_{n-1}^* \right\}$$

de tal manera que el criterio desarrollado es aproximadamente

equivalente a:

$$1 + \frac{2\sigma}{(2\pi)^{1/2} (n-1)} \sum_{i=1}^{n-1} S_{n-i}^* > c$$

que a su vez resulta equivalente a:

$$T(X_1, \dots, X_n) \equiv \sum_{i=1}^{n-1} S_{n-i}^* > k$$

Una expresión más sencilla de la región de rechazo, se obtiene verificando que:

$$\begin{aligned} \sum_{i=1}^{n-1} S_{n-i}^* &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_j \\ &= (X_2 + X_3 + \dots + X_n) \\ &+ (X_3 + \dots + X_n) \\ &\dots \dots \dots \\ &+ (X_n) \\ &= \sum_{i=1}^n (i-1) X_i \end{aligned}$$

De tal forma que, de acuerdo con el razonamiento de los autores, solo es necesario determinar la distribución de:

$$T(X_1 \dots X_n) = \sum_{i=1}^n (i-1) X_i$$

bajo la hipótesis nula de no cambio. Esto es muy simple puesto que en esas condiciones T es una combinación lineal de variables independientes e idénticamente distribuidas según una ley Normal.

Si no se considera la restricción de que μ_n sea cero, mediante un procedimiento análogo es fácil demostrar que la región de rechazo esta definida por:

$$T^*(X_1, \dots, X_n) \equiv \sum_{i=1}^n (X_i - \bar{X}) > C$$

en donde, naturalmente,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

En este caso, también resulta sencillo establecer la distribución de T^* bajo la hipótesis de no cambio, puesto que de nuevo se trata de una combinación lineal de variables aleatorias Normales. Por otro lado, la función potencia es descrita por los autores parametrizada por m , el valor del punto de cambio y δ , la diferencia entre las diferentes medias.

Este procedimiento es comparado, mediante simulación con una

técnica propuesta por Page [48], basada en sumas acumuladas pero que resulta inferior a la que ya se ha presentado en este trabajo y los resultados de la comparación resultan favorables a la técnica de Chernoff y Zacks.

En un trabajo posterior, Kander y Zacks [53], se retoma el el procedimiento de prueba para la hipótesis de no cambio, generalizando el resultado. Se considera una sucesión X_1, X_2, \dots, X_n que tiene una distribución que pertenece a la familia exponencial de un solo parámetro, esto es:

$$f(x_i; \theta) = h(x_i) \exp [\phi_1(\theta) \cup (x_i) \phi_2(\theta)]$$

en donde $\theta \in \Omega$, y $\phi_1(\theta)$ es una función monótona de θ . Adicionalmente se impone la condición de que ambas funciones $\phi_1(\theta)$ y $\phi_2(\theta)$ sean diferenciables respecto a θ en Ω . Las hipótesis bajo consideración son las siguientes:

$$H_0: \theta_1 = \theta_2 \dots = \theta_n \equiv \theta_0 \quad (\theta_0 \text{ conocido})$$

vs

$$H_1: \theta_1 = \dots = \theta_m \equiv \theta_0; \quad \theta_{m+1} = \dots = \theta_n = \theta_0 + \delta$$

en donde δ es mayor que cero pero desconocido, $\theta_0 + \delta \in \Omega$ y m es también desconocido ($m=1, 2, \dots, n-1$).

Las probabilidades apriori para las hipótesis son asignadas de la misma forma que en el trabajo previo. Una diferencia es que en este caso, el parámetro δ que se considera desconocido no se trata en la forma Bayesiana, esto es, no se le asigna una distribución apriori sino que el momio que en analogía al resultado de Chernoff y Zacks, se obtiene en estas condiciones, se trabaja como función de δ para posteriormente emplear una aproximación vía teorema de Taylor de la probabilidad a posteriori de H_1 , para el caso en que $\delta \rightarrow 0$ que produce una función monótona de la estadística $T(X_1, \dots, X_n) = \sum_{i=1}^n (i-1)U(X_i)$. Es importante mencionar, sin embargo, que posteriormente los autores indican que tal resultado también se obtiene si apriori se asigna a δ una distribución exponencial de media $1/t$ y se hace tender t a infinito. De esta forma y de nuevo recurriendo a un razonamiento frecuentista los autores proponen como criterio de rechazo para H_0 :

$$T(X_1, \dots, X_n) > C$$

y proceden a la determinación de la distribución de $T(X_1, \dots, X_n)$ bajo la hipótesis H_0 . En este caso general el problema distri

bucional no es tan simple pero Kander y Zacks demuestran que asintóticamente y bajo condiciones muy generales T se distribuye como una variable Normal. Finalmente, Kander y Zacks tratan los casos particulares de las distribuciones Binomial y Exponencial encontrando que sobre todo en el segundo caso la aproximación Normal es muy pobre aún para tamaños de muestra no excesivamente pequeños. Un par de aproximaciones más a la distribución de T son propuestas, siempre basadas en la Normal, que parecen mejorar los resultados obtenidos, al menos con los dos casos examinados.

En relación al problema de puntos de cambio, uno de los primeros artículos que utiliza el argumento Bayesiano pero que a diferencia de los trabajos de Chernoff y Zacks [14] y Kander y Zacks [15] se aproxima más a la esencia del problema tal como se ha planteado originalmente es el debido a A.F.M. Smith [16] en donde se parte del siguiente esquema:

Sean X_1, \dots, X_T variables aleatorias independientes tales que X_i tiene una distribución $F_1(X|\theta_1)$ para $i=1, 2, \dots, r$ y una distribución $F_2(X|\theta_2)$ para $i=r+1, \dots, T$ ($1 < r < T$). Bajo el supuesto de que ambas distribuciones tienen una forma funcional conocida y que las funciones de densidad correspondientes son $f_1(X|\theta_1)$ y $f_2(X|\theta_2)$ respectivamente, el objetivo principal es la inferencia sobre el valor de r y sobre los valores de θ_1 y

y θ_2 en caso de que estos sean también desconocidos.

Dados los valores de θ_1 , θ_2 y τ , la función de densidad conjunta de las observaciones está dada por la siguiente expresión:

$$f(X_1 \dots X_T | \tau = t, \theta_1, \theta_2) = f_1(X_1 \dots X_t | \theta_1) f_2(X_{t+1}, \dots, X_T | \theta_2)$$

$$= \prod_{i=1}^t f_1(X_i | \theta_1) \prod_{i=t+1}^T f_2(X_i | \theta_2)$$

Ahora bien, Smith considera diversos casos. Primero, suponiendo que tanto θ_1 como θ_2 son valores conocidos y por tanto, el único parámetro que está sujeto al proceso de inferencia es τ , es necesario considerar una función de probabilidad discreta a priori $f_0(t); t=1, \dots, T$ que describa el conocimiento inicial que se posee respecto a los posibles valores de τ . Es importante remarcar que la asignación de la función $f_0(t)$ depende del conocimiento previo sobre los valores de τ , de tal modo que por ejemplo, si la función es constante $f_0(t) = 1/T$ se está asignando igual probabilidad a priori a cada uno de los posibles valores del parámetro. En algunos casos específicos, puede resultar de especial interés la probabilidad asignada al valor $\tau=T$ que representa una situación en que no se verifi

có ningún cambio. De cualquier forma, aplicando el teorema de Bayes a la distribución a priori de τ y la densidad conjunta de las observaciones, se obtiene la distribución a posteriori para τ como sigue:

$$f(t|\underline{X}_T, \theta_1, \theta_2) = \frac{f(X_1 \dots X_T | \tau=t, \theta_1, \theta_2) f_0(t)}{\sum_{t=1}^T f(X_1 \dots X_T | \tau=t, \theta_1, \theta_2) f_0(t)}; \quad t=1, \dots, T$$

en donde $\underline{X}_T = (X_1, \dots, X_T)$

En virtud de que el denominador de esta expresión es una constante que no depende de t , se tiene que:

$$f(t|\underline{X}_T, \theta_1, \theta_2) \propto f(X_1 \dots X_T | \tau=t, \theta_1, \theta_2) f_0(t); \quad t=1, 2, \dots, T$$

esto es,

$$f(t|\underline{X}_T, \theta_1, \theta_2) \propto f_1(X_1 \dots X_T | \theta_1) f_2(X_{t+1}, \dots, X_T | \theta_2) f_0(t); \quad t=1, 2, \dots, T$$

o equivalentemente,

$$f(t|\underline{X}_T, \theta_1, \theta_2) \propto \prod_{i=1}^t f_1(X_i | \theta_1) \prod_{i=t+1}^T f_2(X_i | \theta_2) f_0(t); \quad t=1, 2, \dots, T$$

En el caso en que $f_0(t)$ sea constante, la moda de la distribu

ción a posteriori coincide con el estimador de máxima verosimilitud de r y se obtiene determinando el valor de t que maximice el término del lado derecho en la expresión anterior. Una observación interesante es la siguiente:

$$\begin{aligned} \prod_{i=1}^t f_1(X_i|\theta_1) \prod_{i=t+1}^T f_2(X_i|\theta_2) &= \prod_{i=1}^t f_1(X_i|\theta_1) \prod_{i=t+1}^T f_2(X_i|\theta_2) \prod_{i=t+1}^T f_1(X_i|\theta_1) / \prod_{i=t+1}^T f_1(X_i|\theta_1) \\ &= \prod_{i=1}^T f_1(X_i|\theta_2) \prod_{i=t+1}^T f_1(X_i|\theta_1) / \prod_{i=t+1}^T f_1(X_i|\theta_1) \\ &\propto \prod_{i=t+1}^T f_1(X_i|\theta_1) / \prod_{i=t+1}^T f_1(X_i|\theta_1) \end{aligned}$$

de tal manera que para un valor fijo de $f_0(t)$ la magnitud de $f(t|X_1 \dots X_T, \theta_1, \theta_2)$ es proporcional al cociente de verosimilitudes que se obtienen si se confrontan las hipótesis de no cambio contra un cambio en $r=t$. En cualquier caso la distribución a posteriori contiene toda la información relativa a r y en base a ella se pueden construir regiones de alta probabilidad o efectuar estimaciones puntuales o pruebas de hipótesis mediante el empleo de una función de pérdida adecuada.

Otro caso considerado por Smith es el de que θ_2 sea desconocido en donde es necesario definir una función de densidad

apriori para θ_2 sobre Θ_2 , el espacio de valores posibles de θ_2 , que el autor denota por $f_0(\theta_2|\theta_1)$ remarcando con ello la posibilidad de que esta función dependa del valor θ_1 . En estas condiciones interesa obtener tanto la función de densidad a posteriori de θ_2 como de τ . Si se supone que $f_0(t)$ es independiente de $f_0(\theta_2|\theta_1)$ se tiene que la conjunta a posteriori esta dada por:

$$f(t, \theta_2 | \underline{X}_T, \theta_1) = f_1(X_1 \dots X_t | \theta_1) f_2(X_{t+1}, \dots, X_T | \theta_2) f_0(\theta_2 | \theta_1) f_0(t)$$

para $\theta_2 \in \Theta_2$ y $t=1, 2, \dots, T$.

Para obtener las distribuciones a posteriori para cada parámetro basta con marginalizar la conjunta. Así se tiene que:

$$\begin{aligned} f(t | \underline{X}_T, \theta_1) &= \int_{\Theta_2} f(t, \theta_2 | X_1 \dots X_T, \theta_1) d\theta_2 \\ &= f_0(t) \int_{\Theta_2} f_1(X_1 \dots X_t | \theta_1) f_2(X_{t+1}, \dots, X_T | \theta_2) f_0(\theta_2 | \theta_1) d\theta_2 \\ &= f_0(t) f(X_1, \dots, X_T | \theta_1, t) \end{aligned}$$

y además:

$$\begin{aligned}
 f(\theta_2 | \underline{X}_T, \theta_1) &= \prod_{t=1}^T f(t, \theta_2 | X_1, \dots, X_T, \theta_1) \\
 &= \prod_{t=1}^T f_1(X_1, \dots, X_t | \theta_1) f_2(X_{t+1}, \dots, X_T | \theta_2) f_0(\theta_2 | \theta_1) f_0(t) \\
 &= f_0(\theta_2 | \theta_1) f(X_1, \dots, X_T | \theta_1, \theta_2).
 \end{aligned}$$

Estas distribuciones a posteriori contienen, como en el caso previo, toda la información respecto a θ_2 y τ y pueden emplearse para construir intervalos de alta probabilidad y estimar o probar hipótesis.

El caso más general se presenta al suponer que tanto θ_1 como θ_2 son desconocidas. En estas circunstancias, el tratamiento es totalmente análogo salvo que es necesario asignar una distribución a priori para θ_1 , $f_0(\theta_1)$, y suponer independencia de esta función respecto a $f_0(\theta_2)$ y $f_0(t)$ o bien asignar una distribución conjunta a priori $f_0(\theta_1, \theta_2)$, para θ_1 y θ_2 independiente de $f_0(t)$. Como resultado se obtiene la distribución a posteriori conjunta de θ_1 , θ_2 y τ de donde por integración o suma se pueden obtener las marginales respectivas.

Los estimadores de los parámetros se pueden obtener minimizando la pérdida esperada para una función de pérdida adecuada. En el caso de prueba de hipótesis, si no se emplea una función de pérdidas, la solución Bayesiana se basa en el cálculo de

los momios correspondientes. Así, por ejemplo si se desea probar:

$$H_0: \tau \in S \quad \text{vs.} \quad H_1: \tau \notin S$$

en donde S es un subconjunto del conjunto $\{1, 2, 3, \dots, T\}$, entonces si por brevedad se denota $f^*(t)$ a la función de probabilidad a posteriori de τ se tiene que:

$$f^*(S) = \sum_{t \in S} f^*(t)$$

y por tanto los momios a posteriori de H_0 contra H_1 están dados por:

$$\eta(S) = \frac{f^*(S)}{1 - f^*(S)}$$

Naturalmente, un valor suficientemente pequeño de $\eta(S)$ sugiere que H_0 se debe rechazar mientras que un valor grande apunta en el sentido de aceptar H_0 .

Este tratamiento, que puede ser aplicado sin complicaciones a variables cuya distribución pertenezca a la familia exponencial si se emplean distribuciones a priori conjugadas cuando los parámetros θ_1 y θ_2 son desconocidos, es ilustrado por Smith para variables con distribución Bernoulli y Normal.

Para variables Bernoulli, se tiene que las primeras r observaciones proceden de un modelo con probabilidad de éxito θ_1 y el resto de uno con probabilidad de éxito θ_2 . De esta forma,

$$f(X_1 \dots X_T | t, \theta_1, \theta_2) = \theta_1^{S_t} (1-\theta_1)^{f_t} \theta_2^{S_{T-t}} (1-\theta_2)^{f_{T-t}}$$

en donde

$$S_t = \sum_{i=1}^t X_i, \quad f_t = t - S_t, \quad S_{T-t} = \sum_{i=t+1}^T X_i, \quad f_{T-t} = T - t - S_{T-t}$$

Si θ_1 y θ_2 son conocidos, la distribución a posteriori de r se obtiene directamente de la expresión

$$f(t | X_T, \theta_1, \theta_2) \propto \theta_1^{S_t} (1-\theta_1)^{f_t} \theta_2^{S_{T-t}} (1-\theta_2)^{f_{T-t}} f_0(t); \quad t=1, 2, \dots, T$$

de donde los nomios para cualquier hipótesis de interés sobre r se calculan fácilmente. Si θ_2 es desconocido, una función de densidad que resulta conveniente en muchos casos para describir el conocimiento a priori sobre este parámetro es la Beta incompleta de cuatro parámetros que se define como:

$$f_0(\theta_2) = K_{a_2, b_2}(\alpha_2, \beta_2) \theta_2^{\alpha_2-1} (1-\theta_2)^{\beta_2-1}$$

para valores tales que $0 < a_2 < \theta_2 < b_2 < 1$ y donde $K_{a_2, b_2}(\alpha_2, \beta_2)$ es una

constante que satisface

$$\left\{ K_{a_2, b_2}(\alpha_2, \beta_2) \right\}^{-1} = \left\{ K_{0, b_2}(\alpha_2, \beta_2) \right\}^{-1} - \left\{ K_{0, a_2}(\alpha_2, \beta_2) \right\}^{-1}$$

con

$$K_{0, \omega}(u, v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)} \sum_{j=0}^{u+v-1} \binom{u+v-1}{j} w^j (1-w)^{u+v-j-1}$$

de tal manera que:

$$\int_{a_2}^{b_2} f_0(\theta_2) d\theta_2 = 1.$$

En virtud de que la forma esencial de la distribución a posteriori es la misma que la a priori se dice que la Beta es una distribución conjugada de la Bernoulli.

La asignación de los parámetros a_2, b_2, α_2 y β_2 permite modelar una variedad de situaciones. En particular, si ninguno de los parámetros depende de θ_1 se obtiene una distribución a priori para θ_2 independiente de θ_1 . Por otra parte, si se sabe que $\theta_2 > \theta_1$, basta con tomar $a_2 = \theta_1$. Un caso análogo se puede describir si $b_2 = \theta_1$. Finalmente, cuando $a_2 = 0$ y $b_2 = 1$ la distribución coincide con la Beta estándar o completa.

En cualquier caso, se puede verificar que la expresión general para la distribución a posteriori de τ esta dada por:

$$f(\tau | X_T, \theta_1) = \theta_1^{S_T} (1-\theta_1)^{T-S_T} \cdot K_{a_2, b_2}(\alpha_2, \beta_2) / K_{a_2, b_2}(\alpha_2 + S_T, \beta_2 + T - S_T) f_0(\tau).$$

Por supuesto, este tipo de distribución a priori puede generalizarse para el caso en que si bien, θ_2 no es siempre menor que θ_1 , ni siempre mayor que θ_1 , el conocimiento sobre θ_2 en cada una de esas regiones se desea describir con un modelo diferente de esta misma clase.

Cuando tanto θ_1 como θ_2 son desconocidos y es razonable suponer independencia entre las correspondientes distribuciones a priori; una solución se obtiene al proponer:

$$f_0(\theta_1, \theta_2) = \prod_{j=1}^2 K_{a_j, b_j}(\alpha_j, \beta_j) \theta_j^{\alpha_j-1} (1-\theta_j)^{\beta_j-1} \quad (a_j < \theta_j < b_j, j=1, 2)$$

Utilizando esta a priori es muy sencillo obtener la correspondiente posteriori conjunta de τ , θ_1 y θ_2 en una forma análoga a los casos anteriores. Asimismo, se puede marginalizar para obtenerla a posteriori individuales. Si de antemano se sabe que $\theta_1 < \theta_2$ el conocimiento a priori se puede describir mediante la función

$$f_0(\theta_1, \theta_2) = K_{a_1, b_2}(\alpha_1, \beta_1, \alpha_2, \beta_2) \theta_1^{\alpha_1-1} (1-\theta_1)^{\beta_1-1} \theta_2^{\alpha_2-1} (1-\theta_2)^{\beta_2-1}$$

para $a_1 < \theta_1 < \theta_2 < b_2$. En este caso la constante K está dada por la expresión:

$$\left\{ K_{a_1, b_2}(\alpha_1, \beta_1, \alpha_2, \beta_2) \right\}^{-1} = \int_{a_1}^{b_2} \left[\int_y^{b_2} X^{\alpha_2-1} (1-X)^{\beta_2-1} d_X \right] y^{\alpha_1-1} (1-y)^{\beta_1-1} dy$$

esto es, K es una constante de normalización de la distribución a priori conjunta de θ_1 y θ_2 .

El procedimiento consiste en calcular de nuevo, la distribución conjunta a posteriori de θ_1, θ_2 y τ de donde se obtienen las marginales. En particular, la a posteriori de τ está dada por:

$$f(t | X_T) = f_0(t) K_{a_1, b_2}(\alpha_1, \beta_1, \alpha_2, \beta_2) \cdot [K_{a_1, b_2}(\alpha_1 + S_t, \beta_1 + f_t, \alpha_2 + S_{T-t}, \beta_2 + f_{T-t})]$$

para $t=1, 2, \dots, T$.

Naturalmente, con esta idea se puede obtener un resultado análogo cuando $\theta_2 < \theta_1$. Un caso general, se puede modelar asignando un modelo de este tipo para la región $\theta_1 < \theta_2$ y otro similar para la región $\theta_2 < \theta_1$, ajustando las constantes correspondientes para que la integral sobre la región total tenga valor 1.

Para sucesiones de variables Normales y considerando $\theta_j = (\mu_j, \phi_j)$ en donde μ y ϕ son la media y la precisión de la distribución respectivamente se tiene que:

$$f(\sum_{i=1}^t X_i | t, \theta_1, \theta_2) = \left(\frac{\phi_1}{2\pi}\right)^{t/2} \left(\frac{\phi_2}{2\pi}\right)^{\frac{T-t}{2}} * \\ * \exp\left\{-\frac{1}{2} \phi_1 \sum_{i=1}^t (X_i - \mu_1)^2\right\} \exp\left\{-\frac{1}{2} \phi_2 \sum_{i=t+1}^T (X_i - \mu_2)^2\right\}$$

de donde la distribución a posteriori de r se puede obtener muy fácilmente si μ_1, μ_2, ϕ_1 y ϕ_2 son todos conocidos. Diferentes resultados se pueden obtener si algunos de estos parámetros son desconocidos. Smith trata varios casos en los cuales emplea distribuciones a priori conjugadas o en ocasiones distribuciones de referencia para μ_1, μ_2, ϕ_1 y ϕ_2 suponiendo siempre, independencia respecto a $f_0(t)$.

Cuando $\phi_1 = \phi_2 = \phi$ conocida, μ_1 es conocida y sólo μ_2 es desconocida, el autor propone el uso de una distribución a priori para μ_2 , Normal truncada en (a_2, b_2) con media α_2 y precisión $\phi \beta_2$. Esto es,

$$f_0(\mu_2) = K_{a_2, b_2}(\alpha_2, \beta_2, \phi) \exp\left\{-\frac{1}{2} \phi \beta_2 \mu_2^2\right\} \exp\left\{\phi \beta_2 \alpha_2 \mu_2\right\}$$

en donde de nuevo $K_{a_2, b_2}(\alpha_2, \beta_2, \phi)$ juega el papel de una cons

tante normalizadora.

La distribución a posteriori de τ , el parámetro de mayor interés, está dada por:

$$f(t | \underline{X}_T; \mu_1, \phi) \propto \left(\frac{\phi}{2\pi}\right)^{\frac{T}{2}} \exp \left\{ -\frac{1}{2} \phi \left[\sum_{i=1}^t (X_i - \mu_1)^2 + \sum_{i=t+1}^T X_i^2 \right] \right\} f_0(t)$$

$$*K_{a_2, b_2}(\alpha_2, \beta_2, \phi) / K_{a_2, b_2}(\alpha_2^{(t)}, \beta_2^{(t)}, t)$$

en donde:

$$\alpha_2^{(t)} = \left(\sum_{i=t+1}^T X_i + \beta_2 \alpha_2 \right) / (T-t+\beta_2), \quad \beta_2^{(t)} = \beta_2 + T - t.$$

Otro caso considerado es el de μ_1 conocido, μ_2 desconocido y $\phi_1 = \phi_2 = \phi$ desconocido también. En estas circunstancias, Smith aborda el caso especial de asignar a ϕ y μ_2 una distribución de referencia

$$f_0(\mu_2, \phi) \propto \phi^{-1}$$

con la cual mediante integración se obtiene que:

$$f(t|X_T, \mu_1) \propto f_0(t) (T-t)^{-\frac{T+1}{2}} \left\{ \sum_{i=1}^t (X_i - \mu_1)^2 + \sum_{i=t+1}^T (X_i - \bar{X}_{T-t})^2 \right\}^{-\frac{T+1}{2}}$$

en donde:

$$\bar{X}_{T-t} = (T-t)^{-1} \sum_{i=t+1}^T X_i .$$

Es interesante notar que si $f_0(t) = 1/T$ para $t=1, 2, \dots, T$, los valores de τ con mayor probabilidad a posteriori tienden a ser aquellos para los cuales, el correspondiente modelo con un cambio en $\tau=t$ proporciona el mejor ajuste a las observaciones en el sentido de que las discrepancias $X_i - \mu_1$ para $i=1, 2, \dots, t$ y $X_i - \bar{X}_{T-t}$ para $i=t+1, \dots, T$ sean pequeñas en sumas de cuadrados.

Asignando una distribución a priori de referencia

$$f_0(\mu_1, \mu_2, \phi) = \phi^{-1}$$

se puede tratar el caso en que también μ_1 es desconocido. Mediante un procedimiento análogo se obtiene que:

$$f(t|X_T) \propto f_0(t) \{t(T-t)\}^{-\frac{1}{2}} \left[\sum_{i=1}^t (X_i - \bar{X}_t)^2 + \sum_{i=t+1}^T (X_i - \bar{X}_{T-t})^2 \right]^{-\frac{T}{2}}$$

en donde

$$\bar{X}_t = t^{-1} \sum_{i=1}^t X_i$$

Naturalmente, si $f_0(t)$ es constante, la interpretación de los valores de r con máxima probabilidad a posteriori es completamente análoga a la anterior sustituyendo μ_1 por \bar{X}_t .

Finalmente, si todos los parámetros son desconocidos y no hay razón para suponer que $\phi_1 = \phi_2$, Smith trata el problema cuando se asigna la a priori conjunta de referencia

$$f_0(\mu_1, \mu_2, \phi_1, \phi_2) = \phi_1^{-1} \phi_2^{-1}$$

en cuyo caso se obtiene que:

$$f(t | X_T) = f_0(t) \Gamma\left(\frac{1}{2}(t+1)\right) \Gamma\left(\frac{1}{2}(T-t+1)\right) \left\{ \sum_{i=1}^t (X_i - \bar{X}_t)^2 \right\}^{-\frac{(t+1)}{2}} \left\{ \sum_{i=t+1}^T (X_i - \bar{X}_{T-t})^2 \right\}^{-\frac{(T-t+1)}{2}}$$

Smith menciona que este tipo de procedimientos puede ser adaptado sin dificultad, para un proceso secuencial en donde después de obtener la m -ésima observación se puede asignar una distribución a priori $f_0(t)$: $t=1, 2, \dots, m$ y proceder al cálculo de la correspondiente a posteriori para evaluar el nomio

$$\eta(m) = \frac{f(m | X_1 \dots X_m)}{1 - f(m | X_1 \dots X_m)}$$

que permite decidir si se verificó un cambio en las primeras m observaciones. La información contenida en este nomio pue de interpretarse de una manera bastante satisfactoria si se calcula la sucesión $\eta(m)$ para los diferentes valores de m . Es claro que una tendencia decreciente, si es suficiente mar cada, sugiere que efectivamente tuvo lugar un cambio en la sucesión completa. Por último, resulta interesante notar el paralelismo de este procedimiento informal, como lo califica el propio Smith, con el tratamiento de Chernoff y Zacks [4] y Kander y Zacks [5] .

En un trabajo posterior, [12], Broemeling retoma el problema pero centra el interés en la predicción de observaciones futuras generadas por el proceso. Es claro que si existiese certeza acerca de que el cambio tuvo lugar, necesariamente la distribución de las observaciones futuras es la correspon diente al modelo que describe el comportamiento de la sucesión después del punto de cambio. Más aún, si se conoce el punto en que se verificó el cambio el problema de que tal mo delo involucre parámetros desconocidos, se puede tratar de una manera estándar, puesto que se conoce exactamente que ob servaciones de la sucesión pueden ser empleadas para efectuar las correspondientes inferencias. Generalmente, como ya se ha mencionado, sucede que no se sabe con exactitud en

que punto se efectuó el cambio y como consecuencia existe incertidumbre acerca de las observaciones que deben ser empleadas para la inferencia. El procedimiento entonces, consiste en determinar la distribución predictiva esperada de las futuras observaciones con respecto al conocimiento a posteriori que se tenga respecto al punto de cambio y los parámetros involucrados. El autor considera diversos casos, siempre suponiendo que efectivamente ocurrió el cambio.

Sucesiones de Variables Normales. En este caso, se considera una sucesión $X_1, X_2, \dots, X_t, X_{t+1}, \dots, X_T$ tal que $X_i \sim N(\phi_0, \sigma^2)$ para $i=1, 2, \dots, t$, $X_i \sim N(\phi_1, \sigma^2)$ para $i=t+1, \dots, T$ con $1 \leq t \leq T-1$ desconocido.

En estas condiciones, las modalidades más generales que se plantean son primero, ϕ_0 y ϕ_1 desconocidos pero σ^2 conocido y después los tres parámetros, ϕ_0 , ϕ_1 y σ^2 desconocidos.

En cualquier caso se tiene que la función de densidad conjunta de la muestra esta dada por:

$$f(\underline{X}_T | t, \phi_0, \phi_1, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^t (X_i - \phi_0)^2 + \sum_{i=t+1}^T (X_i - \phi_1)^2 \right\}\right]$$

Si se desea obtener la función de densidad predictiva para

K observaciones futuras Y_1, \dots, Y_K del proceso, es claro que si ϕ_1, σ^2 y t fuesen conocidas la densidad conjunta estaría dada por:

$$f(Y_1, \dots, Y_K | \phi_1, \sigma^2, t) = (2\pi\sigma^2)^{-K/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^K (Y_j - \phi_1)^2 \right\}$$

Sin embargo, en general estos parámetros no se conocen, de manera que es necesario construir la densidad predictiva tomando en cuenta la información que respecto a todos los parámetros desconocidos proporciona la muestra inicial \underline{X}_T . Esto es, la función predictiva se marginaliza respecto a los parámetros involucrados en base a la distribución posteriori respectiva.

Así, cuando σ^2 es conocida es necesario asignar una distribución inicial o a priori para los posibles valores de los parámetros ϕ_0, ϕ_1 y t . Broemeling emplea, en el trabajo mencionado, la siguientes función de densidad a priori conjunta:

$$f_0(\phi_0, \phi_1, t) \propto (T-1)^{-1}; \phi_0 \in \mathbb{R}, \phi_1 \in \mathbb{R}, t=1, 2, \dots, T-1$$

Esta función pertenece a la clase de las llamadas funciones de densidad impropias en virtud de que con respecto a ϕ_0 y ϕ_1 es constante y esta definida para todo valor en los reales de modo que no integra a 1. El empleo de este tipo de funciones

es usual en condiciones en las cuales se pretende que las distribuciones a posteriori dependan básicamente de la verosimilitud de los datos observados. La a posteriori para ϕ_0, ϕ_1, t resulta entonces proporcional a la verosimilitud, esto es:

$$f(\phi_0, \phi_1, t | \underline{X}_T) = (2\pi\sigma^2)^{-T/2} \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^t (X_i - \phi_0)^2 + \sum_{i=t+1}^T (X_i - \phi_1)^2 \right\} \right]$$

para $t=1, 2, \dots, T-1$; $\phi_0 \in \mathbb{R}$, $\phi_1 \in \mathbb{R}$. O bien, equivalentemente:

$$\begin{aligned} f(\phi_0, \phi_1, t) &= (2\pi\sigma^2)^{-T/2} \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^t (X_i - \bar{X}_t + \bar{X}_t - \phi_0)^2 + \sum_{i=t+1}^T (X_i - \bar{X}_{T-t} + \bar{X}_{T-t} - \phi_1)^2 \right\} \right] \\ &= (2\pi\sigma^2)^{-T/2} \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^t (X_i - \bar{X}_t)^2 + t(\bar{X}_t - \phi_0)^2 + \sum_{i=t+1}^T (X_i - \bar{X}_{T-t})^2 + (T-t)(\bar{X}_{T-t} - \phi_1)^2 \right\} \right] \end{aligned}$$

Así, se obtiene la función de densidad predictiva a posteriori de acuerdo a la siguiente expresión:

$$f(\underline{Y}_k | \underline{X}_T) = \sum_{t=1}^{T-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\underline{Y}_k | \phi_0, \phi_1, t) f(\phi_0, \phi_1, t | \underline{X}_T) d\phi_0 d\phi_1$$

en donde, naturalmente $\underline{Y}_k = Y_1 \dots Y_k$.

El resultado que Broemeling obtiene es que $f(Y_k | X_T)$ es una mezcla de distribuciones normales. El procedimiento, es el siguiente. En virtud de que σ^2 es conocida, se puede suponer que es igual a 1 y entonces

$$f(Y_k | \phi_0, \phi_1, t) f(\phi_0, \phi_1, t | X_T) = \exp \left[-\frac{1}{2} \left\{ C_1(t) + t(\phi_0 - \bar{X}_t)^2 + (T-t)(\phi_1 - \bar{X}_{T-t})^2 \right\} \right] \\ * \exp \left[-\frac{1}{2} \left\{ \sum_{i=1}^k (Y_i - \phi_1)^2 \right\} \right]$$

con $C_1(t) = \sum_{i=1}^t (X_i - \bar{X}_t)^2 + \sum_{i=t+1}^T (X_i - \bar{X}_{T-t})^2$. Por lo tanto,

$$f(Y_k | X_T) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\exp \left[-\frac{1}{2} \left\{ C_1(t) + (T-t)(\phi_1 - \bar{X}_{T-t})^2 + \sum_{i=1}^k (Y_i - \phi_1)^2 \right\} \right] \right] \\ * \left[\exp \left[-\frac{1}{2} t (\phi_0 - \bar{X}_t)^2 \right] d\phi_0 \right] d\phi_1$$

$$= \int_{-\infty}^{\infty} t^{-1/2} \exp \left[-\frac{1}{2} \left\{ C_1(t) + (T-t)(\phi_1 - \bar{X}_{T-t})^2 + \sum_{i=1}^k (Y_i - \phi_1)^2 \right\} \right] d\phi_1$$

$$= \sum_{t=1}^T t^{-1/2} \exp \left[-\frac{1}{2} \{C_2(t)\} \right] \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} \left\{ (T-t) (\phi_1 - \bar{X}_{T-t})^2 + \sum_{i=1}^k (Y_i - \phi_1)^2 \right\} \right] d\phi_1$$

Ahora bien, la expresión en el exponente dentro de la integral puede ser reexpresada para completar una densidad normal en ϕ_1 de la siguiente manera:

$$\begin{aligned} (T-t) (\phi_1 - \bar{X}_{T-t})^2 + \sum_{i=1}^k (Y_i - \phi_1)^2 &= (T-t) \phi_1^2 - 2(T-t) \phi_1 \bar{X}_{T-t} + (T-t) \bar{X}_{T-t}^2 \\ &\quad + \sum_{i=1}^k Y_i^2 - 2 \phi_1 \sum_{i=1}^k Y_i + k \phi_1^2 \\ &= (T-t+k) \phi_1^2 - 2 \left(\sum_{i=1}^k Y_i + (T-t) \bar{X}_{T-t} \right) \phi_1 \\ &\quad + \sum_{i=1}^k Y_i^2 + (T-t) \bar{X}_{T-t}^2 \\ &= (T-t+k) \left[\phi_1 - \frac{\sum_{i=1}^k Y_i + (T-t) \bar{X}_{T-t}}{T-t+k} \right]^2 \\ &\quad - (T-t+k) \left[\frac{\sum_{i=1}^k Y_i + (T-t) \bar{X}_{T-t}}{T-t+k} \right]^2 \\ &\quad + \sum_{i=1}^k Y_i^2 + \bar{X}_{T-t}^2 (T-t) \end{aligned}$$

De esta manera se tiene que:

$$\begin{aligned}
 f(Y_k | X_T) &= \sum_{t=1}^{T-1} t^{-1/2} (T-t+k)^{-1/2} \exp \left[-\frac{1}{2} \left\{ C_1(t) \right\} \right] \\
 & \times \exp \left[-\frac{1}{2} \left\{ \sum_{i=1}^k Y_i^2 + (T-t) \bar{X}_{T-t}^2 - (T-t+k) \left[\frac{\sum_{i=1}^k Y_i + (T-t) \bar{X}_{T-t}}{T-t+k} \right]^2 \right\} \right] \\
 & = \sum_{t=1}^{T-1} \ell_1(t) \cdot n \left(\frac{Y_k, \bar{X}_{T-t}}{T-t+k} \mathbf{J}_k, \left(\frac{S_t}{T-t+k} \right)^{-1} \right)
 \end{aligned}$$

en donde $\ell_1(t) = t^{-1/2} (T-t)^{-1/2} \exp \left[-\frac{1}{2} C_1(t) \right]$, \mathbf{J}_k es un vector de dimensión k con todas sus componentes iguales a 1, S_t es una matriz de dimensión $k \times k$ con elementos iguales a $T-t+k$ en la diagonal e iguales a -1 fuera de ella y la expresión $n(\underline{Y}_k, \underline{\mu}, \Sigma)$ denota una densidad normal K -variada con vector de medias $\underline{\mu}$ y matriz de varianzas y covarianzas Σ .

Para el caso en que σ^2 también es desconocida, Broemeling propone una densidad a priori

$$f_0(\phi_0, \phi_1, \sigma^2, t) \propto \sigma^{-2} (T-1)^{-1}; \quad \sigma^2 > 0, \phi_1 \in \mathbb{R}, \phi_2 \in \mathbb{R}$$

y mediante un procedimiento análogo obtiene que:

$$f(\underline{Y}_k | \underline{X}_T) = \sum_{t=1}^{T-1} k_1(t) h(\underline{Y}_k; T-2, \bar{X}_{T-t} \underline{J}_k, \Sigma_t)$$

en donde $K_1(t) = C_1(t)^{-\frac{T-2}{2}} t^{-1/2} (T-t)^{-\frac{1}{2}}$ y Σ es una matriz $k \times k$ tal que $\Sigma_t = S_t(T-2) [C_1(t) (T-t+k)]^{-1}$; con $C_1(t)$, \underline{J}_k y S_t definidos como en el caso anterior. Por su parte, $h(\underline{Y}_k; p, \underline{\mu}, \Sigma)$ es una densidad t de Student K -variada con p grados de libertad, vector de medias $\underline{\mu}$ y matriz de precisión Σ . Esto es, la función predictiva de \underline{Y}_k está dada por una mezcla de distribuciones multivariadas t de Student. Para algunas otras situaciones, ϕ_0 conocido pero ϕ_1 desconocido o viceversa y dependiendo de si σ^2 es a su vez conocida o desconocida se obtienen resultados análogos en donde la diferencia estriba en que tanto los coeficientes de las mezclas como los parámetros de las distribuciones mezcladas difieren, reflejando el conocimiento que se posee sobre el proceso.

En el caso de sucesiones de variables Bernoulli se tiene que $\underline{X}_T = (X_1, \dots, X_T)$ es una colección de variables aleatorias independientes tales que $X_i \sim \text{Blli}(\phi_i); i=1, 2, \dots, t$ y $X_i \sim \text{Blli}(\phi_i); i=t+1, \dots, T$. Naturalmente, se supone que: $1 \leq t \leq T-1$ y $\phi_0, \phi_1 \in (0, 1)$.

Bajo estas condiciones, la función de densidad conjunta de \underline{X}_T

dados los valores de ϕ_0 , ϕ_1 y t esta dada por

$$f(X_T | \phi_0, \phi_1, t) = \phi_0^{\Delta_t} (1-\phi_0)^{t-\Delta_t} \phi_1^{\Delta_{T-t}} (1-\phi_1)^{T-t-\Delta_{T-t}}$$

en donde $\Delta_t = \sum_{i=1}^t X_i$ y $\Delta_{T-t} = \sum_{i=t+1}^T X_i$.

Broemeling considera la situación más general en la que tanto ϕ_0 como ϕ_1 son desconocidos. Adicionalmente supone que ϕ_0, ϕ_1 y t son independientes a priori, que ϕ_0 y ϕ_1 son idénticamente distribuidas y que el conocimiento respecto a estos dos parámetros se puede describir adecuadamente mediante una distribución conjugada Beta (α, β). Así, la distribución a priori conjunta resulta ser la siguiente:

$$f_0(\phi_0, \phi_1, t) = (T-1)^{-1} \left[\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right]^2 \phi_0^{\alpha-1} (1-\phi_0)^{\beta-1} \phi_1^{\alpha-1} (1-\phi_1)^{\beta-1}$$

De esta manera, se obtiene la posteriori conjunta de ϕ_0, ϕ_1 y t según la expresión:

$$f(\phi_0, \phi_1, t | X_T) \propto \phi_0^{\Delta_t + \alpha - 1} (1-\phi_0)^{t + \beta - \Delta_t - 1} \phi_1^{\Delta_{T-t} + \alpha - 1} (1-\phi_1)^{T-t + \beta - \Delta_{T-t} - 1}$$

Por otra parte, la función de densidad de K observaciones futuras está dada por:

$$f(Y_k | \phi_1) = \phi_1^{\sum_{i=1}^k Y_i} (1 - \phi_1)^{k - \sum_{i=1}^k Y_i}$$

Entonces la predictiva de Y_k dado X_T se obtiene según el siguiente procedimiento:

$$f(Y_k | X_T) = \sum_{t=1}^{T-1} \int_0^1 \int_0^1 f(Y_k | \phi_1) f(\phi_0, \phi_1, t | X_T) d\phi_0 d\phi_1$$

Ahora bien,

$$\begin{aligned} \int_0^1 f(Y_k | \phi_1) f(\phi_0, \phi_1, t | X_T) d\phi_0 &\propto \int_0^1 \phi_0^{\delta t + \alpha - 1} (1 - \phi_0)^{t + \beta - \delta t - 1} \frac{\phi_0^{\delta T - t + \sum_{i=1}^k Y_i - 1}}{\phi_1} \\ &\quad * (1 - \phi_1)^{T - t + \beta + k - \delta T - t - \sum_{i=1}^k Y_i - 1} d\phi_0 \\ &= \phi_1^{\delta T - t + \alpha} \frac{\phi_1^{\sum_{i=1}^k Y_i - 1}}{(1 - \phi_1)^{T - t + \beta - \delta T - t - \sum_{i=1}^k Y_i - 1}} \\ &\quad * \frac{\Gamma(\delta t + \alpha) \Gamma(t + \beta - \delta t)}{\Gamma(t + \beta + \alpha)} \end{aligned}$$

integrando respecto a ϕ_1 se obtiene que:

$$f(Y_k | X_T) = \sum_{t=1}^{T-1} \left\{ \frac{\Gamma(\delta_t + \alpha) \Gamma(t + \beta - \delta_t)}{\Gamma(\alpha + \beta + t)} \right. \\ \left. * \int_0^1 \phi_1^{\delta_{T-t} + \alpha + \sum_{i=1}^k Y_i - 1} (1 - \phi_1)^{T-t + \beta + k - \delta_{T-t} + \sum_{i=1}^k Y_i - 1} d\phi_1 \right\}$$

$$= \sum_{t=1}^{T-1} \frac{\Gamma(\delta_t + \alpha) \Gamma(t + \beta - \delta_t)}{\Gamma(\alpha + \beta + t)} * \frac{\Gamma(\delta_{T-t} + \alpha + \sum_{i=1}^k Y_i) \Gamma(T-t + \beta + k - \delta_{T-t} - \sum_{i=1}^k Y_i)}{\Gamma(T-t + \alpha + \beta + k)}$$

De tal forma que en este caso la predictiva resulta una mezcla de funciones de probabilidad del tipo:

$$f(Y_k | t) = \frac{\Gamma(\delta_{T-t} + \alpha + \sum_{i=1}^k Y_i) \Gamma(T-t + \beta + k - \delta_{T-t} - \sum_{i=1}^k Y_i)}{\Gamma(T-t + \alpha + \beta + k)}$$

con coeficientes de mezcla:

$$k(t) = \frac{\Gamma(\delta_t + \alpha) \Gamma(t + \beta - \delta_t)}{\Gamma(\alpha + \beta + t)}$$

Finalmente, Broemeling trata el caso de sucesiones de variables con distribución exponencial. Considera $X_1 \dots X_T$ una colección de variables tales que:

$$f(X_i | \phi_0) = \phi_0 e^{-\phi_0 X_i}, \quad i=1, 2, \dots, r,$$

$$f(X_i | \phi_1) = \phi_1 e^{-\phi_1 X_i}, \quad i=r+1, \dots, T.$$

La función de densidad conjunta o equivalentemente, la verosimilitud, esta dada entonces por

$$f(\underline{X}_T | \phi_0, \phi_1, t) = \phi_0^t \phi_1^{(T-t)} e^{-\phi_0 \Delta t} e^{-\phi_1 \Delta T-t}$$

Suponiendo a priori independencia y bajo el supuesto de que ϕ_0 y ϕ_1 son idénticamente distribuidas según un modelo Gama (α, β) que es el conjugado para el exponencial, se tiene que

$$f_0(\phi_0, \phi_1, t) = (T-1)^{-1} \beta^{2\alpha} \frac{\phi_0^{\alpha-1} \phi_1^{\alpha-1} e^{-\phi_0 \beta} e^{-\phi_1 \beta}}{\Gamma^2(\alpha)}$$

Combinando la verosimilitud con la a priori se obtiene la siguiente expresión para la a posteriori:

$$f(\phi_0, \phi_1, t | \underline{X}_T) = \phi_0^{\alpha+t-1} \phi_1^{\alpha+T-t-1} e^{-\phi_0(\Delta_t+\beta)} e^{-\phi_1(\Delta_{T-t}+\beta)}$$

Para este caso, la función de densidad conjunta de k observaciones futuras, dado el valor de ϕ_1 esta descrita por:

$$f(Y_k | \phi_1) = \phi_1^k e^{-\phi_1 \sum_{i=1}^k Y_i}$$

Por consiguiente la predictiva se calcula como sigue:

$$\begin{aligned} f(Y_k | \underline{X}_T) &= \sum_{t=1}^{T-1} \int_0^{\infty} \int_0^{\infty} f(Y_k | \phi_1) f(\phi_0, \phi_1, t | \underline{X}_T) d\phi_0 d\phi_1 \\ &= \sum_{t=1}^{T-1} \int_0^{\infty} \int_0^{\infty} \phi_1^k e^{-\phi_1 \sum_{i=1}^k Y_i} \phi_0^{\alpha+t-1} \phi_1^{\alpha+T-t-1} e^{-\phi_0(\Delta_t+\beta)} e^{-\phi_1(\Delta_{T-t}+\beta)} d\phi_0 d\phi_1 \\ &= \sum_{t=1}^{T-1} \int_0^{\infty} \phi_1^{\alpha+T-t+k-1} e^{-\phi_1(\Delta_{T-t}+\beta + \sum_{i=1}^k Y_i)} \int_0^{\infty} \phi_0^{\alpha+t-1} e^{-\phi_0(\Delta_t+\beta)} d\phi_0 d\phi_1 \\ &= \sum_{t=1}^{T-1} \int_0^{\infty} \phi_1^{\alpha+T-t+k-1} e^{-\phi_1(\Delta_{T-t}+\beta + \sum_{i=1}^k Y_i)} \frac{\Gamma(\alpha+t)}{(\Delta_t+\beta)^{\alpha+t}} d\phi_1 \\ &= \sum_{t=1}^{T-1} \frac{\Gamma(\alpha+t)}{(\Delta_t+\beta)^{\alpha+t}} \int_0^{\infty} \phi_1^{\alpha+T-t+k-1} e^{-\phi_1(\Delta_{T-t}+\beta + \sum_{i=1}^k Y_i)} d\phi_1 \end{aligned}$$

$$= \sum_{t=1}^{T-1} \frac{\Gamma(\alpha+t)}{\{\delta_t \beta\}^{\alpha+t}} \cdot \frac{\Gamma(\alpha+k+T-t)}{\{\delta_{T-t} \beta + \sum_{i=1}^k Y_i\}^{\alpha+k+T-t}}$$

De nuevo entonces, se obtiene que la función predictiva de \underline{Y}_k es una mezcla de distribuciones de probabilidad de la forma:

$$h(\underline{Y}_k | t) \propto \left\{ \delta_{T-t} \beta + \sum_{i=1}^k Y_i \right\}^{-(\alpha+k+T-t)}$$

con coeficientes de mezcla:

$$g(t) = \frac{\Gamma(\alpha+t) \Gamma(\alpha+k+T-t)}{\{\delta_t \beta\}^{\alpha+t}}$$

Como una observación final a los resultados obtenidos por Broemeling, debe notarse que si bien el objetivo central, en todos los casos, es encontrar la distribución predictiva de k observaciones futuras, como un subproducto se obtiene siempre la función de densidad a posteriori del punto de cambio τ . Mediante el procedimiento Bayesiano la predictiva resulta una mezcla de distribuciones en donde los coeficientes de la mez-

cla, que dependen de t , son proporcionales a la probabilidad posterior de que $r=t$.

En un trabajo más reciente, Holbert y Broemeling [49], se trata el problema de estimar el punto de cambio τ , en una sucesión de variables aleatorias Normales. Los autores parten de la suposición de que un cambio se efectuó en el proceso. Consideran variables independientes tales que X_i se distribuye $N(\phi_0, \sigma^2)$ para $i=1, 2, \dots, \tau$ y X_i se distribuye $N(\phi_1, \sigma^2)$ para $i=\tau+1, \dots, T$ con $1 \leq \tau \leq T-1$. En virtud de que el proceso de estimación Bayesiano consiste en elegir el valor paramétral que minimice la pérdida esperada respecto a la distribución del parámetro, el objetivo se centra en la determinación de la distribución a posteriori de τ .

En el caso en que ϕ_0 y ϕ_1 son valores conocidos y si se asignan a priori las distribuciones independientes:

$$f_0(\tau) = (T-1)^{-1} ; \tau = 1, 2, \dots, T-1$$

y

$$f_0(\sigma^2) = \frac{1}{\sigma^2} ; 0 < \sigma^2 < \infty$$

se obtiene que

$$f(\tau | \underline{X}_T) = \left\{ \sum_{i=1}^{\tau} (X_i - \phi_0)^2 + \sum_{i=\tau+1}^T (X_i - \phi_1)^2 \right\}^{-T/2}; \tau=1, 2, \dots, T-1$$

a través del procedimiento usual de obtener la conjunta a posteriori $f(\tau, \sigma^2 | \underline{X}_T)$ y marginalizar respecto a σ^2 .

Cuando ϕ_0 y ϕ_1 son desconocidos, Holbert y Broemeling abordan el caso en que $\phi_0 < \phi_1$ y asignan las distribuciones independientes

$$f_0(\phi_0, \phi_1) \propto 1 \quad ; \quad -\infty < \phi_0 < \phi_1 < \infty$$

$$f_0(\sigma^2) \propto 1/\sigma^2 \quad ; \quad 0 < \sigma^2 < \infty$$

$$f_0(\tau) = (T-1)^{-1} \quad ; \quad 1 \leq \tau \leq T-1$$

La distribución a posteriori conjunta queda dada por la expresión:

$$f(\phi_0, \phi_1, \sigma^2, \tau | \underline{X}_T) \propto (\sigma^2)^{-(\frac{n}{2} + 1)} \exp \left\{ \frac{1}{2\sigma^2} \left[\sum_{i=1}^{\tau} (X_i - \phi_0)^2 + \sum_{i=\tau+1}^T (X_i - \phi_1)^2 \right] \right\}$$

a partir de la cual, nuevamente marginalizando respecto a σ^2 , ϕ_1 y ϕ_0 , se tiene que

$$f(\tau | \underline{X}_T) = (\tau(T-\tau))^{-\frac{1}{2}} [C_1^\tau + C_{\tau+1}^\tau]^{-\frac{(T-2)}{2}}$$

$$E_{\phi_0} \{ 1 - \phi_{n-1}(H(\tau, \phi_0)) \}$$

en donde,

$$C_1^\tau = \sum_{i=1}^{\tau} (X_i - \bar{X}_\tau)^2, \quad C_{\tau+1}^\tau = \sum_{i=\tau+1}^T (X_i - \bar{X}_{T-\tau})^2$$

$$H(\tau, \phi_0) = \left\{ \frac{(T-1)(T-\tau)}{\tau(T-2)} \right\}^{\frac{1}{2}} \frac{[C_1^\tau + C_{\tau+1}^\tau]^{1/2} \phi_0 + \tau^{1/2} (T-2)^{1/2} (\bar{X}_\tau - \bar{X}_{T-\tau})}{[C_1^\tau + C_{\tau+1}^\tau]^{1/2} [1 + \phi_0^2 / (T-2)]^{1/2}}$$

$\psi_{n-1}(x)$ es la función de distribución t de Student con n-1 grados de libertad, valuada en x y $E_{\phi_0}(f(\phi_0))$ es la esperanza de $f(\phi_0)$ respecto a una distribución t de Student con T-2 grados de libertad.

Así pues, en el caso más complejo considerado por los autores, la distribución a posteriori de τ resulta una función que si bien tiene una expresión relativamente complicada y requiere para su cálculo de rutinas de integración numérica, es exacta y depende de funciones de distribución estándar.

Tres artículos más pueden ser mencionados que tratan problemas de puntos de cambio para sucesiones de variables aleatorias. El primero de ellos, Smith [9], propone un procedimiento para establecer el valor del punto de cambio en términos de un problema de selección de modelos que en algunos casos puede resultar más ilustrativo. El autor considera la siguiente situación:

Sea X_1, \dots, X_T una sucesión de variables aleatorias independientes. Sea M_r el modelo tal que para $i \leq r$, X_i tiene función de densidad $f(X_i | \theta_1)$ mientras que para $i > r$, X_i tiene función de densidad $f(X_i | \theta_2)$. El índice r corresponde al punto de donde se lleva a cabo el cambio de θ_1 a θ_2 y naturalmente debe satisfacer que $1 \leq r \leq T$ de tal manera que existen T diferentes modelos M_r . Como en el caso ya mencionados anteriormente, si los parámetros θ_1 y θ_2 son desconocidos es necesario asignar una función de probabilidad conjunta $f_0(\theta_1, \theta_2)$. Así pues, si el objetivo que se persigue es la estimación del punto de cambio, el problema se puede plantear en términos de seleccionar el modelo M_r más adecuado. Naturalmente, desde la perspectiva Bayesiana, la forma de proceder está completamente especificada y consiste en calcular las probabilidades a posteriori:

$$P(M_r | \underline{X}_T) = P(\underline{X}_T | M_r) P(M_r); \quad r=1, 2, \dots, T$$

en donde, $P(M_t)$ representa la probabilidad a priori asignada al modelo M_t y el término $P(\underline{X}_T | M_t)$, que se determina con la siguiente relación:

$$P(\underline{X}_T | M_t) = \int_{\Theta_1} \int_{\Theta_2} \prod_{i=1}^t f(X_i | \theta_1) \prod_{i=t+1}^T f(X_i | \theta_2) f_0(\theta_1, \theta_2) d\theta_1 d\theta_2$$

Se puede interpretar como la verosimilitud esperada de M_t respecto al conocimiento que se posee acerca de θ_1 y θ_2 .

En base a esas probabilidades a posteriori y una adecuada función de utilidades o pérdidas se selecciona la opción (el modelo) con mayor utilidad esperada o equivalentemente, con menor pérdida esperada. Sin embargo, en algunas ocasiones cuando la especificación de la función de pérdidas o de utilidades resulta particularmente difícil, algunos autores sugieren la selección de la opción más probable. Este procedimiento puede efectuarse calculando los nomios a posteriori para cada pareja de modelos:

$$\eta \left(\frac{M_j}{M_k} \mid \underline{X}_T \right) = \frac{P(\underline{X}_T | M_j) P(M_j)}{P(\underline{X}_T | M_k) P(M_k)}$$

En el trabajo que se discute, Smith va aún más lejos y propone un método de comparación que elimina la contribución de las probabilidades a priori, o dicho de otra forma, produce re

sultados equivalentes al caso en que a priori todos los modelos tengan la misma probabilidad. El tratamiento consiste en comparar los modelos a través de los factores de Bayes. Los factores se definen como el cociente de los nomios a posteriori entre los nomios a priori. Más claramente se puede expresar como sigue:

$$\begin{aligned}
 \eta \left(\frac{M_j}{M_k} \mid \underline{X}_T \right) &= \frac{P(M_j \mid \underline{X}_T)}{P(M_k \mid \underline{X}_T)} \\
 &= \frac{P(\underline{X}_T \mid M_j) P(M_j)}{P(\underline{X}_T \mid M_k) P(M_k)} \\
 &= \frac{P(\underline{X}_T \mid M_j)}{P(\underline{X}_T \mid M_k)} * \frac{P(M_j)}{P(M_k)} \\
 &= F \left(\frac{M_j}{M_k} \mid \underline{X}_T \right) \eta \left(\frac{M_j}{M_k} \right)
 \end{aligned}$$

en donde $\eta \left(\frac{M_j}{M_k} \right)$ son los nomios a priori y naturalmente,

$F \left(\frac{M_j}{M_k} \mid \underline{X}_T \right)$ es el factor de Bayes. Así pues, Smith propone la comparación de los modelos a través de las cantidades

$$F \left(\frac{M_j}{M_k} \mid \underline{X}_T \right) = \frac{P(\underline{X}_T \mid M_j)}{P(\underline{X}_T \mid M_k)}$$

Es importante insistir en que este procedimiento no es el que se desprende del conjunto de principios de coherencia que determinan el enfoque Bayesiano, pero puede ser de utilidad en algunos casos. El tratamiento puede ser generalizado al caso en que se considere la posibilidad que ocurran más de un cambio.

Smith ilustra el procedimiento propuesto con datos binomiales en los siguientes términos. Sea $M_{(r_1, \dots, r_k)}$ el modelo con k cambios que ocurren en los puntos r_1, r_2, \dots, r_k . Si se definen $r_0 \equiv 1$ y $r_{k+1} \equiv T$ y se considera que para $r_{j-1} < i < r_j$, $f(X_i, \theta_j) = \binom{m_i}{X_i} \theta_j^{S_j} (1-\theta_j)^{f_j}$ en donde

$$S_j = \sum_{i=r_{j-1}+1}^{r_j} X_i, \quad f_j = \sum_{i=r_{j-1}+1}^{r_j} X_i - S_j$$

y m_i representa el número de observaciones entre r_{j-1} y r_j , entonces, si a priori se supone independencia entre los parámetros $\theta_1, \dots, \theta_{k+1}$, una posible representación de la función de densidad conjunta a priori es,

$$f_0(\theta_1, \dots, \theta_{k+1}) = \prod_{j=1}^{k+1} \theta_j^{\alpha_j - 1} (1-\theta_j)^{\beta_j - 1} / B(\alpha_j, \beta_j)$$

con $B(\alpha_j, \beta_j)$ la función Beta evaluada en α_j, β_j .

Con esta estructura se tiene que:

$$P(\underline{X}_T | M_{(r_1, \dots, r_k)}) = \prod_{i=1}^T \binom{m_i}{X_i} \int \dots \int \prod_{j=1}^{k+1} \theta_j^{S_j} (1-\theta_j)^{f_j} \\ \cdot f_0(\theta_1, \dots, \theta_{k+1}) d\theta_1, \dots, d\theta_{k+1}$$

$$\begin{aligned}
&= \prod_{i=1}^T (x_i^{n_i}) \int \dots \int_{j=1}^{k+1} \theta_j^{S_j + \alpha_j - 1} (1 - \theta_j)^{F_j + \beta_j - 1} / B(\alpha_j, \beta_j) d\theta_1, \dots, d\theta_{k+1} \\
&= \prod_{i=1}^T (x_i^{n_i}) \prod_{j=1}^{k+1} \left\{ \int \theta_j^{S_j + \alpha_j - 1} (1 - \theta_j)^{F_j + \beta_j - 1} / B(\alpha_j, \beta_j) d\theta_j \right\} \\
&= \prod_{i=1}^T (x_i^{n_i}) \prod_{j=1}^{k+1} \left\{ [\beta(\alpha_j, \beta_j)]^{-1} \int \theta_j^{S_j + \alpha_j - 1} (1 - \theta_j)^{F_j + \beta_j - 1} d\theta_j \right\} \\
&= \prod_{i=1}^T (x_i^{n_i}) \prod_{j=1}^{k+1} \left\{ \frac{B(S_j + \alpha_j, F_j + \beta_j)}{B(\alpha_j, \beta_j)} \right\}
\end{aligned}$$

De tal manera que si se denota por M_0 de nuevo el modelo sin cambio, se puede verificar que

$$F \left(\frac{M_0}{N(r_1, \dots, r_k)} \mid \underline{X}_T \right) = \frac{B(\alpha_1 + \sum_{i=1}^T X_i, \beta_1 + \sum_{i=1}^T m_i - \sum_{i=1}^T X_i)}{B(\alpha_1, \beta_1)}$$

$$\left[\prod_{j=1}^{k+1} \frac{B(\alpha_j + S_j, \beta_j + F_j)}{B(\alpha_j, \beta_j)} \right]^{-1}$$

Asimismo se tiene que:

$$F\left(\frac{M(r_1, \dots, r_k)}{M(t_1, \dots, t_n)} \mid \underline{X}_T\right) = \frac{F\left(\frac{M_0}{M(t_1, \dots, t_n)} \mid \underline{X}_T\right)}{F\left(\frac{M_0}{M(r_1, \dots, r_k)} \mid \underline{X}_T\right)}$$

De forma que basta con poder calcular los factores con respecto al modelo sin cambios para poder comparar todos los modelos. Más aún, Smith propone una aproximación para el cálculo de los factores en el caso en que $k=1$ y $\alpha_j = \beta_j = 1$ en base a la aproximación de Stirling para coeficientes binomiales. En el ejemplo que se emplea el autor para ilustrar el procedimiento, se presentan las probabilidades a posteriori para diferentes números de puntos de cambio y en base a ellas se explora el caso con $k=2$ puesto que su probabilidad es prácticamente 1. Como segunda etapa, se examinan las probabilidades a posteriori de los diversos posibles pares de puntos de cambio y con esa información se sugiere el modelo que parece ser, el más adecuado. Resulta interesante mencionar que en el ejemplo numérico Smith solo emplea al final y de manera accesoria los Factores de Bayes para tratar de reafirmar la evidencia encontrada.

Un trabajo más, es el debido a Hsu [50], que guarda similitud con la ya expuesta contribución de Kander y Zacks [63], no solo porque parte de los principales resultados obtenidos por estos autores sino que, al igual que ellos y como consecuencia de que se basa en sus resultados, parte de un planteamiento Bayesiano pero manipula la regla decisión para la prueba de las hipótesis relevantes en el problema, desde una perspectiva Frecuentista, centrando su trabajo en la obtención de una aproximación asintótica para la distribución de la estadística de prueba. Concretamente, Hsu considera el problema de contar con una sucesión de variables aleatorias independientes X_1, \dots, X_T tales que para $i=1, 2, \dots, r$, la distribución de X_i es Gama con parámetros α y β_0 , mientras que para $i=r+1, \dots, T$, X_i tiene una distribución Gama de parámetros α y β_1 , donde α es conocido, mientras que β_0 , β_1 y r son parámetros desconocidos. Reexpresando a β_1 como $\beta_0 + \delta$, Hsu recupera un caso particular del esquema planteado en general, por Kander y Zacks para el caso de distribuciones de la familia exponencial de un parámetro. Esto es, obtiene que el nomio H_1 contra H_0 , asignando igual probabilidad a cada uno de los posibles casos que componen a H_1 , tiene asintóticamente, cuando δ/β_0 tiende a cero y δ es positivo, la siguiente expresión:

$$\eta(H_1, H_0) = 1 + \left(\frac{\delta}{\beta_0}\right) \left[\frac{1}{(T-1)\beta_0} \sum_{i=1}^T (i-1) X_i - \frac{\alpha T}{2} \right]$$

que es una función monótona de

$$T(X_1, \dots, X_T) \equiv \frac{1}{\beta_0} \sum_{i=1}^T (i-1) X_i$$

de modo que, en el caso en que β_0 fuese conocido, la hipótesis H_0 se rechazaría, en ausencia de una función de pérdida y de acuerdo con el tratamiento Bayesiano si esta cantidad fuese, en opinión del decisor, suficientemente grande. En el esquema planteado por Hsu, sin embargo, el valor β_0 se supone desconocido de manera que la decisión no puede llevarse a cabo de esta forma. Más aún, si la función $T(X_1, \dots, X_T)$ se manipula como una estadística de prueba en el sentido frecuentista, tal como propone Kander y Zacks, naturalmente su distribución no queda totalmente especificada bajo H_0 puesto que β_0 es desconocido. No obstante siguiendo la misma línea de esos autores, Hsu propone para este caso la sustitución de β_0 por su estimador de máxima verosimilitud

$$\hat{\beta}_0 = \frac{T}{\sum_{i=1}^T X_i} / T\alpha$$

de tal manera que el criterio, que vale la pena recordar no es estrictamente Bayesiano, consiste en rechazar H_0 si la es-

tadística

$$\bar{T}(X_1, \dots, X_T) \equiv T \alpha \frac{\sum_{i=1}^T (i-1) X_i}{\sum_{i=1}^T X_i}$$

o equivalentemente,

$$\hat{T}(X_1, \dots, X_T) \equiv \frac{\sum_{i=1}^T (i-1) X_i}{\sum_{i=1}^T (T-1) X_i}$$

es suficientemente grande.

En estas condiciones, visualizando el problema desde una perspectiva frecuentista, Hsu establece la distribución asintótica de \hat{T} . Básicamente argumenta que \hat{T} es una combinación lineal de variables Dirichlet que toma valores en el intervalo $(0, 1]$ y bajo H_0 , es simétrica respecto al valor de su esperanza que es igual a 0.5. Más aún, afirma que bajo H_0 , la distribución de \hat{T} tiende asintóticamente a una Normal. De esta forma propone el empleo de la estadística

$$T^* \equiv (\hat{T} - 0.5) / \{ \text{Var}(\hat{T}) \}^{1/2}$$

que asintóticamente tiene una distribución Normal estándar. Basado en el hecho de que T^* se obtiene a partir de un cociente de verosimilitudes aproximado para valores pequeños de δ , el

autor hace notar que el criterio para rechazar H_0 que consiste en descartar la hipótesis si el valor observado de la estadística es suficientemente grande, es el localmente más potente si $\delta > 0$ tiende a cero. Análogamente, en el caso de una prueba de dos colas, el criterio de rechazar H_0 si $|T^*|$ excede una constante preestablecida es el localmente más potente si $|\delta|$ tiende a cero. Así pues, Hsu propone un criterio para detección de puntos de cambio en el parámetro de escala de sucesiones de variables aleatorias Gama. Tal criterio se basa en una estadística que bajo la hipótesis nula y asintóticamente, tiene una distribución Normal estándar. El problema que en este punto es comentado por el autor es la evaluación del criterio cuando el tamaño muestral es moderado. Para complementar el estudio, Hsu calcula una tabla de cuantiles aproximados para la distribución de T^* en el caso de variables exponenciales y $\chi^2_{(1)}$, obteniendo que para prácticamente todos los tamaños de muestra considerados, la aproximación Normal resulta satisfactoria. Partiendo de ese hecho, el autor efectúa una serie de simulaciones para esos dos modelos, en que modifica la posición del punto de cambio y la magnitud relativa $(\beta_0 + \delta) / \beta_0$, encontrando que para un tamaño muestral $T=30$, la potencia de la prueba es mayor para aquellos casos en que el punto de cambio se localiza en la parte media de la sucesión y que además, la potencia calculada es uniformemente mayor

cuando se consideran variables exponenciales que cuando se trata de variables $X_{(1)}^2$.

Hsu comenta en el trabajo revisado, la relación que guarda la situación por él discutida con los problemas de prueba de igualdad de varianzas en general y de igualdad de tasas de ocurrencia en procesos Poisson. Este tipo de problemas resultan de interés en la aplicación de algunos modelos de uso común y las técnicas usualmente empleadas para detectarlos requieren, como regla general, de la formación de grupos de observaciones en base al criterio del interesado. El autor afirma que esas técnicas y el procedimiento que propone pueden emplearse como herramientas alternativas en virtud de que si bien algunas de las técnicas para probar igualdad de varianzas son robustas ante la falta de Normalidad mientras que no puede afirmarse lo mismo respecto a T^* , en los casos en que esta estadística pueda ser empleada, no se requiere un agrupamiento de los datos para probar la hipótesis de interés.

La última contribución que se revisará en esta sección es la debida a Díaz, [28], que también aborda el problema de detección de puntos de cambio en el parámetro de escala en sucesiones de variables aleatorias independientes con distribución Gama. Esto es, trata el mismo problema planteado por Hsu pero ahora sí, desde una perspectiva totalmente Bayesiana.

El autor considera el caso en que a lo más ocurrió un cambio y se ocupa de probar la hipótesis nula de no cambio contra la alternativa de que un cambio tuvo lugar en la sucesión. El problema se plantea en los siguientes términos:

Sea $X_1, \dots, X_{r-1}, X_r, X_{r+1}, \dots, X_T$ una sucesión de variables aleatorias independientes tales que X_i se distribuye Gama (θ_1, k) para $i=1, 2, \dots, r$ mientras que X_i se distribuye Gama (θ_2, k) para $i=r+1, \dots, T$. Esto es, se tiene que, la función de densidad de X_i está dada por:

$$f(X_i; \theta_1, k) = \frac{X_i^{k-1}}{\theta_1^k \Gamma(k)} e^{-X_i/\theta_1} ; \theta_1 > 0, k > 0, X_i > 0$$

para $i=1, 2, \dots, r$ en tanto que

$$f(X_i; \theta_2, k) = \frac{X_i^{k-1}}{\theta_2^k \Gamma(k)} e^{-X_i/\theta_2} ; \theta_2 > 0, k > 0, X_i > 0$$

para $i=r+1, \dots, T$

Díaz supone que k es un parámetro conocido pero θ_1 y θ_2 son desconocidos. El objetivo inmediato, como es común en el enfoque Bayesiano, es la determinación de la función de probabilidades del parámetro de interés que en este caso es r . En

el artículo de referencia se utilizan distribuciones a priori conjugadas para θ_1 y θ_2 y adicionalmente se supone independencia entre estos dos parámetros y entre ellos y τ para el cual se asigna una función de densidad de probabilidades

$$f_0(\tau) = \begin{cases} P & \text{si } \tau=T \\ \frac{1-P}{T-1} & \text{si } \tau=1, 2, \dots, T-1 \end{cases}$$

De acuerdo con el planteamiento del problema se tiene que la función de verosimilitud de τ , θ_1 y θ_2 está dada por:

$$f(X_T | \tau, \theta_1, \theta_2) \propto \begin{cases} \theta_1^{-T k} \exp \left\{ -\sum_{i=1}^T X_i / \theta_1 \right\} ; & \tau = T \\ \theta_1^{-\tau k} \theta_2^{-(T-\tau)k} \exp \left\{ -\sum_{i=1}^{\tau} X_i / \theta_1 - \sum_{i=\tau+1}^T X_i / \theta_2 \right\} ; & \tau \neq T \end{cases}$$

Por su parte, la distribución conjunta a priori de τ, θ_1 y θ_2 tiene la siguiente expresión:

$$f_0(\tau, \theta_1, \theta_2) \propto \begin{cases} \frac{P \exp \{-1/\theta_1 a_1\}}{a_1^{r_1} \Gamma(r_1) \theta_1^{r_1+1}} ; & \tau = T \\ \frac{(1-P) \exp\{-1/\theta_1 a_1\} \exp\{-1/\theta_2 a_2\}}{(T-1) a_1^{r_1} \Gamma(r_1) \theta_1^{r_1+1} a_2^{r_2} \Gamma(r_2) \theta_2^{r_2+1}} ; & \tau \neq T \end{cases}$$

De modo que la distribución posteriori queda determinada por la siguiente relación:

$$f(r, \theta_1, \theta_2 | \underline{X}_T) = f(\underline{X}_T | r, \theta_1, \theta_2) f_0(r, \theta_1, \theta_2)$$

Así pues para $r=T$,

$$f(T, \theta_1 | \underline{X}_T) = p \theta_1^{-(Tk+r_1+1)} \exp \left\{ -\left(\sum_{i=1}^T X_i + 1/\alpha_1 \right) / \theta_1 \right\}$$

mientras que para $r \neq T$,

$$f(r, \theta_1, \theta_2 | \underline{X}_T) = \frac{(1-p)}{(T-1)} \theta_1^{-(rk+r_1+1)} \theta_2^{-(Tk-rk+r_2+1)} \exp \left\{ -\left(\sum_{i=1}^T X_i - 1/\alpha_1 \right) / \theta_1 \right\}$$

$$\cdot \exp \left\{ -\left(\sum_{i=r+1}^T X_i - 1/\alpha_2 \right) / \theta_2 \right\} (\Gamma(r_2))^{-1} \alpha_2^{-r_2}$$

de manera que:

$$\begin{aligned} f(T | \underline{X}_T) &= \int_0^\infty p \theta_1^{-(Tk+r_1+1)} \exp \left\{ -\left(\sum_{i=1}^T X_i - 1/\alpha_1 \right) / \theta_1 \right\} d\theta_1 \\ &= p \int_0^\infty \theta_1^{-(Tk+r_1+1)} \exp \left\{ -\left(\sum_{i=1}^T X_i - 1/\alpha_1 \right) / \theta_1 \right\} d\theta_1 \end{aligned}$$

$$= \frac{P \Gamma(Tk+r_1)}{\left(\sum_{i=1}^T X_i + 1/\alpha_1\right)^{Tk+r_1}} \int_0^{\infty} \left[\frac{\left(\sum_{i=1}^T X_i + 1/\alpha_1\right)^{Tk+r_1} \theta_1^{-(Tk+r_1+1)}}{\Gamma(Tk+r_1)} \cdot \exp\left\{-\left(\sum_{i=1}^T X_i + 1/\alpha_1\right) \theta_1\right\} \right] d\theta_1$$

$$= \frac{P \Gamma(Tk+r_1)}{\left(\sum_{i=1}^T X_i + 1/\alpha_1\right)^{Tk+r_1}}$$

Por otra parte se tiene que de forma análoga, completando las expresiones en θ_1 y θ_2 para obtener densidades Gama inversas, la distribución marginal a posteriori de $r \neq T$ es tal que:

$$f(r | \frac{X}{T}) \propto \frac{(1-P) \Gamma(rk+r_1) \Gamma(Tk-rk+r_2) \alpha_2^{-r_2} \{\Gamma(r_2)\}^{-1}}{(T-1) \left(\sum_{i=1}^r X_i + 1/\alpha_1\right)^{rk+r_1} \left(\sum_{i=r+1}^T X_i + 1/\alpha_2\right)^{Tk-rk-r_2}}$$

El autor indica que en el caso en que se utilicen distribuciones iniciales de referencia para θ_1 y θ_2 , de la forma:

$$f_0(\theta_i) = \theta_i^{-1} \quad ; \quad i=1,2.$$

la distribución posteriori marginal de r está dada por:

$$f(r|X_T) = \begin{cases} P \Gamma(T_k) \left\{ \sum_{i=1}^T X_i \right\}^{-T_k} & \text{Si } r = T \\ \frac{(1-P) \Gamma(r_k) \Gamma(T_k - r_k)}{(T-1) \left\{ \sum_{i=1}^r X_i \right\}^{r_k} \left\{ \sum_{i=r+1}^T X_i \right\}^{T_k - r_k}} & \text{Si } r \neq T \end{cases}$$

Un procedimiento Bayesiano informal, que no utiliza función de pérdida alguna para probar las hipótesis $H_0: r=T$ vs. $H_1: r \neq T$, (no cambio contra un cambio) es ilustrado con dos ejemplos. El procedimiento consiste, de acuerdo con la proposición de Lindley [59], que en cierta forma está sugerida por la relación que en el enfoque clásico guardan algunas regiones de rechazo con los intervalos de confianza, en rechazar H_0 si el punto $r=T$ no está contenido en todas las regiones de máxima probabilidad de tamaño relativamente grande. En los dos ejemplos tratados se introducen artificialmente los puntos de cambio y se ensaya con dos distintas magnitudes del cambio en el parámetro θ . Los resultados sugieren que entre mas grande sea la magnitud del cambio o mayor sea el número de observaciones a ambos lados del punto de cambio, más facilmente se obtiene evidencia para rechazar H_0 .

11.3 DISCUSIÓN

De acuerdo al material revisado en este capítulo parece claro que en relación al problema de puntos de cambio en sucesiones de variables aleatorias, resulta de interés establecer primero si efectivamente tuvo lugar un cambio para que en el caso de que así sea, proceder a realizar inferencias sobre el punto donde se produjo el cambio así como sobre cualesquiera otros parámetros no especificados que resulten de interés.

De los trabajos que se han presentado, los primeros en la literatura son los de Page y Hinkley que abordan el problema desde una perspectiva Frecuentista y emplean el método de máxima verosimilitud tanto para decidir si es razonable suponer que se verificó un cambio como para estimar el punto de cambio. Los trabajos de estos dos autores presentan diversas similitudes. Primero, sus resultados más importantes se obtienen para el caso en que sólo hay un punto de cambio y éste es el único parámetro desconocido. Por otro lado, la caracterización de los criterios de prueba así como en su caso de los estimadores, se realiza en base a aproximaciones asintóticas. Estas aproximaciones son relativamente complejas, como consecuencia de que resultados más generales no pueden aplicarse debido a la naturaleza del parámetro de interés. De esta forma, el problema para una situación con un tamaño moderado de muestra no es resuelto en su totalidad. Una característica interesante de estos trabajos es que exhiben muy explícitamente la estrecha re-

lación que guardan los problemas de estimación, pruebas de hipótesis y selección de modelos.

Por lo que concierne a los trabajos revisados que emplean el enfoque Bayesiano, ocurre que fundamentalmente se ocupan, como es de esperarse en este enfoque, de obtener una distribución final que describa el conocimiento que se posee sobre las cantidades de interés.

Aún en los planteamientos más simples se considera que el punto de cambio no es el único parámetro desconocido y en algunas contribuciones se examinan casos donde se contempla la posibilidad de más de un cambio e inclusive se aborda el problema de la predicción de observaciones en presencia de posibles puntos de cambio. La formulación general del procedimiento para obtener la distribución final es extremadamente simple y exacta para cualquier tamaño de muestra aunque debe de tomarse en cuenta que para algunos casos específicos, la evaluación de las funciones de densidad involucradas puede requerir de aproximaciones numéricas.

En términos generales, el enfoque Bayesiano permite un planteamiento y tratamiento único, simple y claro que para los casos de las distribuciones más usuales, si se modela el conocimiento apriori adecuadamente, produce distribuciones finales que

son fácilmente interpretables aunque, como ya se mencionó, pueden encontrarse algunas dificultades menores de cómputo en la evaluación de las referidas distribuciones finales. Por su parte, el enfoque Frecuentista no produce resultados exactos para el tratamiento del problema si no solo aproximaciones asintóticas. Asimismo, y a pesar de que en los dos trabajos revisados se emplea el principio de máxima verosimilitud, no puede garantizarse que este sea el único procedimiento que se aplica para este tipo de problemas.

Un comentario especial merecen las contribuciones de Chernoff y Zacks, Kander y Zacks y la de Hsu, en donde se parte de un razonamiento Bayesiano pero las reglas de decisión que finalmente se proponen son examinadas desde una perspectiva Frecuentista. Es claro que las contribuciones de estos autores no pueden ser clasificadas estrictamente como Bayesianas o Frecuentistas. Este hecho, desde un punto de vista metodológico, puede ser criticado en virtud de que las suposiciones que sobre los parámetros se hacen en ambos enfoques no son compatibles. Un punto a favor de este tipo de trabajos, que algunos estadísticos han dado en llamar cuasibayesianos, podría ser argumentado si procediendo con ese espíritu pragmático se lograra conjuntar las ventajas prácticas de ambos enfoques y disminuir las dificultades que por separado presentan. Sin

embargo, en el caso del problema que se discute, la principal desventaja del enfoque Frecuentista persiste en tanto se requieren aproximaciones asintóticas para casi cualquier situación específica y simultáneamente se pierde la ventaja Bayesiana de contar con un procedimiento único de resultados exactos.

CAPITULO III

Modelos de Regresión
con puntos de cambio.

III. 1 RESULTADOS DEL ENFOQUE FRECUENTISTA.

El análisis de modelos de regresión con puntos de cambio ha sido extensamente tratado en la literatura Estadística. De hecho parece existir un número de contribuciones sensiblemente mayor para el tratamiento de este problema que en relación al de puntos de cambio en sucesiones de variables aleatorias. Esta situación no resulta sorprendente si se toma en cuenta el impresionante éxito que desde sus orígenes, han tenido los modelos de regresión para la caracterización de los más diversos fenómenos.

En la búsqueda bibliográfica realizada para desarrollar este trabajo, las primeras contribuciones registradas para el tratamiento de puntos de cambio en modelos de regresión desde un punto de vista frecuentista, son las debidas a Quandt, [12] y [73]. En su primer trabajo, este autor introduce la situación discutiendo la aplicación que tiene en el campo del análisis económico. Así pues, menciona que en ocasiones, la relación que guardan algunas variables económicas como por ejemplo ingreso y consumo, puede ser satisfactoriamente descrita, bajo ciertas condiciones, por modelos lineales. Sin embargo, tales

modelos pueden cambiar si las condiciones del entorno económico se modifican. Una forma posible de tomar en cuenta estas modificaciones podría ser la inclusión en el modelo lineal de algunas variables que describan el comportamiento de los factores que presumiblemente afectan la relación. Desgraciadamente, en muchos casos no es posible identificar de antemano tales factores o, cuando esto se logra, el efecto no siempre es linealizable, de modo que no puede incluirse en el modelo. En estas condiciones, es necesario considerar el modelo lineal original y examinarlo para determinar si sufre algún cambio que, en general, será atribuido al efecto de algún factor exógeno. La formulación estadística del problema se plantea, en el caso más sencillo, de regresión lineal simple, con un solo punto de cambio, como sigue:

Sean Y_1, Y_2, \dots, Y_T observaciones tales que

$$Y_i = a_1 + b_1 X_i + \epsilon_{1i}; \quad i=1, 2, \dots, \tau$$

$$Y_i = a_2 + b_2 X_i + \epsilon_{2i}; \quad i=\tau+1, \dots, T$$

en donde, a_1, a_2, b_1 y b_2 son parámetros desconocidos, ϵ_{1i} y ϵ_{2i} son errores independientes con distribución Normal de

media cero y varianza σ_1^2 y σ_2^2 respectivamente. Los valores X_1, \dots, X_T son valores fijos y conocidos. Esto es, las parejas (X_i, Y_i) para $i=1, 2, \dots, t$ se ajustan a un modelo de regresión lineal simple con parámetros a_1 y b_1 y con errores cuya varianza es σ_1^2 mientras que las parejas para $i=t+1, t+2, \dots, T$ se ajustan a otro modelo de parámetros a_2, b_2 y σ_2^2 . El primer problema que aborda Quandt es la estimación de r . Para tal efecto propone el empleo del método de máxima verosimilitud como sigue. Para un valor fijo de r , por ejemplo t , la función de verosimilitud tiene la siguiente expresión:

$$L(t, a_1, a_2, b_1, b_2, \sigma_1^2, \sigma_2^2) = \prod_{i=1}^t (2\pi\sigma_1^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_1^2}(Y_i - b_1 X_i - a_1)^2\right\} \\ \prod_{i=t+1}^T (2\pi\sigma_2^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_2^2}(Y_i - b_2 X_i - a_2)^2\right\}$$

Tomando logaritmos se tiene que,

$$\ln L = -\frac{T}{2} \ln(2\pi) - t \ln \sigma_1 - (T-t) \ln \sigma_2 \\ - \frac{1}{2\sigma_1^2} \sum_{i=1}^t (Y_i - b_1 X_i - a_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=t+1}^T (Y_i - b_2 X_i - a_2)^2$$

en donde, considerando a t fijo y maximizando respecto a $b_1, a_1, b_2, a_2, \hat{\sigma}_1^2$ y $\hat{\sigma}_2^2$ se obtienen para estos parámetros los estimadores usuales de mínimos cuadrados considerando los dos modelos de regresión por separado. El valor así maximizado de la verosimilitud, que naturalmente depende de t , tiene la siguiente expresión:

$$\ell(t) = \frac{T}{2} \ln(2\pi) - t \ln \hat{\sigma}_1^2 - (T-t) \ln \hat{\sigma}_2^2 - \frac{T}{2}$$

en donde:

$$\hat{\sigma}_1^2 = \frac{1}{t} \sum_{i=1}^t (Y_i - \hat{a}_1 - \hat{b}_1 X_i)^2; \quad \hat{\sigma}_2^2 = \frac{1}{T-t} \sum_{i=t+1}^T (Y_i - \hat{a}_2 - \hat{b}_2 X_i)^2$$

El problema de encontrar el estimador de máxima verosimilitud de τ consiste entonces en determinar el valor que maximiza $\ell(t)$. La naturaleza del parámetro τ , al igual que en el caso de sucesiones de variables aleatorias, no permite la aplicación de la técnica usual de diferenciación. Así pues, el procedimiento que propone Quandt es el cálculo de $\ell(t)$ para $t=3, \dots, T-3$, a fin de determinar el valor donde se alcanza el máximo de verosimilitud. Es importante observar que la evaluación de $\ell(t)$ para cada valor de t implica el ajuste de un modelo de regresión lineal simple con las

primeras t observaciones y otro por separado para la $T-t$ restantes. Es claro que en estas condiciones, el rango de t se reduce a los valores para los cuales tiene sentido ajustar una recta por mínimos cuadrados en base a t y $T-t$ observaciones. Sin discutir las propiedades del estimador de τ así producido, Quandt procede a plantearse la inquietud de desarrollar un criterio de prueba para la hipótesis de que efectivamente ocurrió un cambio. El autor propone el empleo del cociente de verosimilitudes para tal efecto.

Naturalmente, el máximo de la versosimilitud bajo la hipótesis nula de no cambio $L(H_0)$, se obtiene ajustando una sola recta a todos los datos mientras que el correspondiente a la hipótesis alternativa $L(H_1)$ se deriva mediante el procedimiento descrito por Quandt para la estimación de τ . Es fácil verificar que el cociente de interés tiene la siguiente expresión:

$$\lambda = \frac{\hat{\sigma}_1^{t^*} \hat{\sigma}_2^{T-t^*}}{\hat{\sigma}^T}$$

en donde, t^* es el valor estimado de τ por máxima verosimilitud y

$\hat{\sigma}^2$ es el estimador correspondiente de la varianza en el régimen de regresión ajustado con todas las observaciones.

La distribución exacta de λ bajo la hipótesis nula no ha sido de terminada pero Quandt sugiere la utilización de la aproximación Ji-cuadrada para $-2 \ln \lambda$. El propio autor indica que las condiciones bajo las cuales tal aproximación es adecuada no se satisfacen en este caso, debido esencialmente a que t es una variable discreta. Sin embargo, conjetura que el resultado es apropiado aún en esas circunstancias ya que en su opinión, es de esperarse que a medida que T se incremente las distorsiones debidas a la naturalidad de t sean de menor importancia. Alternativamente, Quandt propone en el mismo artículo, otro procedimiento para el caso en que el tamaño de muestra sea pequeño. Este consiste en dividir el conjunto de observaciones en dos grupos y probar la hipótesis de que las regresiones ajustadas son iguales. Naturalmente, en el contexto del problema la división de las observaciones en dos grupos debe hacerse de tal forma que se obtengan conjuntos de observaciones contiguas respecto al orden de registro. Esto es, los grupos deben ser de la forma $\{(X_i, Y_i) \mid i \leq t\}$, $\{(X_i, Y_i) \mid i > t\}$ para al

gún valor de t que permita el ajuste de las dos regresiones por separado. Con una partición de este tipo, Quandt propone la utilización de la prueba descrita por Rao [78], para probar la igualdad de las regresiones. Literalmente, indica: "Denotando por S_0 la suma de sus cuadrados de desviaciones de las dos regresiones separadas, ajustadas a los dos grupos de observaciones y por S_1 la suma de cuadrados de desviaciones de una regresión común basada en todas las observaciones, la cantidad $(S_1 - S_0)/S_0$ tiene una distribución F con 4 y $T-8$ grados de libertad".

En relación a esta afirmación es interesante hacer notar que efectivamente, en la referencia mencionada, Rao presenta una prueba de igualdad de coeficientes de regresión pero que supone, de antemano, igualdad entre las respectivas varianzas de modo que en realidad la proposición de Quandt no constituye, en el sentido más amplio, una alternativa para el caso de tamaño de muestra pequeña. Por otra parte y como se hace notar en otro trabajo, Mendoza [63], que también se presenta en este capítulo, la construcción de la estadística de prueba es incorrecta ya que el cociente que tiene una distribución F, de acuerdo a Rao, es el que conside

ra los cuadrados medios correspondientes. Asimismo, los grados de libertad no son los que afirma el autor. En la referencia que menciona, efectivamente se utilizan 4 y T-8 grados de libertad pero en aquel caso se tiene un modelo de regresión lineal múltiple con tres variables independientes. Fácilmente se puede verificar que los grados de libertad correctos para un modelo de regresión lineal simple son 2 y T-4.

Aparentemente sin haber detectado estos errores, Quandt comenta el problema adicional que representa la elección del punto para dividir las observaciones en dos grupos. Apunta el hecho de que resulta natural utilizar el valor estimado de τ por máxima verosimilitud aunque este procedimiento implica la prueba de una hipótesis sugerida por el mismo conjunto de datos, de suerte que el criterio resulta en el rechazo de la hipótesis nula más a menudo de lo esperado. Una alternativa en este sentido es, de acuerdo al autor, la utilización de un valor determinado de forma exógena a los datos, por ejemplo, T/2 si T es par o bien algunos de los valores (T+1)/2 o (T-1)/2 si T es impar. El problema de esta proposición, según lo comenta Quandt, es que probablemente al

guno de los conjuntos puede contener observaciones que en realidad pertenecen al otro régimen de regresión. Esto es, existe un fenómeno de contaminación de los datos. Una propuesta final que hace el autor es la posible supresión de algunas de las observaciones centrales de modo de reducir el riesgo de la contaminación. En cualquier forma es claro que la potencia de estos dos últimos procedimientos depende de si realmente el punto de cambio se encuentra cerca de las observaciones centrales.

Quandt termina su trabajo aplicando sus ideas a un conjunto de datos generados por él, que posteriormente ha sido utilizado por diversos autores para ilustrar otras técnicas para este tipo de problemas. El autor genera veinte observaciones de las cuales las primeras doce siguen la siguiente relación:

$$Y = 2.5 + 0.7 X + U_1$$

mientras que las ocho restantes fueron generadas a partir del modelo

$$Y = 5 + 0.5 X + U_2$$

en donde, tanto U_1 como U_2 son errores independientes Normales de

media cero y varianza uno.

El estimador de máxima verosimilitud para τ en este caso resulta ser $t^*=12$ que coincide con el valor verdadero del punto de cambio. Tanto el criterio de distribución asintótica como el correspondiente a un tamaño de muestra pequeña para la prueba de la hipótesis de no cambio son aplicadas por el autor en este conjunto de datos. Con el criterio asintótico, la estadística de prueba tiene un valor de 14.610 que es significativo al 1% comparado con tablas de Ji cuadrada con 4 grados de libertad. En el caso del otro criterio, el autor reporta que solo es significativo al 50%. Sin embargo, debe tomarse en cuenta que tanto la expresión para el cálculo de la estadística de prueba como los grados de libertad son incorrectos. En el ya mencionado trabajo de Mendoza [6], se corrigen los cálculos y se encuentra que la estadística resulta significativa al 1% si se utiliza el valor estimado de τ por máxima verosimilitud, $t^*=12$, para dividir las observaciones y es significativa al 5% si se utiliza el valor $t=T/2=10$. En resumen, en este primer trabajo. Quandt propone la estimación del punto de cambio mediante máxima verosimilitud pero no examina sus propiedades.

Propone dos criterios para probar la hipótesis de no cambio sobre el primero de las cuales conjetura que bajo la hipótesis H_0 la tiene una distribución que se puede aproximar asintóticamente, por una J_i cuadrada. El otro criterio que no requiere de un tamaño de muestra grande es construido con algunos errores evidentes. Al aplicar estos criterios a un conjunto de datos especialmente generado para este propósito y que tiene un punto de cambio, se obtiene que el valor estimado del punto de cambio coincide con un verdadero valor. Los dos criterios producen resultados significativos una vez que el segundo es corregido.

En un artículo posterior, [73], Quandt retoma el problema examinando, como un primer punto, la conjetura de que $-2 \ln \lambda$ tiene una distribución J_i cuadrada con cuatro grados de libertad. Para ello procede a realizar un estudio através de simulación que le permite construir funciones de distribución empíricas para esta variable. Considera tres casos, $T=20, 40$ y 60 . En todos ellos la variable independiente puede tomar veinte diferentes valores; $x=0.05, 0.10, 0.15, \dots, 1.00$. Cada valor es utilizado una vez en el primer caso, dos en el segundo y tres en el tercero. Los errores son indepen-

dientes y Normales de media cero y varianza uno. Quandt simula 200 conjuntos de datos para cada caso y tabula una tabla de frecuencias de $-2 \leq \lambda$ para cada uno. Antes de probar el ajuste de una distribución Ji cuadrada con los cuatro grados de libertad, el autor utiliza tres diferentes pruebas para verificar si las distribuciones empíricas pueden considerarse iguales. Como resultado reporta que la hipótesis de igualdad se rechaza con los tres criterios con niveles de significancia descriptivos menores a 0.007. En vista de este resultado, prueba el ajuste a la distribución Ji cuadrada por separada y reporta que la hipótesis se rechaza en todos los casos con un nivel de significancia de 0.0001. Ante esta evidencia, Quandt descarta la aproximación Ji cuadrada, indica que la distribución de $-2 \leq \lambda$ depende del valor de T y sugiere el empleo de tablas empíricas como las generadas por él para la aplicación del criterio propuesto.

Otros criterios para el caso de muestras pequeñas son propuestas por el autor en el mismo trabajo. La idea general es, como en el trabajo anterior, dividir el conjunto de T observaciones en dos grupos de acuerdo al valor de t^* , el estimador de máxima verosimilitud.

militud de τ y estimar los parámetros de los dos modelos de regresión separados. En base a ellos, el autor propone la construcción de un tipo especial de residuales, que podrían llamarse cruzados, para probar la hipótesis relevante. El procedimiento es el siguiente:

Se calculan las regresiones estimadas

$$\hat{Y}_i = \hat{a}_1 + \hat{b}_1 X_i ; \quad i=1,2,\dots,t^* ,$$

$$\hat{Y}_i = \hat{a}_2 + \hat{b}_2 X_i ; \quad i=t^*+1,\dots,T$$

Y los residuales cruzados se definen como:

$$r_{1i} = \hat{a}_2 + \hat{b}_2 X_i - Y_i ; \quad i=1,2,\dots,t^* ,$$

$$r_{2i} = \hat{a}_1 + \hat{b}_1 X_i - Y_i ; \quad i=t^*+1,\dots,T$$

Quandt afirma que bajo la hipótesis nula de no cambio, $r_{11}, r_{12}, \dots, r_{1t^*}$ forman una colección de variables aleatorias Normales independientes. La misma afirmación se formula sobre r_{2t^*+1}, r_{2T} . Bajo estas condiciones, el autor indica que las estadísticas

$$v_1 = (t^*)^{\frac{1}{2}} \bar{r}_1 / \left[\sum_{i=1}^{t^*} (r_{1i} - \bar{r}_1)^2 / (t^* - 1) \right]$$

$$y \quad v_2 = (T-t^*) \frac{1}{2} \cdot \bar{r}_2 \cdot \left/ \left[\sum_{i=t^*+1}^T (r_{2i} - \bar{r}_2)^2 / (T-t^*-1) \right] \right.$$

$$\text{con } \bar{r}_1 = \sum_{i=1}^{t^*} r_{1i} / t^* \quad , \quad \bar{r}_2 = \sum_{i=t^*+1}^T r_{2i} / (T-t^*)$$

tendrán una distribución de Student con t^*-1 y $T-t^*-1$ grados de libertad respectivamente si t^* no fuese determinado a partir de los mismos datos. El criterio de prueba propuesto consiste en rechazar la hipótesis de no cambio si v_1 o v_2 es significativamente diferente a cero. Una dificultad adicional menos grave que se menciona en relación a este criterio es que los valores esperados de \bar{r}_1 y \bar{r}_2 ,

$$E(\bar{r}_1) = (a_2 - a_1) + (b_2 - b_1) \sum_{i=1}^{t^*} X_i / t^* ,$$

$$E(\bar{r}_2) = (a_1 - a_2) + (b_1 - b_2) \sum_{i=t^*+1}^T X_i / (T-t^*)$$

pueden ser iguales a cero simultáneamente aun bajo la hipótesis alternativa aunque esto es mas bien, improbable.

La dependencia de v_1 y v_2 en t^* puede ser eliminada, de acuerdo a la sugerencia de Quandt, sustituyendo a t^* por un valor determinado exógenamente como por ejemplo, $t=T/2$. Con esta modifica-

ción, Quandt afirma que las estadísticas de prueba siguen una distribución de Student si adicionalmente, en vez de usar la varianza de los residuales cruzados en el denominador, se emplean los estimadores usuales de varianza de los modelos separados de regresión aunque la potencia de la prueba podrían verse disminuída a medida que el verdadero valor del punto de cambio se aleja de t .

Un último criterio es propuesto por el autor. En relación a las cantidades

$$\sum_{i=1}^t (\hat{a}_2 + \hat{b}_2 X_i - Y_i)^2 / \sigma^2$$

y

$$\sum_{i=t+1}^T (\hat{a}_2 + \hat{b}_2 X_i - Y_i)^2 / \sigma^2 \quad ,$$

afirma que estas variables aleatorias tienen, bajo H_0 , una distribución Ji cuadrada con $t-1$ y $T-t-1$ grados de libertad respectivamente, de modo que su cociente se puede completar a una variable F. Un argumento similar es expuesto para las sumas de cuadrados análogos con \hat{a}_1 y \hat{b}_1 . Quandt propone rechazar la hipótesis nula si alguno de los dos cocientes es significativamente grande. En

su opinión, este último procedimiento también evita el problema de probar una hipótesis sugerida por los datos aunque la localización del verdadero punto de cambio puede disminuir su potencia. Es importante notar que en este segundo trabajo se utilizan algunos resultados que son falsos. Particularmente, los grados de libertad de la variable

$$\sum_{i=T-t+1}^T (Y_i - \hat{a}_2 - \hat{b}_2 X_i)^2 / \sigma^2$$

no son $T-t-1$ como afirma sino $T-t-2$. El mismo problema se tiene con

$$\sum_{i=1}^t (Y_i - \hat{a}_1 - \hat{b}_1 X_i)^2 / \sigma^2$$

que tiene en realidad $t-2$ grados de libertad. Por otra parte, los residuales cruzados no son independientes bajo H_0 como lo afirma Quandt, aunque estos pueden ser transformados para que satisfagan esta propiedad. En resumen, Quandt descarta en este segundo trabajo, la conjetura J_1 cuadrada para el criterio asintótico y propone algunos procedimientos alternativos para el caso de tamaño pequeño de muestra. Estos procedimientos han sido cons

truidos en base a algunas afirmaciones que son falsas aunque pueden corregirse. Probablemente la mayor virtud de las contribuciones de este autor radica en que hace patente las dificultades que implica el problema de punto de cambio aun en el caso más simple de regresión cuando se trata mediante el enfoque Frecuentista.

Al igual que en el caso de sucesiones de variables aleatorias, el nombre de D.V.Hinkley aparece asociado a los primeros trabajos, sobre puntos de cambio en regresión. En particular, un artículo que data de 1969, [46], aborda el problema partiendo de la misma estructura propuesta por Quandt pero bajo el supuesto de que efectivamente tuvo lugar uno y solo un cambio, centra el interés en las inferencias sobre el punto en que este cambio se efectuó. Esencialmente considera el procedimiento de máxima verosimilitud para la estimación puntual y una aproximación Ji cuadrada para la distribución asintótica de $-2\ell_n \lambda$, donde λ es el cociente de verosimilitudes para la hipótesis de interés. Como puede apreciarse, en principio el trabajo de Hinkley es muy parecido al de Quandt que ya ha sido descrito. La diferencia,

que permite obtener los resultados propuestos, estriba en que se introducen algunas suposiciones adicionales en la estructura del problema. El modelo considerado es el siguiente:

$$Y_i = a_1 + b_1 X_i + \varepsilon_i ; \quad i = 1, 2, \dots, \tau,$$

$$Y_i = a_2 + b_2 X_i + \varepsilon_i ; \quad i = \tau+1, \dots, T.$$

con $\varepsilon_1, \dots, \varepsilon_T$ errores Normales, independientes de media cero y varianza σ^2 . Los parámetros a_1, a_2, b_1, b_2 y τ se consideran desconocidos. Hasta aquí el modelo coincide, en general, con el propuesto por Quandt; sin embargo, Hinkley introduce dos suposiciones más. Primero que los valores de X_1, \dots, X_T están ordenados respecto al índice de la observación. De esta forma se tiene que, en caso de producirse un cambio, las regresiones no quedan superpuestas en la zona de exploración sino que cada régimen describe la relación entre X y Y en una región distinta. En segundo lugar, considera la magnitud $\gamma = (a_1 - a_2) / (b_2 - b_1)$ que representa el punto donde las dos rectas de regresión se intersectan e impone la restricción de que $X_\tau < \gamma < X_{\tau+1}$. En estas condiciones, se puede suponer que el cambio en realidad constituye una

transición continua de un régimen a otro respecto al valor de la variable independiente. De hecho, el punto de cambio, en un sentido estricto, es precisamente γ . Así pues, Hinkley desarrolla su trabajo con el objetivo de estimar γ , encontrar la distribución del estimador y probar la hipótesis del tipo $H_0: \gamma = \gamma^*$. Es interesante notar que una de las principales consecuencias de la estructura considerada por este autor es que el estimador del punto de cambio no está restringido a tomar valores en un conjunto finito. De esta manera se evita la aparentemente, más notable dificultad del procedimiento propuesto por Quandt.

Para estimar a γ , Hinkley propone el método de máxima verosimilitud y procede de la siguiente manera:

La función de verosimilitud de las observaciones está dada por:

$$L(\underline{Y}_T | a_1, a_2, b_1, b_2, \gamma, \tau, \sigma) = (2\pi\sigma^2)^{-T/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{\tau} (Y_i - a_1 - b_1 X_i)^2 + \sum_{i=\tau+1}^T (Y_i - a_2 - b_2 X_i)^2 \right] \right\}$$

en donde $(a_2 - a_1) + \gamma(b_2 - b_1) = 0$ y además $X_\tau \leq \gamma < X_{\tau+1}$. Para un valor fijo t de τ , la verosimilitud, considerando a σ^2 conocido, puede

de maximizarse respecto a los parámetros a_1 , a_2 , b_1 y b_2 sujeta a la restricción que involucra a γ . Como resultado, se obtiene la función de verosimilitud marginal de γ , $L_t(\gamma)$, en el intervalo $[X_t, X_{t+1}]$. Procediendo de esta manera con los diferentes valores de t , se construye la verosimilitud completa de γ , $L(\gamma)$, definida como

$$L(\gamma) = \prod_{t=2}^{T-1} L_t(\gamma); \quad \gamma \in [X_t, X_{t+1}]; \quad t=2, \dots, T-2$$

Reescribiendo, $L_t(\gamma) = (2\pi\sigma^2)^{-T/2} \exp \{-(2\sigma^2)^{-1} S_t^2(\gamma)\}$ donde $S_t^2(\gamma)$ es la suma de cuadrados de los residuales del modelo con dos rectas de regresión intersectadas en γ , es claro que el problema se reduce a encontrar $\hat{\gamma}$ tal que minimice $S_t^2(\gamma)$ para $\gamma \in [X_t, X_{t+1}]$ y $t=2, \dots, T-2$. Básicamente por facilidad, Hinkley propone el empleo de $Z_t^2(\gamma) = S_0^2 - S_t^2(\gamma)$ como función objetivo, donde S_0^2 representa la suma de cuadrados de los residuales producidos por el ajuste de un modelo de regresión simple a las T observaciones disponibles. Naturalmente, S_0^2 no depende de γ , de modo que el criterio de minimizar $S_t^2(\gamma)$ es equivalente a maximizar la correspondiente $Z_t^2(\gamma)$.

La existencia de $\hat{\gamma}$ queda garantizada en cuanto al autor exhibe la continuidad de $L(\gamma)$ y considera las restricciones estructurales del modelo. Más aún, Hinkley indica que si se define $\tilde{\gamma}_t$ como el valor de γ que maximiza $L_t(\gamma)$ en toda la recta real, entonces y en virtud de la definición de $L(\gamma)$, solo puede ocurrir que $\hat{\gamma}$ coincide con $\tilde{\gamma}_t$ para alguna t , en cuyo caso $\tilde{\gamma}_t \in [X_t, X_{t+1})$, ó $\hat{\gamma}$ es igual a X_t , también para alguna t . Así pues, el procedimiento para calcular $\hat{\gamma}$ consiste en determinar $L_t(\tilde{\gamma}_t)$ para todos aquellos casos en que $\tilde{\gamma}_t \in [X_t, X_{t+1})$ y $L(X_t)$ para el resto, seleccionando como $\hat{\gamma}$ el valor que produzca el mayor valor de $L(\gamma)$. Desafortunadamente, $L(\gamma)$ no es diferenciable en todo su dominio y como consecuencia, no es posible encontrar una expresión explícita para este estimador, de manera que para establecer su distribución asintótica, Hinkley recurre un más detallado examen de la relación que existe entre $\hat{\gamma}$ y la sucesión $\{\tilde{\gamma}_t\}$.

Primeramente, el autor se refiere a un resultado de Feder y Sylvester [32], en donde se exhibe que $\hat{\gamma}$ es asintóticamente distribuida según una ley Normal aunque la aproximación para el caso de un tamaño de muestra pequeño parece no ser adecuada. En cualquier forma,

Hinkley se basa en ese resultado para sugerir que, al menos asintóticamente, la probabilidad de que $\hat{\gamma} = X_t$ para cualquier t es cero, de modo que en general, $\hat{\gamma}$ coincide con $\tilde{\gamma}_t$ para alguna t . Este argumento clarifica notablemente el camino a seguir, puesto que la distribución de $\hat{\gamma}$ debe depender de la correspondiente de los estimadores $\{\tilde{\gamma}_t\}$ que a su vez son obtenidos por máxima verosimilitud en condiciones usuales por lo que su distribución puede aproximarse razonablemente por una Normal.

En base a resultados obtenidos de simulaciones, el autor sugiere que la relación $\hat{\gamma} > X_{t+1}$ implica $\tilde{\gamma}_t > X_{t+1}$ así como que $\hat{\gamma} < X_t$ implica $\tilde{\gamma}_t < X_t$. Por medio del desarrollo formal de la dependencia entre estas relaciones, Hinkley verifica que la distribución de $\hat{\gamma}$ puede ser aproximada, en los puntos X_2, X_1, \dots, X_{T-2} cuando estos están igualmente espaciados, por la función

$$F_1(X_t) = \begin{cases} P(\tilde{\gamma}_t < X_t) & ; t < (T+1)/2 \\ P(\tilde{\gamma}_{t-1} < X_t) & ; t > (T+1)/2 \end{cases}$$

Adicionalmente, propone como aproximación continua una interpolación

ción lineal para cada valor $\gamma \in (X_{t-1}, X_t]$. El autor procede a verificar que como era de esperarse, la distribución de $\tilde{\gamma}_t$ puede aproximarse por una ley Normal cuyos parámetros dependen del verdadero valor de γ , de la diferencia $b = b_2 - b_1$ y naturalmente, de t . De esta forma se tiene que la distribución de $\hat{\gamma}$ puede aproximarse por una mezcla de Normales.

Un argumento más, relativo a que el comportamiento asintótico de los estimadores $\{\tilde{\gamma}_t\}$ es prácticamente el mismo y puede ser aproximado por la distribución de $\tilde{\gamma}_\tau$ donde τ es el verdadero valor del índice de la última observación en el primer régimen de regresión es utilizado. En base a este resultado, Hinkley afirma que:

$$P(\hat{\gamma} < w) \sim F_2(w) = \Phi \left\{ \frac{b(w - \gamma)}{a} \left[\frac{\tau(T - \tau)}{4T} \right] \right\}$$

con $\Phi \{\cdot\}$ la distribución Normal (0,1). El autor menciona que la aproximación con F_1 depende del igual espaciamiento de los valores en X . Sin embargo, afirma que el resultado es esencialmente, válido en general. En base a un estudio de simulación las dos distribuciones son examinadas y se obtiene un mejor comportamiento de F_1 , lo cual resulta razonable si se toma en

cuenta que F_2 requiere de una aproximación adicional. Finalmente, se comenta el procedimiento a seguir para probar las hipótesis

$$H_0: \gamma = \gamma_0 \quad \text{vs.} \quad H_1: \gamma \neq \gamma_0$$

Utilizando el método de cociente de verosimilitudes generalizado se puede verificar que si Λ es cociente involucrado,

$$-2 \ln \Lambda = \frac{1}{\sigma^2} \{ Z_{\tilde{\gamma}}^2 - Z_{\gamma_0}^2 \}$$

donde $X_{\tilde{\gamma}} < \tilde{\gamma} < X_{\tilde{\gamma}+1}$ y $X_{\gamma_0} < \gamma_0 < X_{\gamma_0+1}$. Esta estadística tiene asintóticamente una distribución Ji-cuadrada central con un grado de libertad bajo H_0 . Si el parámetro σ^2 fuese desconocido, se sugiere sustituirlo por su estimador usual y aproximar la distribución por una F.

El trabajo de Hinkley puede describirse brevemente indicando que aborda el problema de puntos de cambio en regresión presuponiendo la existencia de un cambio y considerando algunas suposiciones adicionales que producen un caso particular en donde los métodos frecuentistas usuales pueden aplicarse en mejores condiciones.

Así, el problema de estimar el punto de cambio se traduce en la estimación con restricciones del punto de intersección de dos rectas. La distribución del estimador de máxima verosimilitud se aproxima por una mezcla de Normales y se propone el empleo del cociente de verosimilitudes para probar hipótesis sobre el parámetro de interés. Dos comentarios son oportunos. Primero, el planteamiento de este autor permite pasar de la estimación del punto donde los datos observados cambian de régimen a la estimación del punto en la zona de exploración donde el modelo sufre una modificación. Este tipo de resultados podría resultar de mayor utilidad para la realización de experimentos posteriores. En segundo lugar, las suposiciones de que las rectas no se traslapan y de que la transición es continua propicia que un modelo cuadrático aparezca como un muy natural contendiente del propuesto por Hinkley. Al respecto el autor menciona que un posible procedimiento para confrontar estas alternativas podría ser el análisis de los residuales.

Otro artículo que trata el problema de puntos de cambio en regresión es el debido a Farley y Hinich [3], en donde se desarrolla

un procedimiento para probar la hipótesis nula de no cambio contra la alternativa de que uno y solo un cambio ha tenido lugar. Los autores desarrollan un criterio que es localmente más potente a medida de la magnitud del cambio converge a cero y resulta asintóticamente eficiente.

El modelo considerado es el siguiente:

$$Y_i = a_i + b_i X_i + \varepsilon_i \quad ; \quad i=1,2,\dots,T$$

en donde como de costumbre, X_1, \dots, X_T son constantes fijas y los errores $\{\varepsilon_i\}$ son independientes e idénticamente distribuidos Normal $(0, \sigma^2)$. Por su parte, los parámetros del modelo se definen como sigue:

$$a_i = \begin{cases} a; & 1 \leq i < \tau \\ a - \delta X_\tau; & \tau + 1 \leq i \leq T \end{cases} \quad b_i = \begin{cases} b; & 1 \leq i < \tau \\ b + \delta; & \tau + 1 \leq i \leq T \end{cases}$$

Por supuesto, τ es el punto de cambio en el modelo ($1 < \tau < T$).

Reescribiendo el modelo se tiene que:

$$Y_i = a + b X_i + \varepsilon_i \quad ; \quad i=1,2,\dots,\tau$$

$$Y_i = a - \delta X_\tau + (b + \delta) X_i + \varepsilon_i \quad ; \quad i=\tau+1,\dots,T$$

En estas expresiones es interesante notar algunas características. Primero, el esquema es muy parecido al propuesto por Quandt; sin embargo, los parámetros de las regresiones se transforman de (a, b) en (a_1, b_1) con $a_1 = a - \delta X_T$, $b_1 = b + \delta$, lo cual implica una restricción en el tipo de cambios que se pueden verificar. Particularmente, es de interés observar que el parámetro δ puede interpretarse como una medida de la magnitud del cambio, de hecho, si $\delta = 0$ no existe cambio alguno. Así pues, la hipótesis nula se puede contrastar probando $H_0: \delta = 0$. Por otra parte, la circunstancia de que la ordenada al origen después del cambio, dependa del valor de la observación en donde el cambio se llevó a efecto, impone una restricción muy específica en el modelo. Puede verificarse muy fácilmente que si δ es distinta de cero, y por tanto se efectuó un cambio, las rectas se intersectan precisamente en X_T . De esta manera el trabajo de Farley y Hinich puede considerarse como intermedio entre los de Quandt y Hinkley ya descritos, puesto que si bien no requiere que los valores de la variable independiente estén ordenados respecto al índice de observación, si se reduce a considerar los casos en que el punto de

cambio coincide con el valor donde se intersectan las rectas.

El procedimiento que siguen los autores de este trabajo es relativamente simple y parte de considerar que σ^2 es conocida y τ es una variable aleatoria discreta distribuida uniformemente en $\{1, 2, \dots, T\}$. Esto es,

$$P(\tau = t) = \frac{1}{T} ; \quad t = 1, 2, \dots, T$$

Es importante hacer notar que si bien Farley y Hinich afirman que esta distribución representa la situación en que el investigador no posee información apriori sobre el punto de cambio, el enfoque que utilizan no es Bayesiano ya que en tal caso tendrían que asignar distribuciones iniciales a los demás parámetros (a, b y δ) y seguir un tratamiento distinto del problema.

Los autores proceden a verificar que bajo la hipótesis nula el vector \underline{Y} de observaciones sigue una distribución Normal multivariada de media $\underline{\mu}$ y varianza $\sigma^2 I$ en donde,

$$\underline{\mu} = a \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \gamma \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{bmatrix} = a \underline{1} + \gamma \underline{X}$$

mientras que si se produce un cambio en $\tau=t$, de magnitud δ , la matriz de varianzas y covarianzas permanece igual pero la esperanza de \underline{Y} se modifica como sigue:

$$E(Y[t, a, b, \delta]) = a \underline{1} + \gamma \underline{X} + \delta \underline{Z}(t) = \underline{\mu} + \delta \underline{Z}(t)$$

con $\underline{Z}'(t) = (0, \dots, 0, X_{t+1} - X_t, X_{t+2} - X_t, \dots, X_T - X_t)$.

Para efectuar la prueba de la hipótesis $\delta=0$ proceden a calcular las verosimilitudes de interés.

Bajo H_0 , se tiene que:

$$f_0(\underline{Y} | a, b) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\underline{Y} - \underline{\mu})' (\underline{Y} - \underline{\mu}) \right\}$$

mientras que si el cambio ocurre en $\tau=t$ con una magnitud δ ,

$$f(Y | a, b, t, \delta) = (2\pi\sigma^2)^{-T/2} \exp \left\{ -\frac{1}{2\sigma^2} (\underline{Y} - \underline{\mu} - \delta \underline{Z}(t))' (\underline{Y} - \underline{\mu} - \delta \underline{Z}(t)) \right\}$$

En virtud de que τ es considerado un parámetro aleatorio, y como la hipótesis es relativa a δ , Farley y Hinich proceden a determinar la verosimilitud bajo la hipótesis alternativa, margina-

lizada sobre τ ,

$$f(Y | a, b, \delta) = \prod_{t=1}^T T^{-1} f(Y | a, b, t, \delta) \\ = \prod_{t=1}^T T^{-1} (2\pi\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}(\underline{Y}-\underline{\mu}-\delta \underline{Z}(t))'(\underline{Y}-\underline{\mu}-\delta \underline{Z}(t))\right\}$$

Esta verosimilitud, como es fácil de verificar aún cuando los autores no lo mencionan, tiene un valor esperado dado por la siguiente expresión:

$$\underline{\mu}^* = T^{-1} \sum_{t=1}^T (\underline{\mu} + \delta \underline{Z}(t)) \\ = \underline{\mu} + \delta T^{-1} \sum_{t=1}^T \underline{Z}(t) \\ = \underline{\mu} + \delta \underline{0}$$

Considerando la verosimilitud bajo H_0 y la correspondiente a la alternativa marginalizada sobre τ , Farley y Hinich calculan el cociente respectivo,

$$f(Y | a, b, \delta) / f(\underline{Y} | a, b) = \lambda(\underline{Y} | a, b, \delta) \\ = \prod_{t=1}^T T^{-1} \exp\left\{\frac{\delta}{\sigma^2}(\underline{Y}-\underline{\mu})'\underline{Z}(t) - \frac{\delta^2}{2\sigma^2}\underline{Z}(t)'\underline{Z}(t)\right\}$$

Desarrollando la expansión en series de λ respecto a δ alrededor de $\delta=0$, se obtiene que

$$\lambda = 1 + \frac{\sum_{t=1}^T T^{-1}}{\sigma^2} \frac{\delta}{\sigma^2} (\underline{Y} - \underline{\mu})' \underline{Z} (t)$$

$$\approx 1 + \frac{\delta}{\sigma^2} (\underline{Y} - \underline{\mu})' \underline{Q}$$

de tal manera que si a, b, σ^2 y δ son conocidos y δ además, es cercano a cero, el criterio de prueba sería rechazar H_0 si

$$\lambda^* = 1 + \frac{\delta}{\sigma^2} (\underline{Y} - \underline{\mu})' \underline{Q}$$

es significativamente diferente de cero. Equivalentemente, el criterio se puede definir en términos de $S = \underline{Q}' (\underline{Y} - \underline{\mu})$, rechazando H_0 si S es significativamente diferente de cero. Los autores afirman que si a, b, σ^2 y δ son fijos el criterio es localmente más potente si δ tiende a cero. Por otra parte también el criterio propuesto es asintóticamente eficiente respecto al correspondiente al cociente de verosimilitudes. Esta claro entonces que el criterio desarrollado puede ser empleado para probar $H_0: \delta=0$, si a, b y σ^2 son conocidos, contra cualquier alternativa $H^*: \delta = \delta^*$ siempre que δ^* sea cercana a cero. Un atractivo del

criterio es que bajo H_0 la distribución de S es extremadamente simple puesto que la estadística de prueba es una combinación lineal de variables Normales independientes.

Es interesante notar que la estadística de prueba puede ser interpretada informalmente, de manera muy simple. Como ya se indicó, bajo la hipótesis alternativa, \underline{Y} tiene una distribución de media $\underline{\mu}^* = \underline{\mu} + \delta \underline{\theta}$. De esta forma, el vector $\underline{Y} - \underline{\mu}$ podría interpretarse como una estimación de la diferencia $\underline{\mu}^* - \underline{\mu} = \delta \underline{\theta}$. Naturalmente entonces, la estadística S es cercana a cero si solo si, el parámetro δ es suficientemente cercano a cero.

Los autores mencionan que en general, los parámetros a y b son desconocidos y proponen modificar el criterio sustituyendo estos valores por sus estimadores usuales bajo H_0 . Así la estadística de prueba está dada por

$$\begin{aligned} \tilde{S} &= \underline{\theta}' (\underline{Y} - \hat{\underline{\mu}}) \\ &= \underline{\theta}' (\underline{Y} - \hat{\underline{a}} \underline{1} - \hat{\underline{b}} \underline{X}) \\ &= \underline{\theta}' (I - A) \underline{Y} \end{aligned}$$

que no es más que una combinación lineal de los residuales bajo

H_0 y donde $A = T^{-1} \mathbf{1}\mathbf{1}' + \{X'X + T\bar{X}^2\}^{-1} (X - \bar{X})(X - \bar{X})'$ y \bar{X} es el vector cuyas entradas son todas iguales a \bar{X} , la media de las observaciones,

Es sencillo comprobar que \tilde{S} tiene, bajo H_0 , una distribución Normal de media cero y varianza $\sigma^2 \theta' (I-A) \theta$. De modo que el criterio para probar la hipótesis nula, a un nivel de significancia α , consiste en rechazar H_0 si $|\tilde{S}|$ excede el valor $Z_{\alpha/2} \sigma \{\theta' (I-A) \theta\}^{1/2}$ donde $Z_{\alpha/2}$ es el cuantil de orden $1 - \alpha/2$ de una distribución Normal (0,1). Con la finalidad de analizar la potencia del criterio, los autores proponen el empleo de la medida de detectabilidad

$$D_{\delta}^{\alpha} = \frac{|E_{\delta}(\tilde{S}) - E_0(\tilde{S})|}{\sqrt{V_0(\tilde{S})}}$$

donde $E_{\delta}(\tilde{S})$ es la esperanza de \tilde{S} bajo la hipótesis alternativa cuando la magnitud del cambio es δ mientras que $E_0(\tilde{S})$ y $V_0(\tilde{S})$ son la media y la varianza respectivamente de \tilde{S} bajo la hipótesis nula. Farley y Hinich verifican que de entre todas las posibles combinaciones lineales de los residuales, \tilde{S} es la que produce la mayor detectabilidad. Adicionalmente, indican que si σ^2 es también desconocida pero T es suficientemente grande, este paráme-

tro puede ser sustituido por su estimador usual y la regla de decisión es asintóticamente equivalente.

Finalmente, los autores realizan un estudio mediante simulación para caracterizar la potencia del criterio en función del parámetro δ así como de la configuración de los valores de la variable independiente y la localización del punto de cambio. Los resultados que obtienen sugieren que el patrón de variable independiente, en efecto, influye en la potencia aunque tal influencia parece ser relevante solo para valores de δ cercanos a cero. En general la potencia aumenta a medida que $|\delta|$ crece pero no se aproxima a uno tan rápidamente como se esperaría. Este hecho es comentado en el artículo examinado en términos de que probablemente se debe a que en general se desconoce el punto de cambio. En relación a ello, el propio estudio de simulación revela que es mucho más alta la potencia cuando el punto de cambio se localiza en la parte media de la sucesión que cuando se encuentra cerca de alguno de los extremos. Finalmente, se comparan los resultados cuando σ^2 es conocida contra los que se obtienen en el caso contrario encontrando como era de esperarse, un mejor comportamiento

si se conoce el valor de σ^2 , si bien el deterioro en potencia no es muy grande, en el otro caso sobre todo si el punto de cambio se localiza en la parte media de la sucesión de observaciones y la magnitud del cambio es relativamente pequeña. Resumiendo, estos autores desarrollan un procedimiento frecuentista para probar la hipótesis de no cambio incorporando una suposición adicional en el modelo alternativo. La técnica es extremadamente simple en virtud de que la estadística de prueba resulta ser una combinación lineal de los residuales usuales. De hecho los autores indican que el cálculo de la estadística puede ser implementado en cualquier paquete de computadora como parte de las rutinas de verificación de las suposiciones. Adicionalmente, se sugiere que en caso de detectar un cambio, el punto de cambio puede ser estimado con el procedimiento de Quandt.

La potencia de la prueba depende de la configuración de los valores de la variable independiente y de la magnitud del cambio, si bien los estudios de simulación reportados indican que los resultados son satisfactorios. Como observación final, es interesante notar que los autores no hacen uso del hecho de que la estadísti-

ca de prueba tiene una distribución Normal bajo la hipótesis alternativa. Esta circunstancia podría haber sido empleada para calcular la potencia de la prueba sin recurrir a las simulaciones.

En un trabajo publicado el mismo año que el de Farley y Hinich otros autores, McGee y Carleton [62], presentan un tratamiento para el problema de regresión con puntos de cambio que puede ser considerado mas general en el sentido de que es aplicable no solo al caso de un punto de cambio, sino a aquellas situaciones en las cuales se desconoce el número de cambios en el modelo. Una de las características más interesantes de este artículo es que desarrolla un procedimiento que, con un enfoque pragmático según la propia opinión de sus autores, reduce notablemente la complejidad computacional que implica este tipo de problema . De hecho, McGee y Carleton afirman que la técnica propuesta, una combinación de análisis de regresión y clasificación jerárquica, no garantiza la optimalidad de la solución obtenida en un sentido analítico sino más bien es de carácter constructivo y pretende producir una descripción adecuada del comportamiento registrado en un conjunto es

pecífico de observaciones. El esquema propuesto parte de considerar T vectores de observaciones $(Y_i, X_{1i}, X_{2i}, \dots, X_{ki})$; $i=1, 2, \dots, T$ en donde

$$\begin{aligned}
 Y_i &= \beta_0^{(1)} + \beta_1^{(1)} X_{1i} + \dots + \beta_k^{(1)} X_{ki} + \epsilon_i; \quad i=1, 2, \dots, t_1 \\
 Y_i &= \beta_0^{(2)} + \beta_1^{(2)} X_{1i} + \dots + \beta_k^{(2)} X_{ki} + \epsilon_i; \quad i=t_1+1, \dots, t_2 \\
 &\vdots \\
 Y_i &= \beta_0^{(n)} + \beta_1^{(n)} X_{1i} + \dots + \beta_k^{(n)} X_{ki} + \epsilon_i; \quad i=t_{n-1}+1, \dots, T
 \end{aligned}$$

con $\beta_j^{(t)}$ parámetros desconocidos; t_1, t_2, \dots, t_{n-1} y n también desconocidos y $\epsilon_1, \dots, \epsilon_T$ errores aleatorios independientes e idénticamente distribuidos Normales $(0, \sigma^2)$. El índice i de las observaciones, puede corresponder al orden de registro en el tiempo o la magnitud de alguna de las variables independientes. En virtud de que tanto n como los puntos de cambio son desconocidos, el problema puede presentarse como un caso de agrupamiento en donde es necesario establecer cuantos grupos de observaciones existen, cada uno de los cuales se describe con un modelo de regresión diferente y también cuáles observaciones pertenecen a cada grupo. Aún en el caso en que el número de cambios $n-1$ fuese conocido, para tamaños

relativamente moderados de n y T la cantidad de posibles agrupamientos de T observaciones en n distintos regímenes de regresión puede implicar un volumen excesivo de cálculos. Sin embargo, considerando que las observaciones están ordenadas y que los cambios, en caso de efectuarse, producen grupos de observaciones adyacentes respecto a ese orden, el problema se simplifica. Más aún, los autores proponen el empleo de un algoritmo jerárquico, esto es, un procedimiento de agrupamiento secuencial tal que un grupo, una vez formado no pueda ser dividido en etapas posteriores. En los casos usuales de agrupamiento de observaciones, es necesario definir una medida de distancia o similitud que permita establecer cuales son las observaciones que pueden incluirse en un mismo grupo en cada etapa. Para el problema específico de regresión los autores proponen el empleo de una medida de bondad de ajuste como la estimación de la varianza para el modelo ajustado en cada caso. Así, pues el procedimiento propuesto consiste en considerar inicialmente los grupos minimales, esto es los conjuntos de observaciones adyacentes con el número de puntos mínimo suficiente para ajustar un modelo del tipo establecido y producir

un estimador no trivial de la varianza. En el caso de un modelo lineal con $k+1$ parámetros, los grupos minimales constan de $k+2$ observaciones. De entre estos grupos minimales se considera como un grupo formado, aquel cuya varianza estimada sea menor y se procede con la siguiente etapa considerando que ahora se pueden formar nuevos grupos a partir de datos aislados o bien incluyendo un nuevo punto en el grupo ya establecido. En general, en una etapa intermedia del algoritmo se pueden formar nuevos grupos de tres diferentes formas: a partir de un grupo minimal; incluyendo un dato a un grupo ya formado o combinando dos grupos ya establecidos. De esta forma se procede desde la etapa inicial en que todos los puntos se encuentran aislados hasta la etapa final en que todos ellos forman un grupo al que se ajusta un único modelo. Con el objeto de establecer cual de las etapas es la que brinda una más adecuada descripción de los datos, los autores proponen que cada vez que un nuevo punto se incluya en un grupo se determine si esa inclusión no modifica significativamente el ajuste. Esta determinación se puede realizar a través de la estadística

$$F = (s-k-1) \frac{SC_1 + SC_0}{SC_0}$$

en donde SC_1 es la suma de cuadrados de los residuales del modelo ajustado considerando el nuevo punto junto con los del grupo ya formado, mientras que SC_2 es la correspondiente al grupo formado inicialmente. La estadística propuesta, bajo el supuesto de que el modelo estimado es el mismo, tiene una distribución F con 1 y S-K-1 grados de libertad, donde S es el número de puntos en el grupo preexistente y K+1 es el número de parámetros en el modelo. De esta forma, si el valor calculado de F excede del valor apropiado en tablas de una distribución F(1, S-K-1) no resulta conveniente incluir el nuevo punto aunque esta estrategia sea la que minimice la varianza del nuevo grupo. Un tratamiento similar se puede llevar a cabo cuando la menor varianza se obtenga al combinar dos grupos ya formados. En ese caso se puede utilizar la estadística usual para la prueba de igualdad de coeficientes en dos modelos de regresión

$$F' = \frac{S_1 + S_2 - 2(K+1)}{K+1} \cdot \frac{SC_{1+2} - SC_1 - SC_2}{SC_1 + SC_2}$$

en donde SC_1 , SC_2 y SC_{1+2} representan respectivamente, las sumas de cuadrados de residuales del modelo ajustado al primer grupo

formado, con S_1 puntos, al segundo grupo formado con S_2 puntos y a las observaciones de ambos grupos combinados. Si F' resulta mayor que el cuantil correspondiente de una F con $K+1$ y $S_1+S_2-2(K+1)$ grados de libertad, existe evidencia para no combinar los dos grupos aun cuando el conjunto resultante tenga el mejor ajuste en esa etapa.

Naturalmente, se pueden utilizar algunos otros criterios para conducir el procedimiento. Por ejemplo, si de antemano se posee información acerca del número de grupos, el algoritmo se puede terminar en cuanto se forme ese número de grupos. Asimismo, si se está dispuesto a permitir que algunos puntos permanezcan aislados, también se puede concluir el procedimiento en cuanto se alcance la cota correspondiente.

Los autores mencionan, como ya se indicó, que la técnica no puede evaluarse analíticamente para establecer sus propiedades pero con la finalidad de compararla con otras técnicas, la aplican al conjunto de datos generados por Quandt [72], encontrando que si consideran como máximo permisible un punto aislado y utilizan los cuantiles al 1% para las pruebas involucradas, se obtienen los

mismos resultados que Quandt. Esto es, finalmente se tienen dos grupos, uno de ellos con las primeras 12 observaciones y el otro con las restante ocho. La técnica también es aplicada a una serie de datos económicos reales sobre transacciones bursátiles y se encuentra que el procedimiento sugiere la existencia de tres cambios en el modelo de regresión. El más importante de éstos coincide con la fecha de una modificación en las normas de transacción cuyo efecto ya había sido previsto. Un par de simulaciones más se incluyen en el trabajo descrito en donde se verifica que en general la capacidad para determinar es menor a medida de la varianza de los errores aumenta.

Finalmente, los autores mencionan que de hecho la única restricción de su técnica es que se requiere que los modelos ajustados a los grupos formados sean del mismo tipo con con el fin de que la prueba de igualdad de coeficientes tenga sentido. Asimismo comentan que la idea esencial puede ser aplicada a modelos no necesariamente lineales. El enfoque que este artículo mantiene es su^uamente atractivo desde el punto de vista computacional y si las pruebas estadísticas involucradas son interpretadas adecuadamente,

considerándolas más como indicadores que como pruebas de significancia, la técnica permite obtener una descripción razonable del comportamiento de un conjunto de datos específico y por su puesto, sugerir algunas hipótesis sobre el fenómeno que lo generó.

En un nuevo trabajo [74], Richard E. Quandt ha retomado el problema de regresión con puntos de cambio con un nuevo enfoque. En una variedad de situaciones es posible ordenar las observaciones de tal manera que si ocurrió un cambio, pueda afirmarse que las primeras r observaciones corresponden a un modelo de regresión mientras que las subsecuentes son descritas a través de otro modelo de regresión diferente. Sin embargo, si este orden no es posible o si no es de interés identificar el punto donde se efectuó el cambio sino solo los parámetros asociados a cada modelo, Quandt sugiere que la situación puede ser caracterizada como un problema de mezcla de distribuciones. Así pues, si se tiene $(X_i, Y_i); i=1, 2, \dots, T$ observaciones, algunas de las cuales siguen la relación:

$$Y_i = a_i + b_i X_i + \epsilon_{i,i} ,$$

con a_1 y b_1 parámetros desconocidos y en donde los errores son independientes e idénticamente distribuidos $N(0, \sigma_1^2)$, mientras que el resto de las parejas se ajustan a un modelo

$$Y_i = a_2 + b_2 X_i + \varepsilon_{2i}$$

de parámetros a_2, b_2 desconocidos y errores independientes, Normales de media cero y varianza σ_2^2 con $(a_1, b_1) \neq (a_2, b_2)$, entonces el problema de la estimación de $a_1, b_1, a_2, b_2, \sigma_1^2$ y σ_2^2 puede plantearse considerando que en el fenómeno bajo estudio produce observaciones que mediante un mecanismo aleatorio son generadas de acuerdo a alguno de los dos modelos de regresión. De esta forma la función de densidad de Y_1, Y_2, \dots, Y_T esta dada por la siguiente expresión

$$f(Y_i | X_i, a_1, b_1, a_2, b_2, \sigma_1^2, \sigma_2^2) = \frac{P}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2\sigma_1^2} (Y_i - a_1 - b_1 X_i)^2\right\} \\ + \frac{1-P}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{1}{2\sigma_2^2} (Y_i - a_2 - b_2 X_i)^2\right\}$$

en donde P es la probabilidad con que el fenómeno selecciona el primer régimen de regresión y $1-P$ la probabilidad con que utiliza el segundo. Naturalmente, la distribución de Y_i es una mez-

cla de Normales en donde los coeficientes de la mezcla son precisamente P y $1-P$. Esta formulación de Quandt incorpora a la probabilidad P como un parámetro desconocido más que de ser estimado permitiría establecer de manera aproximada la proporción de observaciones en una muestra específica que pertenecen a uno u otro régimen. Es fácil verificar que el logaritmo de la función de densidad conjunta de Y_1, Y_2, \dots, Y_T está dado por

$$\alpha^n \prod_{i=1}^T \{ \frac{P}{\sqrt{2\pi}\sigma_1} \exp \left\{ -\frac{1}{2\sigma_1^2} (Y_i - a_1 - b_1 X_i)^2 \right\} + \frac{1-P}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{1}{2\sigma_2^2} (Y_i - a_2 - b_2 X_i)^2 \right\} \}$$

Quandt propone la estimación de los parámetros desconocidos por máxima verosimilitud. En estas circunstancias la solución al problema no puede obtenerse analíticamente en virtud de la naturaleza no lineal de la función objetivo, de modo que el autor propone el empleo de métodos numéricos para la maximización. Las propiedades de los estimadores producidos por este método no son discutidas en general en el trabajo revisado, sin embargo, se examina el comportamiento de la técnica mediante simulación en varios casos y con un

conjunto de datos reportados en la literatura.

Con la finalidad de establecer una comparación, el autor introduce un modelo alternativo que es debido a Goldfeld y Quandt [40], en donde se considera que existe una variable exógena Z tal que para un valor $Z = Z_0$ se tiene que si el valor de Z registrado para la observación (X_i, Y_i) es menor que Z_0 entonces la pareja pertenece al primer régimen de regresión mientras que si Z es mayor o igual a Z_0 , el modelo correcto es el segundo. De esta manera, empleando la representación matricial del modelo se puede describir

$$\underline{Y} = (I - D) X \beta_1 + D X \beta_2 + \underline{W}$$

en donde β_1 y β_2 son los vectores de parámetros asociados a los dos distintos regímenes de regresión y \underline{W} es el vector de errores, $\underline{W} = (I - D) \underline{\varepsilon}_1 + D \underline{\varepsilon}_2$ con D una matriz diagonal tal que $d_{ii} = 1$ si y sólo si $Z_i < Z_0$ en tanto que en caso contrario $d_{ii} = 0$. Con el propósito de reducir las dificultades analíticas de este procedimiento, el valor de d_{ii} se aproxima por la ecuación

$$d_{ii} = d(Z_i) = \int_{-\infty}^{Z_i} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(v-Z_0)^2\right\} dv.$$

De esta manera, en este otro modelo se introducen los parámetros σ^2 y Z_0 que en general son desconocidos pero pueden ser estimados

por máxima verosimilitud utilizando también métodos numéricos.

Los resultados de las simulaciones indican que ambos métodos son aparentemente equivalentes en términos de los sesgos medios en la estimación. Sin embargo, considerando el error cuadrático medio los resultados sugieren que el sesgo decrece en forma relativamente rápida a medida que se aumenta el tamaño de muestra cuando se emplea el método de Quandt. En cualquier forma, las simulaciones consideradas solo utilizan treinta repeticiones en cada caso, de modo que la evidencia que se presenta debe ser examinada con extrema cautela pues podría corresponder a un patrón poco representativo. El autor menciona para finalizar, que el método que propone puede ser empleado sin dificultad en modelos de regresión lineal múltiple así como en situaciones en que el número de cambios sea mayor considerando una mezcla que involucre todas las distribuciones necesarias. Por otra parte afirma que el valor estimado de P , además de poder ser interpretado como la proporción de observaciones en un régimen podría ser utilizado para probar la hipótesis de no cambio utilizando un cociente de verosimilitudes para contrastar las hipótesis:

$$H_0: p=0 \text{ ó } p=1 \quad \text{vs} \quad H_1: 0 < p < 1.$$

Es interesante observar que el procedimiento propuesto por Quandt en este trabajo se concentra en el problema de la estimación de los parámetros de los modelos de regresión involucrados en este tipo de problemas y puede ser aplicado en donde, por la naturaleza del fenómeno bajo estudio o de acuerdo a los objetivos del investigador, no se pretende determinar el punto donde el cambio se produce. De hecho, la técnica tampoco permite identificar cuales observaciones corresponden a cada modelo, tan solo a través de la estimación de los coeficientes en la mezcla de distribuciones se puede sugerir la proporción de puntos que en la muestra específica analizada corresponde a cada uno de ellos. En cuanto a las propiedades de la técnica propuesta el autor, como ya se indicó, examina el comportamiento de los estimadores en base a simulaciones que por su reducido volumen producen resultados que deben ser juzgados con precaución y no constituyen evidencia contundente en favor del procedimiento.

Bajo el enfoque Frecuentista se han desarrollado toda una variedad de técnicas para tratar el problema que es objeto de este trabajo.

Las diferencias van desde las suposiciones adicionales que se involucran en los modelos para obtener resultados analíticos más precisos, hasta la consideración de tratamientos que sin perder su carácter inferencial son más bien propuestos con miras a producir un análisis exploratorio de un conjunto de datos específico. Con este último espíritu es abordado el problema en un artículo debido a Brown, Durbin y Evans [15], en donde se considera la posibilidad de que ocurran cambios en el modelo de regresión que describe el comportamiento de una serie de tiempo. La idea es que el análisis mediante de regresión de datos ordenados en el tiempo usualmente se basa en la suposición de que la relación descrita es invariante a lo largo del tiempo. Sin embargo, en algunas aplicaciones esta suposición es cuestionable si se considera un período de tiempo relativamente largo pero puede constituir una alternativa razonable y sobre todo sencilla, si la relación se describe por trozos. De acuerdo a la mención de los autores, las pruebas de significancia que proponen deben ser consideradas esencialmente como indicadores y en general su trabajo pretende exhibir los posibles cambios a través de métodos gráficos sin considerar alternativas específicas para la hipótesis nula de no cambio. Las técnicas propuestas no consideran observaciones

correlacionadas; suponen que las variables independientes en los modelos son determinísticas y la relación con el tiempo es fundamentalmente utilizada para ordenar la sucesión de observaciones. Este orden permite como llevar a cabo un examen secuencial del modelo ajustado a la serie, incorporando uno por uno los datos en la muestra y evaluando la posibilidad de un cambio a través del comportamiento de algunas funciones de los residuales en cada etapa. El problema es planteado de la siguiente manera. Sea

$$Y_i = \beta_{1i}X_{1i} + \beta_{2i}X_{2i} + \beta_{3i}X_{3i} + \dots + \beta_{ki}X_{ki} + \epsilon_i; \quad i=1,2,\dots,T$$

$$= \underline{X}'_i \underline{\beta}_i + \epsilon_i; \quad i=1,2,\dots,T$$

en donde $\underline{X}'_i = (X_{1i}, X_{2i}, \dots, X_{ki})$ es el i -ésimo vector de variables independientes que se consideran determinísticas y $\underline{\beta}'_i = (\beta_{1i}, \beta_{2i}, \dots, \beta_{ki})$ es el vector de parámetros desconocidos en el modelo que también se define con el índice i para indicar que puede cambiar en el tiempo. Los errores $\epsilon_1, \dots, \epsilon_T$ como es usual, se suponen independientes e idénticamente distribuidos según una ley Normal con media común cero y varianzas σ_i^2 respectivamente. La hipótesis nula de interés (no cam

bios] se puede expresar como sigue:

$$\beta_1 = \beta_2 = \dots = \beta_T$$

H₀:

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_T^2$$

Los autores indican que el mayor interés se concentra en detectar cambios en los coeficientes de regresión si bien las técnicas que proponen parecen ser sensibles también a cambios en las varianzas. La idea de las técnicas desarrolladas parte de la consideración de que en general, el exámen de los errores residuales permite verificar las suposiciones de un modelo de regresión. Así, a través de ellos debe ser posible evaluar qué tan razonable resulta la hipótesis de no cambio. Los autores comentan sin embargo, que las gráficas usuales de los residuales no parecen ser sensibles ante este tipo de alteraciones y como consecuencia proponen, en analogía con el tratamiento que se siguen en procesos de control de calidad, el empleo de sumas acumuladas de los residuales y sus cuadrados. Es interesante notar la similitud de este procedimiento con el propuesto por Page [48], [49], para detectar cambios en los parámetros de sucesiones de variables aleatorias.

Desafortunadamente, como se indica en el trabajo descrito, la distribución de esas sumas acumuladas bajo la hipótesis nula resulta intratable y por tanto no es posible establecer cual sería un comportamiento atípico en las gráficas. Con el propósito de evitar esta dificultad, los autores proponen el empleo de un tipo especial de residuales transformados que se calculan de manera secuencial de acuerdo al siguiente razonamiento. Bajo el supuesto de que el vector $\underline{\beta}$ de coeficientes es de dimensión K , se puede calcular la sucesión de estimadores:

$$\hat{\underline{\beta}}_r = (\underline{X}'_r \underline{X}_r)^{-1} \underline{X}'_r \underline{Y}_r ; r = K+1, \dots, T$$

en donde $\underline{X}'_r = (\underline{X}_1, \dots, \underline{X}_r)$ y $\underline{Y}'_r = (Y_1, \dots, Y_r)$. De esta forma se puede definir para cada $r \geq K+1$, el residual

$$w_r = (\underline{Y}_r - \underline{X}'_r \hat{\underline{\beta}}_{r-1}) \{1 + \underline{X}'_r (\underline{X}'_{r-1} \underline{X}_{r-1})^{-1} \underline{X}_r\}^{-1/2}$$

que puede ser interpretado como el error estandarizado de la predicción de Y_r en base al modelo ajustado a las observaciones precedentes. Es fácil de verificar que bajo H_0 , w_{K+1}, \dots, w_r son independientes con distribución Normal $(0, \sigma^2)$. Una de las caracte-

terísticas interesantes de estos residuales es que se pueden calcular mediante un algoritmo secuencial recursivo que evita el ajuste de todos los modelos involucrados. Esta afortunada circunstancia se desprende del hecho de que los estimadores satisfacen la siguiente relación:

$$\hat{\beta}_x = \hat{\beta}_{x-1} + (X_x' X_x)^{-1} X_x (X_x' X_x - X_{x-1}' X_{x-1})^{-1} \hat{\beta}_{x-1}$$

en donde además, el cálculo de $(X_x' X_x)^{-1}$ puede evitarse recurriendo a la recursión

$$(X_T' X_T)^{-1} = (X_{T-1}' X_{T-1})^{-1} - \frac{(X_{T-1}' X_{T-1})^{-1} X_{T-1}' X_T (X_{T-1}' X_{T-1})^{-1}}{1 + X_T' (X_{T-1}' X_{T-1})^{-1} X_T}$$

En virtud de que, si β_i es constante para $i=1,2,\dots,t$ y cambia a partir de ese punto, entonces w_i para $i=1,2,\dots,t$ tienen media cero mientras que la esperanza de w_i ; $i=t+1,\dots,T$ es diferente de cero, es natural pensar en representaciones de los residuales transformados que permitan detectar la posible desviación de cero de sus medias a lo largo del tiempo. Con ese objetivo los autores proponen la construcción de la gráfica de la suma acumulada

$$W_r = \frac{1}{\sigma} \sum_{i=k+1}^T W_i$$

contra el índice i para $i=k+1, \dots, T$; en donde $\hat{\sigma}^2$ es el estimador de la varianza al ajustar un modelo común a todas las T observaciones. Este factor se introduce para estandarizar la magnitud de los residuales y puede ser calculado también por un procedimiento recursivo. Las sumas de cuadrados de los residuales usuales $S_r = (\underline{Y}_r - X_r \hat{\beta}_r)' (\underline{Y}_r - X_r \hat{\beta}_r)$ satisfacen la relación

$$S_r = S_{r-1} + W_r^2 \quad ; \quad r = k+1, \dots, T$$

y dado que $\hat{\sigma}^2 = S_T / (T-k)$ el cálculo es directo. A partir de las propiedades de w_i bajo H_0 , los autores deducen que la sucesión W_{k+1}, \dots, W_T tiene una distribución que puede ser aproximada por una normal tal que

$$E(N_x) = 0, \text{ Var}(N_x) = r-k \text{ y } \text{Cov}(N_x, W_g) = \min(s, r) - K.$$

De esta forma se puede establecer que para cada $t=k+1, \dots, T$, los límites para juzgar si el valor observado de W_t es significativa-

mente diferente a cero son de la forma $C_t = \pm \lambda \sqrt{t-k}$ en donde λ es el cuantil apropiado de una Normal estándar. Considerando a C_t como función de t , se obtienen las bandas de tolerancia del proceso para un nivel establecido de probabilidad. Tal como lo mencionan Brown et al., estos límites pueden ser reemplazados por bandas rectilíneas aunque en ese caso la probabilidad de cruzar las fronteras no es la misma en cada punto del proceso. De hecho, se propone el empleo de rectas que sean tangentes a las curvas originales en el punto medio entre K y T y se deriva la forma de estas rectas para cuando se fija una probabilidad global α de cruzarlas en cualquier punto del proceso. En el trabajo considerado se propone adicionalmente, otro tipo de sumas acumuladas que pueden ser utilizadas de forma complementaria a la sucesión W_t . Los autores consideran la suma acumulada de cuadrados de los residuales transformados;

$$V_x = \frac{\sum_{i=k+1}^x w_i^2}{\sum_{i=k+1}^T w_i^2}$$

$$= S_x / S_T \quad ; \quad i=k+1, \dots, T$$

que puede resultar muy útil, sobre todo en los casos en que se efectuen múltiples cambios que produzcan residuales de signos con

trarlos. En el artículo descrito se afirma que bajo H_0 , V_x tiene una distribución Beta de media $(r+k)/(T-k)$. Con base en este hecho se deriva una forma de construir bandas de tolerancia para el proceso generado por V_x .

Algunos otros procedimientos pueden ser utilizados para examinar la información disponible. Brown et al. sugieren graficar los valores estimados de los coeficientes de regresión cuando se ajusta un modelo a un segmento de n observaciones sucesivas y posteriormente este segmento se recorre a lo largo de la serie. Si no se ha producido ningún cambio, es de esperarse que las gráficas sean aproximadamente constantes. Para obtener un indicador adicional en este sentido pueden considerarse j segmentos disjuntos, cada uno con n observaciones sucesivas y calcular la estadística correspondiente a la prueba usual de igualdad de coeficientes de regresión.

Por otra parte, los autores sugieren que otra forma para determinar si los coeficientes sufren cambios en el tiempo puede consistir en modelar su comportamiento mediante polinomios en el tiempo. Así, en el modelo

$$Y_i = X_i' \beta_i + \epsilon_i, \quad i = 1, 2, \dots, T$$

se puede proponer que $\underline{\beta}_i$ sea expresado como

$$\underline{\beta}_i = \beta_{(1)} + i \beta_{(2)} + i^2 \beta_{(3)} + \dots + i^{m-1} \beta_{(m)}$$

en donde m es un entero seleccionado convenientemente de acuerdo a los objetivos del usuario y al contexto del problema. En estas condiciones la hipótesis de no cambio puede ser formulada en términos de los coeficientes $\underline{\beta}_{(j)}$. Explícitamente se puede probar las hipótesis

$$H_0: \beta_{(1)} = \beta_{(2)} = \dots = \beta_{(m)} = 0 \text{ vs } H_1: \beta_{(j)} \neq 0 \text{ alguna } j \in \{1, m\}$$

a través de los procedimientos usuales de regresión. Mas aún, se puede efectuar un ajuste secuencial de modelos de la forma

$$Y_i = \sum_{j=1}^h \beta_{-i}^{(j)} + e_i \quad ; \quad i = 1, 2, \dots, T$$

con

$$\beta_{-i}^{(h)} = \sum_{j=1}^h i^j \beta_{(j)} \quad \text{para } h = 1, 2, \dots, m$$

de tal forma que como resultado se obtenga el modelo de coeficientes polinomiales que mejor describe la serie de observaciones.

Como una sugerencia final se comenta la utilidad que tendría la

gráfica respecto a t de λ_t , el logaritmo del cociente de verosimilitudes para la prueba de las hipótesis

$$H_1: \beta_1 = \beta_2 = \dots = \beta_T \text{ vs } H_2: \beta_1 = \beta_2 = \dots = \beta_t = \beta^* \\ \beta_{t+1} = \beta_{t+2} = \dots = \beta_T = \beta^{**}; \beta^* \neq \beta^{**}$$

Esta estadística ya ha sido propuesta por Quandt [1] en relación a este problema. En efecto, en uno de sus artículos se plantea la utilización del mínimo de λ_t sobre t como estadística de prueba para la hipótesis de no cambio. Sin embargo, el criterio presenta inconvenientes en el sentido de que su distribución bajo la hipótesis nula no ha sido establecida en general. De cualquier forma Brown, Durbin y Evans sugieren que de nuevo como indicador, puede construirse la gráfica para obtener información sobre la estabilidad de la regresión. Mas aún, indican que todas las gráficas pueden ser construídas siguiendo el orden natural en el tiempo y también el orden inverso de modo que las anomalías comunes puedan ser empleadas para defectar el período en que manifiestan los cambios.

El trabajo de estos autores es particularmente interesante porque han desarrollado un paquete estadístico en el que incluyen todas

las técnicas que proponen. La aplicación que reportan de estas a diversos conjuntos de observaciones, parece indicar que sobre todo a un nivel exploratorio y descriptivo, producen resultados satisfactorios que pueden ser muy útiles. Otro atractivo adicional es que el trabajo analizado incluye una discusión en donde otros autores desarrollan una serie de comentarios complementarios; al margen de las consideraciones técnicas que ahí se consignan, las cuales en su mayoría extienden o confirman algunos de los resultados obtenidos, es ilustrativo mencionar los aspectos conceptuales que se examinan.

En primer lugar, se cuestiona el empleo de algunas estadísticas de prueba como indicadores exclusivamente y en ciertos casos se critica el enfoque exploratorio y descriptivo que emprenden Brown, Durbin y Evans, en términos de que finalmente es necesario proponer procedimientos que permitan realizar inferencias en el sentido estadístico, esto es, con una medida precisa de la incertidumbre que involucran. Respecto a este punto es importante tomar en cuenta, como se menciona en la misma discusión, que independientemente de que para el estudio de un fenómeno específico se cuente con técnicas esta-

dísticas para realizar inferencias precisas y más aún, en los casos en que estas no sean accesibles, los procedimientos exploratorios juegan un papel fundamental en el proceso de adquisición de conocimiento en tanto que, siguiendo la línea de pensamiento de Tukey [102], constituyen instrumentos para el examen de la información disponible desde múltiples y diversos puntos de vista que pueden hacer aparentes características que de otra forma pasarían desapercibidas.

Por otra parte, se menciona y vale la pena enfatizarlo, que la intención de algunas de las técnicas propuestas por los autores tienen notables puntos de contacto con el tratamiento de series de tiempo a través de teoría de control. Esta situación sugiere la importancia que en general, tiene la comunicación de los resultados que se obtienen por diversos especialistas en áreas de interés común. Particularmente, en Estadística sería muy interesante aprovechar la experiencia que en Teoría de Control y Econometría parece existir en relación al problema de puntos de cambio.

En un trabajo más reciente, Hawkins [44], se aborda el problema nuevamente con un enfoque esencialmente computacional. Fundamentalmente se pretende obtener una descripción adecuada de un conjunto de datos específico como en el trabajo de Brown, Durbin y Evans [15] aunque el énfasis es concentrado en las ventajas computacionales de la solución propuesta que son apreciables particularmente en los casos en que se tienen grandes series de datos. En este sentido, el trabajo de Hawkins guarda una estrecha relación con el de McGee y Carleton [62] que ya ha sido revisado.

El autor plantea la estructura básica del problema como sigue. Sea Y_1, Y_2, \dots, Y_T una sucesión de observaciones ordenadas respecto a alguna variable concomitante que puede ser el tiempo. Un modelo de regresión con puntos de cambio o de regresión por segmentos es uno en donde se tiene que:

$$\begin{aligned}
 Y_t &= f_1(t) + e_1(t) & (-\infty < t < \tau_1) \\
 Y_t &= f_2(t) + e_2(t) & (\tau_1 < t < \tau_2) \\
 &\vdots & \\
 Y_t &= f_k(t) + e_k(t) & (\tau_{k-1} < t < \infty)
 \end{aligned}$$

en donde $f_i(t)$ y $e_i(t)$ representan respectivamente, la función de regresión y un término de error en el i -ésimo segmento. Los valores τ_i ; $i=1,2,\dots,k-1$ constituyen los puntos de cambio. Esta estructura general agrupa una variedad de casos particulares. Así por ejemplo, si las funciones $f_i(t)$ son constantes, se tiene una sucesión de variables con puntos de cambio como las que se discuten en el capítulo I. Naturalmente, las diferentes variantes se obtienen modificando la distribución de los errores. El caso de regresión lineal simple se obtiene definiendo $f_i(t)=\alpha_i+\beta_i t$ y usualmentese asigna una distribución Normal a los errores. Los modelos de regresión lineal múltiple se definen de manera análoga. Aún en esas condiciones, se pueden establecer variantes de acuerdo a si se impone la restricción de continuidad sobre $f_i(t)$ y $f_{i+1}(t)$ en el punto τ_i o si se considera que los errores tienen la misma varianza en todos los segmentos.

El autor menciona la dificultad que aparece en los procesos de inferencia particularmente para producir estimadores de máxima verosimilitud cuando existen más de dos segmentos. Esta circunstancia parece deberse a que la función de verosimilitud es una función altamente

no lineal de los puntos de cambio de suerte que es de esperarse que la complejidad computacional para encontrar sus estimadores de máxima verosimilitud aumente exponencialmente con k . El propósito del artículo es presentar métodos de solución que involucren una dificultad mucho menor.

El caso específico considerado en el trabajo de Hawkins es el de modelos de regresión y se presenta como sigue. Sea Y_1, \dots, Y_T una sucesión de observaciones tales que:

$$Y_t = \underline{x}_t' \underline{\beta}_i + e_i(t) \quad ; \quad (\tau_{i-1} < t \leq \tau_i) \quad ; \quad i=1, 2, \dots, k$$

en donde \underline{x}_t es un vector de dimensión p de variables exógenas, $\underline{\beta}_i$ es el vector de dimensión p , de coeficientes de regresión en el i -ésimo segmento y $e_i(t)$ es un error aleatorio de distribución $N(0, \sigma_i^2)$ en el i -ésimo segmento. Adicionalmente se supone independencia entre los errores.

Hawkins no considera restricción alguna de continuidad en las funciones de regresión, supone además que los puntos de cambio coinciden con tiempos de observación y define $\tau_0=0$, $\tau_k=n$. En estas condiciones, considera los valores τ_i conocidos y define $n_i = \tau_i - \tau_{i-1}$, como el número

mero de puntos en el i -ésimo segmento. Más aún, si en el i -ésimo segmento Y_i representa el vector con n_i observaciones y X_i la matriz de diseño de dimensión $n_i \times p$, entonces el autor recuerda que los estimadores de máxima verosimilitud de β_i y σ_i^2 están dados por

$$\hat{\beta}_i = (X_i' X_i)^{-1} X_i' Y_i,$$

$$\hat{\sigma}_i^2 = (Y_i - X_i \hat{\beta}_i)' (Y_i - X_i \hat{\beta}_i) / n_i,$$

de esta forma, para valores $\tau_1, \tau_2, \dots, \tau_k$ de los puntos de cambio, la función de verosimilitud maximizada sobre $\beta_i, \sigma_i^2; i=1, 2, \dots, k$ tiene la siguiente expresión

$$L = \prod_{i=1}^k \frac{\sigma_i^{-n_i}}{\sigma_i^{-n_i}}$$

Por otro lado si en el modelo se incorpora la restricción de igualdad de varianzas en todos los segmentos se tiene que

$$L = \left[\prod_{i=1}^k n_i \hat{\sigma}_i^2 \right]^{-n/2}$$

En cualquiera de esas dos situaciones los estimadores de máxima

verosimilitud de $\tau_1, \tau_2, \dots, \tau_k$, si estos valores son desconocidos, se obtienen maximizando la correspondiente \mathcal{L} o bien minimizando $\sum_{i=1}^k n_i \log \hat{\sigma}_i^2$ ó $\sum_{i=1}^k n_i \hat{\sigma}_i^2$ respectivamente. El autor indica que la similitud que guardan las dos funciones objetivo permite resolver el problema de optimización con un procedimiento común. De hecho, si para un segmento de observaciones Y_x, Y_{x+1}, \dots, Y_s se ajusta un modelo de regresión como los descritos y S es la suma de cuadrados de los residuales correspondiente al ajuste de los $m=s-r+1$ puntos, se puede definir $Q(r;S) = S$ si se supone igualdad de varianzas en todos los segmentos y $Q(r,s) = m \ln(S/m)$ en caso contrario. De cualquier forma, con esta notación, la función objetivo, para encontrar los estimadores de máxima verosimilitud de los puntos de cambio, está dada por la expresión:

$$\mathcal{L} = \sum_{i=1}^k Q(\hat{r}_{i-1} + 1, \hat{\tau}_i)$$

Para la solución del problema puede considerarse la situación más general de ajustar un modelo de regresión con r segmentos a N puntos donde $r \leq k$ y $N \leq n$. Sea $F_x(N)$ el valor mínimo de \mathcal{L} para este problema. Hawkins indica que el modelo que produce ese valor mínimo debe cons

tar de $r-1$ segmentos ajustados a las observaciones Y_1, Y_2, \dots, Y_j y un segmento más ajustado al conjunto Y_{j+1}, \dots, Y_N para alguna j en donde el modelo con $r-1$ segmentos ajustado a Y_1, \dots, Y_j debe ser óptimo, esto es, debe producir el valor $F_{r-1}(j)$ ya que en caso contrario sería posible mejorar el ajuste con r segmentos a los N puntos. Este argumento, menciona el autor, es conocido como el principio de Optimalidad de Bellman [5] y sugiere la solución del problema original a través de programación dinámica. Naturalmente, las ecuaciones de recursión quedan planteadas como sigue:

$$F_1(N) = Q(1, N); \quad F_r(N) = \min_j \{ F_{r-1}(j) + Q(j+1, N) \}; \quad r > 1$$

En el caso de que no se tenga igualdad de varianzas, cada segmento debe contener al menos $p+1$ puntos de modo que el recorrido de j en la minimización es $(r-1)(p+1) \leq j \leq N-p-1$.

El problema original de k segmentos en n observaciones puede resolverse calculando $F_r(N)$ para toda $r \leq k$ y $N \leq n$ siempre que el ajuste tenga sentido. El resultado final es $F_k(n)$ y los estimadores de máxima verosimilitud de los puntos de cambio se obtienen de las ecuaciones en recursión:

$$F_{k-i}(\hat{r}_{k-i}) = F_{k-i-1}(\hat{r}_{k-i-1}) + Q(\hat{r}_{k-i-1} + 1, \hat{r}_{k-i})$$

Este método de solución tiene la ventaja adicional de que produce los valores $F_x(N)$ para $r \leq k$ y $N \leq n$ los cuales pueden ser de utilidad, particularmente si existe duda respecto a que el número de cambios, k , sea el apropiado.

El procedimiento de cómputo requiere, básicamente, el cálculo de las cantidades $Q(i, j)$; $j \geq j + \rho$ a partir de los cuales se obtienen las $F_x(N)$ y los estimadores de los puntos de cambio. En este punto los autores proponen el empleo de un método de cómputo que consideran eficiente y que consiste en calcular las sucesiones de valores $Q(S, n)$, $Q(S, n-1), \dots, Q(S, S+p)$ para $S=1, 2, \dots, n-p$. En una sucesión específica, cada término $Q(i, j)$ implica el ajuste de un modelo de regresión; sin embargo, la diferencia entre $Q(S, n-j)$ y $Q(S, n-j-1)$ solo estriba en que se elimina la última observación del grupo considerado para el primer término en el ajuste indispensable en el segundo. Este patrón permite, como lo indica Hawkins, el empleo de un algoritmo secuencial de actualización similar al presentado por Brown, Durbin y Evans [16],

en donde las estadísticas relevantes para el cálculo de $Q(S, n-j-1)$ pueden obtenerse a partir de las correspondientes a $Q(S, n-j)$. Concretamente, es fácil verificar que si

$$Y = \begin{bmatrix} Y_h \\ Y_{h+1} \end{bmatrix}, \quad X = \begin{bmatrix} X_h \\ X_{h+1} \end{bmatrix}$$

entonces,

$$\hat{\beta} = (X'X)^{-1} X'Y; \quad \hat{\beta}_h = (X_h' X_h)^{-1} X_h' Y$$

$$(h+1) \hat{q}_2 = (Y - X\hat{\beta})' (Y - X\hat{\beta}); \quad h \hat{q}_h^2 = (Y - X_h \hat{\beta}_h)' (Y - X_h \hat{\beta}_h)$$

en donde se satisface que:

$$\hat{\beta}_h = \{I + (X'X)^{-1} X_{h+1}' (I - X_{h+1}' X_{h+1})^{-1} X_{h+1}\}^{-1} \hat{\beta},$$

$$h \hat{q}_h^2 = (h+1) \hat{q}_2 + (I - X_{h+1}' X_{h+1})^{-1} (Y_{h+1} - X_{h+1}' \hat{\beta})^2$$

Con el empleo de estas recursiones es computacionalmente simple el cálculo de los términos de cada una de las sucesiones. El autor menciona que esta técnica de programación dinámica puede ser aplicada también si el criterio de ajuste no es precisamente el def

nimo cuadrados sino algún otro como por ejemplo, el de mínima desviación absoluta media o mínima desviación absoluta máxima e indica que la formulación sigue la misma línea general.

Como una consideración adicional, en el trabajo analizado se afirma que en caso de que para la misma serie de observaciones se ajusten dos modelos, con K y $K-1$ puntos de cambio respectivamente, los correspondientes estimadores de máxima verosimilitud no guardan necesariamente una relación. Más explícitamente, los estimadores del modelo con $K-1$ cambios no tienen porque coincidir con los obtenidos del modelo con k cambios. Sin embargo, según la opinión del autor, en la práctica se esperaría que para un cambio que realmente tuvo lugar, el correspondiente estimador del punto de cambio fuese estable en el sentido de que si se modifica el número de cambios posibles en el modelo, el valor estimado fuese el mismo. Este punto de vista, sugiere inmediatamente una modificación de la técnica propuesta por Hawkins para la estimación. Los puntos de cambio en un modelo con k cambio se pueden estimar, añadiendo a los estimados en un modelo con $K-1$ cambios, un punto más. Naturalmente, la solución en este caso es jerárquica y requiere, por tan-

to, de un volúmen aún más reducido de labores de cómputo. Sin embargo, esta ganancia en facilidad tiene como contraparte dos inconvenientes. En primer lugar, como se desconoce cuales puntos corresponden a cambios que efectivamente tuvieron lugar, el procedimiento jerárquico puede producir estimadores que en la etapa final pierden la importancia que parecían tener en una etapa inicial. Por otra parte y en estrecha relación con lo anterior, es claro que al proceder jerárquicamente, los estimadores que se obtienen no son, necesariamente, los de máxima verosimilitud, sino en el mejor de los casos, una aproximación. Ante estos inconvenientes, el autor propone finalmente una técnica que aunque tampoco garantiza la obtención de los estimadores de máxima verosimilitud, sí permite reevaluar la importancia de los estimadores incorporados en las etapas previas. La idea es ajustar un modelo con un punto de cambio; una vez que el conjunto de datos ha sido dividido, la misma técnica se aplica en cada uno de los subconjuntos resultantes y se selecciona el más adecuado de los dos candidatos obtenidos como el segundo punto de cambio. El procedimiento se prosigue hasta que se completa el número de puntos

de cambio establecidos por el modelo. Adicionalmente, en cada etapa se puede calcular la estadística usual, para probar la igualdad de coeficientes de regresión en los nuevos subconjuntos y determinar con esa información, si es conveniente efectuar la división. Esta nueva modalidad no es estrictamente jerárquica pero mantiene la dificultad de cómputo en un nivel muy bajo. Es muy interesante notar la similitud que guarda el procedimiento propuesto por Hawkins con el de McGee y Carleton [62]. En los dos casos se pretende obtener los estimadores de los puntos de cambio con algoritmos que sean eficientes desde el punto de vista computacional. La diferencia estriba en que se emplean diferentes funciones objetivo y además en la construcción del modelo final se procede en sentido contrario, ya que mientras McGee y Carleton utilizan una técnica de agrupamiento, Hawkins emplea una que se podría llamar de división. La relación entre ambas técnicas es tan natural que el autor incluye en su trabajo un ejemplo donde se aplican a un conjunto de datos reales, analizados previamente por McGee y Carleton, los métodos de programación dinámica original y modificado . Los resultados de ambos métodos

coinciden con los de esos autores cuando se supone igualdad de varianzas. En el caso de heterocedasticidad, los métodos producen resultados similares cuando se consideran 2,3,4 y 5 cambios pero difieren en el modelo con 6 cambios. La evidencia sugiere que el método modificado produce la mejor descripción de los datos.

Un trabajo más donde se aborda el problema de puntos de cambio en regresión, es el de Quandt y Ramsey [75]. En ese artículo, la situación de cambio es planteada en la misma manera que ya ha sido estudiada por Quandt, [74], en una contribución pro
viamente examinada en este capítulo. De hecho, el trabajo de Quandt y Ramsey puede considerarse una secuela del anterior. La idea fundamental es que se parte del supuesto de que si bien el modelo puede incluir cambios, no se posee información sobre las variables exógenas que influyen en ellos. De esta manera el problema puede ser visualizado como uno de mezcla de distribuciones en donde para cada observación, el fenómeno utiliza uno de entre varios modelos para generar la respuesta. El caso más sencillo, que es el considerado por los autores, es el de

regresión lineal simple en donde la naturaleza elige en cada observación, entre dos distintos modelos de regresión. Esta estructura puede describirse como sigue:

Sean Y_1, Y_2, \dots, Y_T observaciones tales que,

$$Y_i = a_1 + b_1 X_i + \varepsilon_{1i} \quad \text{con probabilidad } \lambda$$

$$Y_i = a_2 + b_2 X_i + \varepsilon_{2i} \quad \text{con probabilidad } 1 - \lambda ; i = 1, 2, \dots, T$$

en donde los errores ε_{ji} son independientes y además, $\varepsilon_{1i} \sim N(0, \sigma_1^2)$ y $\varepsilon_{2i} \sim N(0, \sigma_2^2)$. Los valores de la variable X son fijos y los parámetros a_1, b_1, σ_1^2 , a_2, b_2, σ_2^2 y λ son desconocidos. En el trabajo previo de Quandt [74] se propone la estimación de los parámetros mediante el método de máxima verosimilitud. Sin embargo, en esta nueva contribución se afirma que si bien las mezclas de poblaciones normales son identificables, el problema de estimación es generalmente complicado ya sea que se utilice mínimos cuadrados o máxima verosimilitud. La razón es que a pesar de que los estimadores que maximizan la verosimilitud en el interior del espacio parametral son consistentes, su determinación es difícil en la práctica. De hecho, Quandt [] utiliza un méto-

do de Análisis Numérico para la evaluación de los máximos y reporta algunos casos en que el algoritmo no converge después de un número considerable de iteraciones, este hecho se debe, generalmente, a que las ecuaciones normales o de verosimilitud asociadas son singulares. Adicionalmente, los autores comentan que las propiedades de los estimadores de máxima verosimilitud no se conocen para este tipo de problemas cuando se tiene un tamaño de muestra moderado. Por estas razones consideran importante investigar estimaciones alternativas que puedan evitar algunas de esas dificultades.

Quandt y Ramsey afirman que una alternativa, relativamente natural, es el método de momentos y proceden a examinarla primero, para el caso particular en que tanto b_1 como b_2 son conocidos, en cuyo caso el problema se reduce a considerar una mezcla que representa un modelo de cambio en una sucesión de variables aleatorias Normales como los considerados en el capítulo anterior de este trabajo.

En esas condiciones, solo se desconocen λ , σ_1^2 , σ_2^2 y μ_1, μ_2 las medias de las distribuciones. Los autores exhiben la fun-

ción de densidad conjunta de las observaciones que tiene la siguiente expresión:

$$f(\underline{Y}_T | \underline{\gamma}) = \prod_{i=1}^T \left\{ \lambda / (2\pi \sigma_i^2) \right\}^{1/2} \exp \left\{ - (Y_i - \mu_1)^2 / 2 \sigma_i^2 \right\} \\ + [(1-\lambda) / (2\pi \sigma_i^2)]^{1/2} \exp \left\{ - (Y_i - \mu_2)^2 / 2 \sigma_i^2 \right\}$$

en donde $\underline{\gamma}' = (\lambda, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$. El método de momentos implica la solución de cinco ecuaciones simultáneas, igualando los primeros cinco momentos muestrales y poblacionales, para la estimación de $\underline{\gamma}$. El procedimiento de solución requiere la determinación de una raíz negativa de una ecuación de grado nueve cuyos coeficientes dependen de los momentos muestrales. Salvo esta complicación, los autores afirman que se obtienen los estimadores sin mayor dificultad, aunque no se tiene una estimación de sus varianzas de las cuales solo se sabe que en general tienden a ser grandes porque involucran momentos de orden relativamente alto. Por esta razón, Quandt y Ramsey proponen otro método que parte de la consideración de la función generatriz de momentos de las observaciones. Para el caso considerado, la función ge-

neratriz de momentos esta dada por:

$$M_Y(\theta) \equiv E(e^{\theta Y}) = \lambda \exp\{\mu_1 \theta + a_1^2 \theta^2 / 2\} + (1-\lambda) \exp\{\mu_2 \theta + a_2^2 \theta^2 / 2\}$$

Considerando la muestra Y_1, Y_2, \dots, Y_T y k valores $\theta_1, \theta_2, \dots, \theta_k$ de θ en un intervalo (a, b) ; $a < 0 < b$, se tiene que $M_Y(\theta_j)$ puede ser estimado por:

$$\hat{M}_Y(\theta_j) = \frac{1}{T} \sum_{i=1}^T \exp\{\theta_j Y_i\}$$

Es fácil verificar, recurriendo a la ley fuerte de los grandes números que $\hat{M}_Y(\theta_j)$ converge a $M_Y(\theta_j)$ con probabilidad uno. Más aún, dado que para un tamaño de muestra finito la relación entre $\hat{M}_Y(\theta_j)$ y $M_Y(\theta_j)$ se puede expresar como

$$\hat{M}_Y(\theta_j) = M_Y(\theta_j) + u$$

en donde u es un error debido al muestreo, resulta razonable y atractivo desde un punto de vista intuitivo estimar los parámetros involucrados determinando los valores parametrales que minimizan

$$S_T(\underline{\theta}) = \sum_{j=1}^k (\hat{M}_Y(\theta_j) - M_Y(\theta_j))^2$$

para lo cual, los autores afirman que se puede utilizar el criterio de diferenciación respecto a $\underline{\gamma}$. Quandt y Ramsey aseveran que si k es igual a uno, se puede verificar que las ecuaciones normales asociadas resultan singulares. Más aún, aseguran que la singularidad de las ecuaciones depende tanto de k como de los valores específicos de $\theta_1, \theta_2, \dots, \theta_k$.

En estas circunstancias, proponen el empleo del método con $k > 5$ y $\theta_1, \theta_2, \dots, \theta_k$ valores que garanticen la no singularidad de las ecuaciones. Bajo tales supuestos, los autores verifican que el estimador $\hat{\underline{\gamma}}$ converge casi seguramente a $\underline{\gamma}$ y que además, asintóticamente $\sqrt{T}(\hat{\underline{\gamma}} - \underline{\gamma})$ tiene una distribución Normal multivariada de media $\underline{\theta}$ y una matriz de varianzas y covarianzas que depende tanto de $\underline{\gamma}$ como de $\underline{\theta}$.

La extensión de este tratamiento al caso de regresión lineal sigue la misma línea. La alternativa del método de momentos es examinada muy brevemente y se concluye, que debido al aumento en el número de parámetros involucrados, el procedimiento produce estimadores que dependen de momentos muestrales de orden superior a cinco, con la consecuente desventaja en términos de las

varianzas.

El método de la función generatriz de momentos es ilustrado para el caso de regresión lineal simple para el cual se tiene que:

$$M_{Y_j}(\theta) = E(e^{\theta Y_j}) = \lambda \exp\{(a_1 + b_1 x) \theta + \theta^2 \sigma_1^2 / 2\} + (1 - \lambda) \exp\{(a_2 + b_2 x) \theta + \theta^2 \sigma_2^2 / 2\}$$

En virtud de que para este modelo la función generatriz de momentos de Y_i depende de X_i , no se puede utilizar exactamente el mismo argumento que en el caso de sucesiones. Por esa razón los autores proponen una función objetivo que generaliza a la que ya ha sido propuesta. La expresión es la siguiente:

$$S_T(\underline{\theta}) = \sum_{j=1}^k \sum_{i=1}^T (Z_i(\theta_j) - M_{Y_i}(\theta_j))^2$$

en donde $Z_i(\theta_j) = \exp(\theta_j Y_i)$. El criterio de diferenciación respecto al vector de parámetros $\underline{\gamma}' = (\lambda, a_1, b_1, a_2, b_2, \sigma_1^2, \sigma_2^2)$ produce un sistema de siete ecuaciones que deben ser resueltas. Esto indica que en este caso, K debe ser mayor o igual a siete y además es claro que, en general, la solubilidad del sistema depende de la particular elección de $\theta_1, \theta_2, \dots, \theta_k$ que a su vez está con

dicionada por la estructura de los valores de la variable independiente X . De cualquier forma, los autores muestran que aún con estas modificaciones, la distribución asintótica de $\sqrt{T}(\hat{\gamma} - \gamma)$ es una Normal multivariada de media $\underline{0}$ y una matriz de varianzas y covarianzas que depende tanto de γ como de los valores de X .

Con la finalidad de evaluar los resultados que se obtienen con el método de la función generatriz de momentos, Quandt y Romsey desarrollan un estudio de simulación através del cual lo comparan con el método de momentos considerando tanto sucesiones de variables como problemas de regresión. Como un primer resultado reportan que en diversos casos y dependiendo, aparentemente, de los parámetros en la distribución, los dos métodos fallan en el sentido de que no es posible determinar el punto crítico o bien, los valores calculados yacen fuera del espacio parametral. Por esta razón y para los fines de la comparación, los autores repitieron la simulación en cada caso, tantas veces como fué indispensable hasta completar 50 ensayos en que ambos métodos produjeran resultados admisibles. La simulación indica que la elección de los valores $\theta_1, \theta_2, \dots, \theta_k$ resulta fundamental

para los cálculos numéricos. Valores grandes producen indeterminaciones mientras que los valores pequeños parecen producir una función objetivo casi constante. Considerando los tamaños muestrales empleados, $T=50, 100$; Quandt y Ramsey aplicaron la prueba de Shapiro-Wilk para Normalidad y reportan que en aproximadamente la mitad de los casos simulados de sucesiones, el método que proponen produce estimadores que se pueden considerar distribuidos Normalmente mientras que por momentos la fracción se reduce a un tercio aproximadamente.

En el caso de sucesiones, los métodos se comparan mediante el error cuadrático medio y el sesgo al cuadrado, encontrando que en general el método de la función generatriz de momentos produce estimadores más cercanos al valor verdadero de los parámetros. Para los modelos de regresión solo calculan los estimadores del método que proponen y observan que, como era de esperarse, las varianzas disminuyen a medida que el tamaño muestral aumenta y además reportan que los parámetros de la población más intensamente muestreada son, en general, más precisos.

En resumen, los autores proponen un método alternativo al de

máxima verosimilitud para la estimación que se aplica en problemas de puntos de cambio cuando no se cuenta con una variable exógena que permita ordenar las observaciones. En esas circunstancias la situación puede ser tratada como una mezcla de distribuciones. La idea original parece ser la de producir un método de estimación que no presente los inconvenientes de indeterminación asociados a máxima verosimilitud. Sin embargo, los resultados obtenidos por los autores indican que el método de la función generatriz de momentos también sufre de fallas de este tipo aunque presumiblemente se deban a otras causas. En general el método parece ser superior al de momentos pero la crítica más severa y muy natural por otra parte, es que habiendo propuesto su método como una alternativa a la de máxima verosimilitud, los autores no lo contrastan con ese método sino con el de momentos. Es evidente que antes de juzgar la aportación de Quandt y Ramsey será necesario efectuar las comparaciones requeridas.

P.M.Lerman es otro autor que ha contribuido a la literatura sobre el tema de este trabajo. Es una contribución reciente [56], propone un método de ajuste que puede ser aplicado incluso a modelos no lineales con puntos de cambio. La estructura del modelo es descrita

de la siguiente manera. Sea Y una variable aleatoria tal que:

$$E(Y | X) = f_1(X, \underline{\beta}_1); \quad X_0 < X < X_1$$

$$E(Y | X) = f_2(X, \underline{\beta}_2); \quad X_1 < X < X_2$$

⋮

$$E(Y | X) = f_r(X, \underline{\beta}_r); \quad X_{r-1} < X < X_r$$

en donde $E(Y|X)$ es una función continua de X en el intervalo $[X_0, X_r]$, X_0 , X_r y r son conocidos pero los vectores de parámetros $\underline{\beta}_1, \underline{\beta}_2, \dots, \underline{\beta}_r$ y los puntos de cambio $\underline{X} = (X_1, X_2, \dots, X_{r-1})$ se desconocen y deben ser estimados a partir de una muestra. Es evidente que la continuidad de $E(Y|X)$ impone restricciones a las funciones de regresión. Específicamente, se tiene que:

$$f_j(X_j, \underline{\beta}_j) = f_{j+1}(X_j, \underline{\beta}_{j+1}); \quad j=1, 2, \dots, r-1$$

Usualmente, se supone que las observaciones en Y son independientes con distribución Normal y varianza σ_j^2 en el j -ésimo régimen de regresión. Una formulación alternativa consiste en suponer que la varianza de Y_i es de la forma σ^2/w_i , con w_i conocido y σ^2 por estimarse. En esas condiciones y a partir de una muestra aleatoria $(Y_i, X_i); i=1, 2, \dots, T$, los estimadores de mínimos

cuadrados tanto de los parámetros como de los puntos de cambio, se obtienen minimizando la función:

$$S = \sum_{j=1}^r \sum_{x_{j-1} < x_j} w_j \{ Y_i - f_j(x_i, \beta_j) \}^2$$

bajo las restricciones de continuidad. La solución a este problema es en general difícil, como ya se ha discutido en este trabajo. Particularmente, la función objetivo no es diferenciable en todos los casos y más aún, puede existir mínimos locales que distorsionen los resultados de los métodos numéricos utilizados.

Es interesante notar el papel que juegan las restricciones de continuidad. Si éstas son ignoradas, el problema se simplifica en cierto grado y la solución al problema de estimación por mínimos cuadrados puede obtenerse, utilizando la técnica propuesta por Hawkins [44] que ya ha sido descrita. En el caso de modelos lineales con un solo punto de cambio, la restricción correspondiente produce una generalización del modelo de regresión de dos fases tratado por Hinkley [45] en donde la complejidad de la solución es evidente. Una forma de proceder para la estimación en modelos lineales ha sido propuesta por Hudson [21],

y consiste en considerar todas las particiones posibles del conjunto de datos y ajustar en cada segmento la regresión $f(X, \beta)$ correspondiente. Con esta información se calculan los puntos de intersección y el modelo se considera admisible si las intersecciones inducen la misma partición original. Para los modelos admisibles se registra el valor de la función S mientras que con el resto se procede a imponer explícitamente la restricción de continuidad en el cálculo de S . Esta técnica reduce, en general, la dificultad de cómputo, especialmente debido a que en modelos lineales los puntos de intersección se pueden calcular muy fácilmente a partir de los parámetros estimados. Sin embargo, el procedimiento de Hudson resulta poco apropiado, en opinión de Lerman, cuando las funciones de regresión no son lineales ya que en ese caso de cualquier forma se requiere de algoritmos de aproximación sucesiva que suelen ser costosos y complicados. Como alternativa, Lerman propone una técnica que sugiere a partir de la consideración de que en general, cuando los valores X_1, X_2, \dots, X_{r-1} son conocidos, la evaluación de la estadística S es bastante más simple. De esta forma, se propone calcular $S_n(\underline{X})$, el valor mini

mo de S cuando el vector de puntos de intersección es \underline{X} , para cada \underline{X} y a partir de la gráfica de esta función, determinar \hat{S} , el valor mínimo global. Este técnica, menciona el autor, es comunmente empleada en problemas de regresión no lineal y tiene la ventaja de que puede ser empleada con ventaja aún cuando las funciones de regresión sean no lineales minimizando la función objetivo con un algoritmo conveniente y además produce las estadísticas necesarias para realizar inferencias por máxima verosimilitud.

Cuando se pretende probar una hipótesis sobre el modelo con r cambios, el método de cociente de verosimilitudes conduce a la estadística de prueba

$$R_1 = -2 \ell_n \lambda = T \ell_n (S_0 / \hat{S}) .$$

en donde S_0 representa el valor mínimo de S bajo la hipótesis nula. Esta estadística, indica Lerman, es en ocasiones aproximada por:

$$R_2 = (S_0 - \hat{S}) / s^2$$

con $s^2 = \hat{S} / (T-p)$, en donde p es la dimensión del espacio parame-

tral.

El autor se refiere a un trabajo de Feder [53], para afirmar que si efectivamente existen cambios y las regresiones adyacentes son distintas, entonces asintóticamente, R_1 y R_2 tienen una distribución $X^2_{(q)}$, donde q es la reducción en dimensión debida a la hipótesis bajo prueba. Adicionalmente se tiene que, también asintóticamente, los estimadores de $\beta_1, \beta_2, \dots, \beta_r$ y \underline{X} tienen una distribución Normal.

Para tamaños de muestra finitos, Feder [54] ha demostrado que la distribución de R_1 / q y R_2 / q puede ser aproximada por una F con q y $T-p$ grados de libertad. Estos resultados sugieren una forma de construir regiones de confianza para el vector de puntos de cambio \underline{X} . Para un nivel de confianza α una región correspondiente está dada por los valores de \underline{X} que satisfacen la relación

$$S_m(\underline{X}) < C^\alpha.$$

en donde $C^\alpha = C_1^\alpha = \hat{S} \exp \{ (r-1) F_{r-1, T-p}^{1-\alpha} \}$ si se emplea R_1 y $C^\alpha = C_2^\alpha = \hat{S} + s^2 (r-1) F_{r-1, T-p}^{1-\alpha}$. La determinación de la región se puede efectuar de manera muy simple, en los casos de no más de

un cambio, a partir de la gráfica de $S_m(\underline{X})$.

El autor ilustra la técnica propuesta con tres ejemplos en donde se consideran modelos con uno y dos cambios así como con funciones de regresión no lineales. En general, se ilustra la importancia que tiene la gráfica de $S_m(\underline{X})$ particularmente porque como ya se ha comentado, es posible que existan varios mínimos locales que proporcionen información adicional sobre la naturaleza del fenómeno, que deba ser mencionada. Este hecho puede provocar que en ocasiones se obtengan regiones de confianza no conexas. Cuando se tienen dos puntos de cambio, la representación de $S_m(\underline{X})$ se puede realizar mediante el trazado de curvas de nivel de la función en la región $X_1 < X_2$. Cuando el número de cambios excede a dos, el autor sugiere la construcción de gráficas respecto a subconjuntos del vector \underline{X} .

Finalmente, Lerman indica que una simplificación adicional se puede utilizar si las regresiones son lineales y la restricción de continuidad se impone para un conjunto específico de valores de \underline{X} . En ese caso, la estimación de los parámetros β_1, \dots, β_r se puede llevar a cabo utilizando multiplicadores de Lagrange como es

usual en el modelo lineal.

En conclusión, Lerman propone una técnica para la determinación de los estimadores por mínimos cuadrados, de los parámetros de los modelos de regresión y los puntos de cambio que reduce la dificultad de cálculo, produce las estadísticas necesarias para realizar inferencias a través del método de máxima verosimilitud y permite además, considerar funciones de regresión que no sean lineales. La recomendación de representar gráficamente los valores de $S_m(\underline{X})$ parece ser particularmente valiosa para obtener una idea más completa del comportamiento del fenómeno bajo estudio.

Esta sección será concluida comentando un trabajo más, Mendoza [63], en donde se examina el tratamiento, que ya ha sido discutido , propuesto por Quandt [73] para el caso de un modelo de regresión lineal simple con un solo punto de cambio. En ese trabajo Quandt considera el modelo:

$$Y_i = a_1 + b_1 X_i + \epsilon_{1i}; \quad i = 1, 2, \dots, r,$$

$$Y_i = a_2 + b_2 X_i + \epsilon_{2i}; \quad i = r+1, \dots, T$$

en donde, a_1, a_2, b_1 y b_2 son parámetros desconocidos, al igual que τ y los errores ϵ_{1i} y ϵ_{2i} son independientes y Normalmente distribuidos con esperanza cero y varianzas σ_1^2 y σ_2^2 respectivamente. Para probar la hipótesis nula de no cambio, Quandt propone la división del conjunto de datos $(X_1, Y_1), (X_2, Y_2), \dots, (X_T, Y_T)$ en dos grupos de la forma $(X_1, Y_1), \dots, (X_t, Y_t)$ y $(X_{t+1}, Y_{t+1}), \dots, (X_T, Y_T)$ en donde t puede ser el valor estimado de τ por máxima verosimilitud o bien un valor asignado exógenamente. Para cada grupo se ajusta un modelo de regresión lineal por separado,

$$\hat{Y}_i = \hat{a}_1 + \hat{b}_1 X_i ; \quad i=1, 2, \dots, t$$

$$\hat{Y}_i = \hat{a}_2 + \hat{b}_2 X_i ; \quad i= t+1, \dots, T$$

y en el trabajo referido, Quandt propone la construcción de los residuales cruzados:

$$r_{1i} = Y_i - \hat{a}_2 - \hat{b}_2 X_i ; \quad i= 1, 2, \dots, t$$

$$r_{2i} = Y_i - \hat{a}_1 - \hat{b}_1 X_i ; \quad i= t+1, \dots, T$$

en base a los cuales propone algunos criterios para probar la hipótesis nula de no cambio, esto es, $H_0: b_1=b_2, a_1=a_2$ y $\sigma_1^2 = \sigma_2^2$.

La idea básica del autor es que bajo $H_0, r_{1i}; i=1, 2, \dots, t$ es

una colección de variables aleatorias independientes, distribuidas según una ley Normal. Una afirmación similar es postulada por Quandt para los residuales r_{2i} , $i=t+1, \dots, T$.

En esas condiciones, se propone el empleo de las estadísticas:

$$V_1 = t^{1/2} \bar{r}_1 / \left\{ \sum_{i=1}^t (r_{1i} - \bar{r}_1)^2 / (t-1) \right\}^{1/2} \quad \gamma$$

$$V_2 = (T-t)^{1/2} \bar{r}_2 / \left\{ \sum_{i=t+1}^T (r_{2i} - \bar{r}_2)^2 / (T-t-1) \right\}^{1/2}$$

en donde:

$$\bar{r}_1 = \frac{1}{t} \sum_{i=1}^t r_{1i}, \quad \bar{r}_2 = \frac{1}{T-t} \sum_{i=t+1}^T r_{2i}$$

Bajo la hipótesis nula, afirma Quandt, V_1 y V_2 siguen una distribución de Student con $t-1$ y $T-t-1$ grados de libertad si t es un valor asignado exógenamente. El criterio, entonces, sería rechazar H_0 si V_1 ó V_2 son significativamente diferentes de cero. El empleo del valor del estimador de máxima verosimilitud de τ para dividir el conjunto original de datos, modifica la distribución de las estadísticas de prueba pero podría ser utilizando estando consciente de que en general en esa situación

se rechazará la hipótesis nula con mayor probabilidad. Una dificultad adicional que surge a partir de la verificación de que tanto el valor esperado de \bar{r}_1 como el de \bar{r}_2 pueden ser iguales a cero bajo la hipótesis alternativa, si bien con probabilidad prácticamente cero, lleva al autor a proponer un criterio más. En este caso considera los cocientes

$$W_1 = \frac{\sum_{i=1}^t r_{i1}^2}{\sum_{i=t+1}^T e_{i1}^2}, \quad W_2 = \frac{\sum_{i=t+1}^T r_{i1}^2}{\sum_{i=1}^t e_{i1}^2}$$

en donde:

$$e_{1i} = Y_i - \hat{a}_1 - \hat{b}_1 X_i; \quad i = 1, 2, \dots, t$$

$$e_{2i} = Y_i - \hat{a}_2 - \hat{b}_2 X_i; \quad i = t+1, \dots, T$$

esto es, e_{1i} y e_{2i} son los residuales usuales de los dos modelos de regresión.

Quandt afirma que tanto para W_1 como W_2 , bajo la hipótesis nula y si t es un valor determinado exógenamente, el numerador y el denominador en cada cociente son independientes y siguen una distribución X^2 con $t-1$ y $T-t-1$ grados de libertad respectivamente en el caso de W_1 y con $T-t-1$ y $t-1$ para W_2 . Independientemente

te del error evidente en la asignación de los grados de libertad de los denominadores, el autor afirma que la distribución tanto de W_1 como de W_2 se puede completar a una F para rechazar H_0 si cualquiera de los dos valores calculados excede el cuantil apropiado. La contribución del trabajo de Mendoza, consiste en mostrar que las afirmaciones de Quandt respecto a las propiedades distribucionales de las estadísticas propuestas, son en general falsas y además proponer un posible tratamiento alternativo. Para verificar estos hechos resulta conveniente recurrir a la notación matricial. Sean,

$$\underline{Y}_1 = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_t \end{bmatrix} ; \quad X_1 = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_t \end{bmatrix} ; \quad \underline{\beta}_1 = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} ; \quad \underline{\varepsilon}_1 = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1t} \end{bmatrix}$$

$$\underline{Y}_2 = \begin{bmatrix} Y_{t+1} \\ Y_{t+2} \\ \vdots \\ Y_T \end{bmatrix} ; \quad X_2 = \begin{bmatrix} 1 & X_{t+1} \\ 1 & X_{t+2} \\ \vdots & \vdots \\ 1 & X_T \end{bmatrix} ; \quad \underline{\beta}_2 = \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} ; \quad \underline{\varepsilon}_2 = \begin{bmatrix} \varepsilon_{2t+1} \\ \varepsilon_{2t+2} \\ \vdots \\ \varepsilon_{2T} \end{bmatrix}$$

En estos términos, los modelos por ajustar a los dos grupos de observaciones por separado son:

$$\underline{Y}_1 = X_1 \underline{\beta}_1 + \underline{\varepsilon}_1 ,$$

$$\underline{Y}_2 = X_2 \underline{\beta}_2 + \underline{\varepsilon}_2$$

en donde $\underline{\varepsilon}_1$ y $\underline{\varepsilon}_2$ son independientes y distribuidos Normalmente con vector de medias cero y matriz de varianzas y covarianzas $\sigma_1^2 I_t$ y $\sigma_2^2 I_{T-t}$ respectivamente. Como consecuencia se tiene que,

$$\underline{Y}_1 \sim N(X_1 \underline{\beta}_1, \sigma_1^2 I_t) , \underline{Y}_2 \sim N(X_2 \underline{\beta}_2, \sigma_2^2 I_{T-t})$$

y son independientes. La hipótesis nula bajo consideración queda planteada como:

$$H_0: \underline{\beta}_1 = \underline{\beta}_2 = \underline{\beta} , \sigma_1^2 = \sigma_2^2 = \sigma^2 .$$

de tal manera que bajo H_0 ,

$$\underline{Y}_1 \sim N(X_1 \underline{\beta}, \sigma^2 I_t) , \underline{Y}_2 \sim N(X_2 \underline{\beta}, \sigma^2 I_{T-t}) .$$

Procediendo a la estimación usual por mínimos cuadrados, se obtiene que

$$\hat{\underline{\beta}}_1 = (X_1' X_1)^{-1} X_1' \underline{Y}_1 , \hat{\underline{\beta}}_2 = (X_2' X_2)^{-1} X_2' \underline{Y}_2$$

son dos estimadores independientes de $\underline{\beta}$ utilizando los dos modelos de regresión por separado. Más aún si se definen las predicciones cruzadas,

$$\tilde{\underline{Y}}_1 \equiv X_1 \hat{\underline{\beta}}_2 = X_1 (X_2' X_2)^{-1} X_2' \underline{Y}_2 \quad y$$

$$\tilde{\underline{Y}}_2 \equiv X_2 \hat{\underline{\beta}}_1 = X_2 (X_1' X_1)^{-1} X_1' \underline{Y}_1$$

se puede verificar que:

$$\tilde{Y}_1 \sim N(X_1 \underline{\beta}, \sigma^2 X_1 (X_1' X_2)^{-1} X_1'),$$

$$\tilde{Y}_2 \sim N(X_2 \underline{\beta}, \sigma^2 X_2 (X_2' X_1)^{-1} X_2')$$

y que además, son independientes. A partir de estas magnitudes los residuales cruzados de Quandt se pueden calcular como sigue:

$$\underline{r}_1 = \tilde{Y}_1 - \underline{Y}_1 = X_1 (X_1' X_2)^{-1} X_2' \underline{Y}_2 - \underline{Y}_1$$

$$\underline{r}_2 = \tilde{Y}_2 - \underline{Y}_2 = X_2 (X_2' X_1)^{-1} X_1' \underline{Y}_1 - \underline{Y}_2$$

En virtud de la independencia de \underline{Y}_1 y \underline{Y}_2 es sencillo verificar que bajo H_0 ,

$$\underline{r}_1 \sim N(\underline{0}, \sigma^2 (I_t + X_1 (X_1' X_2)^{-1} X_1')),$$

$$\underline{r}_2 \sim N(\underline{0}, \sigma^2 (I_{T-t} + X_2 (X_2' X_1)^{-1} X_2'))$$

Una de las primeras y más importantes afirmaciones de Quandt es que las entradas de \underline{r}_1 son independientes entre sí así como lo son las de \underline{r}_2 . Examinando el caso de \underline{r}_1 , sin embargo, se tiene que si se define,

$$C = (C_{1j}) \equiv I_t + X_1 (X_1' X_2)^{-1} X_1'$$

entonces,

$$C_{ii} = \frac{\sum_{k=t+1}^T (X_k - X_i)^2}{(T-t) \sum_{k=t+1}^T (X_k - \bar{X}_T)^2} + 1; \quad i=1, 2, \dots, t$$

$$C_{ij} = \frac{\sum_{k=t+1}^T (X_k - X_i)(X_k - X_j)}{(T-t) \sum_{k=t+1}^T (X_k - \bar{X}_T)^2}; \quad i \neq j; \quad i, j=1, 2, \dots, t$$

en donde,

$$\bar{X}_T = \frac{\sum_{k=t+1}^T X_k}{(T-t)}$$

En el trabajo de Mendoza, se afirma que a partir de estas expresiones se tiene que C es una matriz de rango completo cuya inversa está dada por:

$$C^{-1} = I_t - X_1 (X_1 X_1 + X_2 X_2)^{-1} X_1'$$

Además, para que las entradas de \underline{x}_1 sean independientes entre sí es indispensable que:

$$\sum_{k=t+1}^T (X_k - X_i)(X_k - X_j) = 0 \quad i \neq j; \quad i, j=1, 2, \dots, t.$$

Sin embargo, como se indica en el mismo trabajo, es fácil de comprobar que para un valor fijo de X_i , el cual determina un renglón de C, existe un único valor de X_j tal que $C_{ij}=0$. Esto im-

plica que, en el mejor de los casos, la matriz C tiene un cero en cada renglón y consecuentemente, las entradas de r_1 no son todas independientes entre sí. Como resultado de este hecho se tiene que, en general, la suma de cuadrados de las desviaciones de los residuales cruzados respecto a su media no sigue una distribución Ji cuadrada con $t-1$ grados de libertad como afirma Quandt. Otro hecho que se puntualiza es que la matriz C no es, en general, idempotente ya que la única matriz idempotente de rango completo es la idéntica. Esto implica que las sumas de cuadrados utilizadas como numeradores en los cocientes W_1 y W_2 de Quandt no son tampoco distribuidas como una Ji-cuadrada. Es tos resultados que invalidan el tratamiento propuesto, pueden ser utilizados, sin embargo, para corregir las fallas. Particularmente, el hecho de que la matriz C sea positiva definida garantiza la existencia de una matriz H única, simétrica y positiva definida tal que:

$$C^{-1} H H = H^2$$

esta matriz es usualmente conocida como la raíz cuadrada positiva de C y puede emplearse para construir un conjunto de residua

les cruzados transformados que satisfagan las afirmaciones de Quandt.

Sea $\underline{Z}_1 = H^{-1} \underline{r}_1$. Bajo la hipótesis nula, \underline{Z}_1 tiene una distribución Normal de media $\underline{0}$ y matriz de covarianzas

$$\begin{aligned} \text{COV}(\underline{Z}_1) &= \sigma^2 H^{-1} C H^{-1} \\ &= \sigma^2 H^{-1} H H H^{-1} \\ &= \sigma^2 I_t I_t \\ &= \sigma^2 I_t \end{aligned}$$

Naturalmente, estos residuales transformados tienen todos la misma varianza y son además, independientes. Un tratamiento completamente análogo puede emplearse para transformar \underline{r}_2 en \underline{Z}_2 . Estos nuevos residuales pueden ser empleados para aplicar las ideas de Quandt con respecto a V_1 y V_2 .

En el artículo considerado, la transformación se ilustra con los datos simulados por Quandt [72] que ya han sido mencionados y se obtiene que las estadísticas V_1 y V_2 resultan significativas a un nivel del 5% si se usa $t = T/2$, un valor asignado exógenamente, mientras que el nivel alcanza el 1% si se emplea

el estimador de máxima verosimilitud de τ , $\hat{\tau}=12$, que coincide con el verdadero valor del punto de cambio. Al considerar las estadísticas W_1 y W_2 , se menciona que si bien la transformación garantiza que los respectivos numeradores siguen en efecto una distribución Ji-cuadrada, no elimina la dificultad de probar que sean independientes de los denominadores respectivos.

Finalmente, se comenta que si bien la matriz H resulta una elección atractiva para efectuar la transformación propuesta, no es la única posibilidad. De hecho cualquier matriz U tal que $UCU' = I_t$ puede ser utilizada para aplicar el procedimiento propuesto. Esto sugiere que es necesaria mayor investigación para definir criterios que conduzcan a una selección óptima de la transformación. Más aún se indica que el procedimiento puede ser generalizado utilizando la distribución conjunta de \underline{z}_1 y \underline{z}_2 . De esta forma se puede utilizar una transformación global que produzca los nuevos residuales con las propiedades requeridas y que además satisfagan la propiedad de independencia entre \underline{z}_1 y \underline{z}_2 .

En resumen, en el trabajo de Mendoza, se examinan con detalle los criterios propuestos por Quandt para probar la hipótesis de

cambio en un modelo de regresión lineal simple con un punto de cambio. Se detectan algunos errores que se corrigen y se propone una transformación que garantiza la validez de las afirmaciones necesarias para la construcción de las estadísticas de prueba.

III. 2. RESULTADOS DEL ENFOQUE BAYESIANO

En el caso de modelos de regresión con puntos de cambio tra
tados desde el punto de vista Bayesiano, uno de los primeros tra
bajos es el debido a Halpern [42], en donde se considera el pro
blema de prueba de hipótesis y predicción en estas condiciones.
El autor presenta el problema con algunos comentarios relativos
al hecho de que, independientemente de que el contexto del fenó
meno bajo estudio sugiera la existencia de puntos de cambio en
un modelo de regresión, es un resultado bien establecido que
cualquier función continua puede ser aproximada por segmentos
mediante líneas rectas, siempre que el número de segmentos sea
suficientemente grande. De esta manera, Halpern motiva el estu
dio de modelos de regresión lineal con puntos de cambio, no solo
para describir fenómenos en donde efectivamente se produzcan ta
les cambios de manera intrínseca, sino también para proporcionar
una descripción razonablemente aproximada de algunos otros cuyo
comportamiento sea no lineal. En cualquier forma, el autor afir
ma que en el ajuste de modelos lineales por trazos, sea cual sea
el objetivo, usualmente se ignora el número de segmentos que se
han de emplear. Sin embargo, como se menciona en este trabajo,

en la literatura generalmente se supone que el número de segmentos es conocido pero se desconoce la localización de los puntos de cambio. Halpern indica que una suposición igualmente razonable es que se conocen todas las posibles localizaciones de los puntos de cambio aunque se ignora en cuales de ellas efectivamente se efectuó un cambio. La estructura del problema es planteada como sigue:

Sean Y_1, Y_2, \dots, Y_T observaciones independientes tales que:

$$Y_i = f(X_i) + \varepsilon_i ; i = 1, 2, \dots, T ; X_i \in [a, b]$$

en donde los errores $\varepsilon_i ; i = 1, 2, \dots, T$ son independientes e idénticamente distribuidos $N(0, \sigma^2)$. Adicionalmente, la función de regresión $f(x)$ es continua y lineal por segmentos, de tal manera que se puede representar como

$$f(x) = M_i + \alpha_i (X - Z_i), \forall X \in (Z_i, Z_{i+1}), i = 1, 2, \dots, q.$$

Naturalmente, los puntos de cambio posibles son conocidos y satisfacen

$$a = Z_1 < Z_2 < \dots < Z_q < Z_{q+1} = b ;$$

Es fácil de verificar que si se definen las funciones

$$U_0(X) \equiv 1$$

$$U_i(X) = \begin{cases} 0 & X < Z_i \\ X - Z_i & X \geq Z_i \end{cases} ; i = 1, 2, \dots, q$$

entonces $f(x)$ puede expresarse como una combinación lineal

$$f(X) = \sum_{i=0}^q \beta_i U_i(X)$$

en donde los parámetros $\beta_0, \beta_1, \dots, \beta_q$ son desconocidos al igual que lo son (M_i, a_i) ; $i = 1, 2, \dots, q$ en la formulación original. El tratamiento propuesto por Halpern consiste en definir, para cada subconjunto S de $\{2, 3, \dots, q\}$, el modelo M_S que considera como puntos de cambio, los correspondientes a los elementos de S . Así y en virtud de que existen $Q = 2^{q-1}$ subconjuntos de puntos de cambio, el problema de inferencia puede ser planteado como la discriminación entre las Q hipótesis H_k ; $k = 1, 2, \dots, Q$ en donde H_k es la hipótesis bajo la cual el modelo que describe el fenómeno bajo estudio es M_{S_k} . Si se denota por $r(k)$ el número de elementos en S_k y estos se designan, en orden creciente, como

$r(k,1), r(k,2), \dots, r(k,r(k))$, entonces bajo H_k los únicos puntos de cambio son

$$Z_{r(k,1)}, Z_{r(k,2)}, \dots, Z_{r(k,r(k))}$$

Más aún, la hipótesis H_k establece que en la expresión

$$f(x) = \sum_{i=0}^q \beta_i U_i(X),$$

$\beta_i \neq 0$ si y solo si $i \in S_k$. En estas condiciones, es claro que, si la hipótesis H_k es cierta, entonces el vector de observaciones \underline{Y} tiene una distribución que puede ser descrita como sigue

$$\underline{Y} \sim N(U B_k \beta_k, \sigma^2 I_t)$$

en donde U es una matriz de dimensiones $T \times (q+1)$ cuyas entradas están dadas por

$$U_{ij} = U_{j-1}(X_i),$$

β_k es una matriz de dimensiones $(q+1) \times (r(k)+2)$ cuyo elemento típico $b_{ji}^{(k)}$ satisface la relación

$$b_{ji}^{(k)} = \begin{cases} 1 & \text{si } j=1=1, j=1=2, j=r(k,i-2)+1 \\ 0 & \text{en otro caso} \end{cases}$$

y el vector $\underline{\beta}_k$ está dado por

$$\underline{\beta}_k = \begin{cases} (\beta_0, \beta_1)' & \text{si } r(k) = 0 \\ (\beta_0, \beta_1, \beta_{j(k,1)}, \dots, \beta_{r(k,r(k))})' & \text{si } r(k) > 0 \end{cases}$$

Con esta estructura, el problema se reduce a la asignación de probabilidades iniciales a las hipótesis H_k , y dada cada hipótesis, de distribuciones iniciales a los parámetros $\underline{\beta}_k$ y σ^2 . Esta información se combina con la proveniente de la muestra para obtener la información final o posteriori y en su caso, tomar las decisiones correspondientes. Halpern menciona que las hipótesis H_1, H_2, \dots, H_Q son excluyentes de tal forma que una y solo una de ellas puede ser cierta a la vez. Como consecuencia, la distribución condicional de $\underline{\beta}_k$ dada H_k debiera de reflejar la imposibilidad de que alguna de las entradas del vector de parámetros sea cero pues en tal caso el modelo coincidiría con el asociado a alguna otra hipótesis. Sin embargo, por facilidad y considerando que el subconjunto del espacio parametral formado por vectores con alguna entrada igual a cero tiene probabilidad nula

para cualquier distribución no degenerada, se ignora la anomalía y se supone que a priori, dadas H_k y σ^2 , $\underline{\beta}_k$ sigue una distribución Normal multivariada de media $\underline{\mu}_k$ y matriz de varianzas y covarianzas $\sigma^2 \Sigma_k$. Por otro lado, a la varianza σ^2 le es asignada dada H_k , una distribución marginal inicial Gamma inversa de parámetros η_k y γ_k mientras que la probabilidad a priori de H_k se denota por P_k . Esta asignación de distribuciones iniciales naturalmente corresponde al esquema de familias conjugadas, que es usual en el enfoque Bayesiano y se emplea por facilidad cuando el conocimiento inicial puede ser descrito de manera razonablemente adecuada en esos términos. Como consecuencia de esta estructura se tiene que a posteriori, respecto a una muestra $(Y_1, X_1), \dots, (Y_T, X_T)$, $\underline{\beta}_k$ tiene una distribución, dadas H_k y σ^2 , Normal multivariada de media $\underline{\mu}_k^*$ y matriz de varianzas y covarianzas $\sigma^2 \Sigma_k^*$, en donde

$$\underline{\mu}_k^* = \Sigma_k^* (\Sigma_k^{-1} \underline{\mu}_k + B_k' U' \underline{Y})^{-1}$$

$$\Sigma_k^* = \{\Sigma_k^{-1} + B_k' U' U B_k\}^{-1}$$

por su parte σ^2 , dada H_k , tiene una distribución Gama inversa de parámetros

$$\gamma_k^* = \gamma_k + T$$

$$\eta_k^* = \frac{1}{\gamma_k^*} \{ \eta_k \gamma_k + \underline{\mu}_k' \Sigma_k^{-1} \underline{\mu}_k - \underline{\mu}_k^* \Sigma_k^{*-1} \underline{\mu}_k^* + \underline{Y}' \underline{Y} \}$$

mientras que la probabilidad posteriori de H_k está dada por

$$P_k^* \propto P_k \frac{(\frac{1}{2} \eta_k \gamma_k)^{\frac{1}{2}} \gamma_k \Gamma(\frac{1}{2} \gamma_k^*) |\Sigma_k^*|^{-\frac{1}{2}}}{(\frac{1}{2} \eta_k^* \gamma_k^*)^{\frac{1}{2}} \gamma_k^* \Gamma(\frac{1}{2} \gamma_k) |\Sigma_k|^{-\frac{1}{2}}}$$

Si antes de obtener la información muestral la información con que se cuenta no es fácil de describir, lo cual puede ocurrir sobre todo en el caso de los vectores $\underline{\beta}_k$, o bien, si la información muestral no se quiere perjudicar con la información inicial, una alternativa es utilizar distribuciones iniciales de referencia que se obtienen, según la regla usual cuando se emplean familias conjugadas, haciendo tender los parámetros de la distribución inicial a un límite apropiado de modo que la distribución final solo dependa de la información muestral. Este procedimiento generalmente produce distribuciones impropias que

en el caso de inferencias en el modelo de regresión lineal común se transforman en propias a posteriori. La estructura de los modelos considerados en este problema no permiten, en general, garantizar que las distribuciones finales sean propias cuando las iniciales no lo son. Halpern utiliza como distribuciones de referencia, $f(\sigma^2) = \sigma^{-2}$ para la varianza y $f(\underline{\beta}_k) = w_k$ para $\underline{\beta}_k$.

En los modelos lineales ordinarios, el valor de w_k es irrelevante pero el autor afirma que en este caso y debido a que la dimensión de $\underline{\beta}_k$ se modifica para diferentes valores de k , la elección arbitraria de las constantes w_k ; $k=1,2,\dots,Q$ puede resultar en una modificación de los resultados finales. Por esta razón propone el empleo del valor $w_k = (2\pi)^{\frac{r(k)}{2}}$ que en su opinión es más justificable en el sentido de que en esas condiciones la varianza generalizada de cualquier $\underline{\beta}_k$ es la misma en cualquier etapa del procedimiento para alcanzar el límite que produce la distribución inicial de referencia. Bajo estas condiciones, Halpern afirma que si la distribución inicial de σ^2 es impropia para la $\underline{\beta}_k$ es propia, la correspondiente final es propia y los

parámetros se calculan con las mismas expresiones que se presentaron, con la salvedad de que γ_k se hace tender a cero. Si la distribución a priori de β_k es impropia con la constante w_k y σ^2 tiene una distribución inicial propia, entonces la distribución final es propia siempre que exista al menos un valor observado de la variable independiente en cada intervalo (z_1, z_{i+1}) y en al menos uno de ellos se observe una cantidad mayor. En ese caso $\gamma_k^* = \gamma_k + T - r(k) - 2$ y los demás parámetros se obtienen haciendo $\Sigma_k = 0$. La probabilidad a posteriori de H_k se obtiene de la expresión

$$P_k^* = P_k (2\pi)^{\frac{r(k)}{2}} N_k \frac{(\frac{1}{2} \gamma_k \eta_k)^{\frac{1}{2} \gamma_k} \Gamma(\frac{1}{2} \gamma_k^*)}{(\frac{1}{2} \gamma_k^* \eta_k^*)^{\frac{1}{2} \gamma_k^*} \Gamma(\frac{1}{2} \gamma_k)} |\Sigma_k^*|^{-1/2}$$

Finalmente, si tanto σ^2 como β_k tienen una distribución inicial impropia, la final es propia si se satisfacen las condiciones descritas sobre las observaciones de la variable independiente y además $T > q + 2$. Los parámetros se obtienen de los correspondientes al caso anterior haciendo tender γ_k a cero.

El autor menciona que los cálculos necesarios para obtener las

distribuciones finales pueden ser de un volúmen inaccesible, especialmente si q es relativamente grande, ya que es necesario obtener las distribuciones para cada una de las 2^{q-1} hipótesis. Por tal razón Halpern introduce la noción de opiniones consistentes.

Sean j y k dos valores en $\{1, 2, \dots, Q\}$ tales que $S_j \subseteq S_k$; se dice que las opiniones sobre β_j y β_k son consistentes si y solo si la distribución a priori de β_k dadas H_k y σ^2 condicional a que $\beta_t = 0$ para toda $t \in S_k \cap S_j^c$ coincide con la distribución a priori de β_j dada H_j y σ^2 . El autor indica que cuando se emplean familias conjugadas en el problema que se discute, un conjunto de opiniones iniciales consistente conduce a opiniones finales consistentes siempre que estas sean propias. Naturalmente, este resultado permite reducir drásticamente el esfuerzo de cómputo pues basta calcular la distribución final para el caso en que $r(k) = q - 1$ y a partir de ella derivar las condicionales respectivas. Este tratamiento, sin embargo es aplicable solo en los casos en que resulte razonable el empleo de opiniones a priori consistentes.

Finalmente, el autor aborda el problema de predicción e indica que en virtud de que la variable Y satisface la relación

$$Y = f(X) + \varepsilon$$

$$= \sum_{j=0}^q \beta_j U_j(X) + \varepsilon,$$

resulta razonable que se utilice un predictor de la forma

$$\hat{Y} = P(X)$$

$$= \sum_{j=0}^q \hat{\beta}_j U_j(X)$$

Más aún, propone que la pérdida debida a utilizar $P(X)$ cuando $f(X)$ es el valor esperado de Y dado X , sea evaluada según la siguiente expresión

$$R(P, f) = \int_{-\infty}^{\infty} (P(X) - f(X))^2 dW(X) + C(I(P), I(f)),$$

en donde $W(X)$ es una medida finita no negativa del intervalo $[a, b]$ tal que la integral es finita para todo $\underline{\beta}^1 = (\beta_0, \beta_1, \dots, \beta_q)$ y todo $\underline{\hat{\beta}}^1 = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_q)$. Adicionalmente, si $I(P)$ e $I(f)$ son tales que $P(X)$ es consistente con $H_{I(P)}$ y $f(x)$ con $H_{I(f)}$, entonces $C(I(P), I(f))$ representa la pérdida debida a utilizar

un predictor consistente con $H_{I(P)}$ dado que $H_{I(f)}$ es cierta. El procedimiento para determinar el predictor óptimo consiste en calcular el valor esperado de $R(P, f)$ respecto a la distribución que describe el conocimiento, tanto sobre $\underline{\beta}_k$ y σ^2 como respecto a H_k ; $k=1, 2, \dots, Q$. El autor afirma que si $C(I(P), I(f))$ es igual a cero, entonces el predictor óptimo es el que utiliza como coeficientes, los valores esperados marginales de β_j ; $j=1, 2, \dots, q$. En general el valor esperado consta de tres términos aditivos, uno constante respecto al predictor, otro que depende de la diferencia entre el vector de parámetros y su valor esperado marginal y uno más que depende de la diferencia entre el número de coeficientes cero en $P(X)$ y $F(X)$. Así, dependiendo del valor relativo que tengan los dos últimos términos, el predictor óptimo coincide con la esperanza marginal del vector de parámetros y satisface que $r(I(P)) = q-1$ cuando $C(I(P), I(f))$ vale cero o bien se acerca al valor medio del vector de parámetros respecto a la distribución condicional dada H_k donde esta hipótesis es la más probable.

El problema de modelos de regresión con puntos de cambio ha

sido tratado, desde el punto de vista Bayesiano, mas recientemente por Ferreira [35], quien con base en una estructura sencilla, de regresión lineal simple en que uno y solo un cambio ha tenido lugar se remite a determinar las distribuciones finales de los parámetros involucrados para efectos de estimación y construcción de regiones de probabilidad. El autor inicia su contribución indicando, como ya se ha exhibido en la sección anterior de este capítulo, que para este tipo de problemas, la construcción de estadísticas de prueba e intervalos de confianza mediante procedimientos no Bayesianos resulta muy complicada y en general, los resultados que se obtienen son de naturaleza asintótica. Partiendo de este razonamiento, Ferreira indica que sin ignorar las ventajas conceptuales que el enfoque Bayesiano ofrece, el análisis de las dificultades computacionales que involucra resulta muy importante. El autor considera el siguiente modelo. Sean $(Y_1, X_1), (Y_2, X_2) \dots (Y_T, X_T)$ parejas tales que $\underline{X} = (X_1, \dots, X_T)$ son valores fijos, mientras que Y_1, \dots, Y_T son variables aleatorias independientes con distribución Normal y tales que:

$$E(Y_i | \underline{X}) = a_1(X_i - \bar{X}_1) + b_1 \quad ; \quad i=1, 2, \dots, \tau$$

$$E(Y_i | \underline{X}) = a_2(X_i - \bar{X}_2) + b_2 \quad ; \quad i=\tau+1, \dots, T$$

$$\text{Var}(Y_i | \underline{X}) = \sigma^2 \quad ; \quad i=1, 2, \dots, T$$

en donde

$$\bar{X}_1 = \frac{1}{\tau} \sum_{i=1}^{\tau} X_i \quad , \quad \bar{X}_2 = \frac{1}{T-\tau} \sum_{i=\tau+1}^T X_i$$

Ferreira menciona el resultado, bien conocido en el enfoque Bayesiano, según el cual bajo condiciones relativamente generales, \underline{X} puede ser considerado como el valor observado de un vector aleatorio y aún así ser tratado como un vector de valores fijos. En cualquier forma a lo largo de su trabajo considera, como ya se indicó, que los valores de la variable independiente son fijos.

El primer objetivo que se persigue en la determinación de la distribución a posteriori conjunta de $\underline{\theta} = (a_1, a_2, b_1, b_2, \sigma^2, \tau)$ y a partir de ella las marginales correspondientes. De la estructura del modelo se tiene que:

$$\begin{aligned}
 f(\underline{Y}_T | \underline{X}, \underline{\theta}) &= \sigma^{-T} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^T (Y_i - a_1 (X_i - \bar{X}_1) - b_1)^2 \right\} \\
 &\cdot \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=\tau+1}^T (Y_i - a_2 (X_i - \bar{X}_2) - b_2)^2 \right\} \\
 &= \sigma^{-T} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^T (Y_i - a_1 (X_i - \bar{X}_1) - b_1)^2 + \right. \right. \\
 &\quad \left. \left. \sum_{i=\tau+1}^T (Y_i - a_2 (X_i - \bar{X}_2) - b_2)^2 \right] \right\} .
 \end{aligned}$$

De acuerdo con la regla de Bayes la distribución a posteriori de $\underline{\theta}$ se obtiene de la expresión:

$$f(\underline{\theta} | \underline{Y}_T, \underline{X}) = f(\underline{Y}_T | \underline{\theta}, \underline{X}) f_0(\underline{\theta} | \underline{X})$$

en donde $f_0(\underline{\theta} | \underline{X})$ representa la distribución a priori para $\underline{\theta}$ tomando en cuenta que las observaciones se efectuaron en los valores \underline{X} . El autor considera, para los fines de su trabajo, una distribución inicial de referencia para $\underline{\theta}$. Más específicamente Ferrelra asigna una distribución conjuntal tal que todos los parámetros son independientes a priori y las distribuciones de a_1, a_2, b_1, b_2 , y $\ln \sigma^2$ son proporcionales a una constante. Respecto a τ considera una función de probabilidad genérica $\pi(t)$ sobre la cual, en principio, solo requiere que satisfaga que:

$$\pi(1) = \pi(T-1) = \pi(T) = 0$$

esto es, que exista un cambio y que haya al menos dos observaciones en cada régimen. La primera condición corresponde a la situación en que se desea estimar la localización del cambio, mientras que la segunda resulta útil en el desarrollo de Ferreira para encontrar expresiones sencillas para las correspondientes distribuciones a posteriori. Con esta distribución inicial, se tiene entonces, que

$$f(\theta | \underline{Y}_T, \underline{X}) = f(\underline{Y}_T | \theta, \underline{X}) \pi(t) / a^2$$

Para proceder a determinar las expresiones específicas de las distribuciones a posteriori, el autor establece que si se definen, para $t=2,3,\dots,T-2$ las cantidades

$$\begin{aligned} \bar{Y}_1 &= \frac{1}{t} \sum_{i=1}^t Y_i & \bar{X}_1 &= \frac{1}{t} \sum_{i=1}^t X_i \\ S_{yy_1} &= \sum_{i=1}^t (Y_i - \bar{Y}_1)^2 & S_{xx_1} &= \sum_{i=1}^t (X_i - \bar{X}_1)^2 \\ S_{yx_1} &= \sum_{i=1}^t (Y_i - \bar{Y}_1)(X_i - \bar{X}_1) & S_f^2 &= S_{yy_1} - S_{xy_1}^2 / S_{xx_1} \end{aligned}$$

entonces,

$$\begin{aligned} \sum_{i=1}^t (Y_i - a_1(X_i - \bar{X}_1) - b_1)^2 &= S_{yy1} + t(\bar{Y}_1 - b_1)^2 + a_1^2 S_{xx1} - 2 a_1 S_{yx1} \\ &= S_1^2 + t(\bar{Y}_1 - b_1)^2 + S_{xx1} (a_1 - S_{yx1} / S_{xx1})^2 \end{aligned}$$

Es interesante notar que la última expresión no tiene sentido si $S_{xx} = 0$ lo cual ocurre particularmente si $t=1$, uno de los casos que Ferreira excluye. Es fácil verificar que definiendo de manera análoga $\bar{Y}_2, \bar{X}_2, S_{xx2}, S_{yy2}, S_{yx2}$ y S_2^2 se tiene que para $t=2, 3, \dots, T-2$.

$$\sum_{i=t+1}^T (Y_i - a_2(X_i - \bar{X}_2) - b_2)^2 = S_2^2 + (T-t)(\bar{Y}_2 - b_2)^2 + S_{xx2} (a_2 - S_{yx2} / S_{xx2})^2$$

De esta forma se tiene que

$$\begin{aligned} f(\underline{Q} | \underline{Y}_T, \underline{X}) &\propto \sigma^{-(T+2)} \pi(t) \exp \left\{ -\frac{1}{2\sigma^2} [t(b_1 - \bar{Y}_1)^2 \right. \\ &\quad \left. + S_{xx1} (a_1 - S_{yx1} / S_{xx1})^2 + (T-t)(b_2 - \bar{Y}_2)^2 \right. \\ &\quad \left. + S_{xx2} (a_2 - S_{yx2} / S_{xx2})^2 + S^2] \right\} \end{aligned}$$

en donde $S^2 = S_1^2 + S_2^2$.

A partir de esta expresión, es fácil determinar las distribuciones a posteriori marginal para cada uno de los parámetros. Basta con integrar o sumar, según sea el caso, sobre los restantes para obtener la expresión correspondiente. Siguiendo la notación propuesta el autor exhibe que a posteriori se tiene que:

$$\pi(t|Y_T, X) \propto (t(T-t)S_{xx1}S_{xx2})^{-\frac{1}{2}} (S^2)^{-\frac{T-4}{2}} \pi(t); t=2,3,\dots,T-2$$

$$f(a_1|Y_T, X) \propto \sum_{t=2}^{T-2} \left(1 + \frac{(a_1 - S_{yx1}/S_{xx1})^2}{S^2/S_{xx1}}\right)^{-\frac{(T-3)}{2}} \frac{\pi(t|Y_t, X)}{(S^2/S_{xx1})^{\frac{1}{2}}}; a_1 \in \mathbb{R}$$

$$f(b_1|Y_T, X) \propto \sum_{t=2}^{T-2} \left(1 + \frac{(b_1 - Y_1)^2}{S^2/t}\right)^{-\frac{(t-3)}{2}} \frac{\pi(t|Y_t, X)}{(S^2/t)^{\frac{1}{2}}}; b_1 \in \mathbb{R}$$

$$f(\sigma^2|Y_T, X) \propto \sum_{t=2}^{T-2} \left(\frac{S^2}{\sigma^2}\right)^{\frac{T-4}{2}} \sigma^{-2} \exp\left\{-\frac{S^2}{2\sigma^2}\right\} \pi(t|Y_t, X); \sigma^2 > 0$$

Naturalmente, las expresiones correspondientes para a_2 y b_2 son análogas a las de a_1 y b_1 . La característica más sobresaliente de estos resultados es que las distribuciones a posteriori resultan mezclas de distribuciones comunes en Estadística. Así, $f(a_1|Y_T, X)$ es una mezcla de distribuciones de Student, de la

misma manera que $f(b_1 | \underline{Y}_T, \underline{X})$ y consecuentemente, $f(a_2 | \underline{Y}_T, \underline{X})$ y $f(b_2 | \underline{Y}_T, \underline{X})$. Por su parte, es fácil verificar que $f(\sigma^2 | \underline{Y}_T, \underline{X})$ es una mezcla de Gammas invertidas, o más específicamente de distribuciones χ^2 invertidas. Como una consecuencia inmediata se tiene que los correspondientes valores esperados a posteriori, que constituyen los estimadores de Bayes si se emplea una función de pérdida cuadrática, resultan relativamente simples y tienen las siguientes expresiones

$$E(a_1 | \underline{Y}_T, \underline{X}) = \sum_{t=2}^{T-2} (S_{yx1} / S_{xx1}) \pi(t | \underline{Y}_T, \underline{X})$$

$$E(a_2 | \underline{Y}_T, \underline{X}) = \sum_{t=2}^{T-2} (S_{yx2} / S_{xx2}) \pi(t | \underline{Y}_T, \underline{X})$$

$$E(b_1 | \underline{Y}_T, \underline{X}) = \sum_{t=2}^{T-2} \bar{Y}_1 \pi(t | \underline{Y}_T, \underline{X})$$

$$E(b_2 | \underline{Y}_T, \underline{X}) = \sum_{t=2}^{T-2} \bar{Y}_2 \pi(t | \underline{Y}_T, \underline{X})$$

$$E(\sigma^2 | \underline{Y}_T, \underline{X}) = \sum_{t=2}^{T-2} (S^2 / (T-6)) \pi(t | \underline{Y}_T, \underline{X})$$

En el caso del parámetro τ , es claro que en virtud de solo puede tomar valores enteros, el estimador Bayesiano correspondiente a una pérdida cuadrática está dado por el valor

$\bar{r} \in \{2, 3, \dots, T-2\}$ mas cercano a la media.

Adicionalmente, se tiene que, a posteriori, para cualquier intervalo (c, d)

$$P(c < a_1 < d | \underline{Y}_T, \underline{X}) = \sum_{t=2}^{T-2} P(c < a_1 < d | \underline{Y}_T, \underline{X}, t) \pi(t | \underline{Y}_T, \underline{X})$$

de modo que para determinar esta probabilidad basta con establecer las probabilidades condicionales a t para $t=2, \dots, T-2$ y calcular el promedio ponderado correspondiente. El aspecto más relevante de esta relación es que $f(a_1 | \underline{Y}_T, \underline{X}, t)$ es una distribución de Student que se puede estandarizar para realizar el cálculo de probabilidades mediante tablas. Claramente, la misma situación se tiene para b_1, a_2 y b_2 . Más aún, un argumento similar puede ser empleado para combinaciones lineales de estos parámetros. En el caso de σ^2 es fácil verificar que si $c > 0$,

$$P(c < \sigma^2 < d | \underline{Y}_T, \underline{X}) = \sum_{t=2}^{T-2} P(S^2/d < Z < S^2/c) \pi(t | \underline{Y}_T, \underline{X})$$

en donde Z es una variable χ^2 con $T-4$ grados de libertad.

Es interesante notar que si bien tanto los valores esperados como los intervalos de probabilidad se obtienen a partir de expresiones sencillas los cálculos para un problema específico,

seguramente requerirán de un auxilio de cómputo, especialmente si T es relativamente grande. Por otra parte, es importante establecer que el procedimiento que propone Ferreira para la determinación de regiones de probabilidad no garantiza que las regiones obtenidas sean de máxima densidad, particularmente debido a que en el caso de mezclas, tales regiones suelen ser discontinuas. Respecto a los valores esperados, como ya se indicó, resultan ser los estimadores Bayesianos si se utiliza una función de pérdida cuadrática pero más aún, en este caso y como consecuencia de la distribución a priori que emplea el autor se pueden interpretar como promedios ponderados de los estimadores de máxima verosimilitud que se obtendrían para cada valor de t . La única salvedad es que para σ^2 aparece la cantidad $S^2/(T-6)$ en vez de la usual $S^2/(T-4)$.

De cualquier forma, el autor desarrolla un estudio de simulación para comparar los estimadores de máxima verosimilitud con los bayesianos correspondientes a tres diferentes distribuciones iniciales para r . Estas son las siguientes:

$$\pi_1(t) \propto 1 \quad ; \quad t=2, 3, \dots, T-2$$

$$\pi_2(t) \propto \{t(T-t)\}^{\frac{1}{2}} \quad ; \quad t=2, 3, \dots, T-2$$

$$\pi_3(t) \propto \{t(T-t)S_{xx1}S_{xx2}\}^{\frac{1}{2}} \quad ; \quad t=2, 3, \dots, T-2$$

La primera corresponde a la situación en que todos los posibles valores de τ son igualmente probables mientras que las dos restantes, como menciona Ferreira, están sugeridas por la forma de $\pi(t|Y_t, X)$. Cuando se utiliza $\pi_3(t)$ la distribución a posteriori resulta de la forma

$$\pi(t|Y_T, X) \propto \{S^2/(T-4)\}^{-\frac{(T-4)}{2}}$$

en donde, como ya se indicó, $S^2/(T-4)$ es el estimador usual de máxima verosimilitud de la varianza si $\tau=t$. De esta forma, la moda a posteriori coincide con el valor que minimiza la estimación de la varianza. La distribución $\pi_2(t)$ representa una situación intermedia entre $\pi_1(t)$ y $\pi_3(t)$ y entre otras características posee la propiedad de asignar las probabilidades mas altas a los valores de τ cercanos a $T/2$.

Para efectos de la comparación el autor considera el modelo empleado en un trabajo ya discutido, por Quandt [72],

$$Y_i = 2.5 - 0.7 X_i + \epsilon_i$$

en donde $T=20$ $X_i = 1; i=1, 2, \dots, 20$ e incluye tres variantes, $\tau=5, 10, 15$. Para cada punto de cambio, se introducen cambios en la pendiente de diferente magnitud con valores similares a los empleados por Farley y Jinich [31] anteriormente.

En cada caso se calculan el error cuadrático medio y el sesgo medio para el estimador de máxima verosimilitud y los tres bayesianos correspondientes a $\pi_1(t)$, $\pi_2(t)$ y $\pi_3(t)$.

Los resultados que se obtienen indican que el error cuadrático medio del estimador clásico es uniformemente mayor al de los tres estimadores Bayesianos pero en términos del sesgo medio, los cuatro producen valores comparables.

Ferreira concluye su estudio aplicando el tratamiento propuesto, utilizando las tres distribuciones apriori para τ , a las observaciones generadas por Quandt [72] y obtiene que, en los tres casos, el estimador Bayesiano es $\hat{\tau}=11$ que difiere de $\tau=12$, el verdadero valor del punto de cambio. Los estimadores para la varianza y los coeficientes de regresión también resul-

tan muy similares en los tres casos.

Finalmente es muy interesante tomar en cuenta algunos aspectos del trabajo de Ferreira. Primero, en el modelo que considere en necesario que $a_1 \neq a_2$ para garantizar que ocurrió un cambio. Si $a_1 = a_2 = a$, la condición

$$b_1 - a\bar{X}_1 = b_2 - a\bar{X}_2 = b$$

es suficiente para que la relación entre X y Y sea

$$Y_i = b + aX_i + \epsilon_i ; i=1,2,\dots,T .$$

Más aún, prácticamente todo modelo de regresión lineal simple puede reescribirse como un modelo con cambio como el descrito por Ferreira. Bajo el supuesto de que

$$Y_i = b + aX_i + \epsilon_i ; i=1,2,\dots,T$$

se tiene que

$$\begin{aligned} Y_i &= b + aX_i - a\bar{X}_1 + a\bar{X}_1 + \epsilon_i \\ &= b + a\bar{X}_1 + a(X_i - \bar{X}_1) + \epsilon_i \\ &= b_1 + a(X_i - \bar{X}) + \epsilon_i ; i=1,2,\dots,T \end{aligned}$$

De la misma manera se puede verificar que

$$Y_i = b_2 + a(X_i - \bar{X}) + \epsilon_i ; i=1,2,\dots,T$$

de tal modo que en particular,

$$Y_i = b_1 + a(X_i - \bar{X}_1) + \varepsilon_i ; i=1,2,\dots,r$$

$$Y_i = b_2 + a(X_i - \bar{X}_2) + \varepsilon_i ; i=r+1,\dots,T$$

para cualquier r . Por esta razón, es recomendable emplear la estructura propuesta por Ferreira con cautela para evitar una formulación errónea del problema. Por otra parte, en un trabajo publicado mas recientemente, Holbert y Broemeling, [49], aplican el mismo tratamiento al modelo planteado en su forma natural, esto es

$$Y_i = b_1 + a_1 X_i + \varepsilon_i ; i=1,2,\dots,r$$

$$Y_i = b_2 + a_2 X_i + \varepsilon_i ; i=r+1,\dots,T.$$

Estos autores solo consideran $\pi(t) \approx 1$ y únicamente derivan la distribución a posteriori para $\tau, \pi(t | \underline{Y}_T, \underline{X})$. El resultado, obviamente es el mismo. Sin embargo, un detalle valioso es que también ilustran la técnica con los datos de Quandt [72], pero reportan la totalidad de los valores de $\pi(t | \underline{Y}_T, \underline{X})$. Con esta información, se puede confirmar que efectivamente $E(\tau | \underline{Y}_T, \underline{X}) = 10.71$ de modo que el estimador de Bayes asociado a

una función pérdida cuadrática es $\tilde{r}=11$ pero como Holbert y Broemeling indican, también es cierto que la mediana y la moda de la distribución son $\tilde{r} = r^* = 12$ que coinciden con el verdadero valor de r . Así, si en vez de utilizar una pérdida cuadrática se hubiese utilizado

$$L_1(\tilde{r}, \tau) = |\tilde{r} - \tau| \quad \text{ó} \quad L_2(\tilde{r}, \tau) = \begin{cases} 1 & \text{si } |\tau - \tilde{r}| > k, k \text{ cercano a } 0 \\ 0 & \text{otro caso} \end{cases}$$

el estimador Bayesiano hubiese coincidido con el de máxima verosimilitud y más aún, en las simulaciones de Ferreira probablemente el sesgo medio sería menor.

De esta manera, en el trabajo de Ferreira se presenta una formulación del problema de un punto de cambio en modelos de regresión lineal simple que tratada desde el punto de vista Bayesiano produce resultados que parecen satisfactorios y tienen un enorme atractivo desde el punto de vista computacional. Algunas precauciones serán necesarias, sin embargo, para emplear esta técnica sin mal especificar el modelo.

Otra con tribución donde se aborda el problema de puntos de

cambio en modelos de regresión es la de Smith [4], en donde se formula matricialmente el problema como sigue. Sean Y_1, Y_2, \dots, Y_T observaciones independientes tales que

$$Y_i = X_i' \underline{\beta}^{(i)} + \epsilon_i \quad ; \quad i=1, 2, \dots, T$$

en donde X_i es el vector de dimensión p , de valores de las variables independientes en el tiempo i , $\underline{\beta}^{(i)}$ es correspondiente vector de coeficientes de regresión y ϵ_i , el término de error de distribución Normal con media cero y varianza σ^2 . El autor considera que las observaciones son descritas por el modelo M_r ($1 \leq r < T$) si y solo si

$$\underline{\beta}^{(1)} = \underline{\beta}^{(2)} = \dots = \underline{\beta}^{(r)} = \underline{\beta} \quad \underline{\beta}^{(r+1)} = \underline{\beta}^{(r+2)} = \dots = \underline{\beta}^{(T)} = \underline{\beta} + \delta$$

con $\delta \neq 0$, desconocido. Por otra parte, si $\delta = 0$ el modelo respectivo, que no presenta cambios, se denota con M_0 . En estos términos, Smith se propone la discriminación entre los modelos M_0, M_1, \dots, M_{T-1} . Naturalmente, este planteamiento corresponde a la situación en que se sabe que ocurre a lo más un cambio en la sucesión y se abordan simultáneamente los problemas de decidir si tuvo efecto y estimar el punto de cambio asociado. Smith indi

ca que bajo el modelo M_t , se tiene que si

$$\underline{Y}'_T = (Y_1, Y_2, \dots, Y_T)$$

$$\underline{X}'_t = (X_1, \dots, X_t) ; \underline{X}'_{T-t} = (X_{t+1}, \dots, X_T)$$

entonces,

$$\underline{Y}_T \sim N(\underline{A}_t \underline{\theta}, \sigma^2 \underline{I}_T)$$

en donde

$$\underline{A}_t = \begin{bmatrix} X_t & 0 \\ & \\ & \\ X_{T-t} & X_{T-t} \end{bmatrix}, \underline{\theta} = \begin{bmatrix} \underline{\beta} \\ \underline{\delta} \end{bmatrix}$$

Por otra parte si M_0 es el modelo que describe el fenómeno

$$\underline{Y}_T \sim N(\underline{A}_0 \underline{\theta}_0, \sigma^2 \underline{I}_T)$$

con

$$\underline{A}_0 = \underline{X}_T, \quad \underline{\theta}_0 = \underline{\beta}.$$

Con el propósito de realizar inferencias sobre el cambio potencial, esto es, para decidir si se efectuó y en donde, basta con determinar las probabilidades a posteriori de M_0, M_1, \dots, M_{T-1} .

A faltade una función de pérdidas mas específica, la elección del modelo se puede efectuar a través de los factores de Bayes de la misma forma en que el autor propone para el caso de sucesiones de variables aleatorias y que ya ha sido discutida en el capítulo precedente. En estas circunstancias, la verosimilitud de las observaciones bajo M_t puede obtenerse de acuerdo a la siguiente expresión:

$$f(Y_T | M_t) = \int \dots \int f(Y_t | \Lambda_t, \underline{\theta}_t, \sigma) f(\underline{\theta}_t, \sigma | \Lambda_t) d\underline{\theta}_t d\sigma$$

en donde es indispensable asignar la distribución conjunta a priori $f(\underline{\theta}_t, \sigma | \Lambda_t)$ para $t=0, 1, 2, \dots, T$. El autor considera el caso en que

$$f(\underline{\theta}_t, \sigma^2 | \Lambda_t) = f(\underline{\theta}_t | \Lambda_t, \sigma) f(\sigma).$$

y más aún, examina las dos posibilidades de σ , conocida y desconocida. En el primer caso, asignando para $\underline{\theta}_t$ una distribución inicial conjugada, esto es, una distribución Normal de media $\underline{\theta}^*$ y matriz de varianzas y covarianzas $\sigma^2 V^*$, en donde

$$\underline{\theta}^* = \begin{bmatrix} \underline{\theta}^* \\ \underline{\delta}^* \end{bmatrix}, \quad V^* = \begin{bmatrix} V_{\underline{\theta}}^* & 0 \\ 0 & V_{\underline{\delta}}^* \end{bmatrix}$$

se tiene que la expresión para $f(Y_T | M_t, \sigma)$ esta dada por

$$f(Y_T | M_t, \sigma) = (2\pi\sigma^2)^{-T} |V^*|^{-\frac{1}{2}} |V^{*-1} + A_t' A_t|^{-\frac{1}{2}} \\ \cdot \exp \left\{ -\frac{1}{2\sigma^2} \left[R_t + (\hat{\theta}_t - \theta^*)' \{V^* + (A_t' A_t)^{-1}\}^{-1} (\hat{\theta}_t - \theta^*) \right] \right\}$$

con $\hat{\theta}_t$ el estimador usual $\hat{\theta}_t$ por mínimos cuadrados y R_t la correspondiente suma de cuadrados de los residuales. Adicionalmente si V^{*-1} puede considerarse casi nula con respecto a $A_t' A_t$ se tiene que

$$f(Y_T | M_t, \sigma) \approx (2\pi\sigma^2)^{-T/2} |V_t^*|^{-T/2} |A_t' A_t|^{-\frac{1}{2}} \exp \{-R_t / 2\sigma^2\}$$

para $t=1, 2, \dots, T-1$ mientras que

$$f(Y_T | M_0, \sigma) \approx (2\pi\sigma^2)^{-\frac{T}{2}} |V_0^*|^{-\frac{1}{2}} |A_0' A_0|^{-\frac{1}{2}} \exp \{-R_0 / 2\sigma^2\}$$

De esta manera y como $|A_t' A_t| = |X_t' X_t| |X_{T-t}' X_{T-t}|$, se tiene que el factor de Bayes para comparar M_0 contra M_t está dado por

$$B_{0,t}(\sigma) = \left\{ \frac{|V_0^*| |X_t' X_t| |X_{T-t}' X_{T-t}|}{|X_T' X_T|} \right\}^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (R_0 - R_t) \right\}; t=1, 2, \dots, T-1$$

Por otra parte si σ es desconocida y se asigna a priori la densidad

$$f(\sigma) = \sigma^{-1}$$

se tiene que el correspondiente factor de Bayes se calcula como

$$B_{0,t} = \frac{f(\underline{Y}_T | M_0)}{f(\underline{Y}_T | M_t)}$$

en donde tanto $f(\underline{Y}_T | M_0)$ como $f(\underline{Y}_T | M_t)$ se obtienen integrando $f(\underline{Y}_T | M_0) f(\sigma)$ y $f(\underline{Y}_T | M_t) f(\sigma)$ respecto a σ . En esas condiciones el factor tiene la siguiente expresión:

$$B_{0,t} = \left\{ \frac{|V_0^*| |X_t' X_t| |X_{T-t}' X_{T-t}|}{|X_T' X_T|} \right\}^{1/2} \left(1 + \frac{P}{(T-2p)} F_t \right)^{-\frac{T}{2}}$$

en donde $F_t = [(R_0 - R_t)/p] / [R_t/(T-p)]$, es la estadística F usual para probar, en el enfoque frecuentista, los modelos M_0 contra M_t .

Al igual que en el caso de sucesiones, el autor indica que los factores de Bayes $B_{0,t}; t=1, 2, \dots, T-1$ pueden ser utilizados para tomar una decisión respecto a si el cambio tuvo lugar mientras que los demás factores $B_t' = B_{0,t} / B_{0,t'}$, aportan información para determinar en donde se efectuó el cambio. La característica más sobresaliente de este tratamiento, es que no se requiere de la asignación de probabilidades iniciales para los mo

delos M_0, M_1, \dots, M_{T-1} . En las aplicaciones que discute Smith, sin embargo, se recurre directamente a las probabilidades a posteriori para llevar a cabo las inferencias correspondientes.

Como un comentario adicional, el autor menciona que en ocasiones, cuando se tiene un modelo de regresión lineal simple, puede resultar de interés efectuar inferencias sobre $\gamma = -\delta_1/\delta_2$, el punto de intersección de las dos rectas de regresión. En esas condiciones, como ya se ha discutido en otros trabajos, la estructura puede incorporar la restricción de que γ se encuentre, cuando el modelo M_t es el correcto, en el intervalo $[X_t, X_{t+1})$ si los vectores de X están ordenados $(X_1 < X_2 < \dots < X_T)$. El tratamiento sin la restricción, se reduce a la consideración de la función de densidad a posteriori

$$f(\gamma | \underline{Y}_T) = \sum f(\gamma | M_t, \underline{Y}_T) p(M_t | \underline{Y}_T)$$

en donde $f(\gamma | M_t, \underline{Y}_T)$ se obtiene mediante la transformación adecuada de $f(\underline{\theta} | M_t, \underline{Y}_T)$. Smith indica que en el caso en que se imponga la restricción en γ , si se denota por C la restricción, se tiene que

$$f(\gamma | C, \underline{Y}_T) = \sum_{t=1}^T P(C | \gamma, t, \underline{Y}_T) f(\gamma, t | \underline{Y}_T) / P(C | \underline{Y}_T)$$

en donde

$$P(C | \gamma, t; \underline{Y}_T) = \begin{cases} 1 & \text{si } \gamma \in [X_t, X_{t+1}) \\ 0 & \text{otro caso} \end{cases}$$

y además,

$$\begin{aligned} P(C | \underline{Y}_T) &= \sum_{t=1}^T \left\{ \int P(C | \gamma, t, \underline{Y}_T) \right\} P(t | \underline{Y}_T) \\ &= \sum_{t=1}^T \int_{X_t}^{X_{t+1}} f(\gamma, t | \underline{Y}_T) d\gamma \end{aligned}$$

Para tomar decisiones respecto al modelo apropiado, la distribución a posteriori, considerando la restricción, se obtiene de la expresión

$$P(t | C, \underline{Y}_T) \propto \int_{X_t}^{X_{t+1}} f(\gamma, t | \underline{Y}_T) d\gamma .$$

Finalmente, el autor aplica este tratamiento a una serie de datos reales y concluye que el análisis es satisfactorio y muy ilustrativo aún a pesar de utilizar una distribución inicial con varianzas grandes para θ y una uniforme para $M_t; t=1, \dots, T-1$.

En este trabajo resulta interesante notar la simplicidad del procedimiento tanto para decidir si el cambio tuvo lugar como para determinar su localización y la relativa facilidad con que se puede involucrar la restricción considerada. Es prudente, sin embargo, tomar en cuenta que se trata de modelos con un solo punto de cambio y que no se discuten los inconvenientes asociados al cálculo numérico de las probabilidades a posteriori sobre todo cuando T es relativamente grande.

En un trabajo más reciente, [92], Smith abunda sobre este problema trabajando con más detalle el caso de regresión lineal simple tanto con restricción como sin ella. La representación que en este caso emplea es la tradicional:

$$Y_i = a_1 + b_1 X_i + \epsilon_i \quad ; \quad i = 1, 2, \dots, r$$

$$Y_i = a_2 + b_2 X_i + \epsilon_i \quad ; \quad i = r+1, r+2, \dots, T$$

en donde los errores son Normales $(0, \sigma^2)$ independientes y los parámetros a_1, b_1, a_2, b_2 , y τ son desconocidos. Por supuesto se supone que $(a_1, b_1) \neq (a_2, b_2)$. Respecto a τ se tiene que $1 < \tau < T$ de modo que se excluye la posibilidad de que no ocurra el cam-

bio o de que solo exista una observación en el primer segmento y entonces, el interés se centra, en principio, en realizar inferencias sobre $\theta^1 = (a_1, a_2, b_1, b_2)$, σ^2 y τ . Adicionalmente, como lo hacen algunos otros autores, se supone que $X_1 < X_2 < \dots < X_T$ lo cual resulta particularmente apropiado en las aplicaciones en que X se relaciona con el tiempo. En estas condiciones nuevamente Smith considera la posibilidad de que el punto de intersección $\gamma = (a_1 - a_2) / (b_2 - b_1)$ satisfaga la restricción $X_T < \gamma < X_{T+1}$. Este parámetro γ , está bien definido siempre que $b_1 \neq b_2$ y como ya ha sido discutido, esta condición implica la continuidad en el modelo con un cambio. Más aún, el valor de γ se puede interpretar como el valor de X , probablemente no observado, en el cual se efectúa la transición de un segmento a otro. El autor menciona que modelos de este tipo, con la restricción en γ , proporcionan descripciones satisfactorias en el estudio de trasplantes renales, Knapp et al. [51], en donde se describe el nivel de funcionamiento del riñón en el tiempo a partir del momento del trasplante. En ese tipo de aplicación se ha observado que una tendencia creciente indica un funcionamiento correc-

to pero un cambio decreciente súbito puede asociarse con el inicio de un proceso de rechazo. El cambio en esas condiciones, aunque súbito, puede ser descrito por un modelo continuo debido a la naturaleza del fenómeno. Esto es, se sabe que el proceso de rechazo es continuo. Smith incluso ilustra los resultados que obtiene con datos de este tipo de aplicaciones.

Por facilidad, como en el artículo anterior, el autor desarrolla el proceso de inferencia sin involucrar la restricción que es impuesta posteriormente. Considerando como una primera aproximación a σ^2 conocida, es claro que

$$f(\underline{\theta} | \underline{Y}_T, \underline{X}_T, \tau, \sigma^2) \propto f(\underline{Y}_T | \underline{X}_T, \underline{\theta}, \tau, \sigma^2) f(\underline{\theta} | \tau, \sigma^2)$$

en donde,

$$\begin{aligned} f(\underline{Y}_T | \underline{X}_T, \underline{\theta}, \tau, \sigma^2) &= (2\pi\sigma^2)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{\tau} (Y_i - a_1 - b_1 X_i)^2 + \sum_{i=\tau+1}^T (Y_i - a_2 - b_2 X_i)^2 \right]\right\} \\ &= (2\pi\sigma^2)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma^2} (\underline{Y}_T - X(\tau)\underline{\theta})' (\underline{Y}_T - X(\tau)\underline{\theta})\right\} \end{aligned}$$

con $\underline{\theta}' = (a_1, a_2, b_1, b_2)$ y

$$X(\tau) = \begin{bmatrix} 1 & 0 & X_1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & X_r & 0 \\ 0 & 1 & 0 & X_{r+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & X_T \end{bmatrix}$$

mientras que $f(\underline{\theta}|\tau, \sigma^2)$ representa el conocimiento a priori sobre $\underline{\theta}$.

Si a priori se supone que θ es independiente de τ y se le asigna una distribución normal de media $\bar{\theta}$ y matriz de varianzas y covarianzas $\sigma^2 \bar{\Omega}^{-1}$, es fácil verificar que $f(\underline{\theta}, | Y_r, X_r, \tau, \sigma^2)$ corresponde a una Normal de media $\bar{\theta}(\tau)$ y matriz de varianzas y covarianzas $\sigma^2 \bar{\Omega}(\tau)$ en donde,

$$\bar{\theta}(\tau) = (\bar{\Omega} + X(\tau)'X(\tau))^{-1} (\bar{\Omega}\bar{\theta} + X(\tau)'X(\tau)\hat{\theta}(\tau))$$

$$\bar{\Omega}(\tau) = (\bar{\Omega} + X(\tau)'X(\tau))$$

con $\hat{\theta}(\tau)$ el estimador usual de mínimos cuadrados de $\underline{\theta}$ cuando efectivamente hay un cambio en X_r . A partir de una transformación lineal de $\underline{\theta}$ se puede determinar la distribución a posteriori de las diferencias $\mu_A = a_1 - a_2$ y $\mu_B = b_2 - b_1$ que naturalmente, resulta también Normal. Con esta distribución una nueva transformación,

$\gamma = \mu_a / \mu_b$ y $\delta = \mu_b$, conduce a la determinación, marginalizando sobre δ , de $f(\gamma | \underline{Y}_T, \underline{X}_T, \tau, \sigma^2)$.

Para calcular $f(\gamma | \underline{Y}_T, \underline{X}_T, \sigma^2)$ se recurre al hecho de que

$$f(\gamma | \underline{Y}_T, \underline{X}_T, \sigma^2) = \sum_{\tau=0}^{T-1} f(\gamma | \underline{Y}_T, \underline{X}_T, \tau, \sigma^2) f(\tau | \underline{Y}_T, \underline{X}_T, \sigma^2)$$

La distribución a posteriori de τ , $f(\tau | \underline{Y}_T, \underline{X}_T, \sigma^2)$ se puede establecer a partir de la conjunta $f(\underline{\theta}, \tau | \underline{Y}_T, \underline{X}_T, \sigma^2)$ integrando respecto a $\underline{\theta}$. La conjunta, a su vez esta dada por

$$f(\underline{\theta}, \tau | \underline{Y}_T, \underline{X}_T, \sigma^2) = f(\underline{\theta} | \underline{Y}_T, \underline{X}_T, \tau, \sigma^2) f(\tau | \sigma^2)$$

en donde $f(\tau | \sigma^2)$ es la distribución a priori de τ que usualmente es independiente de σ^2 , esto es, $f(\tau | \sigma^2) = f(\tau)$.

Smith verifica que la determinación de las distribuciones finales de $\underline{\theta}$ y τ así como la de γ se simplifica si se utilizan a priori distribuciones iniciales de referencia.

El caso en que σ^2 es desconocida es tratado por Smith asignando a este parámetro una distribución inicial conjugada, esto es, una Gamma invertida. Como resultado, se obtiene que condi-

cional a τ , laa posteriori de σ^2 resulta nuevamente una distribuy
 ción de esa familia. Las demás distribuciones se modifican con
 siderando la distribución de σ^2 e integrando respecto a este pa
 rámetro. siendo desconocido σ^2 , también es posible asignar a es
 te parámetro una distribución inicial de referencia que de nue
 vo simplifica en cierta medida, según Smith, los resultados.

La inclusión de la restricción en γ , que se denota por C, es
 tratada por el autor de la misma forma que en el artículo pre
 vio. Condicionando a la restricción se tiene que:

$$f(\gamma|C, \underline{Y}_T, \underline{X}_T) = \int_{\tau} f(\gamma, \tau|C, \underline{Y}_T, \underline{X}_T) \\
 = \int_{\tau} P(C|\gamma, \tau, \underline{X}_T, \underline{Y}_T) f(\gamma, \tau|\underline{Y}_T, \underline{X}_T) / P(C|\underline{Y}_T, \underline{X}_T)$$

en donde,

$$P(C|\gamma, \tau, \underline{Y}_T, \underline{X}_T) = \begin{cases} 1 & \text{si } X_r \leq \gamma < X_{r+1} \\ 0 & \text{en otro caso.} \end{cases}$$

y además

$$P(C|\underline{Y}_T, \underline{X}_T) = \int_{\tau} \left\{ \int P(C|\gamma, \tau, \underline{Y}_T, \underline{X}_T) f(\gamma|\tau, \underline{Y}_T, \underline{X}_T) d\gamma \right\} f(\tau|\underline{Y}_T, \underline{X}_T) \\
 = \int_{X_r}^{X_{r+1}} f(\gamma, \tau|\underline{Y}_T, \underline{X}_T) d\gamma.$$

Para ilustrar estos resultados, el autor los aplica en un primer caso, a un conjunto de observaciones simuladas a través de un modelo que satisface la restricción. Considerando distribuciones iniciales de referencia obtiene las distribuciones finales de τ y γ tanto incorporando la restricción como sin ella. Lo más sobresaliente de este análisis es que dependiendo de si se incluye la restricción, las inferencias sobre los dos parámetros pueden ser apreciablemente distintas. De hecho en el conjunto de datos analizados, cuando no se incorpora la restricción, se tiene que la probabilidad a posteriori de que γ se encuentre en la región con 0.884 de máxima probabilidad de τ , que en los datos simulados coincide con X_τ ya que $X_t = t$ para $t=1$, es prácticamente cero.

Es interesante mencionar que Smith se refiere al procedimiento frecuentista propuesto por Hinkley [43] para este tipo de problemas y conjetura, en base a la aplicación de las dos técnicas a otro conjunto de datos, que probablemente ambas produzcan resultados similares si $f(\gamma|C, \underline{Y}_T, \underline{X}_T)$ tiene asociada una variabilidad reducida pero considera que en otras situaciones el enfoque

Bayesiano tiene una mejor oportunidad de funcionar satisfactoriamente.

Finalmente, en el trabajo discutido se analizan datos provenientes de pacientes de transplante renal con el procedimiento propuesto para estimar el momento en que se manifiesta el rechazo al transplante. De nuevo utilizando distribuciones iniciales de referencia obtiene resultados, que en su opinión, resultan satisfactorios pues describen cuantitativamente las apreciaciones cualitativas de los médicos a cargo de la evaluación del transplante. Como comentario final indica que la estimación de los parámetros en un número razonable de pacientes podría conducir a la formulación de distribuciones que ya no siendo de referencia permitan efectuar inferencias más precisas y probablemente más rápidas si el estudio se hace en tiempo real para un paciente en particular.

En una contribución mas reciente, [22], Chin Choy y Broemeling consideran el modelo lineal general, básicamente en las mismas condiciones en que es tratado por Smith en el trabajo anterior con la excepción de que no consideran restricción alguna sobre

el punto de intersección en el caso de regresión lineal simple. La principal aportación de estos autores consiste en que no solo obtienen las distribuciones finales, sobre las cuales verifican que existen aún en el caso en que la matriz de diseño sea de rango incompleto siempre que a priori se utilicen distribuciones propias, sino que las identifican como mezclas de funciones de distribución extensamente estudiadas, esto es, distribuciones de Student y Gamma. Determinan no solo los valores esperados a posteriori de los parámetros, que coinciden con los estimadores de Bayes para una función de pérdida cuadrática, si no que establecen la varianza del parámetro de precisión y las matrices de varianzas y covarianzas de los coeficientes de regresión. Adicionalmente, indican un procedimiento para el cálculo de regiones de alta densidad a partir de la distribución posteriori.

La representación del modelo lineal general utilizada por estos autores es como sigue. Sean Y_1, Y_2, \dots, Y_T variables aleatorias Normales independientes tales que:

$$Y_i = \underline{X}_i' \underline{\beta}_1 + \varepsilon_i \quad ; \quad i = 1, 2, \dots, r$$

$$Y_i = \underline{X}_i' \underline{\beta}_2 + \varepsilon_i \quad ; \quad i = r+1, \dots, T$$

en donde \underline{X}_i es el vector de dimensión p de variables independientes al tiempo i ; $i=1, 2, \dots, T$. $\underline{\beta}_1$ y $\underline{\beta}_2$ son vectores desconocidos, también de dimensión p , de coeficientes de regresión para los dos regímenes de regresión; ε_i son errores independientes e idénticamente distribuidos $N(0, \sigma^2)$ con σ^2 desconocida. Finalmente r , como de costumbre, representa el punto de cambio del que se supone que $1 < r < T$ y es desconocido

Si se define

$$\underline{Y}_1'(t) = (Y_1, \dots, Y_t)$$

$$\underline{Y}_2'(t) = (Y_{t+1}, \dots, Y_T)$$

$$\underline{\beta} = (\beta_1, \beta_2)'$$

$$X_1(t) = \begin{bmatrix} \underline{X}_1' \\ \vdots \\ \underline{X}_t' \end{bmatrix}, \quad X_2(t) = \begin{bmatrix} \underline{X}_{t+1}' \\ \vdots \\ \underline{X}_T' \end{bmatrix}, \quad \underline{\varepsilon}_1(t) = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_t \end{bmatrix}, \quad \underline{\varepsilon}_2(t) = \begin{bmatrix} \varepsilon_{t+1} \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

para $t=1, 2, \dots, T-1$ entonces el modelo original se puede representar como

$$Y_1(\tau) = X_1(\tau) \beta_1 + \epsilon_1(\tau)$$

$$Y_2(\tau) = X_2(\tau) \beta_2 + \epsilon_2(\tau)$$

o equivalentemente

$$Y_T = X(\tau) \beta + \epsilon$$

en donde

$$Y = (Y_1(\tau), Y_2(\tau))'$$

$$\epsilon = (\epsilon_1(\tau), \epsilon_2(\tau))' \quad y$$

$$X(\tau) = \begin{bmatrix} X_1(\tau) & 0 \\ 0 & X_2(\tau) \end{bmatrix}$$

De esta manera la función de densidad de Y_T , dados β_1, β_2 , $X(t)$ y σ^2 , esta dada por

$$f(Y_T | t, \beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (Y_T - X(t)\beta)' (Y_T - X(t)\beta) \right\}$$

En estas condiciones los autores verifican que si se asigna a τ una distribución a priori uniforme, $f(\tau) = (T-1)^{-1}$ para $\tau = 1, 2, \dots, T-1$ y a β y $R = 1/\sigma^2$ se les asigna una distribución con

junta independiente de τ , tal que dado $R=r$, $\underline{\beta}$ se distribuye Normal con vector de medias $\underline{\beta}$ y matriz de precisión rV con V positiva definida mientras que R sigue una distribución Gamma de parámetros α y δ entonces se tiene que,

$$\pi(t | \underline{Y}_t, \underline{X}_t) \propto D(t)^{-\alpha^*} |X(t)'X(t)+V|^{-1/2}; t=1,2,\dots,T-1$$

en donde,

$$\alpha^* = \alpha + T/2$$

$$D(t) = \delta + \frac{1}{2} \{ \underline{Y}_t' \underline{Y}_t + \underline{\beta}' V \underline{\beta}^*(t) + [X(t)'X(t)+V] \underline{\beta}^*(t) \}$$

$$D(t) = \delta + \frac{1}{2} \{ \underline{Y}_t' \underline{Y}_t + \underline{\beta}' V \underline{\beta} - \underline{\beta}^*(t)' [X(t)X(t)+V] \underline{\beta}^*(t) \}$$

con $\underline{\beta}^*(t) = [X(t)'X(t)+V]^{-1} [V \underline{\beta} + X(t)' \underline{Y}_t]$.

Por otra parte, respecto a R los autores muestran que

$$\pi(r | \underline{Y}_T, \underline{X}_T) = \sum_{t=1}^{T-1} \zeta(r | \alpha^*, D(t)) \pi(t | \underline{Y}_t, \underline{X}_t)$$

donde $\zeta(r | \alpha^*, D(t))$ representa una función de densidad Gamma con parámetros α^* y $D(t)$. En el caso de $\underline{\beta}$ se tiene que

$$\pi(\underline{\beta} | \underline{Y}_T, \underline{X}_T) = \sum_{t=1}^{T-1} h(\underline{\beta} | 2p, 2\alpha^*, \underline{\beta}^*(t), p(t)) \pi(t | \underline{Y}_t)$$

En esta expresión $h(\underline{\beta} | 2p, 2\alpha^*, \underline{\beta}^*(t), p(t))$ representa una distribución de Student multivariada de dimensión $2p$, con $2\alpha^*$ grados de libertad, vector de medias $\underline{\beta}^*(t)$ y matriz de precisión $p(t)$, en donde

$$p(t) = (\alpha^* / D(t)) \{X(t)'X(t) + V\}$$

Estos resultados se pueden verificar aplicando el teorema de Bayes y marginalizando en cada caso sobre los parámetros restantes. Como una consecuencia de la forma de la distribución a posteriori de $\underline{\beta}$ es fácil verificar que si se definen

$$\underline{\beta}^*(t) = \begin{bmatrix} \underline{\beta}_1^*(t) \\ \underline{\beta}_2^*(t) \end{bmatrix}; P(t) = \begin{bmatrix} P_{11}(t) & P_{12}(t) \\ P_{21}(t) & P_{22}(t) \end{bmatrix}$$

con $\underline{\beta}_1^*(t)$, $\underline{\beta}_2^*(t)$ vectores de dimensión p y $P_{ij}(t)$ matrices de $p \times p$; $i, j = 1, 2$, entonces

$$\pi(\underline{\beta}_1 | \underline{Y}_T, \underline{X}_T) = \int_{\underline{\beta}_2}^{\tau-1} h(\underline{\beta}_1 | p, 2\alpha^*, \underline{\beta}_1^*(t), P_1^*(t)) \pi(t | \underline{Y}_T, \underline{X}_T)$$

$$\pi(\underline{\beta}_2 | \underline{Y}_T, \underline{X}_T) = \int_{\underline{\beta}_1}^{\tau-1} h(\underline{\beta}_2 | p, 2\alpha^*, \underline{\beta}_2^*(t), P_2^*(t)) \pi(t | \underline{Y}_T, \underline{X}_T)$$

en donde

$$P_1^*(t) = P_{11}(t) - P_{12}(t)P_{22}(t)^{-1}P_{21}(t).$$

$$P_2^*(t) = P_{22}(t) - P_{21}(t)P_{11}(t)^{-1}P_{12}(t).$$

De manera inmediata se pueden establecer las distribuciones de R , $\underline{\beta}$, $\underline{\beta}_1$ y $\underline{\beta}_2$ condicionales a un valor específico de r . Naturalmente, estas distribuciones coinciden con la correspondiente en el término asociado al valor de r en la expresión de las distribuciones finales no condicionales. De cualquier forma, es claro que el conocimiento sobre R , y $\underline{\beta}$ queda descrito por mezclas de distribuciones Gammas y de Student respectivamente, en donde los coeficientes de las mezclas están dados por $\pi(t|Y_T, X_T)$ la distribución final de r .

Para efectos de estimación puntual, los autores indican que, es usual el empleo de una función de pérdidas y por tal razón exhiben las esperanzas de las distribuciones finales, que en esas condiciones coinciden con los estimadores de Bayes como ya se ha discutido y las matrices de varianzas y covarianzas correspondientes, a partir de las cuales se puede calcular la

pérdida de Bayes asociada.

Condicionando a que $t=t$ es inmediato que

$$\begin{aligned}
 E(\underline{\beta}_1 | \underline{Y}_T, \underline{X}_T, t) &= \underline{\beta}_1^*(t), \\
 E(\underline{\beta}_2 | \underline{Y}_T, \underline{X}_T, t) &= \underline{\beta}_2^*(t), \\
 E(R | \underline{Y}_T, \underline{X}_T, t) &= \alpha^*/D(t), \\
 \text{COV}(\underline{\beta}_1 | \underline{Y}_T, \underline{X}_T, t) &= \alpha^*/(\alpha^*-1) P_1^*(t)^{-1}, \\
 \text{COV}(\underline{\beta}_2 | \underline{Y}_T, \underline{X}_T, t) &= \alpha^*/(\alpha^*-1) P_2^*(t)^{-1}, \\
 \text{Var}(R | \underline{Y}_T, \underline{X}_T, t) &= \alpha^*/D(t)^2
 \end{aligned}$$

y como consecuencia,

$$\begin{aligned}
 E(\underline{\beta}_1 | \underline{Y}_T, \underline{X}_T) &= \sum_{t=1}^{T-1} \underline{\beta}_1^*(t) \pi(t | \underline{Y}_T, \underline{X}_T) \\
 E(\underline{\beta}_2 | \underline{Y}_T, \underline{X}_T) &= \sum_{t=1}^{T-1} \underline{\beta}_2^*(t) \pi(t | \underline{Y}_T, \underline{X}_T) \\
 E(R | \underline{Y}_T, \underline{X}_T) &= \sum_{t=1}^{T-1} \alpha^*/D(t) \pi(t | \underline{Y}_T, \underline{X}_T) \\
 \text{COV}(\underline{\beta}_1 | \underline{Y}_T, \underline{X}_T) &= \sum_{t=1}^{T-1} \alpha^*/(\alpha^*-1) P_1^*(t)^{-1} \pi(t | \underline{Y}_T, \underline{X}_T) \\
 \text{COV}(\underline{\beta}_2 | \underline{Y}_T, \underline{X}_T) &= \sum_{t=1}^{T-1} \alpha^*/(\alpha^*-1) P_2^*(t)^{-1} \pi(t | \underline{Y}_T, \underline{X}_T) \\
 \text{Var}(R | \underline{Y}_T, \underline{X}_T) &= \sum_{t=1}^{T-1} \alpha^* D(t)^2 \pi(t | \underline{Y}_T, \underline{X}_T) .
 \end{aligned}$$

Para la construcción de regiones de alta densidad, los autores comentan que no existe una expresión analítica, precisamente debido a que las distribuciones finales para $\underline{\beta}$ y R son mezclas. De hecho, es posible que resulten en regiones no conexas. El problema tiene que resolverse entonces, mediante integración numérica lo cual no es excesivamente complicado puesto que los componentes de las mezclas son distribuciones bien conocidas. La dificultad mayor se encuentra en determinar las regiones óptimas, de menor tamaño.

Resulta interesante observar, sin embargo, que si para algún valor de τ , $\pi(\tau | \underline{Y}_T, \underline{X}_T)$ es considerablemente mayor que $\pi(t | \underline{Y}_T, \underline{X}_T)$ para cualquier otro valor de t , entonces la distribución final de interés podría ser aproximada satisfactoriamente por la correspondiente condicional a τ y en ese caso, tanto la esperanza, la matriz de varianzas y covarianzas como las regiones de máxima densidad pueden determinarse muy fácilmente.

Los autores aplican, como ilustración, estos resultados a los datos generados por Quandt, [72], utilizando distribuciones a priori propias. De hecho para $\underline{\beta}$ suponen a priori un vec-

tor de medias que coinciden con los verdaderos valores de los parámetros. De igual forma proceden con R mientras que r le asignan una uniforme. Como resultado, obtienen que las distribuciones finales representan muy convenientemente el conocimiento sobre los parámetros. Para esos datos observan que el punto $r=12$ tiene una alta probabilidad a posteriori y en virtud de ello, calculan las distribuciones de los coeficientes de regresión y de R condicionales a ese valor. Como consecuencia, la variabilidad se reduce al igual que la longitud de los intervalos de máxima densidad, pero aún en esas condiciones, las inferencias condicionales parecen proporcionar una descripción razonablemente aproximada del conocimiento a posteriori y resultan, por supuesto, más fáciles de obtener.

En este trabajo Chin Choy y Broemeling suponen que en el modelo se efectuó un cambio y en base a esa consideración desarrollan los procedimientos de inferencia para los parámetros involucrados. El problema de detectar si efectivamente el cambio tuvo lugar ha sido objeto de otro trabajo debido a Chin Choy y Broemeling [23], en donde la detección se realiza en base a la

distribución final del punto de cambio incorporando el caso en que no ocurre cambio alguno. Concretamente, se considera que $1 \leq \tau \leq T$, de modo que si $\tau < T$ el modelo bajo estudio efectivamente involucra un cambio pero si $\tau = T$, entonces todas las observaciones son descritas por el primer y único, régimen de regresión. El problema es planteado por los autores como el correspondiente a probar las hipótesis:

$$H_0: \tau = T \quad \text{vs.} \quad H_1: 1 \leq \tau < T$$

lo cual se puede llevar a cabo de manera informal, esto es sin involucrar una función de pérdida más específica, comparando las probabilidades a posterior $\pi(T | \underline{Y}_T, \underline{X}_T)$ y $\pi(1 \leq \tau < T | \underline{Y}_T, \underline{X}_T)$. Con respecto a la formulación del trabajo anterior, se puede verificar que la principal diferencia estriba en la inclusión de T como un valor posible de τ . De esta forma se tiene que:

$$f(\underline{Y}_T | t, \underline{\beta}, \sigma^2) = \begin{cases} (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\underline{Y}_T - X(t)\underline{\beta})' (\underline{Y}_T - X(t)\underline{\beta}) \right\}; & 1 \leq t < T \\ (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\underline{Y}_T - X_1(T)\underline{\beta}_1)' (\underline{Y}_T - X_1(T)\underline{\beta}_1) \right\}; & t = T \end{cases}$$

Así, mientras que en el caso de que $\tau < T$ los parámetros desconocidos son $\underline{\beta}_1$, $\underline{\beta}_2$, σ^2 y el propio τ , cuando $\tau = T$ solo se desconocen $\underline{\beta}_1$ y σ^2 .

Por esta razón la asignación de distribuciones a priori es distinta dependiendo de si el valor de τ coincide con T . Cuando $\tau < T$, los autores utilizan la misma asignación que se utiliza en el trabajo anterior, mientras que si $\tau = T$, con una estructura similar, se asigna a $\underline{\beta}_1$ y R una distribución conjunta Normal Gamma compatible con la asignación en el caso previo, esto es R tiene una distribución Gamma de parámetros a y δ mientras que condicional a $R = r$, $\underline{\beta}_1$ tiene como distribución, la Normal que se obtiene de marginalizar $\underline{\beta}_2$ en la conjunta cuando $\tau < T$ y que tiene como vector de medias a $\underline{\beta}_1$ y como matriz de precisión $r V_{11}$ en donde,

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

De esta manera y procediendo como en el trabajo anterior, es fácil de verificar que si se asigna a priori a τ una función de probabilidad de la forma:

$$\pi(t) = \begin{cases} q & \text{si } t = T \\ (1-q)/(T-1) & \text{si } 1 \leq t < T \end{cases}$$

entonces, la distribución a posteriori de τ que se obtiene integrando respecto a β_1 , β_2 y R cuando $\tau < T$ e integrando respecto a β_1 y R únicamente cuando $\tau = T$, tiene la siguiente expresión:

$$\begin{cases} q |V_{11}|^{-1/2} D(T)^{-(T+2\alpha)/2} |X_1(T)'X_1(T)+V_{11}|^{-1/2}; \tau = T \\ \frac{(1-q)}{n-1} |V|^{1/2} D(\tau)^{-(T+2\alpha)/2} |X(\tau)'X(\tau)+V|^{-1/2}; 1 \leq \tau < T-1 \end{cases}$$

A partir de esta función de probabilidades es posible probar las hipótesis de interés ya sea a través del nomio $\eta \left(\frac{T-\tau}{T} \mid Y_T, X_T \right)$ o incluso incorporando una función de pérdidas apropiada. Los autores mencionan adicionalmente que Smith ha propuesto un procedimiento para la detección del punto de cambio en tiempo real, esto es, para la detección del cambio al mismo tiempo que se van generando las observaciones del proceso. El procedimiento consiste en calcular, inmediatamente después de que se registró la k -ésima observación la probabilidad $\pi(k \mid Y_k, X_k)$ para

$K=1,2,\dots,T$. $\pi(K|Y_K, X_K)$ representa, por supuesto, la probabilidad de no cambio hasta la observación número K y de acuerdo a Smith una tendencia consistentemente decreciente a partir de cierto punto puede considerarse evidencia a favor del cambio.

ChinChoy y Broemeling utilizan de nuevo para ilustrar la técnica, los datos de Quandt [72] y consideran tanto la detección global, con todas las observaciones, como el procedimiento secuencial de Smith. Para el primer caso consideran cuatro diferentes valores de q , la probabilidad a priori de no cambio, 0.05, 0.50, 0.95 y 0.99. Para el segundo, en adición a estos valores consideran $q=K^{-1}$; $k=1,2,\dots,20$. El análisis de los datos de Quandt, que presentan un cambio en $\tau=12$, revela que utilizando las mismas distribuciones iniciales para β y R del trabajo anterior, se tiene que para la detección global la probabilidad a posteriori de no cambio es relativamente sensible al valor de q a priori, sin embargo, en las cuatro variantes siempre la posteriori resulta menor a la a priori. Para $q=0.05$ la moda a posteriori corresponde con $\tau=12$, el verdadero valor de τ , mientras que para $q=0.50$ la probabilidad mas alta es la de $\tau=20$ (0.3855) aun

difiere de la correspondiente a $r=12$ (0.3830) en una magnitud ex
tremadamente reducida. Para $q=0.95$ y $q=0.99$ el valor inicial
tan grande de la probabilidad de no cambio provoca que la moda
a posteriori nuevamente se localice en $r=20$, en este caso avanta
jando a cualquier otro valor por mucho, sin embargo de todos
los valores de r menores que 20, $r=12$ sigue siendo el más proba
ble.

Para el procedimiento secuencial, en todos los casos la proba
bilidad de no cambio aumenta con k , hasta $k=11$ que en general tiene
un valor parecido al de $k=12$. A partir de ese punto, la probabi
lidad decrece de modo que nuevamente la evidencia apoya la hi
pótesis de cambio y localiza el punto de transición entre $r=11$ y
 $r=12$.

Como un comentario adicional, vale la pena observar que el
tratamiento desarrollado por Chin Choy y Broemeling puede ser
aplicado sin dificultad a la situación en que el cambio no solo
se refiere a los valores de los coeficientes de regresión sino
al más general caso en que los dos regímenes de regresión invo
lucran variables independientes diferentes. De hecho, las di-

mensiones de $\underline{\beta}_1$ y $\underline{\beta}_2$ pueden ser distintas sin que ello represente ningún problema adicional. Es fácil verificar que si $\underline{\beta}_1 \in \mathbb{R}^{p_1}$ y $\underline{\beta}_2 \in \mathbb{R}^{p_2}$, considerando las particiones apropiadas, los resultados obtenidos se pueden reconstruir en general. Asimismo, resulta muy valioso constatar que el procedimiento descrito puede ser aplicado aún cuando la matriz de diseño no sea de rango completo. Las expresiones obtenidas por estos autores hacen evidente tal grado de generalidad.

De esta forma, la contribución de estos autores constituye una breve, pero muy completa presentación del tratamiento Bayesiano para el problema tanto de detección como de estimación en el modelo lineal general con un punto de cambio. Los casos con más de un cambio seguramente presentan complicaciones mayores pero una alternativa puede ser el procedimiento secuencial para identificar los puntos de transición uno por uno con una estrategia similar a la descrita por Brown, Durbin y Evans [15] con el enfoque Frecuentista.

III. 3 DISCUSIÓN

En este capítulo se han presentado las mas relevantes contri
buciones en relación al problema de puntos de cambio en el mode
lo lineal general. De la misma forma que en el capítulo ante-
rior, parece evidente que los procedimientos basados en el enfo
que Frecuentista, si bien recurren a diversas herramientas que
resultan atractivas, como por ejemplo las técnicas de clasifica
ción o agrupamiento propuestas por McGee y Carleton [62], las
de análisis de datos de Brown, Durbin y Evans [63], las de pro
gramación dinámica de Hawkins [44], las de mezclas de distribu
ciones y estimación a través de la función generatriz de momen-
tos de Quandt [74] y Ramsey y Quandt [73], en general los auto
res producen resultados descriptivos o requieren de aproximacio
nes asintóticas que en la realidad pueden ser consideradas ina-
decuadas. Más aún, la diversidad de técnicas manifiesta, ade-
más del interés del problema y el indudable ingenio de sus pro-
motores, la carencia de un método que permita abordar el proble
ma con una estructura general de inferencia.

Por su parte, el enfoque Bayesiano, sin haber resuelto el

problema en su versión más amplia, ha cubierto de sobra, todos los casos considerados por los autores frequentistas con dos ventajas: Primero, los resultados se obtienen para cualquier tamaño de muestra y segundo, en todos los casos el tratamiento es esencialmente, el mismo que se deriva de los principios más elementales del método Bayesiano de Inferencia. Probablemente por esta causa los autores bayesianos ofrecen una menor diversidad en sus contribuciones y de hecho recuperan unos, los resultados de otros. Esta consistencia, sin embargo, puede ser la causa de que las dificultades técnicas de cálculo que en ocasiones se enfrentan, puedan ser estudiadas y resueltas más rápidamente y de que el cúmulo de resultados se consolide más aceleradamente abriendo las perspectivas a problemas más ambiciosos.

CAPITULO IV

Otras contribuciones.

Algunos otros autores han abordado el problema de puntos de cambio considerando tratamientos diferentes y con propósitos diversos. Swamy y Mehta, por ejemplo, han propuesto, [16], el tratamiento de un modelo de regresión lineal múltiple con cambios como un modelo con coeficientes aleatorios en el sentido frecuentista. Así, para cada observación $(Y_i, X_{i1}, X_{i2}, \dots, X_{ik})$, se tiene que

$$Y_i = \sum_{j=1}^k \beta_{jk} X_{ji}$$

en donde X_{i1}, \dots, X_{ik} son los valores fijos de las variables explicativas y el vector de coeficientes $\underline{\beta}_i = (\beta_{i1}, \dots, \beta_{ik})$ está dado por el valor observado de una distribución multivariada en $\underline{\beta}$. Estos autores consideran los casos en que la media de $\underline{\beta}$ y la correspondiente matriz de varianzas y covarianzas son tanto conocidos como desconocidos. También incluyen la posibilidad de que la distribución de $\underline{\beta}$ sea Normal. Cuando el vector de medias es conocido proceden a la estimación conjunta de los vectores de coeficientes para todas las observaciones mediante la minimización del error cuadrático medio, mientras que si la esperanza no se

conoce se le asigna una distribución de probabilidades y se estima con un procedimiento similar.

El problema también es tratado en el mismo trabajo através de máxima verosimilitud para lo cual se requiere de procedimientos iterativos para determinar la solución de las ecuaciones correspondientes.

Otra contribución que puede considerarse relacionada con el problema de puntos de cambio es la presentada por De Alba [25] en donde se propone un tratamiento, mediante el enfoque Bayesiano empírico, para la detección de desprendimientos en sucesiones de variables. En los términos más simples, la detección consiste en un procedimiento para discriminar entre las hipótesis H_0, H_1, \dots, H_T en donde bajo H_0 , las observaciones Y_1, Y_2, \dots, Y_T proceden de una población común con distribución F_1 , mientras que bajo H_i ; $i=1, 2, \dots, T$, las observaciones $Y_1, Y_2, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_T$ provienen de F_1 pero Y_i tiene función de distribución F_2 . Naturalmente, la hipótesis H_0 implica que no ha ocurrido ningún desprendimiento en tanto que bajo H_i , la observación Y_i se ha desprendido del resto. Este problema correspon-

de, en el esquema de puntos de cambio, a la situación en que se producen dos cambios de manera sucesiva antes y después de la i -ésima observación y tiene una clara relación también con las técnicas de detección de observaciones espurias o discordantes. El procedimiento descrito por este autor consiste en asignar probabilidades a priori a las hipótesis, definir una función de pérdidas y seleccionar la decisión (hipótesis) que minimice el riesgo esperado de Bayes. De Alba considera diversos casos particulares y para cada uno de ellos verifica la existencia de una regla de decisión óptima.

Por otra parte, Geisser y Eddy [31], han trabajado en torno al problema de selección de modelos con un enfoque de predicción. En cierto sentido, su propuesta guarda similitudes con el planteamiento de Bromeling [12] para predecir observaciones a partir de modelos con cambios.

Estos autores, si bien no consideran explícitamente el problema de cambios, discuten algunas técnicas para discriminar entre modelos, que pueden ser aplicadas al caso en que alguno de los modelos supone que las observaciones disponibles proceden de más

de una población. El enfoque que desarrollan, como ellos mismos lo comentan, no puede clasificarse como Bayesiano. Los propios Gelsser y Eddy llaman cuasibayesiana a una de las variantes de la técnica y cuasiverosimilitud a otra. La idea fundamental consiste en comparar los modelos bajo estudio en base a su capacidad para predecir o explicar las observaciones disponibles. La modificación esencial de los autores a este principio común en Estadística es que proponen el empleo de funciones de predicción conjunta que se obtienen como el producto de las correspondientes a cada una de las observaciones individuales en base al resto de la información. En el caso de la cuasiverosimilitud los parámetros desconocidos de los modelos son sustituidos por sus estimadores de máxima verosimilitud, mientras que para el procedimiento cuasibayesiano se les asigna una distribución inicial respecto a la cual se marginaliza posteriormente. El criterio de que se propone para la selección es la elección del modelo que maximice la función de predicción conjunta, siempre que la ganancia sobre cualquier otro competidor sea estadísticamente significativa.

La justificación de los autores para presentar este procedimiento como una alternativa a las técnicas usuales tanto Frecuentistas como Bayesianas descansa en argumentos de simplicidad y economía, sin embargo, el hecho de que en el trabajo revisado no comparen sus resultados con los que se obtienen por los otros medios, mantiene abierta la discusión sobre la utilidad de su propuesta.

Para el caso de cambios en sucesiones de variables, como se discute en el capítulo II, el tratamiento Bayesiano ha sido desarrollado de manera muy satisfactoria. Particularmente, para variables Normales, el problema de un cambio en la media ha sido completamente resuelto. Algunas generalizaciones para el caso multivariado pueden encontrarse en las contribuciones de Salazar [80], Sen [83] y Srivastava [84] y [93]. La situación análoga para la varianza que se resuelve de forma similar, no fué tratado con detalle sino hasta muy recientemente por Menzefricke [64], quien deriva la distribución a posteriori para el punto de cambio así como la correspondiente para el cociente de varianzas, cuando la media común es desconocida, así

como cuando se tienen dos medias ya sean conocidas o desconocidas. En relación al planteamiento más general de un posible cambio en un parámetro de escala se puede consultar el trabajo de Talwar y Gentle [47],

Por lo que respecta a los modelos de regresión, el problema también ha seguido siendo objeto de estudio. Así, por ejemplo Salazar, Broemeling y Chi [51], han tratado la situación con un enfoque Bayesiano para el caso de modelos de regresión con errores autocorrelacionados. En este trabajo los autores consideran como desconocidos tanto la varianza, el parámetro de autocorrelación y los coeficientes de regresión, como el punto de cambio.

El procedimiento que utilizan consiste en asignar, como es usual una distribución inicial Normal-Gamma a los coeficientes y a la precisión mientras que a priori, el punto de cambio tiene una distribución uniforme e independiente de la anterior. Por su parte el parámetro de autocorrelación es también considerado independiente del resto de los parámetros y le es asignada una distribución uniforme continua.

Por los procedimientos comunes se determinan las distribuciones a posteriori de los parámetros y concentran la atención en la correspondiente al punto de cambio. Mediante un estudio de simulación verifican que al igual que en otros casos la distribución a posteriori del punto de cambio resulta más sensible a cambios que ocurren en la sección media de la sucesión de observaciones que en los extremos.

Desde el punto de vista Frecuentista, los modelos de regresión con cambios no han sido abandonados. En una contribución publicada hace muy poco tiempo se examina una alternativa muy interesante. Esterby y El-Shaarawi [30] derivan la función de verosimilitud marginal relativa así como la función de verosimilitud marginal condicional para el punto de cambio e indican que cualquiera de estas funciones que toman en cuenta la incertidumbre respecto a los demás parámetros desconocidos, debe ser preferida sobre la usual función de verosimilitud para realizar inferencias sobre la localización del cambio. Los autores examinan con mayor detalle la función de verosimilitud marginal relativa y describen un procedimiento mediante el cual, en el caso de el modelo pueda ser representado por dos polinomios de grado

p y q respectivamente, se puede estimar estos parámetros.

Todayía en relación a modelos de regresión es importante mencionar los esfuerzos que se han realizado, tanto con técnicas Frecuentistas como Bayesianas, para el ajuste y análisis de modelos con cambios en donde éstos se efectúan de manera gradual, esto es, en donde un parámetro, por ejemplo θ , cambia de un valor θ_0 a un valor θ_1 pero esta alteración no se lleva a cabo en forma súbita sino que existe un intervalo para una variable, usualmente el tiempo, a lo largo del cual, de manera continua θ evoluciona de θ_0 hasta θ_1 . Generalmente, este tipo de modelos requiere de la construcción de una función de transición que describa la modificación de θ y los problemas más importantes radican en la estimación de los puntos en que se inicia y termina la transformación así como los valores del parámetro antes y después del cambio. Diversos autores han explorado diferentes formas para la función de transición y han empleado distintas técnicas, tanto frecuentistas como Bayesianas, para efectuar las inferencias de interés. Entre los autores que han trabajado en esta línea se encuentran Bacon y Watts [4], Tsurumi [100] y

[10] y Salazar [80].

Como una consecuencia, que puede considerarse natural, del extenso trabajo desarrollado en relación a los modelos de regresión, ha empezado a ser explorado con éxito desde el punto de vista Bayesiano, el problema de puntos de cambio en series de tiempo. Particular atención han recibido los modelos autoregresivos mas simples para los cuales Smith [91] ha propuesto técnicas tanto para detección como estimación de parámetros. Este autor además de la posibilidad del análisis retrospectivo usual ha considerado también el análisis en tiempo real para sistemas lineales que se representan como filtros de Kalman. Otros autores que han publicado resultados al respecto son Salazar [80] y Broemeling [14]. Este último autor ha comentado que aún es muy poco el trabajo que se ha realizado sobre cambios estructurales en Series de Tiempo. Los casos cubiertos son básicamente los de modelos autorregresivos y de regresión con errores autocorrelacionados. El orden de los modelos autoregresivos parece no complicar de forma adicional el problema y es previsible que las técnicas de predicción en presencia de cambios, puedan desa

rollarse con el enfoque Bayesiano, sin mayor dificultad. Sin embargo, indica que falta aún por desarrollar con detalle el tratamiento de modelos de medias móviles así como el de modelos autoregresivos de medias móviles que resultan, en general, mas complejos.

En otra dirección, es interesante consignar una contribución mas que originada en un problema en Medicina, ha sido publicado por Matthews y Farewell [6] . Estos autores consideran la detección de un cambio en una función de riesgo que describe la tasa de recaídas en pacientes de leucemia. Este enfoque, en opinión de los autores, permite la evaluación de tratamientos terapéuticos alternativos para ese padecimiento.

Es oportuno mencionar que el referido trabajo de Brocmeling [4], constituye una referencia excepcional para los interesados en el tratamiento de modelos lineales con una perspectiva Bayesianay particularmente valiosa para obtener una idea concisa pero general de las técnicas Bayesianas para puntos de cambio en modelos lineales. Si bien en la fecha en que el presente trabajo ha sido concluído, la circulación del manuscrito de

Broemeling es limitada, proxicamente aparecerá como libro con un capítulo íntegramente dedicado al tema que aquí se trata. También como referencia es recomendable indicar la publicación inminente de un número especial del Journal of Econometrics que lleva el título de "Structural Change in Econometrics" en donde se presentan trabajos que discuten el análisis de diferentes problemas relacionados con el tema de interés. Algunos de los trabajos ahí consignados han sido discutidos en este trabajo a partir de reportes preliminares pero las versiones finales ya reunidas probablemente provean un panorama más amplio de la aplicabilidad de este tipo de modelos en Econometría.

Finalmente, un trabajo más relativo a aplicaciones de este tipo, es debido a Hsu, [5] que ha publicado un artículo en que describe un procedimiento Bayesiano robusto para la detección de cambios en la estructura de riesgo en el mercado de valores que en su opinión puede producir resultados empíricos de relevancia en la teoría financiera.

CAPITULO V

Consideraciones finales.

En este trabajo se han presentado diversos procedimientos cuyo objetivo es el análisis de modelos con puntos de cambio. La búsqueda bibliográfica realizada sugiere que este tipo de modelos tienen una vasta aplicabilidad en la descripción de fenómenos económicos y biológicos.

La utilización de modelos con puntos de cambio no solo involucra ventajas en cuanto a la simplicidad del análisis, sino que en algunos casos puede contribuir a un conocimiento más profundo y una interpretación más exacta de la estructura bajo estudio.

La investigación sobre este tópico se inició considerando el caso de sucesiones de variables aleatorias pero rápidamente ha progresado el estudio en modelos de regresión. En años muy recientes se ha proseguido a la generalización del problema en series de tiempo que aún no ha sido totalmente resuelto.

En términos generales, el análisis de modelos con puntos de cambio se concentra en dos aspectos. Por un lado, la detección de los cambios y por otro, la realización de inferencias respecto a todos los parámetros involucrados en el modelo.

En relación al enfoque, en este trabajo se han distinguido, para fines de comparación, las contribuciones que emplean procedimientos Bayesianos y aquellas que pueden clasificarse como Frecuentistas. En general, estas últimas recurren, ya sea a métodos asintóticos basados sobre todo en máxima verosimilitud, o a técnicas fundamentalmente descriptivas. En cualquier forma, la mayor parte de los resultados que producen pueden considerarse solo aproximados. Las técnicas Bayesianas, por otra parte, no solo permiten resolver el problema con un tamaño moderado de muestra, sino que bajo condiciones muy generales, obtienen resultados en términos de distribuciones conocidas que han sido estudiadas en la literatura. La impresión general obtenida de la comparación Frecuentista-Bayesiana, es que a medida de la estructura del problema se complica, es más notable la ventaja de utilizar procedimientos Bayesianos, tanto por su relativa simplicidad técnica como por su consistencia conceptual.

Parece razonable esperar que aumente con un ritmo aún más acelerado la literatura sobre el tema de este trabajo y que las dificultades que actualmente se enfrentan por resolver los problemas

mas generales sean objeto de discusión desde diferentes puntos de vista. En cualquier caso, la utilidad de los modelos con puntos de cambio ha sido muy razonablemente establecida.

BIBLIOGRAFIA

- [1] Alf, M.N. y Giaccotto, C. (1982). The identical distribution hypothesis for stock market prices-location and scale-shift alternatives. Journal of the American Statistical Association 77, 377, 19-28.
- [2] Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.
- [3] Atkinson, A.C. (1978). Posterior probabilities for choosing a regression model. Biometrika, 65, 1, 39-48.
- [4] Bacon, D.W. y Watts, D.G. (1971). Estimating the transition between two intersecting lines. Biometrika, 58, 524-34.
- [5] Bellman, R.E. y Roth, R. (1969). Curve-fitting by segmented straight lines. Journal of the American Statistical Association 53, 789-98.
- [6] Bernardo, J.M. (1977). Memoria sobre conceptos, métodos y fuentes de la Biocstadística. Departamento de Biocstadística, Universidad de Valencia, España.

- [7] Bhattacharyya, G.K. y Johnson, B.A. (1968). Non parametric tests for shift at an unknown time point, Ann. Math. Statist., 39, 1731-34
- [8] Box, G.E.P. y Tiao, G.C. (1965). A change in level of a non-stationary time series. Biometrika, 52, 181-92.
- [9] Box, G.E.P. y Tiao, G.C. (1973). Bayesian Inference in Statistical Analysis, Addison-Wesley Publishing co., Massachusetts.
- [10] Box, G.E.P. (1976). Science and Statistics. Journal of the American Statistical Association, 71, 356, 791-9.
- [11] Broemeling, L. (1972). Bayesian procedures for detecting a change in a sequence of random variables. Metron, Vol. XXX-N-14.
- [12] Broemeling, L. (1977). Forecasting future values of changing sequences. Commun. Statist., -Theor. Meth. A6(1), 87-102,

- [13] Broemeling, L. (ed.) (1982). Structural Change in Econometrics. Número especial del Journal of Econometrics.
- [14] Broemeling, L. (1982). The Bayesian Analysis of Linear Models. Manuscrito no publicado.
- [15] Brown, R.L. y Durbin, J. (1968). Methods of investigating whether a regression relationship is constant over time. Presentado en el European Statistical Meeting.
- [16] Brown, R.L. Durbin, J. y Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time. Journal of the Royal Statistical Society. B37, 149-92.
- [17] Cameron, M.H. y Nash, J.E. (1974). On forecasting the manpower requirements of an organization with homogeneous workloads. Journal of the Royal Statistical Society, A7, 200-18.
- [18] Carter, R.L. y Blight, B.J.N. (1981). A bayesian change-point problem with an application to the prediction and detection of ovulation in women. Biometrics, 37, 743-52.

- [19] Chernoff, H. y Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. Ann. Math Statist. 35, 999-1018.
- [20] Chi, A.Y. (1979). The Bayesian Analysis of Structural Change in Linear Models. Disertación doctoral, Oklahoma State University. Stillwater, Oklahoma.
- [21] Chin-Choy, J.H. (1977). A Bayesian Analysis of a Changing Linear Model. Disertación doctoral, Oklahoma State University. Stillwater, Oklahoma.
- [22] Chin Choy, J.H., Broemeling, L.D. y Holbert, D.H. (1979). Detecting Structural Change in linear models. Reporte Interno, Oklahoma State University
- [23] Chin Choy, J.H. y Broemeling, L. (1980). Some bayesian inferences for a changing linear model. Technometrics, 22, 71-8.
- [24] Cooley, T.F. y Prescott, E.C. (1973). Varying parameter regression: A theory and some applications. Annals of Economic and Social Measurement, 2, 463-79.

- [25] De Alba, E. (1976). Empirical bayes multivariate slippage tests. American Statistical Association Meeting, Boston.
- [26] De Finetti, B. (1970/74). Teoria delle probabilità, sintesi introduttiva con appendice critica. (2 vol.) Torino, Einaudi, Traducido al inglés por Machi y Smith, Theory of Probability. Wiley, New York.
- [27] De Groot, M.H.(1970). Optimal Statistical Decisions. McGraw-Hill, New York.
- [28] Díaz, J. (1980). Detección de cambios de parámetro de escala en sucesiones de variables aleatorias con distribución gama. IDIAS, Cu. Rev., serie naranja, 771. UNAM, México.
- [29] Draper, N.R. y Smith, H. (1966). Applied Regression Analysis. John Wiley and sons, inc. New York.
- [30] Esterby, S.R. y El-Sharawi, A.H. (1981). Inference about the point of change in a regression model. Applied Statistics, 30, 277-85

- [31] Farley, J.V. y Hinich, M.J. (1970). A test for a shifting slope coefficient in a linear model. Journal of the American Statistical Association, 65, 1320-9.
- [32] Feder, P.I. y Sylvester, D.L. (1968). On the asymptotic theory of least squares estimation in segmented regression: identified case. Ann. Math. Statist., 39, 1362.
- [33] Feder, P.I. (1975). On asymptotic distribution theory in segmented regression problems-identified case. Ann. Statist., 3, 49-83.
- [34] Feder, P.I. (1975). The log likelihood ratio in segmented regression. Ann. Statist., 3, 84-97.
- [35] Ferreira, P. (1975). A bayesian analysis of a switching regression model: known number of regimes. Journal of the American Statistical Association, 70, 370-5.
- [36] Gallant, A.R. y Fuller, W.A. (1973). Fitting segmented polynomial regression models whose join point have to be estimated. Journal of the American Statistical Association, 144-147.

- [37] Geiser, S. y Eddy, W.F. (1979). A predictive approach to model selection. Journal of the American Statistical Association, 74, 153-60.
- [38] Goldfeld, S.M. y Quandt, R.E. (1972). Non linear Methods in Econometrics, North-Holland publishing.co., Amsterdam.
- [39] Goldfeld, S.M. y Quandt, R.E. (1973). A markov model for switching regressions. J. of Econometrics, 1, 3-16.
- [40] Goldfeld, S.M. y Quandt, R.E. (1973). The estimation of structural shifts by switching regressions. Annals of Economic and Social Measurement, 2, 475-85.
- [41] Green, C.V. y Fekete, E. (1933). Differential growth in the mouse, J. Esp. Zool., 66, 351-70.
- [42] Halpern, E.F. (1973). Bayesian spline regression when the number of knots is unknown. Journal the Royal Statistical Society, B35, 347-60.
- [43] Halpern, E.F. (1973). Polynomial regression from a bayesian approach. Journal of the American Statistical Association, 68, 137-43.

- [44] Hawkins, D.M. (1976). Point estimation of the parameters of piecewise regression. Appl. Statist., 25, 51-7.
- [45] Hinkley, D.V. (1969). Inference about the intersection in a two-phase regression. Biometrika, 56, 495-504.
- [46] Hinkley, D.V. (1970). Inference about the change-point in a sequence of random variables. Biometrika, 57, 1-17.
- [47] Hinkley, D.V. y Hinkley, E.A. (1970). Inference about the change-point in a sequence of binomial variables. Biometrika, 57, 477-80.
- [48] Hinkley, D.V. (1971). Inference in two-phase regression. Journal of the American Statistical Association, 66, 736-43.
- [49] Holbert, D. y Broemeling, L. (1977). Bayesian inferences related to Shifting sequences and two-phase regression. Commun. Statist. Theor. Meth. A6(3), 265-75.

- [50] Hsu,D.A.(1979). Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis. Journal of the American Statistical Association, 74, 31-40.
- [51] Hsu,D.A.(1982). A Bayesian robust detection of shift in the risk structure of stock market returns, Journal of the American Statistical Association 77, 377, 29-39
- [52] Hudson,D.J.(1966). Fitting segmented curves whose joint points have to be estimated. Journal of the American Statistical Association, 61, 1097-129
- [53] Kander,Z. y Zacks,S.(1966). Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. Ann.Math.Statist., 37, 1196-20.
- [54] Kendall,N.G. y Stuart,A.(1967). The Advanced theory of Statistics, Vol.I, Griffin, London.
- [55] Kiefer,N.(1978). Discrete parameter variation: efficient estimation of a switching regression model. Econometrica, 46, 427-34.

- [56] Khan,M.S.(1974). The stability of the demand for money function in the United States. J.Polit.Econ.,82, 1205-1219.
- [57] Knapp,M.S. Blamey,R.,Cove-Smith,R. y Heath,M.(1977). Monitoring the function of renal transplants. The Lancet, 1183.
- [58] Lerman,P.M.(1980). Fitting segmented regression models by grid search. Appl.Statist. 29, 77-84.
- [59] Lindley,D.V.(1965). An introduction to Probability and Statistics from a Bayesian View point.(2 vol.) Cambridge University, Press, Cambridge.
- [60] Lindley,D.V.(1971). Making Decisions, Wiley, New York.
- [61] Matthews,D.E. y Farewell,V.T.(1982). On testing for a constant hazard against a change-point alternative.Biometrics,38, 463-8.
- [62] McGee,V.E. y Carleton,W.T.(1970). Piecewise regression. Journal of the American Statistical Association, 65, 1109-24.

- [63] Mendoza, M. (1982). A modification to some proposed tests in relation to the problem of switching regression models. Enviado para publicación.
- [64] Menzefricke, U. (1981). A bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point. Applied Statistics, 30, 141-6.
- [65] Mustafi, C.K. (1968). Inference problems about parameters which are subjected to changes over time. Ann.Math.Statist., 39, 840-54.
- [66] Needham, A.E. (1935). On relative growth in the jaws of certain fishes. Proc.Zool.Soc.London. 2, 773-84.
- [67] Ohki, K. (1974). Manganese nutrition of cotton under two boron levels. II. Critical Mn level. Agron.J. 66, 572-5.
- [68] Page, E.S. (1955). A test for a change in a parameter occurring at an unknown point. Biometrika, 42 523-7.

- [69] Page, E.S. (1957). On problems on which a change in parameter occurs at an unknown point. Bionetrika, 44, 248-52.
- [70] Poirier, D.J. (1976). The Econometrics of Structural Change. North-Holland publishing co. Amsterdam.
- [71] Pool, J. & Borchgrevinck, C.F. (1964). Comparison of rat liver response to coumarin administered in vivo versus in vitro. Amer. J. Physiology 206(1), 229-38.
- [72] Quandt, R.E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. Journal of the American Statistical Association, 53, 873-80.
- [73] Quandt, R.E. (1960). Tests of the hypothesis that a model regression system obeys two separate regimes. Journal of the American Statistical Association, 55, 324-30.

- [74] Quandt, R.E. (1972). A new approach to estimating switching regressions. Journal of the American Statistical Association, 67, 306-10.
- [75] Quandt, R.E. y Ramsey, J.B. (1978). Estimating mixtures of normal distributions and switching regressions. Journal of the American Statistical Association, 73, 730-52.
- [76] Ramsey, F.P. (1931/64). Truth and Probability. Reeditado en Studies in Subjective Probability Kyburg y Smokler eds., 61-92, Wiley, New York.
- [77] Rao, C.R. (1948). The utilization of multiple measurements in problems of biological classification. Journal of the Royal Statistical Society, B10, 159-203.
- [78] Rao, C.R. (1973). Linear Statistical Inference and its Applications 2nd. edition. John Wiley and Sons. New York.
- [79] Robinson, D.E. (1964). Estimates for the points of intersection of two polynomial regressions. Journal of the American Statistical Association, 59, 214-24

- [80] Salazar, D. (1980). The Analysis of Structural Changes in Time Series and Multivariate Linear Models. Disertación doctoral, Oklahoma State University, Stillwater, Oklahoma.
- [81] Salazar, D., Broemeling, L. y Chi, A. (1981). Parameter changes in a regression model with auto-correlated errors. Comm. Statist. A10(7), 1751-8.
- [82] Savage, L. J. (1954). The Foundations of Statistics, Wiley, New York.
- [83] Sen, A. K. y Srivastava, M. S. (1973). On multivariate test for detecting change in mean. Sankhya, A35, 173-85.
- [84] Sen, A. K. y Srivastava, M. S. (1975). On tests for detecting a change in mean. Ann. Statist., 3, 98-108.
- [85] Sen, A. K. y Srivastava, M. S. (1975). On tests for detecting change un mean when variance is unknown. Ann. Inst. Statist. Math. 27, 479-86.

- [86] Sen, A.K. y Sriyastava, M.S. (1975). Some one-sided tests for change in level. Technometrics, 17, 61-4.
- [87] Singpurwalla, N.D. (1974). Estimation of the joint point in a heterocedastic regression model arising in accelerated life test. Commun. Statist., 3(9), 853-63.
- [88] Smith, A.F.M. (1973). A general bayesian linear model. Journal of the Royal Statistical Society, B35, 67-75.
- [89] Smith, A.F.M. (1975). A bayesian approach to inference about a change-point in a sequence of random variables. Biometrika, 62, 407-16.
- [90] Smith, A.F.M. y Makov, V.E. (1979). Bayesian detection and estimation of jumps in linear systems. Procc. IMA Conference on the Analysis and Optimization of Stochastic Systems. Academic Press.

- [91] Smith, A.F.M. (1980). Change-point problems: approaches and applications. En Bayesian Statistics (editado por J.M. Bernardo), Imprenta Universitaria, Valencia, España.
- [92] Smith, A.F.M. y Cook, D.G. (1980). Straight lines with a change-point: a bayesian analysis of some renal transplant data. Appl. Statist., 2, 180-9
- [93] Sprent, P. (1961). Some hypothesis concerning two-phase regression lines. Biometrics, 17, 634-45.
- [94] Sprent, P. (1969). Models in Regression and Related Topics. Methuen's monographs on applied probability and statistics, Methuen and co. Ltd, London.
- [95] Srivastava, M.S. (1980). On tests for detecting change in the multivariate mean". Technical Report. No.3 University of Toronto.
- [96] Swamy, P. y Metha, J. (1975). Bayesian and non-bayesian analysis of switching regressions and random coefficients regressions models. Journal of the American Statistical Association, 70, 593-602.

- [97] Talwar, P.T. y Gentle, J.H. (1981). Detecting a scale shift in a random sequence at an unknown time point. Applied Statistics 30, 301-4.
- [98] Tsurumi, H. (1977). A bayesian test of a parameter shift: with a application. J. of Econometrics, 371-80.
- [99] Tsurumi, H. (1978). A bayesian test of a parameter shift in a simultaneous equation with an application to a macro saving function. Economic Studies Quarterly, 24(3), 216-30.
- [100] Tsurumi, H. (1980). A bayesian and maximum likelihood analysis of a gradual switching regression model with sampling experiments. Presentado en la sesión No. 21 del Seminario NSF-NBER Sobre Inferencia Bayesiana en Econometría y Estadística, Universidad de Chicago.
- [101] Tsurumi, H. (1980). A bayesian estimation of structural shifts by gradual switching regressions with an applications to the U.S. gasoline market. A. Zellner (ed.) Bayesian Analysis in Econometrics and Statistics in honor of Harold Jeffreys. North-Holland.

- [102] Tukey, J.W. (1977). Exploratory Data Analysis.
Addison-Wesley, Massachusetts.
- [103] Wichern, D.W., Miller, R.B. y Hsu, D.A. (1976). Changes
of variance in first order auto regressive time
series models-with an application. Journal of
the Royal Statistical Society, C., 25, 248-56.