



Universidade de Vigo

Trabajo Fin de Máster

Regresión Segmentada

Víctor González Carro

Máster en Técnicas Estadísticas

Curso 2016-2017

Propuesta de Trabajo Fin de Máster

Título en galego: Regresión Segmentada
Título en español: Regresión Segmentada
English title: Segmented Regression
Modalidad: A
Autor/a: Víctor González Carro, Universidade de Santiago de Compostela
Director/a: Rosa María Crujeiras Casais, Universidade de Santiago de Compostela
Breve resumen del trabajo: Este trabajo analizará en profundidad los modelos de regresión segmentada en el caso en el que los dos trozos en los que se divide la función de regresión sean rectas. Se adaptarán los métodos clásicos de estimación por máxima verosimilitud y mínimos cuadrados además de uno basado en las aproximaciones de Taylor. Estas tres variantes se compararán en un estudio de simulación en el que se prueban los métodos en diferentes escenarios. Se incluye además una aplicación a datos reales en la que se estima un umbral en la cantidad de hábitat a partir del cual la tasa de extinción de especies decrece bruscamente. Por último se incluyen ampliaciones sobre el estudio del modelo de regresión segmentada que aborda alguno de los problemas de inferencia relacionados con el modelo de regresión segmentada.
Otras observaciones: Propuesta del alumno: Víctor González Carro

Doña Rosa María Crujeiras Casais, profesora de la Universidade de Santiago de Compostela, informa que el Trabajo Fin de Máster titulado

Regresión Segmentada

fue realizado bajo su dirección por don Víctor González Carro para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Santiago de Compostela, a 12 de julio de 2017.

El/la director/a:

El/la autor/a:

Doña Rosa María Crujeiras Casais

Don Víctor González Carro

Índice general

Resumen	IX
1. El modelo de regresión lineal segmentada	1
2. Estimación en regresión segmentada	5
2.1. Métodos de estimación	5
2.1.1. Estimación por mínimos cuadrados	5
2.1.2. Estimación por máxima verosimilitud	5
2.2. Estimadores del punto de cambio	7
2.2.1. Método de Hudson (1966)	8
2.2.2. Método de Hinkley (1969)	14
2.2.3. Método de Muggeo (2003)	16
2.3. Estimación con varios puntos de cambio	19
2.3.1. Método de Hudson para varios puntos de cambio	19
2.3.2. Método de Muggeo para varios puntos de cambio	20
3. Estudio de simulación	21
3.1. Escenarios de simulación	21
3.2. Estimación del punto de cambio	27
3.3. Estimación de la diferencia de pendientes	40
4. Aplicación a datos reales	47
4.1. Características de la muestra	48
4.2. Ajuste del modelo de regresión segmentada	49
5. Ampliación del estudio sobre el modelo de regresión segmentada	51
5.1. Contraste de existencia de punto de cambio	51
5.2. Contraste de no efecto	53
A. Notación	57
Bibliografía	61

Resumen

Resumen en español

Los modelos de regresión segmentada o regresión a trozos se caracterizan por la existencia de un valor en la variable explicativa (llamado punto de cambio) que supone una variación en los valores de los parámetros de tendencia de la función de regresión. El presente documento analizará en profundidad los modelos de regresión segmentada en el caso en el que los dos trozos en los que se divide la función de regresión sean rectas y con la hipótesis a mayores de que la función de regresión sea continua en el punto de cambio. Concretamente, se revisarán las adaptaciones de la estimación mínimo cuadrática y máximo verosímil a este caso, además de un método iterativo basado en aproximaciones de Taylor. Se comparará la actuación de estos métodos en un estudio de simulación y en un ejemplo real en el contexto de la Biología, donde se pretende estimar el punto de cambio en la relación entre la cantidad de hábitat en un paisaje y la probabilidad de existencia de una especie.

English abstract

Segmented regression models (or piecewise regression) are characterised by the existence of a value in the explanatory variable (the so called breakpoint) which implies a change in the values of the trend parameters of the regression function. This document analyses segmented regression models where the two resultant pieces are straight lines deeply, with the additional hypothesis of continuity of the regression function at the breakpoint. More precisely, least squares and maximum likelihood estimators' adaptations to this context are reviewed, as well as an algorithm based on Taylor approximations. The performance of these methods will be compared in a simulation study and a real data example in the context of Biology, where we intend to estimate the breakpoint in the relationship between the amount of habitat in a landscape and the probability of occurrence of a certain species.

Capítulo 1

El modelo de regresión lineal segmentada

En este capítulo se introducen los modelos de regresión lineal para luego particularizar en el caso de la regresión lineal simple y, a partir de la misma, motivar el planteamiento del modelo de regresión segmentada.

Los modelos de regresión surgen con el propósito de explicar la relación de dependencia entre una variable llamada **variable respuesta o dependiente** con otra variable o conjunto de variables denominadas **variables explicativas o independientes**. La modelización de esta relación debe considerar, además, la existencia de errores de medida o de una cierta aleatoriedad en las variables. Por tanto, si Y denota la variable respuesta y X_1, \dots, X_{p-1} son $p-1$ variables explicativas, el planteamiento general del modelo de regresión vendría dado por

$$Y = m(X_1, \dots, X_p) + \varepsilon, \quad (1.1)$$

donde $m(x_1, \dots, x_{p-1}) = E(Y|X_1 = x_1, \dots, X_{p-1} = x_{p-1})$ y $E(\varepsilon|X_1 = x_1, \dots, X_p = x_{p-1}) = 0$, siendo $E(\cdot|\cdot)$ la esperanza condicionada. La función $m(\cdot)$ recibe el nombre de **función de regresión**, mientras que ε es el **error** del modelo.

Existen diferentes tipos de modelos en función de la forma que tenga $m(\cdot)$. Los **modelos de regresión lineal** establecen como hipótesis que la relación entre las variables explicativas y la respuesta se puede expresar en términos de una función lineal en las variables dependientes. Este trabajo se situará en el contexto de la **regresión lineal simple**, que sólo considera una única variable explicativa denotada por X . La ecuación de regresión del modelo se podría escribir¹ como

$$Y = \alpha_1 + \beta_1 X + \varepsilon, \quad \alpha_1, \beta_1 \in \mathbb{R} \quad (1.2)$$

Las principales ventajas de este modelo son su sencillez (en el sentido de que intervienen pocos parámetros) y fácil interpretación. Con esta notación, α_1 representaría el valor de la variable respuesta predicho por el modelo cuando $X = 0$ y β_1 mediría el aumento en la variable Y al incrementar X una unidad. A partir de aquí, α_1 recibirá el nombre de **intercepto** y β_1 el de **pendiente**, pues se está modelizando la relación mediante una recta.

Sin embargo, los modelos de regresión lineal asumen que el ritmo al que crece Y al aumentar X es el mismo independientemente del valor que tenga X . Este hecho es, en ocasiones, poco realista y se antoja necesaria la formulación de un modelo que contemple estos casos. Una primera solución pasaría entonces por suponer que existe un punto $\psi \in \mathbb{R}$ tal que la relación de X e Y es lineal antes y después de ψ pero con diferentes pendientes. De ahora en adelante, el valor ψ será denominado

¹La notación se fija de esta forma para poder relacionar el modelo de regresión lineal simple y el modelo de regresión segmentada.

punto de cambio. De este modo, adaptando la notación usada por Muggeo (2003), la ecuación (1.2) se reescribiría como

$$Y = \alpha_1 + \beta_1 X + \beta_\psi (X - \psi)_+ + \varepsilon, \quad (1.3)$$

donde $(X - \psi)_+ = (X - \psi)I(X > \psi)$, siendo $I(\cdot)$ la función indicadora. Con esta notación, β_1 es la pendiente del modelo cuando $X \leq \psi$ y $\beta_1 + \beta_\psi$ es la pendiente cuando $X > \psi$. Equivalentemente, β_ψ puede ser interpretado como la diferencia entre la primera y la segunda pendiente del modelo. Nótese que, ya que se supone que el punto de cambio existe, por definición se tiene que $\beta_\psi \neq 0$. De no ser así, los modelos (1.2) y (1.3) coincidirían.

Una suposición que lleva implícita la construcción de este modelo y que merece ser comentada es que la función $m(\cdot)$ debe ser continua en el punto $X = \psi$. El caso de no continuidad fue estudiado por Quandt (1958).

Por supuesto, la anterior situación se puede adaptar a contextos más generales en los que haya K puntos de cambio ψ_1, \dots, ψ_K para la variable explicativa. El modelo descrito en (1.3) sería un caso particular del modelo general

$$Y = \alpha_1 + \beta_1 X + \sum_{k=1}^K \beta_{\psi_k} (X - \psi_k)_+ + \varepsilon, \quad (1.4)$$

con $(X - \psi_k)_+ = (X - \psi_k)I(X > \psi_k)$, $k = 1, \dots, K$. En este caso, β_{ψ_k} representa el cambio de pendiente entre $X = \psi_k$ y $X = \psi_{k+1}$ cuando $k = 1, \dots, K - 1$. Equivalentemente, $\beta_{\psi_1} + \dots + \beta_{\psi_k}$ mide el aumento de Y al aumentar una unidad X cuando ésta toma valores en el intervalo $[\psi_k, \psi_{k+1})$, $k \in \{1, \dots, K - 1\}$ ó $[\psi_k, +\infty)$ en el caso de β_{ψ_K} . Del mismo modo que antes, se supondrá que la función de regresión debe ser continua en todos los puntos ψ_1, \dots, ψ_K .

Para poder realizar inferencia sobre el modelo de regresión segmentada, es necesario que éste verifique una serie de hipótesis:

- **Homocedasticidad.** La varianza del error es la misma para cualquier valor de la variable X . Esto es, $\text{Var}(\varepsilon|X = x) = \sigma^2 \quad \forall x \in \text{Sop}(X)$, siendo $\text{Sop}(X)$ el soporte de la variable X .
- **Normalidad.** Habitualmente, se suele exigir también que la distribución de ε sea normal. Juntando esta condición con la anterior, esto se traduciría en que $\varepsilon \sim N(0, \sigma^2)$.
- **Independencia.** Dada una muestra aleatoria simple de tamaño n , las variables aleatorias que representan a los errores $\varepsilon_1, \dots, \varepsilon_n$ son mutuamente independientes.

Estas tres hipótesis también se suelen asumir para hacer inferencia cuando se trabaja con el modelo de regresión lineal (1.2).

Ya que el modelo (1.3) es un modelo paramétrico, la relación entre X e Y depende de unos parámetros teóricos desconocidos en la práctica. Será necesario entonces tomar muestras con observaciones conjuntas de las dos variables para poder estimar los parámetros del mismo. Este muestreo se puede realizar de dos maneras:

- **Diseño fijo.** En este caso los valores de la variable explicativa están fijados de antemano de acuerdo a algún criterio que considere apropiado para sacar conclusiones del mismo.
- **Diseño aleatorio.** Los valores obtenidos de las dos variables al tomar la muestra son aleatorios.

De ahora en adelante se supondrá diseño fijo salvo que se especifique lo contrario y las muestras se denotarán por $\{(x_1, Y_1), \dots, (x_n, Y_n)\}$.

Así pues, dada una muestra aleatoria simple, el modelo (1.3) sugiere que

$$Y_i = \alpha_1 + \beta_1 x_i + \beta_\psi (x_i - \psi)_+ + \varepsilon_i \quad i \in \{1, \dots, n\}, \quad (1.5)$$

donde ε_i es el error asociado a la i -ésima observación de la muestra. Por simplicidad, de ahora en adelante se supondrá que los x_i están ordenados de forma creciente. Es decir, $x_1 \leq x_2 \leq \dots \leq x_n$.

A lo largo de los años, numerosos autores han propuesto métodos para estimar los parámetros del modelo (1.3) en diferentes contextos. En el caso de que el punto de cambio ψ sea conocido, Sprent (1961) estima los parámetros de tendencia mediante mínimos cuadrados haciendo uso de los multiplicadores de Lagrange. Hinkley (1969) se basa en estas ideas y desarrolla un método para estimar los parámetros de tendencia cuando ψ sea desconocido. Dicho método se resume en encontrar el intervalo en el que se maximice la función de verosimilitud asociada al modelo y hallar el punto de corte de dos rectas en él. Este método es muy similar al de Hudson (1966), que no maximiza la función de verosimilitud, sino que minimiza la suma cuadrática de residuos. Muggeo (2003) propone un algoritmo basado en aproximaciones de Taylor para estimar los parámetros del modelo, incluido el punto de cambio ψ . También analiza la posibilidad de que existan varios puntos de cambio y presenta una sencilla forma de extender su algoritmo a este supuesto. Hudson (1966) también contempla esta situación, aunque su propuesta conlleva una gran carga computacional.

Estos métodos suponen la existencia de punto de cambio, que no siempre tiene por qué estar garantizada. Para contrastar esta hipótesis, Hinkley (1971) propone un test de razón de verosimilitudes basado en el estadístico de Hinkley (1969). Por su parte, Davies (1977, 1987) desarrolla un contraste de tipo Kolmogorov–Smirnov y da una cota superior para el p-valor de su contraste. Este último procedimiento es el utilizado por Muggeo (2003) para verificar si realmente existe un punto de corte. Este documento presta especial atención al test de Hinkley (1971) y lo modifica levemente, proponiendo la calibración del mismo mediante Bootstrap.

Además de en los estudios teóricos, la regresión segmentada ha sido aplicada en la práctica en artículos de varias disciplinas, aunque se utiliza principalmente en el biosanitario. Es el caso de Ulm (1991), que la emplea para comprobar si existe un umbral de peligrosidad a partir de la cual la cantidad de polvo en el aire produzca bronquitis crónica. Toms y Lesperance (2003) utilizan para probar la existencia de umbrales en la distancia al borde del bosque en bosques vírgenes y talados en relación con la composición de la vegetación del sotobosque. Muggeo (2003) emplea métodos similares a los presentados en este documento para analizar la relación entre la edad de un paciente y el tiempo en que tarda en rechazar un trasplante. En un ámbito completamente distinto al biosanitario, Ryan et al. (2002) echan mano de la regresión segmentada para probar la existencia de umbrales en el ámbito de la geología. Concretamente, estudian la relación entre el caudal de un río y el porcentaje de sedimentos que arrastra.

Cuando se plantea la utilización de regresión segmentada, es muy razonable preguntarse si verdaderamente aporta algo frente a otros métodos como, por ejemplo, la regresión suavizada. Estos procedimientos no llevan implícita ninguna suposición respecto a los datos ni a la relación existente entre las variables respuesta y explicativa, mientras que la estimación por máxima verosimilitud y la inferencia para el modelo de regresión segmentada exigen que el error del modelo sea normal y la relación sea lineal a trozos. Plantear un modelo paramétrico sin saber a ciencia cierta si realmente correcto es, en ocasiones, arriesgado. Sin embargo, también tienen otras ventajas respecto a estas otras alternativas. La más inmediata es que la regresión segmentada permite cuantificar de manera simple (a través de un parámetro) el crecimiento de una variable respecto a la otra. Además, en general todos los métodos basados en suavizado necesitan de la estimación de un parámetro de ventana. Por último, cabe decir también que si el modelo subyacente realmente es paramétrico, la tasa de convergencia de los estimadores paramétricos es más rápida que la relativa a los métodos no paramétricos.

Este trabajo consiste en comparar las diferentes alternativas para la estimación de parámetros en los modelos de regresión segmentada. Para ello, se introducen teóricamente estos métodos en el Capítulo 2 al tiempo que se comentan y se comparan diferentes propiedades de los mismos. Principalmente, este capítulo fue el resultado de la unificación de tres publicaciones: Hudson (1966), Hinkley (1969) y Muggeo (2003). Es por ello que fue necesario adaptar en el contexto de este trabajo los resultados y la notación de estos y otros artículos (todos ellos tenían la suya propia, no coincidente con ningún otro), que se puede consultar en el Apéndice A.

El Capítulo 3 incorpora un estudio de simulación en el que se pretende probar los tres métodos en diferentes escenarios para analizar sus resultados y compararlos entre ellos, vigilando que los resultados concuerden con lo que se expone en la teoría del Capítulo 2. El estudio se ha llevado a cabo

con el software estadístico R, que incorpora en el paquete `segmented` la programación del método de Muggeo (2003). La programación tanto del método de Hudson (1966) como del de Hinkley (1969) es de elaboración propia. Además, a conocimiento del autor cuando se depositó el trabajo, no existen publicaciones en la literatura que contengan un estudio de simulación comparando métodos de estimación en regresión segmentada, de modo que la realización de este capítulo aporta cierto valor al trabajo. La información se presenta mediante tablas para la estimación del punto de cambio y del tiempo de ejecución, mientras que los resultados de la diferencia de pendientes se ilustrarán con Diagramas de Caja (Boxplots).

También se ilustrarán las técnicas desarrolladas a lo largo del trabajo en un ejemplo con datos reales basado en el artículo de Betts et. al (2007), que estudian los hábitats de las aves cantoras. Concretamente, examinan si existe un umbral en la cantidad de habitats en una zona a partir del cual la tasa de colonización de estas aves se comporte de manera distinta.

Esta parte del trabajo no habría sido posible sin la inestimable colaboración del profesor Matthew Betts de la universidad de Oregon, que de forma completamente desinteresada aportó los datos de su estudio sobre la especie *Seiurus Aurocapilla*. Concretamente, aportó los valores de las variables respuesta y explicativa así como las coordenadas de cada una de ellas. Este profesor mostró en todo momento su predisposición a colaborar, pidiéndonos que lo mantuviésemos al tanto de los avances y que reconociésemos a las entidades que hicieron posible su proyecto gracias a la financiación:

Financial support was provided by the New Brunswick Department of Natural Resources, the New Brunswick Environmental Trust Fund, the Fundy Model Forest, Fundy National Park, and doctoral and postdoctoral fellowships to M.G. Betts from Natural Sciences and Engineering Research Council.

Una vez vista la estimación de parámetros de tendencia y el punto de cambio en el modelo, con el estudio de simulación y su aplicación a casos reales, el último capítulo pretende ampliar el estudio del modelo de regresión segmentada. Muestra dos contrastes de hipótesis relativos al mismo. El primero de ellos consiste en contrastar si verdaderamente hay un punto de corte y es una adaptación de Hinkley (1971), ya que dicho autor se basa en evidencias empíricas para proponer la distribución del estadístico de contraste. Frente a esto, en este documento se propone la aproximación de los valores críticos del test mediante Bootstrap. El segundo contraste también está obtenido de Hinkley (1971) y se usa para contrastar si se puede asumir no efecto a partir del punto de cambio supuesto que éste exista.

Capítulo 2

Estimación en regresión segmentada

En este capítulo se revisarán los principales métodos para estimar el punto de cambio ψ en un modelo de dos fases o segmentos presentado en la ecuación (1.3), a partir de la cual vendrá dada también de manera implícita la estimación de los parámetros de tendencia. Dichos métodos están basados en la estimación por máxima verosimilitud o mínimos cuadrados. Es por ello que, inicialmente, se introducirán brevemente estos métodos de estimación para luego aplicarlos en este contexto concreto.

2.1. Métodos de estimación

Inicialmente, se introducen la estimación de mínimos cuadrados y máxima verosimilitud, que serán de utilidad a lo largo de toda la sección.

2.1.1. Estimación por mínimos cuadrados

Dada la muestra aleatoria simple $\{(x_1, Y_1), \dots, (x_n, Y_n)\}$, los *estimadores de mínimos cuadrados* buscan los parámetros que hagan mínima la diferencia en distancia cuadrática entre las observaciones de la respuesta y las esperadas por el modelo. En el modelo lineal simple (1.2), esto sería resolver el problema de optimización:

$$(\hat{\alpha}_1^{LS}, \hat{\beta}_1^{LS}) = \arg \min_{\alpha_1, \beta_1} \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \arg \min_{\alpha_1, \beta_1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \arg \min_{\alpha_1, \beta_1} \sum_{i=1}^n (Y_i - \alpha_1 - \beta_1 x_i)^2,$$

siendo $\mathbf{y} = (y_1, \dots, y_n)'$, \mathbf{X} la matriz de diseño del modelo y $\boldsymbol{\beta} = (\alpha_1, \beta_1)'$. La suma que interviene en la anterior ecuación recibe el nombre de **suma cuadrática de residuos**. Su minimización en este caso concreto da lugar a los estimadores

$$\hat{\alpha}_1^{LS} = \bar{Y} - \frac{S_{xY}}{S_x^2} \bar{x} \quad \text{y} \quad \hat{\beta}_1^{LS} = \frac{S_{xY}}{S_x^2}, \quad (2.1)$$

que tendrán especial relevancia en posteriores secciones. Un contenido más detallado sobre la estimación mínimo-cuadrática se puede obtener consultando Seber y Wild (1989), Cap. 2 y 12.

2.1.2. Estimación por máxima verosimilitud

Habiendo presentado ya los estimadores de mínimos cuadrados, se exponen a continuación los de máxima verosimilitud. La razón para presentarlos en este orden es que los estimadores máximo-verosímiles están sujetos a las hipótesis de homocedasticidad, normalidad e independencia, mientras que los mínimo-cuadráticos se pueden definir en cualquier contexto.

Definición 2.1 Sea $\{X_1, \dots, X_n\}$ una muestra aleatoria simple de una distribución poblacional F . La distribución del vector aleatorio de n componentes formado por cada una de las observaciones de la muestra anterior se denomina *distribución conjunta*. El valor de su función de densidad asociada, la *densidad conjunta*, para la realización $\{x_1, \dots, x_n\}$ se denotará por $f(x_1, \dots, x_n)$.

Normalmente, las funciones de densidad dependen de ciertos parámetros que son desconocidos en la práctica. En líneas generales, el método de máxima verosimilitud consiste en utilizar como estimador de esos parámetros el valor que haga más probable lo observado en la muestra. La expresión formal de esta idea se muestra con las siguientes definiciones:

Definición 2.2 Sea $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p) \in \Theta \subset \mathbb{R}^p$ un parámetro del que depende la densidad conjunta, donde Θ es un espacio de parámetros admisibles. La función de verosimilitud de la muestra viene dada por $f_{\boldsymbol{\theta}}(x_1, \dots, x_n) = f(\boldsymbol{\theta}, x_1, \dots, x_n)$.

Definición 2.3 Se dice que $\hat{\boldsymbol{\theta}}$ es el *estimador de máxima verosimilitud* de $\boldsymbol{\theta}$ si

$$f_{\hat{\boldsymbol{\theta}}}(x_1, \dots, x_n) = \max_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(x_1, \dots, x_n),$$

para cada realización muestral $\{x_1, \dots, x_n\}$.

Bajo la hipótesis de que la función $f_{\boldsymbol{\theta}}(\cdot)$ es suficientemente regular, hallar el estimador de máxima verosimilitud para $\boldsymbol{\theta}$ se traduce en obtener los valores que constituyan un máximo de la función de verosimilitud. Así surgen las llamadas **ecuaciones de verosimilitud**, cuyas soluciones serán las candidatas a ser los estimadores de máxima verosimilitud de los parámetros:

$$\frac{\partial f_{\boldsymbol{\theta}}(x_1, \dots, x_n)}{\partial \theta_l} = 0, \quad l = 1, \dots, p.$$

En el contexto de la regresión, la densidad involucrada en los cálculos será la del error ε y los parámetros a estimar son los coeficientes del modelo. Por ejemplo, en el caso de la regresión lineal simple (1.2), las hipótesis de homocedasticidad, normalidad e independencia propician que la función de densidad asociada a la muestra tenga la forma:

$$f(\varepsilon_1, \dots, \varepsilon_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \prod_{i=1}^n e^{-\frac{\varepsilon_i^2}{2\sigma^2}}.$$

Dado que $\varepsilon_i = y_i - \alpha_1 - \beta_1 x_i$, el estimador de máxima verosimilitud de α_1 y β_1 será:

$$(\hat{\alpha}_1^{ML}, \hat{\beta}_1^{ML}) = \arg \max_{\alpha_1, \beta_1} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \prod_{i=1}^n e^{-\frac{(y_i - \alpha_1 - \beta_1 x_i)^2}{2\sigma^2}}. \quad (2.2)$$

Puesto que el logaritmo neperiano es una función creciente en todo su dominio, maximizar la anterior expresión es equivalente a maximizar:

$$\arg \max_{\alpha_1, \beta_1} - \sum_{i=1}^n (Y_i - \alpha_1 - \beta_1 x_i)^2.$$

Finalmente, sabiendo que el máximo de una función es el mínimo de su opuesta, se llega a que

$$(\hat{\alpha}_1^{ML}, \hat{\beta}_1^{ML}) = \arg \min_{\alpha_1, \beta_1} \sum_{i=1}^n (Y_i - \alpha_1 - \beta_1 x_i)^2,$$

que es exactamente el estimador de mínimos cuadrados de α_1 y β_1 . Puesto que en el modelo de regresión segmentada también se da esta equivalencia, de aquí en adelante no se distinguirá entre el estimador de mínimos cuadrados y el de máxima verosimilitud en dicho contexto. Más sobre este tema se puede encontrar en Seber y Wild (1989), Cap. 2 y en Vélez-Ibarrola y García-Pérez (1993), Cap. 6.

2.2. Estimadores del punto de cambio

Una vez introducidos los principales métodos de estimación, el siguiente paso será estudiar su aplicación en el contexto de la regresión segmentada con un único punto de cambio. Todos los parámetros involucrados en el modelo (1.5) son desconocidos en la práctica y, consecuentemente, es necesario estimarlos. Una situación muy habitual es aquella que supone conocida la existencia de un punto de cambio pero no su valor exacto.

Asimismo, es razonable suponer que el punto de cambio se va a encontrar entre dos valores consecutivos de la variable explicativa en la muestra. De lo contrario, el modelo ajustado a partir de la misma sería lineal simple. Así pues, a partir de ahora se supondrá que $x_i \leq \psi < x_{i+1}$ para algún $i \in \{2, \dots, n-1\}$. Una primera aproximación al problema de estimar ψ podría ser el proponer un candidato a estimador de ψ en cada intervalo y compararlos de acuerdo a algún criterio. En este caso, el valor elegido será el que maximice la función de verosimilitud asociada a los parámetros del modelo o, lo que es lo mismo, el que minimice la suma cuadrática de residuos.

Supuesto que $x_\tau \leq \psi < x_{\tau+1}$ para un cierto τ , el modelo (1.5) se puede reescribir como

$$Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_i & i = 1, \dots, \tau \\ \alpha_2 + \beta_2 x_i + \varepsilon_i & i = \tau + 1, \dots, n, \end{cases} \quad (2.3)$$

donde τ denota el índice de la muestra que marca el punto de cambio, $\alpha_2 = \alpha_1 - \beta_\psi \psi$ y $\beta_2 = (\beta_1 + \beta_\psi)$ son el intercepto y la pendiente del modelo de regresión lineal después del punto de cambio. Los cuatro parámetros de tendencia que intervienen en el modelo (2.3) recibirán el nombre de **parámetros de tendencia**. El objetivo es estimarlos y hallar el punto de cambio a partir de esas estimaciones. Con esta notación, la condición de continuidad fuerza a que ψ tome el valor

$$\psi = \frac{\alpha_1 - \alpha_2}{\beta_2 - \beta_1}. \quad (2.4)$$

Ahora, por las hipótesis de normalidad y homocedasticidad del modelo, se puede deducir que $Y_i | X_i \sim N(\alpha_1 + \beta_1 x_i + \beta_\psi (x_i - \psi)_+, \sigma^2)$, para todo $i = 1, \dots, n$. Por lo tanto, es posible calcular la función de verosimilitud asociada a la muestra, que sigue la expresión:

$$L(Y_1, \dots, Y_n; \sigma, \tau, \psi, \alpha_1, \alpha_2, \beta_1, \beta_2) = \left(\frac{1}{\sqrt{2\pi}\sigma^2} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{\tau} (Y_i - \alpha_1 - \beta_1 x_i)^2 - \frac{1}{2\sigma^2} \sum_{i=\tau+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2 \right\}, \quad (2.5)$$

con $(\alpha_1 - \alpha_2) + \psi(\beta_2 - \beta_1) = 0$ por la condición (2.4).

La función de verosimilitud (2.5) puede ser maximizada respecto de $\alpha_1, \beta_1, \alpha_2, \beta_2$ con la restricción de continuidad (2.4) de la función y condicionada a un τ fijo, como se ha dicho antes. De este modo, para cada valor de τ se obtendría $L_\tau(\psi)$, que es la función de verosimilitud asociada a que ψ esté en el intervalo $[x_\tau, x_{\tau+1})$. Además, se tiene:

$$L(\psi) = L_\tau(\psi) \quad x_\tau \leq \psi < x_{\tau+1}; \quad \tau = 2, \dots, n-1.$$

Puesto que maximizar la expresión (2.5) es equivalente a minimizar el exponente de la exponencial, el estimador de máxima verosimilitud de $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ de será el que minimice

$$RSS_\tau(\psi) = \sum_{i=1}^{\tau} (Y_i - \alpha_1 - \beta_1 x_i)^2 + \sum_{i=\tau+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2 \quad (2.6)$$

sujeta a la restricción de continuidad, siendo $RSS_\tau(\cdot)$ la suma cuadrática de residuos de un modelo de regresión segmentada cuyo punto de cambio pertenece al intervalo $[x_\tau, x_{\tau+1})$. Este problema de optimización fue abordado de dos maneras por Hudson (1966) y Hinkley (1969).

2.2.1. Método de Hudson (1966)

Hudson (1966) analiza un contexto más general, puesto que permite considerar funciones suficientemente regulares a ambos lados del punto de cambio que no sean necesariamente rectas. El autor distingue varios tipos de uniones entre los modelos antes y después del punto de cambio ψ :

- La unión de tipo 1 se produce cuando el punto de cambio se encuentra en el interior del intervalo $(x_\tau, x_{\tau+1})$ para algún $\tau \in \{2, \dots, n-1\}$ y la función de regresión no es diferenciable en $x = \psi$. El caso que se trata en este trabajo de fin de máster entra en esta categoría.
- La unión de tipo 2 supone que el cambio se produce en $\psi = x_\tau$ para algún $\tau \in \{2, \dots, n-1\}$, sea o no diferenciable la función de regresión en $x = \psi$. Este tipo de unión también engloba al modelo (1.5).
- En la mayoría de los casos no se conoce a ciencia cierta el tipo de unión que se debe producir, de modo que se analizarán los casos en que el tipo de unión sea desconocida.

Una vez presentados los escenarios, Hudson (1966) propone diferentes procedimientos que dependerán del grado de conocimiento del tipo de unión que se produce en el punto de cambio. Se expondrá el procedimiento a seguir cuando el tipo de unión es conocido primero y el procedimiento cuando la unión no es conocida después. Antes de explicarlos, es necesario introducir la notación que se va a usar.

Dado el modelo (2.3), se denotará por $g_1 = \{x_1, \dots, x_\tau\}$ y por $g_2 = \{x_{\tau+1}, \dots, x_n\}$. Esto es, g_1 y g_2 son los conjuntos de observaciones de la variable explicativa antes y después del punto de cambio.

Asimismo, se denotará por $\hat{\beta}_j = (\hat{\alpha}_j, \hat{\beta}_j)$ al vector con los estimadores de máxima verosimilitud de los parámetros del modelo (2.3) en g_j y $\tilde{\beta}_{j,\tau} = (\tilde{\alpha}_{j,\tau}, \tilde{\beta}_{j,\tau})$ al estimador local de los coeficientes de regresión sin la restricción de continuidad de los datos con abscisa en g_j , $j = 1, 2$. Estos últimos vienen dados por la ecuación (2.1).

Por último, $\hat{\psi}$ denotará al estimador de máxima verosimilitud del punto de cambio y $\hat{\psi}_\tau$ el candidato a estimador en el intervalo $[x_\tau, x_{\tau+1})$

Unión conocida de tipo 1. En el caso de que la unión sea de tipo 1, Hudson (1966) propone estimar por separado los parámetros de las dos rectas, considerando el siguiente teorema

Teorema 2.4 *Bajo las condiciones del modelo (1.3), si $x_\tau < \psi < x_{\tau+1}$, entonces $\hat{\beta}_j = \tilde{\beta}_{j,\tau}$ y $\hat{\psi} = \hat{\psi}_\tau$.*

El enunciado del teorema de Hudson (1966) sugiere que el punto de cambio está en un intervalo conocido $(x_\tau, x_{\tau+1})$, lo cual no tiene por qué ocurrir. Es necesario, entonces, disponer de un método para cuando no se conozca el intervalo que contiene al punto de cambio. Dicho método se resume en los siguientes pasos:

1. Seleccionar $\tau \in \{2, \dots, n-1\}$.
2. Calcular $\tilde{\beta}_{1,\tau}$ y $\tilde{\beta}_{2,\tau}$.
3. Calcular $\hat{\psi}_\tau = \frac{\tilde{\alpha}_{1,\tau} - \tilde{\alpha}_{2,\tau}}{\tilde{\beta}_{2,\tau} - \tilde{\beta}_{1,\tau}}$.
4. Si $x_\tau < \hat{\psi}_\tau < x_{\tau+1}$, se define la cantidad $T(\tau) = RSS_\tau^1 + RSS_\tau^2$, donde RSS_τ^j es la suma cuadrática de residuos asociada al modelo local en g_j , $j = 1, 2$. En caso contrario, se fija $T(\tau) = +\infty$. Es decir, $T(\tau) = RSS_\tau(\hat{\psi}_\tau)$ si $\hat{\psi}_\tau \in (x_\tau, x_{\tau+1})$ y $T(\tau) = +\infty$ en otro caso.

Computando el anterior algoritmo para todos los valores posibles de τ se tendrían $(n-2)$ valores de $T(\tau)$, de modo que $\hat{\psi}$ sería el $\hat{\psi}_\tau$ cuyo valor $T(\tau)$ sea mínimo.

Unión conocida de tipo 2. La unión de tipo 2 establecía que el punto de cambio era uno de los $x_\tau, \tau \in \{2, \dots, n-1\}$, lo cual no altera radicalmente el procedimiento anterior. Ahora la solución pasa por ajustar dos modelos de regresión lineal con la restricción de que en el punto $\hat{\psi}_\tau = x_\tau$ la función de regresión debe tomar el valor y_τ . El procedimiento para ajustar estas rectas se presenta a continuación.

Se plantea el problema de minimizar $(Y - X\boldsymbol{\beta})'(Y - X\boldsymbol{\beta})$ sujeto, por hipótesis a la restricción $R\boldsymbol{\beta} = 0$, donde $R = (1, \psi, -1, -\psi)$ y $\boldsymbol{\beta} = (\alpha_1, \beta_1, \alpha_2, \beta_2)$. Aplicando multiplicadores de Lagrange, se llega a que el estimador tendría la forma

$$\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}_\tau - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\tilde{\boldsymbol{\beta}}_\tau)$$

La matriz de diseño en este caso sería

$$X = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_\tau & 0 & 0 \\ 0 & 0 & 1 & x_{\tau+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_n \end{pmatrix}.$$

Por lo tanto, se tiene también que

$$X'X = \begin{pmatrix} \tau & \sum_{i=1}^{\tau} x_i & 0 & 0 \\ \sum_{i=1}^{\tau} x_i & \sum_{i=1}^{\tau} x_i^2 & 0 & 0 \\ 0 & 0 & n - \tau & \sum_{i=\tau+1}^n x_i \\ 0 & 0 & \sum_{i=\tau+1}^n x_i & \sum_{i=\tau+1}^n x_i^2 \end{pmatrix}.$$

De este modo, se divide el problema de calcular la inversa de $X'X$ en dos más pequeños, puesto que¹

$$(X'X)^{-1} = \begin{pmatrix} M_1^{-1} & 0 \\ 0 & M_2^{-1} \end{pmatrix}, M_1 = \begin{pmatrix} \tau & \sum_{i=1}^{\tau} x_i \\ \sum_{i=1}^{\tau} x_i & \sum_{i=1}^{\tau} x_i^2 \end{pmatrix}, M_2 = \begin{pmatrix} n - \tau & \sum_{i=\tau+1}^n x_i \\ \sum_{i=\tau+1}^n x_i & \sum_{i=\tau+1}^n x_i^2 \end{pmatrix}.$$

Finalmente, la suma cuadrática de residuos en este caso sigue la expresión

$$RSS_\tau(x_\tau) = RSS_\tau^1 + RSS_\tau^2 + \frac{(R\tilde{\boldsymbol{\beta}}_\tau)^2}{R(X'X)^{-1}R'}$$

Definiendo $S(\tau) = RSS_\tau(x_\tau)$, se elegirá como $\hat{\psi}$ a aquel $\hat{\psi}_\tau$ cuya $S(\tau)$ asociada sea mínima.

Tipo de unión desconocido. Esta es la situación más frecuente en la práctica, puesto que no se suele saber si el punto de cambio es o no uno de los elementos muestrales. La manera de afrontar

¹Los 0 de la matriz $(X'X)^{-1}$ son matrices nulas 2×2 .

este problema pasa por combinar los dos procedimientos previos del siguiente modo: Inicialmente, se prueba con uniones de tipo 1, es decir, suponiendo que el punto de cambio esté en el interior del intervalo $(x_\tau, x_{\tau+1})$. Posteriormente, se prueban también con uniones de tipo 2. La gran ventaja de este procedimiento es que, aunque parece que hay que realizar muchos cálculos, en realidad se ahorran muchos de ellos gracias a los teoremas que se exponen a continuación.

Teorema 2.5 *Si $T(\tau) < +\infty$, entonces $S(\tau) \geq T(\tau)$ y $S(\tau + 1) \geq T(\tau)$.*

La demostración de este teorema (y de los dos siguientes) se basa en que la suma cuadrática de residuos de un modelo con restricciones siempre es mayor o igual que la del mismo modelo construido sin restricciones. Por tanto, $RSS_\tau(x_\tau) = S(\tau) \geq RSS_\tau(\hat{\psi}_\tau) = T(\tau)$ y $RSS_{\tau+1}(x_{\tau+1}) = S(\tau + 1) \geq RSS_\tau(\hat{\psi}_\tau) = T(\tau)$.

Teorema 2.6 *Dado $i \in \{1, \dots, n\}$, si $T(i) \geq \min_{\tau \leq i} \{T(\tau)\} = T(\tau_1)$, entonces $S(i) \geq T(\tau_1)$ y $S(i+1) \geq T(\tau_1)$.*

Por el teorema anterior, se tiene que $S(i) \geq T(i)$ y $S(i+1) \geq T(i)$. Como, por definición, $T(i) \geq T(\tau_1)$, entonces $S(i) \geq T(\tau_1)$ y $S(i+1) \geq T(\tau_1)$.

Teorema 2.7 *Si ya se han determinado n_1 valores de $S(\cdot)$, siendo $S(N_1)$ es el mínimo de todos ellos, y $T(\tau) \geq S(N_1)$, entonces $S(\tau) \geq S(N_1)$ y $S(\tau + 1) \geq S(N_1)$.*

Del mismo modo que antes, $S(\tau) \geq T(\tau) \geq S(N_1)$ y $S(\tau + 1) \geq T(\tau) \geq S(N_1)$.

Teorema 2.8 *Fijado τ , si \bar{x}_{g_j} denota la media de los elementos de $g_j, j = 1, 2$, entonces $RSS_\tau(w)$ crece al aumentar $|w - \hat{\psi}_\tau|$, siempre que $w \in [\bar{x}_{g_1}, \bar{x}_{g_2}]$.*

Para la demostración de este teorema, véase Hudson (1966).

Estos teoremas tienen como cosecuencia que, en caso de que $\hat{\psi}_\tau \in [x_\tau, x_{\tau+1})$, entonces no será necesario calcular $S(\tau)$ ni $S(\tau + 1)$ porque no mejorarán la suma cuadrática de residuos relativa a $\hat{\psi}_\tau$. Del mismo modo, tampoco será necesario calcularlos cuando la suma cuadrática de residuos asociada a $\hat{\psi}_\tau$ no sea más pequeña que la menor suma cuadrática de residuos hallada hasta ese momento.

A modo de esquema, se podría resumir el procedimiento anterior en el **Algoritmo de Hudson**, que contiene los siguientes pasos:

Algoritmo 1 Algoritmo de Hudson.

1. Seleccionar un τ inicial.
 2. Calcular $\tilde{\beta}_{1,\tau}$ y $\tilde{\beta}_{2,\tau}$.
 3. Calcular $\hat{\psi}_\tau = \frac{\tilde{\alpha}_{1,\tau} - \tilde{\alpha}_{2,\tau}}{\tilde{\beta}_{2,\tau} - \tilde{\beta}_{1,\tau}}$.
 4. Si $\hat{\psi}_\tau \in [x_\tau, x_{\tau+1})$, entonces calcular $T(\tau)$, pasar al siguiente $\tau \in \{2, \dots, n - 1\}$.
 5. Repetir los Pasos 2 y 3 hasta que haya un $\tau = \tau'$ tal que $\hat{\psi}_{\tau'} \notin [x_{\tau'}, x_{\tau'+1})$. En ese caso, se calculan $S(\tau')$ y $S(\tau' + 1)$. Hecho esto, pasar al siguiente τ y volver al paso 2 con el siguiente valor de τ .
 6. Elegir como $\hat{\psi}$ el $\hat{\psi}_\tau$ de los anteriores que minimice la función $RSS_\tau(\cdot)$.
-

Además de los dos tipos de unión considerados, Hudson (1966) trata también el caso de unión continua y derivable en el punto de cambio (**unión de tipo 3**) y menciona el caso degenerado de unión fuera del intervalo $[x_1, x_n]$ (**unión de tipo 4**). Como ya ocurría con las uniones de tipo 1 y

2, existen una serie de resultados que permiten ahorrar cálculos en estas situaciones y pueden ser consultados en Hudson (1966).

Este estimador tiene buenas propiedades, como son la insesgadez asintótica y la consistencia. Además, su distribución asintótica es normal. Todo ello fue probado por Feder (1975) para modelos con funciones suficientemente regulares (no necesariamente rectas) a los lados del punto de corte. Antes de mostrar el caso de dos rectas, se supondrá inicialmente el contexto considerado por Feder (1975), denotando por $f_1(\boldsymbol{\theta}_1, x)$ y $f_2(\boldsymbol{\theta}_2, x)$ a las funciones a la izquierda y derecha del punto de cambio ψ , respectivamente. A este contexto, se añade también la suposición de que $E(|\varepsilon_i|^{2(1+\delta)}) < \infty$ para algún $\delta > 0$, $i = 1, \dots, n$. Para tratar el problema, se introducen una serie de conceptos y términos que tendrán vital importancia posteriormente.

Definición 2.9 *Se dirá que el parámetro $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ está **identificado** por el vector $\mathbf{x} = (x_1, \dots, x_n)'$ si determina una única solución posible para el estimador $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$.*

En el caso de que las funciones sean rectas, el intercepto y pendiente de ambas estará identificado si hay al menos dos puntos a cada lado de ψ , ya que se necesitan como mínimo dos puntos para construir una recta.

Definición 2.10 *Sean $G_n(\cdot)$ la distribución empírica de la variable explicativa X y $G(\cdot)$ tal que $G(0) = 0$, $G(1) = 1$ y $G_n(\cdot) \rightarrow G(\cdot)$ en distribución cuando $n \rightarrow \infty$. Un **centro de observaciones** es un punto de incremento de $G(\cdot)$.*

Feder (1975) crea a partir de la muestra un problema alternativo, que es el de calcular la distribución asintótica de los estimadores de mínimos cuadrados (tanto del punto de cambio como de los parámetros del modelo) tras haber quitado al modelo los datos pertenecientes a un intervalo estratégicamente colocado I_j , $j = 1, 2$. y obtiene resultados asintóticos para él. Puesto que estos estimadores y los que incluyen toda la muestra tienen la misma distribución asintótica, los resultados asintóticos obtenidos a un caso serán extrapolables al otro. Las demostraciones de los resultados aquí expuestos se pueden consultar en dicho artículo.

Si $d_j(n)$ denota la longitud del intervalo I_j , estos intervalos deben verificar que $d_j(n) \rightarrow 0$ y también que $(n/\log \log n)^{m_1/2} \rightarrow \infty$ cuando $n \rightarrow \infty$, siendo $m_1 - 1$ el orden de la primera derivada evaluada en $X = \psi$ que no comparten $f_1(\cdot)$ y $f_2(\cdot)$ (es decir, las primeras $m_1 - 1$ derivadas de $f_1(\cdot)$ y $f_2(\cdot)$ evaluadas en $X = \psi$ toman el mismo valor). Será necesario asumir también que al eliminar observaciones, la cantidad que se ha perdido es $o(n/\log \log n)^2$.

A continuación se proporcionan condiciones para garantizar que los estimadores de los parámetros y del punto de corte son asintóticamente insesgados.

Para cada $n \in \mathbb{N}$, sea $p(n)$ tal que $0 < p(n) \leq n$ y $p(n) \rightarrow \infty$ cuando $n \rightarrow \infty$ y sea N_n una secuencia de variables aleatorias tal que $1 \leq N_n \leq n$ verificando $N_n = O_p(p(n))$. Asimismo, sean ζ_i , $1 \leq i \leq \infty$ un conjunto prefijado de constantes, y ζ_{n_i} , $i = 1, \dots, n$ constantes tal que $\sup_n \max_{i=1, \dots, n} |\zeta_{n_i}| < \infty$ verificando además $n^\alpha \zeta_{n_i} = \zeta_i(1 + \rho_{n_i})$ para $\alpha \geq 0$, con $\max_{i=1, \dots, n} |\rho_{n_i}| = o(p^{-\frac{1}{2}}(n))$.

Lema 2.11 *Supuesto que para todo $K' > 0$ tal que $K'p(n) < n$ para n suficientemente grande y para $\delta > 0$ se verifique³⁴*

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{K'p(n)} \left(\frac{\zeta_i^2 \log \log \sum_{u=1}^i \zeta_u^2}{h(i) \sum_{u=1}^i \zeta_u^2} \right)^{1+\delta} < \infty,$$

con $h(i) = o(1)$ cuando $i \rightarrow \infty$, entonces

$$\frac{\sum_{i=1}^{N_n} \zeta_{n_i} \varepsilon_i}{\sqrt{\sum_{i=1}^{N_n} \zeta_{n_i}^2}} = O_p((\log \log n)^{\frac{1}{2}}) \quad \text{cuando } n \rightarrow \infty,$$

²Según Feder (1975) basta con exigir que G sea continua y que la pendiente sea finita en ψ para que esta condición se cumpla

³en caso de que $\log \log \sum_{u=1}^i \zeta_u^2$ no esté definido o sea negativo, se le asigna el valor 0.

⁴El δ que aparece en esta condición es el mismo que antes cuando se exigió $E(|\varepsilon_i|^{2(1+\delta)}) < \infty$.

Con el Lema 2.11 y las definiciones dadas anteriormente, el siguiente teorema muestra que los estimadores de los parámetros del modelo y del punto de cambio son asintóticamente insesgados. Asimismo, establece el ratio de convergencia de los estimadores de mínimos cuadrados a los verdaderos valores de parámetros y punto de cambio.

Teorema 2.12 *Si θ está identificado por un vector $\mathbf{x} = (x_1, \dots, x_n)$ cuyas componentes son centros de observación, se dan las condiciones del Lema 2.11 y además f_1 y f_2 tienen $(m_1 - 1)$ derivadas iguales en $X = \psi$, entonces*

$$\hat{\theta} - \theta = O_p(n^{-\frac{1}{2}}),$$

y

$$(\hat{\psi} - \psi)^{m_1} = O_p(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}}).$$

Lo siguiente es mostrar que estos estimadores son asintóticamente normales y que además son consistentes. Para ello, es necesario introducir alguna notación. De ahora en adelante, se denotará por $D^+(j, \psi, m_1)$ y $D^-(j, \psi, m_1)$ a las derivadas por la derecha e izquierda de orden m_1 de f_j evaluadas en ψ , $j = 1, 2$.

El lema que se presenta a continuación servirá de apoyo para enunciar el teorema que garantiza la normalidad asintótica del estimador de mínimos cuadrados de θ , el cual se enuncia justo después del lema. Para entenderlo mejor, recordemos que f_1 y f_2 tenían $(m_1 - 1)$ derivadas comunes en $X = \psi$.

Lema 2.13 *Si $D^+(1, \psi, m_1) = D^-(1, \psi, m_1)$, $D^+(2, \psi, m_1) = D^-(2, \psi, m_1)$ y m_1 es impar, entonces θ es un punto interior de Θ . Si m_j es par, entonces θ está en la frontera de Θ .*

Teorema 2.14 *Si θ está identificado por un vector $\mathbf{x} = (x_1, \dots, x_n)$ cuyas componentes son centros de observación, se dan las condiciones del lema (2.11) y θ es un punto interior de Θ , entonces $n^{\frac{1}{2}}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{I}^{-1})$, donde \xrightarrow{d} denota convergencia en distribución e \mathcal{I} es la matriz de información de Fisher de los parámetros del modelo.*

Juntando el Lema 2.13 y el Teorema 2.14 se obtiene el siguiente Corolario, que será de gran utilidad cuando se analice el caso de dos rectas:

Corolario 2.15 *Si $D^+(1, \psi, m_1) = D^-(1, \psi, m_1)$, $D^+(2, \psi, m_1) = D^-(2, \psi, m_1)$ y m_1 es impar, entonces $n^{\frac{1}{2}}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{I}^{-1})$.*

El Corolario (2.15) proporciona las condiciones para garantizar normalidad y consistencia del estimador de mínimos cuadrados de los parámetros de tendencia del modelo, de modo que el siguiente objetivo es buscar un resultado similar para el estimador del punto de corte de las rectas. Antes de enunciar dicho resultado, es necesario introducir la notación de los elementos que intervienen en él.

Se denotará por $D^\pm = \frac{1}{m_1}(D^+(2, \psi, m_1) - D^-(1, \psi, m_1))$. Intuitivamente, D^\pm es la diferencia de derivadas por la izquierda y por la derecha de la función de regresión en $X = \psi$. \mathbf{A} denotará una matriz $1 \times 2q$ formada por la concatenación de $\frac{\partial f_1(\theta, \psi)}{\partial \theta_l} / D^\pm$ y $-\frac{\partial f_2(\theta, \psi)}{\partial \theta_l} / D^\pm$, $l = 1, \dots, q$ verificando

$$n^{\frac{1}{2}}(\hat{\psi} - \psi)^{m_1} = \mathbf{A}[n^{\frac{1}{2}}(\hat{\theta} - \theta)].$$

Teorema 2.16 *Si $D^+(1, \psi, m_1) = D^-(1, \psi, m_1)$, $D^+(2, \psi, m_1) = D^-(2, \psi, m_1)$ y m_1 es impar, entonces $n^{\frac{1}{2}}(\hat{\psi} - \psi)^{m_1} \xrightarrow{d} N(0, \mathbf{A}\mathbf{I}^{-1}\mathbf{A}')$.*

Por tanto, queda garantizada la normalidad asintótica y la consistencia de los estimadores del punto de cambio entre dos funciones suficientemente regulares $f_1(\cdot)$ y $f_2(\cdot)$. Esto incluye también al supuesto de que $f_1(\cdot)$ y $f_2(\cdot)$ sean rectas.

En este último caso, Feder (1975) establece que si los $x_i, i = 1, \dots, n$ están equiespaciados las condiciones del Lema 2.11 se cumplen tomando $p(n) = n$, $N_n = |\hat{\tau} - \tau_v|$, $\zeta_{ni} = \frac{(\beta_1 - \beta_2)i}{n(1+o(1))}$, $\zeta_{ni} \equiv (\beta_1 - \beta_2)i$

y $h(i) = i^{-\delta/2}$, donde $\hat{\tau} = \max\{\tau | x_\tau \leq \hat{\psi} < x_{\tau+1}\}$ y $\tau_v = \max\{\tau | x_\tau \leq \psi < x_{\tau+1}\}$. Esto es, $\hat{\tau}$ es el mayor índice de los elementos muestrales menores que el estimador $\hat{\psi}$ y τ_v es el mayor índice de los elementos muestrales menores que el verdadero valor de ψ .

Dado que la primera derivada de cada función es la pendiente de la recta, el orden de la primera derivada no coincide en $X = \psi$ es $m_1 = 1$ (puesto que, por hipótesis, las pendientes son diferentes).

Se verifica también la condición de igualdad de derivadas, ya que $D^+(1, \psi, m_1) = D^-(1, \psi, m_1) = \beta_1$ y $D^+(2, \psi, m_1) = D^-(2, \psi, m_1) = \beta_2$. Por tanto, $D^\pm = D^+(2, \psi, 1) - D^-(1, \psi, 1) = \beta_2 - \beta_1$. Además, puesto que $\theta = (\alpha_j, \beta_j)$ y $f_j = \alpha_j + \beta_j x$ para $j = 1, 2$, se tiene que $A = \frac{1}{\beta_2 - \beta_1}(1, \psi, -1, -\psi)$. Quedaría únicamente por resolver el cálculo de la matriz de información de Fisher, que se efectuará en las siguientes líneas teniendo en cuenta que

$$\mathcal{I}(\alpha_1, \beta_1, \alpha_2, \beta_2; x_1, \dots, x_n) = \mathbb{E}(-H(\alpha_1, \beta_1, \alpha_2, \beta_2)),$$

siendo $H(\cdot)$ la matriz Hessiana de la función de log-verosimilitud

$$\log L(\alpha_1, \beta_1, \alpha_2, \beta_2) = -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left(\sum_{i=1}^{\tau} (Y_i - \alpha_1 - \beta_1 x_i)^2 + \sum_{i=\tau+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2 \right).$$

Lo siguiente es calcular las derivadas parciales en $\alpha_1, \beta_1, \alpha_2$ y β_2 , que vendrían dadas por

$$\frac{\partial \log L}{\partial \alpha_1} = -\frac{1}{2\sigma^2} \sum_{i=1}^{\tau} (-2Y_i + 2\alpha_1 + 2\beta_1 x_i), \quad (2.7)$$

$$\frac{\partial \log L}{\partial \beta_1} = -\frac{1}{2\sigma^2} \sum_{i=1}^{\tau} (-2Y_i x_i + 2\alpha_1 x_i + 2\beta_1 x_i^2), \quad (2.8)$$

$$\frac{\partial \log L}{\partial \alpha_2} = -\frac{1}{2\sigma^2} \sum_{i=\tau+1}^n (-2Y_i + 2\alpha_2 + 2\beta_2 x_i), \quad (2.9)$$

$$\frac{\partial \log L}{\partial \beta_2} = -\frac{1}{2\sigma^2} \sum_{i=\tau+1}^n (-2Y_i x_i + 2\alpha_2 x_i + 2\beta_2 x_i^2), \quad (2.10)$$

y las derivadas parciales de orden 2 para completar la matriz Hessiana, que serían

$$\frac{\partial^2 \log L^2}{\partial^2 \alpha_1} = \frac{-\tau}{\sigma^2}, \quad (2.11)$$

$$\frac{\partial^2 \log L^2}{\partial \alpha_1 \partial \beta_1} = \frac{-\sum_{i=1}^{\tau} x_i}{\sigma^2}, \quad (2.12)$$

$$\frac{\partial^2 \log L^2}{\partial^2 \beta_1} = \frac{-\sum_{i=1}^{\tau} x_i^2}{\sigma^2}, \quad (2.13)$$

$$\frac{\partial^2 \log L^2}{\partial^2 \alpha_2} = \frac{(\tau - n)}{\sigma^2}, \quad (2.14)$$

$$\frac{\partial^2 \log L^2}{\partial \alpha_2 \partial \beta_2} = \frac{-\sum_{i=\tau+1}^n x_i}{\sigma^2}, \quad (2.15)$$

$$\frac{\partial^2 \log L^2}{\partial^2 \beta_2} = \frac{-\sum_{i=\tau+1}^n x_i^2}{\sigma^2}. \quad (2.16)$$

Por último, se tiene que

$$\mathcal{I} = \frac{1}{\sigma^2} \mathbb{E} \begin{bmatrix} \tau & \sum_{i=1}^{\tau} x_i & 0 & 0 \\ \sum_{i=1}^{\tau} x_i & \sum_{i=1}^{\tau} x_i^2 & 0 & 0 \\ 0 & 0 & n - \tau & \sum_{i=\tau+1}^n x_i \\ 0 & 0 & \sum_{i=\tau+1}^n x_i & \sum_{i=\tau+1}^n x_i^2 \end{bmatrix}$$

Recordando que $\boldsymbol{\beta} = (\alpha_1, \beta_1, \alpha_2, \beta_2)'$ y por $\hat{\boldsymbol{\beta}} = (\hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2, \hat{\beta}_2)'$, se tiene:

$$n^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \mathcal{I}^{-1}) \quad \text{cuando } n \rightarrow \infty,$$

y

$$n^{-\frac{1}{2}}(\hat{\psi} - \psi) \xrightarrow{d} N(0, \mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}') \quad \text{cuando } n \rightarrow \infty,$$

2.2.2. Método de Hinkley (1969)

Hinkley (1969) propone un procedimiento similar al de Hudson (1966), pero modificando la función a maximizar. El autor intenta encontrar primero una expresión general de la función $RSS_{\tau}(\cdot)$, que, recordemos, era la suma cuadrática de residuos asociada a un punto de cambio en el intervalo $[x_{\tau}, x_{\tau+1})$. Para ello, se basa en los cálculos de Sprent (1961), que resuelve este problema para el caso en que ψ sea conocido utilizando multiplicadores de Lagrange. En el contexto considerado por Sprent (1961), la función a minimizar se convierte en

$$L_1(\alpha_1, \alpha_2, \beta_1, \beta_2, \lambda) = \sum_{i=1}^{\tau} (Y_i - \alpha_1 - \beta_1 x_i)^2 + \sum_{i=\tau+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2 + 2\lambda[\alpha_1 - \alpha_2 + \psi(\beta_2 - \beta_1)],$$

donde λ es el multiplicador de Lagrange. Derivando la anterior ecuación respecto de $\alpha_1, \alpha_2, \beta_1$ y β_2 se obtienen e igualando a 0, se obtienen unas ecuaciones muy similares a las obtenidas en el cálculo de la matriz de información de Fisher:

$$\frac{\partial L_1}{\partial \alpha_1} = 0 \Rightarrow \sum_{i=1}^{\tau} (-2Y_i + 2\alpha_1 + 2\beta_1 x_i) - 2\lambda = 0, \quad (2.17)$$

$$\frac{\partial L_1}{\partial \alpha_2} = 0 \Rightarrow \sum_{i=\tau+1}^n (-2Y_i + 2\alpha_2 + 2\beta_2 x_i) + 2\lambda = 0, \quad (2.18)$$

$$\frac{\partial L_1}{\partial \beta_1} = 0 \Rightarrow \sum_{i=1}^{\tau} (-2Y_i x_i + 2\alpha_1 x_i + 2\beta_1 x_i^2) - 2\lambda\psi = 0, \quad (2.19)$$

$$\frac{\partial L_1}{\partial \beta_2} = 0 \Rightarrow \sum_{i=\tau+1}^n (-2Y_i x_i + 2\alpha_1 x_i + 2\beta_1 x_i^2) + 2\lambda\psi = 0. \quad (2.20)$$

Despejando α_1 y α_2 de (2.17) y (2.18) y usando la condición de continuidad, se tiene que el multiplicador de Lagrange toma el valor

$$\lambda = \frac{\tau(n-\tau)}{n} [\bar{y}_{g_2} - \bar{y}_{g_1} + \beta_1(\bar{x}_{g_1} - \psi) - \beta_2(\bar{x}_{g_2} - \psi)]^5.$$

⁵Consultar apéndice de notación para \bar{y}_{g_j}

Sustituyendo los valores de α_1 , α_2 y λ en (2.19) y (2.20), se obtiene el sistema de dos ecuaciones:

$$\left[S_{g_1, \tau}^2 + \frac{\tau(n-\tau)}{n} (\bar{x}_{g_1} - \psi)^2 \right] \beta_1 - \frac{\tau(n-\tau)}{n} (\bar{x}_{g_1} - \psi)(\bar{x}_{g_2} - \psi) \beta_2 = C_{g_1, \tau} - \frac{\tau(n-\tau)}{n} (\bar{y}_{g_2} - \bar{y}_{g_1})(\bar{x}_{g_1} - \psi) \quad (2.21)$$

$$- \frac{\tau(n-\tau)}{n} (\bar{x}_{g_1} - \psi)(\bar{x}_{g_2} - \psi) \beta_1 + \left[S_{g_2, \tau}^2 + \frac{\tau(n-\tau)}{n} (\bar{x}_{g_2} - \psi)^2 \right] \beta_2 = C_{g_2, \tau} + \frac{\tau(n-\tau)}{n} (\bar{y}_{g_2} - \bar{y}_{g_1})(\bar{x}_{g_1} - \psi) \quad (2.22)$$

donde $C_{g_1, \tau} = \sum_{i=1}^{\tau} (x_i - \bar{x}_{g_1})(y_i - \bar{y}_{g_1})$, $C_{g_2, \tau} = \sum_{i=\tau+1}^n (x_i - \bar{x}_{g_2})(y_i - \bar{y}_{g_2})$, $S_{g_1, \tau}^2 = \sum_{i=1}^{\tau} (x_i - \bar{x}_{g_1})^2$ y $S_{g_2, \tau}^2 = \sum_{i=\tau+1}^n (x_i - \bar{x}_{g_2})^2$. Intuitivamente, $C_{g_1, \tau}$ y $C_{g_2, \tau}$ son una medida de covarianza entre X e Y y $S_{g_1, \tau}^2$ y $S_{g_2, \tau}^2$ representan una medida de la variabilidad de X a ambos lados del punto de cambio.

Las soluciones de este sistema, $\hat{\beta}_{1, \tau}$ y $\hat{\beta}_{2, \tau}$, son los estimadores de máxima verosimilitud (y también de mínimos cuadrados) de β_1 y β_2 en caso de que ψ sea conocido. Finalmente, para obtener los estimadores de máxima verosimilitud de α_1 y α_2 , denotados como $\hat{\alpha}_{1, \tau}$ y $\hat{\alpha}_{2, \tau}$, sólo quedaría sustituir las soluciones de este sistema en (2.17) y (2.18) y resolver las ecuaciones lineales.

A partir de la fórmula anterior, Hinkley (1969) llega a que $RSS_{\tau}(\cdot)$ tiene la siguiente expresión para un punto de cambio w :

$$\begin{aligned} RSS_{\tau}(w) &= \sum_{i=1}^n (y_i - \bar{y})^2 + \frac{\tau(n-\tau)}{n} (\bar{y}_{g_1} - \bar{y}_{g_2})^2 \\ &+ \hat{\beta}_{1, \tau} \left\{ \frac{\tau(n-\tau)}{n} (\bar{y}_{g_2} - \bar{y}_{g_1})(\bar{x}_{g_1} - w) - C_{g_1, \tau} \right\} \\ &+ \hat{\beta}_{2, \tau} \left\{ \frac{\tau(n-\tau)}{n} (\bar{y}_{g_1} - \bar{y}_{g_2})(\bar{x}_{g_2} - w) - C_{g_2, \tau} \right\}. \end{aligned}$$

No obstante, como ya se ha dicho previamente, ésta no será la función usada para hallar $\hat{\psi}$. En su lugar, Hinkley (1969) propone emplear

$$Z_{\tau}^2(w) = RSS^0 - RSS_{\tau}(w), \quad (2.23)$$

siendo $RSS^0 = \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{C_{g_1, n}^2}{S_{g_1, n}^2}$. De la ecuación (2.23) es necesario comentar que RSS^0 es la suma cuadrática de residuos del modelo de regresión lineal simple (sin punto de cambio) ajustado con la muestra $\{(x_i, Y_i)\}_{i=1}^n$. Por tanto, es un término que permanece constante sea cual sea el valor de w . De este modo, minimizar (2.6), que es lo mismo que maximizar la función de verosimilitud, es estadísticamente equivalente⁶ a maximizar (2.23), por lo que el objetivo ahora es encontrar el máximo de esta última función. Esto quiere decir que el método de Hinkley (1969) heredará todas las propiedades que se prueben para la distribución del método de Hudson (1966).

Para este último cometido, Hinkley (1969) desarrolla la expresión alternativa para $Z_{\tau}^2(\cdot)$:

$$Z_{\tau}^2(w) = \frac{(\tilde{\beta}_{2, \tau} - \tilde{\beta}_{1, \tau})^2 \{F_{\tau} - 2D_{\tau}(w + \hat{\psi}_{\tau}) + E_{\tau} \hat{\psi}_{\tau} w\}^2}{S_{g_1, n}^2 (F_{\tau} - 2D_{\tau} w + E_{\tau} w^2)}, \quad (2.24)$$

donde $D_{\tau} = \frac{\tau(n-\tau)}{n} (\bar{x}_{g_1} S_{g_2, \tau}^2 + \bar{x}_{g_2} S_{g_1, \tau}^2)$, $E_{\tau} = \frac{\tau(n-\tau)}{n} (S_{g_1, \tau}^2 + S_{g_2, \tau}^2)$ y $F_{\tau} = S_{g_1, \tau}^2 S_{g_2, \tau}^2 + \frac{\tau(n-\tau)}{n} (\bar{x}_{g_1}^2 S_{g_2, \tau}^2 + \bar{x}_{g_2}^2 S_{g_1, \tau}^2)$. Este estadístico

Un vistazo a la ecuación (2.24) revela que, en general, maximizar $Z_{\tau}^2(\cdot)$ irá asociado a maximizar $(\tilde{\beta}_{2, \tau} - \tilde{\beta}_{1, \tau})^2$. No obstante, en muchas ocasiones la diferencia de pendientes será máxima en algún punto

⁶Dos estadísticos estadísticamente equivalentes son aquéllos que tienen la misma distribución.

cerca de los extremos del intervalo $[x_1, x_n]$. Por ello, tiene sentido que aparezca en la fórmula el término que multiplica a esta diferencia de pendientes, ya que se puede interpretar como una ponderación de la variabilidad global de la variable explicativa. Esta variabilidad será mayor si el punto de cambio está en el centro, de modo que se compensa el desequilibrio originado por la diferencia de pendientes en los extremos.

El **Algoritmo de Hinkley** empleado para buscar el máximo de $Z_\tau^2(\cdot)$ es análogo al procedimiento de Hudson (1966), salvo porque se usa una función diferente. Se presenta a continuación:

Algoritmo 2 Algoritmo de Hinkley.

1. Seleccionar un τ inicial.
 2. Calcular $\tilde{\beta}_{1,\tau}$ y $\tilde{\beta}_{2,\tau}$.
 3. Calcular $\hat{\psi}_\tau = \frac{\tilde{\alpha}_{1,\tau} - \tilde{\alpha}_{2,\tau}}{\tilde{\beta}_{2,\tau} - \tilde{\beta}_{1,\tau}}$.
 4. Si $\hat{\psi}_\tau \in [x_\tau, x_{\tau+1}]$, entonces calcular $Z_\tau^2(\hat{\psi}_\tau)$, pasar al siguiente $\tau \in \{2, \dots, n-1\}$.
 5. Repetir los pasos 2 y 3 hasta que haya un $\tau = \tau'$ tal que $\hat{\psi}_{\tau'} \notin [x_{\tau'}, x_{\tau'+1}]$. En ese caso, descartar $\hat{\psi}_{\tau'}$ y calcular $Z_{\tau'}^2(x_{\tau'})$ y $Z_{\tau'}^2(x_{\tau'+1})$. Hecho esto, pasar al siguiente τ y se vuelve al paso 2.
 6. Elegir como $\hat{\psi}$ el $\hat{\psi}_\tau$ de los anteriores que maximice la función $Z_\tau^2(\cdot)$.
-

Sería razonable preguntarse por qué se pueden realizar los mismos pasos que se usan en el algoritmo propuesto por Hudson (1966) si la función que se utiliza es diferente. Esto se debe a que el funcional $L(\cdot)$ es continuo en todo el dominio y convexo en los intervalos $[x_\tau, x_{\tau+1}]$, $\tau = 2, \dots, n-1$. La continuidad en el intervalo $[x_1, x_n]$ garantizaría que existe solución dentro del rango de valores de la muestra, pues la imagen continua de un compacto es compacta y aquí es un intervalo cerrado, de modo que el máximo se alcanza. En particular, esto también es cierto para cada intervalo $[x_\tau, x_{\tau+1}]$, $\tau = 2, \dots, n-1$, donde además la convexidad garantiza que esta solución es global (es decir, sólo hay un máximo en cada intervalo).

Este procedimiento tiene la ventaja de que no hay que recurrir a funciones auxiliares como $T(\cdot)$ y $S(\cdot)$, ya que todos los cálculos se hacen sobre la función de verosimilitud.

2.2.3. Método de Muggeo (2003)

Los dos algoritmos tratados hasta el momento buscaban el valor del punto de cambio ψ perteneciente al modelo (2.3) en una rejilla de valores determinada por la muestra. Estos valores se obtenían dividiendo la muestra en dos partes y hallando el punto de corte entre dos rectas, lo cual permitía estimar también los parámetros de tendencia del modelo (que, recordemos, eran el intercepto α_1 y la pendiente β_1 antes del punto de cambio y el intercepto α_2 y pendiente β_2 después del punto de cambio). Muggeo (2003) propone una metodología distinta basada en la aproximación mediante un polinomio de Taylor que se expondrá primero de un modo general y, a continuación, aplicado al contexto de la regresión segmentada.

Sea $\theta^{(0)}$ una aproximación al estimador de mínimos cuadrados $\hat{\theta}$ de los parámetros del modelo. Ya que la función de regresión $m(\cdot)$ depende de θ , en este contexto se la denotará por $m(\cdot, \theta)$. Aproximando mediante un polinomio de Taylor de grado 1, $m(\cdot, \theta)$ se puede expresar como

$$m(x, \theta) \simeq m(x, \theta^{(0)}) + m'(x, \theta^{(0)})(\theta - \theta^{(0)}), \quad (2.25)$$

siendo $m'(x, \theta) = \frac{\partial m}{\partial \theta}(x, \theta)$, donde la diferencial respecto de θ denota la derivada parcial respecto a cada una de las q componentes del vector de parámetros. Esto permite reescribir el error del modelo

en los siguientes términos:

$$\varepsilon = y - m(x, \boldsymbol{\theta}) = \varepsilon^{(0)} - m'(x, \boldsymbol{\theta}^{(0)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}),$$

siendo $\varepsilon^{(0)} = y - m(x, \boldsymbol{\theta}^{(0)})$. Entonces, el estimador de mínimos cuadrados, que viene dado por $\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$, toma la forma

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \{\boldsymbol{\varepsilon}^{(0)'}\boldsymbol{\varepsilon}^{(0)} - 2(\boldsymbol{\varepsilon}^{(0)})'m'(x, \boldsymbol{\theta}^{(0)}) + (\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)})'[m'(x, \boldsymbol{\theta}^{(0)})]'[m'(x, \boldsymbol{\theta}^{(0)})](\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)})\}, \quad (2.26)$$

donde $\boldsymbol{\varepsilon}^{(0)} = (y_1 - m(x_1, \boldsymbol{\theta}^{(0)}), \dots, y_n - m(x_n, \boldsymbol{\theta}^{(0)}))'$ y $\mathbf{m}'(x, \boldsymbol{\theta}^{(0)}) = (m'(x_1, \boldsymbol{\theta}^{(0)}), \dots, m'(x_n, \boldsymbol{\theta}^{(0)}))$.

La expresión (2.26) alcanza su mínimo cuando

$$\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)} = ([\mathbf{m}'(x, \boldsymbol{\theta}^{(0)})]'[\mathbf{m}'(x, \boldsymbol{\theta}^{(0)})])^{-1}[\mathbf{m}'(x, \boldsymbol{\theta}^{(0)})]'\boldsymbol{\varepsilon}^{(0)} = \boldsymbol{\delta}^{(0)}.$$

Por tanto, el siguiente iterante será $\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}^{(0)} + \boldsymbol{\delta}^{(0)}$. Esto daría pie a calcular $\boldsymbol{\delta}^{(1)}$, para entonces obtener $\boldsymbol{\theta}^{(2)} = \boldsymbol{\theta}^{(1)} + \boldsymbol{\delta}^{(1)}$, y así sucesivamente.

En el caso de la regresión segmentada, considerando $(X - \psi)_+$ en la ecuación (1.3) como función de ψ y usando el teorema de Taylor, se tiene que

$$\beta_{\psi}(X - \psi)_+ = \beta_{\psi}(X - \psi^{(0)})I(X > \psi^{(0)}) - \gamma I(X > \psi^{(0)}),$$

con $\gamma = \beta_{\psi}(\psi - \psi^{(0)})$. De este modo, despejando en la expresión de γ , se obtiene la expresión para ψ :

$$\psi = \psi^{(0)} + \frac{\gamma}{\beta_{\psi}}. \quad (2.27)$$

Un vistazo a la ecuación (2.27) sugiere que la estrategia a seguir es encontrar un modelo en el que γ fuese lo más cercano a 0 posible. Se podrían estimar los términos de la ecuación (2.27) por mínimos cuadrados o máxima verosimilitud y con eso daría lugar al $\hat{\psi}$, el estimador de ψ , que vendría dado por

$$\hat{\psi} = \psi^{(0)} + \frac{\hat{\gamma}}{\hat{\beta}_{\psi}}. \quad (2.28)$$

De este modo, se podría construir un proceso iterativo que tomara como semilla inicial $\psi^{(0)}$ y que converja a ψ del modo descrito anteriormente. Por tanto, el **Algoritmo de Muggeo** puede resumirse en los siguientes pasos:

Este procedimiento tiene la ventaja de que proporciona estimadores de máxima verosimilitud para todos los parámetros que intervienen en el modelo en cada paso. Consecuentemente, el límite también lo será. El inconveniente es que este algoritmo no tiene asegurada la convergencia para un $\psi^{(0)}$ arbitrario, aunque rara vez se tomará como semilla inicial un punto que diste mucho del verdadero valor de ψ .

Puesto que, en cada paso, el algoritmo de Muggeo ajusta un modelo de regresión múltiple, una opción para averiguar cuándo no converge el algoritmo es intentar dar alguna condición para la cual la minimización de su suma cuadrática de residuos no tenga solución. Esto es, cuando el producto de la matriz de diseño del modelo y su traspuesta, $X'X$, no tenga inversa. Por lo tanto, a continuación se calcula dicho producto.

Dada la semilla inicial $\psi^{(0)}$, sea τ_0 el mayor entero tal que $x_{\tau_0} \leq x < x_{\tau_0+1}$. Puesto que $I(X > \psi^{(0)}) = 0$, las variables $U^{(0)}$ y $V^{(0)}$ tomarán el valor 0 para cada x_i menor que x_{τ_0} . En caso de que $i \in \{\tau_0 + 1, \dots, n\}$, $U^{(0)} = (x_i - \psi^{(0)})$ y $V^{(0)} = 1$. De este modo, la matriz de diseño tomaría la siguiente

Algoritmo 3 Algoritmo de Muggeo.

1. Seleccionar $\psi^{(0)}$ inicial.
2. Calcular $U^{(0)} = (X - \psi^{(0)})_+$ y $V^{(0)} = -I(X > \psi^{(0)})$.
3. Estimar los coeficientes relativos a $U^{(0)}$ y $V^{(0)}$ en el modelo

$$Y = \alpha_1 + \beta_1 X + \beta_\psi U^{(0)} + \gamma V^{(0)}$$

mediante $\hat{\beta}_\psi$ y $\hat{\gamma}$.

4. Calcular el nuevo iterante $\psi^{(1)} = \psi^{(0)} + \frac{\hat{\gamma}}{\hat{\beta}_\psi}$.
5. Volver al Paso 1 tomando $\psi^{(1)}$ como iterante inicial.
6. El proceso termina cuando $|\hat{\gamma}|$ es más pequeño que un número llamado *tolerancia*. Cuando $|\hat{\gamma}|$ es menor que la tolerancia, lo que realmente ocurre es que toma un valor muy cercano a 0 y no se considera necesario seguir con el algoritmo, dando por buena la aproximación de ψ obtenida. En este trabajo, la tolerancia será 0.0001 en todos los casos.

forma:

$$X = \begin{bmatrix} 1 & x_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{\tau_0} & 0 & 0 \\ 1 & x_{\tau_0+1} & x_{\tau_0+1} - \psi^{(0)} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n - \psi^{(0)} & 1 \end{bmatrix}$$

Por lo tanto, se sigue que

$$X'X = \begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=\tau_0+1}^n (x_i - \psi^{(0)}) & n - \tau_0 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=\tau_0+1}^n x_i(x_i - \psi^{(0)}) & \sum_{i=\tau_0+1}^n (x_i - \psi^{(0)}) \\ \sum_{i=\tau_0+1}^n (x_i - \psi^{(0)}) & \sum_{i=\tau_0+1}^n x_i(x_i - \psi^{(0)}) & \sum_{i=\tau_0+1}^n (x_i - \psi^{(0)})^2 & \sum_{i=\tau_0+1}^n (x_i - \psi^{(0)}) \\ n - \tau_0 & \sum_{i=1}^n x_i & \sum_{i=\tau_0+1}^n (x_i - \psi^{(0)}) & n - \tau_0 \end{bmatrix},$$

y, por tanto, el método no convergerá cuando esta matriz tenga determinante nulo.

Un caso particular se produce cuando el último iterante toma un valor entre x_1 y x_2 en el algoritmo⁷. En estos casos, la matriz de diseño tiene rango 3 y, por tanto, $X'X$ también lo tendrá, de modo que no será inversible y no se puede continuar el algoritmo dado que no es posible ajustar un modelo de regresión lineal múltiple en ese paso.

En la práctica, una estrategia es habitual es representar los datos gráficamente y proporcionar al método un $\psi^{(0)}$ inicial basado en la gráfica de la función.

⁷El razonamiento es análogo si el iterante tomase un valor entre x_{n-1} y x_n

Además, cabe decir también que este método sólo es válido cuando hay una única variable que presenta segmentación en el modelo. Es decir, si se pretende hacer regresión sobre dos variables explicativas X_1, X_2 y se supone que ambas presentan punto de cambio, el método no será válido. En la práctica, alguna simulación no incluida en el estudio confirmaba este hecho. Si bien esto es un inconveniente, los otros dos algoritmos presentados en este capítulo padecen el mismo problema.

2.3. Estimación con varios puntos de cambio

Hasta ahora se ha supuesto que sólo existe un único punto de cambio de tendencia ψ en todo el dominio. Esto deriva de que apenas hay ejemplos de situaciones en las que el efecto de la explicativa cambie varias veces de tendencia y pueda ser modelizable por rectas en todos los casos. No obstante, no hay nada que impida modelizar matemáticamente esta situación.

Así pues, como ya se dijo anteriormente, el modelo (1.3) sería un caso particular de

$$Y = \alpha_1 + \beta_1 X + \sum_{k=1}^K \beta_{\psi_k} (X - \psi_k)_+ + \varepsilon,$$

con $K = 1$ y $\psi_1 = \psi$. Por lo tanto, dada una muestra aleatoria simple, el modelo (1.4) sugiere que

$$Y_i = \alpha_1 + \beta_1 x_i + \sum_{k=1}^K \beta_{\psi_k} (x_i - \psi_k)_+ + \varepsilon_i, i = 1, \dots, n.$$

Nótese que este planteamiento exige saber exactamente cuántos puntos de cambio hay en el eje de abscisas. Ya que los tres métodos presentados hasta el momento son específicos para el caso de $K = 1$ y no contemplan este escenario, surge ahora el problema de estimar los parámetros en esta situación. Lo primero a plantearse es ver si es posible extender de algún modo los métodos de Hudson, Hinkley y Muggeo a este supuesto. La respuesta es afirmativa en el primer y tercer caso, de modo que se presentan ambos métodos a continuación.

2.3.1. Método de Hudson para varios puntos de cambio

El método de Hudson cuando hay varios puntos de cambio resuelve el problema de la localización de los K puntos de cambio por fuerza bruta. Dado el número de puntos K , el método consiste en calcular todos los posibles modelos con 0, 1 y hasta K puntos de cambio calculados con los n datos de la muestra y escoger aquél con menor suma cuadrática de residuos. Hudson (1966) proporciona el número total de modelos a comparar, que viene dado por la expresión

$$\sum_{k=0}^K 2^k \binom{K}{k} \binom{n - K - 2}{k}.$$

Naturalmente, el gran inconveniente de este método es su gran carga computacional. Esto quiere decir que en una muestra de 50 datos el método tendría que calcular y comparar la suma cuadrática de residuos de 4325 modelos, lo cual es muy costoso computacionalmente. No obstante, dependiendo de cómo sea la muestra, se pueden ahorrar algunos cálculos. Por ejemplo, supongamos que disponemos de una muestra y que se han probado los modelos sin puntos de cambio y con un punto de cambio, resultando que el que minimiza la suma cuadrática de residuos es el punto $X = x_4$ y obtiene un resultado de $RSS_4 = 10$. Lo siguiente sería probar con los modelos con dos puntos de cambio. Si, por ejemplo, para $X = x_3$ se tiene que $RSS_3^1 > 10$, no tendría sentido seguir probando todos los modelos con dos puntos de cambio y tales que uno de ellos es $X = x_3$ porque la suma siempre va a ser mayor que 10 y, por tanto, peor que el modelo con un punto de cambio.

2.3.2. Método de Muggeo para varios puntos de cambio

El método de Muggeo (2003) cuando existen varios puntos de cambio es completamente análogo al mostrado para un único punto de cambio, aunque exige saber cuántos cambios de tendencia presenta la función de regresión de antemano y proporcionar un valor inicial para cada uno de ellos. No obstante, una representación gráfica de la función de regresión puede subsanar este problema. Presentado en forma de algoritmo, el método estaría compuesto por los siguientes pasos:

Algoritmo 4 Algoritmo de Muggeo con varios puntos de cambio.

1. Seleccionar $\psi_1^{(0)}, \dots, \psi_K^{(0)}$ iniciales.
2. Calcular $U_k^{(0)} = (X - \psi_k^{(0)})_+$ y $V_k^{(0)} = -I(X > \psi_k^{(0)})$, $k = 1, \dots, K$.
3. Estimar los coeficientes relativos a $U_k^{(0)}$ y $V_k^{(0)}$, $k = 1, \dots, K$, en el modelo

$$Y = \alpha_1 + \beta_1 X + \sum_{k=1}^K \beta_{\psi_k} U_k^{(0)} + \sum_{k=1}^K \gamma_k V_k^{(0)}$$

mediante $\hat{\beta}_{\psi_k}$ y $\hat{\gamma}_k$.

4. Calcular el nuevo iterante $\psi_k^{(1)} = \psi_k^{(0)} + \frac{\hat{\gamma}_k}{\hat{\beta}_{\psi_k}}$, $k = 1, \dots, K$.
 5. Volver al Paso 1 tomando $\psi_k^{(1)}$, $k = 1, \dots, K$ como iterante inicial.
-

Muggeo (2003) sugiere no emplear un número demasiado elevado de puntos de cambio en el modelo porque éstos pasarían a tener un significado menos claro. Concretamente, en dicho artículo se recomienda no usar más de tres y utilizar funciones de suavizado en caso de que se vean necesarios más puntos.

Como se puede observar, no se ha comentado nada sobre el método de Hinkley (1969) en este contexto. Esto se debe a que la función que se maximiza está basada específicamente en la suposición de que sólo hay un punto de cambio y su generalización no parece inmediata. Una posible solución sería la de hallar inicialmente el primer punto de cambio con el método de Hinkley (1969), luego el siguiente con el mismo procedimiento y así sucesivamente. No obstante, lo ideal sería hallar todos los puntos a la vez puesto que pueden existir varios puntos que maximicen la verosimilitud del modelo pero que, sin embargo, al hacerlo por pasos no coincidan con los devueltos por este método.

A modo de resumen, los métodos de Hudson (1966) y Hinkley (1969) buscan el punto de cambio en una rejilla de valores determinada por la muestra, mientras que el método de Muggeo (2003) se basa en un algoritmo iterativo que recuerda al método de Newton. El resultado devuelto por el método de Hudson (1966) es el estimador de mínimos cuadrados, mientras que Hinkley (1969) y Muggeo (2003) aproximan el estimador de máxima verosimilitud. Al principio de este capítulo, se probó la consistencia del estimador de mínimos cuadrados, de modo que los otros dos también son consistentes⁸. No obstante, la insesgadez es asintótica, lo cual deja entrever que estos estimadores serán sesgados en muestras pequeñas. Todas estas cuestiones se abordan en el estudio de simulación, así como otras mucho más específicas.

⁸Como se dijo previamente, los métodos de máxima verosimilitud y mínimos cuadrados son equivalentes en este contexto.

Capítulo 3

Estudio de simulación

Una vez presentados los métodos de estimación del punto de cambio y la inferencia sobre el modelo de regresión segmentada, es necesario verificar que dichos métodos funcionan adecuadamente en la práctica con muestras finitas. Para ello, se presenta un estudio de simulación donde se generarán muestras de varios escenarios en los que se estimará el punto de cambio. En este sentido, es necesario también explicar que los estimadores se evaluarán mediante su sesgo, desviación típica y error cuadrático medio.

3.1. Escenarios de simulación

Como se ha dicho en la introducción de esta sección, antes de examinar los resultados es necesario conocer los modelos a partir de los que se generan los datos en los que se prueban los métodos. Para esto, hay que pensar unos escenarios que pongan a prueba los métodos cuando teóricamente funcionan bien, pero también otros en los que se incumpla alguna hipótesis del modelo para ver cómo de afectados se ven los procedimientos por este contratiempo. Por tanto, es necesario tener en cuenta las siguientes consideraciones:

- No se ha mencionado que los procedimientos de estimación funcionen mejor o peor en función de la localización del punto de cambio en el intervalo $[x_1, x_n]$. Puede parecer que si el punto de cambio está en un extremo los métodos deberían fallar más, pero será necesario verificarlo.
- Tampoco se ha dicho que los métodos vayan a detectar mejor un cambio brusco de pendiente que uno débil, aunque intuitivamente parece que será así. No obstante, esta afirmación tiene algo de subjetivo detrás, puesto que cada persona puede tener consideraciones diferentes sobre lo que es fuerte y débil.
- Durante el trabajo se ha supuesto que la varianza del error es la misma a ambos lados del punto de cambio. Sería interesante comprobar qué ocurre cuando esto no se da.
- Como en cualquier estudio de simulación, también se pretende ver cómo afecta el tamaño de la muestra al rendimiento de estos procedimientos. Ya que la mayoría de resultados presentados son asintóticos, se espera que afecte mucho, en particular en muestras de pequeño tamaño.
- Por último, los métodos deberían ser inconsistentes cuando el modelo subyacente que genera los datos no es de regresión segmentada. Para comprobar esto, se añade también un escenario con datos generados bajo un modelo de regresión lineal simple.

Por todo esto, se han planteado los siguientes escenarios:

Escenario 1: Punto de cambio en el medio del intervalo y cambio débil con misma varianza a ambos lados.

$$Y = \begin{cases} 2X + \varepsilon & \text{si } X \leq 5 \\ -5 + 3X + \varepsilon & \text{si } X > 5, \end{cases} \quad (3.1)$$

con $\text{Sop}(x) = [0, 10]$ y $\varepsilon \sim N(0, 1)$. Para la variable explicativa se han generado puntos equiespaciados entre 0 y ψ a un lado del punto de cambio y puntos equiespaciados entre ψ y 10 al otro lado para cada tamaño muestral. Así, por ejemplo, en el Escenario 1, X estaba compuesta por 15 puntos equiespaciados entre 0 y 5, y 25 puntos equiespaciados entre 5 y 10 para $n_1 = 15$ y $n_2 = 25$. Esto se ha hecho así en todos los escenarios, de modo que ya no se repetirá en el futuro.

Escenario 2: Punto de cambio en el medio del intervalo y cambio fuerte con misma varianza a ambos lados.

$$Y = \begin{cases} 2X + \varepsilon & \text{si } X \leq 5 \\ -25 + 7X + \varepsilon & \text{si } X > 5, \end{cases} \quad (3.2)$$

con $\text{Sop}(x) = [0, 10]$ y $\varepsilon \sim N(0, 1)$.

Escenario 3: Punto de cambio en un extremo del intervalo y cambio débil con misma varianza a ambos lados.

$$Y = \begin{cases} 2X + \varepsilon & \text{si } X \leq 2.5 \\ -2.5 + 3X + \varepsilon & \text{si } X > 2.5, \end{cases} \quad (3.3)$$

con $\text{Sop}(x) = [0, 10]$ y $\varepsilon \sim N(0, 1)$.

Escenario 4: Punto de cambio en un extremo del intervalo y cambio fuerte con misma varianza a ambos lados.

$$Y = \begin{cases} 2X + \varepsilon & \text{si } X \leq 2.5 \\ -12.5 + 7X + \varepsilon & \text{si } X > 2.5, \end{cases} \quad (3.4)$$

con $\text{Sop}(x) = [0, 10]$ y $\varepsilon \sim N(0, 1)$.

Escenario 5: Punto de cambio en el medio del intervalo y cambio débil con distinta varianza a ambos lados.

$$Y = \begin{cases} 2X + \varepsilon_a & \text{si } X \leq 5 \\ -5 + 3X + \varepsilon_b & \text{si } X > 5, \end{cases} \quad (3.5)$$

con $\text{Sop}(x) = [0, 10]$, $\varepsilon_a \sim N(0, 2)$ y $\varepsilon_b \sim N(0, 1)$. - **Escenario 6:** Punto de cambio en el medio del intervalo y cambio fuerte con distinta varianza a ambos lados.

$$Y = \begin{cases} 2X + \varepsilon_a & \text{si } X \leq 5 \\ -25 + 7X + \varepsilon_b & \text{si } X > 5, \end{cases} \quad (3.6)$$

con $\text{Sop}(x) = [0, 10]$, $\varepsilon_a \sim N(0, 2)$ y $\varepsilon_b \sim N(0, 1)$.

El hecho de que sea el primer tramo o el segundo el que tiene la varianza más alta no afectará al comportamiento de los estimadores porque el dominio de la variable explicativa queda dividido en dos partes iguales. Cuando esto no ocurre, es decir, si el punto de cambio está en el extremo, es necesario distinguir entre el caso en que el lado corto tenga la varianza más alta y el caso en que la varianza más alta está en el lado largo. Esto se recoge en los siguientes escenarios:

Escenario 7: Punto de cambio en un extremo del intervalo y cambio débil con mayor varianza en el lado corto.

$$Y = \begin{cases} 2X + \varepsilon_a & \text{si } X \leq 2.5 \\ -2.5 + 3X + \varepsilon_b & \text{si } X > 2.5, \end{cases} \quad (3.7)$$

con $\text{Sop}(x) = [0, 10]$, $\varepsilon_a \sim N(0, 2)$ y $\varepsilon_b \sim N(0, 1)$.

Escenario 8: Punto de cambio en un extremo del intervalo y cambio débil con mayor varianza en el lado largo.

$$Y = \begin{cases} 2X + \varepsilon_a & \text{si } X \leq 2.5 \\ -2.5 + 3X + \varepsilon_b & \text{si } X > 2.5, \end{cases} \quad (3.8)$$

con $\text{Sop}(x) = [0, 10]$, $\varepsilon_a \sim N(0, 1)$ y $\varepsilon_b \sim N(0, 2)$.

Escenario 9: Punto de cambio en un extremo del intervalo y cambio fuerte con mayor varianza en el lado corto.

$$Y = \begin{cases} 2X + \varepsilon_a & \text{si } X \leq 2.5 \\ -12.5 + 7X + \varepsilon_b & \text{si } X > 2.5, \end{cases} \quad (3.9)$$

con $\text{Sop}(x) = [0, 10]$, $\varepsilon_a \sim N(0, 2)$ y $\varepsilon_b \sim N(0, 1)$.

Escenario 10: Punto de cambio en un extremo del intervalo y cambio fuerte con mayor varianza en el lado largo.

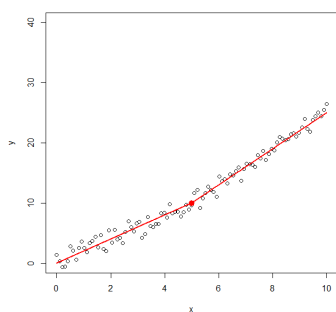
$$Y = \begin{cases} 2X + \varepsilon_a & \text{si } X \leq 2.5 \\ -12.5 + 7X + \varepsilon_b & \text{si } X > 2.5, \end{cases} \quad (3.10)$$

con $\text{Sop}(x) = [0, 10]$, $\varepsilon_a \sim N(0, 1)$ y $\varepsilon_b \sim N(0, 2)$.

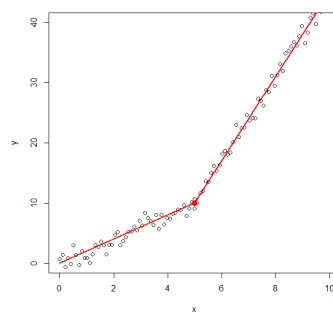
Escenario 11: No existe punto de cambio.

$$Y = 2X + \varepsilon \quad (3.11)$$

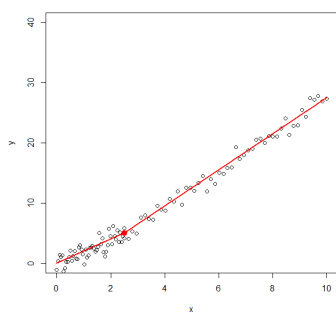
con $\text{Sop}(x) = [0, 10]$, $\varepsilon \sim N(0, 1)$.



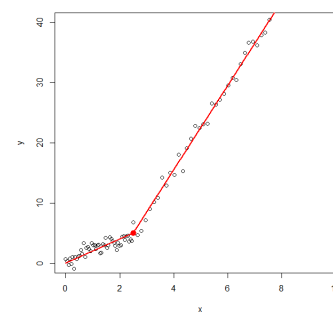
(a) Escenario 1



(b) Escenario 2

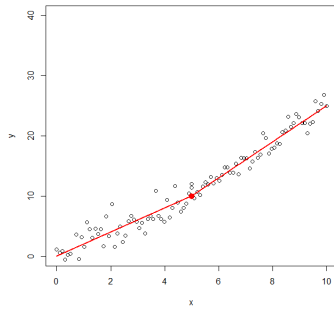


(c) Escenario 3

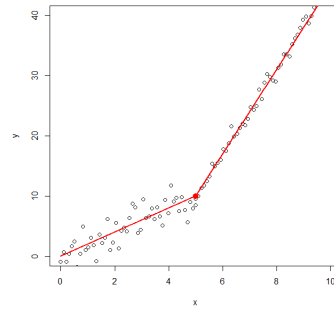


(d) Escenario 4

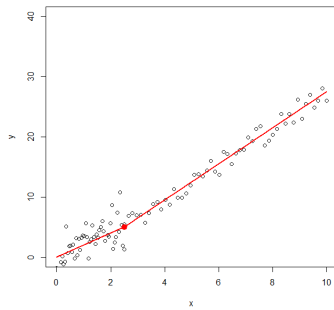
Figura 3.1: Gráfica de la verdadera función de regresión sobre la representación de una muestra generada en los escenarios homocedásticos. (a) $\psi = 5$ y $\beta_\psi = 1$, (b) $\psi = 5$ y $\beta_\psi = 5$, (c) $\psi = 2.5$ y $\beta_\psi = 1$, (d) $\psi = 2.5$ y $\beta_\psi = 5$.



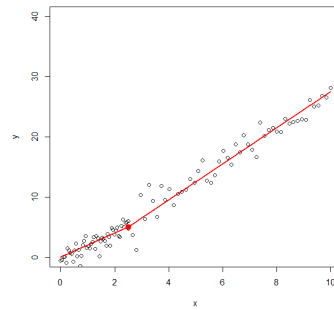
(a) Escenario 5



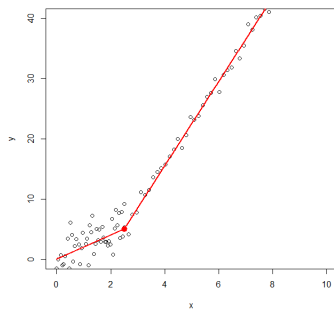
(b) Escenario 6



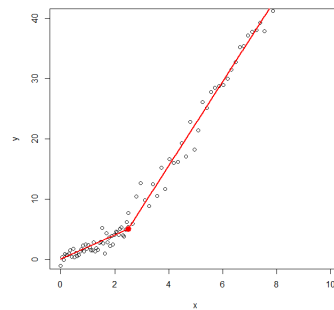
(c) Escenario 7



(d) Escenario 8



(e) Escenario 9



(f) Escenario 10

Figura 3.2: Gráfica de la verdadera función de regresión sobre la representación de una muestra generada en los escenarios heterocedásticos. (a) $\psi = 5$, $\sigma_a = 2$ y $\beta_\psi = 1$, (b) $\psi = 5$, $\sigma_a = 2$ y $\beta_\psi = 5$, (c) $\psi = 2.5$, $\sigma_a = 2$ y $\beta_\psi = 1$, (d) $\psi = 2.5$, $\sigma_b = 2$ y $\beta_\psi = 1$, (e) $\psi = 2.5$, $\sigma_a = 2$ y $\beta_\psi = 5$, (f) $\psi = 2.5$, $\sigma_b = 2$ y $\beta_\psi = 5$.

Todos estos escenarios de los que se han simulado muestras aparecen representados en las Figuras 3.1 y 3.2. Cada una de las 10 gráficas que aparecen en ellas contiene una muestra generada por ese modelo y la verdadera función de regresión pintada en rojo. Se puede ver que el cambio de pendiente es casi imperceptible a la vista en aquellos escenarios donde $\beta_\psi = 1$.

Para cada escenario se han generado 1000 muestras en las que se aplicaban los tres métodos de estimación. Es importante notar que las muestras en las que se aplica cada método deben ser las mismas, pues no tendría sentido comparar el funcionamiento de estimadores en muestras diferentes. El método de Muggeo precisa además una semilla inicial para arrancar el algoritmo, de modo que para generarla se ha seguido el mismo criterio que utiliza Muggeo (2003). Dicho autor sostiene que no hay razón para empezar el algoritmo muy lejos del punto de cambio y, en cada paso, simula aleatoriamente la semilla inicial de una distribución Uniforme definida en un intervalo que contiene a ψ . En este caso, en cada escenario, esta semilla se generó aleatoriamente a partir de la distribución $U(\psi - 0.5, \psi + 0.5)$.

Hinkley (1969) y Muggeo (2003) incorporan estudios de simulación con los que prueban sus métodos en la práctica, mientras que Hudson (1966) no presenta un estudio de simulación en su artículo. Sin embargo, no hay artículos en la literatura que comparen varios métodos de estimación en el contexto de Regresión Segmentada. Ya que Muggeo basa su estudio de simulación en un modelo de regresión de Poisson con diseño aleatorio, la validez de los resultados del presente estudio tendrá que ser comparada con los presentados por Hinkley (1969). Hudson (1966) se basa en un criterio de optimización similar al de Hinkley, de modo que lo razonable sería que devolviese resultados también similares.

Es preciso explicar también la notación empleada en las tablas. En ellas, n_1 y n_2 denotan la cantidad de observaciones a la izquierda y a la derecha del punto de cambio, respectivamente. Para cada muestra generada, cada método devuelve un estimador del punto de cambio (denotado como $\hat{\psi}$). Lo que se pretende es evaluar el sesgo y la varianza de los estimadores para comprobar si se cumplen las propiedades enunciadas en capítulos anteriores. Por tanto, considerando las 1000 muestras en su conjunto, se tendrán 1000 valores de $\hat{\psi}$. La media de estos valores se denota por $\bar{\psi}$ y es la medida utilizada para cuantificar el sesgo, ya que se compara con el verdadero valor de ψ . Este valor aquí es conocido, puesto que se conoce el modelo generador de los datos. Si el estimador es insesgado, el valor de $\bar{\psi}$ debería ser muy cercano a ψ . Por último, para comprobar si el estimador es consistente, se debe analizar la convergencia hacia 0 del **Error Cuadrático Medio**, dado por

$$ECM(\hat{\psi}) = (E(\hat{\psi}) - \psi)^2 + \text{Var}(\hat{\psi}),$$

La anterior expresión no se tiene en la práctica, ya que no se conoce $E(\hat{\psi})$. Es por ello que se ha intentado aproximar el error cuadrático medio mediante

$$\widehat{ECM}(\hat{\psi}) = (\bar{\psi} - \psi)^2 + \text{Var}(\hat{\psi}).$$

Por tanto, se han incluido en las tablas tres columnas para cada estimador de ψ con la media ($\bar{\psi}$), la desviación típica ($SD(\hat{\psi})$) y la estimación del error cuadrático medio ($\widehat{ECM}(\hat{\psi})$) de la muestra de 1000 valores de $\hat{\psi}$.

Muggeo (2008) sostiene que procedimientos como el de Hinkley (1971) (y, consecuentemente, también el de Hudson (1966)) tienen un gran inconveniente en la carga computacional. Es por esto que, además de incluir los resultados de estimación, se incluye también una tabla con los tiempos de computación en cada escenario.

La programación de los métodos de Hudson y Hinkley es de elaboración propia, mientras que para el método de Muggeo se ha utilizado la función del paquete de R `segmented`. Más detalles sobre el paquete `segmented` pueden ser consultados en Muggeo (2008). Es necesario decir que la programación de los dos primeros métodos no es óptima puesto que se utiliza la fuerza bruta. Es decir, en cada paso del algoritmo se obvian las comprobaciones teóricas que permitirían ahorrar cálculos y, por tanto, los tiempos de computación son levemente superiores a lo que debieran.

Por último, varias tablas contienen un número entre paréntesis en algunas celdas relacionadas con el método de Muggeo. Eso quiere decir que la función del paquete de R no ha convergido en tantos casos

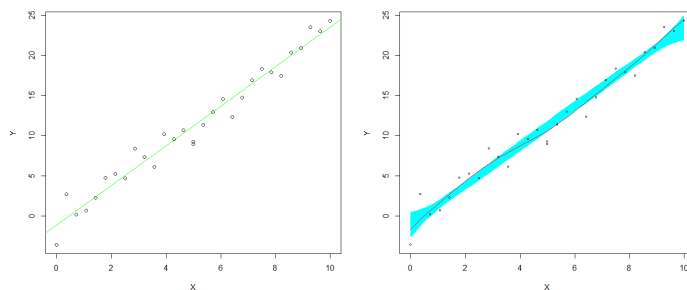


Figura 3.3: Muestra donde el algoritmo de Muggee falla con el ajuste función de regresión lineal simple en verde (izqda.) y bandas de referencia en azul junto con un ajuste de regresión no paramétrica en rojo (drcha.).

como aparece en el número entre paréntesis. El motivo es que, para alguna muestra, el último iterante del Algoritmo de Muggee es un punto del intervalo (x_1, x_2) , de modo que deja $n - 1$ observaciones a la izquierda y una a la derecha¹. En estos casos, como ya se indicó en el capítulo relativo a la estimación, el producto de la matriz de diseño del modelo y su traspuesta no es inversible, luego no se puede ajustar un modelo de regresión en el siguiente paso.

Para ilustrar este tipo de casos, tomemos como ejemplo la muestra representada en la Figura 3.3. En ella, se ajustó y representó un modelo de regresión lineal simple sin punto de cambio (en verde). La gráfica arroja bastantes dudas acerca de la existencia de un punto de cambio, ya que parece que el modelo sin él se ajusta bastante bien. Una forma de comprobarlo es contrastar si se acepta la hipótesis de linealidad en base a los datos de la muestra, aunque también se podrían representar las bandas de referencia para el verdadero valor de la función de regresión en cada punto (Figura 3.3, drcha.). El comando `sm.regression` del paquete `sm` en `R` permite hacer ambas cosas, de modo que se ha empleado para este análisis. En esta ocasión, los dos procedimientos concluyen que no hay evidencias suficientes para rechazar la linealidad del modelo, pues el p-valor es de 0.429 y las bandas de referencia contienen a una recta. Puesto que no hay evidencias para rechazar la linealidad, el método tiene problemas para encontrar el punto de cambio y, finalmente, acaba no convergiendo.

3.2. Estimación del punto de cambio

Una vez introducidos los objetivos del estudio de simulación, se procederá a comentar los resultados, primero para el caso homocedástico, posteriormente para el caso heterocedástico y, por último, se recogerán unas breves conclusiones generales.

Caso homocedástico (Escenarios 1 a 4). Los resultados se corresponden con las Tablas 3.2 a 3.5, mientras que los tiempos de ejecución aparecen en la Tabla 3.14. Se puede apreciar que los resultados obtenidos en la práctica por los tres métodos muestran que son asintóticamente insesgados y consistentes, ya que la media de los estimadores del punto de cambio converge al verdadero valor en todos los casos y su varianza error cuadrático medio converge a 0 al aumentar el tamaño de muestra. Comparando los resultados de las Tablas 3.2-3.4 y 3.3-3.5, donde se mantiene el cambio pendiente pero cambia el verdadero valor de ψ , se puede ver que el alejamiento de ψ del centro del intervalo produce estimadores sesgados en muestras pequeñas, aunque el efecto es mucho mayor si β_ψ es pequeño. Del mismo modo, comparando las Tablas 3.2-3.3 y 3.4-3.5, donde se fija el punto de cambio y se estudia el efecto de la pendiente, los resultados muestran que los cambios bruscos provocan una gran disminución en la varianza del estimador.

¹También puede ocurrir que el iterante caiga en el intervalo (x_{n-1}, x_n) y el problema es el mismo.

Además, no parece haber un método que destaque sobre los otros en términos de sesgo o varianza. No obstante, la Tabla 3.4 refleja que el Escenario 3 es aquél en el que los resultados son peores, ya que el sesgo y la varianza son los más altos de los 4 escenarios considerados para los tres métodos. Además, la mayoría de casos en los escenarios homocedásticos donde el Algoritmo de Muggeo (2003) no converge se dan en este escenario.

Cabría preguntarse aquí si los métodos serían capaces de detectar cambios de pendiente más pequeños. Hinkley (1969) analizó varios cambios de pendiente inferiores a 1, así que sería bueno replicar algún caso y comprobar que las conclusiones que se sacan son similares.

Por ejemplo, fijemos nuestra atención en el caso $\beta_\psi = 0.4$ que propone Hinkley (1969). En su artículo, él toma $x_i = i$, $i \in \{1, \dots, 50\}$, $\psi = 15.5$ y simula 500 muestras. Ya que no especifica el valor de los parámetros α_1 y β_1 que, recordemos, eran el intercepto y la pendiente antes de ψ , aquí se ha tomado $\alpha_1 = 0$ y $\beta_1 = 2$. La razón para adoptar estos valores se fundamenta en que éstos son los valores que se han considerado en el resto del estudio y así las conclusiones que aquí se obtengan podrán ir alineadas con el resto de resultados.

Los resultados de la réplica se muestran en la Tabla 3.1. En general, parece que los métodos de Hudson y Hinkley devuelven unas conclusiones similares a las que se sacan del artículo de Hinkley (1969): el sesgo de los estimadores es reducido pero a costa de presentar una gran varianza. De esto se deduce que los estimadores no son especialmente fiables en casos como éste. Es preciso indicar, también, que no el algoritmo de Muggeo no converge en este caso y, por tanto, no se ha incluido en la tabla.

Método	$\bar{\psi}$	$\text{Var}(\hat{\psi})$	\widehat{ECM}
Hudson (1966)	15.6434	2.1346	20.7836
Hinkley (1969)	15.5823	2.1092	19.7997
Artículo de Hinkley (1969)	15.6400	1.9468	14.3837

Tabla 3.1: Comparación de los resultados obtenidos con la programación de los métodos de este trabajo y los del artículo de Hinkley (1969) para 500 muestras de tamaño $n = 50$ con $\psi = 15.5$.

Caso heterocedástico (Escenarios 5 a 10). Aquí, lo que se espera es que los estimadores no funcionen bien, especialmente en muestras pequeñas. Además, lo que se pretende con esta simulación, (y por ese motivo se han incluido muestras con distintos tamaños y varianzas a ambos lados), es comprobar si cambios de tamaño o varianza se comportan igual en el lado corto o lado largo.

Cuando el punto de cambio está en el medio, la heterocedasticidad empeora manifiestamente los resultados en términos de sesgo y varianza, como puede observarse ojeando las Tablas 3.6 y 3.7. Aún así, los resultados de ambas tablas sugieren que todos los estimadores son consistentes. Nuevamente, el Algoritmo de Muggeo (2003) tiene problemas para converger en las mismas situaciones que en el caso homocedástico: muestras pequeñas y cambio de pendiente débil.

Si además de haber heterocedasticidad el punto de cambio está en un extremo (Tablas 3.8 a 3.11) los métodos funcionan especialmente mal en muestras pequeñas. El ejemplo más claro es el Escenario 8 (Tabla 3.10), aunque los resultados también sugieren que los problemas de sesgo y varianza que tienen los estimadores en estos contextos se solucionan aumentando el tamaño de muestra.

No existencia de punto de cambio. Como cabía esperar, aquí los métodos de estimación se comportan de manera errática. La Tabla 3.12 muestra que, según aumenta el tamaño muestral, la media de las estimaciones con los métodos de Hudson y Hinkley tiende a irse hacia el centro del intervalo de definición. Esto tiene una explicación muy clara: al no haber punto de cambio, el punto con el que se minimice la suma cuadrática de residuos (o se maximice la verosimilitud) va a depender

únicamente de los valores que tome el error en esa muestra. Como la generación del error es aleatoria, también lo será la estimación que realicen estos métodos y, por lo tanto, los valores que salgan van a seguir una distribución Uniforme (en este caso, $U(0, 10)$). La media de una distribución uniforme es el centro del intervalo en el que está definida y por eso salen estos resultados. Aquí es donde el método de Muggeo (2003) presenta más problemas de convergencia, puesto que, por ejemplo, para muestras de tamaño $n = 30$ no converge en algo más del 20% de los casos.

Como se ha dicho antes, Muggeo (2008) cita como un inconveniente de los métodos de Hudson (1966) y Hinkley (1969) su pesada carga computacional. Ya que los escenarios de mayor interés en la práctica son los homocedásticos, sólo se han incluido las tablas con los tiempos relativos a estos casos (Tablas 3.13 y 3.14)². Los resultados presentados en las mismas concuerdan con sus afirmaciones, ya que en prácticamente todos los casos el método de Muggeo (2003) es más rápido que los otros dos. Únicamente en muestras de pequeño tamaño se puede dar el caso contrario. Esto se debe a que la rejilla de valores en la que los métodos de Hudson (1966) y Hinkley (1969) aumenta con el tamaño de la muestra, pero el método de Muggeo (2003) no realiza necesariamente más cálculos en muestras más grandes. Además, se puede establecer un orden y concluir que el método de Muggeo es (2003) manifiestamente más rápido que el de Hinkley (1969), que sería el segundo, y el de Hudson (1966), que es el más lento. Este orden también tiene su explicación. El de Muggeo es el más rápido porque generalmente sólo suele necesitar ajustar unos pocos modelos de regresión. El de Hinkley como mínimo tiene que calcular $n - 3$ valores del estadístico $Z_{\tau}^2(\cdot)$, para los cuales es necesario calcular $2(n - 3)$ modelos de regresión. A estos se añadirían todos aquellos cálculos cuando el candidato inicial a punto de cambio caiga fuera del intervalo $(x_{\tau}, x_{\tau+1})$. El método de Hudson (1966) es el más lento porque tiene que calcular los mismos modelos de regresión que el de Hinkley y realiza un mayor número de operaciones cuando el candidato inicial a punto de cambio cae fuera de $(x_{\tau}, x_{\tau+1})$.

Si bien el método de Muggeo (2003) tiene grandes ventajas computacionales, como muestran las Tablas 3.13 y 3.14, también tiene la desventaja de que a veces no converge. En el caso homocedástico, esto ocurre cuando el punto de cambio está en un extremo y la diferencia de pendientes es pequeña, como se puede ver en la Tabla 3.4. En el contexto heterocedástico, la no convergencia se da en los escenarios 5, 7 y 8, que son aquéllos en que la variabilidad en el lado corto es mayor, la diferencia de pendientes es pequeña o ambos. Esto se muestra en las Tablas 3.6, 3.8 y 3.9. Por tanto, los resultados sugieren que lo ideal para que este método converja es que el punto de cambio no esté muy lejos del centro del intervalo dominio de la variable explicativa y que la diferencia de pendientes sea elevada (en especial esto último), lo cual parece razonable. Aparentemente, influye en menor medida que la variabilidad sea constante en todo el dominio.

A modo de resumen, se puede decir que, en general, cuanto más evidencia haya en los datos, mejor será la estimación que los tres métodos hagan del punto de cambio tanto en sesgo como en varianza. El método propuesto por Muggeo (2003) es el más rápido pero con el inconveniente de que es el que más sufre cuando no está claro si hay punto de cambio. En vista de los resultados, parece más recomendable usar los métodos de Hudson (1966) y Hinkley (1969) en muestras de tamaño pequeño o moderado y el método de Muggeo (2003) para muestras grandes en base a la diferencia de tiempos de computación.

²Las conclusiones obtenidas para estos escenarios son extrapolables al resto.

(n_1, n_2)	Hudson			Hinkley			Muggeo		
	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$
(15, 15)	5.0024	1.1583	1.3417	5.0127	1.1917	1.4204	4.9837	1.1613	1.3488(2)
(25, 25)	5.0156	0.7866	0.6190	5.0186	0.8100	0.6565	5.0361	0.8174	0.6695
(50, 50)	5.0099	0.5262	0.2770	5.0067	0.5299	0.2808	5.0127	0.5695	0.3245
(250, 250)	5.0077	0.1974	0.0390	5.0077	0.1979	0.0392	5.0030	0.2039	0.0416
(15, 25)	4.9981	0.9893	0.9787	4.9937	1.0140	1.0282	5.0164	0.9128	0.8334
(25, 50)	4.9790	0.6615	0.4380	4.9666	0.6695	0.4494	5.0191	0.6595	0.4353
(50, 250)	5.0153	0.3507	0.1232	5.0154	0.3531	0.1249	4.9850	0.3446	0.1190

Tabla 3.2: Media, desviación típica y error cuadrático medio en el Escenario 1 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Tamaños de muestra a ambos lados del punto de cambio $\psi = 5$ dados por (n_1, n_2) . El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

(n_1, n_2)	Hudson			Hinkley			Muggeo		
	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$
(15, 15)	4.9898	0.1428	0.0205	5.0040	0.1825	0.0333	5.0023	0.1442	0.0208
(25, 25)	5.0053	0.1186	0.0141	5.0063	0.1410	0.0199	5.0057	0.1568	0.0246
(50, 50)	5.0008	0.0836	0.0070	4.9996	0.0890	0.0079	4.9978	0.0845	0.0071
(250, 250)	5.0019	0.0354	0.0013	5.0020	0.0358	0.0013	4.9969	0.0344	0.0012
(15, 25)	5.0006	0.1277	0.0163	4.9942	0.1576	0.0249	5.0016	0.1314	0.0173
(25, 50)	5.0023	0.1033	0.0107	4.9995	0.1104	0.0122	4.9983	0.1008	0.0102
(50, 250)	4.9989	0.0643	0.0041	4.9981	0.0667	0.0045	4.9995	0.0622	0.0039

Tabla 3.3: Media, desviación típica y error cuadrático medio en el Escenario 2 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Tamaños de muestra a ambos lados del punto de cambio $\psi = 5$ dados por (n_1, n_2) . El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

(n_1, n_2)	Hudson			Hinkley			Muggeo		
	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$
(15, 15)	2.6899	1.5115	2.3205	2.7079	1.5478	2.4390	2.7288	1.5129	2.3412 (21)
(25, 25)	2.6327	1.2048	1.4691	2.6393	1.2099	1.4833	2.6397	1.0602	1.1436 (3)
(50, 50)	2.5088	0.6258	0.3917	2.5047	0.6235	0.3888	2.6077	0.8111	0.6694
(250, 250)	2.5040	0.2046	0.0419	2.5047	0.2052	0.0421	2.5085	0.2179	0.0476
(15, 25)	2.7874	1.6006	2.6444	2.7994	1.6186	2.7096	2.6259	1.0640	1.1480 (14)
(25, 15)	2.6637	1.3606	1.8780	2.6873	1.4231	2.0602	2.5913	1.2154	1.4857 (3)
(25, 50)	2.6015	0.9843	0.9791	2.6091	1.0176	1.0475	2.5517	0.9061	0.8236
(50, 25)	2.5469	0.6945	0.4845	2.5539	0.6955	0.4866	2.5495	0.7853	0.6191
(50, 250)	2.5354	0.3880	0.1518	2.5539	0.3883	0.1537	2.5495	0.4127	0.1728
(250, 50)	2.5017	0.3112	0.0969	2.5028	0.3118	0.0973	2.4804	0.3273	0.1075

Tabla 3.4: Media, desviación típica y error cuadrático medio en el Escenario 3 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Tamaños de muestra a ambos lados del punto de cambio $\psi = 2.5$ dados por (n_1, n_2) . El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

(n_1, n_2)	Hudson			Hinkley			Muggeo		
	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$
(15, 15)	2.4978	0.1553	0.0241	2.5215	0.2020	0.0413	2.4913	0.1508	0.0228
(25, 25)	2.5054	0.1201	0.0144	2.5062	0.1297	0.0169	2.5052	0.1131	0.0128
(50, 50)	2.4968	0.0854	0.0073	2.5021	0.0972	0.0095	2.5034	0.0844	0.0071
(250, 250)	2.5016	0.0379	0.0014	2.5014	0.0384	0.0015	2.4991	0.0363	0.0013
(15, 25)	2.4978	0.1364	0.0186	2.4975	0.1430	0.0204	2.4943	0.1344	0.0181
(25, 15)	2.5045	0.1322	0.0175	2.5244	0.1770	0.0319	2.4959	0.1575	0.0248
(25, 50)	2.4998	0.1058	0.0112	2.5031	0.1165	0.0136	2.4961	0.1032	0.0107
(50, 25)	2.4981	2.4981	6.2406	2.4980	2.4980	6.2398	2.4959	2.4959	6.2295
(50, 250)	2.5005	2.5005	6.2525	2.5003	2.5003	6.2514	2.5024	2.5024	6.2621
(250, 50)	2.5002	2.5002	6.2508	2.5014	2.5014	6.2572	2.4917	2.4917	6.2088

Tabla 3.5: Media, desviación típica y error cuadrático medio en el Escenario 4 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Tamaños de muestra a ambos lados del punto de cambio $\psi = 5$ dados por (n_1, n_2) . El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

(n_1, n_2)	Hudson			Hinkley			Muggeo		
	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$
(15, 15)	4.2838	1.8914	4.0903	4.3040	1.8830	4.0300	4.5398	1.6708	3.0032 (19)
(25, 25)	4.4939	1.5798	2.7520	4.4911	1.5768	2.7452	4.6527	1.5476	2.5157 (2)
(50, 50)	4.8256	1.0588	1.1516	4.8246	1.0554	1.1446	4.8438	0.9794	1.0038
(250, 250)	4.9667	0.3407	0.1172	4.9662	0.3402	0.1169	4.9561	0.3614	0.3633
(15, 25)	4.5033	1.7966	3.4744	4.5188	1.8098	3.5067	4.6452	1.5425	2.5052 (20)
(25, 15)	4.3384	1.8069	3.7025	4.3459	1.8093	3.7014	4.4988	1.5857	2.7658 (8)
(25, 50)	4.6765	1.2927	1.7759	4.6721	1.2973	1.7904	4.7285	1.1731	1.4499 (4)
(50, 25)	4.5985	1.2748	1.4360	4.5974	1.2783	1.4404	5.0019	0.1437	0.1437
(50, 250)	4.8887	0.6003	0.3728	4.8837	0.6061	0.3809	5.0409	0.4794	0.4811
(250, 50)	4.8914	0.5520	0.3165	4.8922	0.5539	0.3184	4.9856	0.2604	0.2606

Tabla 3.6: Media, desviación típica y error cuadrático medio en el Escenario 5 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Tamaños de muestra a ambos lados del punto de cambio $\psi = 5$ dados por (n_1, n_2) . El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

(n_1, n_2)	Hudson			Hinkley			Muggeo		
	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$
(15, 15)	4.9594	0.2467	0.0625	4.9545	0.2799	0.0804	4.9850	0.2317	0.0539
(25, 25)	4.9893	0.1952	0.0382	4.9876	0.2181	0.0477	4.9863	0.2492	0.0623
(50, 50)	4.9932	0.1396	0.0195	4.9900	0.1439	0.0208	4.9938	0.1377	0.0190
(250, 250)	5.0017	0.0565	0.0032	5.0019	0.0575	0.0033	4.9960	0.0593	0.0035
(15, 25)	4.9817	0.2221	0.0497	4.9594	0.2558	0.0671	4.9786	0.2254	0.0513
(25, 15)	4.9915	0.2037	0.0416	4.9884	0.2358	0.0557	4.9788	0.2122	0.0455
(25, 50)	4.9823	0.1800	0.0327	4.9776	0.1920	0.0374	4.9899	0.1697	0.0289
(50, 350)	4.9924	0.1624	0.0264	4.9921	0.1624	0.0264	0.0000	0.0000	25.0000
(50, 250)	4.9988	0.1080	0.0117	4.9949	0.1138	0.0130	4.9921	0.1148	0.0132
(250, 50)	4.9905	0.0898	0.0082	4.9908	0.0923	0.0086	4.9832	0.0955	0.0094

Tabla 3.7: Media, desviación típica y error cuadrático medio en el Escenario 6 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Tamaños de muestra a ambos lados del punto de cambio $\psi = 5$ dados por (n_1, n_2) . El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

(n_1, n_2)	Hudson			Hinkley			Muggeo		
	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$
(15, 15)	2.1995	1.8057	3.3508	2.2493	1.8578	3.5143	2.4442	1.7785	3.1662 (71)
(25, 25)	2.2694	1.6096	2.6439	2.2936	1.6517	2.7707	2.5155	1.6373	2.6810 (32)
(50, 50)	2.2896	1.2383	1.5776	2.2847	1.2192	1.5327	2.4084	1.0249	1.0588 (8)
(250, 250)	2.4786	0.3771	0.1427	2.4787	0.3775	0.1429	2.5479	0.3792	0.1461
(15, 25)	2.3189	1.9508	3.8385	2.3427	1.9637	3.8807	2.5028	1.6754	2.8072 (63)
(25, 15)	2.1260	1.6968	3.0189	2.1531	1.7442	3.1626	2.4496	1.6529	2.7348 (38)
(25, 50)	2.3188	1.4567	2.1549	2.3299	1.4914	2.2533	2.4034	1.3662	1.8759 (15)
(50, 25)	2.2579	1.2723	1.6774	2.2595	1.2774	1.6897	2.3999	1.3379	1.7999 (12)
(50, 250)	2.3754	0.7341	0.5544	2.3743	0.7332	0.5533	2.4147	0.7014	0.4992
(250, 50)	2.3871	0.5383	0.3025	2.3875	0.5412	0.3056	2.3746	0.5263	0.2927

Tabla 3.8: Media, desviación típica y error cuadrático medio en el Escenario 7 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Tamaños de muestra a ambos lados del punto de cambio $\psi = 2.5$ dados por (n_1, n_2) . El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

(n_1, n_2)	Hudson			Hinkley			Muggeo		
	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$
(15, 15)	4.0573	2.6210	9.2952	4.3402	2.7404	10.8962	3.8505	2.4215	7.6879 (65)
(25, 25)	3.7121	2.4283	7.3656	3.8244	2.5144	8.0763	3.4992	2.1920	5.8034 (31)
(50, 50)	3.1268	1.8003	3.6341	3.1849	1.8676	3.9572	3.0719	1.6867	3.1722 (6)
(250, 250)	2.5453	0.3535	0.1270	2.5452	0.3525	0.1263	2.5118	0.3429	0.1177
(15, 25)	4.1859	2.7559	10.4375	4.3692	2.8624	11.6874	3.9362	2.5288	8.4575 (53)
(25, 15)	3.7309	2.4107	7.3267	3.9335	2.5789	8.7058	3.5510	2.2066	5.9740 (42)
(25, 50)	3.6535	2.3277	6.7490	3.7311	2.4159	7.3522	3.6125	2.2086	6.1157 (9)
(50, 25)	3.2165	1.8389	3.8949	3.2794	1.9193	4.2909	3.1750	1.7775	3.6150 (9)
(50, 250)	2.8661	1.0876	1.3168	2.8734	1.1109	1.3736	2.7947	0.9514	0.9920
(250, 25)	2.6261	0.6744	0.4707	2.6274	0.6724	0.4683	2.6134	0.6375	0.4193

Tabla 3.9: Media, desviación típica y error cuadrático medio en el Escenario 8 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Tamaños de muestra a ambos lados del punto de cambio $\psi = 2.5$ dados por (n_1, n_2) . El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

(n_1, n_2)	Hudson			Hinkley			Muggeo		
	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	$SD(\hat{\psi})$	$\widehat{ECM}(\hat{\psi})$
(15, 15)	2.4648	0.2748	0.0768	2.4792	0.2994	0.0901	2.4641	0.2634	0.0707
(25, 25)	2.4908	0.2023	0.0410	2.4934	0.2163	0.0468	2.4814	0.2166	0.0473
(50, 50)	2.4909	0.1435	0.0207	2.4917	0.1490	0.0223	2.4955	0.1461	0.0214
(250, 250)	2.5015	0.0633	0.0040	2.5014	0.0643	0.0041	2.4960	0.0618	0.0038
(15, 25)	2.4761	0.2497	0.0629	2.4742	0.2555	0.0659	2.4624	0.2596	0.0688
(25, 15)	2.4804	0.2154	0.0468	2.4878	0.2369	0.0563	2.4688	0.2498	0.0634
(25, 50)	2.4894	0.1850	0.0343	2.4822	0.1936	0.0378	2.4952	0.1935	0.0375
(50, 25)	2.4751	0.1702	0.0296	2.4760	0.1771	0.0319	2.4742	0.1652	0.0279
(50, 250)	2.4950	0.1119	0.0126	2.4947	0.1138	0.0130	2.4924	0.1171	0.0138
(250, 50)	2.4870	0.0939	0.0090	2.4884	0.0972	0.0096	2.4787	0.1010	0.0107

Tabla 3.10: Media, desviación típica y error cuadrático medio en el Escenario 9 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Tamaños de muestra a ambos lados del punto de cambio $\psi = 2.5$ dados por (n_1, n_2) . El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

(n_1, n_2)	Hudson			Hinkley			Muggeo		
	$\bar{\psi}$	SD($\hat{\psi}$)	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	SD($\hat{\psi}$)	$\widehat{ECM}(\hat{\psi})$	$\bar{\psi}$	SD($\hat{\psi}$)	$\widehat{ECM}(\hat{\psi})$
(15, 15)	2.5286	0.2607	0.0688	2.5569	0.2936	0.2968	2.5172	0.2371	0.2374
(25, 25)	2.5205	0.1914	0.0370	2.5238	0.2053	0.2059	2.5094	0.1757	0.1758
(50, 50)	2.5025	0.1319	0.0174	2.5107	0.1403	0.1404	2.5099	0.1386	0.1387
(250, 250)	2.5024	0.0789	0.0062	2.5014	0.0643	0.0643	2.5023	0.0582	0.0582
(15, 25)	2.5187	0.2177	0.0478	2.5226	0.2306	0.2311	2.5768	0.1972	0.2031
(25, 15)	2.5285	0.2153	0.0472	2.5561	0.2482	0.2513	2.5303	0.2298	0.2307
(25, 50)	2.5145	0.1638	0.0270	2.5177	0.1735	0.1739	2.5150	0.1602	0.1604
(50, 25)	2.5129	0.1644	0.0272	2.5137	0.1677	0.1679	2.5246	0.1876	0.1882
(50, 250)	2.5085	0.0889	0.0080	2.5089	0.0919	0.0920	2.5022	0.0825	0.0825
(250, 50)	2.5130	0.1046	0.0111	2.5165	0.1094	0.1097	2.5075	0.1082	0.1083

Tabla 3.11: Media, desviación típica y error cuadrático medio en el Escenario 10 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Tamaños de muestra a ambos lados del punto de cambio $\psi = 2.5$ dados por (n_1, n_2) . El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

n	Hudson		Hinkley		Muggeo	
	$\bar{\psi}$	SD($\hat{\psi}$)	$\bar{\psi}$	SD($\hat{\psi}$)	$\bar{\psi}$	SD($\hat{\psi}$)
30	3.8693	3.3798	4.2027	3.4511	5.0631	3.2165 (208)
50	4.1241	3.6244	4.4572	3.7151	4.9816	3.4489 (151)
100	3.9594	3.8075	4.2801	3.9151	5.1383	3.7094 (98)
500	4.3092	4.0968	4.5114	4.1546	5.1971	3.9101 (48)

Tabla 3.12: Media y desviación típica el Escenario 11 de los estimadores obtenidos a través de los tres métodos presentados con 1000 simulaciones. Muestras de tamaño n generadas a través de un modelo de regresión lineal simple. El número entre paréntesis indica la cantidad de casos en los que el algoritmo de Muggeo no converge.

	(15, 15)	(25, 25)	(50, 50)	(250, 250)	(15, 25)	(25, 50)	(50, 250)
Hudson	81.32	146.39	485.83	2490.39	107.49	221.74	1050.84
Hinkley	45.08	78.42	162.40	1006.34	59.31	118.47	550.10
Muggeo	112.00	127.62	163.8	213.28	175.92	128.48	190.67
	(15, 15)	(25, 25)	(50, 50)	(250, 250)	(15, 25)	(25, 50)	(50, 250)
Hudson	79.64	142.80	308.37	1954.58	111.35	224.42	1056.97
Hinkley	43.36	76.34	161.83	1006.81	59.54	118.01	549.68
Muggeo	93.55	97.63	108.52	238.25	102.44	127.58	130.06

Tabla 3.13: Tiempos de ejecución en segundos para las simulaciones de los escenarios 1 y 2.

Esc. 3	(15, 15)	(25, 25)	(50, 50)	(250, 250)	(15, 25)	(25, 15)	(25, 50)	(50, 25)	(50, 250)	(250, 50)
Hudson	81.36	141.47	309.19	1959.09	108.26	108.19	220.89	221.82	1054.02	1056.22
Hinkley	43.14	76.26	161.70	1003.30	59.81	59.43	118.11	118.86	548.80	549.80
Muggeo	175.62	118.34	140.98	228.69	100.63	98.63	136.09	121.33	201.03	188.44
Esc. 4	(15, 15)	(25, 25)	(50, 50)	(250, 250)	(15, 25)	(25, 15)	(25, 50)	(50, 25)	(50, 250)	(250, 50)
Hudson	107.86	177.38	387.74	2413.66	112.15	112.67	227.12	227.28	1064.49	1065.75
Hinkley	43.23	76.53	162.35	1009.67	59.71	59.63	118.56	118.62	549.91	550.09
Muggeo	91.39	92.25	118.42	178.91	86.25	88.09	91.93	99.09	135.78	145.84

Tabla 3.14: Tiempos de ejecución en segundos para las simulaciones de los escenarios 3 a 10.

3.3. Estimación de la diferencia de pendientes

Todos los resultados comentados hasta el momento estaban relacionados con la estimación del punto de cambio. No obstante, para tener la completa seguridad de que los métodos funcionan adecuadamente en la práctica, es necesario también analizar lo que ocurre con la diferencia de pendientes, pues de nada sirve saber dónde cambia de tendencia la función de regresión si no se cuantifica correctamente el cambio.

Puesto que el caso principal es el homocedástico, solamente se estudiarán los Escenarios 1, 2, 3 y 4. En esta ocasión se presentan los resultados a través de **Diagramas de Caja o Boxplots** y no se mostrarán tablas. Este nuevo formato permite hacerse una idea de la distribución de valores del cambio de pendiente e identificar los atípicos con mucha mayor facilidad. Además, en esta sección no se incluyen los tiempos de computación puesto que resulta redundante con las Tablas 3.13 y 3.14 presentadas en la anterior sección. Por último, cabe destacar también que las muestras sobre las que se han calculado las diferencias de pendiente en cada caso son las mismas que las utilizadas para estimar el punto de cambio.

Escenario 1

Recordemos que el Escenario 1 se caracterizaba por tener $\psi = 5$ y una diferencia de pendientes $\beta_\psi = 1$. Se puede observar que para los tres métodos ocurre prácticamente lo mismo: en muestras de tamaño 30 el estimador de la diferencia de pendientes presenta un pequeño sesgo, posiblemente influenciado por la gran cantidad de atípicos que se observan. Según se aumenta el tamaño muestral, este sesgo va desapareciendo y la cantidad de atípicos se reduce considerablemente. Es necesario comentar también que, a tenor de lo que se ve en la imagen, la varianza del estimador se reduce también con el tamaño muestral. Todos estos hechos van en la línea de lo que se había probado en el capítulo de estimación, que garantizaba la consistencia de los estimadores.

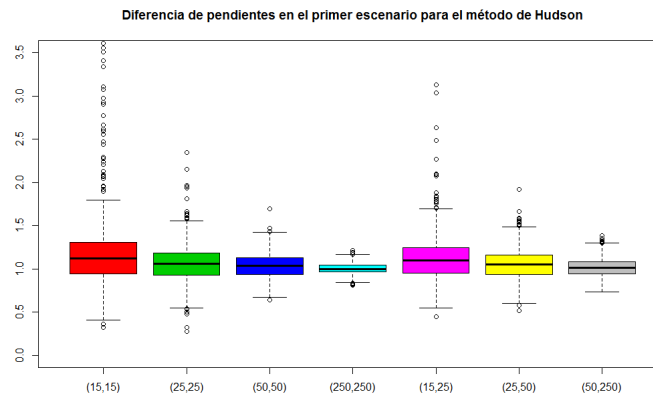


Figura 3.4: Diferencia de pendientes para cada tamaño muestral para el método de Hudson en el Escenario 1.

Escenario 2

En el Escenario 2, donde $\psi = 2.5$ y $\beta_\psi = 5$, las conclusiones que se pueden sacar son muy similares a la del Escenario 1. No obstante, lo que sí que varía es la magnitud de los problemas que presentaban los tres métodos en el Escenario 1: ahora, aunque el sesgo en muestras pequeñas existe, es menor que en el Escenario 1 y la cantidad de atípicos es manifiestamente menor. Claramente, esto es consecuencia

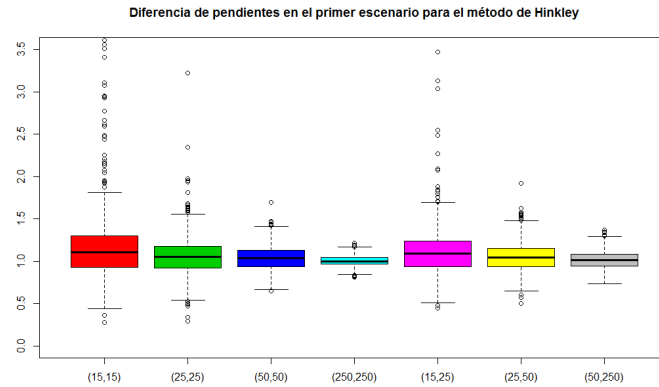


Figura 3.5: Diferencia de pendientes para cada tamaño muestral para el método de Hinkley en el Escenario 1.

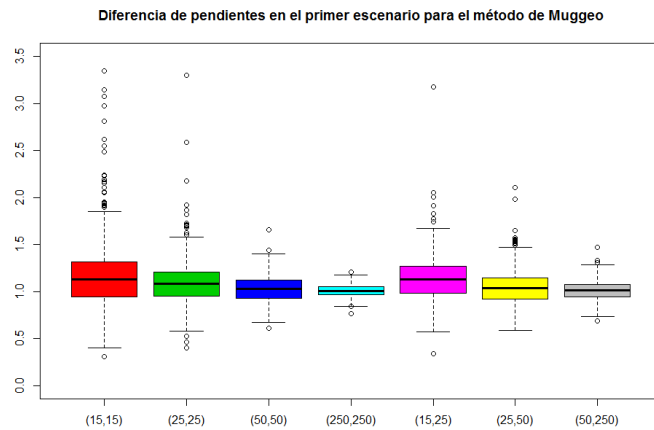


Figura 3.6: Diferencia de pendientes para cada tamaño muestral para el método de Muggeo en el Escenario 1.

de que los tres métodos estiman mucho mejor ψ cuando la diferencia de pendientes es elevada y así los métodos pueden estimar mejor las pendientes de las rectas a cada lado de ψ .

Escenario 3

Habiendo mostrado lo que ocurre cuando el punto de cambio está en el centro del dominio de definición, pasemos ahora a los dos escenarios en los que $\psi = 2.5$. El tercero de ellos estaba construido con $\beta_\psi = 1$. De los cuatro escenarios considerados, éste era donde la estimación del punto de cambio funcionaba peor para los tres métodos, así que cabe esperar que con la diferencia de pendientes ocurra lo mismo. Efectivamente, el sesgo en muestras pequeñas es mucho más elevado y la cantidad de atípicos es ingente. Como siempre, estos problemas desaparecen al aumentar el tamaño muestral, aunque incluso para muestras de tamaño 500 hay más atípicos que en los otros escenarios. Comparando la Imagen 3.12 con las Imágenes 3.10 y 3.11, puede parecer que las cajas para el método de Muggeo son más

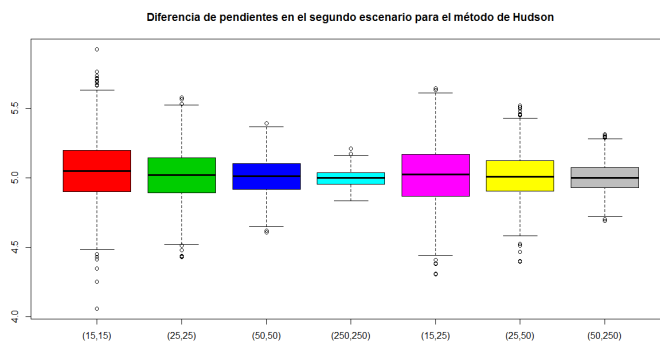


Figura 3.7: Diferencia de pendientes para cada tamaño muestral para el método de Hudson en el Escenario 2.

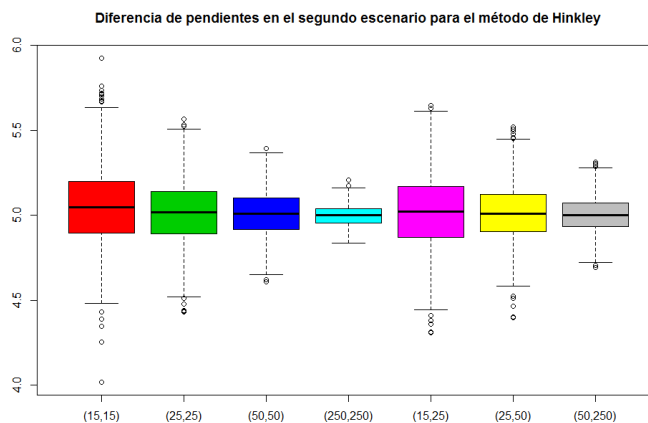


Figura 3.8: Diferencia de pendientes para cada tamaño muestral para el método de Hinkley en el Escenario 2.

estrechas o que hay menos atípicos, pero hay que recordar también que el Algoritmo de Muggeo (2003) no convergió en 21 casos.

Escenario 4

Por último, cuando $\psi = 2.5$ y $\beta_\psi = 5$, los problemas de la estimación en el Escenario 3 se reducen mucho, como yo ocurría con el Escenario 2 en comparación con el Escenario 1.

El resumen que se expuso a continuación de la parte de estimación del punto de cambio en el Estudio de Simulación aplica también aquí: los métodos estimarán mejor la diferencia de pendientes cuanto mayor sea la evidencia de que el modelo de regresión segmentada es adecuado. Es decir, se estimará mejor β_ψ cuanto mayor sea su valor y cuanto más cerca del centro del dominio de definición esté ψ .

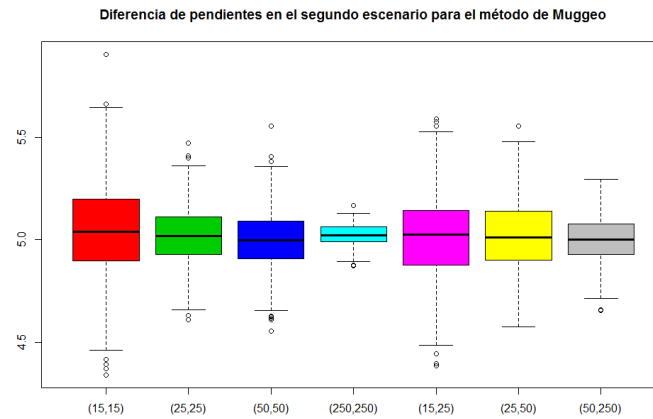


Figura 3.9: Diferencia de pendientes para cada tamaño muestral para el método de Muggeo en el Escenario 2.

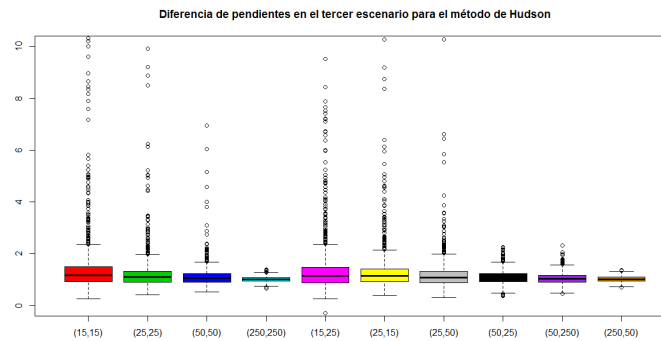


Figura 3.10: Diferencia de pendientes para cada tamaño muestral para el método de Hudson en el Escenario 3.

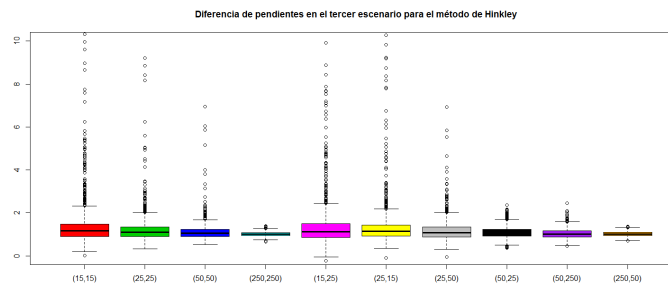


Figura 3.11: Diferencia de pendientes para cada tamaño muestral para el método de Hinkley en el Escenario 3.

Es necesario comentar también que la escala escogida para las figuras relativas a cada escenario es diferente puesto que se pretende mostrar también cuántos atípicos hay y su dimensión, algo que no sería posible poniendo todas en la misma escala (ya que el hecho de que en algún caso haya que coger

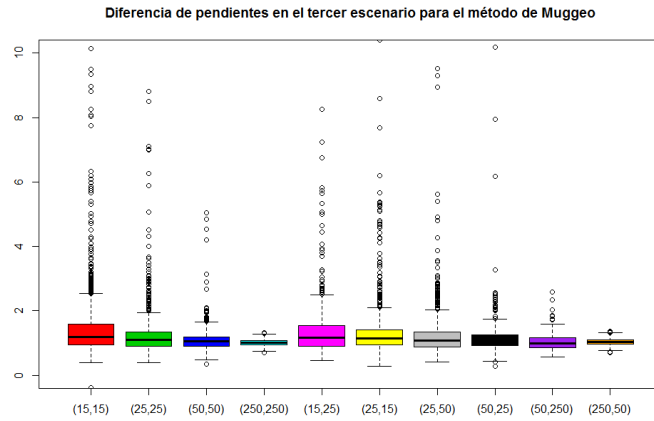


Figura 3.12: Diferencia de pendientes para cada tamaño muestral para el método de Muggeo en el Escenario 3.

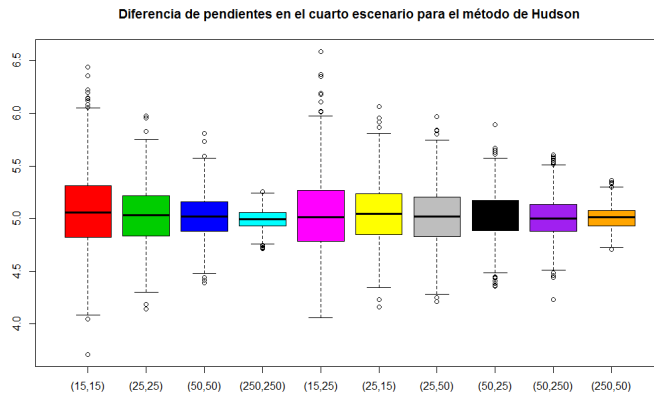


Figura 3.13: Diferencia de pendientes para cada tamaño muestral para el método de Hudson en el Escenario 4.

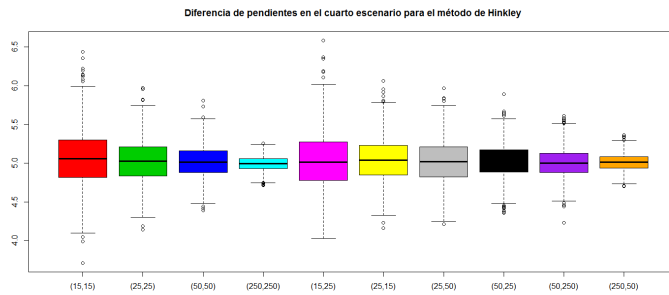


Figura 3.14: Diferencia de pendientes para cada tamaño muestral para el método de Hinkley en el Escenario 4.

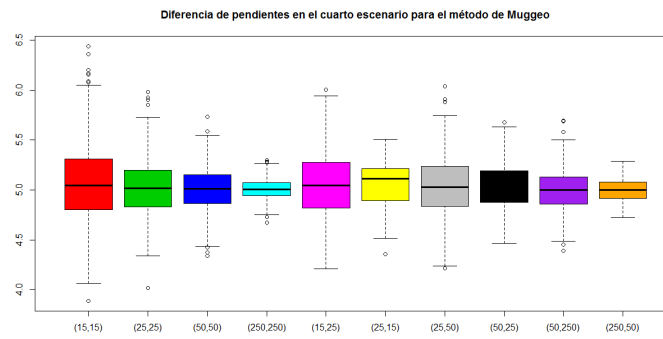


Figura 3.15: Diferencia de pendientes para cada tamaño muestral para el método de Muggeo en el Escenario 4.

rangos muy grandes haría que las cajas más finas fuesen prácticamente indistinguibles.)

Capítulo 4

Aplicación a datos reales

Según pasa el tiempo, es cada vez más evidente la influencia del ser humano en el paisaje y en las relaciones entre especies animales. Desgraciadamente, esta influencia es negativa y en algunos casos provoca la pérdida de hábitat. Es por ello que numerosos investigadores se han centrado en investigar las condiciones que propician la desaparición de las especies.

Rígurosamente, se define la **pérdida o destrucción del hábitat** como el proceso por el cual un hábitat natural es transformado y se vuelve incapaz de mantener a las especies originarias del mismo. Dentro de la diversidad de teorías existentes sobre cómo se produce este fenómeno, este trabajo prestará especial atención a la **Teoría de la fragmentación del hábitat**. En líneas generales, se basa en que el proceso de pérdida de hábitat se produce mediante la división de una gran superficie de terreno en un número de parches más pequeños que se encuentran aislados entre sí por una matriz con propiedades diferentes a la del hábitat original. A medida que pasa el tiempo, los parches van separándose cada vez más entre sí, lo cual llevaría al aislamiento de las poblaciones que los componen y, finalmente, a su extinción. Por tanto, en base a esta teoría se generaron dos hipótesis:

- **La hipótesis del umbral de fragmentación** afirma que los umbrales de pérdida de hábitat (a nivel especie) son consecuencia de una interacción entre la fragmentación del paisaje y los niveles de cantidad de hábitat, entendiendo como cantidad de hábitat el porcentaje de especies remanentes en el paisaje. Más detalles sobre esta hipótesis pueden encontrarse en Andrén (1994).
- **La hipótesis del umbral de extinción** (Fahrig (2003)), por su parte, sostiene que la fragmentación es independiente de la pérdida de hábitat. Es decir, los umbrales de pérdida de hábitat (también a nivel especie) se deberían exclusivamente a los bajos niveles de cantidad de hábitat.

En esta dirección se enmarca el trabajo de Betts. et al (2007), que analiza datos de 15 especies distintas de aves cantoras en un bosque canadiense. Para cada especie, los autores ajustan un modelo de regresión segmentada donde la variable explicativa es la cantidad de hábitat del paisaje correspondiente y la respuesta es la aparición o no del animal en un radio de 150 metros. Esto es diferente a lo planteado hasta el momento en este documento ya que solamente se ha trabajado con respuestas de tipo continuo. De los tres métodos estudiados en este trabajo, sólo Muggeo (2003) contempla la posibilidad de una respuesta de tipo no continuo, de modo que se trabajará con este método.

La idea subyacente en Betts. et al (2007) era comprobar si existían umbrales en la cantidad de hábitat del paisaje a partir de los cuales la existencia de las aves fuese mucho menos probable. Una vez hecho esto, lo siguiente que hicieron fue ajustar modelos de regresión logística múltiple con interacción donde las variables explicativas fueran la cantidad de hábitat y la fragmentación del paisaje, medida en términos del tamaño de los parches¹.

¹Se entiende que, cuanto más grandes sean los parches, menos parches va a haber para la misma unidad de terreno y, consecuentemente, menos fragmentado estará el hábitat.

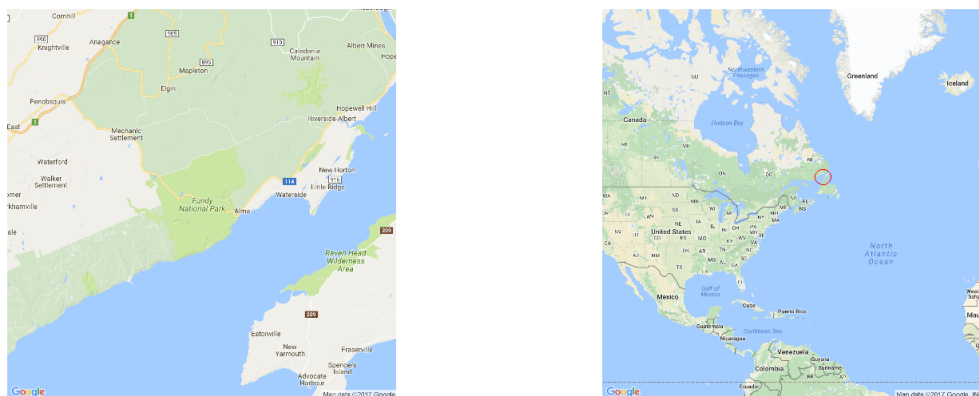


Figura 4.1: Vista aérea del Fundy National Park en New Brunswick, Canadá (izqda.) y Situación del Fundy National Park dentro de Canadá (drcha.).

De las 15 modelos ajustados (uno para cada especie), sólo uno tuvo una interacción significativa. La especie en cuestión se denomina *Seirus aurocappilla* y se han obtenido sus datos para este trabajo por cortesía del profesor Betts. Los resultados presentados en el artículo apoyan fuertemente la hipótesis del umbral de extinción, si bien no son definitivas dado que pueden existir otras variables que midan la fragmentación del paisaje que interactúen con la cantidad de hábitat y no estén consideradas por los autores.

No obstante, aquí no se analizará la interacción del tamaño del parche con la cantidad de hábitat, únicamente se intentará ajustar un modelo de regresión lineal segmentada entre la probabilidad de aparición del ave y la cantidad de hábitat del entorno. Esto permitiría obtener el umbral a partir del cual la pérdida de hábitat se incrementa, independientemente de cuál de las dos teorías sea cierta.

4.1. Características de la muestra

El estudio fue llevado a cabo en el Fundy National Park, situado en la provincia de New Brunswick, Canadá ($66.08^{\circ}O$ a $64.96^{\circ}O$, $46.08^{\circ}N$ a $45.47^{\circ}N$). El paisaje de este área está mayoritariamente formado por bosque (89 %) con un clima marítimo (Betts et al. (2003)). Además, es heterogéneo como consecuencia de la actividad humana en la zona (tala, plantación de diferentes especies, etc.). La Figura 4.1 (izqda.) muestra una imagen extraída de google maps a través del paquete `ggmap` de R, mientras que la Figura 4.1 (drcha.) ofrece una panorámica global de la situación del Fundy National Park dentro de Canadá y global.

Para registrar la existencia o no de aves, se establecieron dos grids: uno en 2000 y otro en 2002 hasta juntar un tamaño muestral de $n = 425$ puntos. Se fijaron tres horas entre las 05 : 30 y las 11 : 00 en las cuales se hacía un recuento de 5 minutos. Inicialmente, se registró la cantidad de aves macho que emitían algún sonido en esos 5 minutos. Este recuento rara vez tomaba valores mayores que 2, así que se optó por transformar este recuento en una variable dicotómica (aparición o no aparición). Dado que lo que interesa es la probabilidad de aparición de la especie en cuestión, los dos valores de esta variable los llamaremos **éxito** (hay un ave en la localización especificada) y **fracaso** (no hay ningún ave en la localización especificada).

La Figura 4.2 muestra la distribución espacial de los éxitos y fracasos en la región de estudio. A

simple vista, parece que los éxitos generalmente se sitúan por encima de la recta de 45° y los fracasos por debajo. Esto indica la posibilidad de que exista **correlación espacial**. Para determinarlo, se aplicará un test de independencia sobre los residuos del modelo ajustado especial para correlación de tipo espacial. De ser así, dado que una de las premisas básicas del modelo de regresión segmentada es la independencia de los errores, sería necesario eliminar esta componente espacial.

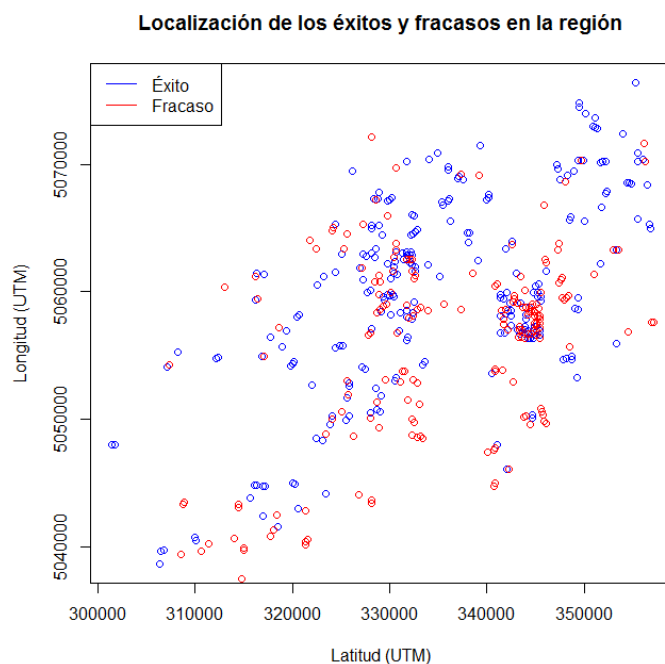


Figura 4.2: Distribución espacial de los éxitos y fracasos en la muestra.

4.2. Ajuste del modelo de regresión segmentada

Variable	Estimación	Error estándar	Estadístico Z	P-valor
(Intercept)	-1.8650	0.3188	-5.85	0.0000
Cantidad de hábitat	0.0352	0.0049	7.13	0.0000

Tabla 4.1: Resumen del ajuste del modelo de regresión logística de la probabilidad de existencia de la especie sobre la cantidad de hábitat.

En la Tabla 4.1 se puede ver el ajuste del modelo de regresión logística (sin punto de cambio) sobre la probabilidad de existencia de la especie utilizando el método de Muggeo. Vemos que los coeficientes son muy significativos. Este modelo tiene un AIC de 526.47. Por su parte, el modelo presentado en la Tabla 4.2 tiene un punto de cambio en el punto $X = 47.68$. Lo que devuelve R en este caso es la diferencia de pendientes antes y después del punto de cambio ($\hat{\beta}_\psi$). Al contrario que en el primer caso,

este modelo refleja que si sólo permanece la mitad del hábitat o menos, la probabilidad de existencia se reduce de forma mucho más brusca al considerar cantidades de hábitat menores. En ambos casos la relación es creciente: a mayor cantidad de hábitat, mayor probabilidad de existencia del ave. No obstante, preferimos el segundo por su AIC de 522.15 y su interpretación más acorde a la realidad.

Variable	Estimación	Error estándar	Estadístico Z	P-valor
(Intercepto)	-3.2230	0.7349	-4.39	0.0000
Cantidad de hábitat	0.0751	0.0215	3.49	0.0005
Cantidad de hábitat > 47.68	-0.0609	0.0239	-2.55	0.0107

Tabla 4.2: Resumen del ajuste del modelo de regresión segmentada logística de la probabilidad de existencia de la especie sobre la cantidad de hábitat.

Lo último que queda por comprobar es si existe dependencia espacial. Para ello, se aplicará un test basado en el variograma existente en la librería `sm` de R.

El variograma se utiliza en la rama de la estadística espacial para cuantificar la varianza en el valor de las realizaciones de un proceso estocástico a una determinada distancia de las observaciones de la muestra. Por tanto, la idea aquí es estimar empíricamente el variograma y ver cómo de probable es bajo la hipótesis nula de independencia espacial. Bajo dependencia espacial, observaciones cercanas tomarán valores cercanos y la similitud de estas observaciones decrece con la distancia. Por lo tanto, una forma de estimar el variograma podría ser medir la diferencia entre los valores de las observaciones a diferentes distancias y representar esta diferencia en función de la distancia. Esto se consigue en la Figura 4.3.

En base a lo dicho hasta ahora, bajo independencia espacial no debería haber un aumento en la diferencia de valores con la distancia, de modo que debiera ser plano. Las bandas de referencia para el verdadero valor del variograma en la Figura 4.3 pueden contener a una recta de pendiente 0 y, por lo tanto, no hay evidencias para rechazar la independencia espacial. La función `sm.variogram` de R proporciona este gráfico y un p-valor de 0.46 que permiten sostener que la dependencia espacial no está presente, en contra de lo que inicialmente pudiera parecer.

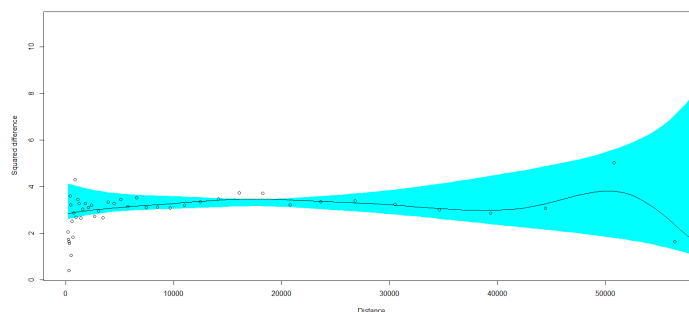


Figura 4.3: Estimación empírica del variograma y bandas de confianza para el verdadero valor del variograma.

Capítulo 5

Ampliación del estudio sobre el modelo de regresión segmentada

Habiendo explorado los métodos de estimación a este capítulo se expondrán los procedimientos de inferencia relativos a los parámetros del modelo. Concretamente, primero se presentarán diferentes contrastes de hipótesis existentes en la literatura: uno para contrastar el caso de igualdad de pendientes (es decir, $\beta_\psi = \beta_1$) y otro para el no efecto a partir del punto de cambio ($\beta_\psi = 0$).

5.1. Contraste de existencia de punto de cambio

En el segundo capítulo se presentaron los métodos de estimación del punto de cambio supuesto que éste exista. En ocasiones, esto último no está del todo claro y es precisa una herramienta que permita garantizar la presencia de un cambio de tendencia. Por simplicidad, se trabajará con la notación del modelo (2.3), que es equivalente al modelo (1.3), y tomaba la forma

$$Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_i & i = 1, \dots, \tau \\ \alpha_2 + \beta_2 x_i + \varepsilon_i & i = \tau + 1, \dots, n, \end{cases}$$

con $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$, $\varepsilon_i \sim N(0, \sigma^2)$ y sujeto a la restricción de continuidad $\alpha_2 - \alpha_1 + \psi(\beta_1 - \beta_2) = 0$. Para garantizar la existencia del punto de cambio se utilizarán contrastes de hipótesis basados en el anterior modelo, con hipótesis nula y alternativa

$$\begin{aligned} H_0 : \beta_1 &= \beta_2 \\ H_a : \beta_1 &\neq \beta_2. \end{aligned} \tag{5.1}$$

El contraste propuesto por Hinkley (1971) para la existencia de punto de cambio recuerda mucho a los test F que se efectúan en regresión lineal simple para comprobar la significación de los parámetros del modelo. El estadístico de contraste tomará la forma

$$FT = \frac{RSS^0 - RSS_{\hat{\tau}}(\hat{\psi})/3}{RSS_{\hat{\tau}}(\hat{\psi})/(n-4)}, \tag{5.2}$$

siendo $\hat{\tau}$ tal que $x_{\hat{\tau}} \leq \hat{\psi} \leq x_{\hat{\tau}+1}$. Recordemos que $RSS^0 = \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{C_{g_1, n}^2}{S_{g_1, n}}$, $RSS_{\tau}(\psi) = \sum_{i=1}^{\tau} (Y_i - \alpha_1 - \beta_1 x_i)^2 + \sum_{i=\tau+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2$. La expresión final del numerador de la ecuación 5.2 es consecuencia de la propia definición de Z_{τ}^2 .

Lo que se pretende mediante este estadístico es cuantificar las diferencias entre la suma cuadrática de residuos con punto de cambio y sin él para ver si la inclusión del mismo proporciona una mejora clara del modelo.

Para hallar la distribución del estadístico FT , es necesario echar la vista atrás y recordar la expresión de la suma cuadrática de residuos que da lugar al estadístico de Hinkley (1969). De acuerdo a la expresión (2.23), se tiene

$$Z_\tau^2(w) = RSS^0 - RSS_\tau(w) \quad \forall w \in [x_\tau, x_{\tau+1}),$$

lo cual da lugar a la expresión

$$FT = \frac{Z_\tau^2(\hat{\psi})/3}{RSS_{\hat{\tau}}(\hat{\psi})/(n-4)}.$$

Recordemos que $Z_\tau^2(\cdot)$ se podía escribir también como

$$Z_\tau^2(w) = \frac{(\tilde{\beta}_{2,\tau} - \tilde{\beta}_{1,\tau})^2 \{F_\tau - 2D_\tau(w + \hat{\psi}_\tau) + E_\tau \hat{\psi}_\tau w\}^2}{S_{g_1,n}^2(F_\tau - 2D_\tau w + E_\tau w^2)}, \quad (5.3)$$

donde $D_\tau = \frac{\tau(n-\tau)}{n}(\bar{x}_{g_1} S_{g_2,\tau}^2 + \bar{x}_{g_2} S_{g_1,\tau}^2)$, $E_\tau = \frac{\tau(n-\tau)}{n}(S_{g_1,\tau}^2 + S_{g_2,\tau}^2)$ y $F_\tau = S_{g_1,\tau}^2 S_{g_2,\tau}^2 + \frac{\tau(n-\tau)}{n}(\bar{x}_{g_1}^2 S_{g_2,\tau}^2 + \bar{x}_{g_2}^2 S_{g_1,\tau}^2)$, como ya se definió en el capítulo de estimación.

El numerador de la expresión 5.3 será el motivador de la idea que inspira este test. De no existir un cambio de pendiente, entonces los estimadores $\tilde{\beta}_{1,\tau}$ y $\tilde{\beta}_{2,\tau}$ (es decir, los estimadores de los parámetros de tendencia a ambos lados del punto) serían el mismo para un punto arbitrario $w \in \mathbb{R}$. En tal caso, se tendría además que $Z_\tau^2(\cdot) = 0$ y, en virtud de (2.23), $RSS_\tau(\cdot) = RSS^0$.

Considerando el estimador de máxima verosimilitud del punto de cambio, $\hat{\psi}$, y desarrollando la expresión (2.23), se llega a que

$$RSS_{\hat{\tau}}(\hat{\psi}) = \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{C_{g_1,n}^2}{S_{g_1,n}} - Z_{\hat{\tau}}^2(\hat{\psi}) = \sum_{i=1}^n (y_i - \bar{y})^2 - \tilde{\beta}_{1,n}^2 S_{g_1,n} - Z_{\hat{\tau}}^2(\hat{\psi}).$$

Ahora, por la suposición de normalidad, se tiene que $\sum_{i=1}^n (y_i - \bar{y})^2$ sigue una distribución $\sigma^2 \chi_{n-1}^2$. Como, además, la expresión de $RSS_{\hat{\tau}}(\hat{\psi})$ contiene tres parámetros más a estimar, se deduce que

$$\frac{RSS_{\hat{\tau}}(\hat{\psi})}{\sigma^2} \sim \chi_{n-4}^2.$$

En cuanto al numerador del estadístico FT , Hinkley (1971) sostiene que hay evidencias empíricas suficientes como para afirmar que

$$\frac{Z_{\hat{\tau}}^2(\hat{\psi})}{\sigma^2} \sim \chi_3^2,$$

aunque este resultado es asintótico. De este modo, ya que el estadístico FT es cociente de dos estadísticos con distribución chi-cuadrado, FT seguirá una distribución F de Snédecor. Concretamente, se sigue que

$$FT = \frac{Z_{\hat{\tau}}^2(\hat{\psi})/3}{RSS_{\hat{\tau}}(\hat{\psi})/(n-4)} \sim \mathcal{F}_{3,n-4},$$

donde $\mathcal{F}_{3,n-4}$ denota la distribución F de Snédecor con 3 grados de libertad en el numerador y $n-4$ grados de libertad en el denominador. Por tanto, trabajando con un nivel de significación α , se rechazará la hipótesis nula cuando $FT > \mathcal{F}_{\alpha,3,n-4}$, siendo $\mathcal{F}_{\alpha,3,n-4}$ el cuantil $1-\alpha$ de la distribución $\mathcal{F}_{3,n-4}$.

No obstante, la distribución propuesta para el estadístico FT está basada en una consideración subjetiva y cabría preguntarse si la evidencia empírica referida anteriormente se puede refinar un

poco más. Para ello, en este documento se propone un procedimiento basado en el bootstrap, que se describirá posteriormente en forma de algoritmo.

El bootstrap permite aproximar la distribución en el muestreo de un estadístico. Si la distribución poblacional a la que pertenece la muestra fuese conocida, lo ideal sería generar muestras de dicha distribución, calcular el estadístico en todas ellas y aproximar la distribución en el muestreo del estadístico con la distribución de probabilidad dada por sus valores en esas muestras. No obstante, no siempre se conoce la distribución poblacional. La idea que propone el bootstrap es generar remuestras a partir de un estimador de la distribución poblacional.

Por ejemplo, en el caso del modelo (1.3), se supone que $\varepsilon \in N(0, \sigma^2)$. Una posible idea sería estimar el error en la muestra y generar las remuestras a partir de este estimador. Con las remuestras del error y fijando los valores de la variable explicativa, se tendrían nuevos valores de la respuesta y , por tanto, nuevos valores del estadístico en cuestión que permitirán aproximar su distribución en el muestreo. Este método es útil para calcular los valores críticos del contraste y consistiría en lo siguiente

1. Dada la muestra $\{(x_i, Y_i)\}_{i=1}^n$ y el modelo (2.3), estimar los parámetros $\psi, \alpha_1, \beta_1, \alpha_2$ y β_2 mediante el método de Hinkley, dando lugar a $\hat{\psi}, \hat{\alpha}_1, \hat{\beta}_1, \hat{\alpha}_2$ y $\hat{\beta}_2$.
2. Estimar σ^2 mediante $\hat{\sigma}^2$.
3. Generar n valores $\varepsilon_1^*, \dots, \varepsilon_n^* \sim N(0, \hat{\sigma}^2)$. De ahora en adelante, * indicará que hay un proceso bootstrap implicado. Otra opción sería remuestrear de los residuos recentrados, $\tilde{\varepsilon}_i = \varepsilon_i - \bar{\varepsilon}$, siendo $\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i$.
4. Construir $Y_i^* = \hat{\alpha}_1 + \hat{\beta}_1 x_i + \varepsilon_i^*, i = 1, \dots, n$.
5. Con la muestra $\{(x_i, Y_i^*)\}_{i=1}^n$, calcular los estimadores bootstrap de $\psi, \alpha_1, \beta_1, \alpha_2$ y β_2 por el método de Hinkley dados por $\hat{\psi}^*, \hat{\alpha}_1^*, \hat{\beta}_1^*, \hat{\alpha}_2^*$ y $\hat{\beta}_2^*$. Asimismo, sea $\hat{\tau}^* = \max\{i; x_i \leq \hat{\psi}^* < x_{i+1}\}$.
6. Calcular $FT^* = \frac{Z_{\hat{\tau}^*}(\hat{\psi}^*)}{RSS_{\hat{\tau}^*}(\hat{\psi}^*)}$.
7. Repetir B veces los Pasos 3 a 5 para obtener FT_1^*, \dots, FT_B^* .
8. Sea $FT_{(1)}^*, \dots, FT_{(B)}^*$ la muestra ordenada de valores bootstrap de FT . A un nivel de significación α , el valor crítico del test será $FT_{[(B(1-\alpha))]}^*$, donde $[a]$ denota la parte entera de a .

Es necesario destacar que la hipótesis nula del test es unilateral a la derecha, pues el estadístico U está midiendo diferencias entre un modelo con punto de cambio y sin él. Por tanto, recoge si hay cambios grandes (región de rechazo, valores muy positivos de U) o no hay ninguno (valores cercanos a 0).

Cabe decir también que este método propuesto tiene diversas variantes en función del procedimiento elegido para generar las remuestras en el Paso 3. Otra opción podría ser utilizar la distribución empírica del error o usar el conocido como **Wild Bootstrap**, aunque es más indicado en contextos heterocedásticos. Este procedimiento consiste en encontrar una variable aleatoria P^* tal que $E(P^*) = 0$, $E(P^{*2}) = 1$ y $E(P^{*3}) = 1$ para, posteriormente, generar P_1^*, \dots, P_n^* y definir $\varepsilon_i^* = \hat{\varepsilon}_i P_i^*$. Tanto si se generan las remuestras del error con el método propuesto o con estos dos, los siguientes Pasos son los mismos para todos los casos. Más sobre el bootstrap se puede encontrar en Efron y Tibshirani (1994) y, en particular, del Wild Bootstrap en Wu (1986).

5.2. Contraste de no efecto

En el contexto de la regresión segmentada, si bien el contraste de existencia de punto de cambio es el primero que viene a la mente, no siempre es el más importante. Podría darse el caso de que el

investigador quisiese contrastar alguna hipótesis sobre los parámetros de tendencia del modelo. En esta sección se trata el caso concreto de que no haya relación entre la variable explicativa y la respuesta a partir del punto de cambio. En términos de la notación del Modelo (2.3), esto se traduce en realizar el contraste:

$$H_0 : \beta_2 = 0 \quad \text{vs.} \quad H_a : \beta_2 \neq 0.$$

Un ejemplo de este caso puede verse en Hinkley (1971), donde se intenta comprobar que hay un umbral en los niveles de warfarina (variable explicativa) y la producción del Factor de Coagulación VII (variable respuesta), cuya ausencia provoca problemas de coagulación. Por lo tanto, a la hora de tratar a pacientes con déficit de Factor de Coagulación VII, es importante determinar si a partir de un determinado valor el suministro de warfarina deja de ser efectivo. A modo de curiosidad, Hinkley (1971) transforma los niveles de warfarina con un logaritmo, resultando que el punto de cambio es $\hat{\psi} = 4.65$ y el contraste sugiere que no hay evidencias para rechazar la hipótesis nula de no efecto, si bien las conclusiones que se pueden sacar son moderadas dado que la muestra tiene $n = 15$ datos.

Antes de comenzar a exponer el procedimiento llevado a cabo para realizar el contraste, es necesario decir que, ya que en el contexto homocedástico máxima verosimilitud y mínimos cuadrados son equivalentes, las propiedades de sesgo, consistencia y normalidad del estimador de Hudson (1966) se conservan también para el de Hinkley (1969). Esto tendrá una importancia clave posteriormente.

Pensado en perspectiva, el problema de contrastar el no efecto a partir del punto de cambio es equivalente a contrastar que la pendiente de una recta de regresión es nula, de modo que se presentará primero la teoría para este contraste en la regresión lineal simple y luego se adaptará al caso de regresión segmentada.

Dada una muestra $\{(x_1, Y_1), \dots, (x_n, Y_n)\}$, se vio al inicio del Capítulo 2 (Ecuación 2.1) que los estimadores de máxima verosimilitud y mínimos cuadrados de la pendiente en regresión lineal simple tomaban la forma:

$$\hat{\beta}_1 = \frac{S_{xY}}{S_x^2} \bar{x} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{nS_x^2},$$

donde S_{xY} denota la covarianza entre X e Y y S_x^2 es la varianza de X .

El objetivo del procedimiento que sigue es encontrar la distribución de $\hat{\beta}_1$ para poder obtener un estadístico de contraste y las regiones de aceptación y rechazo. Ya que $\hat{\beta}_1$ es una combinación lineal de variables normales (pues Y_1, \dots, Y_n lo son por hipótesis), $\hat{\beta}_1$ seguirá una **distribución normal**, cuya esperanza y varianza se calcularán en lo que sigue:

$$E(\hat{\beta}_1) = E\left(\frac{S_{xY}}{S_x^2}\right) = E\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{nS_x^2}\right) = \frac{\sum_{i=1}^n (x_i - \bar{x})}{nS_x^2} E(Y_i - \bar{Y}).$$

Puesto que, por hipótesis, $Y = \beta_0 + \beta_1 X$, la última igualdad se convierte en

$$E(\hat{\beta}_1) = \frac{\sum_{i=1}^n (x_i - \bar{x})}{nS_x^2} (x_i - \bar{x}) \beta_1 = \beta_1,$$

de modo que $\hat{\beta}_1$ es un estimador insesgado de la pendiente.

Por su parte, la varianza del estimador $\hat{\beta}_1$ vendría dada por las siguientes igualdades:

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \text{Var}\left(\frac{S_{xY}}{S_x^2}\right) = \text{Var}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{nS_x^2}\right) = \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n^2 S_x^4} \text{Var}(Y_i) = \frac{\sigma^2}{nS_x^2}, \end{aligned}$$

donde la última igualdad es consecuencia de la hipótesis de homocedasticidad.

En resumen, la distribución del estimador de la pendiente será

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{nS_x^2}\right).$$

Ahora bien, la distribución de este estimador depende de la varianza del error, σ^2 , que no siempre es conocida. En base a este hecho, pueden darse dos casos:

Si σ^2 es conocida, entonces el estadístico usado para el test será:

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma/S_x\sqrt{n}} \sim N(0, 1). \quad (5.4)$$

Si σ^2 es desconocida, entonces será necesario estimar dicha varianza. El estimador empleado aquí será

$$\hat{\sigma} = \frac{\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2}.$$

Ya que $\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}$ el estadístico usado para el test será:

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/S_x\sqrt{n}} \sim t_{n-2}, \quad (5.5)$$

siendo t_{n-2} la distribución T de Student con $n-2$ grados de libertad.

A diferencia de la hipótesis nula del contraste de existencia de punto de cambio, la hipótesis nula de este test es bilateral. Así, si la varianza es conocida, se rechazará la hipótesis nula de no efecto en regresión lineal simple si $\frac{|\hat{\beta}_1|}{\sigma/S_x\sqrt{n}} > z_{\alpha/2}$, donde $z_{\alpha/2}$ es el cuantil $1 - \frac{\alpha}{2}$ de la distribución $N(0, 1)$.

Por contra, si no se conoce σ^2 , se rechazará la hipótesis nula si $\frac{|\hat{\beta}_1|}{\hat{\sigma}/S_x\sqrt{n}} > t_{n-2, \alpha/2}$, siendo $t_{n-2, \alpha/2}$ el cuantil de orden $1 - \frac{\alpha}{2}$ de la distribución T de Student con $n-2$ grados de libertad.

Ahora que se ha desarrollado el caso de regresión lineal simple, analicemos la analogía con el caso de regresión lineal segmentada.

Con la notación adoptada en este trabajo, el estimador de la pendiente a la derecha del punto de cambio toma la forma

$$\hat{\beta}_2 = \frac{C_{g_2, \hat{\tau}}}{S_{g_2, \hat{\tau}}^2},$$

donde, recordemos, $C_{g_2, \tau} = \sum_{i=\tau+1}^n (x_i - \bar{x}_{g_2})(y_i - \bar{y}_{g_2})$ y $S_{g_2, \tau}^2 = \sum_{i=\tau+1}^n (x_i - \bar{x}_{g_2})^2$.

Puesto que el estimador de Hinkley (1969) y el de Hudson (1966) son equivalentes en este contexto, el estimador de Hinkley (1969) hereda las propiedades de consistencia y normalidad. Es decir:

$$\hat{\beta}_2 \rightarrow N(\beta_2, \mathcal{I}^{-1}),$$

siendo $\beta_2 = (\alpha_2, \beta_2, \sigma^2)^t$, $\hat{\beta}_2 = (\hat{\alpha}_2, \hat{\beta}_2, \hat{\sigma}^2)^t$ e \mathcal{I}^{-1} la inversa de la matriz de información de Fisher de los parámetros del modelo. Un sencillo cálculo permite ver que la varianza es la misma que en el caso de regresión lineal simple. Por tanto, se sigue que $\text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{S_{g_2, \hat{\tau}}^2}$.

Por tanto, la distribución del estimador de la pendiente en el segundo tramo será

$$\hat{\beta}_2 \sim N\left(\beta_2, \frac{\sigma^2}{S_{g_2, \hat{\tau}}^2}\right).$$

En virtud de lo descrito en las líneas anteriores sobre el estimador de la pendiente en regresión lineal simple, pueden darse dos situaciones:

Si σ^2 es conocida, entonces el estadístico usado para el test será análogo a la expresión 5.4:

$$\frac{\hat{\beta}_2 - \beta_2}{\sigma/S_{g_2, \hat{\tau}}} \sim N(0, 1). \quad (5.6)$$

Si σ^2 es desconocida, entonces será necesario estimarla. Aquí radica la principal diferencia con el caso de regresión lineal simple. En el primer contraste presentado en este capítulo se probó que

$$\frac{RSS_{\hat{\tau}}(\hat{\psi})}{\sigma^2} \sim \chi_{n-4}^2.$$

Consecuentemente, el estadístico de contraste pierde dos grados de libertad respecto a la regresión lineal simple mostrado en la Ecuación 5.5.

$$\frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}/(n-4)S_{g_2, \hat{\tau}}} \sim t_{n-4}. \quad (5.7)$$

Por lo tanto, si la varianza poblacional es conocida, se rechazará la hipótesis nula de no efecto en regresión lineal segmentada si $\frac{|\hat{\beta}_2|}{\sigma/S_{g_2, \hat{\tau}}} > z_{\alpha/2}$. Por contra, si no se conoce σ^2 , se rechazará la hipótesis nula si $\frac{|\hat{\beta}_2|}{\hat{\sigma}/(n-4)S_{g_2, \hat{\tau}}} > t_{n-4, \alpha/2}$.

Apéndice A

Notación

En un modelo de regresión lineal simple:

- α_1 denota el intercepto del modelo.
- β_1 denota la pendiente del modelo.
- $\varepsilon \sim N(0, \sigma^2)$ denota el error del modelo.

En el modelo de regresión segmentada:

- α_1 denota el intercepto del modelo antes del punto de cambio.
- β_1 denota la pendiente del modelo antes del punto de cambio.
- ψ denota el punto de cambio de pendiente del modelo.
- β_ψ denota la diferencia de pendientes antes y después del punto de cambio.
- $(X - \psi)_+ = (X - \psi)I(X > \psi)$
- $\alpha_2 =$ denota el intercepto tras el punto de cambio.
- $\beta_2 =$ denota la pendiente tras el punto de cambio.
- $\varepsilon \sim N(0, \sigma^2)$ denota el error del modelo.
- ε_i denota el error relativo al i -ésimo individuo.
- ε_i^* denota la i -ésima realización del remuestreo Bootstrap de los residuos muestrales.
- $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)'$ denota el vector de residuos.
- $g_1 = \{x_1, \dots, x_\tau\}$.
- $g_2 = \{x_{\tau+1}, \dots, x_n\}$.
- $L(\cdot)$ denota la función de verosimilitud asociada al modelo.
- $L_\tau(\cdot)$ denota la función de verosimilitud asociada a que el punto de cambio esté en el intervalo $[x_\tau, x_{\tau+1})$, $\tau = 2, \dots, n - 1$.
- $RSS_\tau^1 = \sum_{i=1}^\tau (Y_i - \alpha_1 - \beta_1 x_i)^2$ es la suma de residuos del modelo antes del punto de cambio $\psi \in [x_\tau, x_{\tau+1})$.

- $RSS_\tau^2 = \sum_{i=\tau+1}^n (Y_i - \alpha_2 - \beta_2 x_i)^2$ es la suma de residuos del modelo después del punto de cambio $\psi \in [x_\tau, x_{\tau+1})$.
- $RSS_\tau(\cdot) = RSS_\tau^1 + RSS_\tau^2$ es la suma cuadrática de residuos global de un modelo en el que $\psi \in [x_\tau, x_{\tau+1})$.
- $\hat{\beta}_j = (\hat{\alpha}_j, \hat{\beta}_j)$ denota el vector con los estimadores de máxima verosimilitud (mínimos cuadrados) de $(\alpha_j, \beta_j), j = 1, 2$.
- $\tilde{\beta}_{j,\tau} = (\tilde{\alpha}_{j,\tau}, \tilde{\beta}_{j,\tau})$ denota el estimador local de los coeficientes de regresión sin la restricción de continuidad de los datos con abscisa en $g_j, j = 1, 2$ calculado en el método de Hudson (1966).
- $\hat{\psi}$ denota al estimador de máxima verosimilitud (mínimos cuadrados) del punto de cambio.
- $\hat{\tau}$ es el índice del único elemento muestral que verifica $x_{\hat{\tau}} \leq \hat{\psi} \leq x_{\hat{\tau}+1}$.
- $\hat{\psi}_\tau$ el candidato a estimador en el intervalo $[x_\tau, x_{\tau+1})$.
- $T(\tau) = RSS_\tau(\hat{\psi}_\tau)$ si $\hat{\psi}_\tau \in (x_\tau, x_{\tau+1})$ y $T(\tau) = +\infty$ en otro caso.
- $S(\tau) = RSS_\tau(x_\tau)$ es la suma cuadrática de residuos del modelo de regresión segmentada con punto de cambio x_τ .
- $G_n(\cdot)$ es la distribución empírica de la variable explicativa X en la muestra.
- $G(t) = \lim_{n \rightarrow \infty} G_n(t)$.
- $p(n)$ es un número real tal que $0 < p(n) \leq n$ y $p(n) \rightarrow \infty$ cuando $n \rightarrow \infty$.
- N_n es una secuencia de variables aleatorias verificando $1 \leq N_n \leq n$ y $N_n = O_p(p(n))$.
- ζ_i es un conjunto prefijado de constantes.
- ζ_{ni} es un conjunto de constantes tal que $\sup_n \max_{1 \leq i \leq n} |\zeta_{ni}| < \infty$ y tal que $n^\alpha \zeta_{ni} = \zeta_i(1 + \rho_{ni})$ para $\alpha \geq 0$, con $\max_{1 \leq i \leq n} |\rho_{ni}| = o(p^{-\frac{1}{2}}(n))$.
- m_1 es el orden de la primera derivada de $f_1(\cdot)$ y $f_2(\cdot)$ que no coincide en $X = \psi$.
- $D^+(j, \psi, m_1)$ denota la derivada por la derecha de orden m_1 de f_j evaluadas en $\psi, j = 1, 2$.
- $D^-(j, \psi, m_1)$ denota la derivada por la izquierda de orden m_1 de f_j evaluadas en $\psi, j = 1, 2$.
- \mathcal{I} es la matriz de información de Fisher de los parámetros del modelo.
- \mathbf{A} denotará una matriz $1 \times 2q$ formada por la concatenación de $\frac{\partial f_1(\boldsymbol{\theta}, \psi)}{\partial \theta_l} / D^\pm$ y $-\frac{\partial f_2(\boldsymbol{\theta}, \psi)}{\partial \theta_l} / D^\pm$, $l = 1, \dots, q$.
- \bar{x}_{g_j} denota la media muestral de los elementos de $g_j, j = 1, 2$.
- \bar{y}_{g_j} denota la media muestral de las respuestas con abscisa en $g_j, j = 1, 2$.
- $C_{g_1, \tau} = \sum_{i=1}^\tau (x_i - \bar{x}_{g_1})(y_i - \bar{y}_{g_1})$.
- $C_{g_2, \tau} = \sum_{i=\tau+1}^n (x_i - \bar{x}_{g_2})(y_i - \bar{y}_{g_2})$.
- $S_{g_1, \tau}^2 = \sum_{i=1}^\tau (x_i - \bar{x}_{g_1})^2$.
- $S_{g_2, \tau}^2 = \sum_{i=\tau+1}^n (x_i - \bar{x}_{g_2})^2$.
- $\hat{\alpha}_{1,\tau}$ estimador de máxima verosimilitud (mínimos cuadrados) de α_1 caso de que $\psi \in [x_\tau, x_{\tau+1})$.

- $\hat{\alpha}_{2,\tau}$ estimador de máxima verosimilitud (mínimos cuadrados) de α_2 caso de que $\psi \in [x_\tau, x_{\tau+1})$.
- $\hat{\beta}_{1,\tau}$ estimador de máxima verosimilitud (mínimos cuadrados) de β_1 caso de que $\psi \in [x_\tau, x_{\tau+1})$.
- $\hat{\beta}_{2,\tau}$ estimador de máxima verosimilitud (mínimos cuadrados) de β_2 caso de que $\psi \in [x_\tau, x_{\tau+1})$.
- $RSS^0 = \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{C_{g_1,n}^2}{S_{g_1,n}^2}$.
- $Z_\tau^2(w) = RSS^0 - RSS_\tau(w)$.
- $D_\tau = \frac{\tau(n-\tau)}{n} (\bar{x}_{g_1} S_{g_2,\tau}^2 + \bar{x}_{g_2} S_{g_1,\tau}^2)$.
- $E_\tau = \frac{\tau(n-\tau)}{n} (S_{g_1,\tau}^2 + S_{g_2,\tau}^2)$.
- $F_\tau = S_{g_1,\tau}^2 S_{g_2,\tau}^2 + \frac{\tau(n-\tau)}{n} (\bar{x}_{g_1}^2 S_{g_2,\tau}^2 + \bar{x}_{g_2}^2 S_{g_1,\tau}^2)$.
- $m'(x, \boldsymbol{\theta}) = \frac{\partial m}{\partial \boldsymbol{\theta}}(x, \boldsymbol{\theta})$.
- $\varepsilon^{(0)} = y - m(x, \boldsymbol{\theta}^{(0)})$.
- $\boldsymbol{\varepsilon}^{(0)} = (y_1 - m(x_1, \boldsymbol{\theta}^{(0)}), \dots, y_n - m(x_n, \boldsymbol{\theta}^{(0)}))'$.
- $\mathbf{m}'(x, \boldsymbol{\theta}^{(0)}) = (m'(x_1, \boldsymbol{\theta}^{(0)}), \dots, m'(x_n, \boldsymbol{\theta}^{(0)}))$.
- $\gamma = \beta_\psi(\psi - \psi^{(0)})$.
- $\hat{\gamma}$ es el estimador de máxima verosimilitud (mínimos cuadrados) de γ .
- $\psi^{(0)}$ es la semilla inicial del algoritmo de Muggeo.
- $U^{(0)} = (X - \psi^{(0)})_+$.
- $V^{(0)} = -I(X > \psi^{(0)})$.
- K denota el número de puntos de cambio del modelo. En un modelo con K puntos, éstos se denotarán por ψ_1, \dots, ψ_K .
- $\psi_1^{(0)}, \dots, \psi_K^{(0)}$ son las semillas iniciales del algoritmo de Muggeo para un modelo con K puntos de cambio.
- $U_k^{(0)} = (X - \psi_k^{(0)})_+, k = 1, \dots, K$.
- $V_k^{(0)} = -I(X > \psi_k^{(0)}), k = 1, \dots, K$.
- $ECM(\hat{\psi}) = (E(\hat{\psi}) - \psi)^2 + \text{Var}(\hat{\psi})$.
- $\bar{\psi}$ es la media de los B estimadores $\hat{\psi}$ calculados para cada escenario en el estudio de simulación.
- $\widehat{ECM}(\hat{\psi}) = (\bar{\psi} - \psi)^2 + \text{Var}(\hat{\psi})$.
- $\mathcal{F}_{3,n-4}$ denota la distribución F de Snédecor con 3 grados de libertad en el numerador y $(n-4)$ en el denominador.
- t_{n-2} denota la distribución t de Student $(n-2)$ grados de libertad.
- $[a] = \max\{i \in \mathbb{N}/i \leq a\}$.
- $S_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) = \frac{C_{g_1,n}}{n}$ es la covarianza de X e Y .
- $S_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{S_{g_1,n}^2}{n}$ es la varianza muestral de X .

Bibliografía

- [1] Andren, H. (1994). Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review. *Oikos*, 355-366.
- [2] Betts, M. G., Franklin, S. E., y Taylor, R. G. (2003). Interpretation of landscape pattern and habitat change for local indicator species using satellite imagery and geographic information system data in New Brunswick, Canada. *Canadian Journal of Forest Research*, 33(10), 1821-1831.
- [3] Betts, M. G., Forbes, G. J., y Diamond, A. W. (2007). Thresholds in songbird occurrence in relation to landscape structure. *Conservation Biology*, 21(4), 1046-1058.
- [4] Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2), 247-254.
- [5] Efron, B. y Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC press, Boca Raton.
- [6] Fahrig, L. 2003. Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology Evolution and Systematics*, 34(1), 487-515.
- [7] Feder, P. I. (1975). On asymptotic distribution theory in segmented regression problems—identified case. *The Annals of Statistics*, 3(1), 49-83.
- [8] Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika*, 56(3), 495-504.
- [9] Hinkley, D. V. (1971). Inference in two-phase regression. *Journal of the American Statistical Association*, 66(336), 736-743.
- [10] Hudson, D. J. (1966). Fitting segmented curves whose join points have to be estimated. *Journal of the American Statistical Association*, 61(316), 1097-1129.
- [11] Kahle, D. and Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, 5(1), 144-161. URL <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>
- [12] Muggeo, V. (2003). Estimating regression models with unknown break-points, *Statistics in Medicine*, 22 (19), 3055-3071.
- [13] Muggeo, V. (2008). Segmented: an R package to fit regression models with broken-line relationships, R package VERSION 1.0.0. <http://www.cran.r-project.org/package=segmented>. Accedido 28 de noviembre de 2016.
- [14] Pardini, R., de Arruda Bueno, A., Gardner, T. A., Prado, P. I., y Metzger, J. P. (2010). Beyond the fragmentation threshold hypothesis: regime shifts in biodiversity across fragmented landscapes. *PloS one*, 5(10), e13666.
- [15] 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(284), 873-880.

- [16] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [17] Ryan, S. E., Porth, L. S., y Troendle, C. A. (2002). Defining phases of bedload transport using piecewise regression. *Earth Surface Processes and Landforms*, 27(9), 971-990.
- [18] Seber, G. A. F, y Wild, C. J. (1989). *Nonlinear Regression*. Wiley, New York.
- [19] Sprent, P. (1961). Some hypotheses concerning two phase regression lines. *Biometrics*, 17(4), 634-645.
- [20] Toms, J. D., y Lesperance, M. L. (2003). Piecewise regression: a tool for identifying ecological thresholds. *Ecology*, 84(8), 2034-2041.
- [21] Ulm, K. (1991). A statistical method for assessing a threshold in epidemiological studies. *Statistics in medicine*, 10(3), 341-349.
- [22] Vélez-Ibarrola, R. y García-Pérez, A. (1993). *Principios de Inferencia Estadística*. UNED, Madrid.
- [23] Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14(4), 1261-1295.