

*Modelos lineales con cambio estructural: una perspectiva
Bayesiana*

YEFERSON ANDRÉS CASTAÑO TAFUR
ESTADÍSTICO



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
2019

*Modelos lineales con cambio estructural: una perspectiva
Bayesiana*

YEFERSON ANDRÉS CASTAÑO TAFUR
ESTADÍSTICO

DISERTACIÓN PRESENTADA PARA OPTAR AL
TÍTULO DE
MAESTRÍA EN CIENCIAS ESTADÍSTICA

DIRECTOR
EDILBERTO CEPEDA CUERVO, PH. D.
DOCTOR EN MATEMÁTICA



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
2019

Título en español

Modelos lineales con cambio estructural: una perspectiva Bayesiana

Title in English

Linear models with structural change: a Bayesian perspective

Resumen: En esta tesis, se presenta una propuesta Bayesiana para la estimación de modelos de regresión lineal normal y modelos longitudinales con cambio estructural. Primero, se introduce la propuesta de estimación de modelos de regresión lineal con un punto de cambio. Luego, se amplía gradualmente este concepto a modelos de regresión con varianza heterocedástica y cambio estructural. Posteriormente, se presenta la propuesta para modelos longitudinales con un punto de cambio y estructuras paramétricas de covarianza SC , $AR(1)$ y $ARMA(1,1)$. La metodología Bayesiana se implementa mediante el uso de simulaciones estocásticas MCMC a través del algoritmo Metropolis-Hastings within Gibbs. Por otro lado, la estimación del punto de cambio se realiza con una búsqueda sobre todos los posibles valores, esta es optimizada por los núcleos de transición propuestos. Después, el rendimiento de los algoritmos es investigado mediante estudios de simulación y se concluye que se detecta el punto de cambio con una gran precisión. Finalmente, los modelos propuestos se aplican sobre datos reales sugeridos en la literatura y se comparan con modelos sin cambio estructural; se encuentra que los modelos propuestos ajustan mejor los datos, para esto se utilizan los estadísticos de bondad de ajuste AIC y BIC , y el análisis de residuales.

Abstract: In this thesis, a Bayesian proposal for the estimation of normal linear regression models and longitudinal models both with structural change is presented. First, the proposal for estimating linear regression models with a change-point is introduced. Afterwards, this concept is gradually extended to regression models with heterocystic variance and structural change. Subsequently, the proposal for longitudinal models with a change-point and parametric covariance structures SC , $AR(1)$ and $ARMA(1,1)$ is being presented. The Bayesian methodology is implemented through the use of MCMC stochastic simulations across the Metropolis-Hastings within Gibbs algorithm. Additionally, the estimation of the change-point is made by a search on all possible values, this is optimized by the Transition Kernel proposed. Then, the performance of the algorithms is analyzed through simulation studies which allows concluding that the change-point is detected with great precision. Finally, the proposed models are applied on real data suggested in the literature and those are compared with models without structural change; it is found that the proposed models fit the data better. Thus, it is necessary to use the AIC and BIC goodness of fit statistics and the residual analysis.

Palabras clave: Cambio estructural, punto de cambio, Bayesiana, heterocedasticidad, datos longitudinales, estructura de covarianza, Metropolis-Hastings

Keywords: Structural change, change-point, Bayesian, heterocedasticity, longitudinal data, covariance structure, Metropolis-Hastings

Dedicado a

A mi abuelita, la luz de mi vida, por su amor y apoyo incondicional.

Agradecimientos

A Dios, por todas sus bendiciones. A mis padres y hermanos, por su apoyo incondicional y enseñarme las cosas buenas que realmente importan en la vida. Un agradecimiento especial a mi padre por sus esfuerzos para ayudarme a alcanzar mis sueños. También quiero agradecer a Nelly, por ser mi mejor amiga y no dejarme derrumbar ante las adversidades, y creer en mí. A mi hermana Alejandra, por su compañía y disposición para escuchar mis quejas y frustraciones. A Ingrid, por su amor y paciencia durante la culminación de esta maestría. A mis amigos, por sus consejos siempre apropiados y respaldo absoluto.

A todos los profesores que me guiaron por este camino académico. Principalmente a mi director de tesis el profesor Edilberto Cepeda, por sus aportes y comentarios permanentes, y las muchas discusiones y reuniones que tuvieron un impacto directo en el resultado final de este trabajo. Un agradecimiento especial a la profesora Luz Mery González por todas sus enseñanzas.

1.3.6. Algoritmo de Metropolis-Hastings	17
1.3.7. Algoritmo caminata aleatoria Metropolis-Hastings	18
2. Modelos de regresión con cambio estructural	20
2.1. Modelos de regresión con un punto de cambio	21
2.1.1. Estimación clásica del modelo	21
2.1.1.1. Ejemplo de estimación clásica del modelo	22
2.1.2. Enfoque por razón de verosimilitud	24
2.1.3. Enfoque por residuales recursivos	26
2.1.4. Enfoque informativo	29
3. Propuestas Bayesianas	31
3.1. Propuesta Bayesiana para modelos lineales con un punto de cambio y va- rianza constante	32
3.2. Propuesta Bayesiana para modelos lineales con un punto de cambio y va- rianza heterogénea	35
3.3. Propuesta Bayesiana para modelos longitudinales con un punto de cambio y estructura de covarianza SC y AR(1)	38
3.4. Propuesta Bayesiana para modelos longitudinales con un punto de cambio y estructura de covarianza ARMA(1,1)	42
4. Simulaciones y aplicaciones	44
4.1. Simulaciones	44
4.1.1. Modelo lineal con un punto de cambio y varianza homocedástica . . .	44
4.1.2. Modelo lineal con un punto de cambio y varianza heterocedástica . .	47
4.1.3. Modelo longitudinal con estructura de covarianza de Simetría Com- puesta	48
4.1.4. Modelo longitudinal con estructura de covarianza Autoregresiva . . .	50
4.1.5. Modelo longitudinal con estructura de covarianza Autoregresiva de Promedio Móvil	52
4.2. Aplicaciones	54
4.2.1. Aplicación: Datos de expansión térmica del cobre	54
4.2.2. Aplicación: Datos de crecimiento de pollos	59
4.2.3. Aplicación: Datos de tolerancia a la insulina en pacientes obesos . . .	63
4.2.3.1. Modelo longitudinal con estructura de covarianza SC	65
4.2.3.2. Modelo longitudinal con estructura de covarianza AR(1) . . .	65
4.2.3.3. Modelo longitudinal con estructura de covarianza AR- MA(1,1)	66

Conclusiones	69
Trabajo futuro	71
Bibliografía	72

Índice de tablas

2.1. Datos generados por Quandt	23
2.2. Log-verosimilitud (2.6) y (2.7) para todos los posibles puntos de cambio τ , asumiendo varianzas diferentes e iguales, respectivamente.	24
4.1. Resumen de resultados a posteriori para la simulación con varianza homocedástica	45
4.2. Resumen de resultados a posteriori para la simulación con varianza heterocedástica	47
4.3. Resumen de resultados a posteriori para la simulación de un modelo longitudinal con estructura de covarianza SC	50
4.4. Resumen de resultados a posteriori para la simulación de un modelo longitudinal con estructura de covarianza $AR(1)$	52
4.5. Resumen de resultados a posteriori para la simulación de un modelo longitudinal con estructura de covarianza $ARMA(1,1)$	54
4.6. Resumen de resultados a posteriori para los datos de expansión térmica del cobre	58
4.7. Resumen de resultados a posteriori para los datos de crecimiento de pollos	60
4.8. Resultados a posteriori para datos de tolerancia a la insulina con estructura de covarianza SC	65
4.9. Resultados a posteriori para datos de tolerancia a la insulina con estructura de covarianza $AR(1)$	66
4.10. Resultados a posteriori para datos de tolerancia a la insulina con estructura de covarianza $ARMA(1,1)$	67
4.11. Criterios de bondad de ajustes para el modelo longitudinal con datos de tolerancia a la insulina en pacientes obesos	67

Índice de figuras

2.1. Gráfico de datos generados por Quandt	23
4.1. Comportamiento de las cadenas para cada uno de los parámetros del estudio de simulación de un modelo de regresión con varianza homocedástica	46
4.2. Histograma de las distribuciones marginales a posteriori en el estudio de simulación de un modelo de regresión con varianza homocedástica	46
4.3. Comportamiento de las cadenas para cada uno de los parámetros del estudio de simulación de un modelo de regresión con varianza heterocedástica, $n = 400$	48
4.4. Histograma de las distribuciones marginales a posteriori en el estudio de simulación de un modelo de regresión con varianza heterocedástica, $n = 400$	49
4.5. Comportamiento de las cadenas para cada uno de los parámetros del estudio de simulación de un modelo longitudinal con estructura de covarianza SC . .	51
4.6. Histograma de las distribuciones marginales a posteriori en el estudio de simulación de un modelo longitudinal SC	51
4.7. Comportamiento de las cadenas para cada uno de los parámetros del estudio de simulación de un modelo longitudinal con estructura de covarianza AR(1)	53
4.8. Histograma de las distribuciones marginales a posteriori en el estudio de simulación de un modelo longitudinal AR(1)	53
4.9. Comportamiento de las cadenas para los parámetros de media del estudio de simulación de un modelo longitudinal con estructura ARMA(1,1)	55
4.10. Comportamiento de las cadenas para los parámetros de covarianza y el punto de cambio en un estudio de simulación de un modelo longitudinal con estructura ARMA(1,1)	55
4.11. Histograma de las distribuciones marginales a posteriori de los parámetros de media en el estudio de simulación de un modelo longitudinal ARMA(1,1)	56
4.12. Histograma de las distribuciones marginales a posteriori de los parámetros de covarianza en el estudio de simulación de un modelo longitudinal ARMA(1,1)	56
4.13. Coeficiente de expansión térmica del cobre en función de la temperatura en grados Kelvin para un estudio NIST	57

4.14. Comportamiento de las cadenas para cada uno de los parámetros del modelo de expansión del cobre	58
4.15. Histograma de las distribuciones marginales a posteriori para el modelo de expansión del cobre	59
4.16. Gráfico de datos de crecimiento de pollos	60
4.17. Comportamiento de las cadenas para cada uno de los parámetros del modelo de crecimiento de pollos	61
4.18. Histograma de las distribuciones marginales a posteriori para el modelo de crecimiento de pollos	62
4.19. Residuales para los modelos de crecimiento de pollos	62
4.20. Gráfico de perfiles para datos de tolerancia a la insulina en pacientes obesos	63
4.21. Dispersograma datos de tolerancia a la insulina en pacientes obesos para los diferentes tiempos de medición en pacientes obesos	64
4.22. Comportamiento de las cadenas para cada uno de los parámetros del modelo de los datos de tolerancia a la insulina con estructura de covarianza $AR(1)$.	68
4.23. Histograma de las distribuciones marginales a posteriori para el modelo de los datos de tolerancia a la insulina con estructura de covarianza $AR(1)$. . .	68

Introducción

Es común pensar que los modelos de regresión explican todo el conjunto de datos, pero esto no siempre es cierto. Por el contrario, los datos pueden cambiar después de un punto específico que los divide en dos segmentos. Lo cual indica, que en muchos casos cuando se modelan datos con las metodologías de regresión clásicas, no es posible encontrar un ajuste adecuado, ni un modelo capaz de describir el comportamiento de los datos. En modelos lineales clásicos el conjunto de observaciones se denota como un vector $\mathbf{y} = (y_1, y_2, \dots, y_n)'$. El conjunto de variables explicativas o covariables se organiza como una matriz $\mathbf{X}_{n \times k}$. Cada fila $\mathbf{x}'_i = (x_{i1}, \dots, x_{ik})$ de \mathbf{X} refiere a diferentes observaciones, y cada columna a diferentes covariables. El conjunto de coeficientes de regresión es un vector denotado por $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ y $\boldsymbol{\epsilon}$ de tamaño $n \times 1$, se conoce como el vector de errores. Formalmente, se tiene la siguiente relación entre estos elementos

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon};$$

con media $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ y se supone que el vector $\boldsymbol{\epsilon}$ tiene componentes ϵ_i , normalmente, idénticamente e independientemente distribuidos, cada uno con media cero y varianza común σ^2 . Cuando se presentan un cambio en la estructura de la media y heterogeneidad de varianza, estas premisas no se cumplen, el problema de cambio en la estructura de la media se puede solucionar ajustando modelos lineales con cambio estructural. Por otro lado, la violación del supuesto de homogeneidad de varianza se puede solucionar con una transformación (Box & Cox, 1964) de la variable respuesta. Cuando esto no ayuda a estabilizar la varianza es conveniente hacer un análisis de modelado explícito de la varianza (Harvey, 1976).

Por lo anterior, surge la necesidad de introducir los modelos de regresión con cambio estructural, en donde el cambio se puede presentar en la estructura de media, varianza o en ambas. Con frecuencia, factores difíciles de controlar como cambios en políticas económicas, guerras, cambios repentinos en las dinámicas de los mercados, entre muchos otros, pueden conducir cambios estructurales y heterogeneidad en los datos. Estos modelos han ganado gran atención no solo en análisis econométrico y series temporales, sino también en muchas otras áreas como en ciencias sociales, físicas, químicas y biológicas (Chaturvedi & Shrivastava, 2016). Para detectar el cambio estructural, los datos deben tener algún orden natural, puede ser el tiempo de observación o alguna otra variable (por ejemplo, tiempo, espacio, edad, entre otros). En esta investigación, un cambio estructural es entendido como un cambio en los parámetros que describen la media o la varianza, pero la función de enlace entre la respuesta y las variables explicativas permanece igual.

Muchos autores han abordado el problema del punto de cambio asociado a un modelo de regresión. Quandt (1958) y Quandt (1960) determinó una prueba de razón de verosimilitud para probar la necesidad de un punto de cambio y estimar un modelo de regresión lineal en dos regímenes diferentes. Hinkley (1970) derivó la distribución asintótica del estadístico de la prueba de razón de verosimilitud y lo aplicó para detectar un cambio en la media dentro de una muestra de observaciones distribuidas normalmente. Ferreira (1975) estudió el modelo de regresión con cambio estructural desde el punto de vista Bayesiano. Brown et al. (1975) presentaron una metodología basada en residuales recursivos para probar la existencia de puntos de cambio en modelos de regresión múltiple. Posteriormente, aparecen los modelos con cambio estructural empleados en econometría por Bai (1997), Bai & Perron (1998), Bai & Perron (2003) en donde la estimación de los parámetros se realiza por el método de mínimos cuadrados, los errores pueden ser homocedásticos o heterocedásticos y dependientes sobre el tiempo. Chen & Gupta (2012) abordaron el tema del punto de cambio para el modelo de regresión lineal simple y múltiple mediante el uso del criterio de información de Schwarz y por un enfoque Bayesiano.

Los datos longitudinales consisten en observaciones repetidas de una característica de interés en múltiples unidades de observación. La respuesta generalmente se mide sobre puntos discretos en el tiempo, aunque también se podría medir en puntos en el espacio unidimensional. Es posible observar una o más covariables en los individuos, éstas pueden ser dependientes del tiempo, independientes del tiempo o una combinación de ambas (Zimmerman & Núñez Antón, 2009). Los métodos estadísticos para el análisis de datos longitudinales han existido desde hace mucho tiempo. Sin embargo, es importante resaltar que los modelos paramétricos para la matriz de covarianza y los modelos de efectos aleatorios se han desarrollado de forma paralela, pero se consideran típicamente distintos. Algunos de estos modelos son presentados por Davis (2002), Diggle et al. (2002), Verbeke & Molenberghs (2009) y Zimmerman & Núñez Antón (2009). En esta investigación se estudia el ajuste de modelos de regresión con estructura longitudinal en los cuales se especifica la estructura de covarianza.

En el análisis de datos longitudinales el cambio se puede presentar de forma natural. Además, las intervenciones específicas también pueden causar cambios, por ejemplo, los niveles de colesterol pueden disminuir con nuevos medicamentos; los puntajes de alguna prueba pueden aumentar después de un entrenamiento. Al medir y trazar cambios tanto naturales como inducidos experimentalmente, se descubre la naturaleza temporal del desarrollo. Singer & Willett (2003) menciona que no todos los estudios longitudinales son adecuados para realizar análisis de cambio, estos deben compartir tres características metodológicas que los hacen particularmente idóneos para esta tarea, cada uno debe tener: (i) dos o más segmentos de datos, (ii) un resultado cuyos valores cambien sistemáticamente con el tiempo y (iii) una métrica razonable de medición en el tiempo.

Singer & Willett (2003) utilizan los modelos multinivel para el cambio, explican cómo estos permiten abordar los efectos intra y entre individuos simultáneamente sobre el cambio. Definen un submodelo de nivel 1 que describe cómo cambia cada individuo con el tiempo y un submodelo de nivel 2 que relaciona cómo estos cambios difieren entre individuos. Además, presentan una definición compuesta del modelo multinivel que combina los submodelos de nivel 1 y nivel 2 en una sola ecuación, para este describen los métodos de estimación por mínimos cuadrados generalizados y mínimos cuadrados generalizados iterativos y distinguen los enfoques completo y restringido. Por otro lado, exponen estrategias para ajustar modelos en los que el cambio individual es explícitamente discontinuo o no lineal sobre el tiempo, estos autores describen formas de incorporar discontinuidades

abruptas en la trayectoria de los individuos. Además, amplían estos conceptos al especificar trayectorias que son funciones polinómicas del tiempo y examinan varios modelos de cambio en los que el resultado es una función explícita no lineal de los parámetros de crecimiento. Finalmente, se centran en los efectos aleatorios del modelo multinivel para el cambio, tal como están incorporados en la estructura de covarianza, para esto examinan los efectos aleatorios y demuestran que el término de error compuesto es heterocedástico y autocorrelacionado, perfecto para el análisis de datos longitudinales. Comparan varias estructuras de covarianza y proporcionan herramientas para elegir la que mejor se ajuste a los datos.

Gamerman (1997) propone una metodología de estimación bayesiana por bloques de parámetros, para el ajuste sobre modelos generalizados mixtos. Cepeda & Gamerman (2001) ajustan de manera conjunta la media y la estructura de varianza para un modelo de regresión lineal heterocedástico, empleando a prioris normales para los parámetros de la media y la varianza. Una generalización de la propuesta de estimación conjunta de la estructura de media y varianza es presentada por Cepeda (2001), Cepeda & Gamerman (2004) y Cepeda & Gamerman (2005). Por otro lado, Daniels & Pourahmadi (2002) desarrollaron propuestas bayesianas para ajustar modelos dinámicos longitudinales suponiendo a prioris normales para los parámetros de media y varianza, además emplearon la Descomposición de Cholesky presentada por Pourahmadi (1999). Cepeda & Núñez Antón (2007) y Cepeda & Núñez Antón (2009) proponen metodologías bayesianas para ajustar de manera conjunta los parámetros de media y varianza en modelos longitudinales, para la estructura de varianza suponen un modelo de antedependencia no estructurado, este enfoque es aplicado sobre modelos lineales y no lineales. Más recientemente, Castillo et al. (2019) presenta un propuesta bayesiana para la estimación conjunta de los parámetros de media y la matriz de varianzas covarianzas para modelos longitudinales con estructuras de covarianza SC , $AR(1)$, $ARMA(1, 1)$ y algunos casos de estructuras de antedependencia no estructurada $ADS(1)$.

Esta tesis se presenta en cuatro capítulos, inicia con el abordaje de los modelos de regresión lineal con varianza homocedástica y heterocedástica, además se presenta la metodología de estimación de los parámetros desde la perspectiva clásica y bayesiana. Continúa con la definición del modelo de regresión para datos longitudinales y algunas estructuras de covarianza, se exponen las metodologías de estimación clásica y bayesiana. Finalmente, se describen las características de la inferencia bayesiana y presentan los algoritmos de simulación estocástica que se utilizan a lo largo de la tesis.

En el capítulo dos se presenta el modelo de regresión con cambio estructural y se expone la metodología clásica de estimación de detección del punto de cambio desde diferentes enfoques: razón de verosimilitud, residuales recursivos y criterio de información de Schwarz. Posteriormente, en el capítulo tres se presentan las propuestas bayesianas para la estimación conjunta de la media, varianza y punto de cambio en modelos de regresión homocedásticos y heterocedásticos, y modelos longitudinales con estructura de covarianza paramétrica SC , $AR(1)$ y $ARMA(1, 1)$.

En el capítulo cuatro se ejemplifica la funcionalidad de las propuestas bayesianas a través de simulaciones y aplicaciones con datos reales para los modelos de regresión lineal normal con varianza homocedástica y heterocedásticos, y modelos longitudinales. Se implementan distribuciones a priori no informativas que permiten mostrar la ganancia en el ajuste de los diferentes modelos al incluir el punto de cambio. Finalmente, en el capítulo 5 se presentan las conclusiones del trabajo y algunos ideas para desarrollar en el futuro.

Marco teórico

En este capítulo se hace un resumen del modelo lineal, se incluyen elementos que abordan la estimación por máxima verosimilitud y bayesiana de los parámetros. Se introducen los modelos de regresión heterocedásticos y su estimación clásica y bayesiana. Posteriormente, se presentan los modelos de regresión para datos longitudinales, algunas estructuras paramétricas de covarianza que se utilizan a lo largo del documento y las metodologías de estimación de los parámetros desde los enfoques clásico y bayesiano. Finalmente, se realiza un resumen sobre inferencia bayesiana y se exponen algunos algoritmos de simulación estocástica.

1.1. Modelo de regresión normal

En modelos lineales clásicos, un conjunto de datos es denotado por un vector de observaciones $\mathbf{Y} = (y_1, \dots, y_n)'$. El conjunto de covariables o variables explicativas son denotadas como una matriz \mathbf{X} de tamaño $n \times p$. Cada fila de \mathbf{X} representa una observación, y cada columna a una covariable. El conjunto de parámetro es un vector denotado por $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ y $\boldsymbol{\varepsilon}$ es un vector de errores observados. Se obtiene la siguiente relación entre estos elementos:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1.1}$$

La ecuación (1.1) es conocida como modelo de regresión lineal. Cuando $p = 1$, es decir, se cuenta con tan solo una variable explicativa, se obtiene un modelo de regresión lineal simple. Adicionalmente, (1) si se asume varianza constante $V(y_i) = \sigma^2$ y los errores después de observada la muestra se obtienen como

$$e_i = y_i - (\beta_1 x_{1i} + \dots + \beta_p x_{pi}), \quad i = 1, \dots, n,$$

entonces los parámetros $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ pueden ser estimados por Mínimos Cuadrados. Por lo tanto, el estimador es dado por el valor de $\boldsymbol{\beta}$ que minimiza $\sum_{i=1}^n e_i^2$, en este caso no es necesario realizar supuestos sobre la distribución de y_i o los errores.

Si además del supuesto de varianza constante, (2) se asume la hipótesis de normalidad e independencia de la distribución de y_i , $i = 1, \dots, n$, con $E(y_i) = \mu_i$, es posible obtener

la función de verosimilitud de la forma

$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\}. \quad (1.2)$$

y el logaritmo de la función de verosimilitud es,

$$l(\mathbf{y}; \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (1.3)$$

1.1.1. Estimación clásica del modelo

Para $\boldsymbol{\beta}$ se tiene que las estimaciones por mínimos cuadrados y máxima verosimilitud coinciden y son dados por los valores de $\boldsymbol{\beta}$ que minimizan la forma cuadrática $S(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$. Diferenciando $S(\boldsymbol{\beta})$ con respecto a el vector $\boldsymbol{\beta}$ se obtiene

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 2(\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - \mathbf{X}'\mathbf{Y})$$

Cuando la matriz de segundas derivadas es definida positiva, la solución de $\partial S(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = \mathbf{0}$ define el mínimo $\hat{\boldsymbol{\beta}}$ de $S(\boldsymbol{\beta})$ y produce el conjunto de ecuaciones normales $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$, estas tienen solución si la matriz $\mathbf{X}'\mathbf{X}$ es de rango completo, o si las columnas de \mathbf{X} son linealmente independientes. La estimación de máxima verosimilitud de $\boldsymbol{\beta}$ es dado por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (1.4)$$

Para la estimación de σ^2 , la forma cuadrática $S(\boldsymbol{\beta})$ se puede expresar como

$$S(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + S_e,$$

donde

$$\begin{aligned} S_e &= \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}' \mathbf{X}'\mathbf{X} \hat{\boldsymbol{\beta}} \\ &= (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \end{aligned}$$

Por lo tanto, la estimación por máxima verosimilitud de σ^2 se obtiene al solucionar la siguiente ecuación

$$\frac{\partial l(\hat{\boldsymbol{\beta}}, \sigma^2; \mathbf{Y})}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\hat{\sigma}^2} + \frac{S(\hat{\boldsymbol{\beta}})}{2\hat{\sigma}^4} = 0,$$

la estimación de máxima verosimilitud para σ^2 está dado por

$$\hat{\sigma}^2 = \frac{S_e}{n} = \frac{1}{n} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (1.5)$$

Se tiene que $\hat{\beta}$ es una función lineal de \mathbf{Y} , por tanto, la distribución muestral de $\hat{\beta}$ es una normal multivariada con media y varianza dadas por

$$\begin{aligned} E(\hat{\beta}|\beta, \sigma^2) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}|\hat{\beta}, \sigma^2) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta \\ &= \beta \\ \text{Var}(\hat{\beta}|\beta, \sigma^2) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{Y}|\hat{\beta}, \sigma^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

luego, este es un estimador insesgado. Además, la forma cuadrática dada por $[S(\beta) - S_e]/\sigma^2$ tiene distribución muestral χ_p^2 . Por otra parte, se tiene que S_e es independiente de $\hat{\beta}$ y por consiguiente independiente de $S(\beta) - S_e$. Luego $S(\beta)/\sigma^2 \sim \chi_n^2$, $S_e/\sigma^2 \sim \chi_{n-p}^2$, por lo tanto, $s^2 = S_e/(n-p)$ es un estimador insesgado de σ^2 .

1.1.2. Estimación Bayesiana del modelo

Para implementar una metodología Bayesiana con el objetivo de estimar los parámetros β y σ^2 del modelo de regresión (1.1) es necesario especificar una distribución a priori para ellos. Se asume que β tiene una distribución a priori condicional $N(\mathbf{b}_0, (\phi\mathbf{B}_0)^{-1})$ donde $\phi = \sigma^{-2}$ y que $v_0\lambda_0\phi \sim \chi_{v_0}^2$. Por lo tanto, la densidad de la distribución a priori está dada por (Migon & Gamerman, 1999)

$$\begin{aligned} p(\beta, \phi) &= (2\pi)^{-p/2}|\phi\mathbf{B}_0|^{1/2}\exp\left\{-\frac{\phi}{2}(\beta - \mathbf{b}_0)'\mathbf{B}_0(\beta - \mathbf{b}_0)\right\} \\ &\quad \times \frac{(n_0\sigma_0^2/2)^{n_0/2}}{\Gamma(n_0/2)}\phi^{(n_0/2)-1}\exp\left\{-\frac{n_0\sigma_0^2}{2}\phi\right\} \\ &\propto \phi^{[(n_0+p)/2]-1}\exp\left\{-\frac{\phi}{2}[n_0\sigma_0^2 + (\beta - \mathbf{b}_0)'\mathbf{B}_0(\beta - \mathbf{b}_0)]\right\} \end{aligned}$$

Entonces, la distribución condicional de $\phi|\beta$ se puede obtener de la distribución a priori conjunta de β y ϕ tomando solo los términos que involucran a ϕ . Esta es dada por $[n_0\sigma_0^2 + (\beta - \mathbf{b}_0)'\mathbf{B}_0(\beta - \mathbf{b}_0)]\phi|\beta \sim \chi_{n_0+p}^2$. Se tiene que la distribución marginal de β está dada por

$$p(\beta) \propto [n_0\sigma_0^2 + (\beta - \mathbf{b}_0)'\mathbf{B}_0(\beta - \mathbf{b}_0)]^{(n_0+p)/2},$$

que corresponde a la función de densidad de una distribución $t_{n_0}(\mathbf{b}_0, \sigma_0^2\mathbf{B}_0^{-1})$. Por otra parte, la función de verosimilitud de β y ϕ está dada por

$$\begin{aligned} L(\beta, \phi) &= (\phi/2\pi)^{n/2}\exp\left\{-\frac{\phi}{2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\right\} \\ &\propto \phi^{n/2}\exp\left\{-\frac{\phi}{2}[S_e + (\beta - \hat{\beta})'\mathbf{X}'\mathbf{X}(\beta - \hat{\beta})]\right\} \end{aligned} \quad (1.6)$$

se observa que 1.6 tiene la misma forma de la distribución a priori. Entonces, la distribución a posteriori está dada por

$$\begin{aligned} p(\boldsymbol{\beta}, \phi | \mathbf{y}) &\propto \phi^{[(n+n_0+p)/2]-1} \\ &\times \exp \left\{ -\frac{\phi}{2} [n_0\sigma_0^2 + S_e + (\boldsymbol{\beta} - \mathbf{b}_0)' \mathbf{B}_0 (\boldsymbol{\beta} - \mathbf{b}_0) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})] \right\} \\ &\propto \phi^{p/2} \exp \left\{ -\frac{\phi}{2} (\boldsymbol{\beta} - \mathbf{b}_1)' \mathbf{B}_1 (\boldsymbol{\beta} - \mathbf{b}_1) \right\} \phi^{(n_1/2)-1} \exp \left\{ -\frac{\phi}{2} n_1 \sigma_1^2 \right\} \end{aligned} \quad (1.7)$$

donde $\mathbf{b}_1 = \mathbf{B}_1^{-1}(\mathbf{B}_0 \mathbf{b}_0 + \mathbf{X}' \mathbf{y})$, $\mathbf{B}_1 = \mathbf{B}_0 + \mathbf{X}' \mathbf{X}$, $n_1 = n + n_0$ y $n_1 \sigma_1^2 = n_0 \sigma_0^2 + (\mathbf{y} - \mathbf{X} \mathbf{b}_1)' \mathbf{y} + (\mathbf{b}_0 - \mathbf{b}_1)' \mathbf{B}_0 \mathbf{b}_0$. La distribución a posteriori para $\boldsymbol{\beta}$ está dada por $\boldsymbol{\beta} | \mathbf{y} \sim t_{n_1}(\mathbf{b}_1, \sigma_1^2 \mathbf{B}_1^{-1})$, la media y matriz de varianza-covarianza a posteriori de $\boldsymbol{\beta}$ están dadas, respectivamente, por \mathbf{b}_1 y $\frac{n_1}{n_1 - 2} \sigma_1^2 \mathbf{B}_1^{-1}$. Por otro lado, la distribución a posteriori de ϕ es $n_1 \sigma_1^2 \phi | \mathbf{y} \sim \chi_{n_1}^2$ con media σ_1^{-2} .

La distribución a priori no informativa se puede implementar para representar la ausencia de información inicial (Migon & Gamerman, 1999). Se tiene que $\boldsymbol{\beta}$ es un parámetro de localización y ϕ es un parámetro de escala, por lo tanto, la distribución a priori conjunta no informativa está dada por $p(\boldsymbol{\beta}, \phi) \propto \phi^{-1}$. Realizando algunos desarrollos algebraicos se obtiene que la función de densidad a posteriori es

$$\begin{aligned} p(\boldsymbol{\beta}, \phi | \mathbf{y}) &\propto \phi^{(n/2)-1} \exp \left\{ -\frac{\phi}{2} [S_e + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})] \right\} \\ &\propto \phi^{p/2} \exp \left\{ -\frac{\phi}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\} \phi^{((n-p)/2)-1} \exp \left\{ -\frac{\phi}{2} (n-p) s^2 \right\}, \end{aligned}$$

donde $s^2 = S_e / (n - p)$ es un estimador insesgado de σ^2 . Por lo tanto, la distribución a posteriori es similar a (1.7), los cambios se encuentran en los valores de los hiperparámetros de las distribuciones a priori. Así, $\boldsymbol{\beta} | \mathbf{y} \sim t_{n-p}(\hat{\boldsymbol{\beta}}, s^2 (\mathbf{X}' \mathbf{X})^{-1})$ y $(n-p) s^2 \phi | \mathbf{y} \sim \chi_{n-p}^2$. Con esto se obtienen resultados Bayesianos paralelos a los obtenidos usando una enfoque de estimación clásico.

1.2. Modelo de regresión normal con varianza heterogénea

La homogeneidad de varianza es uno de los supuestos estándar en el análisis de regresión normal. Cuando este no se cumple, puede a veces ser rectificado después de aplicar una transformación Box & Cox (1964) de la variable respuesta. Como esto no siempre sucede, es conveniente considerar un análisis explícito de la varianza. Además, el modelo puede extenderse para incluir posibles explicaciones de heterogeneidad de la varianza.

Se considera el modelo de regresión normal en donde se modela a través de covariables la heterogeneidad de la varianza. Esto significa que

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (1.8)$$

$\epsilon_i \sim N(0, \sigma_i^2)$, donde $\boldsymbol{\mu} = \mathbf{X} \boldsymbol{\beta}$ es el modelo de media y $g(\sigma_i^2) = \mathbf{z}_i' \boldsymbol{\gamma}$ es el modelo de varianza, $i = 1, \dots, n$, \mathbf{X} es una matriz $n \times p$ de variables explicativas de la media y $\mathbf{z}_i = (z_{i1}, \dots, z_{ir})$ un vector de variables explicativas de la varianza que puede contener algunas o todas las variables incluidas en \mathbf{X} . $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ y $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_r)'$ son

los vectores de parámetros de los modelos de media y de varianza, respectivamente. La función g debe ser monótona, diferenciable y debe garantizar que la varianza sea positiva. La opción más común es considerar $g = \log$.

1.2.1. Estimación clásica del modelo

Dadas las observaciones (y_i, x_i, z_i) , $i = 1, \dots, n$, siguiendo el modelo 1.8 con $g = \log$, el kernel de la función de verosimilitud es

$$L(\beta, \gamma) = \prod_{i=1}^n \frac{1}{\sigma_i} \exp \left\{ -\frac{1}{2\sigma_i^2} (y_i - x_i' \beta)^2 \right\}, \quad (1.9)$$

con $\log(\sigma_i^2) = z_i' \gamma$.

La matriz de información observada para el vector de parámetros completo (β, γ) es

$$J = \begin{bmatrix} \mathbf{X}' \mathbf{W}_{11} \mathbf{X} & \mathbf{X}' \mathbf{W}_{12} \mathbf{Z} \\ \mathbf{Z}' \mathbf{W}_{12} \mathbf{X} & \mathbf{Z}' \mathbf{W}_{12} \mathbf{Z} \end{bmatrix},$$

donde $\mathbf{Z}' = (z_1, \dots, n)$, $\mathbf{W}_{11} = \text{diag}(1/\hat{\sigma}_i^2)$, $\mathbf{W}_{12} = \text{diag}(\hat{e}_i/\hat{\sigma}_i^2)$, $\mathbf{W}_{22} = 1/2 \text{diag}(\hat{e}_i^2/\hat{\sigma}_i^2)$, $\hat{e}_i = \mathbf{Y}_i - x_i' \hat{\beta}$ y $\hat{\sigma}_i^2 = \exp(z_i' \hat{\gamma})$, donde $\hat{\beta}$ y $\hat{\gamma}$ son las estimaciones de máxima verosimilitud simultaneas de β y γ . Entonces, la matriz de información esperada estimada es $\hat{I} = \hat{E}(J)$ dada por

$$\hat{I} = \begin{bmatrix} \mathbf{X}' \mathbf{W}_{11} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2} \mathbf{Z}' \mathbf{Z} \end{bmatrix}.$$

Por lo tanto, un algoritmo de Fisher Scoring para la estimación de máxima verosimilitud simultanea de β y γ conduce a dos algoritmos separados para β y γ . Como \mathbf{W}_{11} depende de γ y \hat{e}_i depende de β , el algoritmo scoring completo se puede formular como un algoritmo iterativo alternado (Aitkin, 1987). Cuando el algoritmo es iterado hasta obtener convergencia, proporciona simultáneamente estimaciones de máxima verosimilitud $(\hat{\beta}, \hat{\gamma})$ y la matriz de información esperada estimada \hat{I} evaluada en la estimación de máxima verosimilitud. La teoría asintótica es usada para obtener distribuciones aproximadas de las estimaciones de máxima verosimilitud como

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \sim N \left[\begin{pmatrix} \beta \\ \gamma \end{pmatrix}, \hat{I}^{-1} \right].$$

Esta distribución aproximada es usada para construir intervalos de confianza o regiones para funciones de β y γ , este procedimiento fue sugerido por Harvey (1976).

1.2.2. Estimación Bayesiana del modelo

Para implementar una metodología Bayesiana con el objetivo de estimar los parámetros del modelo 1.8 es necesario especificar una distribución a priori para ellos. Por simplicidad se define la distribución a priori $p(\beta, \gamma)$ como

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} \sim \left[\begin{pmatrix} \mathbf{b}_0 \\ \mathbf{g}_0 \end{pmatrix}, \begin{pmatrix} \mathbf{B}_0 & \mathbf{C}_0 \\ \mathbf{C}'_0 & \mathbf{G}_0 \end{pmatrix} \right] \quad (1.10)$$

Entonces, empleando el teorema de Bayes, se obtiene la distribución condicional completa $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma})$ como

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}) &\propto L(\boldsymbol{\beta}, \boldsymbol{\gamma})p(\boldsymbol{\beta}, \boldsymbol{\gamma}) \\ &\propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right\} \end{aligned} \quad (1.11)$$

donde $\boldsymbol{\Sigma} = \text{diag}(\sigma_i^2)$, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})'$ y $\boldsymbol{\theta}_0 = (\mathbf{b}_0, \mathbf{g}_0)'$.

Como 1.11 es intratable analíticamente, se aplica un procedimiento de inferencia en donde se utilizan muestras aproximadas de $\boldsymbol{\theta}$, basado en una metodología de muestreo por bloques de parámetros propuesta por Cepeda & Gamerman (2001). Entonces, resulta importante obtener las distribuciones condicionales completas para los bloques $\boldsymbol{\beta}$ y $\boldsymbol{\gamma}$. Estas son respectivamente dados por π_β y π_γ .

La distribución condicional completa π_β está dada por

$$\pi_\beta \sim N(\mathbf{b}_1, \mathbf{B}_1), \quad (1.12)$$

con $\mathbf{b}_1 = \mathbf{B}_1(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y})$ y $\mathbf{B}_1 = (\mathbf{B}_0^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$, $\boldsymbol{\Sigma} = \text{diag}(\sigma_i^2)$ donde $\sigma_i^2 = \exp(\mathbf{z}'_i\boldsymbol{\gamma})$, $i = 1, \dots, n$. Por lo tanto, para obtener los valores estimados de β empleando el muestreador de Gibbs, se tomaran muestras de la distribución 1.12 con probabilidad de aceptación 1.

A diferencia de la distribución condicional completa de $\boldsymbol{\beta}$, la distribución condicional completa de $\boldsymbol{\gamma}$ es intratable analíticamente y no es fácil de generar. En este caso Cepeda & Gamerman (2001) utiliza la metodología de Gamerman (1997) y elementos del algoritmo de Fisher Scoring para construir propuestas adecuadas.

El algoritmo requiere variables de trabajo para aproximar la transformación de las observaciones en torno a las estimaciones de los parámetros actuales. Para la iteración de $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ se fija en el valor actual $\boldsymbol{\beta}^{(j)}$ y el modelo observacional es (Cepeda & Gamerman, 2001)

$$t_i = (y_i - \mathbf{x}'_i\boldsymbol{\beta}^{(j)})^2 \sim \sigma_i^2\chi_1^2, \text{ para } i = 1, \dots, n.$$

Entonces, las observaciones t_i tienen media $E(t_i) = \sigma_i^2$, varianza $\text{Var}(t_i) = 2\sigma_i^4$ y están relacionadas al parámetro de regresión $\boldsymbol{\gamma}$ a través de $g[E(t_i)] = \mathbf{z}'_i\boldsymbol{\gamma}$. Dada la diferenciabilidad de g , $g(t_i) \simeq g[E(t_i)] + g'[E(t_i)][t_i - E(t_i)] := \tilde{y}_i$ y $\text{Var}[g(t_i)] \simeq \text{Var}(\tilde{y}_i) = \text{Var}\{g[E(t_i)] + g'[E(t_i)][t_i - E(t_i)]\} = g'[E(t_i)]\text{Var}(t_i)$. Por lo tanto, si $\boldsymbol{\beta}^{(j)}$ y $\boldsymbol{\gamma}^{(j)}$ son los valores actuales de $\boldsymbol{\beta}$ y $\boldsymbol{\gamma}$, la observación de trabajo apropiada es

$$\tilde{y}_i = \mathbf{z}'_i\boldsymbol{\gamma}^{(j)} + g'[g^{-1}(\mathbf{z}'_i\boldsymbol{\gamma}^{(j)})][(y_i - \mathbf{x}'_i\boldsymbol{\beta}^{(j)})^2 - g^{-1}(\mathbf{z}'_i\boldsymbol{\gamma}^{(j)})]$$

tiene varianza de trabajo asociada $\text{Var}(\tilde{y}_i) = \{g'[g^{-1}(\mathbf{z}'_i\boldsymbol{\gamma}^{(j)})]\}^2\text{Var}(t_i) = 2\{g'[g^{-1}(\mathbf{z}'_i\boldsymbol{\gamma}^{(j)})]g^{-1}(\mathbf{z}'_i\boldsymbol{\gamma}^{(c)})\}^2$. Donde $g = \log$, las expresiones anteriores se simplifican y el vector de observaciones de trabajo es $\tilde{\mathbf{Y}} = (\tilde{y}_1, \dots, \tilde{y}_n)$, con

$$\tilde{y}_i = \mathbf{z}'_i\boldsymbol{\gamma}^{(j)} + \frac{(y_i - \mathbf{x}'_i\boldsymbol{\beta}^{(j)})^2}{\exp(\mathbf{z}'_i\boldsymbol{\gamma}^{(j)})} - 1, \quad i = 1, \dots, n,$$

y varianza de trabajo asociada igual a 2. El kernel de transición q_γ se obtiene como

$$q_\gamma(\boldsymbol{\gamma}^{(j)}, \boldsymbol{\gamma}^{(n)}) = N(\mathbf{g}_1, \mathbf{G}_1)$$

donde $\mathbf{g}_1 = \mathbf{G}_1(\mathbf{G}_0^{-1}\mathbf{g}_0 + 1/2\mathbf{Z}'\tilde{\mathbf{Y}})$ y $\mathbf{G}_1 = (\mathbf{G}_0^{-1} + 1/2\mathbf{Z}'\mathbf{Z})^{-1}$. Esta distribución propuesta es la misma distribución a posteriori de $\boldsymbol{\gamma}$ que se habría obtenido de la combinación de la a priori con el modelo con las observaciones de trabajo (Cepeda & Gamerman, 2001)

$$\tilde{y}_i \sim N(\mathbf{z}_i'\boldsymbol{\gamma}, 2), \quad i = 1, \dots, n.$$

Esta idea fue introducida por Gamerman (1997) para definir un algoritmo basado en el muestreo de propuestas generales para realizar la inferencia Bayesiana en problemas de modelos lineales generalizados y extendido por Cepeda & Gamerman (2001) para el análisis de modelos lineales con varianza heterogénea.

1.3. Datos longitudinales

Los datos longitudinales, consisten en observaciones repetidas de una característica de interés en múltiples unidades de observación. La respuesta puede ser categórica, discreta o continua; además, esta puede ser univariada o multivariada (Zimmerman & Núñez Antón, 2009). Sin embargo, en este documento, supondremos que la respuesta es continua univariada. La respuesta generalmente se mide en puntos discretos en el tiempo, aunque también se podría medir en puntos en el espacio unidimensional. Es posible observar una o más covariables en los individuos, éstas pueden ser dependientes del tiempo, independientes del tiempo o una combinación de ambas (Zimmerman & Núñez Antón, 2009).

Los métodos estadísticos para el análisis de datos longitudinales han existido desde hace mucho tiempo. Sin embargo, es importante resaltar que los modelos paramétricos para la matriz de varianzas covarianzas y los modelos de efectos aleatorios se han desarrollado de forma paralela, pero se consideran típicamente distintos. Algunos de estos modelos son presentados por Davis (2002), Diggle et al. (2002), Verbeke & Molenberghs (2009) y Zimmerman & Núñez Antón (2009).

En los modelos de efectos aleatorios se ve el origen de la estructura de covarianza como regresiones que varían entre los individuos o unidades de observación en el estudio (Núñez Antón & Zimmerman, 2001). Por otro lado, los modelos en donde se definen estructuras paramétricas para la matriz de varianzas covarianzas, tiene algunas ventajas las cuales son descritas por Núñez Antón & Zimmerman (2001). Sin embargo, los modelos de efectos aleatorios pueden emplearse y ser de mucha utilidad en el análisis de datos longitudinales. En este documento, se estudia el ajuste de modelos de regresión lineal con estructura longitudinal en los cuales se especifica la estructura de covarianza.

1.3.1. Modelo de regresión para datos longitudinales

Se realizan m mediciones a lo largo del tiempo en cada uno de los n individuos o unidades de observación. Sea \mathbf{y}_{it} la medición realizada en el i -ésimo individuo en el tiempo t . Sea $\mathbf{y}_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{im})'$ el vector de las m mediciones que se han realizado en el i -ésimo individuo y sea $\mathbf{t}_i = (t_{i1}, t_{i2}, \dots, t_{im})'$ el vector que contiene los tiempos correspondientes en los que se han realizado dichas mediciones. Además, se supone que se observa un vector p variante de covariables, \mathbf{x}_{ij} , asociado con \mathbf{y}_{ij} . En forma matricial, sean $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_n)'$, $\mathbf{t} = (\mathbf{t}'_1, \mathbf{t}'_2, \dots, \mathbf{t}'_n)'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{im})'$, $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_n)'$ y $N = nm$.

Según Núñez Antón & Zimmerman (2001), las hipótesis generales para el modelo paramétrico de varianza y covarianza son las siguientes:

1. Las mediciones realizadas en individuos distintos son independientes entre sí, aunque en un mismo individuo podrán ser dependientes.
2. La respuesta media para los individuos, es una combinación lineal de funciones conocidas de las covariables o del tiempo.
3. Las respuestas están normalmente distribuidas.
4. La naturaleza de la variación de segundo orden intra-individuos es la misma para todos los individuos.
5. Si hay datos perdidos, estos se asumen ignorables.

De lo anterior, se obtiene el modelo $\mathbf{y} \sim NMV(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, donde $\boldsymbol{\beta}$ es un vector p -dimensional de parámetros fijos y desconocidos, $\boldsymbol{\Sigma} = \sigma^2\mathbf{V}$ es una matriz diagonal por bloques con bloques no nulos $\sigma^2\mathbf{V}_0$ de tamaño $m \times m$, donde cada bloque no nulo representa la matriz de varianzas y covarianzas para cada individuo.

Para obtener una modelización paramétrica de la estructura de covarianzas, es necesario establecer un modelo para $\boldsymbol{\Sigma}$,

$$\mathbf{y} \sim NMV(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\mathbf{t}, \boldsymbol{\theta})), \quad (1.13)$$

donde $\boldsymbol{\theta}$ es un vector q -dimensional de parámetros desconocidos, restringido a un espacio de parámetros $\boldsymbol{\Theta}$, el cual es el conjunto de todos los vectores $\boldsymbol{\theta}$ para los que $\boldsymbol{\Sigma}$ es definida positiva o algún subconjunto de este conjunto. Los elementos de $\boldsymbol{\Sigma}$ pueden ser funciones de \mathbf{t} , pero no de \mathbf{X} . Además, $\boldsymbol{\Sigma}$ es definida positiva si y sólo si $\sigma^2\mathbf{V}_0$ es definida positiva para cada individuo.

En (Weiss, 2005) y (Núñez Antón & Zimmerman, 2001) se discuten algunas estructuras para la matriz de varianzas y covarianzas; entre estas se destacan la Simétrica Compuesta *SC*, Intercepto Aleatorio *IA*, Autogresiva de Orden p *AR(p)*, Independiente *ID*, Incrementos Independientes *II*, Antedependientes *ADP*, Toeplitz *TL*, Factor Analítico *FA*, Autogresiva de Promedio Móvil de orden (p, q) *ARMA(p, q)*. Algunas de estas estructuras serán definidas con más detalle en la siguiente subsección 1.3.2.

1.3.2. Estructuras de covarianza

En la presente subsección se plasman las estructuras de covarianza que se utilizarán en el desarrollo de este trabajo. Para cada una, se exponen sus propiedades principales, ventajas y desventajas. Se define la forma teórica de la matriz de varianzas covarianzas y se realizan ejemplos numéricos que ayudan a explicar el comportamiento de la correlación o covarianza intra-individuos. Además de las estructuras que se exponen en esta subsección, otros tipos de modelos paramétricos para la covarianza se pueden consultar en (Weiss, 2005) y (Núñez Antón & Zimmerman, 2001). Para ilustrar las matrices, se asume que los datos fueron tomados en puntos del tiempo igualmente espaciados $\mathbf{t}_i = \mathbf{t} = (1, 2, 3, 4, 5)'$ sin datos perdidos.

1.3.2.1. Estructura de Simetría Compuesta

La estructura de simetría compuesta es la más simple, esta se define por los parámetros σ^2 y ρ , es decir, $\boldsymbol{\theta} = (\sigma^2, \rho)'$. El primer parámetro es la varianza de las observaciones $Var(y_{ij}) = \sigma^2$, y bajo este modelo las observaciones y_{ij} tienen varianza constante a través del tiempo. El parámetro ρ es la correlación entre dos observaciones del mismo individuo $Corr(y_{ij}, y_{il}) = \rho$, para $j \neq l$. La característica principal de esta estructura para la matriz de varianzas covarianzas es la homogeneidad, es decir, la varianza a lo largo de la diagonal principal se mantiene constante.

La forma algebraica de la matriz de varianzas covarianzas con estructura de simetría compuesta, cuando se consideran 5 observaciones en el tiempo es:

$$Var(\mathbf{y}_i | \sigma^2, \rho) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{pmatrix} \quad (1.14)$$

La característica clave de la estructura de simetría compuesta es que, en cualquier intervalo de tiempo $t_{ij} - t_{il} \neq 0$ grande o pequeño, la correlación $Corr(y_{ij}, y_{il})$ es la misma. Es decir, las observaciones tomadas con pocos minutos de diferencia o pocos años de diferencia tienen la misma correlación. Es difícil que este comportamiento se presente en datos reales medidos durante periodos de tiempo suficientemente largos. En la práctica, las medidas que son persistentes sobre el tiempo pueden seguir una estructura de simetría compuesta para la matriz de varianzas y covarianzas (Weiss, 2005).

Si se supone un modelo con la estructura de simetría compuesta donde $\sigma^2 = 1$ y $\rho = 0.7$, con cinco observaciones en el tiempo, siguiendo la estructura 1.14, la matriz de varianzas y covarianzas para un individuo es,

$$Var(\mathbf{y}_i) = \begin{pmatrix} 1 & 0.7 & 0.7 & 0.7 & 0.7 \\ 0.7 & 1 & 0.7 & 0.7 & 0.7 \\ 0.7 & 0.7 & 1 & 0.7 & 0.7 \\ 0.7 & 0.7 & 0.7 & 1 & 0.7 \\ 0.7 & 0.7 & 0.7 & 0.7 & 1 \end{pmatrix}$$

La estructura de simetría compuesta puede encontrarse con otros nombres. También se le conoce como equicovarianza o equicorrelación (*EC*). Esta última en algunos casos es usada para referirse a la versión de estructura heterogénea, es decir, varianzas no constantes en la diagonal principal de la matriz de varianzas y covarianzas (Weiss, 2005).

1.3.2.2. Estructura Autoregresiva

La estructura de covarianza Autoregresivo o *AR*(1), al igual que en la estructura de simetría compuesta, también tiene dos parámetros, $\boldsymbol{\theta} = (\sigma^2, \rho)'$. La varianza de y_{ij} es σ^2 para todo i y j y se asume varianza constante. En este caso ρ es el parámetro de correlación entre las mediciones de un individuo en m tiempos t_1, t_2, \dots, t_m y la correlación entre las mediciones tomadas en el i -ésimo individuo en los tiempos j y l viene dada por $Corr(y_{ij}, y_{il}) = \rho^{|t_{ij} - t_{il}|}$. La diferencia fundamental de este tipo de correlación serial con la

expuesta en 1.14 es que la covarianza decrece de manera exponencial a medida que avanza el tiempo, es decir, cuanto más separadas están dos observaciones en el tiempo, menor es la correlación entre ellas.

Weiss (2005) expone que en teoría el parámetro de correlación ρ puede ser negativo cuando las observaciones son balanceadas, aunque menciona que en la práctica es poco común encontrar correlaciones negativas, es por esto que sugiere considerar el parámetro de correlación positivo $\rho \in [0, 1]$. Si se tiene que $\mathbf{t} = (1, 2, 3, 4, 5)'$ la forma algebraica que representa la matriz de varianzas y covarianzas es,

$$Var(\mathbf{y}_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix} \quad (1.15)$$

En la matriz 1.15 se observa que las correlaciones de rezago 1 son todas iguales a ρ . Al alejarse de la diagonal principal, el rezago aumenta y las correlaciones disminuyen. Además, es importante resaltar que las observaciones y_{i1} y y_{i5} tienen la correlación más pequeña ρ^4 . Si se supone como en la subsección 1.3.2.1 $\sigma^2 = 1$ y $\rho = 0.7$, y se consideran cinco observaciones en el tiempo, siguiendo la estructura 1.15 la matriz de varianzas y covarianzas para un individuo es,

$$Var(\mathbf{y}_i) = \sigma^2 \begin{pmatrix} 1 & 0.7 & 0.49 & 0.343 & 0.2401 \\ 0.7 & 1 & 0.7 & 0.49 & 0.343 \\ 0.49 & 0.7 & 1 & 0.7 & 0.49 \\ 0.343 & 0.49 & 0.7 & 1 & 0.7 \\ 0.2401 & 0.343 & 0.49 & 0.7 & 1 \end{pmatrix}$$

Es difícil observar la diferencia entre las estructuras $AR(1)$ y SC visualizando los datos desde un gráfico de perfiles, en particular cuando la correlación ρ es alta o cuando se tienen pocas observaciones en el tiempo. Para escoger entre la estructura $AR(1)$ y SC para la matriz de varianzas y covarianzas es más útil analizar la matriz de dispersión (Weiss, 2005).

Núñez Antón & Zimmerman (2001) exponen que es posible obtener una generalización directa de los modelos $AR(1)$ y SC definiendo las varianzas en la diagonal principal distintas, y denominan estos modelos como una extensión homogénea de los anteriores, $ARH(1)$ y SCH , respectivamente.

1.3.2.3. Estructura Autoregresiva de Media Móvil

La estructura Autoregresiva de Promedio Móvil o $ARMA(1, 1)$ es una generalización de las estructuras $AR(1)$ y SC . La estructura $ARMA(1, 1)$ tiene tres parámetros $\boldsymbol{\theta} = (\sigma^2, \rho, \gamma)'$, un parámetro de varianza $Var(y_{ij}) = \sigma^2$ y dos parámetros de correlación, γ y ρ . El parámetro γ determina la correlación de rezago uno $Corr(y_{ij}, y_{i(j-i)})$, mientras ρ es la disminución adicional en la correlación en relación al aumenta del rezago entre las observaciones. La correlación para el rezago k es $Corr(y_{ij}, y_{i(j-k)}) = \gamma\rho^{k-1}$.

Si se tiene que $\mathbf{t} = (1, 2, 3, 4, 5)'$ y se suponen observaciones balanceadas, la forma algebraica que representa la matriz de varianzas y covarianzas con estructura $ARMA(1, 1)$ para un individuo es:

$$Var(\mathbf{y}_i) = \sigma^2 \begin{pmatrix} 1 & \gamma & \gamma\rho & \gamma\rho^2 & \gamma\rho^3 \\ \gamma & 1 & \gamma & \gamma\rho & \gamma\rho^2 \\ \gamma\rho & \gamma & 1 & \gamma & \gamma\rho \\ \gamma\rho^2 & \gamma\rho & \gamma & 1 & \gamma \\ \gamma\rho^3 & \gamma\rho^2 & \gamma\rho & \gamma & 1 \end{pmatrix} \quad (1.16)$$

Weiss (2005) presenta un ejemplo numérico del modelo de covarianza $ARMA(1, 1)$, el cual está basado en un estudio psicológico, se consideran cinco observaciones balanceadas sobre cada individuo en estudio y se supone que $\sigma^2 = 1$, $\gamma = 0.65$ y $\rho = 0.95$, entonces la matriz de varianzas y covarianzas teórica es,

$$Var(\mathbf{y}_i) = \begin{pmatrix} 1 & 0.65 & 0.6175 & 0.587 & 0.557 \\ 0.65 & 1 & 0.65 & 0.6175 & 0.587 \\ 0.6175 & 0.65 & 1 & 0.65 & 0.6175 \\ 0.587 & 0.6175 & 0.65 & 1 & 0.65 \\ 0.557 & 0.587 & 0.6175 & 0.65 & 1 \end{pmatrix} \quad (1.17)$$

Existen al menos tres casos especiales de la estructura $ARMA(1, 1)$.

1. $\rho = 1$ es la estructura de simetría compuesta SC ,
2. $\gamma = \rho$ es la estructura autoregresiva $AR(1)$, y
3. $\rho = 0$ es la estructura de promedio móvil $MA(1)$.

La estructura de promedio móvil $MA(1)$ tiene matriz de varianzas y covarianzas dada por,

$$Var(\mathbf{y}_i) = \sigma^2 \begin{pmatrix} 1 & \gamma & 0 & 0 & 0 \\ \gamma & 1 & \gamma & 0 & 0 \\ 0 & \gamma & 1 & \gamma & 0 \\ 0 & 0 & \gamma & 1 & \gamma \\ 0 & 0 & 0 & \gamma & 1 \end{pmatrix} \quad (1.18)$$

Núñez Antón & Zimmerman (2001) presenta una estructura más general, conocido como el modelo $ARIMA$ en donde se definen las diferencias entre mediciones consecutivas en la estructura de estacionariedad. Si se necesita conocer más aspectos teóricos y prácticos de la estructura $ARMA(1, 1)$ se recomienda consultar a Weiss (2005).

1.3.3. Estimación clásica del modelo

La función de verosimilitud para datos de n individuos, que se basa en el supuesto de la normalidad multivariada para los errores aleatorios en el modelo dado por 1.13, es

$$L(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = -\frac{1}{2} \sum_{i=1}^n \log(|\boldsymbol{\Sigma}_i|) - \frac{1}{2} \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}), \quad (1.19)$$

donde $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{V}_i$. El objetivo ahora es maximizar esta función con respecto a $\boldsymbol{\beta}$, σ^2 y \mathbf{V} . La estrategia es concentrar la verosimilitud para maximizar con respecto a $\boldsymbol{\beta}$ y σ^2 , manteniendo \mathbf{V}_0 fijo y conocido,

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\mathbf{V}_0) &= (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{Y} \\ \hat{\sigma}^2(\mathbf{V}_0) &= \frac{SCE(\mathbf{V}_0)}{nm}. \end{aligned}$$

Donde $SCE(\mathbf{V}_0) = (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\mathbf{V}_0))' \mathbf{V}_0^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}(\mathbf{V}_0))$. Al reemplazar los valores estimados de $\hat{\boldsymbol{\beta}}(\mathbf{V}_0)$ y $\hat{\sigma}^2(\mathbf{V}_0)$ en la verosimilitud 1.19 y al aplicar el logaritmo se obtiene,

$$L(\hat{\boldsymbol{\beta}}(\mathbf{V}_0), \hat{\sigma}^2(\mathbf{V}_0), \mathbf{V}_0) = -\frac{m}{2} \{n \log(SCE(\mathbf{V}_0)) + \log(|\mathbf{V}_0|)\} \quad (1.20)$$

Las estimaciones de $\boldsymbol{\beta}$ y σ^2 se obtienen al optimizar la función 1.20 respecto a los diferentes parámetros que contenga \mathbf{V} . Esta optimización se puede realizar aplicando métodos numéricos, por ejemplo el algoritmo de Newton-Raphson. Según Núñez Antón & Woodworth (1994) la estimación por máxima verosimilitud de $\boldsymbol{\Sigma}_i$ produce resultados insesgados hacia abajo. Además, Verbeke & Molenberghs (2009) expone que el método de máxima verosimilitud no siempre funcionan de forma adecuada, algunos problemas como la alta dimensionalidad de la matriz de covariables o estudios desbalanceados pueden ocasionar que el método sea poco efectivo. Los autores anteriormente citados, aconsejan usar la metodología alterna de máxima verosimilitud restringida (MVR).

El enfoque de estimación MVR fue propuesto por Patterson & Thompson (1971) y fue revisada por Harville (1977), entre otros. El método maximiza la verosimilitud sobre un conjunto de parámetros restringidos, se aplica para la estimación de una gran cantidad de modelos, principalmente para estimar los componentes de varianza en un modelo lineal general con errores distribuidos de manera normal multivariada. El procedimiento se basa en el hecho de que los efectos fijos son desconocidos al estimar los componentes de varianza. Este proceso remueve el efecto de las variables fijas al aplicar el proceso de máxima verosimilitud sobre los residuales obtenidos mediante mínimos cuadrados.

Los estimadores de MVR se obtienen al sobre una transformación lineal del vector respuesta. $\mathbf{Y}^* = \mathbf{A} \mathbf{Y}$, de modo que la distribución de \mathbf{V}^* no depende de $\boldsymbol{\beta}$. Para esta transformación se considera,

$$\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'.$$

Así, \mathbf{Y}^* tiene distribución normal multivariada con media cero sin importar los distintos valores de $\boldsymbol{\beta}$, de modo que las estimaciones de σ^2 y \mathbf{V}_0 no dependen de la dimensión de $\boldsymbol{\beta}$. La matriz usada para la transformación \mathbf{A} puede ser cualquier matriz de rango completo que cumpla con la condición que $E(\mathbf{Y}^*) = 0$.

En el modelo 1.13 se asume que la matriz de varianza y covarianza $\boldsymbol{\Sigma}$ depende de un conjunto de parámetros $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Sea $\mathbf{A} = \mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'$ y \mathbf{B} una matriz de tamaño

$nm \times (nm - p)$ tales que, $\mathbf{B}\mathbf{B}' = \mathbf{A}$ y $\mathbf{B}'\mathbf{B} = \mathbf{I}_{(nm-p)}$, sea $\mathbf{Z} = \mathbf{B}'\mathbf{Y}$, si el conjunto de parámetros $\boldsymbol{\theta}$, es fijado, se obtiene

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$$

Diggle et al. (2002) demuestra que bajo las condiciones anteriormente descritas $E(\mathbf{Z}) = 0$ y $Cov(\mathbf{Z}, \hat{\boldsymbol{\beta}})$, es decir, el supuesto de normalidad implica la independencia de \mathbf{Z} y $\hat{\boldsymbol{\beta}}$. Entonces se obtiene que la función de log-verosimilitud restringida para $\boldsymbol{\theta}$ es

$$L^*(\boldsymbol{\theta}) = \frac{1}{2}\log(|\boldsymbol{\Sigma}|) - \frac{1}{2}\log|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (1.21)$$

Por lo tanto, si los bloques de la matriz \mathbf{V}_0 son fijos y conocidos, se tiene

$$\hat{\boldsymbol{\beta}}(\mathbf{V}_0) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})'\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \quad (1.22)$$

y

$$\hat{\sigma}^2(\mathbf{V}_0) = \frac{SCE(\mathbf{V}_0)}{nm - p}. \quad (1.23)$$

La estimación de \mathbf{V}_0 resulta al maximizar la función,

$$L^*(\mathbf{V}_0) = -\frac{1}{2}m\{n \log(SCE(\mathbf{V}_0)) + \log|\mathbf{V}_0|\} - \frac{1}{2}\log(|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|) \quad (1.24)$$

Finalmente, al reemplazar los valores de la estimación de los bloques de la matriz de varianzas y covarianzas $\hat{\mathbf{V}}_0$ en los valores 1.22 y 1.23 se obtienen las estimaciones de $\hat{\boldsymbol{\beta}}$ y $\hat{\sigma}^2$.

1.3.4. Estimación Bayesiana del modelo

Gamerman (1997) propone una metodología de estimación bayesiana por bloques de parámetros, para el ajuste sobre modelos generalizados mixtos. Cepeda & Gamerman (2001) ajustan de manera conjunta la media y la estructura de varianza para un modelo de regresión lineal heterocedástico, empleando a prioris normales para los parámetros de la media y la varianza. Una generalización de la propuesta de estimación conjunta de la estructura de media y varianza es presentada por Cepeda (2001), Cepeda & Gamerman (2004) y Cepeda & Gamerman (2005). Por otro lado, Daniels & Pourahmadi (2002) desarrollaron propuestas bayesianas para ajustar modelos dinámicos longitudinales suponiendo a prioris normales para los parámetros de media y varianza, además emplearon la Descomposición de Cholesky presentada por Pourahmadi (1999). Cepeda & Núñez Antón (2007) y Cepeda & Núñez Antón (2009) proponen metodologías bayesianas para ajustar de manera conjunta los parámetros de media y varianza en modelos longitudinales, para la estructura de varianza suponen un modelo de antedependencia no estructurado, este enfoque es aplicado sobre modelos lineales y no lineales. Más recientemente, Castillo et al. (2019) presenta un propuesta bayesiana para la estimación conjunta de los parámetros de media y la matriz de varianzas covarianzas para modelos longitudinales con estructuras de

covarianza SC , $AR(1)$, $ARMA(1, 1)$ y algunos casos de estructuras de antedependencia no estructurada $ADS(1)$.

Si se tiene el modelo de regresión longitudinal presentado en 1.13, dado que $\mathbf{y} \sim NMV(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\mathbf{t}, \boldsymbol{\theta}))$, se considera $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})$. Si se supone la independencia entre $\boldsymbol{\beta}$, σ^2 y $\boldsymbol{\theta}$, la distribución a posteriori está dada por $\pi(\boldsymbol{\Theta}) = L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}|\mathbf{y})p(\boldsymbol{\beta})p(\sigma^2)p(\boldsymbol{\theta})$. Por lo anterior, se suponen las distribución a priori para el parámetro de media $\boldsymbol{\beta}$,

$$p(\boldsymbol{\beta}) \sim N(\mathbf{b}_0, \mathbf{B}_0) \quad (1.25)$$

Donde la distribución a priori $p(\boldsymbol{\theta})$ se define según la estructura de la matriz de varianzas y covarianzas. De 1.25, se obtiene la distribución $p(\boldsymbol{\beta})$,

$$p(\boldsymbol{\beta}) \propto \exp \left\{ -\frac{1}{2}(\boldsymbol{\beta} - \mathbf{b}_0)' \mathbf{B}_0^{-1}(\boldsymbol{\beta} - \mathbf{b}_0) \right\},$$

y a través del Teorema de Bayes, se obtiene la distribución condicional completa a posteriori $\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\theta}, \mathbf{y})$, dada por

$$\pi_{\boldsymbol{\beta}} \propto \exp \left\{ -\frac{1}{2}(\boldsymbol{\beta} - \mathbf{b}_1)' \mathbf{B}_1(\boldsymbol{\beta} - \mathbf{b}_1) \right\}$$

en consecuencia,

$$\pi_{\boldsymbol{\beta}} \sim N(\mathbf{b}_1, \mathbf{B}_1), \quad (1.26)$$

donde $\mathbf{b}_1 = \mathbf{B}_1(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y})$ y matriz de varianza $\mathbf{B}_1 = (\mathbf{B}_0^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}$ (Cepeda & Gamerman, 2004). Para obtener los valores estimados de $\boldsymbol{\beta}$ mediante el muestreador de Gibbs, se obtienen muestras de la distribución dada en 1.26 con probabilidad de aceptación $r = 1$.

1.3.4.1. Estimación Bayesiana de modelos longitudinales con estructura de covarianza SC y AR(1)

Las estructuras de la matriz de covarianza $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, presentadas en 1.14 y 1.15, depende de dos parámetros $\boldsymbol{\theta} = (\sigma^2, \rho)'$. Por lo anterior, es necesario establecer las distribuciones a posteriori condicionales completas para σ^2 y ρ .

Para obtener la distribución condicional completa a posteriori para σ^2 , es importante mencionar que $\boldsymbol{\Sigma} = \sigma^2\mathbf{V}$, donde \mathbf{V} es una matriz diagonal por bloques que depende únicamente del parámetro de correlación ρ , es decir, representa la matriz de correlación. Lo primero es definir la distribución a priori $p(\sigma^2)$,

$$p\left(\frac{1}{\sigma^2}\right) = p(\phi) \sim G\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right) \quad (1.27)$$

Dado que en 1.27 se supone una distribución a priori para ϕ , se tiene que

$$p(\phi) = \frac{1}{\Gamma\left(\frac{n_0}{2}\right)} \left(\frac{n_0\sigma_0^2}{2}\right)^{n_0/2} \phi^{(n_0/2-1)} \exp\left(-\frac{n_0\sigma_0^2}{2}\phi\right) I_{(0,\infty)}.$$

La verosimilitud se puede escribir como,

$$L(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \rho) \propto \phi^{n/2} \exp \left\{ -\frac{\phi}{2} \mathbf{R} \right\},$$

donde $\mathbf{R} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$. Por lo tanto, la distribución condicional completa a posteriori de ϕ es

$$\pi_\phi \propto \phi^{((n+n_0)/2-1)} \exp \left\{ -\phi \frac{n_0\sigma_0^2 + R}{2} \right\}$$

Por lo tanto, los valores de ϕ se pueden obtener de la distribución $\Gamma\left(\frac{n+n_0}{2}, \frac{n_0\sigma_0^2 + R}{2}\right)$, por lo tanto, los valores de σ^2 se obtienen directamente a través del muestreo de una distribución Gamma inversa.

$$\phi(\sigma^2) \sim \Gamma\left(\frac{n_1}{2}, \frac{n_1\sigma_1^2}{2}\right) \quad (1.28)$$

donde $n_1 = n + n_0$ y $n_1\sigma_1^2 = n_0\sigma_0^2 + \mathbf{R}$.

Se supone la distribución a priori para el parámetro de correlación ρ ,

$$p(\rho) \sim \text{Beta}(a, b) \quad (1.29)$$

y su función de densidad de probabilidad es

$$p(\rho) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \rho^{a-1} (1-\rho)^{b-1} I_{(0,1)}$$

Por lo tanto, la distribución condicional completa a posteriori para ρ está dada por

$$\pi_\rho \propto \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + (a-1) \ln(\rho) + (b-1) \ln(1-\rho) \right\} \quad (1.30)$$

La distribución condicional completa para el parámetro de correlación ρ , definida en 1.30, no tiene forma de una distribución conocida. De modo que para obtener muestras de esta distribución Castillo et al. (2019) propone como kernel de transición la caminata aleatoria

$$q(\rho^*|\rho^k) = \begin{cases} \rho^* \sim U(0, 2\rho^k) & \rho^k \leq 0.5 \\ \rho^* \sim U(2\rho^k - 1, 1) & \rho^k > 0.5 \end{cases} \quad (1.31)$$

Castillo et al. (2019) propone el algoritmo para la estimación conjunta de los parámetros de media y varianza en modelos longitudinales con estructura de covarianza SC o $AR(1)$.

1.3.4.2. Estimación Bayesiana de modelos longitudinales con estructura AR-MA(1,1)

Para un modelo longitudinal 1.13 con estructura de covarianza como la presentada en la matriz 1.16, $\Sigma(\boldsymbol{\theta})$ depende de tres parámetros $\boldsymbol{\theta} = (\sigma^2, \rho, \gamma)$. Como γ es un parámetro de correlación con características similares a ρ , se suponen las distribuciones a priori $Beta(a, b)$ y $Beta(c, d)$ para ρ y γ , respectivamente. Las distribuciones a posteriori para σ^2 y ρ están dadas por 1.28 y 1.30, respectivamente.

Para el parámetro de correlación γ , la distribución condicional completa a posteriori está dada por,

$$\pi_\gamma \propto \exp \left\{ -\frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\Sigma(\boldsymbol{\theta})^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + (c-1)\ln(\rho) + (d-1)\ln(1-\rho) \right\} \quad (1.32)$$

Para γ , Castillo et al. (2019) propone el kernel de transición

$$q(\gamma^*|\gamma^k) = \begin{cases} \gamma^* \sim U(0, 2\gamma^k) & \gamma^k \leq 0.5 \\ \gamma^* \sim U(2\gamma^k - 1, 1) & \gamma^k > 0.5 \end{cases} \quad (1.33)$$

Castillo et al. (2019) propone el algoritmo para la estimación conjunta de los parámetros de media y varianza en modelos longitudinales con estructura de covarianza ARMA(1, 1).

1.3.5. Inferencia Bayesiana

La Inferencia Bayesiana es el proceso de ajustar un modelo de probabilidad a un conjunto de datos y resumir el resultado por una distribución de probabilidad en los parámetros del modelo y en cantidades no observadas tales como predicciones para nuevas observaciones (Gelman et al., 2014). En algunas ocasiones, es posible tener información previa de los valores de los parámetros. Para incluir esta información en el análisis estadístico, se emplea la función de densidad para $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$. Esta función de probabilidad dependerá de un conjunto de parámetros $\boldsymbol{\theta}'$, denominados hiperparámetros, los cuales se asumen conocidos. $p(\boldsymbol{\theta})$ es conocida como la distribución a priori (Cepeda, 2001).

Entonces, luego de obtener los valores observados de la variable de interés \mathbf{Y} , se cuenta con dos fuentes de información sobre los parámetros: la función de verosimilitud $L(\boldsymbol{\theta}|\mathbf{Y}) = \prod f(y_i|\boldsymbol{\theta})$ y la distribución a priori $p(\boldsymbol{\theta})$. Por lo tanto, en el análisis estadístico la inferencia se puede centrar en la distribución de $\boldsymbol{\theta}$ después de observados los datos. Por lo anterior, se obtiene la distribución a posteriori de $\boldsymbol{\theta}$, se denota $\pi(\boldsymbol{\theta})$ y se puede obtener por el teorema de Bayes

$$\pi(\boldsymbol{\theta}) \propto L(\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

Desde la perspectiva Bayesiana, la inferencia sobre $\boldsymbol{\theta}$ está basada en la distribución a posteriori $\pi(\boldsymbol{\theta})$. Dado que no siempre es posible obtener información a posteriori de manera analítica, en este trabajo se usan métodos basados en simulación estocástica empleando

cadenas de Markov (MCMC) para obtener muestras de la distribución a posteriori $\pi(\theta)$ (Gamerman & Lopes, 2006).

1.3.6. Algoritmo de Metropolis-Hastings

El algoritmo de Metropolis-Hastings, inicialmente propuesto por Metropolis et al. (1953) y posteriormente extendido por Hastings (1970), es un método de simulación estocástica, que usa cadenas de Markov. Para aplicarlo, es necesario construir un núcleo de transición $p(\theta, \phi)$, de tal manera que π sea la distribución de equilibrio de la cadena. Una forma fácil de comprobar esto es cuando p satisface la condición de reversibilidad de la cadena

$$\pi(\theta)p(\theta, \phi) = \pi(\phi)p(\phi, \theta), \text{ para todo } \theta, \phi,$$

esta es una condición suficiente para que π sea una distribución de equilibrio de la cadena. Si se calcula la integral en ambos lados de la igualdad, se obtiene:

$$\int \pi(\theta)p(\theta, \phi)d\theta = \pi(\phi), \text{ para todo } \phi.$$

Para la construcción del núcleo de transición se deben tener en cuenta dos partes: un núcleo arbitraria de transición $q(\theta, \phi)$ que satisfaga $\int q(\theta, \phi)d\phi = 1$ y una probabilidad de aceptación $\alpha(\theta, \phi)$ tal que

$$p(\theta, \phi) = \begin{cases} q(\theta, \phi)\alpha(\theta, \phi) & \theta \neq \phi \\ 1 - \int q(\theta, \phi)\alpha(\theta, \phi)d\theta & \theta = \phi. \end{cases}$$

El núcleo de transición $q(\theta, \phi)$ propone el movimiento de la cadena de Markov del punto θ a un nuevo valor ϕ . La probabilidad de aceptación se define como

$$\alpha(\theta, \phi) = \min \left\{ 1, \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)} \right\}.$$

El algoritmo para la simulación de una muestra de una distribución π empleando métodos de cadenas de Markov se puede describir como sigue

1. Iniciar el contador de la cadena, $j = 1$ y asignar un valor inicial $\theta^{(0)}$ para el parámetro θ .
2. Proponer un nuevo valor ϕ de θ generado de la densidad propuesta $q(\theta^{(j-1)}, \cdot)$.
3. Calcular la probabilidad de aceptación $\alpha(\theta, \phi)$, si se acepta $\theta^j = \phi$, en caso contrario $\theta^j = \theta^{j-1}$.
4. Cambiar el contador de j a $j + 1$ y retornar al paso 2 hasta obtener convergencia.

El algoritmo es implementado generando una cantidad u de una distribución uniforme en el intervalo $(0, 1)$, independiente de θ . Si $u \leq \alpha$ se acepta el movimiento de la cadena y si $u > \alpha$ se rechaza el movimiento (Gamerman & Lopes, 2006). Los valores obtenidos se pueden considerar como una muestra de la distribución π únicamente después de que la cadena pase del estado transitorio y el efecto de los valores iniciales es demasiado pequeño como para ser ignorado.

Gamerman & Lopes (2006) presentan una variación del algoritmo Metropolis-Hastings cuando el conjunto de parámetros θ es demasiado grande. Esta consiste en particionar θ en bloques de dimensiones más pequeñas y de esta manera obtener la estimación para cada partición de manera sucesiva. En este caso, el algoritmo toma la siguiente forma

1. Iniciar el contador de la cadena, $j = 1$ y asignar un valor inicial $\theta^{(0)}$ para el parámetro θ .
2. Iniciar un contador de componentes i aleatoriamente o en una forma fija.
3. Proponer un movimiento para el i -ésimo movimiento del vector θ a un nuevo valor ϕ_i generado de la densidad propuesta $q(\theta^{j-1}, \cdot)$.
4. Calcular la probabilidad de aceptación del movimiento, $\alpha_i(\theta_i^{(j-1)}, \phi_i)$. Si el movimiento es aceptado $\theta_i^j = \phi_i$, en caso contrario $\theta_i^j = \theta_i^{j-1}$.
5. Cambiar el contador $j = j + 1$ y retornar al paso 2 hasta obtener convergencia.

Las probabilidades de aceptación son dadas por

$$\alpha(\theta, \phi) = \min \left\{ 1, \frac{\pi_i(\phi_i)q(\phi_i, \theta_i)}{\pi_i(\theta_i)q(\theta_i, \phi_i)} \right\},$$

donde $\pi_i(\theta_i) = \pi(\theta_i | \theta_{-i})$ y θ_{-i} es el vector θ sin la i -ésima componente θ_i . π es conocida como la distribución condicional completa de θ_i .

1.3.7. Algoritmo caminata aleatoria Metropolis-Hastings

El algoritmo caminata aleatoria Metropolis-Hastings es un caso especial del algoritmo de Metropolis-Hastings. Una caminata aleatoria es una cadena de Markov con movimiento dada por $\theta^{(j)} = \theta^{(j-1)} + w_j$ donde w_j es una variable aleatoria con distribución independiente de la cadena. La perturbación w_j es independiente e idénticamente distribuida con densidad f_w . Si f_w es simétrica al rededor de 0, la cadena es simétrica. Por lo tanto, tiene movimientos propuesto de acuerdo a $q(\theta, \phi) = f_w(\phi - \theta)$ entonces $q(\theta, \phi) = q(\phi, \theta)$ (Gamerman & Lopes, 2006). Un ejemplo de cadenas simétricas es el Algoritmo de Metropolis (Metropolis et al., 1953). Entonces se tiene que la probabilidad de aceptación no depende de q y está dada por

$$\alpha(\theta, \phi) = \min \left\{ 1, \frac{\pi(\phi)}{\pi(\theta)} \right\}.$$

El algoritmo puede describirse en los siguientes pasos

1. Iniciar el contador de la cadena $j = 1$ y asignar un valor inicial $\theta^{(0)}$ para el parámetro θ .
2. Proponer un movimiento $\phi = \theta^{(j-1)} + w_j$ generado de la densidad propuesta $q(\theta^{(j-1)}, \cdot) = q(\phi - \theta^{(j-1)})$.
3. Calcular la probabilidad de aceptación $\alpha(\theta, \phi)$, si se acepta $\theta^{(j)} = \phi$, en caso contrario $\theta^{(j)} = \theta^{(j-1)}$.

4. Cambiar el contador de j a $j + 1$ y retornar al paso 2 hasta obtener convergencia.

Esta es una opción muy común y la mayoría de aplicaciones del algoritmo de Metropolis-Hastings implementan este esquema. Las elecciones de distribuciones de transición propuestas simétricas incluyen a la distribución Normal, distribución t-student y distribución uniforme centradas en el origen. Los valores propuestos están basados en los valores previos de la cadena. En cuanto a la dispersión de q , los valores grandes de varianza permiten movimientos muy distantes de los valores previos, pero a un costo de tasa de aceptación muy pequeña. Por otro lado, los valores pequeños de varianza genera movimientos cercanos a los valores previos con altas tasas de aceptación (Gamerman & Lopes, 2006).

Modelos de regresión con cambio estructural

Los modelos lineales clásicos son empleados en el análisis de datos normales, su utilidad es el resultado de usar una función para expresar la relación entre una variable respuesta y una o más variables explicativas. Se pueden encontrar aplicaciones casi en cualquier campo, incluyendo ingeniería, ciencias físicas y químicas, economía, administración, ciencias biológicas, ciencias sociales, entre otras (Montgomery et al., 2001). Existe una gran cantidad de literatura sobre modelos lineales clásicos, Draper & Smith (1998), Ravishanker & Dey (2001) y McCullagh & Nelder (1989) son algunas referencias que abordan este tema con detalle.

Por lo general, se supone que el modelo explica todo el conjunto de datos, pero esto no siempre es cierto. Por el contrario, el modelo puede cambiar después de un punto específico que divide los datos en dos segmentos y surge la necesidad de introducir los modelos lineales con cambio estructural, en donde el cambio se puede presentar en la estructura de media, varianza o en ambas. En la literatura, muchos autores han abordado el problema del punto de cambio asociado a un modelo de regresión. Quandt (1958) y Quandt (1960) determinó una prueba de razón de verosimilitud para probar la necesidad de un punto de cambio y estimar un modelo de regresión lineal en dos regímenes diferentes. Hinkley (1970) derivó la distribución asintótica del estadístico de la prueba de razón de verosimilitud y lo aplicó para detectar un cambio en la media dentro de una muestra de observaciones distribuidas normalmente. Ferreira (1975) estudió el modelo de regresión con cambio estructural desde el punto de vista Bayesiano. Brown et al. (1975) presentaron una metodología basada en residuales recursivos para probar la existencia de puntos de cambio en modelos de regresión múltiple. Chen & Gupta (2012) abordaron el tema del punto de cambio para el modelo de regresión lineal simple y el modelo de regresión lineal múltiple mediante el uso del criterio de información de Schwarz, y por un enfoque Bayesiano.

Para detectar el cambio estructural, los datos deben tener algún orden natural, puede ser el tiempo de observación o alguna otra variable (por ejemplo; tiempo, espacio, edad, entre otros). En esta investigación, un cambio estructural es entendido como un cambio en los parámetros que describen la media o la varianza, pero la función de enlace entre la respuesta y las variables explicativas permanece igual.

2.1. Modelos de regresión con un punto de cambio

Sea \mathbf{y}_i , $i = 1, 2, 3, \dots, n$ que denota las observaciones de la variable respuesta, sea $\mathbf{x}_i = (1, x_{i2}, \dots, x_{ip})'$ denota un vector columna de p covariables o variables explicativas independientes, con $x_{i1} = 1$ para todo i , para incluir un intercepto en el modelo. Entonces, el modelo lineal clásico con un punto de cambio puede escribirse como (Quandt, 1958),

$$\mathbf{y}_i = \begin{cases} \mathbf{x}_i' \boldsymbol{\beta}_1 + \epsilon_{1i} & i = 1, \dots, \tau \\ \mathbf{x}_i' \boldsymbol{\beta}_2 + \epsilon_{2i} & i = \tau + 1, \dots, n, \end{cases} \quad (2.1)$$

donde $\boldsymbol{\beta}_d$, $d = 1, 2$ son vectores $p \times 1$ desconocidos y ϵ_{di} son errores independientes e idénticamente distribuidos con $\epsilon_{di} \sim N(\mathbf{0}, \boldsymbol{\sigma}_d^2)$. Para realizar una estimación válida para $\boldsymbol{\beta}_d$ y $\boldsymbol{\sigma}_d^2$ se plantea la restricción $\tau \in \{p+1, \dots, n-p-1\}$. Se asume que los primeros $p+1$ y últimos $n-p-1$ vectores de \mathbf{x}_i son linealmente independientes. El modelo (2.1) puede escribirse en forma matricial como dos modelos de regresión por separado,

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1 \\ \mathbf{Y}_2 &= \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2, \end{aligned} \quad (2.2)$$

donde \mathbf{Y}_1 y \mathbf{Y}_2 son vectores columna de las primeras τ y las últimas $n-\tau$ observaciones de la variable respuesta, respectivamente. Las matrices \mathbf{X}_1 y \mathbf{X}_2 son los primeros τ y los últimos $n-\tau$ vectores fila de la matriz diseño, respectivamente. Son dadas por $\mathbf{X}_1 = (x_1, \dots, x_\tau)'$ y $\mathbf{X}_2 = (x_{\tau+1}, \dots, x_n)'$. Además, el vector de errores sigue una distribución normal, es decir, $\boldsymbol{\epsilon}_d \sim N(\mathbf{0}, \sigma_d^2 \mathbf{I}_d)$, donde \mathbf{I}_d es la matriz identidad con rango τ para $d = 1$ y rango $n-\tau$ para $d = 2$. No hay restricciones de continuidad para los modelos (2.2) en el punto de cambio y éstos son independientes, por lo tanto, pueden ser escritos como

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{pmatrix}. \quad (2.3)$$

2.1.1. Estimación clásica del modelo

La matriz diseño en (2.3) es diagonal en bloques, lo que indica independencia entre las estimaciones de $\boldsymbol{\beta}_1$ y $\boldsymbol{\beta}_2$. Si el punto de cambio τ es conocido, la función de log-verosimilitud puede ser dividida en dos términos,

$$\ell(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 | \tau, \mathbf{y}) = \ell(\boldsymbol{\beta}_1, \sigma_1^2 | y_1, \dots, y_\tau) + \ell(\boldsymbol{\beta}_2, \sigma_2^2 | y_{\tau+1}, \dots, y_n), \quad (2.4)$$

con $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)'$. Estos términos corresponden a la log-verosimilitud de las primeras τ observaciones y las últimas $n-\tau$ observaciones y ambos términos son mutuamente independientes. Así, las estimaciones para $\boldsymbol{\beta}$ y $\boldsymbol{\sigma}^2$ son las estimaciones de máxima verosimilitud (1.4) y (1.5), respectivamente, para el modelo (2.3) en los dos segmentos y están dadas por,

$$\hat{\boldsymbol{\beta}}_d = (\mathbf{X}_d' \mathbf{X}_d)^{-1} \mathbf{X}_d' \mathbf{y}_d, \quad d = 1, 2 \quad (2.5)$$

y

$$\sigma_1^2 = \frac{1}{\tau} \hat{S}_1^2, \quad \sigma_2^2 = \frac{1}{n-\tau} S_2^2,$$

donde,

$$S_d^2 = (\mathbf{y}_d - \mathbf{X}_d \hat{\boldsymbol{\beta}}_d)' (\mathbf{y}_d - \mathbf{X}_d \hat{\boldsymbol{\beta}}_d),$$

es la suma residual de cuadrados para el modelo en el d -ésimo segmento.

El estimador de máxima verosimilitud de τ es el valor que maximiza la log-verosimilitud (2.4) con los estimadores de máxima verosimilitud $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)'$ y $\hat{\boldsymbol{\sigma}}^2 = (\hat{\sigma}_1^2, \hat{\sigma}_2^2)'$ del modelo en los dos segmentos. La log-verosimilitud con estos estimadores es dada por

$$\ell(\tau|\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}, \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{\tau}{2} \log(\hat{\sigma}_1^2) - \frac{n-\tau}{2} \log(\hat{\sigma}_2^2) - \frac{1}{2\hat{\sigma}_1^2} \tau \hat{\sigma}_1^2 - \frac{1}{2\hat{\sigma}_2^2} (n-\tau) \hat{\sigma}_2^2.$$

Por lo tanto, $\hat{\tau}$ se obtiene al maximizar

$$\ell(\tau|\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}, \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{\tau}{2} \log(\hat{\sigma}_1^2) - \frac{n-\tau}{2} \log(\hat{\sigma}_2^2) - \frac{n}{2}, \quad (2.6)$$

con respecto a $\tau = p+1, \dots, n-p-1$.

Bajo el supuesto $\sigma_1^2 = \sigma_2^2 = \sigma^2$ el estimador de máxima verosimilitud de σ^2 es,

$$\sigma^2 = \frac{1}{n} (\hat{S}_1^2 + \hat{S}_2^2)$$

y (2.6) se reduce a

$$\ell(\tau|\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}, \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\hat{\sigma}^2) - \frac{n}{2}. \quad (2.7)$$

Luego, en el caso de varianzas iguales, $\hat{\tau}$ minimiza $\hat{S}_1^2 + \hat{S}_2^2$. Por otro lado, si no es de interés estimar el punto de cambio, el modelo puede ser interpretado como la mixtura de dos modelos de regresión. (Leisch, 2004) presenta una metodología general para la estimación de modelos de mixturas finitas, basada en el algoritmo EM (Dempster et al., 1977).

2.1.1.1. Ejemplo de estimación clásica del modelo

En éste ejemplo presentado por Quandt (1958), se considera un modelo de regresión lineal con un punto de cambio. Los datos fueron generados de manera artificial. La variable explicativa x se generó a través de una permutación aleatoria de los primeros 20 números enteros. Los errores ϵ_i fueron generados como 20 variable normales independientes e idénticamente distribuidas con media cero y varianza uno. Estos errores fueron agregados a la media de un modelo que cambia después de las primeras 12 observaciones, este se presenta a continuación,

$$y_i = \begin{cases} 2.5 + 0.7x_i + \epsilon_{1i} & i = 1, \dots, 12 \\ 5.0 + 0.5x_i + \epsilon_{2i} & i = 13, \dots, 20. \end{cases}$$

Los datos se presentan en la Tabla 2.1, y el modelo ajustado con un punto de cambio en $\hat{\tau} = 12$ es presentado en la Figura 2.1. Resulta de interés mencionar que en este conjunto de datos el orden es determinado por el índice i , es decir, la variable explicativa no necesariamente debe estar ordenada, como se puede apreciar en la Tabla 2.1.

El punto de cambio τ se estima con una búsqueda sobre la grillas de todos los posibles valores $\tau \in \{3, 4, \dots, 17\}$. Los valores de la log-verosimilitud (2.6) y (2.7) bajo ambos supuestos en la varianza se presentan en la Tabla 2.2. En ambos caso el máximo se encuentra en $\hat{\tau} = 12$, que efectivamente corresponde al verdadero valor definido por Quandt

i	x_i	y_i	i	x_i	y_i
1	4	3.473	11	15	13.036
2	13	11.555	12	11	8.264
3	5	5.714	13	3	7.612
4	2	5.710	14	14	11.802
5	6	6.046	15	16	12.551
6	8	7.650	16	10	10.296
7	1	3.140	17	7	10.014
8	12	10.312	18	19	15.472
9	17	13.353	19	18	15.650
10	20	17.197	20	9	9.871

TABLA 2.1. Datos generados por Quandt

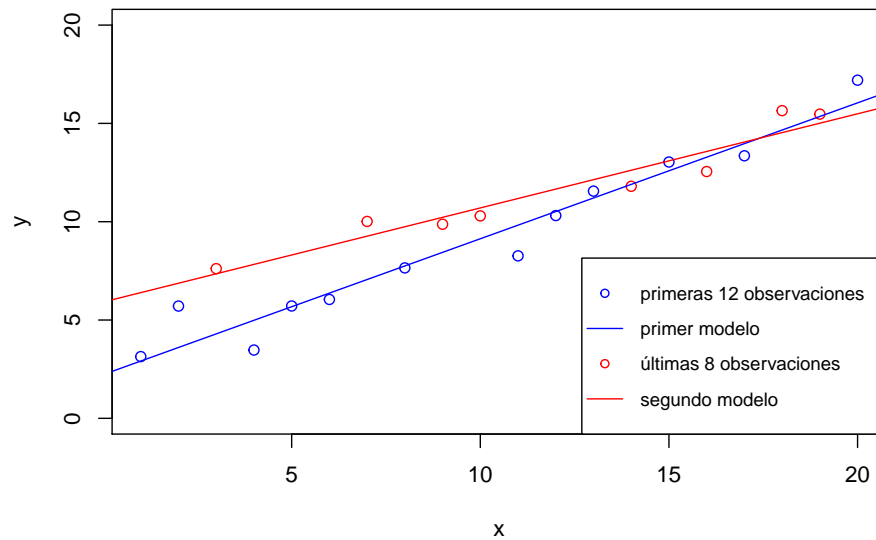


FIGURA 2.1. Gráfico de datos generados por Quandt

τ	$\ell_{\sigma_1^2 \neq \sigma_2^2}(\tau \hat{\beta}, \hat{\sigma}, \mathbf{y})$	$\ell_{\sigma_1^2 = \sigma_2^2 = \sigma^2}(\tau \hat{\beta}, \hat{\sigma}, \mathbf{y})$	τ	$\ell_{\sigma_1^2 \neq \sigma_2^2}(\tau \hat{\beta}, \hat{\sigma}, \mathbf{y})$	$\ell_{\sigma_1^2 = \sigma_2^2 = \sigma^2}(\tau \hat{\beta}, \hat{\sigma}, \mathbf{y})$
3	-29.34	-30.25	11	-28.26	-28.55
4	-32.06	-32.08	12	-25.38	-25.82
5	-31.54	-31.54	13	-29.40	-30.46
6	-31.07	-31.12	14	-29.25	-30.16
7	-29.63	-29.70	15	-27.74	-29.54
8	-29.01	-29.18	16	-28.09	-29.77
9	-28.56	-28.91	17	-29.21	-31.91
10	-28.42	-28.53			

TABLA 2.2. Log-verosimilitud (2.6) y (2.7) para todos los posibles puntos de cambio τ , asumiendo varianzas diferentes e iguales, respectivamente.

(1958). Por lo tanto, las estimaciones de los parámetros del modelo son $\hat{\beta}_{01} = 2.2215$, $\hat{\beta}_{11} = 0.6912$, $\hat{\beta}_{02} = 5.9141$, $\hat{\beta}_{12} = 0.4787$, $\hat{\sigma}_1^2 = 0.95$ y $\hat{\sigma}_2^2 = 0.5114$.

2.1.2. Enfoque por razón de verosimilitud

Quandt (1960) fue el primero que propuso utilizar una prueba de razón de verosimilitud para probar la necesidad de incluir un punto de cambio en el modelo de regresión. Si se considera el modelo (2.3), y la hipótesis nula de interés es $H_0 : \beta_1 = \beta_2$, la hipótesis alterna es $H_1 : \beta_1 \neq \beta_2$, entonces se obtiene que el test de razón de verosimilitud puede ser escrito como

$$\lambda(\mathbf{y}) = \frac{\sup_{\Theta_0} L(\boldsymbol{\theta} | \mathbf{y})}{\sup_{\Theta} L(\boldsymbol{\theta} | \mathbf{y})} = \frac{L(\hat{\boldsymbol{\theta}}_0 | \mathbf{y})}{L(\hat{\boldsymbol{\theta}} | \mathbf{y})},$$

donde $L(\boldsymbol{\theta} | \mathbf{y})$ es la función de verosimilitud del vector de parámetros $\boldsymbol{\theta}$ para la muestra de datos \mathbf{y} y Θ es el espacio de parámetros completo. El conjunto Θ_0 es el espacio de parámetros restringido en H_0 y es necesariamente un subconjunto de Θ . Usando el método de máxima verosimilitud para estimar el parámetro $\boldsymbol{\theta}$, se obtiene que $\hat{\boldsymbol{\theta}}$ es la estimación de máxima verosimilitud sin restricciones de $\boldsymbol{\theta}$ que se puede realizar en el espacio de parámetros completo Θ , y $\hat{\boldsymbol{\theta}}_0$ es la estimación de máxima verosimilitud restringida donde la maximización es restringida a Θ_0 . Bajo algunas condiciones de regularidad, se obtiene

$$\Lambda(\mathbf{y}) = -2 \log \lambda(\mathbf{y}),$$

sigue asintóticamente una distribución χ^2 con q grados de libertad, donde q es la diferencia del número de parámetros en los modelos bajo H_0 y H_1 , respectivamente (Casella & Berger, 2001).

Un supuesto para aplicar la prueba de razón de verosimilitud es que los modelos bajo H_0 y H_1 deben estar anidados. No es obvio que un modelo de regresión sin un punto de cambio esté anidado en el modelo 2.2, este supuesto es validado por Hofrichter (2007). La máxima log-verosimilitud para el modelo de regresión lineal es

$$(\hat{\beta}, \tilde{\sigma}^2 | \mathbf{y}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\tilde{\sigma}^2) - \frac{n}{2}, \quad (2.8)$$

donde $\tilde{\sigma}^2$ es la estimación de máxima verosimilitud de σ^2 basado en todas las observaciones. Entonces $\Lambda(\mathbf{y})$ para el modelo (2.2) se obtiene restando (2.7) de (2.8) como

$$\Lambda(\mathbf{y}) = \left[n \log \frac{\tilde{S}^2}{n} - \tau \log \frac{\hat{S}_1^2}{\tau} - (n - \tau) \log \frac{\hat{S}_2^2}{n - \tau} \right]_{\tau=\hat{\tau}}, \quad (2.9)$$

donde

$$\tilde{S}^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

es la suma de cuadrados residual para el modelo (2.2) bajo la hipótesis nula. Este estadístico puede ser empleado para probar cambio en la varianza y en los coeficientes de regresión (Worsley, 1983). En caso de varianzas iguales, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, se obtiene

$$\Lambda(\mathbf{y}) = n \log \left[\frac{\tilde{S}}{\hat{S}_1^2 + \hat{S}_2^2} \right]_{\tau=\hat{\tau}}.$$

Para τ conocido y $\sigma_1^2 = \sigma_2^2$ la estadística de prueba F para $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ es

$$F_\tau = \frac{[\tilde{S}^2 - (\hat{S}_1^2 + \hat{S}_2^2)]/p}{(\hat{S}_1^2 + \hat{S}_2^2)/(n - 2p)},$$

que bajo H_0 sigue una distribución F con p y $n - 2p$ grados de libertad. Worsley (1983) sugiere usar una generalización de la estadística de prueba F , esta es

$$F_{max} = \max_{p < \tau < n-p} F_\tau$$

para probar $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Worsley (1983) presenta una aproximación a la distribución de F_{max} bajo la hipótesis nula basada en una mejora de la desigualdad de Bonferroni (Worsley, 1982). Además, proporciona un límite superior para el nivel de significación de la prueba. Los ejemplos numéricos indican que sus límites superiores aproximados son razonablemente precisos para muestras pequeñas y para tamaños moderados de nivel de significancia. Sin embargo, la aproximación pierde su precisión para muestras grandes. Finalmente, Worsley (1983) define el estadístico de razón de verosimilitud bajo una variedad de supuestos. Kim (1994a) introduce una expresión alternativa para la prueba de razón de verosimilitud; estudia la potencia y robustez del estadístico e indica su aplicabilidad en la práctica.

Las pruebas de hipótesis generales, donde la varianza cambia en el punto de cambio, fueron introducidas por primera vez por Brown et al. (1975) empleando residuales recursivos; se propone una prueba basada en la cusum y cusum cuadrada de los residuales recursivos. Kim (1994b) investiga algunas propiedades analíticas de la prueba de razón de verosimilitud y compara su desempeño con la prueba cusum de Brown et al. (1975), concluye que la prueba de razón de verosimilitud es mucho más potente que la prueba cusum y propone un test basado en la backward cusum para mejorar la potencia de la prueba. Una discusión más detallada sobre pruebas para determinar la necesidad de un punto de cambio en modelos de regresión es dada por Chen & Gupta (2012).

Killick & Eckley (2014) desarrollaron el paquete de R llamado **change**point cuya principal contribución es que el usuario tiene la capacidad de seleccionar el método de búsqueda de múltiples puntos de cambio. Los tres métodos implementados son: Segmentos Vecinos (Auger & Lawrence, 1989), Segmentación Binaria (Edwards & Cavalli-Sforza, 1965) y Pruned Exact Linear Time (PELT) (Killick et al., 2012). Éstos están disponibles para

cambios en la media y/o varianza empleando supuestos distributivos o sin distribución. En particular, éste paquete se puede aplicar para buscar un punto de cambio en la media y/o varianza incluyendo la prueba de razón de verosimilitud (Worsley, 1983) o la prueba cusum (Brown et al., 1975).

2.1.3. Enfoque por residuales recursivos

Brown et al. (1975) fueron los primeros en introducir los residuales recursivos para detectar cambio en la estructura de media o varianza de un modelo lineal. En general, los residuales son definidos como la diferencia entre los valores observados y_i y los valores ajustados μ_i , cabe aclarar que la estimación de los valores ajustados μ_i se realiza con todos los datos. En contraste, en los residuales recursivos los valores ajustados se obtienen con una parte de los datos. En particular, el residual recursivo de la r -ésima observación se define como la diferencia entre y_r y la media ajustada $\hat{\mu}_{r-}$, donde $\hat{\mu}_{r-}$ está basado en las primeras $r - 1$ observaciones. Es importante mencionar que es necesario ordenar los datos.

Se considera un modelo líneal clásico donde los residuales recursivos son definidos como

$$u_r = y_r - \hat{\mu}_{r-} = y_r - \mathbf{x}'_r \hat{\boldsymbol{\beta}}_{r-}, \quad r = p + 1, \dots, n. \quad (2.10)$$

donde $\hat{\boldsymbol{\beta}}_{r-}$ es la estimación de máxima verosimilitud de $\boldsymbol{\beta}$ basada en las primeras $r - 1$ observaciones y p es el número de parámetros en el predictor lineal. La esperanza de los residuales recursivos es

$$\begin{aligned} E[U_r] &= E[y_r - \mathbf{x}'_r \hat{\boldsymbol{\beta}}_{r-}] \\ &= E[y_r] - \mathbf{x}'_r E[\hat{\boldsymbol{\beta}}_{r-}] \\ &= \mu_r - \mathbf{x}'_r \boldsymbol{\beta} = 0 \end{aligned}$$

Por lo tanto, $\hat{\boldsymbol{\beta}}_{r-}$ es un estimador insesgado de $\boldsymbol{\beta}$.

La estimación de máxima verosimilitud $\hat{\mu}_{r-}$ se realiza con las primeras $r - 1$ observaciones y es independiente de las observaciones restantes y de y_r . Por lo tanto, las dos componentes de u_r son independientes y su varianza es la suma de las varianzas de y_r y $\hat{\mu}_{r-}$. Se tiene que la varianza de y_r es σ^2 y la varianza de $\hat{\mu}_{r-}$ es

$$\begin{aligned} \text{Var}(\hat{\mu}_{r-}) &= \text{Var}(\mathbf{x}'_r \hat{\boldsymbol{\beta}}_{r-}) \\ &= \mathbf{x}'_r \text{Var}(\hat{\boldsymbol{\beta}}_{r-}) \mathbf{x}_r \\ &= \sigma^2 \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r \end{aligned}$$

donde \mathbf{X}_{r-1} es la matriz diseño que contiene los primeros $r - 1$ vectores fila de \mathbf{X} . De lo anterior, se obtiene que

$$\text{Var}(u_r) = \sigma^2 (1 + \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r) \quad (2.11)$$

y los residuales recursivos estandarizados son definidos como

$$u_r^* = \frac{y_r - \mathbf{x}'_r \hat{\boldsymbol{\beta}}_{r-}}{\sqrt{1 + \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r}}. \quad (2.12)$$

Como y_r y $\mathbf{x}'_r \hat{\boldsymbol{\beta}}_{r-}$ tienen distribución normal, entonces $u_r^* \sim N(0, \sigma^2)$. Bajo el supuesto que no hay cambio de régimen en un modelo lineal homocedástico, estos residuales recursivos son independientes, entonces $u_r^* \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$. Como los u_r son normalmente distribuidos, se puede demostrar que la covarianza $Cov[u_r, u_s]$, $r \neq s$, de dos residuales recursivos es cero (Brown et al., 1975), es decir,

$$Cov[u_r, u_s] = E[u_r, u_s] - E[u_r]E[u_s] = 0, \quad r \neq s,$$

y los residuales recursivos son no correlacionados.

Para probar si es necesario un punto de cambio y determinar su ubicación Brown et al. (1975) no aplicaron el residual estandarizado completo, en su lugar usaron la siguiente variación

$$U_r = \frac{1}{\sigma} \sum_{j=p+1}^r u_j^*, \quad r = p+1, \dots, n.$$

Si σ es desconocido, este puede ser hallado como la estimación de la desviación estándar basada en todos los datos (1.5), por lo tanto, $\boldsymbol{\beta}$ es constante. Es decir, se puede emplear un modelo lineal para ajustar todos los datos y la media de U_r es cero.

Por otro lado, cuando los datos tienen un punto de cambio, la relación entre la respuesta y las variables explicativas cambia, por lo tanto U_r tiene una media diferente de cero y por consiguiente la secuencia U_{p+1}, \dots, U_n se aleja de cero. Por lo tanto, la secuencia puede ser usada para probar si es necesario un punto de cambio.

La hipótesis nula considerada, H_0 , es que no hay cambios estructurales y no es necesario un punto de cambio. Bajo la hipótesis H_0 , se obtiene que U_{p+1}, \dots, U_n es una secuencia de variables normales con $E[U_r] = 0$, $Var[U_r] = r - p$ y para $r < s$ la $Cov[U_r, U_s] = r - p$.

Para probar la necesidad de un punto de cambio, es necesario definir bandas tal que la probabilidad de que la secuencia U_{p+1}, \dots, U_n cruce a través de estas es α . Para un punto arbitrario $p+1 < t < n$, la varianza de u_t es $t - p$, por lo tanto, las bandas deben ser curvas de la forma $\pm \lambda \sqrt{t - p}$, donde λ es una constante.

Tomando como base el enfoque de Brown et al. (1975) las líneas deben ser tangentes a las curvas en el punto $t^* = p + (n - p)/2$, que se encuentra entre p y n . Se asume que la probabilidad de que la trayectoria cruce ambas bandas es insignificante. Por lo tanto, solo se considera la curva positiva $\lambda \sqrt{t - p}$ y $\lambda/2$ se toma como la probabilidad de cruce en lugar de α . Para tal línea se define la función

$$\xi(t) = a + bt.$$

Primero se calcula la pendiente b de esta función, que es la derivada de la función $f(t) = \lambda \sqrt{t - p}$ en el punto t^* la derivada es

$$f'(t) = \frac{\lambda}{2\sqrt{t - p}};$$

que en el punto de interés es

$$b = f'(t^*) = \frac{\lambda \sqrt{2}}{2} \frac{\sqrt{n - p}}{n - p}$$

Ahora, se calcula el intercepto a . Se considera $f(t^*)$, que es

$$f(t^*) = \frac{\lambda\sqrt{2}}{2}\sqrt{n-p}$$

Se realizan las respectivas sustituciones

$$\xi(t) = a + bt \Rightarrow a = \frac{\lambda\sqrt{2}}{2}\sqrt{n-p} \left[\frac{n-3p}{2(n-p)} \right]$$

La función $\xi(t)$ se convierte en

$$\xi(t) = \frac{\lambda\sqrt{2}}{4} \frac{\sqrt{n-p}}{n-p} (n-3p+2t), \quad p \leq t \leq n;$$

esta función puede ser descrita por dos puntos bien definidos, los cuales son p y n . Sustituyendo $d = \lambda\sqrt{2}/4$ estos dos puntos son $(p, d\sqrt{n-p})$ y $(n, 3d\sqrt{n-p})$. Para obtener una línea recta general para cualquier n y $p < n$ arbitrarios, se transforma t en el intervalo $[0, 1]$. En este camino algún índice $p \leq i \leq n$ puede ser dado por

$$i = [t(n-p) + p] \quad \text{para} \quad 0 \leq t \leq 1.$$

Entonces este punto comienza en $(0, d\sqrt{n-p})$ y $(1, 3d\sqrt{n-p})$, sustituyendo $v = d\sqrt{n-p}$, puede ser expresado como

$$\xi(t) = v + 2vt \quad \text{pero} \quad 0 \leq t \leq 1.$$

Para calcular v como una probabilidad cruzada, se tiene

$$U_n(t) = \frac{1}{\sigma} \sum_{j=p+1}^{p+t(n-p)} u_j^*, \quad 0 \leq t \leq 1.$$

Krämer et al. (1988) muestran que para $n \rightarrow \infty$

$$U_n(t) \rightarrow B(t)$$

converge en distribución, donde $B(t)$ es un movimiento Browniano estándar. La probabilidad asintótica para $n \rightarrow \infty$ que $U_n(t)$ cruce la línea $\xi(t)$ es

$$P(U_n(t) > v + 2vt \quad \text{para algún} \quad 0 \leq t \leq 1) = 1 - \Phi(3v) + \exp(-4v^2)\Phi(v) \quad (2.13)$$

Se asume que la probabilidad de que $U_n(t)$ cruce la línea más de una vez es cero. La observación después de la cual $U_n(t)$ cae fuera de estas líneas proporciona una estimación del punto de cambio.

Zeileis et al. (2002) revisaron las pruebas de cambio estructural en los modelos de regresión lineal desde el marco de la prueba de fluctuación generalizada y la prueba F. Introducen un marco unificado para implementar estas pruebas y evidencian la materialización de estas ideas a través de un paquete de R llamado **strucchange**. El paquete contiene métodos para ajustar, probar y trazar procesos de fluctuación empíricos (como cusum, mosum y procesos basados en estimaciones) y para calcular, probar y trazar secuencias de estadísticas F. Con este se dispone de potentes herramientas para mostrar

información sobre los cambios estructurales en las relaciones de regresión y evaluar su importancia.

2.1.4. Enfoque informativo

Un enfoque alternativo para estimar el punto de cambio en modelos de regresión lineal es emplear el criterio de información de Schwarz, SIC. Es un enfoque simple que reduce los cálculos en comparación con el enfoque de razón de verosimilitud (Chen & Gupta, 2012). En este caso, se debe probar la hipótesis nula $H_0 : \mu_i = \mathbf{x}'_i \boldsymbol{\beta}$ para $i = 1, \dots, n$, versus la hipótesis alternativa $H_1 : \mu_i = \mathbf{x}'_i \boldsymbol{\beta}_1$ para $i = 1, \dots, \tau$, y $\mu_i = \mathbf{x}'_i \boldsymbol{\beta}_2$ para $i = \tau + 1, \dots, n$.

La función de verosimilitud bajo H_0 , $L_0(\boldsymbol{\beta}, \sigma^2)$, es equivalente a la función (1.2), por lo tanto, los estimadores de máxima verosimilitud de $\boldsymbol{\beta}$ y σ^2 son, (1.4) y (1.5), respectivamente. Entonces, la máxima verosimilitud bajo H_0 es

$$L_0(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = (2\pi)^{-n/2} \left[\frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right]^{-n/2} \exp \left(-\frac{n}{2} \right).$$

Por lo tanto, bajo H_0 el criterio de información de Schwarz, denotado por $SIC(n)$, se obtiene como

$$\begin{aligned} SIC(n) &= -2 \log L_0(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) + (p+2) \log n \\ &= n \log [(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})] + n(\log 2\pi + 1) + (p+2-n) \log n. \end{aligned}$$

La hipótesis alternativa corresponde al modelo (2.2). En este caso, se encuentra que la función de verosimilitud es

$$\begin{aligned} L_1(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2) &= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp -\frac{1}{2\sigma^2} (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1)' (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1) \\ &\quad \times \exp -\frac{1}{2\sigma^2} (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)' (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2), \end{aligned}$$

y la estimación de máxima verosimilitud para β_d , $d = 1, 2$ y σ^2 están dadas por (2.5) y

$$\hat{\sigma}^2 = \frac{1}{n} [(\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1)' (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1) + (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)' (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)],$$

respectivamente. Entonces la máxima verosimilitud es

$$\begin{aligned} L_1(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}^2) &= (2\pi)^{-n/2} \left\{ \frac{1}{n} [(\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1)' (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1) \right. \\ &\quad \left. + (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)' (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)] \right\}^{-n/2} \exp \left(-\frac{n}{2} \right). \end{aligned}$$

Por lo tanto, bajo H_1 el criterio de información de Schwarz, denotado por $SIC(\tau)$ para $\tau = p+1, \dots, n-p$, se obtiene como

$$\begin{aligned} SIC(\tau) &= -2 \log L_1(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}^2) + (2p+3) \log n \\ &= n \log [(\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1)' (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1) + (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)' (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)] \\ &\quad + n(\log 2\pi + 1) + (2p+3-n) \log n. \end{aligned}$$

De acuerdo al principio de criterio información en la selección del modelo, H_0 debe ser aceptado si $SIC(n) \leq \min_{p+1 \leq \tau \leq n-p} SIC(\tau)$, y H_1 debe ser aceptado si $SIC(n) > \min_{p+1 \leq \tau \leq n-p} SIC(\tau)$. Cuando H_1 es aceptado, la estimación del punto de cambio debe ser $\hat{\tau}$ tal que

$$SIC(\hat{\tau}) = \min_{p+1 \leq \tau \leq n-p} SIC(\tau)$$

Propuestas Bayesianas

La Inferencia Bayesiana es el proceso de ajustar un modelo de probabilidad a un conjunto de datos y resumir el resultado por una distribución de probabilidad en los parámetros del modelo y en cantidades no observadas tales como predicciones para nuevas observaciones (Gelman et al., 2014). En algunas ocasiones, es posible tener información previa de los valores de los parámetros.

Gamerman (1997) propone una metodología de estimación bayesiana por bloques de parámetros, para el ajuste sobre modelos generalizados mixtos. Cepeda & Gamerman (2001) ajustan de manera conjunta la media y la estructura de varianza para un modelo de regresión lineal heterocedástico, empleando a priori normales para los parámetros de la media y la varianza. Una generalización de la propuesta de estimación conjunta de la estructura de media y varianza es presentada por Cepeda (2001) y Cepeda & Gamerman (2004). Por otro lado, Daniels & Pourahmadi (2002) desarrollaron propuestas bayesianas para ajustar modelos dinámicos longitudinales suponiendo a priori normales para los parámetros de media y varianza, además emplearon la Descomposición de Cholesky presentada por Pourahmadi (1999). Cepeda & Núñez Antón (2007) y Cepeda & Núñez Antón (2009) proponen metodologías bayesianas para ajustar de manera conjunta los parámetros de media y varianza en modelos longitudinales, para la estructura de varianza suponen un modelo de antedependencia no estructurado, este enfoque es aplicado sobre modelos lineales y no lineales. Más recientemente, Castillo et al. (2019) presenta un propuesta bayesiana para la estimación conjunta de los parámetros de media y la matriz de varianzas covarianzas para modelos longitudinales con estructuras de covarianza SC , $AR(1)$, $ARMA(1,1)$ y algunos casos de estructuras de antedependencia no estructurada $ADS(1)$.

Siguiendo la metodología introducida por Cepeda (2001), en este capítulo se presentan las propuestas Bayesianas para la estimación conjunta de la media, varianza y punto de cambio en modelos de regresión con respuesta normal homocedásticos y heterocedásticos. Posteriormente, en base a la metodología propuesta por Castillo et al. (2019) se expone la propuesta para la estimación conjunta de la media, varianza y punto de cambio para modelos longitudinales con respuesta normal, suponiendo que la matriz de covarianzas sigue alguna de las siguientes estructuras de covarianza: SC , $AR(1)$ o $ARMA(1,1)$. La estimación de los modelos se realiza usando el muestreador de Gibbs, el algoritmo de Metropolis Hastings por bloques y el algoritmo de Caminata Aleatoria Metropolis-Hastings, los cuales se exponen con detalle en el capítulo 1.

3.1. Propuesta Bayesiana para modelos lineales con un punto de cambio y varianza constante

Se tiene el modelo de regresión lineal con un punto de cambio,

$$y_i = \begin{cases} x_i' \beta_1 + \epsilon_{1i} & i = 1, \dots, \tau \\ x_i' \beta_2 + \epsilon_{2i} & i = \tau + 1, \dots, n, \end{cases} \quad (3.1)$$

donde \mathbf{y}_i , $i = 1, 2, 3, \dots, n$ denota las observaciones de la variable respuesta, $\mathbf{x}_i = (1, x_{i2}, \dots, x_{ip})'$ denota un vector columna de p covariables o variables explicativas independientes, sea β_d , vectores de tamaño $p \times 1$ desconocidos y sea ϵ_{di} errores independientes e idénticamente distribuidos con $\epsilon_{di} \sim N(\mathbf{0}, \sigma_d^2)$ con $d = 1, 2$. Si se denota $\mathbf{y}_1 = (y_1, \dots, y_\tau)'$ y $\mathbf{y}_2 = (y_{\tau+1}, \dots, y_n)'$ son vectores columna de las primeras τ y las últimas $n - \tau$ observaciones de la variable respuesta, respectivamente. $\mathbf{X}_1 = (x_1, \dots, x_\tau)'$ y $\mathbf{X}_2 = (x_{\tau+1}, \dots, x_n)'$ son los primeros τ y los últimos $n - \tau$ vectores fila de la matriz diseño, respectivamente. Además, $\epsilon_d \sim N(\mathbf{0}, \sigma_d^2 \mathbf{I}_d)$, donde \mathbf{I}_d es la matriz identidad con rango τ para $d = 1$ y rango $n - \tau$ para $d = 2$. Dado que no hay restricciones de continuidad y los modelos son independientes, estos pueden ser definido como

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}. \quad (3.2)$$

Para el modelo 3.2, dado que $\mathbf{Y}_d \sim (\mathbf{X}_d \beta_d, \sigma_d^2 \mathbf{I}_d)$, la función de verosimilitud está dada por

$$L(\beta, \sigma^2, \tau) = L(\beta_1, \sigma_1^2 | y_1, \dots, y_\tau) L(\beta_2, \sigma_2^2 | y_{\tau+1}, \dots, y_n).$$

donde $\beta = (\beta_1, \beta_2)'$, $\sigma^2 = (\sigma_1^2, \sigma_2^2)'$,

$$L(\beta_1, \sigma_1^2 | y_1, \dots, y_\tau) = (2\pi)^{-\tau/2} (\sigma_1^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_1^2} (\mathbf{y}_1 - \mathbf{X}_1 \beta_1)' (\mathbf{y}_1 - \mathbf{X}_1 \beta_1) \right\},$$

y

$$L(\beta_2, \sigma_2^2 | y_{\tau+1}, \dots, y_n) = (2\pi)^{-(n-\tau)/2} (\sigma_2^2)^{-(n-\tau)/2} \exp \left\{ -\frac{1}{2\sigma_2^2} (\mathbf{y}_2 - \mathbf{X}_2 \beta_2)' (\mathbf{y}_2 - \mathbf{X}_2 \beta_2) \right\}.$$

Estos términos corresponden a la verosimilitud de las primeras τ y las siguientes $n - \tau$ observaciones y ambos términos son mutuamente independientes. Si se considera $\alpha_d = (\beta_d, \sigma_d^2, \tau)'$, dado que se desea realizar una propuesta Bayesiana, es necesario definir una distribución a priori $p(\alpha_d)$. Si se supone independencia entre β_d , σ_d^2 y τ , entonces la distribución a posteriori $\pi(\alpha_d) = L(\beta_d, \sigma_d^2 | \mathbf{Y}_d) p(\beta_d) p(\sigma_d^2) p(\tau)$. Por lo tanto, si se suponen las distribuciones a priori

$$p(\beta_d) \sim N(\mathbf{b}_{0d}, \mathbf{B}_{0d}) \quad (3.3)$$

$$p\left(\frac{1}{\sigma_d^2}\right) = P(\phi_d) \sim G\left(\frac{n_{0d}}{2}, \frac{n_{0d}\sigma_{0d}^2}{2}\right) \quad (3.4)$$

$$p(\tau) \sim U(p+1, n-p-1) \quad (3.5)$$

La función de densidad conjunta a posteriori es intratable analíticamente, por lo tanto, se propone un procedimiento de inferencia utilizando un enfoque basado en muestreo empleando los métodos de cadenas de Markov en bloques. Por lo tanto, un paso intermedio

es obtener distribuciones condicionales completas a posteriori para los bloques β_d , σ_d^2 y τ , estos son dados por π_{β_d} , $\pi_{\sigma_d^2}$ y π_{τ} , respectivamente. De 3.3, se deduce que la distribución $p(\beta_d)$ está dada por

$$p(\beta_d) \propto \exp \left\{ -\frac{1}{2}(\beta_d - \mathbf{b}_{0d})' \mathbf{B}_{0d}^{-1}(\beta_d - \mathbf{b}_{0d}) \right\}$$

Entonces, empleando el teorema de Bayes, se concluye que la distribución condicional completa π_{β_d} está dada por

$$\pi_{\beta} \propto \exp \left\{ -\frac{1}{2}(\beta_d - \mathbf{b}_{1d})' \mathbf{B}_{1d}^{-1}(\beta_d - \mathbf{b}_{1d}) \right\} \quad (3.6)$$

en consecuencia,

$$\pi_{\beta} \sim N(\mathbf{b}_{1d}, \mathbf{B}_{1d}) \quad (3.7)$$

donde $\mathbf{b}_{1d} = \mathbf{B}_{1d}(\mathbf{B}_{0d}^{-1}\mathbf{b}_{0d} + \mathbf{X}'_d\mathbf{y}_d)$, $\mathbf{B}_{1d} = (\mathbf{B}_{0d}^{-1} + \mathbf{X}'_d\mathbf{X}_d)^{-1}$. Por lo tanto, para obtener los valores estimados de β_d empleando el muestreador de Gibbs, se tomaran muestras de la distribución 3.7 con probabilidad de aceptación 1. De 3.4, se obtiene que la distribución $p(\sigma_d^2)$ está dada por

$$p(\phi_d) = \frac{1}{\Gamma(n_{0d}/2)} \left(\frac{\sigma_{0d}^2}{2} \right)^{n_{0d}/2} \phi_d^{(n_{0d}/2-1)} \exp \left(-\frac{n_{0d}\sigma_{0d}^2}{2} \phi_d \right) I_{(0,\infty)}$$

Por lo tanto, la distribución condicional completa π_{ϕ_d} , se puede escribir como

$$\pi_{\phi_d} \propto \phi_d^{n_{1d}/2-1} \exp \left\{ -\frac{n_{1d}\sigma_{1d}^2}{2} \phi_d \right\}$$

que corresponde al kernel de la distribución gamma, notado como

$$\pi_{\phi_d} \sim G \left(\frac{n_{1d}}{2}, \frac{n_{1d}\sigma_{1d}^2}{2} \right). \quad (3.8)$$

Debido a que $\phi_d = 1/\sigma_d^2$ por propiedades de la distribución gamma, se obtiene que los valores de σ_d^2 se pueden obtener directamente empleando el muestreador de Gibbs de una distribución inversa gamma, definida como

$$\pi_{\sigma_d^2} \sim IG \left(\frac{n_{1d}}{2}, \frac{n_{1d}\sigma_{1d}^2}{2} \right), \quad (3.9)$$

donde $n_{11} = \tau + n_{01}$, $n_{12} = (n - \tau) + n_{02}$ y $n_{1d}\sigma_{1d}^2 = n_{0d}\sigma_{0d}^2 + (\mathbf{y}_d - \mathbf{X}_d\beta_d)'(\mathbf{y}_d - \mathbf{X}_d\beta_d)$.

Para construir una distribución a posteriori para el punto de cambio τ , debido a la restricción $\tau \in \{p + 1, \dots, n - p - 1\}$, Holbert (1982) propone una distribución a priori $p(\tau) \sim U(p + 1, n - p - 1)$ y de 3.5 se deduce que la distribución $p(\tau)$ está dada por

$$p(\tau) = \begin{cases} \frac{1}{n - 2(p + 1)}, & \tau = p + 1, \dots, n - p - 1 \\ 0, & \text{en otro caso} \end{cases}$$

Por lo tanto, la distribución a posteriori condicional completa para τ está dada por

$$\begin{aligned} \pi_\tau &= (2\pi)^{-n/2} \phi_1^{\tau/2} \exp \left\{ -\frac{\phi_1}{2} (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1)' (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1) \right\} \\ &\quad \times \phi_2^{(n-\tau)/2} \exp \left\{ -\frac{\phi_2}{2} (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)' (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2) \right\} \frac{1}{n-2(p+1)} \\ &\propto \phi_1^{\tau/2} \phi_2^{(n-\tau)/2} \exp \left\{ -\frac{1}{2} \left[\phi_1 (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1)' (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1) + \phi_2 (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)' (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2) \right] \right\}. \end{aligned} \quad (3.10)$$

Bajo el supuesto $\sigma_1^2 = \sigma_2^2 = \sigma^2$ la función de verosimilitud es

$$L(\boldsymbol{\beta}, \sigma^2, \tau) = L(\boldsymbol{\beta}_1, \sigma^2 | y_1, \dots, y_\tau) L(\boldsymbol{\beta}_2, \sigma^2 | y_{\tau+1}, \dots, y_n).$$

Entonces, la distribución condicional completa para ϕ está dada por

$$\pi_\phi \sim G\left(\frac{n_1}{2}, \frac{n_1 \sigma_1^2}{2}\right).$$

donde $n_1 = n + n_0$ y $n_1 \sigma_1^2 = n_0 \sigma_0^2 + (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1)' (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1) + (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)' (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)$. De lo anterior, se obtiene que la distribución condicional completa para τ está definida como

$$\pi_\tau \propto \phi^{n/2} \exp \left\{ -\frac{\phi}{2} \left[(\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1)' (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1) + (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)' (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2) \right] \right\}. \quad (3.11)$$

Como la función a posteriori 3.10 y 3.11 no tiene forma de una distribución conocida para obtener simulaciones de Monte Carlo Cadenas de Markov (MCMC), se propone una caminata aleatoria como kernel de transición. Por lo tanto, el kernel de transición para τ se define como

$$q(\tau^{(*)} | \tau^{(j)}) = \tau^{(*)} \sim U(\tau^{(j)} - \delta, \tau^{(j)} + \delta),$$

Se encontró que este tipo de kernel obtiene valores fuera del rango $\tau \in \{p+1, \dots, n-p-1\}$. Para evitar obtener valores fuera del intervalo objetivo y no realizar transformaciones sobre los valores obtenidos, se propone la siguiente caminata aleatoria como kernel de transición

$$q(\tau^{(*)} | \tau^{(j)}) = \begin{cases} \tau^{(*)} \sim U(p+1, \tau^{(j)} + 1), & |\tau^{(j)} - (p+1)| \leq 1 \\ \tau^{(*)} \sim U(\tau^{(j)} - 1, n-p-1), & |(n-p-1) - \tau^{(j)}| \leq 1 \\ \tau^{(*)} \sim U(\tau^{(j)} - 1, \tau^{(j)} + 1), & \text{en otro caso} \end{cases} \quad (3.12)$$

Por lo tanto, para obtener las estimaciones conjuntas de los parámetros de media y varianza para modelos con un punto de cambio, el algoritmo es el siguiente

1. Iniciar el contador en $j = 1$ y dar valores iniciales $\boldsymbol{\beta}_d^{(0)}$, $\sigma_d^{2(0)}$ y $\tau^{(0)} \in \mathbb{Z}$.
2. Mediante el muestreador de Gibbs obtener $\boldsymbol{\beta}_1^{(j+1)}$ y $\boldsymbol{\beta}_2^{(j+1)}$, a partir de la distribución normal multivariada 3.7.
3. Mediante el muestreador de Gibbs obtener $\sigma_1^{2(j+1)}$ y $\sigma_2^{2(j+1)}$, a partir de la distribución inversa gamma 3.9.
4. Proponer un nuevo valor $\tau^{(*)}$ generado de la densidad propuesta 3.12.

5. Aproximar $\tau^{(*)}$ al entero más cercano y calcular la probabilidad de aceptación

$$\alpha(\tau^{(j)}, \tau^{(*)}) = \min \left(1, \frac{\pi(\mathbf{y}|\boldsymbol{\beta}_1^{(j+1)}, \boldsymbol{\beta}_2^{(j+1)}, \sigma_1^{2(j+1)}, \sigma_2^{2(j+1)}, \tau^{(*)})}{\pi(\mathbf{y}|\boldsymbol{\beta}_1^{(j+1)}, \boldsymbol{\beta}_2^{(j+1)}, \sigma_1^{2(j+1)}, \sigma_2^{2(j+1)}, \tau^{(j)})} \right).$$

Si el movimiento es aceptado, entonces $\tau^{(j+1)} = \tau^{(*)}$. Si el movimiento es rechazado, entonces $\tau^{(j+1)} = \tau^{(j)}$ y la cadena no se mueve.

6. Retornar al paso 2 hasta obtener convergencia.

El paso 5 se implementa computacionalmente generando una cantidad u de una distribución uniforme en el intervalo $(0, 1)$, independiente de los parámetros. Si $u \leq \alpha$ el movimiento es aceptado y si $u > \alpha$ el movimiento es rechazado.

3.2. Propuesta Bayesiana para modelos lineales con un punto de cambio y varianza heterogénea

Se tiene el modelo de regresión normal con un punto de cambio en donde se modela a través de covariables la heterogeneidad de la varianza,

$$y_i = \begin{cases} x_i' \boldsymbol{\beta}_1 + \epsilon_{1i} & i = 1, \dots, \tau \\ x_i' \boldsymbol{\beta}_2 + \epsilon_{2i} & i = \tau + 1, \dots, n, \end{cases} \quad (3.13)$$

con $\epsilon_{di} \sim N(0, \sigma_{di}^2)$, $d = 1, 2$. Se tiene que \mathbf{y}_i , $i = 1, \dots, n$ denota las observaciones de la variable respuesta, $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})'$ denota un vector columna de variables explicativas, $\mathbf{z}_i = (z_{i1}, \dots, z_{ir})'$ es un vector columna de variables explicativas de la varianza que puede contener algunas o todas las variables incluidas en \mathbf{x}_i . Donde $\boldsymbol{\mu}_1 = \mathbf{x}_1' \boldsymbol{\beta}_1$ y $\boldsymbol{\mu}_2 = \mathbf{x}_2' \boldsymbol{\beta}_2$ son los modelos de media de las primeras τ y las últimas $n - \tau$ observaciones, respectivamente. Por otro lado, se tiene que $g(\sigma_1^2) = \mathbf{z}_1' \boldsymbol{\gamma}_1$ y $g(\sigma_2^2) = \mathbf{z}_2' \boldsymbol{\gamma}_2$ son los modelos de varianza de las primeras τ y las últimas $n - \tau$ observaciones, respectivamente. La función g debe ser monótona, diferenciable y debe garantizar que la varianza sea positiva. En esta investigación, siguiente la metodología expuesta por Cepeda & Gamerman (2001) se considera $g = \log$.

Para el modelo 3.13 con $g = \log$, dado que $Y_d \sim (X_d \boldsymbol{\beta}_d, \Sigma_d)$ con $\Sigma_d = \text{diag}(\sigma_{di}^2)$ y por independencia la función de verosimilitud está dado por

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}^2, \tau) = L(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1 | y_1, \dots, y_\tau) L(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2 | y_{\tau+1}, \dots, y_n).$$

donde $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)'$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)'$,

$$L(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1 | y_1, \dots, y_\tau) \propto |\boldsymbol{\Sigma}_1|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1)' \boldsymbol{\Sigma}_1^{-1} (\mathbf{Y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1) \right\}$$

y

$$L(\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2 | y_{\tau+1}, \dots, y_n) \propto |\boldsymbol{\Sigma}_2|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)' \boldsymbol{\Sigma}_2^{-1} (\mathbf{Y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2) \right\}.$$

Estos términos corresponden a la verosimilitud de las primeras τ y las siguientes $n - \tau$ observaciones y ambos son mutuamente independientes. Si se considera $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau)'$, dado que se desea realizar una propuesta Bayesiana, es necesario definir una distribución a priori $p(\boldsymbol{\alpha})$. Si se supone independencia entre $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ y τ y usando el teorema de Bayes se obtiene la distribución a posteriori $\pi(\boldsymbol{\alpha}) = L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau | \mathbf{Y})p(\boldsymbol{\beta})p(\boldsymbol{\gamma})p(\tau)$. Por lo tanto, si se suponen las distribuciones a priori

$$p(\boldsymbol{\beta}_d) \sim N(\mathbf{b}_{0d}, \mathbf{B}_{0d}) \quad (3.14)$$

$$p(\boldsymbol{\gamma}_d) \sim N(\mathbf{g}_{0d}, \mathbf{G}_{0d}) \quad (3.15)$$

$$p(\tau) \sim U(p + 1, n - p - 1) \quad (3.16)$$

La función de densidad conjunta a posteriori es intratable analíticamente, por lo tanto, se propone un procedimiento de inferencia utilizando un enfoque basado en muestreo empleando los métodos de cadenas de Markov en bloques. Al considerar $\boldsymbol{\alpha}$ conformado por los bloques $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ y τ , un paso intermedio es obtener distribuciones condicionales completas a posteriori para cada uno de ellos, estos son dados por $\pi_{\beta_d}, \pi_{\gamma_d}$ y π_{τ} , respectivamente. De 3.14, se deduce que la distribución $p(\boldsymbol{\beta}_d)$ está dada por

$$p(\boldsymbol{\beta}_d) \propto \exp \left\{ -\frac{1}{2}(\boldsymbol{\beta}_d - \mathbf{b}_{0d})' \mathbf{B}_{0d}^{-1}(\boldsymbol{\beta}_d - \mathbf{b}_{0d}) \right\}$$

Entonces, empleando el teorema de Bayes, se concluye que la distribución condicional completa π_{β_d} está dada por

$$\pi_{\beta} \propto \exp \left\{ -\frac{1}{2}(\boldsymbol{\beta}_d - \mathbf{b}_{1d})' \mathbf{B}_{1d}^{-1}(\boldsymbol{\beta}_d - \mathbf{b}_{1d}) \right\} \quad (3.17)$$

en consecuencia,

$$\pi_{\beta} \sim N(\mathbf{b}_{1d}, \mathbf{B}_{1d}) \quad (3.18)$$

donde $\mathbf{b}_{1d} = \mathbf{B}_{1d}(\mathbf{B}_{0d}^{-1}\mathbf{b}_{0d} + \mathbf{X}'_d\boldsymbol{\Sigma}_d^{-1}\mathbf{y}_d)$, $\mathbf{B}_{1d} = (\mathbf{B}_{0d}^{-1} + \mathbf{X}'_d\boldsymbol{\Sigma}_d^{-1}\mathbf{X}_d)^{-1}$. Por lo tanto, para obtener los valores estimados de $\boldsymbol{\beta}_d$ empleando el muestreador de Gibbs, se tomaran muestras de la distribución 3.18 con probabilidad de aceptación 1.

A diferencia de la distribución condicional completa de $\boldsymbol{\beta}$, la distribución condicional completa de $\boldsymbol{\gamma}$ es intratable analíticamente y no resulta fácil de generar. Para esto se emplea la metodología propuesta por Cepeda & Gamerman (2001) y algunos elementos del algoritmo de Fisher Scoring. Con $g = \log$, se obtienen los vectores de observaciones de trabajo $\tilde{\mathbf{Y}}_1 = (\tilde{y}_1, \dots, \tilde{y}_\tau)$ y $\tilde{\mathbf{Y}}_2 = (\tilde{y}_{\tau+1}, \dots, \tilde{y}_n)$, con variables de trabajo para los dos segmentos

$$\tilde{y}_{i1} = \mathbf{z}'_{i1}\boldsymbol{\gamma}_1^{(j)} + \frac{(y_{i1} - \mathbf{x}'_{i1}\boldsymbol{\beta}_1^{(j)})^2}{\exp(\mathbf{z}'_{i1}\boldsymbol{\gamma}_1^{(j)})} - 1, \quad i = 1, \dots, \tau,$$

y

$$\tilde{y}_{i2} = \mathbf{z}'_{i2}\boldsymbol{\gamma}_2^{(j)} + \frac{(y_{i2} - \mathbf{x}'_{i2}\boldsymbol{\beta}_2^{(j)})^2}{\exp(\mathbf{z}'_{i2}\boldsymbol{\gamma}_2^{(j)})} - 1, \quad i = \tau + 1, \dots, n.$$

y varianzas de trabajo asociadas igual a 2. Los kernel de transición q_{γ_1} y q_{γ_2} se obtienen como

$$q_{\gamma_1}(\boldsymbol{\gamma}^{(j)}, \boldsymbol{\gamma}^{(\tau)}) = N(\mathbf{g}_{11}, \mathbf{G}_{11}) \quad (3.19)$$

y

$$q_{\gamma_2}(\boldsymbol{\gamma}^{(j)}, \boldsymbol{\gamma}^{(n)}) = N(\mathbf{g}_{12}, \mathbf{G}_{12}) \quad (3.20)$$

donde $\mathbf{g}_{1d} = \mathbf{G}_{1d}(\mathbf{G}_{0d}^{-1}\mathbf{g}_{0d} + 1/2\mathbf{Z}'_d\tilde{\mathbf{Y}}_d)$ y $\mathbf{G}_{1d} = (\mathbf{G}_{0d}^{-1} + 1/2\mathbf{Z}'_d\mathbf{Z}_d)^{-1}$. Esta distribución propuesta es la misma distribución a posteriori de $\boldsymbol{\gamma}_d$ que se habría obtenido de la combinación de la a priori con el modelo con las observaciones de trabajo

$$\tilde{y}_{i1} \sim N(\mathbf{z}'_{i1}\boldsymbol{\gamma}_1, 2), \quad i = 1, \dots, \tau \quad \text{y} \quad \tilde{y}_{i2} \sim N(\mathbf{z}'_{i2}\boldsymbol{\gamma}_2, 2), \quad i = \tau + 1, \dots, n.$$

La distribución condicional completa para τ es intratable analíticamente y resulta difícil de generar. Se propone tomar como kernel de transición la caminata aleatoria planteada en 3.12, este tiene la ventaja de generar valores dentro del intervalo objetivo y por lo tanto, no es necesario aplicar transformaciones sobre los resultados obtenidos.

Con estas condiciones, para encontrar las estimaciones conjuntas de los parámetros de media y varianza para modelos lineales con un punto de cambio y varianza heterogénea el algoritmo es el siguiente:

1. Iniciar el contador en $j = 1$ y dar valores iniciales $\boldsymbol{\beta}_d^{(0)}, \gamma_d^{(0)}$ y $\tau^{(0)} \in \mathbb{Z}$.
2. Mediante el muestreador de Gibbs obtener $\boldsymbol{\beta}_1^{(j+1)}$ y $\boldsymbol{\beta}_2^{(j+1)}$, a partir de la distribución normal multivariada 3.18.
3. Obtener muestras de $q(\gamma_1^{(*)}|\gamma_1^{(j)})$ y $q(\gamma_2^{(*)}|\gamma_2^{(j)})$ desde los kernel de transición 3.19 y 3.20, respectivamente.
4. Calcular la probabilidad de aceptación para γ_1 y γ_2 , respectivamente.

$$\alpha_1(\gamma_d^{(j)}, \gamma_d^{(*)}) = \min \left(1, \frac{\pi(\mathbf{y}|\boldsymbol{\beta}_d^{(j+1)}, \gamma_d^{(*)}, \tau^{(j)})q(\gamma_d^{(*)}, \gamma_d^{(j)})}{\pi(\mathbf{y}|\boldsymbol{\beta}_d^{(j+1)}, \gamma_d^{(j)}, \tau^{(j)})q(\gamma_d^{(j)}, \gamma_d^{(*)})} \right).$$

5. Generar u_1 de una distribución uniforme en el intervalo $(0, 1)$, independiente de los parámetros. Si $u_1 \leq \alpha_1$ el movimiento es aceptado y $\gamma_d^{(j+1)} = \gamma_d^{(*)}$, si $u_1 > \alpha_1$ el movimiento es rechazado y $\gamma_d^{(j+1)} = \gamma_d^{(j)}$.
6. Proponer un nuevo valor $\tau^{(*)}$ generado de la densidad propuesta 3.12.
7. Aproximar $\tau^{(*)}$ al entero más cercano y calcular la probabilidad de aceptación para τ

$$\alpha_2(\tau^{(j)}, \tau^{(*)}) = \min \left(1, \frac{\pi(\mathbf{y}|\boldsymbol{\beta}_1^{(j+1)}, \boldsymbol{\beta}_2^{(j+1)}, \gamma_1^{(j+1)}, \gamma_2^{(j+1)}, \tau^{(*)})}{\pi(\mathbf{y}|\boldsymbol{\beta}_1^{(j+1)}, \boldsymbol{\beta}_2^{(j+1)}, \gamma_1^{(j+1)}, \gamma_2^{(j+1)}, \tau^{(j)})} \right).$$

8. Generar u_2 de una distribución uniforme en el intervalo $(0, 1)$, independiente de los parámetros. Si $u_2 \leq \alpha_2$ el movimiento es aceptado y $\tau^{(j+1)} = \tau^{(*)}$, si $u_2 > \alpha_2$ el movimiento es rechazado y $\tau^{(j+1)} = \tau^{(j)}$.
9. Retornar al paso 2 hasta obtener convergencia.

3.3. Propuesta Bayesiana para modelos longitudinales con un punto de cambio y estructura de covarianza SC y AR(1)

Se tiene el modelo de regresión longitudinal para el i -ésimo individuo con un punto de cambio,

$$\mathbf{y}_{it} = \begin{cases} \mathbf{x}'_{it}\boldsymbol{\beta}_1 + \epsilon_{1t} & t = 1, \dots, \tau \\ \mathbf{x}'_{it}\boldsymbol{\beta}_2 + \epsilon_{2t} & t = \tau + 1, \dots, m, \end{cases} \quad (3.21)$$

donde \mathbf{y}_{it} es el vector de mediciones realizadas en el i -ésimo individuo en el tiempo $t = 1, 2, \dots, m$. Se tiene que $\epsilon_{td} \sim N(0, \sigma_d^2 \mathbf{V}_{0d})$, donde $d = 1, 2$, además se tiene un vector de covariables asociadas \mathbf{x}_{it} donde $\mathbf{X}_{i1} = (1, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{i\tau})'$ de tamaño $\tau \times (p + 1)$ y $\mathbf{X}_{i2} = (1, \mathbf{x}_{i(\tau+1)}, \mathbf{x}_{i(\tau+2)}, \dots, \mathbf{x}_{im})'$ son las matrices de covariables de las primeras τ y las últimas $m - \tau$ mediciones asociadas al i -ésimo individuo, respectivamente, y $\boldsymbol{\beta}_d$ son los vectores de parámetros desconocidos para los dos segmentos del modelo. Si se denota $\mathbf{y}_{i1} = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{i\tau})'$ y $\mathbf{y}_{i2} = (\mathbf{y}_{i(\tau+1)}, \mathbf{y}_{i(\tau+2)}, \dots, \mathbf{y}_{im})'$, entonces $\mathbf{y}_1 = (\mathbf{y}_{11}, \mathbf{y}_{21}, \dots, \mathbf{y}_{n1})'$ y $\mathbf{y}_2 = (\mathbf{y}_{12}, \mathbf{y}_{22}, \dots, \mathbf{y}_{n2})'$ son los vectores que contienen $N_1 = n \times \tau$ y $N_2 = n \times (m - \tau)$, observaciones de la variable respuesta, respectivamente. $\mathbf{X}_1 = (\mathbf{X}_{11}, \mathbf{X}_{21}, \dots, \mathbf{X}_{n1})'$ y $\mathbf{X}_2 = (\mathbf{X}_{12}, \mathbf{X}_{22}, \dots, \mathbf{X}_{n2})'$, sea $\sigma_d^2 \mathbf{V}_d = \boldsymbol{\Sigma}_d(\sigma_d^2 \boldsymbol{\theta}_d)$ una matriz diagonal con bloques no nulos $\sigma_d^2 \mathbf{V}_{d0}$ de tamaño $\tau \times \tau$ para las primeras τ mediciones y tamaño $(m - \tau) \times (m - \tau)$ para las últimas $m - \tau$, donde cada bloque representa la matriz de covarianzas para cada individuo en los dos segmentos, con $\boldsymbol{\theta}_d$ un conjunto de parámetros dependiendo del tipo de estructura de covarianza que se ajuste para modelar los datos.

Para el modelo 3.21, dado que $\mathbf{y}_d \sim N(\mathbf{X}_d \boldsymbol{\beta}_d, \boldsymbol{\Sigma}_d(\mathbf{t}_d, \boldsymbol{\theta}_d))$ la función de verosimilitud está dada por

$$L(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \tau) = L(\boldsymbol{\beta}_1, \sigma_1^2, \boldsymbol{\theta}_1 | y_{11}, \dots, y_{n\tau}) L(\boldsymbol{\beta}_2, \sigma_2^2, \boldsymbol{\theta}_2 | y_{1(\tau+1)}, \dots, y_{nm}).$$

donde $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)'$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'$,

$$L(\boldsymbol{\beta}_1, \sigma_1^2, \boldsymbol{\theta}_1 | y_{11}, \dots, y_{n\tau}) \propto \prod_{i=1}^n \prod_{t=1}^{\tau} |\boldsymbol{\Sigma}_1|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_{it} - \mathbf{X}_{it} \boldsymbol{\beta}_1)' \boldsymbol{\Sigma}_1^{-1} (\mathbf{y}_{it} - \mathbf{X}_{it} \boldsymbol{\beta}_1) \right\}$$

y

$$L(\boldsymbol{\beta}_2, \sigma_2^2, \boldsymbol{\theta}_2 | y_{1(\tau+1)}, \dots, y_{nm}) \propto \prod_{i=1}^n \prod_{t=\tau+1}^m |\boldsymbol{\Sigma}_2|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_{it} - \mathbf{X}_{it} \boldsymbol{\beta}_2)' \boldsymbol{\Sigma}_2^{-1} (\mathbf{y}_{it} - \mathbf{X}_{it} \boldsymbol{\beta}_2) \right\}$$

Estos términos corresponden a la verosimilitud de las primeras $n\tau$ y las siguientes $n(m - \tau)$ observaciones y ambos son mutuamente independientes. Si se considera $\boldsymbol{\alpha} = (\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \tau)$, es necesario definir una distribución a priori $p(\boldsymbol{\alpha})$. Si se supone independencia entre $\boldsymbol{\beta}$, $\boldsymbol{\sigma}^2$, $\boldsymbol{\theta}$ y τ al usar el Teorema de Bayes se obtiene la distribución a posteriori $\pi(\boldsymbol{\alpha}) = L(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \tau | \mathbf{y}) p(\boldsymbol{\beta}) p(\boldsymbol{\sigma}^2) p(\boldsymbol{\theta}) p(\tau)$. Por lo anterior, es necesario definir las distribuciones a priori

$$p(\boldsymbol{\beta}_d) \sim N(\mathbf{b}_{0d}, \mathbf{B}_{0d}) \quad (3.22)$$

$$p\left(\frac{1}{\sigma_d^2}\right) = p(\phi_d) \sim \Gamma\left(\frac{n_{0d}}{2}, \frac{n_{0d}\sigma_{0d}^2}{2}\right) \quad (3.23)$$

$$p(\tau) \sim U(2, m - 1) \quad (3.24)$$

Donde la distribución a priori $p(\boldsymbol{\theta})$ se define según la estructura de la matriz de covarianza. De 3.22, se deduce que la distribución a priori para $\boldsymbol{\beta}$ es

$$p(\boldsymbol{\beta}_d) \propto \exp \left\{ -\frac{1}{2}(\boldsymbol{\beta}_d - \mathbf{b}_{0d})' \mathbf{B}_0^{-1}(\boldsymbol{\beta}_d - \mathbf{b}_{0d}) \right\}.$$

Entonces, empleando el teorema de Bayes, se concluye que la distribución condicional completa a posteriori $\pi(\boldsymbol{\beta}_d | \sigma_d^2, \boldsymbol{\theta}_d, \mathbf{y}_d)$ está dada por

$$\pi_{\beta_d} \propto \exp \left\{ -\frac{1}{2}(\boldsymbol{\beta}_d - \mathbf{b}_{1d})' \mathbf{B}_{1d}^{-1}(\boldsymbol{\beta}_d - \mathbf{b}_{1d}) \right\},$$

por lo tanto,

$$\pi_{\beta_d} \sim N(\mathbf{b}_{1d}, \mathbf{B}_{1d}) \quad (3.25)$$

donde $\mathbf{b}_{1d} = \mathbf{B}_{1d}(\mathbf{X}'_d \boldsymbol{\Sigma}_{1d}^{-1} \mathbf{y} + \mathbf{B}_{0d}^{-1} \mathbf{b}_{0d})$ y matriz de varianza $\mathbf{B}_{1d} = (\mathbf{X}'_d \boldsymbol{\Sigma}_{1d}^{-1} \mathbf{X}_d + \mathbf{B}_{0d}^{-1})^{-1}$. Por lo tanto, para obtener los valores estimados de $\boldsymbol{\beta}_d$ empleando el muestreador de Gibbs, se tomaran muestras de la distribución 3.25 con probabilidad de aceptación 1.

Las estructuras de la matriz de covarianzas $\boldsymbol{\Sigma}_d(\sigma_d^2, \boldsymbol{\theta}_d)$ *SC* y *AR*(1), presentadas en 1.14 y 1.15, respectivamente, dependen de los parámetros σ_d^2 y ρ_d . Por lo tanto, es necesario establecer las distribuciones condicionales completas a posteriori para σ_d^2 y ρ_d .

Para obtener la distribución condicional completa a posteriori para σ_d^2 es importante notar que $\boldsymbol{\Sigma}_d = \sigma_d^2 \mathbf{V}_d$ y $\boldsymbol{\Sigma}_d^{-1} = \frac{1}{\sigma_d^2} \mathbf{V}_d^{-1}$, donde \mathbf{V}_d es una matriz diagonal por bloques que depende del parámetro de correlación ρ_d . De 3.23 se obtiene que la distribución $p(\phi_d)$ está dada por

$$p(\phi_d) = \frac{1}{\Gamma(n_{0d}/2)} \left(\frac{n_{0d}\sigma_{0d}^2}{2} \right)^{n_{0d}/2} \phi_d^{(n_{0d}/2-1)} \exp \left(-\frac{n_{0d}\sigma_{0d}^2}{2} \phi_d \right) I_{(0,\infty)}$$

Por lo tanto, la distribución condicional completa π_{ϕ_d} , se puede escribir como

$$\pi_{\phi_d} \propto \phi_d^{n_{1d}/2-1} \exp \left\{ -\frac{n_{1d}\sigma_{1d}^2}{2} \phi_d \right\},$$

que corresponde al kernel de la distribución gamma, notado como

$$\pi_{\phi_d} \sim G \left(\frac{n_{1d}}{2}, \frac{n_{1d}\sigma_{1d}^2}{2} \right). \quad (3.26)$$

Debido a que $\phi_d = \frac{1}{\sigma_d^2}$ por propiedades de la distribución gamma, se obtiene que los valores de σ_d^2 se pueden obtener directamente empleando el muestreador de Gibbs de una distribución inversa gamma, definida como

$$\pi_{\sigma_d^2} \sim IG \left(\frac{n_{1d}}{2}, \frac{n_{1d}\sigma_{1d}^2}{2} \right), \quad (3.27)$$

donde $n_{11} = n\tau + n_{01}$, $n_{12} = n(m - \tau) + n_{02}$ y $n_{1d}\sigma_{1d}^2 = n_{0d}\sigma_{0d}^2 + (\mathbf{y}_d - \mathbf{X}_d\boldsymbol{\beta}_d)'(\mathbf{y}_d - \mathbf{X}_d\boldsymbol{\beta}_d)$.

Para definir la distribución a posteriori del parámetro de correlación ρ , Gelman et al. (2014) propone reemplazar la distribución a priori $U(-1, 1)$ con $p(\rho) \propto (1 - \rho)(1 + \rho)$,

que es equivalente a la distribución $Beta(2, 2)$ en el parámetro transformado $\frac{\rho+1}{2}$. Por otro lado, Hoff (2009) propone una distribución a priori $p(\rho) \sim U(0, 1)$. Posteriormente, Castillo et al. (2019) en busca de una propuesta más flexible y con el objetivo de poder construir distribución a priori informativas, supone la distribución a priori para el parámetro de correlación $p(\rho) \sim Beta(a, b)$.

Para definir una distribución condicional completa a posteriori para el parámetro de correlación en los dos segmentos ρ_d , siguiendo la propuesta de Castillo et al. (2019) se supone la distribución a priori,

$$p(\rho_d) \sim Beta(a_d, b_d), \quad (3.28)$$

en caso de no tener información sobre el comportamiento de la correlación ρ_d , se puede hacer uso del caso especial de la distribución $Beta(1, 1)$, que coincide con la distribución uniforme en el intervalo $[0, 1]$. En general, la función de densidad de probabilidad está dada por

$$p(\rho_d) = \frac{\Gamma(a_d + b_d)}{\Gamma(a_d)\Gamma(b_d)} \rho_d^{a_d-1} (1 - \rho_d)^{b_d-1} I_{(0,1)}.$$

Por lo tanto, la distribución condicional completa a posteriori para ρ_d está dada por,

$$\pi_{\rho_d} \propto \exp \left\{ -\frac{1}{2}(\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}_d)' \boldsymbol{\Sigma}_d^{-1} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}_d) + (a_d - 1) \ln(\rho_d) + (b_d - 1) \ln(1 - \rho_d) \right\} \quad (3.29)$$

La función 3.29 no tiene forma de una distribución conocida. Para hacer uso de la simulación MCMC y obtener muestras a posteriori del parámetro de correlación es necesario definir un kernel de transición en los dos segmentos del modelo.

Hoff (2009) propone una caminata aleatoria uniforme como kernel de transición, con este tipo de metodologías se obtienen valores fuera del rango de la función y es necesario realizar transformaciones sobre los valores muestreados.

Para no obtener valores fuera del intervalo de interés y, siguiendo la metodología de Castillo et al. (2019) se propone como kernel de transición para cada uno de los segmentos del modelo de regresión la caminata aleatoria.

$$q(\rho_d^{(*)} | \rho_d^{(j)}) = \begin{cases} \rho_d^{(*)} \sim U(0, 2\rho_d^{(j)}), & \rho_d^{(j)} \leq 0.5 \\ \rho_d^{(*)} \sim U(2\rho_d^{(j)} - 1, 1), & \rho_d^{(j)} > 0.5 \end{cases} \quad (3.30)$$

Para construir una distribución a posteriori para el punto de cambio τ , debido a la restricción $\tau \in \{2, \dots, m-1\}$, se propone una distribución a priori $p(\tau) \sim U(2, m-1)$ y de 3.24 se deduce que la distribución $p(\tau)$ está dada por

$$p(\tau) = \begin{cases} \frac{1}{m-3}, & \tau = 2, \dots, m-1 \\ 0, & \text{en otro caso} \end{cases}$$

Por lo tanto, la distribución a posteriori condicional completa para τ está dada por

$$\pi_{\tau} \propto |\boldsymbol{\Sigma}_1|^{N_1/2} |\boldsymbol{\Sigma}_2|^{N_2/2} \exp \left\{ -\frac{1}{2} \left[(\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1)' \boldsymbol{\Sigma}_1^{-1} (\mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1) + (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2)' \boldsymbol{\Sigma}_2^{-1} (\mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2) \right] \right\} \quad (3.31)$$

Como la función a posteriori 3.31 no tiene forma de una distribución conocida, para obtener simulaciones de Monte Carlo Cadenas de Markov (MCMC) se propone una caminata aleatoria como kernel de transición. Por lo tanto, el kernel de transición para τ se define como

$$q(\tau^{(*)}|\tau^{(j)}) = \tau^{(*)} \sim U(\tau^{(j)} - \delta, \tau^{(j)} + \delta),$$

Se encontró que este tipo de kernel obtiene valores fuera del rango $\tau \in \{2, \dots, m-1\}$. Para evitar obtener valores fuera del intervalo objetivo y no realizar transformaciones sobre los valores obtenidos, se propone la siguiente caminata aleatoria como kernel de transición

$$q(\tau^{(*)}|\tau^{(j)}) = \begin{cases} \tau^{(*)} \sim U(2, \tau^{(j)} + 1), & |\tau^{(j)} - 2| \leq 1 \\ \tau^{(*)} \sim U(\tau^{(j)} - 1, m - 1), & |(m - 1) - \tau^{(j)}| \leq 1 \\ \tau^{(*)} \sim U(\tau^{(j)} - 1, \tau^{(j)} + 1), & \text{en otro caso} \end{cases} \quad (3.32)$$

Por lo tanto, para obtener las estimaciones conjuntas de los parámetros de media y varianza para modelos con un punto de cambio, el algoritmo es el siguiente

1. Iniciar el contador $j = 1$ y dar valores iniciales $\beta_d^{(0)}$, $\sigma_d^{2(0)}$, $\rho_d^{(0)}$ y $\tau^{(0)} \in \mathbb{Z}$.
2. Mediante el muestreador de Gibbs obtener $\beta_1^{(j+1)}$ y $\beta_2^{(j+1)}$, a partir de la distribución normal multivariada 3.25.
3. Mediante el muestreador de Gibbs obtener $\sigma_1^{2(j+1)}$ y $\sigma_2^{2(j+1)}$, a partir de la distribución gamma 3.27.
4. Obtener muestras $q(\rho_d^{(*)}|\rho_d^{(j)})$ desde el kernel de transición 3.30.
5. Se calcula el radio de aceptación

$$\alpha(\rho_d^{(j)}, \rho_d^{(*)}) = \min \left(1, \frac{P(\mathbf{y}_d | \beta_d^{(j+1)}, \sigma_d^{2(j+1)}, \rho_d^{(*)}, \tau^j) p(\rho_d^{(*)})}{P(\mathbf{y}_d | \beta_d^{(j+1)}, \sigma_d^{2(j+1)}, \rho_d^{(j)}, \tau^j) p(\rho_d^{(j)})} \right).$$

Si el movimiento es aceptado, entonces $\rho_d^{(j+1)} = \rho_d^{(*)}$. Si el movimiento es rechazado, entonces $\rho_d^{(j+1)} = \rho_d^{(j)}$ y la cadena no se mueve.

6. Proponer un nuevo valor $\tau^{(*)}$ generado de la densidad propuesta 3.32.
7. Aproximar $\tau^{(*)}$ al entero más cercano y calcular la probabilidad de aceptación

$$\alpha(\tau^{(j)}, \tau^{(*)}) = \min \left(1, \frac{\pi(\mathbf{y} | \beta_1^{(j+1)}, \beta_2^{(j+1)}, \sigma_1^{2(j+1)}, \sigma_2^{2(j+1)}, \rho_1^{(j+1)}, \rho_2^{(j+1)}, \tau^{(*)})}{\pi(\mathbf{y} | \beta_1^{(j+1)}, \beta_2^{(j+1)}, \sigma_1^{2(j+1)}, \sigma_2^{2(j+1)}, \rho_1^{(j+1)}, \rho_2^{(j+1)}, \tau^{(j)})} \right).$$

Si el movimiento es aceptado, entonces $\tau^{(j+1)} = \tau^{(*)}$. Si el movimiento es rechazado, entonces $\tau^{(j+1)} = \tau^{(j)}$ y la cadena no se mueve.

8. Retornar al paso 2 hasta obtener convergencia.

Los pasos 5 y 7 se implementan computacionalmente generando una cantidad u de una distribución uniforme en el intervalo $(0, 1)$, independiente de los parámetros. Si $u \leq \alpha$ el movimiento es aceptado y si $u > \alpha$ el movimiento es rechazado.

3.4. Propuesta Bayesiana para modelos longitudinales con un punto de cambio y estructura de covarianza ARMA(1,1)

Se tiene un modelo longitudinal 3.21 con estructura para la matriz de covarianzas ARMA(1,1) como la presentada en 1.16, $\Sigma_d(\theta_d)$ depende de tres parámetros, $\theta = (\sigma_d^2, \rho_d, \gamma_d)$. Como γ_d es un parámetro de correlación con las mismas características de ρ_d , se suponen distribuciones a priori $Beta(a_{1d}, b_{1d})$ y $Beta(a_{2d}, b_{2d})$ para ρ y γ , respectivamente y se obtienen las siguientes distribuciones condicionales completas a posteriori para los parámetros de la matriz de covarianzas y el punto de cambio τ

$$\pi_{\sigma_d^2} \sim IG\left(\frac{n_{1d}}{2}, \frac{n_{1d}\sigma_{1d}^2}{2}\right),$$

donde $n_{11} = n\tau + n_{01}$, $n_{12} = n(m - \tau) + n_{02}$ y $n_{1d}\sigma_{1d}^2 = n_{0d}\sigma_{0d}^2 + (\mathbf{y}_d - \mathbf{X}_d\boldsymbol{\beta}_d)'(\mathbf{y}_d - \mathbf{X}_d\boldsymbol{\beta}_d)$.

$$\pi_{\rho_d} \propto \exp\left\{-\frac{1}{2}(\mathbf{y}_d - \mathbf{X}_d\boldsymbol{\beta}_d)' \Sigma_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\boldsymbol{\beta}_d) + (a_{1d} - 1) \ln(\rho_d) + (b_{1d} - 1) \ln(1 - \rho_d)\right\}$$

$$\pi_{\gamma_d} \propto \exp\left\{-\frac{1}{2}(\mathbf{y}_d - \mathbf{X}_d\boldsymbol{\beta}_d)' \Sigma_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\boldsymbol{\beta}_d) + (a_{2d} - 1) \ln(\gamma_d) + (b_{2d} - 1) \ln(1 - \gamma_d)\right\} \quad (3.33)$$

Para γ_d se propone el kernel de transición

$$q(\gamma_d^{(*)} | \gamma_d^{(j)}) = \begin{cases} \gamma_d^{(*)} \sim U(0, 2\gamma_d^{(j)}), & \gamma_d^{(j)} \leq 0.5 \\ \gamma_d^{(*)} \sim U(2\gamma_d^{(j)} - 1, 1), & \gamma_d^{(j)} > 0.5 \end{cases} \quad (3.34)$$

$$\pi_{\tau} \propto |\Sigma_1|^{N_1/2} |\Sigma_2|^{N_2/2} \exp\left\{-\frac{1}{2}\left[(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1)' \Sigma_1^{-1}(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1) + (\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2)' \Sigma_2^{-1}(\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2)\right]\right\} \quad (3.35)$$

Siguiendo la metodología utilizada en la sección 3.3, al suponer una estructura para la matriz de covarianzas ARMA(1,1), al algoritmo para la estimación conjunta de la estructura de media, varianza y covarianza en los dos segmentos del modelo es:

1. Iniciar el contador $j = 1$ y dar valores iniciales $\boldsymbol{\beta}_d^{(0)}$, $\sigma_d^{2(0)}$, $\rho_d^{(0)}$, $\gamma_d^{(0)}$ y $\tau^{(0)} \in \mathbb{Z}$.
2. Mediante el muestreador de Gibbs obtener $\boldsymbol{\beta}_1^{(j+1)}$ y $\boldsymbol{\beta}_2^{(j+1)}$, a partir de la distribución normal multivariada 3.25.
3. Mediante el muestreador de Gibbs obtener $\sigma_1^{2(j+1)}$ y $\sigma_2^{2(j+1)}$, a partir de la distribución gamma 3.27.
4. Obtener muestras $q(\rho_d^{(*)} | \rho_d^{(j)})$ desde el kernel de transición 3.30.

5. Se calcula el radio de aceptación

$$\alpha(\rho_d^{(j)}, \rho_d^{(*)}) = \min \left(1, \frac{P(\mathbf{y}_d | \boldsymbol{\beta}_d^{(j+1)}, \sigma_d^{2(j+1)}, \rho_d^{(*)}, \gamma_d^{(j)}, \tau^j) p(\rho_d^{(*)})}{P(\mathbf{y}_d | \boldsymbol{\beta}_d^{(j+1)}, \sigma_d^{2(j+1)}, \rho_d^{(j)}, \gamma_d^{(j)}, \tau^j) p(\rho_d^{(j)})} \right).$$

Si el movimiento es aceptado, entonces $\rho_d^{(j+1)} = \rho_d^{(*)}$. Si el movimiento es rechazado, entonces $\rho_d^{(j+1)} = \rho_d^{(j)}$ y la cadena no se mueve.

6. Obtener muestras $q(\gamma_d^{(*)} | \gamma_d^{(j)})$ desde el kernel de transición 3.34.

7. Se calcula el radio de aceptación

$$\alpha(\gamma_d^{(j)}, \gamma_d^{(*)}) = \min \left(1, \frac{P(\mathbf{y}_d | \boldsymbol{\beta}_d^{(j+1)}, \sigma_d^{2(j+1)}, \rho_d^{(j+1)}, \gamma_d^{(*)}, \tau^j) p(\gamma_d^{(*)})}{P(\mathbf{y}_d | \boldsymbol{\beta}_d^{(j+1)}, \sigma_d^{2(j+1)}, \rho_d^{(j+1)}, \gamma_d^{(j)}, \tau^j) p(\gamma_d^{(j)})} \right).$$

Si el movimiento es aceptado, entonces $\gamma_d^{(j+1)} = \gamma_d^{(*)}$. Si el movimiento es rechazado, entonces $\gamma_d^{(j+1)} = \gamma_d^{(j)}$ y la cadena no se mueve.

8. Proponer un nuevo valor $\tau^{(*)}$ generado de la densidad propuesta 3.32.

9. Aproximar $\tau^{(*)}$ al entero más cercano y calcular la probabilidad de aceptación

$$\alpha(\tau^{(j)}, \tau^{(*)}) = \min \left(1, \frac{\pi(\mathbf{y} | \boldsymbol{\beta}_1^{(j+1)}, \boldsymbol{\beta}_2^{(j+1)}, \sigma_1^{2(j+1)}, \sigma_2^{2(j+1)}, \rho_1^{(j+1)}, \rho_2^{(j+1)}, \gamma_1^{(j+1)}, \gamma_2^{(j+1)}, \tau^{(*)})}{\pi(\mathbf{y} | \boldsymbol{\beta}_1^{(j+1)}, \boldsymbol{\beta}_2^{(j+1)}, \sigma_1^{2(j+1)}, \sigma_2^{2(j+1)}, \rho_1^{(j+1)}, \rho_2^{(j+1)}, \gamma_1^{(j+1)}, \gamma_2^{(j+1)}, \tau^{(j)})} \right).$$

Si el movimiento es aceptado, entonces $\tau^{(j+1)} = \tau^{(*)}$. Si el movimiento es rechazado, entonces $\tau^{(j+1)} = \tau^{(j)}$ y la cadena no se mueve.

8. Retornar al paso 2 hasta obtener convergencia.

Los pasos 5, 7 y 9 se implementan computacionalmente generando una cantidad u de una distribución uniforme en el intervalo $(0, 1)$, independiente de los parámetros. Si $u \leq \alpha$ el movimiento es aceptado y si $u > \alpha$ el movimiento es rechazado.

Simulaciones y aplicaciones

En este capítulo se presentan simulaciones y aplicaciones para algunos conjuntos de datos, empleando las propuestas bayesianas descritas en el capítulo 3. Se ejemplifica la funcionalidad de las propuestas de identificación del punto de cambio a través de simulaciones para los modelos de regresión con varianza homocedástica y heterocedástica, y los modelos longitudinales con matriz de covarianzas SC , $AR(1)$ y $ARMA(1, 1)$, a través de la implementación de distribuciones a priori no informativas.

Para ejemplificar la aplicación del modelo de regresión con cambio estructural y varianza homocedástica se utilizan los datos de expansión térmica del cobre presentados por Cooper & B. (1970). Por otro lado, para presentar la funcionalidad de la propuesta del modelo de regresión con cambio estructural y varianza heterocedástica se emplean los datos de crecimiento de pollos reportados por Riazoshams & Miri (2005).

Finalmente, para los modelos longitudinales con matriz de covarianzas SC , $AR(1)$ y $ARMA(1, 1)$, se usa los datos de tolerancia a la insulina en pacientes obesos reportados inicialmente por Chi & Reinsel (1989) y utilizados posteriormente por Acosta (2010), los cuales se comparan a través de las estadísticas de bondad de ajuste AIC y BIC para determinar cuál de las estructuras de covarianza ajusta mejor los datos vía la metodología Bayesiana propuesta.

4.1. Simulaciones

4.1.1. Modelo lineal con un punto de cambio y varianza homocedástica

El siguiente estudio de simulación fue desarrollado para examinar el comportamiento del procedimiento de estimación comparado con los parámetros originales. Se consideran dos modelos de regresión con un punto de cambio. La variable respuesta en ambos modelos sigue una distribución normal con parámetro de dispersión igual en los dos segmentos. Por lo tanto, los dos modelos solo difieren en la magnitud del cambio.

Se consideran las variables X_1 y X_2 simuladas con $n = 40$ valores, $x_{1i} = 1$ para definir un modelo con intercepto, x_{2i} fue generado de una distribución uniforme en el intervalo $(0, 20)$ y y_i de una distribución normal con media $\mu_{1i} = 40 - 0.4x_{2i}$ y varianza $\sigma_{1i}^2 = 4$ para

n	Parámetro	Valor simulado	Media	Mediana	D. Estándar
40	β_{01}	40	40.501	40.482	0.412
	β_{11}	-0.4	-0.509	-0.509	0.039
	β_{02}	27.5	29.140	29.142	0.509
	β_{12}	0.3	0.238	0.238	0.043
	σ_1^2	4	3.953	3.668	1.395
	σ_2^2	4	4.262	3.962	1.533
	τ	20	20.082	20.220	0.648
40	β_{01}	40	38.797	38.795	0.495
	β_{11}	-0.4	-0.350	-0.350	0.042
	β_{02}	20	20.485	20.481	0.452
	β_{12}	0.3	0.247	0.248	0.038
	σ_1^2	4	3.329	3.111	1.155
	σ_2^2	4	4.874	4.534	1.775
	τ	20	20.498	20.501	0.291

TABLA 4.1. Resumen de resultados a posteriori para la simulación con varianza homocedástica

el primer segmento y con media $\mu_{2i} = 27.5 + 0.3x_{2i}$ y varianza $\sigma_{2i}^2 = 4$ para el segundo segmento, el punto de cambio se encuentra en $\tau = 20$. Se obtuvieron las estimaciones bayesianas de la media, la mediana y la desviación estándar presentadas en la Tabla 4.1. Se asignaron las distribuciones a priori $\beta_d \sim N(0, 10^k I_p)$ donde I_p es una matriz identidad de tamaño $p \times p$ y k es un exponente diseñado para definir varianzas a priori grandes, $\sigma_d^2 \sim G(0.001/2, 0.001 * 10^k/2)$ y para el punto de cambio se asignó la distribución a priori $\tau \sim U(p + 1, n - p - 1)$. Se presentan los resultados basados en $k = 4$, pero también se utilizaron valores más grandes que condujeron a cambios despreciables en los resultados a posteriori.

El principal interés de éste estudio de simulación es verificar el comportamiento de las cadenas que determinan la estimación del punto de cambio τ , el cual se puede observar en la Figura 4.1. En la Figura 4.2 se presentan los histogramas de las distribuciones marginales a posteriori de los parámetros, los gráficos anteriores están relacionados al primer modelo. Los resultados de la simulación del primer modelo muestran que el punto de cambio es detectado con una gran precisión (Figura 4.1).

El segundo modelo considerado difiere del primer modelo solo en la intersección del segundo segmento, es decir, y_i fue generado de una distribución normal con media $\mu_{1i} = 40 - 0.4x_{2i}$ y varianza $\sigma_{1i}^2 = 4$ para el primer segmento y con media $\mu_{2i} = 20 + 0.3x_{2i}$ y varianza $\sigma_{2i}^2 = 4$ para el segundo segmento, el punto de cambio se encuentra en $\tau = 20$. Se obtuvieron las estimaciones bayesianas de la media, la mediana y la desviación estándar presentadas al final de la Tabla 4.1. En la simulación bajo el segundo modelo se obtienen resultados similares, nuevamente el punto de cambio es detectado con precisión. En este caso, se puede observar que la estimación se encuentra más alejada del valor verdadero, pero la variabilidad del punto de cambio estimado es menor.

Se encontró que la detección del punto de cambio depende en gran medida de la desviación entre los modelos, sin embargo, el rezago entre la estimación y el verdadero valor no aumenta significativamente. Por lo tanto, en la mayoría de los casos el método propuesto detecta el punto de cambio con una gran precisión.

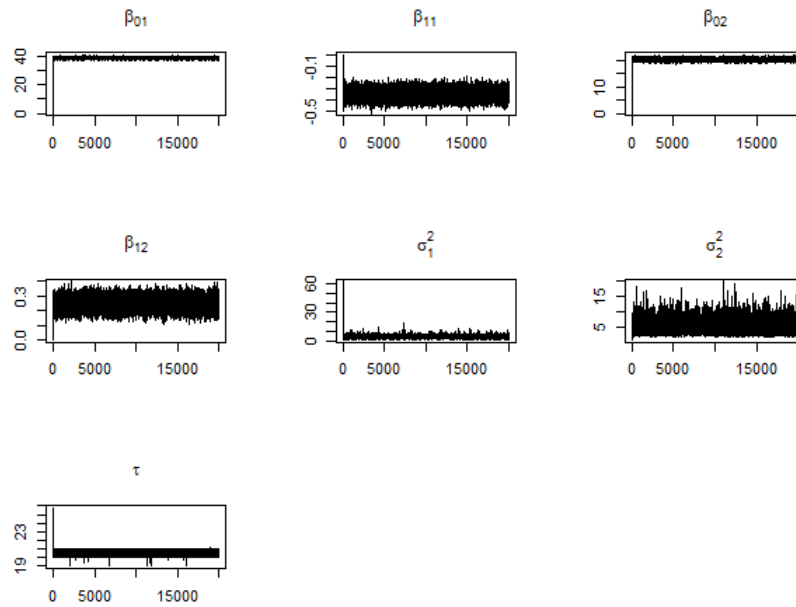


FIGURA 4.1. Comportamiento de las cadenas para cada uno de los parámetros del estudio de simulación de un modelo de regresión con varianza homocedástica

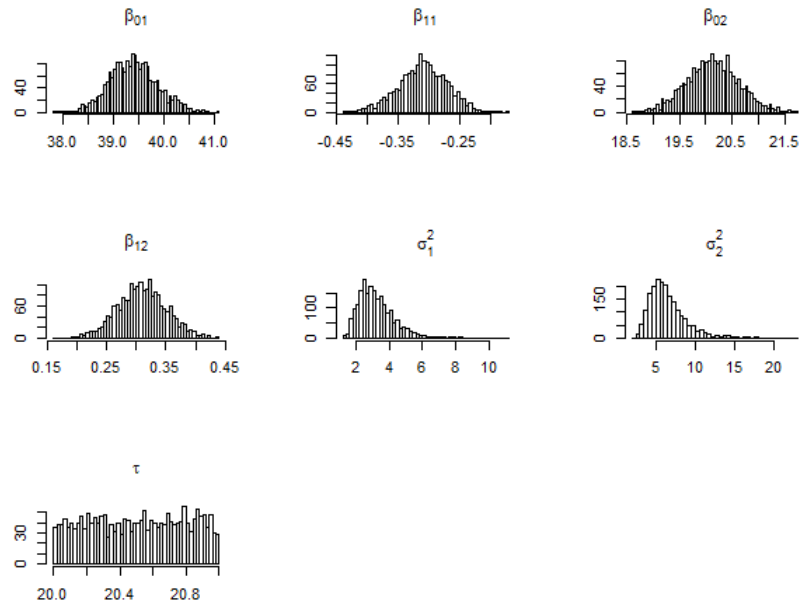


FIGURA 4.2. Histograma de las distribuciones marginales a posteriori en el estudio de simulación de un modelo de regresión con varianza homocedástica

n	Parámetro	Valor simulado	Media	Mediana	D. Estándar
40	β_{01}	2.5	2.519	2.520	0.050
	β_{11}	0.7	0.699	0.698	0.009
	β_{02}	5.0	4.685	4.674	0.484
	β_{12}	0.5	0.671	0.651	0.028
	γ_{01}	-6.5	-5.848	-5.922	0.640
	γ_{11}	0.25	0.223	0.225	0.023
	γ_{02}	-1.5	-0.929	-0.892	0.788
	γ_{12}	0.08	0.063	0.058	0.039
	τ	20	20.422	20.489	0.6309
400	β_{01}	2.5	2.488	2.488	0.011
	β_{11}	0.7	0.700	0.700	0.002
	β_{02}	5.0	4.996	4.997	0.094
	β_{12}	0.5	0.506	0.506	0.007
	γ_{01}	-6.5	-6.426	-6.428	0.120
	γ_{11}	0.25	0.242	0.242	0.005
	γ_{02}	-1.5	-1.475	-1.477	0.115
	γ_{12}	0.08	0.084	0.084	0.005
	τ	200	200.297	200.387	0.492

TABLA 4.2. Resumen de resultados a posteriori para la simulación con varianza heterocedástica

4.1.2. Modelo lineal con un punto de cambio y varianza heterocedástica

Un estudio de simulación fue conducido para examinar el comportamiento del procedimiento de estimación comparado con los parámetros originales. Las variables X_1 y X_2 fueron simuladas con $n = 40$ valores, $x_{1i} = 1$ para definir un modelo con intercepto, x_{2i} fue generado de una distribución uniforme en el intervalo $(0, 40)$ y y_i de una distribución normal con media $\mu_{1i} = 2.5 + 0.7x_{2i}$ y varianza $\sigma_{1i} = \exp(-6.5 + 0.25x_{2i})$ para las primeras 20 observaciones y con media $\mu_{2i} = 5 + 0.5x_{2i}$ y varianza $\sigma_{2i} = \exp(-1.5 + 0.08x_{2i})$ para las siguientes 20 observaciones, es decir, el punto de cambio se encuentran en $\tau = 20$. Se obtuvieron las estimaciones bayesianas de la media, la mediana y la desviación estándar presentadas en la Tabla 4.2. Se asignaron las distribuciones a priori $(\beta, \gamma) \sim N(0, 10^k I_p)$ donde I_p es una matriz identidad de tamaño $p \times p$ y k es una exponente diseñado para definir varianzas a priori grandes. Se presentan los resultados basados en $k = 4$, pero también se utilizaron valores más grandes que condujeron a cambios despreciables en los resultados a posteriori. Para el punto de cambio se asignó la distribución a priori $\tau \sim U(p+1, n-p-1)$.

A continuación, se generó un conjunto extra de 360 datos, se definió $\tau = 200$ y se estimaron los parámetros del modelo en el mismo camino con $n = 400$. Los valores de los parámetros y sus estimaciones (incluida la desviación estándar) se presentan en la Tabla 4.2. Se pueden observar mejores estimaciones de los parámetros y desviaciones estándar más pequeñas en el segundo estudio, como consecuencia de mayor información disponible en la verosimilitud.

La Figura 4.3 muestra el comportamiento de la cadena obtenida para los parámetros y en la Figura 4.4 se observan los histogramas de las distribuciones marginales a posteriori de los parámetros, los gráficos anteriores están relacionados a la simulación con $n = 400$. Para realizar la simulación, se generaron varias cadenas, con valores iniciales diferentes. En todas se encontró el mismo comportamiento cualitativo después de un período transitorio

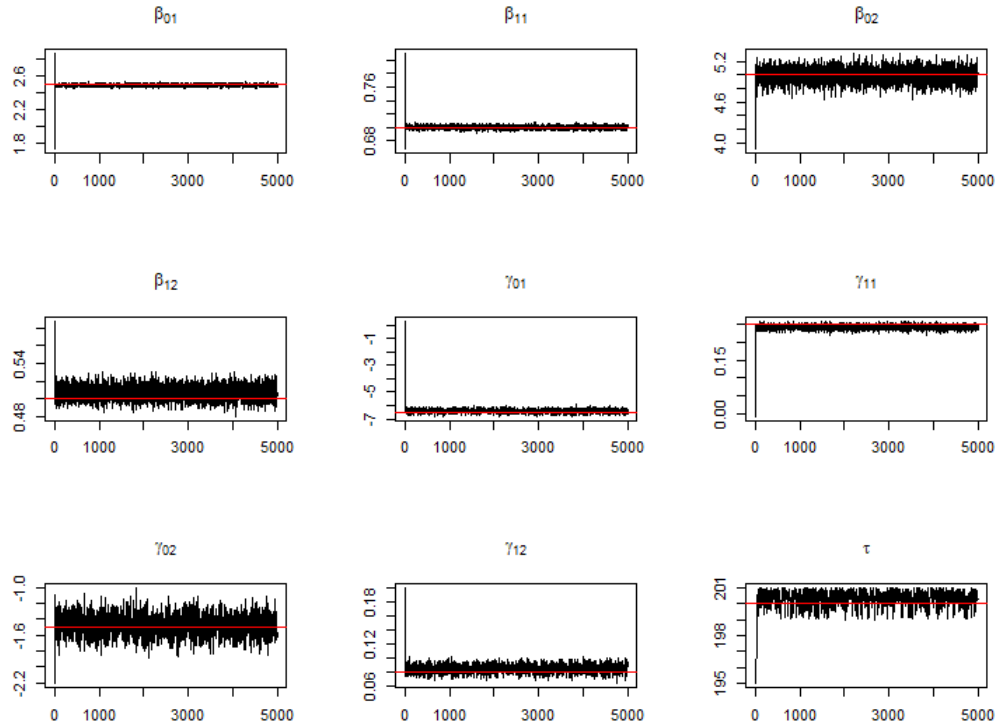


FIGURA 4.3. Comportamiento de las cadenas para cada uno de los parámetros del estudio de simulación de un modelo de regresión con varianza heterocedástica, $n = 400$

inicial, lo cual indica una aproximación a estacionariedad. Por otro lado, se encontró que la elección de un valor inicial para el punto de cambio lejano al parámetro no afecta el procedimiento de estimación, pues la convergencia al valor verdadero es rápida. Estos comportamientos se ilustran para una de las cadenas en la Figura 4.3. Los histogramas presentados en la Figura 4.4 parecen mostrar que las distribuciones marginales a posteriori para los parámetros de media y varianza son aproximadamente normal. Además, resulta importante mencionar que en esta simulación, la tasa de aceptación para la estimación del punto de cambio τ está al rededor de 35 %.

4.1.3. Modelo longitudinal con estructura de covarianza de Simetría Compuesta

Se realiza un estudio de simulación para observar el comportamiento del procedimiento de estimación descrito en la subsección 3.3 para un modelo longitudinal con un punto de cambio y estructura de covarianza SC .

$$y_{ij} = \begin{cases} \beta_{01} + \beta_{11}t_{ij1} + \epsilon_{ij1} & j = 1, \dots, \tau \\ \beta_{02} + \beta_{12}t_{ij2} + \epsilon_{ij2} & j = \tau + 1, \dots, m, \end{cases}$$

Donde $i = 1, 2, \dots, 20$ indica los individuos y $j = 1, 2, \dots, 10$ representa el número de observaciones para el i -ésimo individuo, además se considera un modelo balanceado.

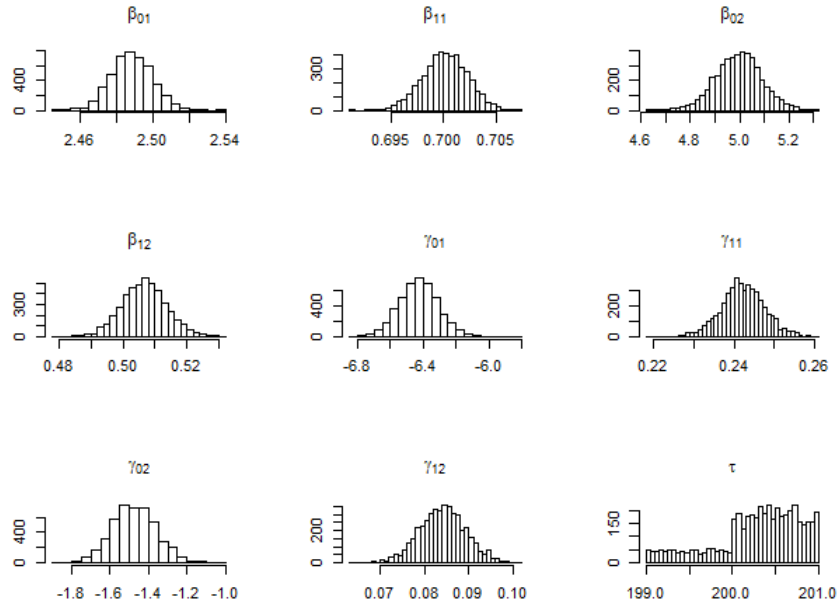


FIGURA 4.4. Histograma de las distribuciones marginales a posteriori en el estudio de simulación de un modelo de regresión con varianza heterocedástica, $n = 400$

Se supone que $\epsilon_{ijd} \sim N(0, \Sigma_d(\sigma_d^2, \rho_d))$, donde $\Sigma_d(\sigma_d^2, \rho_d)$ sigue una estructura como la presentada en la matriz 1.14 y $d = 1, 2$, por lo tanto, $\mathbf{y}_d \sim NMV(\mathbf{X}_d \beta_d, \Sigma_d)$.

Para obtener el i -ésimo componente de $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{20})'$, donde $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{i10})$ es generados de una distribución normal con media $\mu_{ij1} = 200 - 25t_{ij1}$ y matriz de covarianzas con $\sigma_1^2 = 16$ y $\rho_1 = 0.6$ para las primeras 5 observaciones y $\mu_{ij2} = 7 + 12t_{ij2}$ y matriz de covarianza con $\sigma_2^2 = 9$ y $\rho_1 = 0.8$ para las siguientes 5 observaciones, por lo tanto, el punto de cambio está dado por $\tau = 5$. Se realizaron 20.000 iteraciones y se descartaron las primeras 2.000, por lo tanto, se obtuvieron las estimaciones bayesianas de la media, la mediana y la desviación estándar presentadas en la Tabla 4.3. Se asignaron las distribuciones a priori $\beta_d \sim N(0, 10^k \mathbf{I}_p)$ donde \mathbf{I}_p es una matriz identidad de tamaño $p \times p$. Para el parámetro de varianza $\sigma_d^2 \sim IG(10^{-k}, 10^{-k} * 100)$ y k es un exponente para definir distribuciones a priori no informativas. Se presentan los resultados basados en $k = 4$, pero también se utilizaron valores más grandes que condujeron a cambios despreciables en los resultados. Finalmente, para el parámetro de correlación se supuso la distribución a priori $\rho_d \sim Beta(1, 1)$ con $d = 1, 2$ y para el punto de cambio $\tau \sim U(2, 9)$.

La Figura 4.5 muestra el comportamiento de la cadena para los parámetros y en la Figura 4.6 se presentan los histogramas de las distribuciones marginales a posteriori de los parámetros. Para realizar la simulación se generaron varias cadenas, con valores iniciales diferentes, en todos se encontró el mismo comportamiento cualitativo después de un periodo transitorio inicial, lo cual sugiere una aproximación a estacionariedad. Además, la Figura 4.5 muestra que la elección del punto de cambio lejano al parámetro no afecta el procedimiento de estimación, pues converge rápidamente al verdadero valor. La Figura 4.6 evidencian que las distribuciones marginales a posteriori para los parámetros de media son aproximadamente normal y los parámetros de varianza siguen una distribución Gamma.

n	m	Parámetro	Valor simulado	Media	Mediana	D. Estándar
20	8	β_{01}	200	200.277	200.284	0.915
		β_{11}	-25	-25.020	-25.019	0.186
		β_{02}	7	6.310	6.304	1.064
		β_{12}	12	12.100	12.102	0.109
		σ_1^2	16	16.574	15.995	3.555
		σ_2^2	9	9.367	8.934	2.496
		ρ_1	0.6	0.551	0.554	0.090
		ρ_2	0.8	0.729	0.731	0.069
		τ	5	5.503	5.506	0.289

TABLA 4.3. Resumen de resultados a posteriori para la simulación de un modelo longitudinal con estructura de covarianza SC

Finalmente, la tasa de aceptación para ρ_1 fue de 23.68%, ρ_2 de 23.96% y para τ fue de 50.11%.

4.1.4. Modelo longitudinal con estructura de covarianza Autoregresiva

Se realiza un estudio de simulación para observar el comportamiento del procedimiento de estimación descrito en la subsección 3.3 para un modelo longitudinal con un punto de cambio y estructura de covarianza $AR(1)$.

$$y_{ij} = \begin{cases} \beta_{01} + \beta_{11}t_{ij1} + \epsilon_{ij1} & j = 1, \dots, \tau \\ \beta_{02} + \beta_{12}t_{ij2} + \epsilon_{ij2} & j = \tau + 1, \dots, m, \end{cases}$$

Donde $i = 1, 2, \dots, 20$ indica los individuos y $j = 1, 2, \dots, 8$ representa el número de observaciones para el i -ésimo individuo, además se considera un modelo balanceado. Se supone que $\epsilon_{ijd} \sim N(0, \Sigma_d(\sigma_d^2, \rho_d))$, donde $\Sigma_d(\sigma_d^2, \rho_d)$ sigue una estructura como la presentada en la matriz 1.15 y $d = 1, 2$, por lo tanto, $\mathbf{y}_d \sim NMV(\mathbf{X}_d\beta_d, \Sigma_d)$.

Para obtener el i -ésimo componente de $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{20})'$, donde $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{i8})$ es generados de una distribución normal con media $\mu_{ij1} = 30 + 20t_{ij1}$ y matriz de covarianzas con $\sigma_1^2 = 25$ y $\rho_1 = 0.5$ para las primeras 4 observaciones y $\mu_{ij2} = 150 - 9t_{ij2}$ y matriz de covarianza con $\sigma_2^2 = 4$ y $\rho_1 = 0.7$ para las siguientes 4 observaciones, por lo tanto, el punto de cambio está dado por $\tau = 4$. Se realizaron 20.000 iteraciones y se descartaron las primeras 2.000, por lo tanto, se obtuvieron las estimaciones bayesianas de la media, la mediana y la desviación estándar presentadas en la Tabla 4.4. Se asignaron las distribuciones a priori $\beta_d \sim N(0, 10^k \mathbf{I}_p)$ donde \mathbf{I}_p es una matriz identidad de tamaño $p \times p$. Para el parámetro de varianza $\sigma_d^2 \sim IG(10^{-k}, 10^{-k} * 100)$ y k es un exponente para definir distribuciones a priori no informativas. Se presentan los resultados basados en $k = 4$, pero también se utilizaron valores más grandes que condujeron a cambios despreciables en los resultados. Finalmente, para el parámetro de correlación se supuso la distribución a priori $\phi_d \sim Beta(1, 1)$ y para el punto de cambio $\tau \sim U(2, 7)$.

La Figura 4.7 muestra el comportamiento de la cadena para los parámetros y en la Figura 4.8 se presentan los histogramas de las distribuciones marginales a posteriori de los parámetros. Para realizar la simulación se generaron varias cadenas, con valores iniciales diferentes, en todos se encontró el mismo comportamiento cualitativo después de

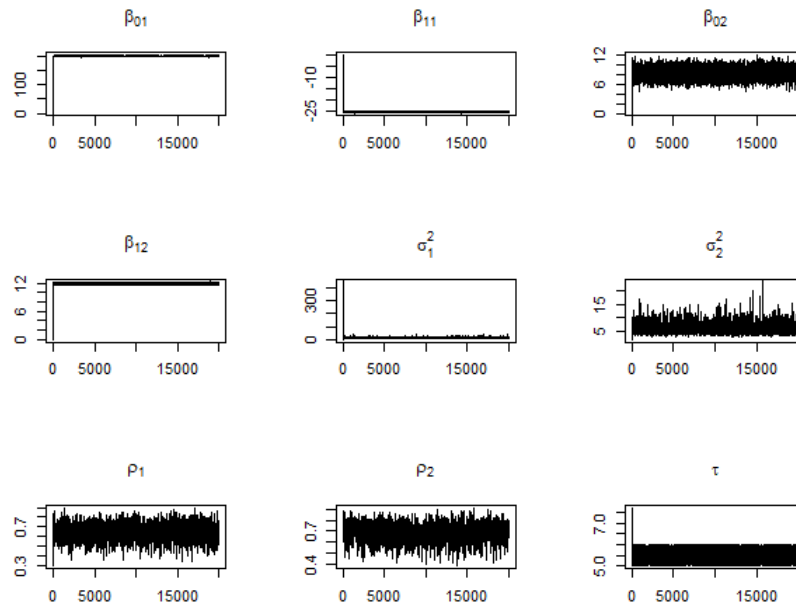


FIGURA 4.5. Comportamiento de las cadenas para cada uno de los parámetros del estudio de simulación de un modelo longitudinal con estructura de covarianza SC

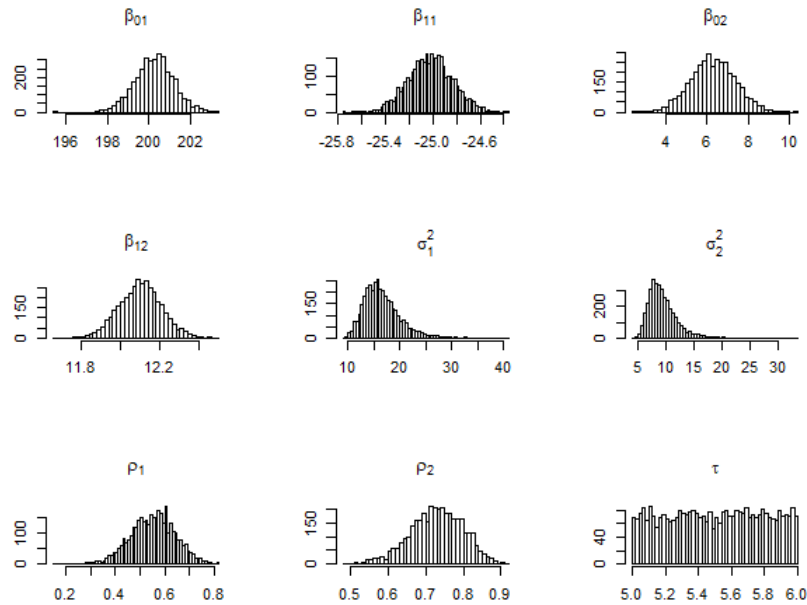


FIGURA 4.6. Histograma de las distribuciones marginales a posteriori en el estudio de simulación de un modelo longitudinal SC

n	m	Parámetro	Valor simulado	Media	Mediana	D. Estándar
20	8	β_{01}	30	31.307	31.317	1.467
		β_{11}	20	19.866	19.865	0.488
		β_{02}	150	148.518	148.526	1.230
		β_{12}	-9	-8.843	-8.843	0.179
		σ_1^2	25	25.766	24.928	5.373
		σ_2^2	4	4.373	4.204	1.019
		ρ_1	0.5	0.527	0.531	0.099
		ρ_2	0.7	0.680	0.684	0.074
		τ	4	4.496	4.494	0.287

TABLA 4.4. Resumen de resultados a posteriori para la simulación de un modelo longitudinal con estructura de covarianza AR(1)

un periodo transitorio inicial, lo cual sugiere una aproximación a estacionariedad. Además, la Figura 4.7 muestra que la elección del punto de cambio lejano al parámetro no afecta el procedimiento de estimación, pues converge rápidamente al verdadero valor. La Figura 4.8 evidencian que las distribuciones marginales a posteriori para los parámetros de media son aproximadamente normal y los parámetros de varianza siguen una distribución Gamma. Finalmente, la tasa de aceptación para ρ_1 fue de 31 %, ρ_2 de 28 % y de τ de 49 %.

4.1.5. Modelo longitudinal con estructura de covarianza Autoregresiva de Promedio Móvil

Se realiza un estudio de simulación para observar el comportamiento del procedimiento de estimación descrito en la subsección 3.4 para un modelo longitudinal con un punto de cambio y estructura de covarianza $ARMA(1, 1)$.

$$y_{ij} = \begin{cases} \beta_{01} + \beta_{11}x_{2ij} + \beta_{21}x_{3ij} + \epsilon_{ij1} & j = 1, \dots, \tau \\ \beta_{02} + \beta_{12}x_{2ij} + \beta_{22}x_{3ij} + \epsilon_{ij2} & j = \tau + 1, \dots, m. \end{cases}$$

Donde $i = 1, 2, \dots, 20$ indica los individuos y $j = 1, 2, \dots, 10$ representa el número de observaciones para el i -ésimo individuo, además se considera un modelo balanceado. Se supone que $\epsilon_{ijd} \sim N(0, \Sigma_d(\sigma_d^2, \rho_d, \gamma_d))$, donde $\Sigma_d(\sigma_d^2, \rho_d, \gamma_d)$ sigue una estructura $ARMA(1, 1)$ como la presentada en la matriz 1.16 y $d = 1, 2$, por lo tanto, $\mathbf{y}_d \sim NMV(\mathbf{X}_d\beta_d, \Sigma_d)$.

Se consideran las variables \mathbf{X}_1 , \mathbf{X}_2 y \mathbf{X}_3 simuladas con $n = 20$ individuos observados en $m = 6$ mediciones, donde $x_{1ij} = 1$ para definir un modelo con intercepto, $x_{2ij} = j/8$, x_{3ij} fue generada de una distribución uniforme en el intervalo $(0, 1)$ y y_{ij} de una distribución normal multivariada con media $\mu_{ij1} = 6.8 - 2.5x_{1ij} + 0.05x_{2ij}$ y matriz de varianza y covarianza dada por $\sigma_1^2 = 0.53$, $\rho_1 = 0.9$ y $\gamma_1 = 0.6$ para las primeras 5 observaciones en el i -ésimo individuo y con media $\mu_{ij2} = 2.5 + 4.5x_{1ij} - 0.05x_{2ij}$ y matriz de varianza y covarianza dada por $\sigma_2^2 = 0.74$, $\rho_2 = 0.8$ y $\gamma_2 = 0.9$ para las siguientes 3 observaciones en el i -ésimo individuo, por lo tanto, el punto de cambio es $\tau = 5$. Se realizaron 20.000 iteraciones y se descartaron las primeras 2.000, las estimaciones bayesianas de la media, la mediana y la desviación estándar se presentan en la Tabla 4.5. Se asignaron las distribuciones a priori $\beta_d \sim N(0, 10^k \mathbf{I}_p)$ donde \mathbf{I}_p es una matriz identidad de tamaño $p \times p$. Para el parámetro de varianza $\sigma_d^2 \sim IG(10^{-k}, 10^{-k} * 100)$ y k es un exponente para definir distribuciones a priori

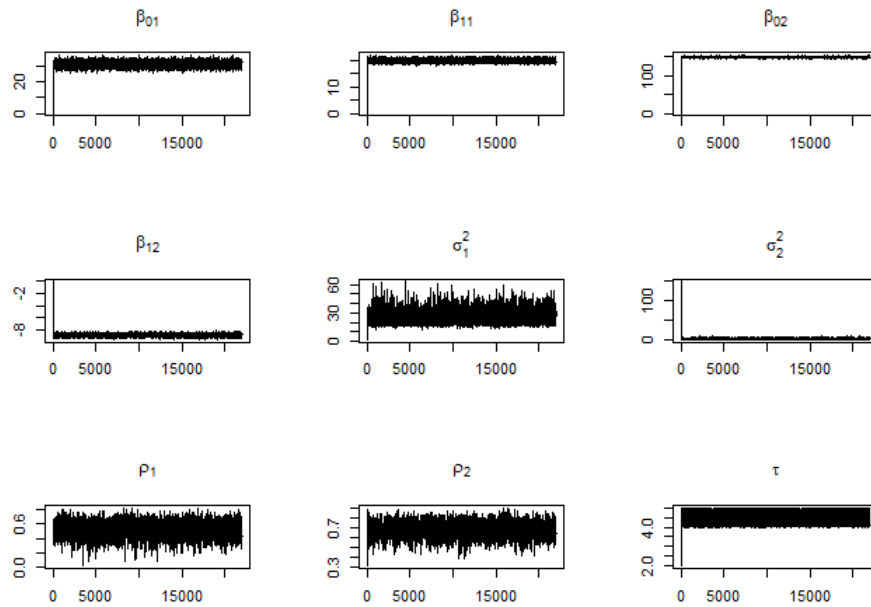


FIGURA 4.7. Comportamiento de las cadenas para cada uno de los parámetros del estudio de simulación de un modelo longitudinal con estructura de covarianza AR(1)

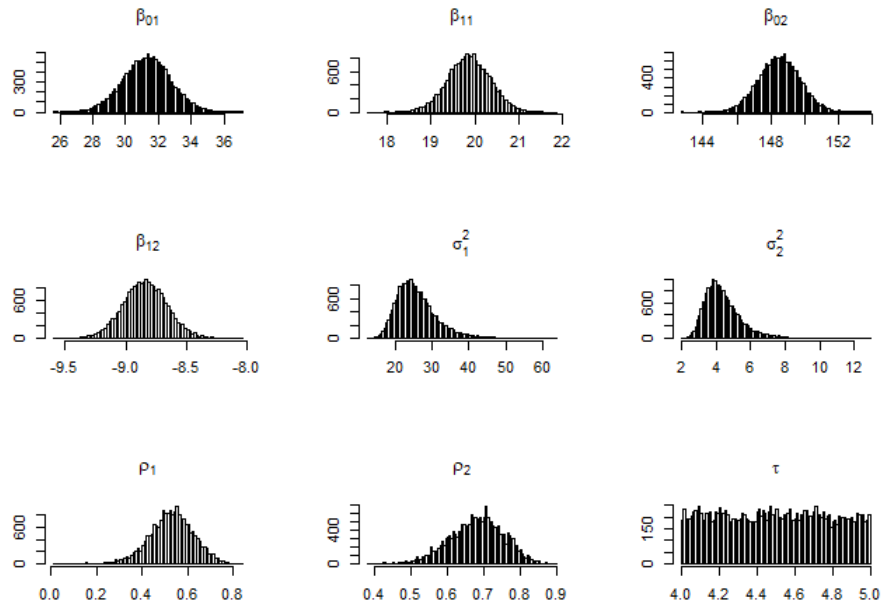


FIGURA 4.8. Histograma de las distribuciones marginales a posteriori en el estudio de simulación de un modelo longitudinal AR(1)

n	m	Parámetro	Valor simulado	Media	Mediana	D. Estándar
20	8	β_{01}	6.8	6.372	6.370	0.209
		β_{11}	-2.5	-2.258	-2.256	0.389
		β_{21}	0.05	0.278	0.278	0.199
		β_{02}	2.5	3.137	3.144	0.709
		β_{12}	4.5	3.881	3.878	0.792
		β_{22}	-0.05	-0.278	-0.276	0.277
		σ_1^2	0.53	0.484	0.474	0.083
		σ_2^2	0.74	0.596	0.582	0.119
		ρ_1	0.9	0.681	0.684	0.107
		ρ_2	0.8	0.721	0.727	0.111
		γ_1	0.6	0.485	0.489	0.084
		γ_2	0.9	0.817	0.823	0.053
		τ	5	5.444	5.417	0.283

TABLA 4.5. Resumen de resultados a posteriori para la simulación de un modelo longitudinal con estructura de covarianza ARMA(1,1)

no informativas. Se presentan los resultados basados en $k = 4$, pero también se utilizaron valores más grandes que condujeron a cambios despreciables en los resultados. Finalmente, para los parámetros de correlación se supuso la distribución a priori $\rho_d \sim Beta(7, 5)$, $\gamma_d \sim Beta(1, 1)$ y para el punto de cambio $\tau \sim U(2, 7)$.

Las Figuras 4.9 y 4.10 muestra el comportamiento de la cadena para los parámetros de media y covarianza, respectivamente. En las Figuras 4.11 y 4.12 se presentan los histogramas de las distribuciones marginales a posteriori de los parámetros. Para realizar la simulación se generaron varias cadenas, con valores iniciales diferentes, en todos se encontró el mismo comportamiento cualitativo después de un periodo transitorio inicial, lo cual sugiere una aproximación a estacionariedad. Además, la Figura 4.10 muestra que la elección del punto de cambio lejano al parámetro no afecta el procedimiento de estimación, pues converge rápidamente al verdadero valor. La Figura 4.11 evidencian que las distribuciones marginales a posteriori para los parámetros de media son aproximadamente normal y en la Figura 4.12 se observa que los parámetros de varianza siguen una distribución Gamma. Finalmente, la tasa de aceptación para los parámetros de correlación ρ_1 fue de 46 %, ρ_2 de 42 %, γ_1 de 31 %, γ_2 de 29 % y de τ de 64 %.

4.2. Aplicaciones

4.2.1. Aplicación: Datos de expansión térmica del cobre

Para mostrar una aplicación de los resultados teóricos presentados en la sección 3.1, se usaron los datos del resultado de un estudio NIST que involucra la expansión térmica del cobre presentados por Cooper & B. (1970). La variable respuesta es el coeficiente de expansión térmica y la variable predictiva es la temperatura en grados Kelvin. Se obtuvieron un total de $n = 236$ observaciones.

De la figura 4.13 se puede observar que las variables involucradas en el estudio no están asociadas linealmente, ya que al comenzar el estudio y hasta aproximadamente los 200 grados Kelvin hay un incremento acelerado en la expansión térmica del cobre, pero

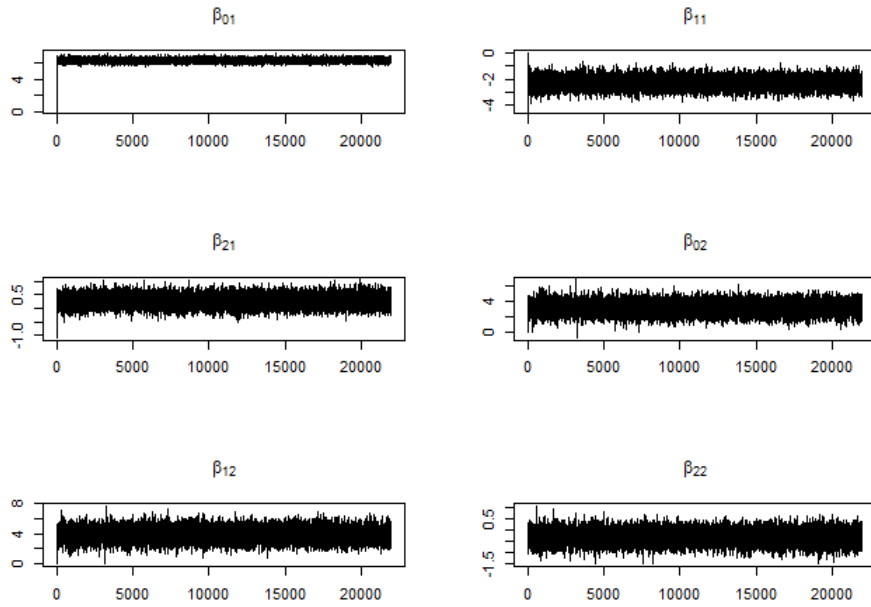


FIGURA 4.9. Comportamiento de las cadenas para los parámetros de media del estudio de simulación de un modelo longitudinal con estructura ARMA(1,1)

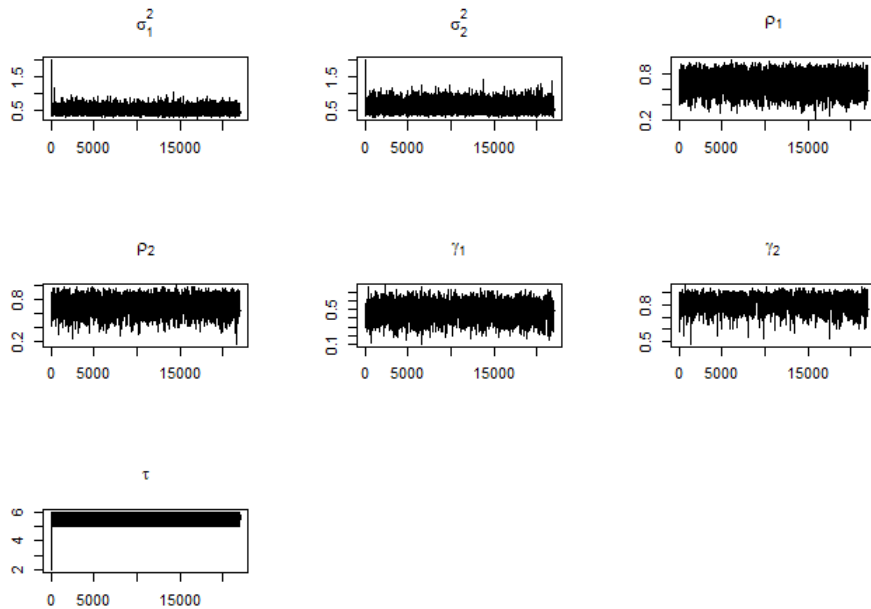


FIGURA 4.10. Comportamiento de las cadenas para los parámetros de covarianza y el punto de cambio en un estudio de simulación de un modelo longitudinal con estructura ARMA(1,1)

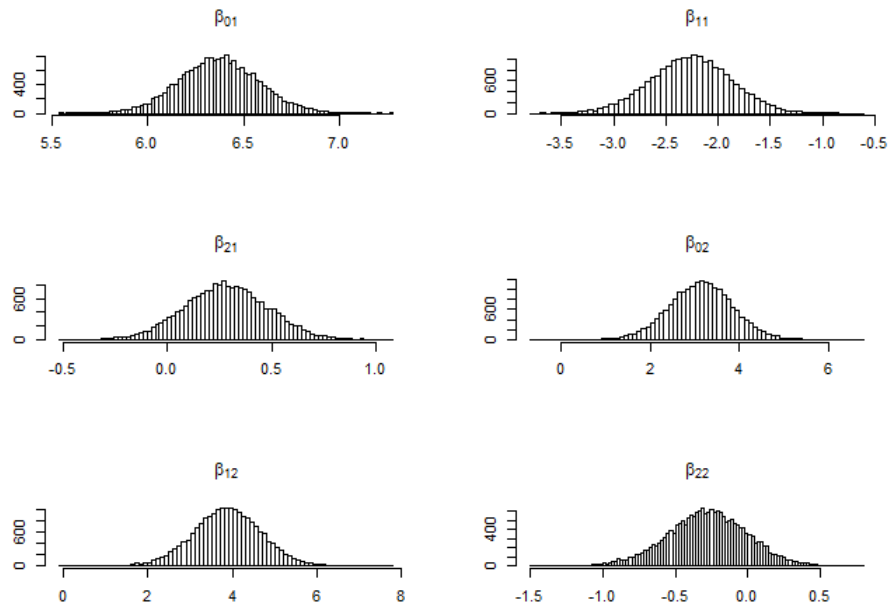


FIGURA 4.11. Histograma de las distribuciones marginales a posteriori de los parámetros de media en el estudio de simulación de un modelo longitudinal ARMA(1,1)

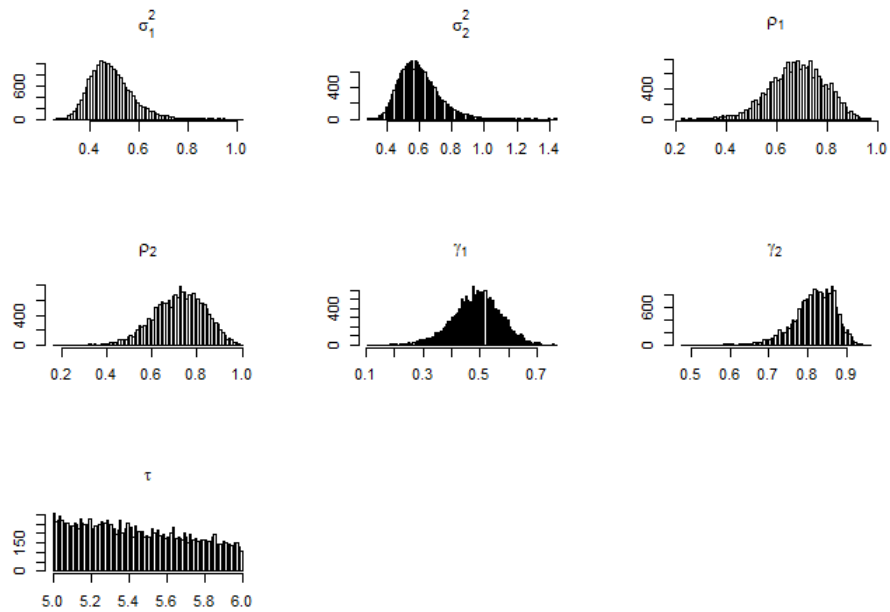


FIGURA 4.12. Histograma de las distribuciones marginales a posteriori de los parámetros de covarianza en el estudio de simulación de un modelo longitudinal ARMA(1,1)

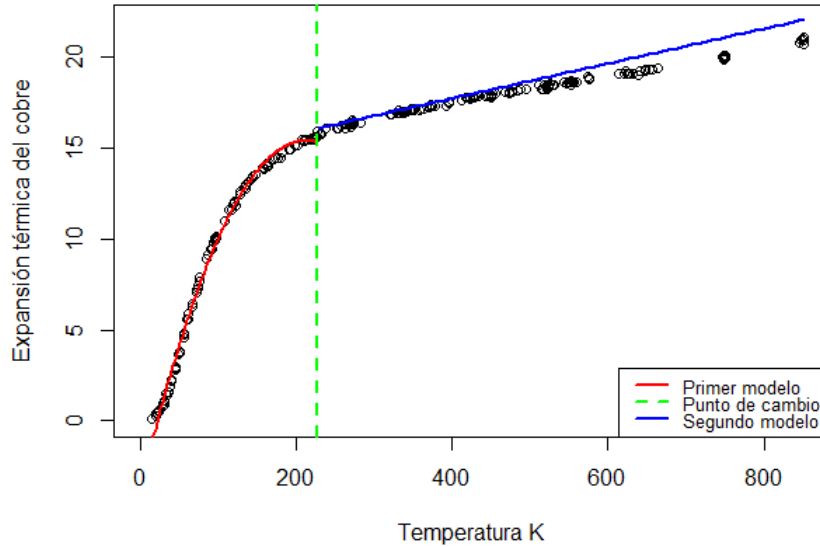


FIGURA 4.13. Coeficiente de expansión térmica del cobre en función de la temperatura en grados Kelvin para un estudio NIST

a partir de este punto y hasta el final de las observaciones, empieza a estabilizarse dicho comportamiento.

Para el presente caso, como la relación entre las variables no es lineal, se utiliza el modelo de regresión 4.1 con cambio estructural y varianza homocedastica, además, la estimación de los parámetros se realiza a través de la metodología propuesta en la sección 3.1

$$y_i = \begin{cases} \beta_{01} + \beta_{11}x_i + \beta_{21}x_i^2 + \epsilon_{i1} & i = 1, \dots, \tau \\ \beta_{02} + \beta_{12}x_i + \epsilon_{i2} & i = \tau + 1, \dots, n. \end{cases} \quad (4.1)$$

Para encontrar la estimación de los parámetros se asignaron las distribuciones a priori no informativas $\beta_d \sim N(0, 10^4 \mathbf{I}_3)$, $\sigma_d^2 \sim IG(10^{-4}, 10^{-4} \times 100)$, y $\tau \sim U(p + 1, n - p - 1)$. Se realizaron 20.000 iteraciones y se descartaron las primeras 10.000, de este modo se obtuvieron las estimaciones bayesianas de la media, mediana y desviación estándar a posteriori presentadas en la tabla 4.6. Se observa que el punto de cambio está ubicado en $\tau = 100$, lo que corresponde a temperatura de 227 grados Kelvin. En la figura 4.14 se puede observar que el punto de cambio es detectado con una gran precisión. En la figura 4.15 se presentan los histogramas de las distribuciones marginales a posteriori de los parámetros.

La tasa de aceptación para el parámetro τ es de 79.7%, los valores obtenidos para los criterios de información son $AIC = -76.586$ y $BIC = -59.267$. Para mostrar la utilidad de la inclusión del punto de cambio en la propuesta Bayesiana 3.1, se realiza la comparación contra un modelo de regresión lineal polinómico de la forma $\mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$. La estimación se hace con la metodología Bayesiana presentada en el capítulo 1.1.2. Se obtienen los parámetros estimados $\hat{\beta}_0 = 3.176874$, $\hat{\beta}_1 = 0.057091$, $\hat{\beta}_2 = -0.000047$ y los estadísticos $AIC = 1005.217$ y $BIC = 1015.609$.

n	Parámetro	Media	Mediana	D. Estándar
236	β_{01}	-3.8048	-3.8059	3.4413×10^{-1}
	β_{11}	0.1815	0.1813	7.1332×10^{-3}
	β_{21}	-0.0004	-0.0004	2.9803×10^{-5}
	β_{02}	1.3927	1.3923	7.4251×10^{-1}
	β_{12}	9.5881×10^{-3}	9.5902×10^{-3}	3.1322×10^{-3}
	σ_1^2	0.1591	0.1534	0.0349
	σ_2^2	0.0355	0.0310	0.0194
	τ	100.7930	101.2519	2.3392

TABLA 4.6. Resumen de resultados a posteriori para los datos de expansión térmica del cobre

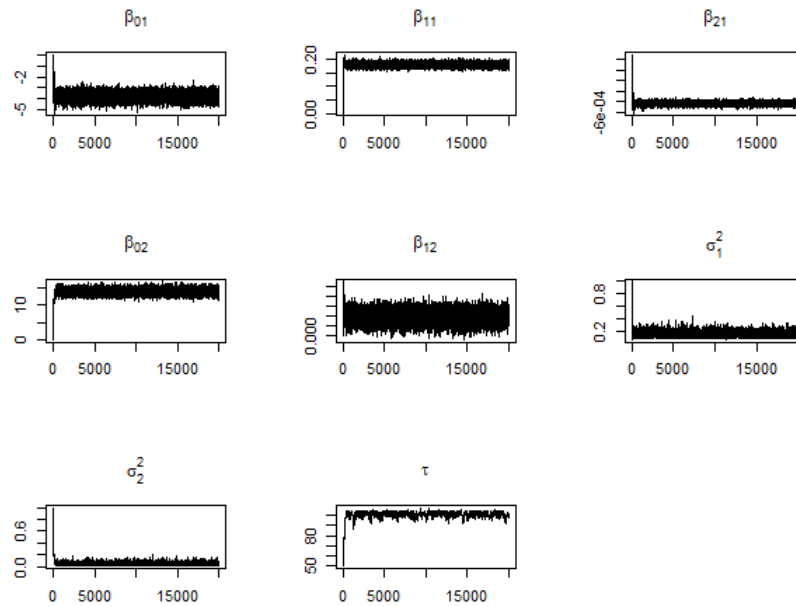


FIGURA 4.14. Comportamiento de las cadenas para cada uno de los parámetros del modelo de expansión del cobre

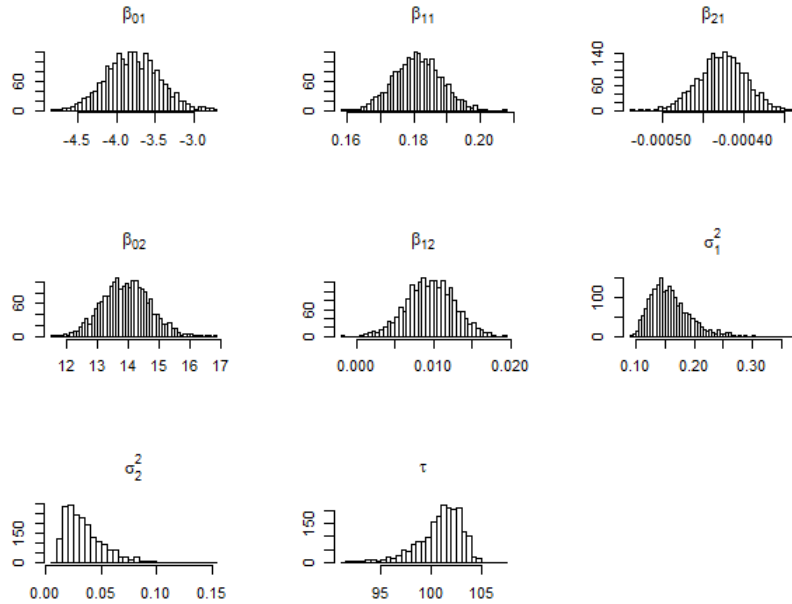


FIGURA 4.15. Histograma de las distribuciones marginales a posteriori para el modelo de expansión del cobre

4.2.2. Aplicación: Datos de crecimiento de pollos

Los datos fueron reportados inicialmente por Riazoshams & Miri (2005), se recolectaron en un periodo de 51 días de suministro de pollos de engorde en Irán. Los datos fueron tomados de forma aleatoria de una población en la cámara de engorde de 7300 pollos. Durante los primeros 25 días se seleccionaron al azar 10 pollos y se registró su peso. Desde el día 27 hasta el día 51 (último día), se seleccionaron 20 pollos cada dos días y se registró su peso. Para cada día de medición, los pollos se recolectaron aleatoriamente de una parte diferente de la cámara de engorde. En este caso es importante detectar el punto de cambio de la curva, que corresponde a la tasa más alta de crecimiento de los pollos entre los días 10 y 20.

Riazoshams et al. (2018) presentan los detalles del método de recolección y los errores de muestreo para los datos. Como se muestra en la Figura 4.16, la variación de los datos aumenta con el tiempo. Este es un fenómeno que ocurre debido al aumento de peso durante la vida útil de los pollos, lo que hace que aumente la variación de los datos. La heterocedasticidad debido a este comportamiento es obvia y debe tenerse en cuenta en el modelado de los datos.

Riazoshams et al. (2018) explican que la tasa de crecimiento observada en los datos tiene forma de S: en los animales jóvenes, el peso aumenta lentamente, luego rápidamente en la juventud antes de disminuir la tasa de crecimiento en su etapa adulta. Finalmente, el peso tiende a comportarse de manera constante. En la práctica se pueden ajustar muchos modelos a los datos. Riazoshams & Miri (2005) eligieron el modelo no lineal logístico. En el presente trabajo se utiliza el modelo de regresión 4.2 con un punto de cambio y varianza heterocedastica, además la estimación de los parámetros se realiza a través de la metodología propuesta en la sección 3.2

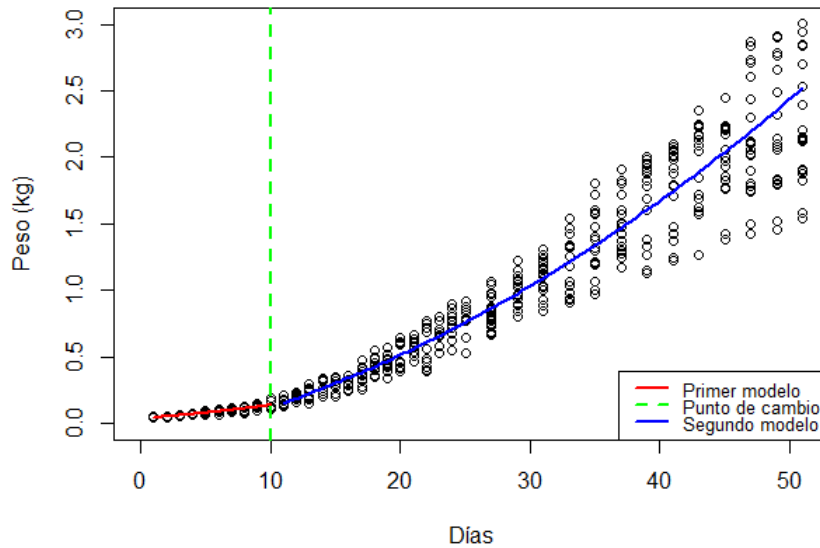


FIGURA 4.16. Gráfico de datos de crecimiento de pollos

$$y_i = \begin{cases} \beta_{01} + \beta_{11}x_i + \beta_{21}x_i^2 + \epsilon_{i1} & i = 1, \dots, \tau \\ \beta_{02} + \beta_{12}x_i + \beta_{22}x_i^2 + \epsilon_{i2} & i = \tau + 1, \dots, n. \end{cases} \quad (4.2)$$

con $\sigma_{i1}^2 = \exp \gamma_{01} + \gamma_{21}z_i^2$, $i = 1, \dots, \tau$ y $\sigma_{i2}^2 = \exp \gamma_{02} + \gamma_{12}z_i + \gamma_{22}z_i^2$, $i = \tau + 1, \dots, n$. Para encontrar la estimación de los parámetros se asignaron las distribuciones a priori no informativas $(\beta, \gamma) \sim N(0, 10^4 I_3)$ y $\tau \sim U(p + 1, n - p - 1)$, donde $n = 510$ y $p = 3$. Se realizaron 20.000 iteraciones y se descartaron las primeras 10.000, de este modo se obtuvieron las estimaciones bayesianas de la media, mediana y desviación estándar a posteriori presentadas en la Tabla 4.7.

n	Parámetro	Media	Mediana	D. Estándar
510	β_{01}	0.0372	0.0372	0.0019
	β_{11}	0.0071	0.0071	0.0012
	β_{21}	0.0003	0.0003	0.0001
	β_{02}	-0.1575	-0.1564	0.0341
	β_{12}	0.0215	0.0215	0.0031
	β_{22}	0.0006	0.0006	6.25×10^{-5}
	γ_{01}	-11.1593	-11.1633	0.3428
	γ_{21}	0.0387	0.0391	0.0132
	γ_{02}	-8.7875	-8.7906	0.3351
	γ_{12}	0.2059	0.2063	0.0231
	γ_{22}	-0.0012	-0.0012	0.0003
	τ	101.3449	101.9341	4.9550

TABLA 4.7. Resumen de resultados a posteriori para los datos de crecimiento de pollos

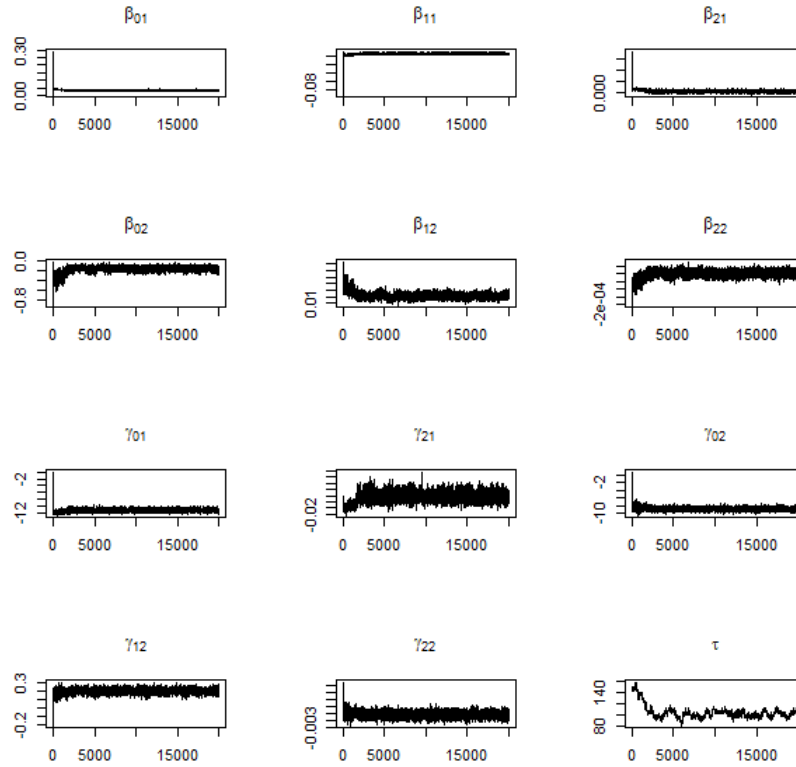


FIGURA 4.17. Comportamiento de las cadenas para cada uno de los parámetros del modelo de crecimiento de pollos

Se observa que el punto de cambio está ubicado en $\tau = 101$, lo que corresponde a las mediciones del peso de los pollos en el día 11. En la figura 4.14 se puede observar que el punto de cambio es detectado con una gran precisión. En la figura 4.15 se presentan los histogramas de las distribuciones marginales a posteriori de los parámetros.

La tasa de aceptación para los parámetros γ_1 , γ_2 y τ fue de 43.9%, 43.1% y 92.1%, respectivamente. Los valores obtenidos para los criterios de información son $AIC = -1002.723$ y $BIC = 951.909$. Para mostrar la utilidad de la inclusión del punto de cambio en la propuesta Bayesiana 3.2, se realiza la comparación contra un modelo de regresión lineal polinómico heterocedástico con media $\mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$ y varianza $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 z_i + \gamma_2 z_i^2)$. La estimación se hace a través de la metodología Bayesiana presentada en el capítulo 1.2.2. Se obtienen los parámetros estimados $\hat{\beta}_0 = 0.043810$, $\hat{\beta}_1 = 0.001778$, $\hat{\beta}_2 = 0.000991$, $\hat{\gamma}_0 = -10.608356$, $\hat{\gamma}_1 = 0.332329$, $\hat{\gamma}_2 = -0.003032$ y los estadísticos $AIC = -913.7416$ y $BIC = -888.335$.

Con el objetivo de presentar una mejor explicación de la utilidad de incluir el punto de cambio, en la figura 4.19 se presentan los gráficos de residuales para el modelo con cambio estructural y el modelo polinómico. Se observa que el comportamiento de los residuales en los modelos ajustados es similar. Además, la suma de cuadrados del error calculada sobre el modelo de regresión polinómico es $SCE = 29.99567$, mientras la suma de cuadrados del error del modelo propuesto es de $SCE = 24.72274$. En conclusión, el modelo propuesto brinda un mejor ajuste al obtenido con el modelo polinómico.

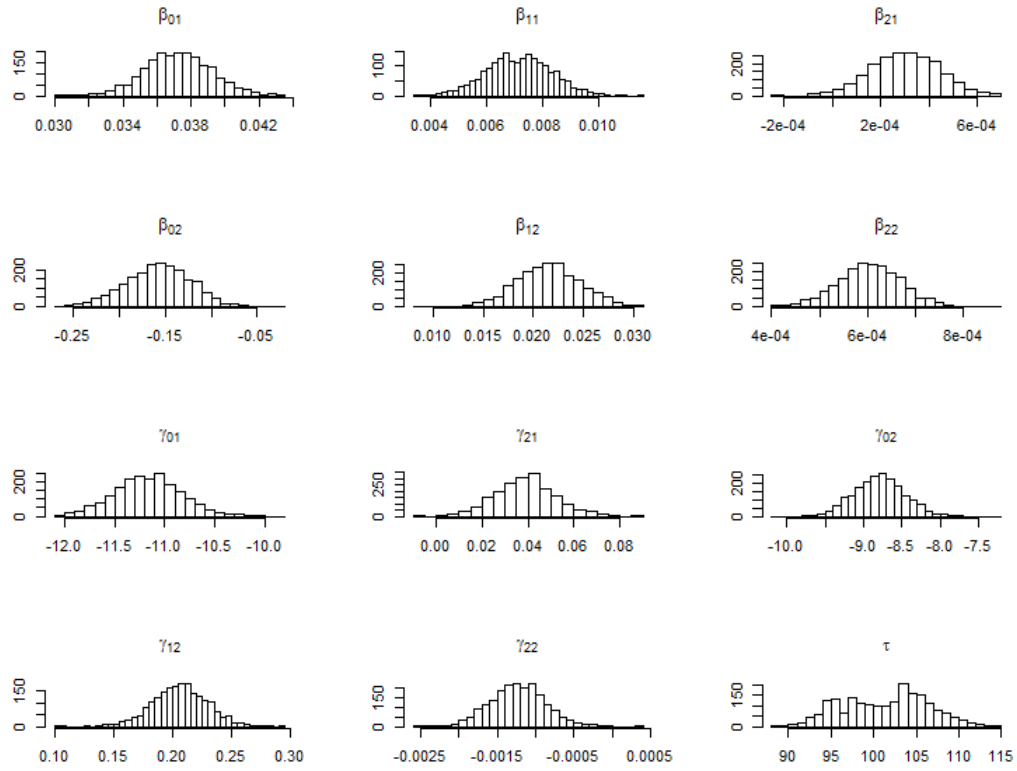


FIGURA 4.18. Histograma de las distribuciones marginales a posteriori para el modelo de crecimiento de pollos

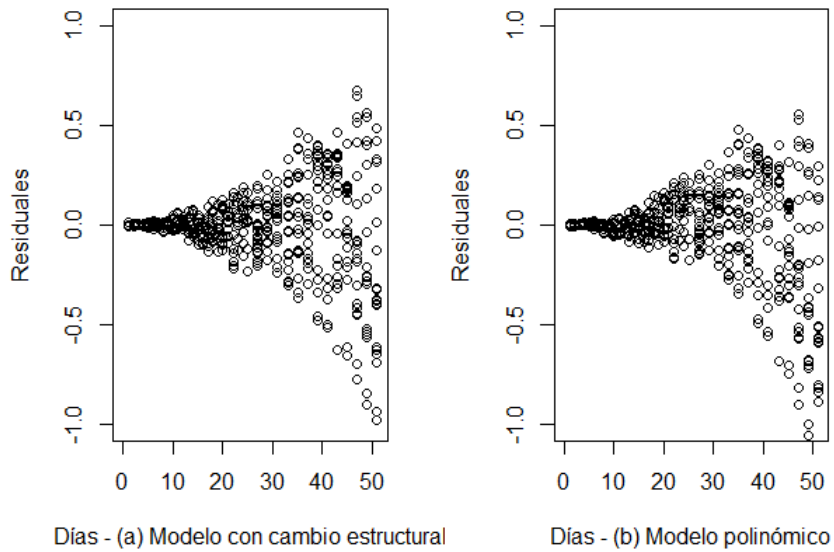


FIGURA 4.19. Residuales para los modelos de crecimiento de pollos

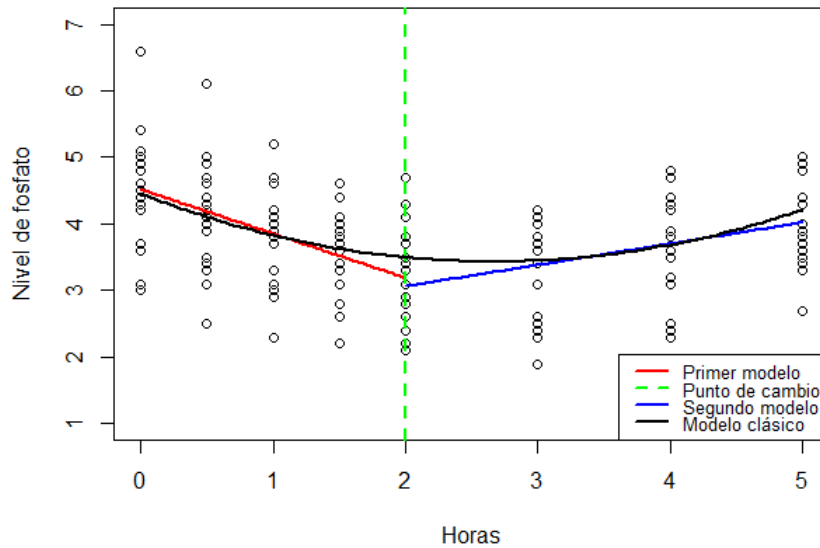


FIGURA 4.20. Gráfico de perfiles para datos de tolerancia a la insulina en pacientes obesos

4.2.3. Aplicación: Datos de tolerancia a la insulina en pacientes obesos

El departamento de Pediatría de la Escuela de Medicina en la Universidad de Colorado, en un estudio de la asociación de hiperglucemia e hiperinsulina, realizó pruebas sobre la tolerancia a la insulina. Estas pruebas se realizaron en 13 pacientes control y 20 pacientes obesos, en esta investigación se tomaron las mediciones de fosfato inorgánico en plasma (mg/dl), obtenidas a partir de muestras de sangre extraídas a 0, 0.5, 1, 1.5, 2, 3, 4 y 5 horas después de administrar la dosis estándar de glucosa. Los datos fueron reportados inicialmente por Chi & Reinsel (1989) en donde se encontró que las curvas de pacientes obesos y pacientes control difieren significativamente solamente durante las primeras tres horas después de administrada la glucosa. El consumo de glucosa determina una disminución en el fosfato inorgánico del plasma debido exclusivamente a la acción de la insulina.

En el gráfico 4.20 se observa que en el tiempo cero hay niveles de fosfato un tanto atípicos, disminución de las mediciones después de la media hora de administrada la glucosa hasta la segunda hora, luego en la hora tres se presenta un incremento en los niveles de fosfato hasta la hora cinco. Por el comportamiento descrito anteriormente, se presume que después de la segunda hora se presenta un cambio en la estructura de los datos. En la figura 4.21, se presenta el dispersograma de los niveles de fosfato para los diferentes tiempos de medición, se observa que la correlación en tiempos adyacentes es mayor que en tiempos posteriores. En la matriz 4.3 se calcula la correlación exacta entre los diferentes tiempos de medición, se obtiene una correlación alta en la primera banda, lo que sugiere

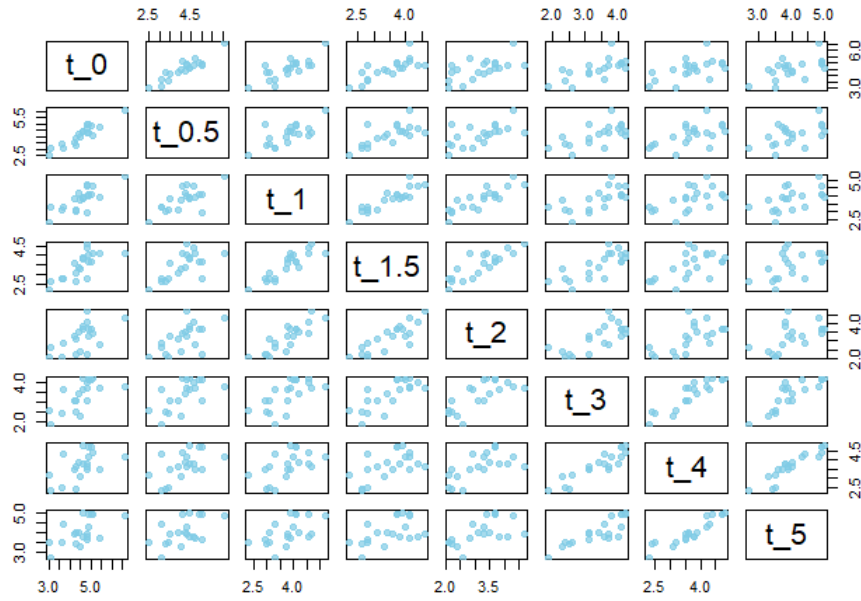


FIGURA 4.21. Dispersograma datos de tolerancia a la insulina en pacientes obesos para los diferentes tiempos de medición en pacientes obesos

que la estructura $AR(1)$ es la más conveniente para la matriz de covarianzas.

$$\begin{pmatrix}
 1 & 0.919 & 0.769 & 0.742 & 0.594 & 0.657 & 0.537 & 0.575 \\
 0.919 & 1 & 0.704 & 0.671 & 0.601 & 0.624 & 0.560 & 0.537 \\
 0.769 & 0.704 & 1 & 0.896 & 0.860 & 0.669 & 0.543 & 0.548 \\
 0.742 & 0.671 & 0.896 & 1 & 0.848 & 0.768 & 0.573 & 0.519 \\
 0.594 & 0.601 & 0.860 & 0.848 & 1 & 0.734 & 0.639 & 0.588 \\
 0.657 & 0.624 & 0.669 & 0.768 & 0.734 & 1 & 0.911 & 0.859 \\
 0.537 & 0.560 & 0.543 & 0.573 & 0.639 & 0.911 & 1 & 0.926 \\
 0.575 & 0.537 & 0.548 & 0.519 & 0.588 & 0.859 & 0.926 & 1
 \end{pmatrix} \quad (4.3)$$

Dado el comportamiento acampanado de los datos, Acosta (2010) propone ajustar un modelo de curvas de crecimiento por máxima verosimilitud con polinomios clásicos y polinomios fraccionarios, además emplea las estructuras de covarianza no estructurada (NE), uniforme, aleatoria, de Rao (ECR) y autoregresiva de orden uno $AR(1)$, para finalmente concluir que el modelo por polinomios fraccionarios de grado dos y potencias (1,1) con estructura de covarianza $AR(1)$ se ajusta mejor a los datos.

A través del mismo conjunto de datos, se realiza el ajuste del modelo 4.4 utilizando la metodología bayesiana propuesta para las estructuras de covarianza SC , $AR(1)$ y $ARMA(1,1)$ y se calculan los criterios de bondad de ajuste AIC y BIC . Finalmente, los resultados se comparan con la metodología propuesta por Castillo et al. (2019) en la que no se considera cambio estructural.

$$y_{ij} = \begin{cases} \beta_{01} + \beta_{11}t_{ij} + \epsilon_{ij1} & j = 1, \dots, \tau \\ \beta_{02} + \beta_{12}t_{ij} + \epsilon_{ij2} & j = \tau + 1, \dots, m. \end{cases} \quad (4.4)$$

n	m	Parámetro	Media	Mediana	D. Estándar
20	8	$\hat{\beta}_{01}$	4.490	4.489	0.159
		$\hat{\beta}_{11}$	-0.665	-0.665	0.054
		$\hat{\beta}_{02}$	2.418	2.419	0.213
		$\hat{\beta}_{12}$	0.319	0.319	0.038
		$\hat{\sigma}_1^2$	0.569	0.543	0.150
		$\hat{\sigma}_2^2$	0.477	0.448	0.147
		$\hat{\rho}_1$	0.726	0.729	0.071
		$\hat{\rho}_2$	0.870	0.876	0.045
		$\hat{\tau}$	5.444	5.423	0.281

TABLA 4.8. Resultados a posteriori para datos de tolerancia a la insulina con estructura de covarianza SC

4.2.3.1. Modelo longitudinal con estructura de covarianza SC

Para la estimación del modelo 4.4 se supone una estructura de covarianza SC como la presentada en la matriz 1.14, por lo tanto, se utiliza la metodología propuesta en la sección 3.3. Se asignaron las distribuciones a priori no informativas $\beta_d \sim N(0, 10^4 \mathbf{I}_2)$, $\sigma_d^2 \sim IG(10^{-4}/2, 10^{-4} * 100/2)$, $\rho_d \sim Beta(1, 1)$ y $\tau \sim U(2, 7)$.

Se realizaron 20.000 iteraciones y se descartaron las primeras 10.000, de este modo se obtuvieron las estimaciones bayesianas de la media, mediana y desviación estándar a posteriori presentadas en la tabla 4.8. Se encontró que la estimación del punto de cambio es $\hat{\tau} = 5$, por lo que el cambio estructural del modelo comienza en la medición de fosfato inorgánico a las 3 horas de aplicada la dosis estándar de glucosa.

La tasa de aceptación para los parámetro τ , ρ_1 y ρ_2 es de 66.87 %, 23.99 % y 34.07 %, respectivamente. Los valores obtenidos para los criterios de información son $AIC = 204.3$ y $BIC = 216.6$. Para mostrar la utilidad de la inclusión del punto de cambio en la propuesta Bayesiana 3.3 con estructura de covarianza SC , se realiza la comparación contra un modelo de regresión longitudinal polinómico de segundo orden sobre el tiempo en donde la estimación se hace a través de la metodología Bayesiana propuesta por Castillo et al. (2019). Se obtienen los parámetros estimados $\hat{\beta}_0 = 4.5874041$, $\hat{\beta}_1 = -0.8505952$, $\hat{\beta}_2 = 0.1594368$, $\hat{\sigma}^2 = 0.6655265$, $\hat{\rho} = 0.7261547$ y los estadísticos $AIC = 244.2953$ y $BIC = 259.6712$.

4.2.3.2. Modelo longitudinal con estructura de covarianza $AR(1)$

Para la estimación del modelo 4.4 se supone una estructura de covarianza $AR(1)$ como la presentada en la matriz 1.15, por lo tanto, se utiliza la metodología propuesta en la sección 3.3. Se asignaron las distribuciones a priori no informativas $\beta_d \sim N(0, 10^4 \mathbf{I}_2)$, $\sigma_d^2 \sim IG(10^{-4}/2, 10^{-4} * 100/2)$, $\rho_d \sim Beta(1, 1)$ y $\tau \sim U(2, 7)$.

Se realizaron 20.000 iteraciones y se descartaron las primeras 10.000, de este modo se obtuvieron las estimaciones bayesianas de la media, mediana y desviación estándar a posteriori presentadas en la tabla 4.9. Se encontró la misma estimación del punto de cambio a la presentada en el modelo con estructura de covarianza SC . El cambio estructural del modelo comienza en la medición de fosfato inorgánico a las 3 horas de aplicada la dosis estándar de glucosa.

n	m	Parámetro	Media	Mediana	D. Estándar
20	8	$\hat{\beta}_{01}$	4.516	4.516	0.170
		$\hat{\beta}_{11}$	-0.666	-0.666	0.085
		$\hat{\beta}_{02}$	2.417	2.417	0.239
		$\hat{\beta}_{12}$	0.319	0.318	0.047
		$\hat{\sigma}_1^2$	0.587	0.561	0.150
		$\hat{\sigma}_2^2$	0.458	0.432	0.138
		$\hat{\rho}_1$	0.827	0.829	0.044
		$\hat{\rho}_2$	0.886	0.892	0.045
		$\hat{\tau}$	5.444	5.423	0.281

TABLA 4.9. Resultados a posteriori para datos de tolerancia a la insulina con estructura de covarianza $AR(1)$

La tasa de aceptación para los parámetro τ , ρ_1 y ρ_2 es de 66.5 %, 25.095 % y 33.485 %, respectivamente. Los valores obtenidos para los criterios de información son $AIC = 189.1$ y $BIC = 201.4$. Para mostrar la utilidad de la inclusión del punto de cambio en la propuesta Bayesiana 3.3 con estructura de covarianza $AR(1)$, se realiza la comparación contra un modelo de regresión longitudinal polinómico de segundo orden sobre el tiempo, en donde la estimación se hace a través de la metodología Bayesiana propuesta por Castillo et al. (2019). Se obtienen los parámetros estimados $\hat{\beta}_0 = 4.5839297$, $\hat{\beta}_1 = -0.8634981$, $\hat{\beta}_2 = 0.1469383$, $\hat{\sigma}^2 = 0.5324053$, $\hat{\rho} = 0.8475753$ y los estadísticos $AIC = 204.4927$ y $BIC = 219.8685$.

4.2.3.3. Modelo longitudinal con estructura de covarianza $ARMA(1,1)$

Para la estimación del modelo 4.4 se supone una estructura de covarianza $ARMA(1,1)$ como la presentada en la matriz 1.16, por lo tanto, se utiliza la metodología propuesta en la sección 3.4. Se asignaron las distribuciones a priori no informativas $\beta_d \sim N(0, 10^4 \mathbf{I}_2)$, $\sigma_d^2 \sim IG(10^{-4}/2, 10^{-4} * 100/2)$, $\rho_d \sim Beta(1, 1)$, $\gamma_d \sim Beta(1, 1)$ y $\tau \sim U(2, 7)$.

Se realizaron 20.000 iteraciones y se descartaron las primeras 10.000, de este modo se obtuvieron las estimaciones bayesianas de la media, mediana y desviación estándar a posteriori presentadas en la tabla 4.10. Se encontró la misma estimación del punto de cambio a la presentada en el modelo con estructura de covarianza SC y $AR(1)$. El cambio estructural del modelo comienza en la medición de fosfato inorgánico a las 3 horas de aplicada la dosis estándar de glucosa.

La tasa de aceptación para los parámetro τ , ρ_1 y ρ_2 , γ_1 , γ_2 es de 66.86 %, 48.36 %, 67.32 %, 24.04 % y 34.41 %, respectivamente. Los valores obtenidos para los criterios de información son $AIC = 198.1$ y $BIC = 225.8$. Para mostrar la utilidad de la inclusión del punto de cambio en la propuesta Bayesiana 3.4 con estructura de covarianza $ARMA(1,1)$, se realiza la comparación contra un modelo de regresión longitudinal polinómico de segundo orden sobre el tiempo en donde la estimación se hace a través de la metodología Bayesiana propuesta por Castillo et al. (2019). Se obtienen los parámetros estimados $\hat{\beta}_0 = 4.6361314$, $\hat{\beta}_1 = -0.9150355$, $\hat{\beta}_2 = 0.1559074$, $\hat{\sigma}^2 = 0.6155763$, $\hat{\rho} = 0.8865394$, $\hat{\gamma} = 0.8421563$ y los estadísticos $AIC = 203.1732$ y $BIC = 221.6243$.

n	m	Parámetro	Media	Mediana	D. Estándar
20	8	$\hat{\beta}_{01}$	4.512	4.512	0.154
		$\hat{\beta}_{11}$	-0.667	-0.666	0.074
		$\hat{\beta}_{02}$	2.419	2.417	0.226
		$\hat{\beta}_{12}$	0.319	0.318	0.046
		$\hat{\sigma}_1^2$	0.521	0.504	0.119
		$\hat{\sigma}_2^2$	0.350	0.337	0.138
		$\hat{\rho}_1$	0.876	0.881	0.046
		$\hat{\rho}_2$	0.915	0.931	0.079
		$\hat{\gamma}_1$	0.797	0.803	0.050
		$\hat{\gamma}_1$	0.843	0.848	0.043
		$\hat{\tau}$	5.444	5.426	0.277

TABLA 4.10. Resultados a posteriori para datos de tolerancia a la insulina con estructura de covarianza $ARMA(1,1)$

Modelo	AIC	BIC
SC	204.3	216.6
$AR(1)$	189.1	201.4
$ARMA(1,1)$	198.1	225.8

TABLA 4.11. Criterios de bondad de ajustes para el modelo longitudinal con datos de tolerancia a la insulina en pacientes obesos

Con el propósito de realizar un comparación para los modelos estimados y determinar cuál de la estructuras de covarianza SC , $AR(1)$ y $ARMA(1,1)$ ajusta mejor los datos vía la metodología bayesiana propuesta, se presentan los valores de AIC y BIC .

El mejor AIC y BIC corresponde al ajuste por la metodología bayesiana propuesta con estructura de covarianza $AR(1)$. En la Figura 4.22, se muestra el comportamiento de las cadenas y en la Figura 4.23 se presentan los histogramas de las distribuciones marginales a posteriori para los parámetros del modelo longitudinal con estructura $AR(1)$.

En general, el modelo ajustado indica que en los primeros tiempos después de aplicar la dosis de glucosa en pacientes obesos, se observa una disminución en la cantidad promedio de fosfato en el plasma inorgánico y luego de la quinta medición, que corresponde a tres horas después de aplicado el tratamiento, los niveles promedio de fosfato aumentan.

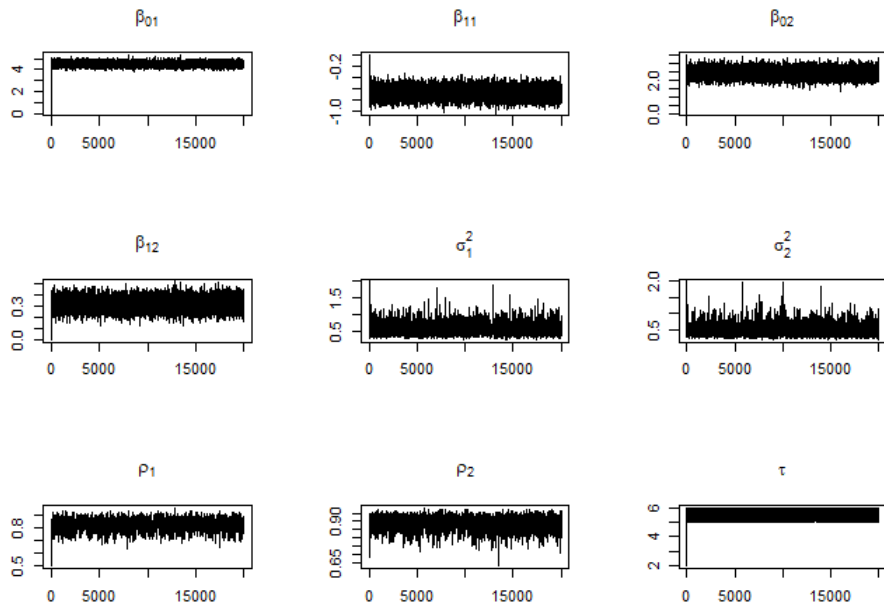


FIGURA 4.22. Comportamiento de las cadenas para cada uno de los parámetros del modelo de los datos de tolerancia a la insulina con estructura de covarianza $AR(1)$

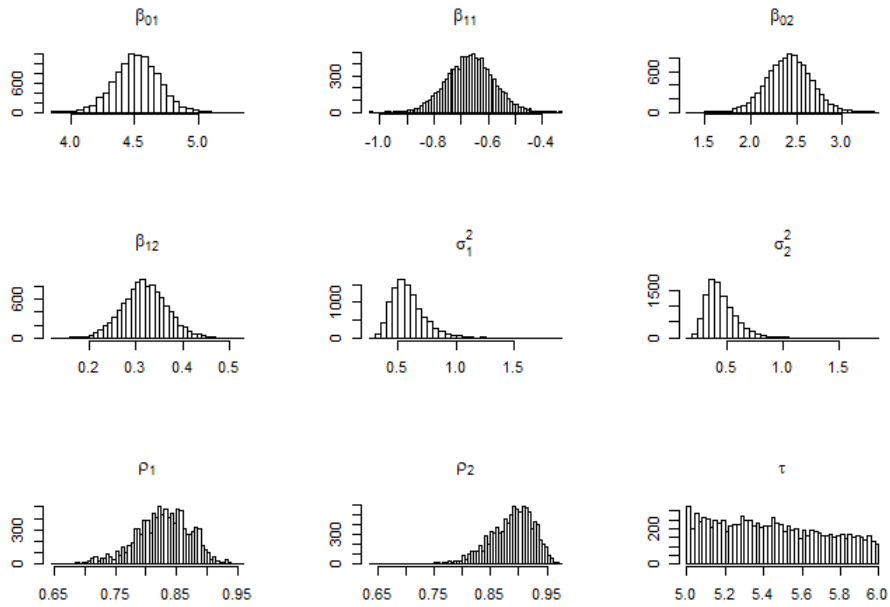


FIGURA 4.23. Histograma de las distribuciones marginales a posteriori para el modelo de los datos de tolerancia a la insulina con estructura de covarianza $AR(1)$

Conclusiones

- En este trabajo se propone modelar de forma novedosa el cambio en la estructura de media y varianza en modelos de regresión y modelos longitudinales a través de la estimación del parámetro de punto de cambio τ , esta metodología resulta ser mejor en términos de interpretación.
- Los kernel de transición 3.12 y 3.32 permiten la estimación del punto de cambio dentro del intervalo objetivo sin transformar o ignorar valores que caigan fuera de este. Estos kernel presentan buenas tasas de aceptación y permiten estimar el cambio con una gran precisión.
- Se encontró que la detección del punto de cambio depende en gran medida de la desviación entre las estructuras de los datos, aún así, en la mayoría de los casos el método propuesto detecta el punto de cambio con una gran precisión. Además, la elección de un valor inicial para τ lejano al parámetro no afecta el procedimiento de estimación, pues la convergencia al verdadero valor es rápida.
- Si no es de interés estimar el punto de cambio, el modelo de regresión con un punto de cambio puede ser interpretado como la mixtura de dos modelos de regresión. Por lo anterior, la metodología bayesiana propuesta es un buen acercamiento al estudio de mixturas de modelos.
- La implementación de las propuestas Bayesianas para la estimación conjunta de la estructura de media, matriz de covarianzas y punto de cambio para modelos longitudinales con cambio estructural en los cuales se consideran estructuras de SC , $AR(1)$ y $ARMA(1,1)$, retorna los valores de las simulaciones realizadas. La elección de una distribución a priori informativa para el punto de cambio y los parámetros de correlación de estos modelos proporciona un ajuste más preciso y mejora las tasas de aceptación de algoritmo MCMC.
- Los datos de expansión térmica del cobre presentan cambio en la estructura de media, al inicio de las observaciones y hasta aproximadamente los 200 grados Kelvin hay un incremento acelerado, posteriormente se estabiliza el comportamiento de la variable. Mediante la metodología bayesiana propuesta para modelos de regresión con cambio estructural y varianza homocedástica, se obtuvieron mejores resultados aproximados a los dados por el modelo de regresión lineal. Los criterios de comparación AIC y BIC , obtenidos desde el ajuste del modelo con la metodología propuesta son mejores a los obtenidos con el modelo.
- Para los datos de crecimiento de pollos, la variación aumenta con el tiempo y la tasa de crecimiento tiene forma de S. A través del modelo de regresión con cambio

estructural y varianza heterocedástica propuesto se obtienen resultados similares a los obtenidos a través de un modelo de regresión lineal polinómico sobre la covariable con varianza heterocedástico. Los criterios de comparación *AIC* y *BIC*, obtenidos desde la propuesta son ligeramente mayores a los obtenidos con el ajuste del modelo sin considerar cambio estructural.

- El ajuste de los datos de tolerancia a la insulina en pacientes obesos mediante la metodología bayesiana propuesta para modelos de longitudinales, retornó mejores resultados a los presentados por Chi & Reinsel (1989) y Acosta (2010), con buenas tasas de aceptación desde el punto de vista teórico. Adicional, se realiza la comparación contra un modelo longitudinal polinómico en donde el modelo propuesta presenta un mejor ajuste a los datos. Los criterios de comparación *AIC* y *BIC*, obtenidos desde el ajuste del modelo con la metodología propuesta son menores a los obtenidos con el modelo polinómico.

Trabajo futuro

- Generalizar los algoritmos presentados para la estimación bayesiana de modelos de regresión y modelos longitudinales con más de un punto de cambio.
- Realizar una propuesta bayesiana para modelos de regresión y longitudinales con puntos de cambio continuos, como los presentados por Hofrichter (2007).
- Extender la propuesta bayesiana para el ajuste de datos longitudinales con otras estructuras de covarianza.
- Implementar, en un paquete en R las funciones y algoritmos propuestos en este trabajo.

Bibliografía

- Acosta, L. C. (2010). *Ajuste de Polinomios Fraccionarios a Curvas de Crecimiento*, Msc thesis Universidad Nacional de Colombia.
- Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using glim, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **36**.
- Auger, I. E. & Lawrence, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods, *Bulletin of Mathematical Biology* **51**.
- Bai, J. (1997). Estimation of a change point in multiple regression models, *The Review of Economics and Statistics* **4**(79).
- Bai, J. & Perron, P. (1998). Estimating and testing linear models with multiple structural changes, *Econometrica* **1**(66).
- Bai, J. & Perron, P. (2003). Computation and analysis of multiple structural change models, *Journal of Applied Econometrics* (18).
- Box, G. E. P. & Cox, D. R. (1964). An analysis of transformation (with discussion), *Journal of the Royal Statistical Society B* **26**.
- Brown, R. L., Durbin, J. & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society Series B (Methodological)* **37**.
- Casella, G. & Berger, R. L. (2001). *Statistical Inference*, 2nd edn, Duxbury Press.
- Castillo, E., Cepeda, E. & Núñez Antón, V. (2019). *Bayesian Structured Antedependence Model Proposals for Longitudinal Data*, unpublished paper.
- Cepeda, E. (2001). *Modelagem da Variabilidade em Modelos Lineares Generalizados*, PhD thesis Universidade Federal do Rio de Janeiro.
- Cepeda, E. (2011). Generalized spatio-temporal models, *SORT* **35**.
- Cepeda, E. & A., A. J. (2009). Regression models with heteroscedasticity using bayesian approach, *Revista Colombiana de Estadística* **32**.
- Cepeda, E. & Gamerman, D. (2001). Bayesian modeling of variance heterogeneity in normal regression models, *Brazilian Journal of Probability and Statistics* **11**.
- Cepeda, E. & Gamerman, D. (2004). Bayesian modeling of joint regressions for the mean and covariance matrix, *Biometrical Journal* **46**.

- Cepeda, E. & Gamerman, D. (2005). Bayesian methodology for modeling parameters in the two parameter exponential family, *Estadística* **57**(168 y 169).
- Cepeda, E. & Núñez Antón, V. (2007). Bayesian joint modelling of the mean and covariance structures for normal longitudinal data, *SORT* **31**.
- Cepeda, E. & Núñez Antón, V. (2009). Bayesian modelling of the mean and covariance matrix in normal nonlinear models, *Journal of Statistical Computation and Simulation* **79**.
- Chaturvedi, A. & Shrivastava, A. (2016). Bayesian analysis of a linear model involving structural changes in either regression parameters or disturbances precision, *Communications in Statistics-Theory and Methods* **2**(45).
- Chen, J. & Gupta, A. K. (2012). *Parametric Statistical Change Point Analysis With Applications to Genetics, Medicine, and Finance*, 2nd edn, Birkhauser.
- Chi, E. M. & Reinsel, G. C. (1989). Models for longitudinal data with random effects and ar(1) errors, *Journal of the American Statistical Association* **84**(406).
- Cooper, R. F. & B., Y. (1970). Low temperature interferometric measurements of the thermal expansion of copper and nitralloy, *Cryogenics* **10**.
- Daniels, M. J. & Pourahmadi, M. (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data, *Biometrika* **89**.
- Davis, C. (2002). *Statistical Methods for the Analysis of Repeated Measurements*, Springer.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em-algorithm, *Journal of the Royal Statistical Society* **39**.
- Diggle, P., Heagerty, P., Liang, K.-Y. & Zeger, S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*, 2nd edn, Oxford University Press.
- Draper, N. R. & Smith, H. (1998). *Applied Regression Analysis*, 3rd edn, John Wiley & Sons.
- Edwards, A. W. F. & Cavalli-Sforza, L. L. (1965). A method for cluster analysis, *Biometrics* **21**.
- Ferreira, P. E. (1975). A bayesian analysis of a switching regression model: Known number of regimes, *Journal of the American Statistical Association* **70**(350).
- Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models, *Statistics and Computing* **7**.
- Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd edn, Chapman & Hall/CRC. Text in Statistical Science Series.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2014). *Bayesian Data Analysis*, 3rd edn, Chapman & Hall/CRC.
- Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity, *Econometrica* **44**.

-
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems, *Journal of the American Statistical Association* **72**.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications, *Biometrika* **57**.
- Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables, *Biometrika* **57**(1): 1–17.
- Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*, Springer Texts in Statistics.
- Hofrichter, J. (2007). *Change Point Detection in Generalized Linear Models*, PhD thesis TU Graz.
- Holbert, D. (1982). A bayesian analysis of a switching linear model, *Journal of Econometrics* **19**.
- Killick, R. & Eckley, I. A. (2014). changepoint: An R package for changepoint analysis, *Journal of Statistical Software* **58**(3): 1–19.
- Killick, R., Fearnhead, P. & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost, *Journal of the American Statistical Association* **107**.
- Kim, H.-J. (1993). Two-phase regression with nonhomogeneous errors, *Communication in Statistics- Theory and Methods* **22**.
- Kim, H.-J. (1994a). Change-point problems. tests for a change-point in linear regression, *IMS Lecture Notes–Monograph Series* **23**.
- Kim, H.-J. (1994b). Likelihood ratio and cumulative sum tests for a change-point in linear regression, *Journal of Multivariate Analysis* **51**.
- Krämer, W., Ploberger, W. & Alt, R. (1988). Testing for structural change in dynamic models, *Econometrica* **56**.
- Leisch, F. (2004). Flexmix: A general framework for finite mixture models and latent class regression in r, *Journal of Statistical Software* **11**.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn, Springer-Science+Business Media, B.V.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equation of state calculations by fast computing machines, *The Journal of Chemical Physics* **21**.
- Migon, H. S. & Gamerman, D. (1999). *Statistical Inference: an Integrated Approach*, Arnold, a member of the Hodder Headline Group.
- Montgomery, D. C., Peck, E. A. & Vining, G. G. (2001). *Introduction to linear Regression Analysis*, John Wiley & Sons.
- Núñez Antón, V. A. & Zimmerman, D. L. (2001). Modelización de datos longitudinales con estructuras de covarianza no estacionarias: Mode los de coeficiente aleatorios frente a modelos alternativos, *Questiío* **25**.

- Núñez Antón, V. & Woodworth, G. G. (1994). Analysis of longitudinal data with unequally spaced observations and time-dependent correlated errors, *Biometrics* **50**.
- Patterson, H. D. & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal, *Biometrika* **58**.
- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation, *Biometrika* **86**.
- Quandt, R. E. (1958). The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes, *American Statistical Association* **53**(284): 873–880.
- Quandt, R. E. (1960). Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes, *Journal of the American Statistical Association* **55**(290): 324–330.
- Ravishanker, N. & Dey, D. K. (2001). *A First Course in Linear Model Theory*, 1st edn, Chapman & Hall/CRC Texts in Statistical Science.
- Riazoshams, H., Midi, H. & Ghilagaber, G. (2018). *Robust Nonlinear Regression with Applications Using R*, Wiley.
- Riazoshams, H. & Miri, H. (2005). *Investigating growth models using nonlinear regression models*, Technical report, Islamic Azad University, Abade branch, Fars province, Iran.
- Singer, J. D. & Willett, J. B. (2003). *Applied Longitudinal Data Analysis - Modeling Change and Event Occurrence*, Oxford University Press.
- Smith, H. & Dubey, S. (1964). Some reliability problems in the chemical industry, *Industrial Quality Control* **21**.
- Verbeke, G. & Molenberghs, G. (2009). *Linear Mixed Models for Longitudinal Data*, 2nd edn, Springer Series in Statistics.
- Weiss, R. E. (2005). *Modeling Longitudinal Data*, Springer Science & Business Media.
- Worsley, K. J. (1982). An improved bonferroni inequality and applications, *Biometrika* **69**.
- Worsley, K. J. (1983). Testing for a two-phase multiple regression, *Technometrics* **25**.
- Zeileis, A., Leisch, F., Hornik, K. & Kleiber, C. (2002). strucchange: An r package for testing for structural change in linear regression models, *Journal of Statistical Software* **7**(2): 1–38.
- Zimmerman, D. L. & Núñez Antón, V. A. (2009). *Antedependence Models for Longitudinal Data*, CRC Press - Taylor & Francis Group.