



FACULTAD DE CIENCIAS
INSTITUTO DE ESTADÍSTICA

MODELO MIXTO SEMIPARAMÉTRICO
ELÍPTICO CON ERRORES $AR(1)$ PARA
DATOS LONGITUDINALES

Trabajo final presentado por:

Alan Barraza S.

Para optar al grado de:

Magíster en Estadística

Profesor guía:

Dr. Germán Ibacache-Pulgar

Valparaíso, 2016

AGRADECIMIENTOS

Primero que todo, agradecer a dios Jehova = señor de los ejercitos, ante todas las cosas, por brindarme conocimiento y bendecirme con un hijo que llamaremos Matías Daniel. También, mi más sincero amor y agradecimiento a mi esposa Bernardita Pérez por su paciencia y amor incondicional, a mi madre, hermanos, familia y mi suegra Sra. Elsa dedico con mucho cariño el presente trabajo.

Además, una especial dedicación a la memoria del Profesor Renato Allende Olivares, Profesor José Pascual Sánchez y a mi abuelo José Luis Saez Barrios Q.E.P.D.

Alan Phillips B. Saez.

RESUMEN

El presente proyecto considera el estudio de los modelos mixtos semiparamétricos (MMS) con errores aleatorios que poseen una distribución de contornos elípticos para datos longitudinales con una estructura de correlación de tipo AR(1). En este contexto, se propone un método de estimación para los parámetros del modelo basado en el criterio de máxima verosimilitud penalizada, para así obtener las funciones de score y la matriz de información de Fisher esperada asociados a los parámetros del modelo. Además, se describen algunas medidas de diagnóstico de residuos del modelo basados en las distancias de Cook y del valor estimado estandarizado. Como también, se proponen dos test estadísticos para probar las hipótesis de autocorrelación de primer orden intra-unidades y homogeneidad de la autocorrelación entre-unidades, estudiando su comportamiento a través de su potencia empírica por medio de simulaciones Monte Carlo. Finalmente, se aplicará la metodología descrita en datos reales y la implementación computacional de la metodología fue en el software R-Cran¹ 3.0.1 (Versión x64).

Palabras clave: Modelos mixtos semiparamétricos; Distribuciones de elípticas; errores AR(1); Datos longitudinales; Distancia de Cook; DFFITS; Test de Score; Test de autocorrelación.

¹Software de libre distribución, disponible en <https://cran.r-project.org>

CONTENIDOS

RESUMEN	1
1 INTRODUCCIÓN	5
1.1 Una motivación para el uso de modelos semiparamétricos . . .	5
1.2 Modelos para datos longitudinales	9
1.3 Supuesto distribucional del modelo	10
1.4 Algunas propiedades de las distribuciones elípticas	12
HIPÓTESIS DE TRABAJO	17
OBJETIVOS	18
2 MODELO MIXTO SEMIPARAMÉTRICO ELÍPTICO PARA DATOS LONGI- TUDINALES	19
2.1 Especificación del modelo	19
2.2 Representación matricial	20
2.2.1 Matriz de incidencia	21
2.2.2 Modelo mixto semiparamétrico	21
2.3 Estructura de autocorrelación de primer orden	22
2.4 Modelo bajo distribución elíptica	23
2.5 Estructura de covarianza del MMS	25
2.6 Comentarios del capítulo	26
3 ESTIMACIÓN DE MÁXIMA VEROSIMILITUD PENALIZADA	27
3.1 Criterios de verosimilitud penalizada	28

3.1.1	Función de penalización	29
3.1.2	Función de log-verosimilitud penalizada	31
3.2	Funciones de score	31
3.2.1	Score componente paramétrica	32
3.2.2	Score componente no paramétrica	33
3.2.3	Score componentes de varianza	33
3.3	Matriz de Información de Fisher penalizada	34
3.4	Proceso de estimación	36
3.4.1	Estimación de los efectos fijos	38
3.4.2	Estimación de los componentes de varianza	38
3.4.3	Proceso iterativo	39
3.4.4	Estimación de los efectos aleatorios	42
3.4.5	Estimación del error estándar	43
3.5	Selección del modelo	45
3.5.1	Validación cruzada generalizada	45
3.5.2	Criterios de información	46
3.6	Comentarios del capítulo	47
4	ANÁLISIS DIAGNÓSTICO DEL MMS ELÍPTICO PARA DATOS LONGI- TUDINALES	48
4.1	Análisis de residuos del MMS elíptico	49
4.1.1	Residuo marginal del modelo	49
4.1.2	Residuo condicional del modelo	50
4.1.3	Residuos estandarizados	50
4.2	Diagnóstico de influencia global para los efectos fijos	51
4.2.1	Diagnóstico basado en supresión de una observación	52
4.2.2	Influencia de la componente paramétrica	53
4.2.3	Influencia de la componente no paramétrica	54
4.3	Comentarios del capítulo	55
5	TEST DE SCORE PARA AUTOCORRELACIÓN DE PRIMER ORDEN INTRA-UNIDAD	56
5.1	Test para la presencia de autocorrelación	57
5.1.1	Test de score para evaluar $H_0 : \rho = 0$	58
5.1.2	Potencia empírica de SC_a	59
5.2	Test para la homogeneidad de autocorrelación	63
5.2.1	Test de Score para evaluar $H_0 : \rho_i = \rho$	65
5.2.2	Potencia empírica de SC_h	67

5.3	Comentarios del capítulo	68
6	APLICACIONES	72
6.1	Conjunto de datos N°1: fosfato inorgánico	73
6.1.1	Modelo propuesto	75
6.1.2	Ajuste del MMS elíptico	76
6.1.3	Análisis diagnóstico del modelo	81
6.1.4	Análisis de autocorrelación de primer orden	87
6.1.5	Conclusiones del análisis de datos	89
6.2	Conjunto de datos N°2: índices Odontológicos	90
6.2.1	Modelo propuesto	93
6.2.2	Ajuste del MMS elíptico	94
6.2.3	Análisis diagnóstico del modelo	99
6.2.4	Análisis de autocorrelación de primer orden	105
6.2.5	Conclusiones del análisis de datos	107
7	CONSIDERACIONES FINALES	109
7.1	Discusión	109
7.2	Implementación computacional	111
7.3	Consideraciones de trabajo futuro	112
	APENDICE I: GRÁFICOS DEL CAPÍTULO 1 EN R-CRAN	113
	APENDICE II: DEMOSTRACIONES DEL CAPÍTULO 3	119
	APENDICE III: DEMOSTRACIONES DEL CAPÍTULO 4	141
	APENDICE IV: DEMOSTRACIONES CAPITULO 5	147
	APENDICE V: DATOS FOSFATO INORGÁNICO DE PLASMA	153
	APENDICE IV: DATOS DE ÍNDICES ODONTOLÓGICOS	155
	BIBLIOGRAFÍA	161

INTRODUCCIÓN

En este capítulo se describe algunos motivos para el uso de modelos semiparamétricos, dando una descripción general de estos modelos, mostrando algunas ventajas de su uso en comparación con los modelos tradicionales en el contexto de datos longitudinales. Además, se describe la clase de distribuciones de contornos elípticos, la cual es el supuesto base del modelo semiparamétrico propuesto en el presente trabajo, además, se muestran algunas propiedades de las distribuciones elípticas que fueron utilizadas en los capítulos posteriores.

1.1 Una motivación para el uso de modelos semiparamétricos

En el contexto de diseños experimentales, se requiere estudiar la relación entre una variable respuesta y con un conjunto de p variables que explican el comportamiento de y , comúnmente, se propone estudiar dicho comportamiento por medio de un modelo lineal, no lineal, con efectos aleatorios o efectos mixtos, una extensa teoría de todos estos modelos puede ser visto en: Mardia (1979), Rao (2002), Rencher (2002; 2008), Seber (2004), Manor & Zucker (2004), Muller & Stewart (2006), Bates & Watts

(1988), Seber & Wild (2003) y Ritz & Streibig (2008). También se puede considerar un modelo no paramétrico, el cual es mucho más flexible que un modelo paramétrico, debido a que no incorporan muchas restricciones en su estructura sistemática, pero se presume que la función del modelo de la parte sistemática es suficientemente regular, del cual, un supuesto habitual es que la función provenga de una clase \mathfrak{F} , donde \mathfrak{F} es un espacio de funciones continuas y diferenciables, esta función es denominada como una función suave¹. Una extensa teoría de estos modelos puede ser vista en: Eubank (1999), Efromovich (1999), Györfi *et al.* (2002), Takezawa (2002), Govindarajulu (2007) y Tsybakov (2009).

Sin embargo, existe una clase de modelos que unifica los modelos antes mencionados, conocidos como modelos semiparamétricos y son mucho más flexibles que los modelos paramétricos. Estos modelos se utilizan a menudo cuando el uso de modelos paramétricos y no paramétricos por separado no funcionan bien, y esto puede ser debido a que un cierto subconjunto de covariables explica de mejor manera la respuesta a partir de una estructura paramétrica y el resto explica de mejor manera a partir de una estructura no paramétrica. Una característica particular de los modelos semiparamétricos es que el parámetro asociado a la componente paramétrica no se considera el intercepto, puesto que dicho parámetro es absorbido por la componente no paramétrica del modelo. El modelo semiparamétrico más simple es conocido con el nombre de modelo lineal parcial. Las investigaciones de Speckman (1988) y Robinson (1988) fueron pioneras en el uso de los modelos lineales parciales, algunos otros trabajos basados de este modelo pueden ser vistos en Hamilton & Truong (1997), Wang & Jing (2003) y Liang & Li (2009).

Otro modelo semiparamétrico es conocido como no lineal parcial, que es una clase de modelo que generaliza el modelo lineal parcial donde la función $f \in \mathfrak{F}$ y g es una función conocida. En cambio, el modelo aditivo separable es un caso particular de un modelo no paramétrico, donde $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$ pero bajo el supuesto de que $f_1 + f_2 \in \mathfrak{F}$, note que los modelos lineal parcial y no lineal parcial también son modelos aditivos separables pero en el sentido de las componentes paramétrica y no paramétrica.

¹del inglés: smooth

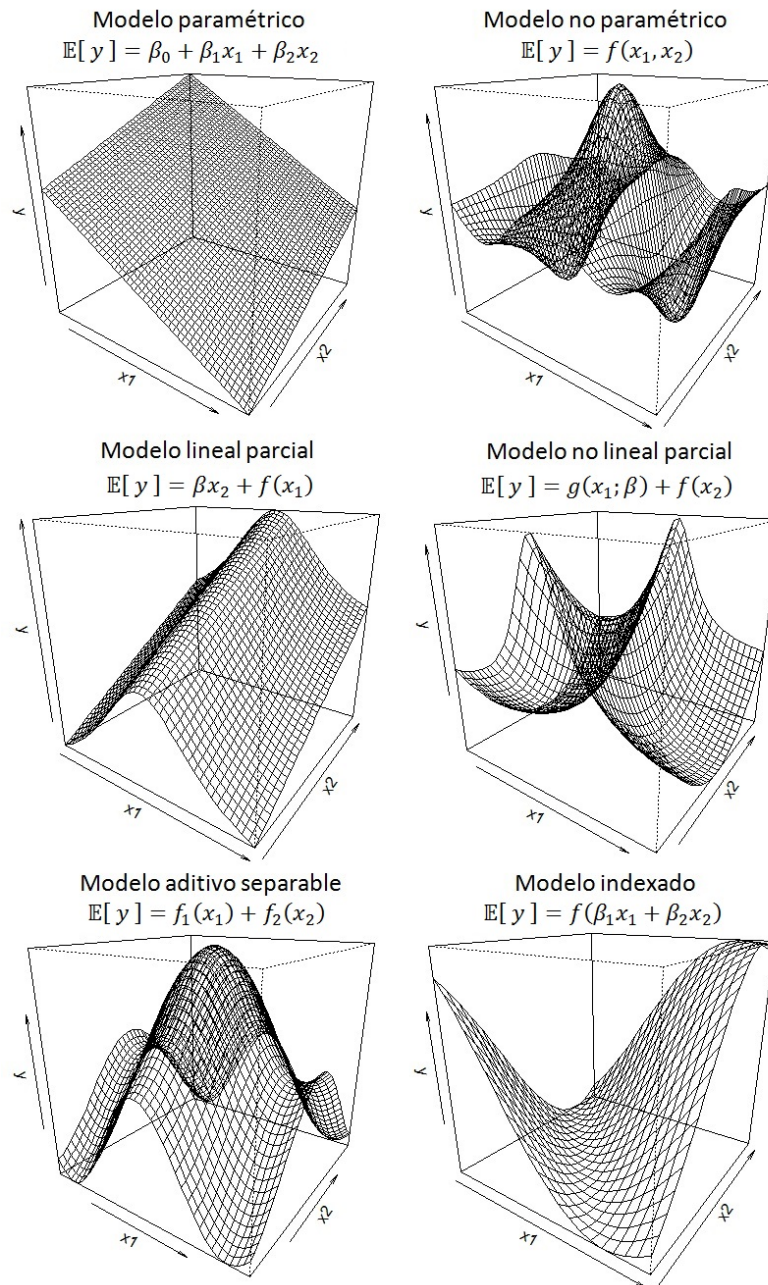


Figura 1.1: Representación de algunos modelos semiparamétricos en \mathbb{R}^2
(Fuente: Elaboración propia)

El modelo indexado es una regresión de tipo no paramétrica la cual está indexada a una combinación lineal de las observaciones y un vector de parámetros desconocidos que corresponde a la parte paramétrica, esto es, la parte no paramétrica esta indexada a la parte paramétrica, la ventaja de estos modelos es que no se ven afectados por la dimensión de las variables explicativas como en el caso de los modelos puramente no paramétricos, también se ha propuesto una generalización de este último modelo llamado modelo indexado lineal parcial.

En la Figura 1.1 se muestran algunas representaciones de modelos para la respuesta media en el software *R-Cran*², una literatura más detallada de los tipos de modelos semiparamétricos, sus propiedades y análisis diagnósticos, pueden ser vistos en: Buja *et al.* (1989), Kim, Park & Kim (2001; 2002), Fung *et al.* (2002), Yatchew (2003), Yatchew (2003), Zhang & Lin (2003), Ruppert *et al.* (2003; 2009), Härdle *et al.* (2004), Härdle *et al.* (2004), Fengler (2005), Sperlich, Härdle & Aydinli (2006), Tsiatis (2006), Keele (2008), Kosorok (2008), Horowitz (2009), Lian & Li (2009), Horowitz (2009) e Ibacache, Paula & Cysneiros (2013).

En el presente trabajo se propone un modelo semiparamétrico con una estructura sistemática más general que el modelo lineal parcial propuesto por Speckman (1988) y Robinson (1988), donde la componente paramétrica es p -dimensional, y tanto la componente paramétrica y no paramétrica representan los efectos fijos del modelo. Además, se propone una medida diagnóstica de la influencia global de las componentes paramétrica y no paramétrica del modelo, basada en la distancia de Cook generalizada y el valor estimado estandarizado, que corresponden a una generalización de la medida diagnóstica propuesta por Fung *et al.* (2002).

²ver Apéndice I

1.2 Modelos para datos longitudinales

Para el análisis de datos longitudinales, los modelos con efectos mixtos son comúnmente los más utilizados, estos modelos son mucho más flexibles para modelar la estructura de covarianza entre las observaciones medidas dentro de una unidad, lo cual permite tratar con unidades de estudio que poseen más de una observación, tales como en los casos balanceados y desbalanceados, una extensa teoría de estos modelos pueden ser vistos en: Laird & Ware (1982), Verbeke & Molenberghs (2001), Diggle (2002), Arnau & Balluerka (2004), Hedeker & Gibbons (2006) y Fitzmaurice *et al.* (2009). Una observación longitudinal puede ser considerada como una medida repetida, en el sentido de que en una unidad, se observa una característica de interés por un determinado período de tiempo, de esta manera, es razonable suponer una estructura autorregresiva a nivel intra-unidad.

En este contexto, varios autores han propuesto algunos modelos más generales para datos longitudinales, éstos pueden ser vistos en: Fitzmaurice *et al.* (1993), Lin & Ying (2003), Chen, Wang & Huang (2004), Molenberghs & Verbeke (2005), Wu & Zhang (2006), Wu & Zhang (2006), Osorio, Paula & Galea (2007), Paula, Medeiros & Vilca-Labra (2009). Además, de incorporar una estructura autorregresiva en el modelo, también es razonable considerar evaluar, la hipótesis de presencia de autocorrelación y si ésta es homogénea a través de las unidades de estudio. Para esto, existen varios tipos de metodologías para la evaluación de la presencia y homogeneidad de autocorrelación intra-unidad, donde el test de score posee cierta ventaja asociada al esfuerzo del proceso de cálculo computacional, respecto de los test de razón de verosimilitud y Wald, algunas aplicaciones de dicho test pueden ser vistas en: Tsai (1986), Chi & Reinsel (1989), Chi & Reinsel (1991), Lin & Wei (2004), Cao & Lin (2012).

En adelante, supondremos que la componente paramétrica tiene asociado un vector de variables explicativas y la componente no paramétrica sólo depende del tiempo de medición en donde la observación fue realizada. Además, se incorpora una estructura de autocorrelación de primer orden en los errores del modelo, lo que implica que la estructura de varianza del modelo, contiene parámetros asociados a la autocorrelación intra-unidad, los efectos y errores aleatorios.

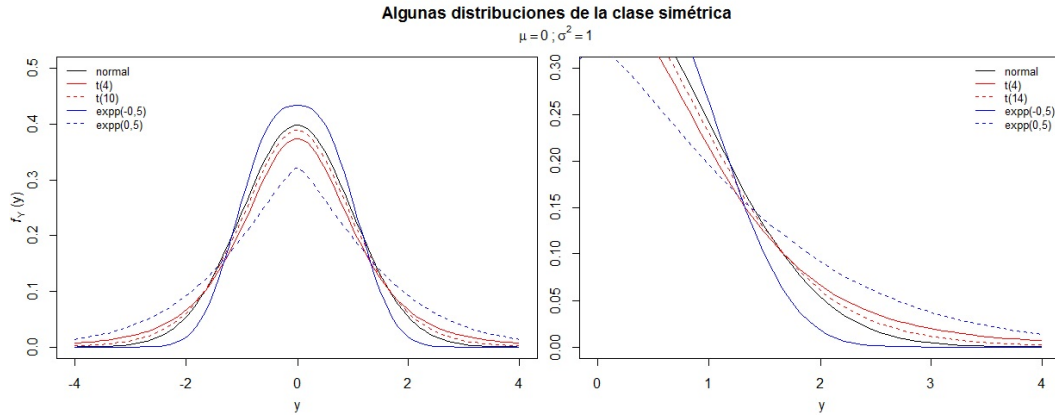


Figura 1.2: Algunas distribuciones simétricas en los reales (Fuente: elaboración propia)

1.3 Supuesto distribucional del modelo

Por muchos años el análisis inferencial y diagnóstico en el contexto de los modelos lineales, se ha basado en el supuesto de normalidad de los errores aleatorios; ver por ejemplo, Mardia *et al.* (1979), Cook D. (1986), Seber G. (2004), Rao C.R. (2004), Rencher A. (2008). Sin embargo, en muchas situaciones, la distribución normal no es el modelo estadístico más adecuado para ajustar a un conjunto de datos, principalmente, cuando éstos presentan observaciones atípicas y potencialmente influyentes.

Algunos autores han propuesto utilizar una clase más amplia de distribuciones que contiene, como caso particular, la distribución normal. Específicamente, nos referimos a la clase de distribuciones de contornos elípticos, la cual contiene distribuciones con colas más pesadas o más livianas que la distribución normal, tales como la distribución *t*-Student, logística I y II, exponencial potencia y normal contaminada, más detalles en: Fang, Kotz & Ng (1989), Fang & Zhang (1990), Gupta & Varga (1993), Genton & Loperfido (2005) y Vernic (2005). En la Figura 1.2, se muestran algunas distribuciones simétricas realizadas en el software *R-Cran*³, para ver otros trabajos relacionados con esta clase de distribuciones, ver por ejemplo: Galea, Paula & Bolfarine (1997), Galea, Riquelme & Paula (2000), Liu

³ver Apéndice I

(2002), Galea, Paula & Uribe-Opazo (2003), Maruyama & Seo (2003), Cordeiro (2004), Cysneiros & Paula (2005), Galea, Paula & Cysneiros (2005), Cysneiros, Paula & Galea (2007) y Riquelme, Leiva, Galea & Sanhueza (2011).

Cuadro 1.1: Algunas distribuciones de la clase elíptica, donde c es una constante de normalización y u la distancia de Mahalanobis

Distribución	Densidad = $cg(u)$	Restricciones
Normal	$c \exp(-u/2)$	$u > 0$
t -Student	$c \left(1 + \frac{u}{\nu}\right)^{-(\nu+n)/2}$	$n, \nu, u > 0$
Exponencial potencia	$c \exp(-\frac{1}{2}u^{1/(1+\vartheta)})$	$-1 < \vartheta \leq 1, u > 0$
Logística	$c \frac{\exp(-u)}{[1 + \exp(-u)]^2}$	$u > 0$
Tipo Kotz	$cu^{N-1} \exp(-ru^s)$	$r, s, u > 0, 2N + n > 2$
Pearson Tipo II	$c(1-u)^m$	$m, u > 0$
Pearson Tipo VII	$c \left(1 + \frac{u}{s}\right)^{-N}$	$u, s > 0, N > \frac{n}{2}$

En virtud de lo anterior, el análisis inferencial del presente trabajo se asume un modelo de regresión semiparamétrico con efectos mixtos (MMS) y errores aleatorios que siguen una distribución de contornos elípticos, esta familia es una clase de distribuciones multivariadas que corresponden a todas las distribuciones simétricas en \mathbb{R}^n .

Si $\mathbf{x} \in \mathbb{R}^n$ es un vector aleatorio con distribución de contornos elípticos entonces su función de densidad está dada por:

$$f_{\mathbf{x}}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, g) = |\boldsymbol{\Sigma}|^{-1/2} g_n([\mathbf{x} - \boldsymbol{\mu}]^t \boldsymbol{\Sigma}^{-1} [\mathbf{x} - \boldsymbol{\mu}]) = |\boldsymbol{\Sigma}|^{-1/2} g_n(u), \quad (1.1)$$

las distribuciones elípticas se caracterizan por un vector de parámetros de posición $\boldsymbol{\mu} \in \mathbb{R}^n$, una matriz de escala $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ la cual es simétrica y semidefinida positiva con $\text{rg}(\boldsymbol{\Sigma}) = n$. Además, $g : \mathbb{R} \rightarrow [0, \infty[$ es una función escalar continua, diferenciable y satisface la condición:

$$\int_0^{\infty} u^{(n/2)-1} g_n(u) du < \infty, \quad (1.2)$$

donde g es conocida como función generadora de densidades y u es la distancia de Mahalanobis. De esta manera, usaremos la notación $\mathbf{x} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, y en el caso de $\boldsymbol{\mu} = \mathbf{0}$ y $\boldsymbol{\Sigma} = \mathbb{I}_n$, la distribución es llamada como distribución esférica, ya que los contornos de este tipo de distribución son de forma esférica y lo denotaremos por $\mathbf{x} \sim S_n(g)$.

Cabe de señalar que, el vector aleatorio \mathbf{x} podría no tener una función de densidad cuando $\text{rg}(\boldsymbol{\Sigma}) = h < n$ (caso singular), por lo que la función de densidad no existe en todo el espacio \mathbb{R}^n , pero siempre es posible definir, una función de densidad en un espacio de dimensión menor o igual al rango de la matriz de escala. En cambio, si $\text{rg}(\boldsymbol{\Sigma}) = n$ (caso no singular), la función de densidad existe en relación a la medida de Lebesgue sobre todo el espacio \mathbb{R}^n . Una extensa teoría de distribuciones elípticas en el contexto del caso singular puede ser vista en Leiva & García (2001).

1.4 Algunas propiedades de las distribuciones elípticas

Una de las características principales de las distribuciones elípticas son su media y varianza, en el caso de que $\text{rg}(\boldsymbol{\Sigma}) = h < n$ sus expresiones son:

$$\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu} \quad (1.3)$$

$$\text{Cov}(\mathbf{x}) = \mathbf{k}\boldsymbol{\Sigma} \quad (1.4)$$

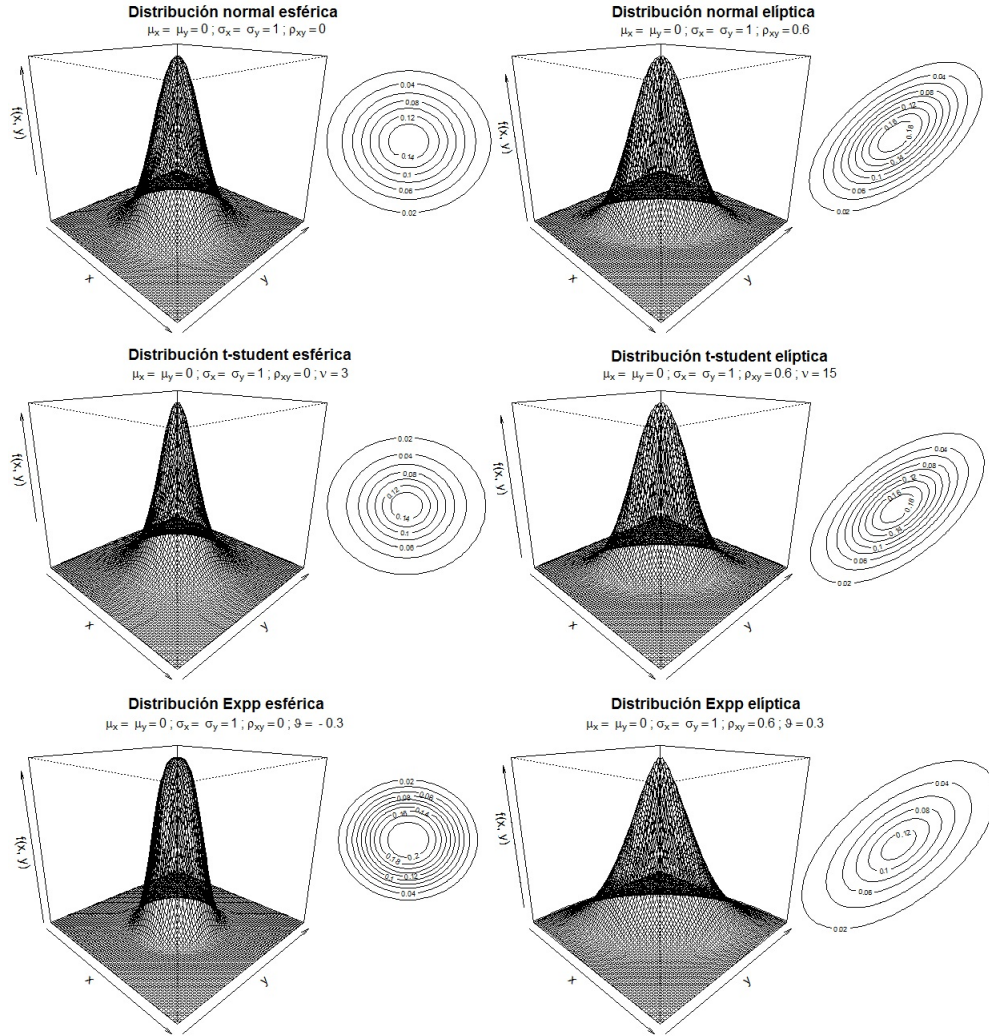


Figura 1.3: Distribuciones esféricas y elípticas: normal, exponencial potencia con parámetro de forma ϑ y t -Student con ν grados de libertad

,

en el cual $k > 0$ y es una constante dada por:

$$k = -2 \left. \frac{\partial g_h(u)}{\partial u} \right|_{u=0} \tag{1.5}$$

En particular para la distribución t -Student con ν grados de libertad y Exponencial potencia (Expp) con parámetro de forma ϑ tenemos los valores de \mathbf{k} son respectivamente:

$$\mathbf{k} = \frac{\nu}{\nu - 2}, \nu > 2 \quad (1.6)$$

$$\mathbf{k} = 2^{1+\vartheta} \frac{\Gamma[(n+2)(1+\vartheta)/2]}{n\Gamma[n(1+\vartheta)/2]} \quad (1.7)$$

donde Γ es la función gamma. En el caso de la distribución Exponencial potencia, cuando $\vartheta > 0$, la densidad posee colas pesadas y cuando $\vartheta < 0$ la densidad posee colas más livianas. En cambio, cuando $\vartheta = 0$ corresponde al caso normal, además, si $\mathbf{k} = 1$ también corresponde al caso normal, donde \mathbf{k} es un parámetro relativo a la curtosis de la distribución. En el caso de $\mathbf{x} \sim \text{EC}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ posea la siguiente partición:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \text{EC}_n \left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}, g \right) \quad (1.8)$$

donde $\mathbf{x}_1 \in \mathbb{R}^{n_1}$, $\mathbf{x}_2 \in \mathbb{R}^{n_2}$ y $n_1 + n_2 = n$. Entonces, se cumple:

$$\mathbf{x}_1 \sim \text{EC}_{n_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, g) \quad (1.9)$$

$$\mathbf{x}_2 \sim \text{EC}_{n_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, g) \quad (1.10)$$

Por lo que, de una distribución elíptica conjunta, sus distribuciones marginales también son elípticas. Además, se puede demostrar que:

$$\mathbf{x}_1 | \mathbf{x}_2 \sim \text{EC}_{n_1}(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{11|2}, g_{u(\mathbf{x}_2)}) \quad (1.11)$$

$$\mathbf{x}_2 | \mathbf{x}_1 \sim \text{EC}_{n_2}(\boldsymbol{\mu}_{2|1}, \boldsymbol{\Sigma}_{22|1}, g_{u(\mathbf{x}_1)}) \quad (1.12)$$

donde:

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} [\mathbf{x}_2 - \boldsymbol{\mu}_2]$$

$$\boldsymbol{\mu}_{2|1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} [\mathbf{x}_1 - \boldsymbol{\mu}_1]$$

$$\boldsymbol{\Sigma}_{11|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$$

$$\boldsymbol{\Sigma}_{22|1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}$$

$$g_{u(\mathbf{x}_1)} = [\mathbf{x}_1 - \boldsymbol{\mu}_1]^t \boldsymbol{\Sigma}_{11}^{-1} [\mathbf{x}_1 - \boldsymbol{\mu}_1]$$

$$g_{u(\mathbf{x}_2)} = [\mathbf{x}_2 - \boldsymbol{\mu}_2]^t \boldsymbol{\Sigma}_{22}^{-1} [\mathbf{x}_2 - \boldsymbol{\mu}_2]$$

Lo que nos muestra que si la distribución conjunta de dos vectores aleatorios es elíptica entonces, las distribuciones marginales y condicionales siguen siendo elípticas. Además, si se reescribe la distancia de Mahalanobis de la siguiente manera:

$$u = \boldsymbol{\varepsilon}^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^t \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\varepsilon} = \mathbf{P}^t \mathbf{P} = \|\mathbf{P}\|^2$$

en el cual $\boldsymbol{\varepsilon} = [\mathbf{x} - \boldsymbol{\mu}]$ y $\mathbf{P} = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\varepsilon}$, asimismo:

$$\mathbb{E} \left[\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\varepsilon} \right] = \mathbf{0} \quad \text{Cov} \left[\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\varepsilon} \right] = \mathbb{I}_n$$

por lo que $\mathbf{P} \sim S_n(g)$. Y según Lange *et al.* (1989) y Fang *et al.* (1990), el vector \mathbf{P} tiene una representación estocástica $\mathbf{P} \stackrel{d}{=} \xi \mathbf{v}^{(u)}$ para algún $\xi > 0$ que es independiente de $\mathbf{v}^{(u)}$, denotado por $\xi \perp \mathbf{v}^{(u)}$, de esta manera, si $\mathbf{P} \stackrel{d}{=} \xi \mathbf{v}^{(u)} \sim S_n(g)$ y $\mathbb{P}(\mathbf{P} = \mathbf{0})$, entonces:

$$\|\mathbf{P}\| \stackrel{d}{=} \xi, \quad \frac{\mathbf{P}}{\|\mathbf{P}\|} \stackrel{d}{=} \mathbf{v}^{(n)}$$

donde $\|\cdot\|$ representa la norma euclidea y al definir las matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{h \times h}$, entonces se cumplen las siguientes propiedades:

$$\mathbb{E} \left[\frac{\mathbf{P}}{\|\mathbf{P}\|} \right] = \mathbf{0}; \quad \text{Cov} \left(\frac{\mathbf{P}}{\|\mathbf{P}\|} \right) = \frac{1}{n} \mathbb{I}_n \quad (1.13)$$

$$\mathbb{E} [w \|\mathbf{P}\|^2] = n; \quad \mathbb{E} [w^2 \|\mathbf{P}\|^2] = 4d_{g_i}; \quad \mathbb{E} [w^2 \|\mathbf{P}\|^4] = 4f_{g_i} \quad (1.14)$$

$$\mathbb{E} \left[\frac{\mathbf{P}^t}{\|\mathbf{P}\|} \mathbf{A} \frac{\mathbf{P}}{\|\mathbf{P}\|} \middle| \|\mathbf{P}\| \right] = \frac{1}{h} \text{tr}(\mathbf{A}) \quad (1.15)$$

$$\mathbb{E} \left[\frac{\mathbf{P}^t}{\|\mathbf{P}\|} \mathbf{A} \frac{\mathbf{P}}{\|\mathbf{P}\|} \frac{\mathbf{P}^t}{\|\mathbf{P}\|} \mathbf{B} \frac{\mathbf{P}}{\|\mathbf{P}\|} \middle| \|\mathbf{P}\| \right] = \frac{[2\text{tr}(\mathbf{A}\mathbf{B}) + \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})]}{h(h+2)} \quad (1.16)$$

donde $w = -2\dot{g}(u)/g(u)$, $\dot{w} = \partial w / \partial u$, $\dot{g}(u) = \partial g(u) / \partial u$ y los valores w, \dot{w}, d_g y f_g son expresiones de gran utilidad para los cálculos de la matriz de información de Fisher penalizada del MMS. Por último, la siguiente tabla se describen dichos valores para cuatro distribuciones de la clase elíptica.

Cuadro 1.2: Valores de w , \dot{w} , d_g y f_g para algunas distribuciones elípticas multivariadas

Distribución	w	\dot{w}	d_g	f_g
Normal	1	0	$\frac{n}{4}$	$\frac{n(n+2)}{4}$
t -Student	$\frac{\nu+n}{\nu+u}$	$-\frac{\nu+n}{(\nu+u)^2}$	$\frac{n}{4} \left(\frac{\nu+n}{\nu+n+2} \right)$	$\frac{n(n+2)}{4} \left(\frac{\nu+n}{\nu+n+2} \right)$
Cauchy	$\frac{1+n}{1+u}$	$-\frac{1+n}{(1+u)^2}$	$\frac{n}{4} \left(\frac{1+n}{3+n} \right)$	$\frac{n(n+2)}{4} \left(\frac{1+n}{3+n} \right)$
Exponencial potencia	$\left(\frac{1}{1+\vartheta} \right) u \left(\frac{1}{1+\vartheta} - 1 \right)$	$-\frac{\vartheta}{(1+\vartheta)^2} u \left(\frac{1}{1+\vartheta} - 2 \right)$	$\frac{2}{(1+\vartheta)^3} \frac{\Gamma \left(\frac{(n+2)(1+\vartheta)}{2} + 2 \right)}{\Gamma \left(\frac{n(1+\vartheta)}{2} \right)}$	$\frac{n}{4} \left(n + \frac{2}{(1+\vartheta)} \right)$

ν =grados de libertad ϑ =parámetro de forma

HIPÓTESIS DE TRABAJO

Las hipótesis del presente trabajo son:

- I. Es posible derivar un proceso de estimación para los parámetros del modelo mixto semiparamétrico elíptico con errores aleatorios AR(1).

Nuestra propuesta se sustenta en la aplicación de los métodos iterativos de Backfitting y Score de Fisher.

- II. Es posible derivar un proceso de diagnóstico para los residuos del modelo mixto semiparamétrico elíptico con errores aleatorios AR(1).

Este punto, nuestra propuesta se sustenta en la construcción de la distancia de Cook generalizada.

- III. Es posible construir un test estadístico asintótico para probar la hipótesis de autocorrelación de tipo AR(1) bajo el modelo mixto semiparamétrico elíptico para datos longitudinales.

Y finalmente, nuestra propuesta se sustenta en la construcción de un test de Score.

OBJETIVOS

Objetivo general

Desarrollar un método de estimación de parámetros en un MMS con errores elípticos AR(1) para datos longitudinales basado en máxima verosimilitud penalizada, medidas de diagnóstico para la estimación de parámetros y test estadísticos para evaluar las hipótesis de presencia de autocorrelación y su homogeneidad a nivel intra-unidad.

Objetivos específicos

1. Encontrar las funciones score y la matriz de Información de Fisher esperada asociadas al MMS elíptico con errores aleatorios de tipo AR(1).
2. Encontrar los residuos marginales y condicionales como también construir las distancias de Cook generalizada y del valor estimado estandarizado asociados al MMS elíptico con errores aleatorios AR(1).
3. Desarrollar test asintóticos para probar las hipótesis de presencia de autocorrelación de tipo AR(1), la homogeneidad de la correlación a nivel intra-unidad y estudiar sus comportamientos a través de un análisis de simulación.
4. Aplicación de las metodologías desarrolladas a un conjunto de datos reales e implementar computacionalmente el proceso de estimación propuesto y las medidas de diagnóstico desarrolladas.

MMS ELÍPTICO PARA DATOS LONGITUDINALES

En este capítulo describiremos el modelo mixto semiparamétrico (MMS) para datos longitudinales y su representación matricial, especificando la forma de la matriz de covarianzas del error aleatorio, la cual incorpora una estructura de autocorrelación de primer orden en las observaciones dentro de cada una de las unidades de estudio (intra-unidad). Asimismo, presentaremos el supuesto distribucional del modelo que se basa en la clase de distribuciones elípticas.

2.1 Especificación del modelo

El MMS posee la siguiente relación entre las variables explicativas y respuesta:

$$y_{ij} = \mathbf{x}_{ij}^t \boldsymbol{\beta} + \mathbf{z}_{ij}^t \mathbf{b}_i + f(t_{ij}) + \varepsilon_{ij} , \quad (2.1)$$

donde $y_{ij} \in \mathbb{R}$ representa la j -ésima observación de la i -ésima unidad de estudio medida al tiempo $t_{ij} \in \mathbb{R}$. Además, $\mathbf{x}_{ij} \in \mathbb{R}^p$ y $\mathbf{z}_{ij} \in \mathbb{R}^q$ corresponden a

dos vectores de variables explicativas asociadas a la (i, j) -ésima observación, $\boldsymbol{\beta} \in \mathbb{R}^p$ es un vector de parámetros fijos y desconocidos, $\mathbf{b}_i \in \mathbb{R}^q$ es un vector de efectos aleatorios asociado a la i -ésima unidad de estudio, mientras que $f(\cdot)$ es una función suave que depende sólo de los tiempos t_{ij} y $\varepsilon_{ij} \in \mathbb{R}$ son los errores aleatorios asociados a la j -ésima observación de la i -ésima unidad de estudio, para $i = 1, \dots, n$ y $j = 1, \dots, n_i$.

El modelo (2.1) es una generalización de otros modelos. Por ejemplo, cuando la parte paramétrica del modelo no se considera, esto implica $\boldsymbol{\beta} = \mathbf{0}$, corresponde al modelo propuesto por Wang (1998), en cambio, cuando la parte no paramétrica no es considerada, entonces corresponde al modelo propuesto por Chi & Reinsel (1989), Osorio, Paula & Galea (2007) y Cao & Lin (2011). Cuando no se consideran los efectos aleatorios, esto es $\mathbf{b}_i = \mathbf{0}$ ($i = 1, \dots, n$), y este corresponde al caso de los modelos propuestos por Speckman (1988), Kim, Park & Kim (2002) y Paula, Medeiros & Vilca-Labra (2009). Asimismo, cuando no se considera tanto la parte no paramétrica como los efectos aleatorios, este caso corresponde a los modelos propuestos por Liu (2000) y Galea, Riquelme & Paula (2000). No obstante, Fung *et al.* (2002) también propone un MMS elíptico para datos longitudinales, pero la diferencia es que incorporaron una variable aleatoria adicional al modelo, que depende del tiempo de medición t_{ij} y con los efectos aleatorios \mathbf{b}_i de cada unidad de estudio, que miden en conjunto la estructura de correlación intra-unidad.

2.2 Representación matricial

Al considerar n_i observaciones de cada una de las n unidades de estudio, el modelo puede ser descrito a través de la siguiente estructura matricial:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{ij} \\ \vdots \\ y_{in_i} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{i1}^t \boldsymbol{\beta} \\ \mathbf{x}_{i2}^t \boldsymbol{\beta} \\ \vdots \\ \mathbf{x}_{ij}^t \boldsymbol{\beta} \\ \vdots \\ \mathbf{x}_{in_i}^t \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{z}_{i1}^t \mathbf{b}_1 \\ \mathbf{z}_{i2}^t \mathbf{b}_2 \\ \vdots \\ \mathbf{z}_{ij}^t \mathbf{b}_j \\ \vdots \\ \mathbf{z}_{in_i}^t \mathbf{b}_{n_i} \end{bmatrix} + \begin{bmatrix} f(t_{i1}) \\ f(t_{i2}) \\ \vdots \\ f(t_{ij}) \\ \vdots \\ f(t_{in_i}) \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{ij} \\ \vdots \\ \varepsilon_{in_i} \end{bmatrix} \quad (2.2)$$

2.2.1 Matriz de incidencia

Si consideramos un vector r -dimensional cuyos elementos son $t_1^0, t_2^0, \dots, t_r^0$ que corresponden a los r -valores diferentes de t_{ij} ($i = 1, \dots, n$) de tal manera que $t_1^0 < t_2^0 < \dots < t_r^0$ y según (2.2), es posible definir una matriz $\mathbf{N}_i \in \mathbb{R}^{n_i \times r}$ cuyo (i, j) -ésimo elemento es igual a $N_{ij}^k = I(t_{ij} = t_k^0)$ con $k = 1, \dots, r$, donde \mathbf{N}_i es conocida como matriz de incidencia y tiene la siguiente forma:

$$\begin{bmatrix} f(t_{i1}) \\ f(t_{i2}) \\ \vdots \\ f(t_{ij}) \\ \vdots \\ f(t_{in_i}) \end{bmatrix} = \begin{bmatrix} N_{11}^1 & N_{12}^2 & \cdots & N_{1r}^r \\ N_{21}^1 & N_{22}^2 & \cdots & N_{2r}^r \\ \vdots & \vdots & \vdots & \vdots \\ N_{i1}^1 & N_{i2}^2 & \cdots & N_{ir}^r \\ \vdots & \vdots & \vdots & \vdots \\ N_{n_i1}^1 & N_{n_i2}^2 & \cdots & N_{n_i r}^r \end{bmatrix} \begin{bmatrix} f(t_1^0) \\ f(t_2^0) \\ \vdots \\ f(t_j^0) \\ \vdots \\ f(t_r^0) \end{bmatrix} = \mathbf{N}_i \mathbf{f}, \quad (2.3)$$

donde $\mathbf{N}_i = [\mathbf{n}_{i1}, \mathbf{n}_{i2}, \dots, \mathbf{n}_{ir}]$ y \mathbf{n}_{ij} corresponde a la j -ésima columna de esta matriz. En adelante, denotaremos por $f_j^0 = f(t_j^0)$ al correspondiente parámetro asociado a la función f evaluada en el tiempo t_j^0 con $\mathbf{f} = [f_1^0, \dots, f_r^0]^t$.

2.2.2 Modelo mixto semiparamétrico

Además, tenemos que (2.2) puede ser escrito como:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{N}_i \mathbf{f} + \boldsymbol{\varepsilon}_i, \quad (2.4)$$

donde $\mathbf{y}_i \in \mathbb{R}^{n_i}$ es el vector de observaciones de la i -ésima unidad de estudio, $\mathbf{X}_i \in \mathbb{R}^{n_i \times p}$ es una matriz de diseño de efectos fijos asociada a la i -ésima unidad y $\mathbf{Z}_i \in \mathbb{R}^{n_i \times q}$ es una matriz de diseño de efectos aleatorios asociados a la i -ésima unidad. $\boldsymbol{\beta} \in \mathbb{R}^p$ es un vector de parámetros fijos y desconocidos, $\mathbf{b}_i \in \mathbb{R}^q$ es un vector de efectos aleatorios asociados a la i -ésima unidad y $\boldsymbol{\varepsilon}_i \in \mathbb{R}^{n_i}$ es un vector errores aleatorios asociados a la i -ésima unidad de estudio, donde $i = 1, \dots, n$. Al considerar el total de observaciones, el modelo (2.4) puede ser reescrito como:

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{b} + \mathbf{N} \mathbf{f} + \boldsymbol{\varepsilon}, \quad (2.5)$$

donde $\mathbf{y} \in \mathbb{R}^m$ es el vector aleatorio de respuestas observadas de las n unidades de estudio el cual poseen $m = \sum_{i=1}^n n_i$ observaciones. Además,

$\mathbf{X} \in \mathbb{R}^{m \times p}$ es una matriz de diseño de efectos fijos. $\mathbf{Z} \in \mathbb{R}^{m \times nq}$ y $\mathbf{b} \in \mathbb{R}^{nq}$ corresponden a una matriz de diseño de bloque diagonal y un vector de parámetros asociados a los efectos aleatorios de las n unidades de estudio respectivamente. $\mathbf{N} \in \mathbb{R}^{m \times r}$ es una matriz de incidencia y $\boldsymbol{\varepsilon} \in \mathbb{R}^m$ es un vector de errores aleatorios independientes de los efectos aleatorios de cada unidad de estudio, en donde: $\mathbf{y} = [\mathbf{y}_1^t, \dots, \mathbf{y}_n^t]^t$, $\mathbf{X} = [\mathbf{X}_1^t, \dots, \mathbf{X}_n^t]^t$, $\mathbf{N} = [\mathbf{N}_1^t, \dots, \mathbf{N}_n^t]^t$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$, $\mathbf{b} = [\mathbf{b}_1^t, \dots, \mathbf{b}_n^t]^t$, y $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1^t, \dots, \boldsymbol{\varepsilon}_n^t]^t$. La expresión (2.5) es útil para simplificar la notación utilizada en las expresiones las funciones de Score y de la matriz de información de Fisher esperada presentadas en mayor detalle en el capítulo 3.

2.3 Estructura de autocorrelación de primer orden

En el caso de tener observaciones que fueron medidas de manera secuencial a través del tiempo, entonces se podría considerar incorporar una estructura autorregresiva en las observaciones a nivel intra-unidad. Suponiendo que los errores aleatorios del modelo descrito en (2.4), poseen una estructura AR(1) dentro de la i -ésima unidad de estudio, entonces:

$$\varepsilon_{i1} = \epsilon_{i1} \quad \varepsilon_{ij} = \rho \varepsilon_{i(j-1)} + \epsilon_{ij} \quad j = 2, \dots, n_i, \quad (2.6)$$

donde ρ corresponde al coeficiente de autocorrelación en la i -ésima unidad de estudio, note que este coeficiente es el mismo para todas las unidades y es conocido como correlación de tipo homogénea. Pero, en el caso de poseer una estructura AR(1) de tipo:

$$\varepsilon_{i1} = \epsilon_{i1} \quad \varepsilon_{ij} = \rho_i \varepsilon_{i(j-1)} + \epsilon_{ij} \quad j = 2, \dots, n_i \quad (2.7)$$

donde ρ_i corresponde al coeficiente de autocorrelación dentro de la i -ésima unidad de estudio, y en este caso, el coeficiente de autocorrelación es específico para cada unidad de estudio, y es conocido como correlación de tipo heterogénea.

Para la incorporación de la estructura autorregresiva de primer orden en el MMS, se asume que el vector de errores aleatorios $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1} \ \varepsilon_{i2} \ \dots \ \varepsilon_{ij} \ \dots \ \varepsilon_{in_i}]^t$ tiene $\mathbb{E}[\boldsymbol{\varepsilon}_i] = \mathbf{0}$ y $\text{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Upsilon}_{n_i}$, donde $\boldsymbol{\Upsilon}_{n_i} = \text{diag}[(1 - \rho^2)^{-1}, 1, 1, \dots, 1]$ con $\boldsymbol{\Upsilon}_{n_i} \in \mathbb{R}^{n_i \times n_i}$. Correspondientemente, se tiene que $\mathbb{E}[\boldsymbol{\varepsilon}_i] = \mathbf{0}$ y $\text{Var}(\boldsymbol{\varepsilon}_i) = \mathbf{Q}_i$, donde $\mathbf{Q}_i = \phi \mathbf{C}_i^{-1} \mathbf{C}_i^{-t}$,

$\mathbf{C}_i^{-t} = [\mathbf{C}_i^{-1}]^t$, $\phi > 0$ es una constante desconocida y la matriz $\mathbf{C}_i \in \mathbb{R}^{n_i \times n_i}$ esta dada por:

$$\mathbf{C}_i = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & -\rho & 1 & 0 \\ 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix} \quad (2.8)$$

esta matriz incorpora la correlación tipo AR(1) de tipo homogénea a nivel intra-unidad descrita en (2.6), para el caso (2.7) se ve con mayor detalle en el capítulo 5. Este tipo de estructura de covarianzas con errores normales fue utilizado por Chi & Reinsel (1989), Chi & Reinsel (1991), Lin & Wei (2007), Paula, Medeiros & Vilca-Labra (2009) y Cao & Lin (2011; 2012). Lin & Wei (2004) generalizan la estructura de autocorrelación incorporando al modelo errores con estructura AR(p).

2.4 Modelo bajo distribución elíptica

Comúnmente la distribución normal ha sido utilizada para modelar muchos fenómenos de interés pero también han sido criticado este supuesto por proporcionar estimaciones de máxima verosimilitud de los parámetros que son sensibles a la presencia de outliers. Con el fin de encontrar un proceso de estimación más robusto en el sentido de modelar observaciones atípicas.

En adelante, se asume que los errores del MMS poseen una distribución de contornos elípticos, debido a que esta clase incluye distribuciones con niveles de Kurtosis mucho más flexibles. Es por esto que, en el presente trabajo se consideró la siguiente estructura distribucional:

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{b}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} \sim EC_{2n_i+q} \left(\begin{bmatrix} \boldsymbol{\mu}_i \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_i & \mathbf{Z}_i \mathbf{D} & \mathbf{Q}_i \\ \mathbf{D} \mathbf{Z}_i^t & \mathbf{D} & \mathbf{0} \\ \mathbf{Q}_i & \mathbf{0} & \mathbf{Q}_i \end{bmatrix}, g \right) \quad (2.9)$$

El MMS se basa en la especificación del modelo marginal de las observaciones de la expresión (2.9), el cual es: $\mathbf{y}_i \sim EC_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, g)$, donde el vector de

Cuadro 2.1: Funciones de densidad para las distribuciones normal, t -Student con ν grados de libertad y exponencial potencia con parámetro de forma ϑ

Distribución	Función de densidad
Normal	$f_{\mathbf{y}}(\mathbf{y}_i) = \frac{1}{(2\pi)^{\frac{n_i}{2}} \boldsymbol{\Sigma}_i ^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} u_i \right\}$
t -Student	$f_{\mathbf{y}}(\mathbf{y}_i) = \frac{\Gamma \left(\frac{n_i + \nu}{2} \right) \boldsymbol{\Sigma}_i ^{-\frac{1}{2}}}{\pi^{\frac{n_i}{2}} \Gamma \left(\frac{\nu}{2} \right)} \left(1 + \frac{u_i}{\nu} \right)^{-\frac{n_i + \nu}{2}}$
Exponencial potencia	$f_{\mathbf{y}}(\mathbf{y}_i) = \frac{n_i \Gamma \left(\frac{n_i}{2} \right) \boldsymbol{\Sigma}_i ^{-\frac{1}{2}}}{\pi^{\frac{n_i}{2}} \Gamma(\vartheta^*) 2^{\vartheta^*}} \exp \left\{ -\frac{1}{2} u_i^{\frac{1}{1+\vartheta}} \right\}$

donde $\vartheta^* = 1 + \frac{n_i}{2} (1 + \vartheta)$

parámetros de localización es $\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{N}_i \mathbf{f}$ y matriz de escala $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t + \mathbf{Q}_i$. Además, la función de densidad de las observaciones en la i -ésima unidad de estudio está dada por:

$$f_{\mathbf{y}}(\mathbf{y}_i, |\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, g) = |\boldsymbol{\Sigma}_i|^{-1/2} g(u_i) \quad (2.10)$$

donde $u_i = [\mathbf{y}_i - \boldsymbol{\mu}_i]^t \boldsymbol{\Sigma}_i^{-1} [\mathbf{y}_i - \boldsymbol{\mu}_i]$ es la distancia de Mahalanobis de las observaciones de la i -ésima unidad de estudio ($i = 1, \dots, n$), g es una función generadora de densidades que satisface la condición (1.2) y en la Tabla 2.1 se describen las funciones de densidad para tres distribuciones elípticas.

En el contexto de modelos mixtos, es sabido que un modelo también puede ser formulado desde una perspectiva jerárquica lo cual implicaría suponer que $\mathbf{y}_i | \mathbf{b}_i$, \mathbf{b}_i y $\boldsymbol{\varepsilon}_i$ posean cada una por separado una distribución elíptica. Sin embargo, la desventaja de utilizar esa representación es que la distribución conjunta -de dos o tres vectores aleatorios- puede no pertenecer

a la clase elíptica, lo cual dificulta la obtención del modelo marginal de las observaciones. Una forma de enfrentar esta dificultad analítica es suponer inicialmente la distribución conjunta (2.9), por lo que la distribución marginal de las observaciones también sería de la clase elíptica. No obstante, se tiene el inconveniente de que los efectos aleatorios no tienen la misma interpretación de efectos fijos como en el tipo jerárquico, pero en ambas formulaciones del modelo el vector de medias es el mismo.

2.5 Estructura de covarianza del MMS

Puesto que la estructura de covarianza del MMS incorpora la variabilidad debido a los efectos y errores aleatorios, además de depender de la autocorrelación de primer orden a nivel intra-unidad, y según la estructura distribucional descrita en (2.9), se tiene que:

$$\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t + \mathbf{Q}_i, \quad (2.11)$$

$$\mathbf{Q}_i = \phi \mathbf{C}_i^{-1} \mathbf{C}_i^{-t}, \quad (2.12)$$

Y siguiendo con la notación del MMS descrito en la expresión (2.5) y suponiendo que las matrices \mathbf{D} y \mathbf{Q}_i ($i = 1, \dots, n$) son definidas no negativas y la matriz de covarianzas $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$, esta asociada a las n unidades de estudio, y sean $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_n)$ y $m = \sum_{i=1}^n n_i$, $\mathbf{Q} = \text{diag}(\mathbf{Q}_1, \dots, \mathbf{Q}_n)$, $\mathbf{C} = \text{diag}(\mathbf{C}_1, \dots, \mathbf{C}_n)$ y $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$, con $\mathbf{Z} \in \mathbb{R}^{m \times nq}$, $\mathbf{Q} \in \mathbb{R}^{m \times m}$ y $\mathbf{C} \in \mathbb{R}^{m \times m}$, donde \mathbf{C}_i fue descrita en (2.8).

De lo anterior, se puede demostrar que la estructura de covarianza del MMS elíptico para las n observaciones esta dada por:

$$\begin{aligned} \boldsymbol{\Sigma} &= \begin{bmatrix} \mathbf{Z}_1 \mathbf{D} \mathbf{Z}_1^t + \mathbf{Q}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \mathbf{D} \mathbf{Z}_2^t + \mathbf{Q}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_n \mathbf{D} \mathbf{Z}_n^t + \mathbf{Q}_n \end{bmatrix} \\ &= \mathbf{Z} [\mathbb{I}_n \otimes \mathbf{D}] \mathbf{Z}^t + \mathbf{Q} \end{aligned} \quad (2.13)$$

donde:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Q}_n \end{bmatrix} = \phi [\mathbf{C}^t \mathbf{C}]^{-1} \quad (2.14)$$

De esta manera, para las m observaciones de las n unidades de estudio podemos escribir que $\mathbf{y} \sim EC_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, donde $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Nf}$ y $\boldsymbol{\Sigma} = \mathbf{Z} [\mathbb{I}_n \otimes \mathbf{D}] \mathbf{Z}^t + \mathbf{Q}$. Para ver otros tipos de estructuras de la matriz de covarianzas en los errores, pueden ser vistas en Verbeke (2001), Diggle (2002), Hedeker & Gibbons (2006) y Muller & Stewart (2006).

2.6 Comentarios del capítulo

En el presente capítulo, fue presentado el MMS elíptico para datos longitudinales, describiendo el modelo y la forma de la matriz de covarianzas, la cual, incorpora una estructura autorregresiva de primer orden a través del tiempo de seguimiento de cada unidad de estudio, como también, el supuesto distribucional del modelo basado en la clase de distribuciones elípticas, donde fue descrita la distribución marginal del MMS elíptico.

En el siguiente Capítulo, se discute el problema de estimación de los parámetros asociados al MMS elíptico, donde se propone un método de estimación basado en un criterio de verosimilitud que incorpora una componente no paramétrica.

ESTIMACIÓN DE MÁXIMA VEROSIMILITUD PENALIZADA

En el presente capítulo se aborda el problema de estimación de los parámetros asociados al MMS elíptico para datos longitudinales con errores AR(1), para lo cual, se utiliza la función de log-verosimilitud, y así, a ésta se incorpora una función de penalización para la estimación de la componente no paramétrica asociado al modelo. Asimismo, se presenta las funciones de score y matriz de Información de Fisher. También, se presenta un método de estimación de los parámetros asociados al modelo basado en el score de Fisher y el algoritmo Backfitting. Por último, se presentan expresiones para los errores estándar de los parámetros estimados. En adelante, todos los resultados de algebra matricial, productos de Kronecker, diferenciación de matrices y propiedades en matrices particionadas pueden ser vistas con mayor detalle en Wong (1985), Vargas (1996), Lu & Shiou (2002), Khuri (2003), Kollo & Von Rosen (2005), Magnus & Neudecker (2007), Harville (2008) y Seber (2008).

3.1 Criterios de verosimilitud penalizada

En adelante denotaremos al vector de parámetros del MMS elíptico como $\boldsymbol{\theta} = [\boldsymbol{\beta}^t \ \mathbf{f}^t \ \boldsymbol{\tau}^t]^t$ donde los vectores $\boldsymbol{\beta} \in \Theta_{\boldsymbol{\beta}}$ y $\mathbf{f} \in \Theta_{\mathbf{f}}$ corresponden a los parámetros de los efectos fijos del MMS que están asociados a los espacios paramétricos de la parte paramétrica y función no paramétrica respectivamente. Respecto de los parámetros de componentes de varianza de los efectos aleatorios y de la estructura de autocorrelación a nivel intra-unidad, pueden ser descritos a través del vector $\boldsymbol{\tau} = [\boldsymbol{\delta}^t \ \boldsymbol{\eta}^t]^t$ tal que los espacios paramétricos correspondientes son $\boldsymbol{\tau} \in \Theta_{\boldsymbol{\tau}}, \boldsymbol{\delta} \in \Theta_{\boldsymbol{\delta}}, \boldsymbol{\eta} \in \Theta_{\boldsymbol{\eta}}$ y $\Theta_{\boldsymbol{\tau}} \equiv \Theta_{\boldsymbol{\delta}} \cup \Theta_{\boldsymbol{\eta}}$. Por lo que la estructura de varianza de la i -ésima unidad de estudio está parametrizada como $\boldsymbol{\Sigma}_i(\boldsymbol{\tau}) = \mathbf{Z}_i \mathbf{D}(\boldsymbol{\delta}) \mathbf{Z}_i^t + \mathbf{Q}_i(\boldsymbol{\eta})$ donde $\boldsymbol{\delta} = [\boldsymbol{\delta}_1 \ \cdots \ \boldsymbol{\delta}_s]^t$ es compuesta por los elementos de la matriz triangular superior¹ de la matriz \mathbf{D} con $s = q(q+1)/2$ y $\boldsymbol{\eta} = [\phi \ \rho]^t$, por ende, el espacio paramétrico asociado al MMS elíptico está dado por:

$$\Theta = \left(\boldsymbol{\theta} \in \mathbb{R}^{p^*} \mid \boldsymbol{\beta} \in \Theta_{\boldsymbol{\beta}} \subseteq \mathbb{R}^p, \mathbf{f} \in \Theta_{\mathbf{f}} \subseteq \mathbb{R}^r, \boldsymbol{\tau} \in \Theta_{\boldsymbol{\tau}} \subseteq \mathbb{R}^{s+2} \right)$$

, donde $p^* = p + r + s + 2$. Suponiendo que se obtiene una muestra aleatoria de n grupos mutuamente independientes entre sí, entonces, tenemos que la función de log-verosimilitud asociada al MMS elíptico es:

$$\begin{aligned} l(\boldsymbol{\theta} ; \mathbf{y}_1, \dots, \mathbf{y}_n) &= \ln f_{\mathbf{y}}(\mathbf{y}_1, \dots, \mathbf{y}_n ; \boldsymbol{\theta}) = \ln \left[\prod_{i=1}^n f_{\mathbf{y}}(\mathbf{y}_i) \right] \\ &= \sum_{i=1}^n \ln \{ |\boldsymbol{\Sigma}_i|^{-1/2} g(u_i) \} = \sum_{i=1}^n l_i(\boldsymbol{\theta} | \mathbf{y}_i) \end{aligned} \quad (3.1)$$

Green (1987) muestra un ejemplo en el que una maximización directa de $l(\mathbf{y}_1, \dots, \mathbf{y}_n ; \boldsymbol{\theta})$ sin imponer restricciones sobre la función f puede causar un sobre-ajuste² y la no identificación del parámetro $\boldsymbol{\beta}$. Sin embargo, existe un procedimiento que permite corregir este problema y se basa en el método de la función penalizada.

¹Una matriz \mathbf{A} se dice ser triangular superior si su elemento $a_{ij} = 0$ para $i > j$ donde los índices i y j corresponden a la fila y columna de \mathbf{A} respectivamente

²También conocido como interpolación

Varios autores han incorporado funciones de penalización en sus análisis, como por ejemplo, Silverman (1982) propone un estimador máximo verosímil penalizado de la función de densidad, Klonia (1982) estudia la consistencia de los dos estimadores no paramétricos máximo verosímil penalizado de la función de densidad, Hastie *et al.* (2000) incorporan una función de penalización pero en el contexto de análisis discriminante, Eggermont & LaRiccia (2009) muestran algunas funciones de penalización utilizados en varios tipos de modelos no paramétricos y series temporales.

El método de penalización consiste en incorporar una función sobre la función de log-verosimilitud de la forma: $l_p(\boldsymbol{\theta}, \alpha) = l(\boldsymbol{\theta}) + \alpha^* \mathcal{P}(f)$, donde $\mathcal{P}(f)$ denota la función de penalidad sobre la función f y α^* es una constante que depende del parámetro de suavizamiento $\alpha > 0$ que controla el equilibrio entre la bondad del ajuste y la suavidad de la función estimada, para más detalles ver Green & Silverman (1994), en adelante asumiremos α^* como un parámetro fijo.

3.1.1 Función de penalización

En el contexto no paramétrico, si asumimos que la función de penalización posee la siguiente forma:

$$\mathcal{P}(f) = \alpha^* \int_b^a [f^{(k)}(t)]^2 dt, \quad (3.2)$$

donde $f^{(k)}(t) = \frac{\partial^k}{\partial t^k} f(t)$, $t_j^0 \in [a, b]$ y f pertenece a un espacio de funciones de Sobolev dado por:

$$\mathcal{Z}_2^k[a, b] = \{ f : f^{(k)} \in L^2[a, b] \text{ y } f^{(1)}, \dots, f^{(k-1)} \text{ son absolutamente continuas} \}$$

En el caso particular $k = 2$, la estimación de f conduce a un suavizamiento Spline cúbico con nodos en los puntos $t_1^0, t_2^0, \dots, t_r^0$ y es posible mostrar que la curva estimada $\hat{f}(t_k^0)$ es un polinomio cúbico en cada intervalo $[t_{k_j}^0, t_{k_{j+1}}^0]$ evaluada en cada punto t_k^0 de la curva y sus primeras derivadas son continuas, aunque puede ser discontinua en su tercera derivada y en los intervalos $]-\infty, t_1^0]$ y $[t_r^0, \infty[$ tienen segunda derivada igual a cero, de modo que la estimación es lineal fuera del dominio de los datos.

Green & Silverman (1994) demuestran que la función de penalización puede escribirse como una forma cuadrática dada por:

$$\mathcal{P}(f) = \alpha^* \int_b^a [f^{(2)}(t)]^2 dt = \alpha^* \mathbf{f}^t \mathbf{K} \mathbf{f} \quad (3.3)$$

donde el vector \mathbf{f} es definido como en (2.3) y la matriz de penalización $\mathbf{K} \in \mathbb{R}^{r \times r}$ es definida no negativa que depende solamente de los nodos t_k^0 ($k = 1, \dots, r$), por lo que la función (3.3) mide la ondulación de la función f . La matriz de penalización posee una estructura dada por $\mathbf{K} = \mathbf{A} \mathbf{B}^{-1} \mathbf{A}^t$ donde $\mathbf{A} \in \mathbb{R}^{r \times (r-2)}$ cuyo elemento a_{ij} ($i = 1, \dots, r; j = 2, \dots, r-1$) está dado por:

$$a_{ij} = \begin{cases} a_{(j-1)j} = h_{j-1}^{-1}, & a_{jj} = -h_{j-1}^{-1} - h_j^{-1}, & a_{(j+1)j} = h_j^{-1} & \text{si } |i-j| < 2 \\ 0, & & & \text{si } |i-j| \geq 2 \end{cases}$$

La matriz \mathbf{A} posee una notación no estándar en sus columnas, la cual denota a la primera columna como $j = 2$, por ende, el primer elemento superior izquierdo corresponde a a_{12} .

En cambio, la matriz simétrica $\mathbf{B} \in \mathbb{R}^{(r-2) \times (r-2)}$ cuyo elemento b_{ij} , con $(i, j = 2, \dots, r-1)$, está dado por:

$$b_{ij} = \begin{cases} b_{ii} = \frac{h_{i-1} + h_i}{3}, & b_{(i+1)i} = b_{i(i+1)} = \frac{h_i}{6} & \text{si } |i-j| < 2 \quad (i = 2, \dots, r-1) \\ 0 & & \text{si } |i-j| \geq 2 \quad (i = 2, \dots, r-2) \end{cases}$$

donde $h_i = t_{i+1}^0 - t_i^0$, recordando que la función f es un parámetro infinito dimensional, si se estima el vector de parámetros $\boldsymbol{\theta}$ bajo el criterio de máxima verosimilitud penalizada se obtiene un estimador Spline cúbico natural para f en los puntos t_i^0 , por lo cual, el proceso de estimación se reduce a un problema finito dimensional. Para ver una demostración más formal de (3.3) se puede ver Green & Silverman (1994), específicamente teorema 2.1.

3.1.2 Función de log-verosimilitud penalizada

De acuerdo con la función de log-verosimilitud definida en (3.1), la función de penalización (3.3) y considerando $\alpha^* = -\alpha/2$, tenemos que la función de log-verosimilitud penalizada asociada al MMS elíptico esta dado por:

$$\begin{aligned} l_p(\boldsymbol{\theta}, \alpha; \mathbf{y}_1, \dots, \mathbf{y}_n) &= l(\boldsymbol{\theta}; \mathbf{y}_1, \dots, \mathbf{y}_n) - \frac{\alpha}{2} \mathbf{f}^t \mathbf{K} \mathbf{f} \\ &= \sum_{i=1}^n \left[l_i(\boldsymbol{\theta}; \mathbf{y}_i) - \frac{\alpha}{2n} \mathbf{f}^t \mathbf{K} \mathbf{f} \right] = \sum_{i=1}^n l_{p_i}(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) \end{aligned} \quad (3.4)$$

Note que para valores grandes de α , significa que se agrega una mayor ponderación a la suavización de la función no paramétrica, lo que implica, que se obtendrán curvas mucho más suaves que onduladas. En cambio, para valores pequeños de α significa que se agrega una mayor ponderación a la medida de la calidad del ajuste, lo que implica, que se obtendrán curvas mucho más onduladas (llegando al caso de interpolación) y menos suaves en su comportamiento. El proceso de estimación y el análisis inferencial bajo el MMS elíptico estará basado en la expresión (3.4) y en la siguiente sección, se deriva la función de score penalizada asociada al vector de parámetros $\boldsymbol{\theta}$.

3.2 Funciones de score

La función se score penalizada está definida como a la derivada de la función de log-verosimilitud penalizada respecto del parámetro $\boldsymbol{\theta}$, lo cual es:

$$\mathbb{S}_p(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} l_{p_i}(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) = \sum_{i=1}^n \mathbb{S}_{p_i}(\boldsymbol{\theta}), \quad (3.5)$$

donde:

$$\mathbb{S}_{p_i}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbb{S}_{p_i}(\boldsymbol{\beta}) \\ \mathbb{S}_{p_i}(\mathbf{f}) \\ \mathbb{S}_{p_i}(\boldsymbol{\tau}) \end{bmatrix} = \begin{bmatrix} \partial l_{p_i}(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) / \partial \boldsymbol{\beta} \\ \partial l_{p_i}(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) / \partial \mathbf{f} \\ \partial l_{p_i}(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) / \partial \boldsymbol{\tau} \end{bmatrix}$$

Además, note que función se score penalizada es la suma entre la función de Score no penalizada y la derivada de la función de penalización, por lo que

la función de score no penalizada puede ser escrita como:

$$\begin{aligned}\mathbb{S}_i(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} l_i(\boldsymbol{\theta}; \mathbf{y}_i) = \frac{\partial}{\partial \boldsymbol{\theta}} \left(-\frac{1}{2} \ln |\boldsymbol{\Sigma}_i| \right) + \frac{\partial}{\partial \boldsymbol{\theta}} \ln g(u_i) \\ &= -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \boldsymbol{\theta}} \right) + \frac{1}{g(u_i)} \frac{\partial g(u_i)}{\partial u_i} \frac{\partial u_i}{\partial \boldsymbol{\theta}},\end{aligned}$$

y en adelante utilizaremos la siguiente notación:

$$\begin{aligned}\dot{g}(u_i) &= \frac{\partial g(u_i)}{\partial u_i} & \mathcal{W}_g(u_i) &= \frac{\dot{g}(u_i)}{g(u_i)} & \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\theta}} &= \frac{\partial \boldsymbol{\Sigma}_i}{\partial \boldsymbol{\theta}} \\ w_i &= -2\mathcal{W}_g(u_i) & \boldsymbol{\varepsilon}_i &= \mathbf{y}_i - \boldsymbol{\mu}_i\end{aligned}$$

Entonces, tenemos que la función de Score no penalizada asociada a la i -ésima unidad puede ser escrita como:

$$\mathbb{S}_i(\boldsymbol{\theta}) = -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\theta}} \right) + \mathcal{W}_g(u_i) \frac{\partial u_i}{\partial \boldsymbol{\theta}} \quad (3.6)$$

De esta manera, la función de score penalizada asociada a la i -ésima unidad de estudio está dada por:

$$\begin{aligned}\mathbb{S}_{p_i}(\boldsymbol{\theta}) &= \mathbb{S}_i(\boldsymbol{\theta}) + \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\alpha}{2n} \mathbf{f}^t \mathbf{K} \mathbf{f} \right) \\ &= -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\theta}} \right) + \mathcal{W}_g(u_i) \frac{\partial u_i}{\partial \boldsymbol{\theta}} - \frac{\alpha}{2n} \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{f}^t \mathbf{K} \mathbf{f}\end{aligned} \quad (3.7)$$

Asimismo, note que la derivada sobre la función de penalización es distinta de cero sólo en el caso de $\boldsymbol{\theta} = \mathbf{f}$. A continuación se derivan las funciones de score penalizada para los efectos fijos del modelo, tanto la parte paramétrica como la no paramétrica, así también para los componentes de varianza. Por último, las demostraciones de las funciones de score están con mayor detalle en la primera sección del Apéndice II.

3.2.1 Score componente paramétrica

La función de score penalizada de la parte paramétrica para la i -ésima unidad de estudio está dado por:

$$\mathbb{S}_{p_i}(\boldsymbol{\beta}) = w_i \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \quad (3.8)$$

Y la función de score penalizada de las n unidades de estudio está dado por:

$$\mathbb{S}_p(\boldsymbol{\beta}) = \sum_{i=1}^n w_i \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i = \mathbf{X}^t \mathbf{W} \boldsymbol{\varepsilon} \quad (3.9)$$

donde $\mathbf{W} = \text{diag}(w_1 \boldsymbol{\Sigma}_1^{-1}, \dots, w_n \boldsymbol{\Sigma}_n^{-1})$, $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1^t \dots \boldsymbol{\varepsilon}_n^t]^t$, $\mathbf{X} = [\mathbf{X}_1^t \dots \mathbf{X}_n^t]^t$, $\boldsymbol{\varepsilon} = \mathbf{y} - \boldsymbol{\mu}$ y $m = \sum_{i=1}^n n_i$ con $\mathbf{X} \in \mathbb{R}^{m \times p}$, $\mathbf{W} \in \mathbb{R}^{m \times m}$ y $\boldsymbol{\varepsilon} \in \mathbb{R}^m$.

3.2.2 Score componente no paramétrica

La función de score penalizada de la parte no paramétrica de la i -ésima unidad de estudio está dado por:

$$\mathbb{S}_{p_i}(\mathbf{f}) = w_i \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \frac{\alpha}{n} \mathbf{K} \mathbf{f} \quad (3.10)$$

Y la función de score penalizada de las n unidades de estudio está dado por:

$$\mathbb{S}_p(\mathbf{f}) = \sum_{i=1}^n \left(w_i \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \frac{\alpha}{n} \mathbf{K} \mathbf{f} \right) = \mathbf{N}^t \mathbf{W} \boldsymbol{\varepsilon} - \alpha \mathbf{K} \mathbf{f} \quad (3.11)$$

donde $\mathbf{N} = [\mathbf{N}_1^t \dots \mathbf{N}_n^t]^t$, $\mathbf{N} \in \mathbb{R}^{m \times r}$, $\mathbf{W} \in \mathbb{R}^{m \times m}$ y $\boldsymbol{\varepsilon} \in \mathbb{R}^m$.

3.2.3 Score componentes de varianza

La función de score para los parámetros asociados a los componentes de varianza para la i -ésima unidad de estudio está dado por:

$$\mathbb{S}_{p_i}(\boldsymbol{\tau}) = -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \right) - \mathcal{W}_g(u_i) \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \quad (3.12)$$

Y la función de score penalizada de las n unidades de estudio está dado por:

$$\mathbb{S}_p(\boldsymbol{\tau}) = -\frac{1}{2} \sum_{i=1}^n \left[\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \right) - w_i \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right] \quad (3.13)$$

donde $\boldsymbol{\tau} = [\delta_1 \dots \delta_s \phi \rho]^t = [\tau_1 \dots \tau_s \tau_{s+1} \tau_{s+2}]^t$ y $\dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}}$ posee la siguiente forma según el parámetro de la componente de varianza:

$$\dot{\boldsymbol{\Sigma}}_{i\tau_k} = \begin{cases} \mathbf{Z}_i \dot{\mathbf{D}}_k \mathbf{Z}_i^t & \tau_k = \delta_k \\ \mathbf{C}_i^{-1} \dot{\mathbf{C}}_i^{-t} & \tau_k = \phi \\ -\phi \mathbf{C}_i^{-1} \left(\dot{\mathbf{C}}_i \mathbf{C}_i^{-1} + \mathbf{C}_i^{-t} \dot{\mathbf{C}}_i^t \right) \mathbf{C}_i^{-t} & \tau_k = \rho \end{cases}$$

La matriz $\dot{\mathbf{D}}_k \in \mathbb{R}^{q \times q}$ depende de una matriz $\mathbf{E}_{lj} \in \mathbb{R}^{q \times q}$ cuyo (l, j) -ésimo elemento es igual a $e_{lj} = \mathbf{I}(\delta_k = d_{lj})$, donde $k = 1, \dots, s$ y $\dot{\mathbf{D}}_k$ está dado por:

$$\dot{\mathbf{D}}_k = \begin{cases} \mathbf{E}_{lj} + \mathbf{E}_{lj}^t & \forall l \neq j \\ \mathbf{E}_{jj} & \forall l = j = 1, \dots, q \end{cases}$$

Y la matriz $\dot{\mathbf{C}}_i \in \mathbb{R}^{n_i \times n_i}$ posee la siguiente forma:

$$\dot{\mathbf{C}}_i = \begin{bmatrix} \frac{-\rho}{\sqrt{1-\rho^2}} & 0 & 0 & \dots & 0 & 0 \\ -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 0 \end{bmatrix}$$

3.3 Matriz de Información de Fisher penalizada

Al momento de realizar inferencias sobre los estimadores de los parámetros del modelo estadístico asociado, es razonable utilizar como una medida de información la esperanza de la matriz Hessiana³, también conocida como matriz de información de Fisher. Esta matriz esta definida por:

$$\mathcal{I}_p(\boldsymbol{\theta}) = \sum_{i=1}^n -\mathbb{E}[\mathbb{H}_i(\boldsymbol{\theta})] = \sum_{i=1}^n \mathcal{I}_{p_i}(\boldsymbol{\theta}) \quad (3.14)$$

donde $\mathbb{H}_i(\boldsymbol{\theta})$ es la matriz Hessiana y $\mathcal{I}_{p_i}(\boldsymbol{\theta})$ es la matriz de información de Fisher penalizada asociada a la i -ésima unidad de estudio dada por:

$$\mathcal{I}_{p_i}(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{I}_{p_i}(\boldsymbol{\beta}, \boldsymbol{\beta}) & \mathcal{I}_{p_i}(\boldsymbol{\beta}, \mathbf{f}) & \mathcal{I}_{p_i}(\boldsymbol{\beta}, \boldsymbol{\tau}) \\ \mathcal{I}_{p_i}(\mathbf{f}, \boldsymbol{\beta}) & \mathcal{I}_{p_i}(\mathbf{f}, \mathbf{f}) & \mathcal{I}_{p_i}(\mathbf{f}, \boldsymbol{\tau}) \\ \mathcal{I}_{p_i}(\boldsymbol{\tau}, \boldsymbol{\beta}) & \mathcal{I}_{p_i}(\boldsymbol{\tau}, \mathbf{f}) & \mathcal{I}_{p_i}(\boldsymbol{\tau}, \boldsymbol{\tau}) \end{bmatrix} \quad (3.15)$$

Sin embargo, evaluar directamente la esperanza de la matriz Hessiana penalizada dificulta los cálculos para la obtención de $\mathcal{I}_p(\boldsymbol{\theta})$.

³En la segunda sección del Apéndice II se describen las expresiones de la matriz Hessiana para el MMS elíptico

Pero al reescribir la matriz de información de Fisher penalizada de la i -ésima unidad de estudio de la siguiente forma:

$$\begin{aligned}\mathcal{I}_{p_i}(\boldsymbol{\theta}) &= -\mathbb{E}[\mathbb{H}_i(\boldsymbol{\theta}, \boldsymbol{\theta})] = \mathbb{E}\left[\left(-\frac{\partial^2 l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t}\right)\right] = \mathbb{E}\left[\frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta}} \frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta}}\right] \\ &= \mathbb{E}\left\{\mathbb{E}\left[\frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta}} \frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \boldsymbol{\theta}} \middle| \|\mathbf{P}_i\|\right]\right\} = \mathbb{E}\left\{\mathbb{E}\left[\mathbb{S}_{p_i}(\boldsymbol{\theta}) \mathbb{S}_{p_i}(\boldsymbol{\theta}) \middle| \|\mathbf{P}_i\|\right]\right\}\end{aligned}\quad (3.16)$$

donde $\mathbf{P}_i = \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\varepsilon}_i$, se puede demostrar que la matriz de información de Fisher penalizada asociado al MMS elíptico posee una estructura de bloque diagonal con la siguiente forma:

$$\mathcal{I}_p(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{X}^t \boldsymbol{\Pi} \mathbf{X} & \mathbf{X}^t \boldsymbol{\Pi} \mathbf{N} & \mathbf{0} \\ \mathbf{N}^t \boldsymbol{\Pi} \mathbf{X} & \mathbf{N}^t \boldsymbol{\Pi} \mathbf{N} + \alpha \mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}_p(\boldsymbol{\tau}, \boldsymbol{\tau}) \end{bmatrix} = \begin{bmatrix} \mathcal{I}_{[\beta, \mathbf{f}]} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{\boldsymbol{\tau}\boldsymbol{\tau}} \end{bmatrix} \quad (3.17)$$

donde $\mathcal{I}_{[\beta, \mathbf{f}]} \in \mathbb{R}^{(p+r) \times (p+r)}$, $\boldsymbol{\Pi} \in \mathbb{R}^{m \times m}$, $\boldsymbol{\Pi} = \text{diag}\left(\frac{4d_{g_1}}{n_1} \boldsymbol{\Sigma}_1^{-1}, \dots, \frac{4d_{g_n}}{n_n} \boldsymbol{\Sigma}_n^{-1}\right)$, $\mathcal{I}_{\boldsymbol{\tau}\boldsymbol{\tau}} \in \mathbb{R}^{(s+2) \times (s+2)}$ cuyos (j, k) -ésimos elementos de $\mathcal{I}_p(\boldsymbol{\tau}, \boldsymbol{\tau}) = \sum_{i=1}^n \mathcal{I}_{p_i}(\boldsymbol{\tau}, \boldsymbol{\tau})$ están dados por:

$$\mathcal{I}_{p_i}(\tau_j, \tau_k) = \frac{b_{jk}^i}{4} \left(\frac{4f_{g_i}}{n_i(n_i + 2)} - 1 \right) + \frac{2f_{g_i}}{n_i(n_i + 2)} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) \quad (3.18)$$

con $\mathbf{X} \in \mathbb{R}^{m \times p}$, $\mathbf{N} \in \mathbb{R}^{m \times r}$, $m = \sum_{i=1}^n n_i$ y \mathbf{K} es la matriz de penalización definida en la sección 3.1.1, d_{g_i} y f_{g_i} son expresiones que dependen de la distribución elíptica del modelo, ver cuadro 1.2. Las demostraciones del cálculo de cada elemento de $\mathcal{I}_p(\boldsymbol{\theta})$ se basan en la expresión (3.16) y están descritas en la tercera sección del Apéndice II.

Además, note que cada elemento de la matriz de información de Fisher penalizada entre el vector de parámetros de los componentes de varianza y el vector de parámetros asociados a los efectos fijos es cero, debido a esto, se dice que entre dichos vectores de parámetros son ortogonales, y es una propiedad útil en el sentido del esfuerzo computacional para el cálculo numérico.

3.4 Proceso de estimación

Primeramente debemos recordar que $l_p(\boldsymbol{\theta})$ es función del parámetro de suavizamiento, es por esto que el valor de α debe ser considerado en el proceso de estimación (ver sección 3.5.1), denotaremos como $l_p(\boldsymbol{\theta}, \alpha)$ la función de log-verosimilitud penalizada con parámetro de suavizado α fijo y $\hat{\boldsymbol{\theta}}$ al valor de $\boldsymbol{\theta}$ que maximiza $l_p(\boldsymbol{\theta}, \alpha)$ sobre todo el espacio paramétrico Θ . Entonces, el estimador máximo verosímil penalizado (EMVP) del vector de parámetros $\boldsymbol{\theta}$ satisface:

$$l_p(\hat{\boldsymbol{\theta}}, \alpha) = \text{Sup}_{\boldsymbol{\theta} \in \Theta} l_p(\boldsymbol{\theta}, \alpha) \quad (3.19)$$

Puesto que no es posible obtener EMVP de manera explícita como en la teoría clásica de modelos de regresión, existen varios procedimientos para poder realizar la maximización de una función de log-verosimilitud, sin embargo, para la maximización de (3.19) asumiremos que $l_p(\boldsymbol{\theta}, \alpha)$ es una función convexa y que satisface ciertas condiciones de regularidad, una discusión sobre estas condiciones en el contexto de modelos de regresión no paramétricos puede ser vista en Gu (2013) y en el contexto de estimación robusta de parámetros de localización en Huber (1981). La determinación de EMVP de los parámetros asociados al MMS elíptico se realizará mediante maximizaciones sucesivas de la función de log-verosimilitud penalizada como se describe en Ibacache-Pulgar, Paula & Galea (2012). Al asumir que el parámetro de suavizamiento α es fijo, la solución $\hat{\boldsymbol{\theta}}$ al problema de maximización de:

$$\text{Max}_{\boldsymbol{\theta} \in \Theta} l_p(\boldsymbol{\theta}, \alpha) = \text{Max}_{\boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\tau}} l_p(\boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\tau}, \alpha) \quad (3.20)$$

puede ser obtenido mediante tres procedimientos que se describen a continuación:

- I. Como primera etapa, se debe maximizar $l_p(\boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\tau}, \alpha)$ sobre $\boldsymbol{\beta}$ manteniendo los parámetros \mathbf{f} y $\boldsymbol{\tau}$ fijos, el valor máximo lo denotaremos por $\hat{\boldsymbol{\beta}}_{(\mathbf{f}, \boldsymbol{\tau})}$ y es obtenido para valores de $\boldsymbol{\beta}$ en un conjunto $\mathfrak{B}(\mathbf{f}, \boldsymbol{\tau}) \subset \Theta_{\boldsymbol{\beta}}$ que depende de los parámetros \mathbf{f} y $\boldsymbol{\tau}$. Asimismo, si $\boldsymbol{\beta} \in \mathfrak{B}(\mathbf{f}, \boldsymbol{\tau})$ entonces el valor de la función de log-verosimilitud penalizada está dada por:

$$l_p^c(\mathbf{f}, \boldsymbol{\tau}, \alpha) = \text{Max}_{\boldsymbol{\beta}} l_p(\boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\tau}, \alpha) \quad (3.21)$$

La cual, $l_p^c(\mathbf{f}, \boldsymbol{\tau}, \alpha)$ es denominada como función de log-verosimilitud penalizada concentrada en $\boldsymbol{\beta}$.

- II. El segundo procedimiento corresponde en maximizar $l_p^c(\mathbf{f}, \boldsymbol{\tau}, \alpha) = l_p^c(\hat{\boldsymbol{\beta}}_{(\mathbf{f}, \boldsymbol{\tau})}, \mathbf{f}, \boldsymbol{\tau}, \alpha)$ sobre \mathbf{f} manteniendo el vector de parámetros $\boldsymbol{\tau}$ fijo y el valor máximo lo denotaremos por $\hat{\mathbf{f}}_{(\boldsymbol{\tau})}$ que es obtenido para valores de \mathbf{f} en un conjunto $\mathfrak{F}(\boldsymbol{\tau}) \subset \Theta_{\mathbf{f}}$ que depende del parámetro $\boldsymbol{\tau}$. Correspondientemente, si $\mathbf{f} \in \mathfrak{F}(\boldsymbol{\tau})$ entonces el valor de la función de log-verosimilitud penalizada está dada por:

$$l_p^c(\boldsymbol{\tau}, \alpha) = \text{Max}_{\mathbf{f}} l_p^c(\mathbf{f}, \boldsymbol{\tau}, \alpha) \quad (3.22)$$

La cual, $l_p^c(\boldsymbol{\tau}, \alpha)$ es denominada como función de log-verosimilitud penalizada concentrada en $\boldsymbol{\beta}$ y \mathbf{f} .

- III. Y el último procedimiento corresponde a maximizar $l_p^c(\boldsymbol{\tau}, \alpha) = l_p^c(\hat{\boldsymbol{\beta}}_{(\mathbf{f}, \boldsymbol{\tau})}, \hat{\mathbf{f}}_{(\boldsymbol{\tau})}, \boldsymbol{\tau}, \alpha)$ sobre $\boldsymbol{\tau}$, el valor máximo lo denotaremos por $\hat{\boldsymbol{\tau}}$ que es obtenido para valores de $\boldsymbol{\tau}$ en un conjunto $\mathfrak{T} \subset \Theta_{\boldsymbol{\tau}}$ que depende del parámetro $\boldsymbol{\tau}$.

Por lo que estos procedimientos son equivalentes a resolver las siguientes ecuaciones:

$$\mathbb{S}_p(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} l_p(\boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\tau}, \alpha) \quad (3.23)$$

$$\mathbb{S}_p^c(\mathbf{f}) = \frac{\partial}{\partial \mathbf{f}} l_p^c(\mathbf{f}, \boldsymbol{\tau}, \alpha) \quad (3.24)$$

$$\mathbb{S}_p^c(\boldsymbol{\tau}) = \frac{\partial}{\partial \boldsymbol{\tau}} l_p^c(\boldsymbol{\tau}, \alpha) \quad (3.25)$$

Estas ecuaciones son la base de la estimación de los parámetros del MMS elíptico, el cual se implemento a través de un proceso iterativo basado en el algoritmo Backfitting para la estimación de los efectos fijos y el algoritmo de Score de Fisher para la estimación de los componentes de varianza del modelo.

3.4.1 Estimación de los efectos fijos

Tanto la componente paramétrica como la no paramétrica, corresponden a los efectos fijos del MMS elíptico. En primer lugar, para encontrar la estimación del vector de parámetros de la componente paramétrica debemos considerar (3.9) y resolver (3.23) asumiendo α , Σ_i y \mathbf{f} son fijos, entonces se puede demostrar que:

$$\boldsymbol{\beta} = [\mathbf{X}^t \mathbf{W} \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W} (\mathbf{y} - \mathbf{N} \mathbf{f}) \quad (3.26)$$

Y para el caso de la componente no paramétrica, debemos considerar la función de Score dada en (3.11) y resolver (3.24) asumiendo que los parámetros α , Σ_i y $\boldsymbol{\beta}$ son fijos, entonces se puede demostrar que:

$$\mathbf{f} = [\mathbf{N}^t \mathbf{W} \mathbf{N} + \alpha \mathbf{K}]^{-1} \mathbf{N}^t \mathbf{W} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \quad (3.27)$$

De esta manera, se propone el uso del algoritmo Backfitting, el cual nos entrega una solución para resolver (3.23) y (3.24) a través de la operación límite obtenida mediante una repetición cíclica entre las expresiones (3.26) y (3.27). Además, es posible demostrar que:

$$\boldsymbol{\beta} = [\mathbf{X}^t \mathbf{W}_n \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W}_n \mathbf{y} \quad (3.28)$$

con $\mathbf{W}_n = \mathbf{W} - \mathbf{W} \mathbf{N} (\mathbf{N}^t \mathbf{W} \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^t \mathbf{W}$, de igual manera:

$$\mathbf{f} = [\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K} \mathbf{f}]^{-1} \mathbf{N}^t \mathbf{W}_x \mathbf{y} \quad (3.29)$$

donde $\mathbf{W}_x = \mathbf{W} - \mathbf{W} \mathbf{X} (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}$, note que estas las expresiones dependen de la distribución elíptica a través de w_i y las matrices \mathbf{W}_x y \mathbf{W}_n se comportan como matrices de ponderación, y \mathbf{f} es un Spline cúbico con nodos en t_k^0 . Las demostraciones de las expresiones (3.26)-(3.29) pueden se pueden ver en la cuarta sección del Apendice II.

3.4.2 Estimación de los componentes de varianza

Según la ecuación de estimación (3.25) y asumiendo que el parámetro α es fijo, entonces se puede obtener la estimación de máxima verosimilitud penalizada a través del siguiente proceso iterativo:

$$\boldsymbol{\tau}^{(v+1)} = \arg \underset{\boldsymbol{\tau}}{\text{Max}} l_p^c(\boldsymbol{\tau}^{(v)}, \alpha) \quad (3.30)$$

donde $v = 0, 1, \dots$ y corresponde al v -ésimo calculo iterativo. Para la estimación del vector de parámetros de los componentes de varianza del MMS elíptico, se propone el uso del algoritmo de Score de Fisher el cual nos entrega a través de un proceso iterativo, una solución que es la operación límite de:

$$\begin{aligned}\mathcal{I}_p^{(v)}(\boldsymbol{\tau}, \boldsymbol{\tau}) [\boldsymbol{\tau}^{(v+1)} - \boldsymbol{\tau}^{(v)}] &= \mathbb{S}_p^{c(v)} \\ \mathcal{I}_p^{(v)}(\boldsymbol{\tau}, \boldsymbol{\tau}) \boldsymbol{\tau}^{(v+1)} &= \mathcal{I}_p^{(v)} \boldsymbol{\tau}^{(v)} + \mathbb{S}_p^{c(v)} \\ \boldsymbol{\tau}^{(v+1)} &= \boldsymbol{\tau}^{(v)} + [\mathcal{I}_p^{(v)}(\boldsymbol{\tau}, \boldsymbol{\tau})]^{-1} \mathbb{S}_p^{c(v)}\end{aligned}\quad (3.31)$$

donde $\mathcal{I}_p(\boldsymbol{\tau}, \boldsymbol{\tau}) = \sum_{i=1}^n \mathcal{I}_{p_i}(\boldsymbol{\tau}, \boldsymbol{\tau})$, $\mathcal{I}_{p_i}(\boldsymbol{\tau}, \boldsymbol{\tau}) = [\mathcal{I}_{p_i}(\tau_j, \tau_k)]$ con $j, k = 1, \dots, s + 2$ y $\mathcal{I}_{p_i}(\tau_j, \tau_k)$ fue definida en (3.18).

3.4.3 Proceso iterativo

Finalmente, el proceso iterativo conjunto para la obtención del EMVP del vector de parámetros del MMS elíptico basado en los algoritmos de Backfitting y Score de Fisher que está compuesta por dos etapas para $v, v^* = 0, 1, \dots$, en el cual denotaremos como v y v^* al v -ésimo y v^* -ésimo cálculo iterativo de la primera y segunda fase del proceso de estimación, donde $\boldsymbol{\theta}^{(0)}$ y $\boldsymbol{\theta}^{(v+1)}$ corresponden a los valores iniciales y actuales de $\boldsymbol{\theta}$ respectivamente.

Así, para el proceso de estimación de los efectos fijos del MMS elíptico, los cálculos iterativos realizados en ambas fases del proceso es el algoritmo Backfitting, y en el proceso de estimación de los componentes de varianza, los cálculos iterativos sólo utiliza el algoritmo de Score de Fisher en la primera fase evaluada en los valores maximizados de la segunda fase, esto es porque la segunda fase del proceso de estimación sólo afecta a la estimación del vector de parámetros de los efectos fijos del modelo, para ver más detalles del proceso de estimación Backfitting en otros contextos, pueden ser vistos en: Luo (1998), Hastie & Tibshirani (2000), Mammen & Park (2005; 2006), Xia (2009), Yu *et al.* (2008) y Lee *et al.* (2010).

Para las ecuaciones del proceso de estimación consideremos $v = 0, 1, \dots$ que corresponden a las ecuaciones de actualización para $\boldsymbol{\beta}$ y \mathbf{f} , y son obtenidas de las expresiones (3.23) y (3.24) que inducen interiormente un algoritmo

Backfitting con ponderaciones dadas por $\mathbf{W}^{(v)} \in \mathbb{R}^{m \times m}$, donde:

$$\begin{aligned} \boldsymbol{\Sigma}_i^{(v)} &= \mathbf{Z}_i \mathbf{D}^{(v)} \mathbf{Z}_i^t + \mathbf{Q}_i^{(v)} & \mathbf{Q}_i^{(v)} &= \phi^{(v)} [\mathbf{C}_i^{-1} \mathbf{C}_i^{-t}]^{(v)} \\ \mathbf{W}^{(v)} &= \text{diag} (w_1 \boldsymbol{\Sigma}_1^{-1}, \dots, w_n \boldsymbol{\Sigma}_n^{-1}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(v)}} & w_i^{(v)} &= -2\mathcal{W}_g (u_i^{(v)}) \end{aligned}$$

y \mathbf{C}_i es definida en (2.8). Además, el vector de medias y la distancia de Mahalanobis asociados a la i -ésima unidad de estudio son respectivamente:

$$\boldsymbol{\mu}_i^{(v)} = \mathbf{X}_i \boldsymbol{\beta}^{(v)} + \mathbf{N}_i \mathbf{f}^{(v)} \quad u_i^{(v)} = [\mathbf{y}_i - \boldsymbol{\mu}_i^{(v)}]^t \boldsymbol{\Sigma}_i^{(v)} [\mathbf{y}_i - \boldsymbol{\mu}_i^{(v)}]$$

cuya solución del algoritmo es el límite de la iteración obtenida mediante repetición cíclica entre las expresiones (3.26) y (3.27) para $v^* = 0, 1, \dots$

Algoritmo Backfitting

Para $v, v^* = 0, 1, \dots$ la estimación de los parámetros de los efectos fijos del MMS elíptico están dadas por los ecuaciones siguientes:

$$\boldsymbol{\beta}^{(v+1, v^*+1)} = [\mathbf{X}^t \mathbf{W}^{(v)} \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W}^{(v)} [\mathbf{y} - \mathbf{N} \mathbf{f}^{(v+1, v^*)}] \quad (3.32)$$

$$\mathbf{f}^{(v+1, v^*+1)} = [\mathbf{N}^t \mathbf{W}^{(v)} \mathbf{N} + \alpha \mathbf{K}]^{-1} \mathbf{N}^t \mathbf{W}^{(v)} [\mathbf{y} - \mathbf{X} \boldsymbol{\beta}^{(v+1, v^*)}] \quad (3.33)$$

donde las ecuaciones (3.32) y (3.33) deben ser calculadas hasta obtener convergencia.

Algoritmo Score de Fisher

Para $v = 0, 1, \dots$ la estimación de los parámetros de los componentes de varianza del MMS elíptico está dada por la ecuación:

$$\begin{aligned} \boldsymbol{\tau}^{(v+1)} &= \arg \text{Max}_{\boldsymbol{\tau}} l_p^c (\boldsymbol{\beta}^{(v+1)}, \mathbf{f}^{(v+1)}, \boldsymbol{\tau}, \alpha) \\ \boldsymbol{\tau}^{(v+1)} &= \boldsymbol{\tau}^{(v)} + \mathcal{I}_p^{-1} (\boldsymbol{\tau}, \boldsymbol{\tau}) \mathbb{S}_p^c (\boldsymbol{\tau}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(v)}} \end{aligned} \quad (3.34)$$

donde $\boldsymbol{\beta}^{(v+1)} = \widehat{\boldsymbol{\beta}}^{(v+1, v^*+1)}$, $\mathbf{f}^{(v+1)} = \widehat{\mathbf{f}}^{(v+1, v^*+1)}$, $\mathcal{I}_p (\boldsymbol{\tau}, \boldsymbol{\tau}) = \sum_{i=1}^n \mathcal{I}_{p_i} (\boldsymbol{\tau}, \boldsymbol{\tau})$, $\mathcal{I}_{p_i} (\boldsymbol{\tau}, \boldsymbol{\tau}) = [\mathcal{I}_{p_i} (\tau_j, \tau_k)]$ para $j, k = 1, \dots, s + 2$ y $\mathcal{I}_{p_i} (\tau_j, \tau_k)$ fue definida en (3.18).

Cuadro 3.1: Proceso iterativo de estimación del vector de parámetros del MMS elíptico

Paso	Descripción
1.	Inicie el proceso iterativo con los valores iniciales $\boldsymbol{\beta}^{(0)}$, $\mathbf{f}^{(0)}$ y $\boldsymbol{\tau}^{(0)}$
2.	Para $v, v^* = 0, 1, \dots$ se debe calcular:
2.1.	$\boldsymbol{\beta}^{(v+1, v^*+1)} = [\mathbf{X}^t \mathbf{W}^{(v)} \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W}^{(v)} [\mathbf{y} - \mathbf{N} \mathbf{f}^{(v+1, v^*)}]$
2.2.	$\mathbf{f}^{(v+1, v^*+1)} = [\mathbf{N}^t \mathbf{W}^{(v)} \mathbf{N} + \alpha \mathbf{K}]^{-1} \mathbf{N}^t \mathbf{W}^{(v)} [\mathbf{y} - \mathbf{X} \boldsymbol{\beta}^{(v+1, v^*+1)}]$
2.3.	$\boldsymbol{\tau}^{(v+1)} = \boldsymbol{\tau}^{(v)} + \mathcal{I}_p^{-1}(\boldsymbol{\tau}, \boldsymbol{\tau}) \mathbb{S}_p^c(\boldsymbol{\tau}) \Big _{\boldsymbol{\theta}=\boldsymbol{\theta}^{(v)}}$
3.	Repita los pasos 2.1, 2.2 y 2.3 hasta obtener convergencia

Los EMVP de los vectores de parámetros de los efectos fijos y los componentes de varianza del MMS elíptico serán denotados como $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(v+1)}$, $\hat{\mathbf{f}} = \mathbf{f}^{(v+1)}$ y $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^{(v+1)}$, y en este contexto, debemos considerar para el proceso iterativo establecer valores iniciales del vector de parámetros, como por ejemplo: Paula *et al.* (2009) para estimar $\boldsymbol{\theta}$ en un modelo lineal elíptico, proponen usar como valores iniciales de ϕ y ρ por $\phi^{(0)} = \frac{S^2}{\mathbf{k}}$ y $\rho^{(0)} = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2}$ con $S^2 = \frac{\sum_{i=1}^n e_i^2}{n-p}$ donde e corresponde al residuo del modelo lineal normal tradicional y \mathbf{k} es el parámetro relativo a la curtosis de la distribución elíptica.

Para el proceso descrito, denotaremos por $\boldsymbol{\beta}^{(0)}$, $\mathbf{f}^{(0)}$ y $\boldsymbol{\tau}^{(0)}$ los valores iniciales del vector de parámetros del MMS elíptico, el cual se puede utilizar para el algoritmo Backfitting como $\mathbf{f}^{(v,0)} = \mathbf{f}^{(0)}$ y los EMVP provenientes del MMS normal, en particular, tenemos que $w_i = 1$, $d_{g_i} = n_i/4$ y $f_{g_i} = n_i(n_i + 2)/4$ lo cual entrega un estimador que otorga una ponderación uniforme a través de las observaciones.

3.4.4 Estimación de los efectos aleatorios

En los modelos lineales mixtos se utiliza comúnmente la media condicional del vector de efectos aleatorios \mathbf{b}_i dado el vector de observaciones \mathbf{y}_i como estimación de los efectos aleatorios específicos por unidad de estudio, para ver otras estudios donde estiman efectos aleatorios en sus modelos, pueden dirigirse a: Laird & Ware (1982), Zhang *et al.* (1998), Zewotir & Galpin (2005), Savalli *et al.* (2006), Bastos *et al.* (2008), Lin (2008) e Ibacache-Pulgar *et al.* (2012).

Según con lo anterior, se utilizó el estimador empírico de Bayes como estimador de los efectos aleatorios, el cual se basa en la media condicional del vector de efectos aleatorios \mathbf{b}_i dado el vector de observaciones \mathbf{y}_i bajo el supuesto distribucional descrito en (2.9), el cual es:

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{bmatrix} \sim EC_{n_i+q} \left(\begin{bmatrix} \boldsymbol{\mu}_i \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_i & \mathbf{Z}_i \mathbf{D} \\ \mathbf{D} \mathbf{Z}_i^t & \mathbf{D} \end{bmatrix}, g \right)$$

Y así, se puede demostrar a través de la expresión (1.11), que la distribución condicional $\mathbf{b}_i | \mathbf{y}_i$ posee la forma:

$$\mathbf{b}_i | \mathbf{y}_i \sim EC_q \left(\mathbf{D} \mathbf{Z}_i^t \boldsymbol{\Sigma}_i^{-1} [\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{N}_i \mathbf{f}], \mathbf{D} - \mathbf{D} \mathbf{Z}_i^t \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i \mathbf{D}, g \right) \quad (3.35)$$

asumiendo $\boldsymbol{\Sigma}_i$ como un parámetro fijo, entonces el estimador empírico de Bayes de los efectos aleatorios asociados a la i -ésima unidad de estudio esta dado por:

$$\hat{\mathbf{b}}_i = \hat{\mathbb{E}}[\mathbf{b}_i | \mathbf{y}_i] = \mathbf{D} \mathbf{Z}_i^t \hat{\boldsymbol{\Sigma}}_i^{-1} [\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i] \quad (3.36)$$

donde $\hat{\boldsymbol{\mu}}_i = \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{N}_i \hat{\mathbf{f}}$ con $\hat{\boldsymbol{\beta}}$ y $\hat{\mathbf{f}}$ denotan los EMVP. Además, comúnmente en la práctica se tiene que $\boldsymbol{\Sigma}_i(\boldsymbol{\tau})$ es desconocida, el cual suele ser reemplazado por el EMVP $\hat{\boldsymbol{\tau}}$.

De esta manera, el vector de estimaciones del MMS elíptico asociados a la i -ésima unidad de estudio, denotado por $\hat{\mathbf{y}}_i$, puede demostrarse que posee la siguiente forma:

$$\hat{\mathbf{y}}_i = \hat{\mathbf{Q}}_i \hat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\mu}}_i + \left[\mathbb{I}_{n_i} - \hat{\mathbf{Q}}_i \hat{\boldsymbol{\Sigma}}_i^{-1} \right] \mathbf{y}_i \quad (3.37)$$

donde $\hat{\mathbf{y}}_i$ puede interpretarse como una media ponderada entre la media de la población y los datos observados con pesos $\widehat{\mathbf{Q}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1}$ y $[\mathbb{I}_{n_i} - \widehat{\mathbf{Q}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1}]$ respectivamente. Note que, si la variabilidad entre las n unidades de estudio es mayor que la variabilidad específica dentro de las observaciones en la i -ésima unidad de estudio, entonces se dará una mayor ponderación a $\hat{\boldsymbol{\mu}}_i$. En cambio, cuando la variabilidad intra-sujeto es mayor que la variabilidad entre-sujetos, entonces se dará una mayor ponderación al vector de observaciones de la i -ésima unidad de estudio.

Además, sea $\mathbf{b} \in \mathbb{R}^{nq}$ con $\mathbf{b} = [\mathbf{b}_1^t, \dots, \mathbf{b}_n^t]^t$, el vector de estimaciones de los efectos aleatorios de las m observaciones realizadas a las n unidades de estudio y según (7.2), $\hat{\mathbf{b}}$ puede ser escrito de la siguiente manera:

$$\hat{\mathbf{b}} = (\mathbb{I}_n \otimes \mathbf{D}) \mathbf{Z}^t \boldsymbol{\Sigma}^{-1} [\mathbf{y} - \hat{\boldsymbol{\mu}}] \quad (3.38)$$

donde $\mathbf{Z} \in \mathbb{R}^{m \times nq}$, $\mathbf{D} \in \mathbb{R}^{q \times q}$ y $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$. No obstante, en la práctica $\boldsymbol{\Sigma}$ es desconocido, por el cual debe ser reemplazado por el EMVP $\hat{\boldsymbol{\tau}}$. Por último, las demostraciones de las expresiones (3.37) y (3.38) pueden ser vistas en la cuarta sección del Apéndice II.

3.4.5 Estimación del error estándar

Para poder describir la matriz de varianzas y covarianzas del vector de parámetros estimados, comúnmente se puede sustentar en base a resultados asintóticos el cual nos entrega como una aproximación la estimación de la matriz de varianzas y covarianzas, ver por ejemplo: DasGupta (2008), Liu & Bozdogan (2008), Riquelme *et al.* (2011) e Ibacache-Pulgar *et al.* (2013). Pero en el contexto bayesiano, Segal *et al.* (1994) descubre que la estimación de la matriz de varianzas y covarianzas del EMVP desarrolladas por Wahba (1983) y Silverman (1985) corresponden a la inversa de la matriz de información observada obtenida al tratar la verosimilitud penalizada como una función de verosimilitud tradicional.

Y según con lo anterior, se propone como aproximación de la matriz de varianzas y covarianzas de $\hat{\boldsymbol{\theta}}$:

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}}) \approx \mathcal{I}_p^{-1}(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (3.39)$$

y según con la expresión (3.17), tenemos que:

$$\begin{aligned} \mathcal{I}_p^{-1}(\boldsymbol{\theta}) &= \begin{bmatrix} \mathbf{X}^t \boldsymbol{\Pi} \mathbf{X} & \mathbf{X}^t \boldsymbol{\Pi} \mathbf{N} & \mathbf{0} \\ \mathbf{N}^t \boldsymbol{\Pi} \mathbf{X} & \mathbf{N}^t \boldsymbol{\Pi} \mathbf{N} + \alpha \mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}_p(\boldsymbol{\tau}, \boldsymbol{\tau}) \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \begin{bmatrix} \mathbf{X}^t \boldsymbol{\Pi} \mathbf{X} & \mathbf{X}^t \boldsymbol{\Pi} \mathbf{N} \\ \mathbf{N}^t \boldsymbol{\Pi} \mathbf{X} & \mathbf{N}^t \boldsymbol{\Pi} \mathbf{N} + \alpha \mathbf{K} \end{bmatrix}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_p^{-1}(\boldsymbol{\tau}, \boldsymbol{\tau}) \end{bmatrix} = \begin{bmatrix} \mathcal{I}^{[\boldsymbol{\beta}, \mathbf{f}]} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}^{\boldsymbol{\tau}\boldsymbol{\tau}} \end{bmatrix} \end{aligned}$$

donde $\mathcal{I}^{[\boldsymbol{\beta}, \mathbf{f}]} \in \mathbb{R}^{(p+r) \times (p+r)}$ y $\mathcal{I}^{\boldsymbol{\tau}\boldsymbol{\tau}} \in \mathbb{R}^{(s+2) \times (s+2)}$. Sin embargo, para encontrar la matriz inversa, debemos suponer que $\mathbf{X}^t \boldsymbol{\Pi} \mathbf{X}$ y $\mathbf{N}^t \boldsymbol{\Pi} \mathbf{N} + \alpha \mathbf{K}$ son matrices definidas positivas, y aplicando la propiedad de los complementos de Schur cuando las matrices del bloque diagonal son no singulares en el caso de una matriz de bloques de orden 2x2, después de algunas operaciones matriciales, más detalles ver Lu & Shiou (2002), es posible escribir la inversa de la matriz de información de Fisher penalizada como una matriz de bloque diagonal de la forma:

$$\mathcal{I}_p^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} (\mathbf{X}^t \boldsymbol{\Pi}_n \mathbf{X})^{-1} & -\mathbf{L} & \mathbf{0} \\ -\mathbf{L}^t & (\mathbf{N}^t \boldsymbol{\Pi}_x \mathbf{N} + \alpha \mathbf{K})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}^{\boldsymbol{\tau}\boldsymbol{\tau}} \end{bmatrix} \quad (3.40)$$

donde:

$$\begin{aligned} \mathbf{L} &= (\mathbf{X}^t \boldsymbol{\Pi}_n \mathbf{X})^{-1} \mathbf{X}^t \boldsymbol{\Pi} \mathbf{N} (\mathbf{N}^t \boldsymbol{\Pi} \mathbf{N} + \alpha \mathbf{K})^{-1} \\ \boldsymbol{\Pi}_n &= \boldsymbol{\Pi} - \boldsymbol{\Pi} \mathbf{N} (\mathbf{N}^t \boldsymbol{\Pi} \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^t \boldsymbol{\Pi} \\ \boldsymbol{\Pi}_x &= \boldsymbol{\Pi} - \boldsymbol{\Pi} \mathbf{X} (\mathbf{X}^t \boldsymbol{\Pi} \mathbf{X})^{-1} \mathbf{X}^t \boldsymbol{\Pi} \\ \boldsymbol{\Pi} &= \text{diag} \left(\frac{4d_{g_1}}{n_1} \boldsymbol{\Sigma}_1^{-1}, \dots, \frac{4d_{g_n}}{n_n} \boldsymbol{\Sigma}_n^{-1} \right) \end{aligned}$$

En el caso de estar interesados en realizar inferencias para $\boldsymbol{\beta}$ y \mathbf{f} , las matrices de varianzas y covarianzas aproximadas puede ser estimadas por las correspondientes matrices del bloque diagonal de (3.40) que son:

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}) \approx (\mathbf{X}^t \boldsymbol{\Pi}_n \mathbf{X})^{-1} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (3.41)$$

$$\widehat{\text{Cov}}(\hat{\mathbf{f}}) \approx (\mathbf{N}^t \boldsymbol{\Pi}_x \mathbf{N} + \alpha \mathbf{K})^{-1} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (3.42)$$

En la práctica, la matriz $\boldsymbol{\Sigma}_i$ es desconocida por lo que se estimarán por medio de los EMVP de $\widehat{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}_i \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ así como los pesos $\widehat{w}_i = w_i \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$.

3.5 Selección del modelo

Según lo expuesto en las secciones anteriores, al ajustar un MMS elíptico es importante señalar que dichas estimaciones dependen de las matrices de ponderación que fueron descritas en (7.2), la cual depende de los componentes de varianza del modelo $\boldsymbol{\tau}$, de la matriz de penalización y del parámetro de suavizamiento α de la componente no paramétrica del MMS, como también, depende de los parámetros de la distribución elíptica conjunta descrita en (2.9). Es por esto que debemos decidir por medio de algún criterio estadístico el modelo a escoger bajo que valores de los parámetros de interés, considerando que puede ser sensible en mayor o menor medida al tamaño de muestra disponible y características de los datos.

En este contexto, se utilizan criterios de bondad de ajuste que sean capaces de permitir comparar distintos modelos ajustados a una misma muestra, ver García (1996). Como también, se emplea alguna estrategia de selección del parámetro de suavizamiento que permita escoger de manera automática el valor que cumpla con algún criterio para escoger el "mejor", dado que el parámetro de suavizamiento se considero como un parámetro fijo en los procedimientos antes descritos y éste debe ser estimado.

Para enfrentar estos problemas, en la presente sección se considerará el método de validación cruzada generalizada para la selección del parámetro de suavizamiento y del criterio de información de Schwarz para seleccionar los parámetros asociados a la distribución elíptica del MMS.

3.5.1 Validación cruzada generalizada

Considerando que en las secciones previas se asumió que el parámetro α era fijo para realizar inferencia sobre la función no paramétrica f , en la práctica dicho parámetro debe ser estimado a partir de los datos, por lo cual existen varios métodos para estimar α tales como validación cruzada y validación cruzada generalizada, criterio del estimador máximo verosímil modificado, error cuadrático medio de predicción y C_p de Mallows entre otros. Para más detalles pueden ser vistos en: Wahba (1983; 1985), Kay (1991), Stein (1993), Eubank (1999), Lee (2003) y Kauermann (2005).

En adelante, para estimar el parámetro de suavizamiento α del MMS elíptico utilizaremos el método de validación cruzada generalizada (VCG) que es un método basado en una extensión del método de validación cruzada (VC) descrito en Green & Silverman (1984) en el contexto de regresión no paramétrica, cuya motivación básica detrás de éste está en términos de predicción, suponiendo que el error aleatorio posee media cero y que el verdadero modelo tiene la propiedad de que si una observación y_{ij} se toma en un punto t_{ij} entonces \hat{y}_{ij} es el mejor predictor de y_{ij} en términos del error cuadrático medio (ECM), donde la estimación de la función no paramétrica f es un Spline cúbico con matriz de penalización \mathbf{K} . De esta manera, el valor de VCG corresponde a:

$$\text{VCG}(\alpha) = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{(1 - n^{-1}\text{tr}\{\mathbf{H}(\alpha)\})^2} \quad (3.43)$$

Y así, se debe seleccionar el parámetro de suavizamiento cuyo valor minimiza el valor de la VCG, donde $\hat{\mathbf{y}} = \mathbf{H}(\alpha)\mathbf{y}$ corresponde al vector de observaciones ajustadas que depende de α y $\mathbf{H}(\alpha)$ es una matriz denominada de influencia o como matriz *hat* del MMS elíptico y se verán más detalles de esta matriz en el próximo capítulo. No obstante, en el contexto de regresión no paramétrica, la matriz de influencia puede escribirse de manera explícita y depende de las matrices A y B definidas en la sección 3.1.1, ver más detalles en Green & Silverman (1994).

3.5.2 Criterios de información

Recordemos que los parámetros del MMS elíptico se estiman a través de máxima verosimilitud, por lo que es razonable considerar algún criterio de selección de modelos basado en sus funciones de log-verosimilitud penalizada. En el caso paramétrico, se han propuesto varios criterios de selección de modelos que en su mayoría se basan de cierta manera en el AIC⁴ (Akaike, 1974), otros métodos que se relacionan con el AIC son los llamados SIC⁵ (Schwarz, 1978) y TIC⁶.

⁴Criterio de información de Akaike

⁵Criterio de información de Schwarz

⁶Criterio de información de Takeuchi's

EL AIC se basa en la estimación de la información Kullback-Leibler basado en la función de log-verosimilitud empírica evaluada en su punto máximo, de manera análoga al caso paramétrico, se utilizan como criterios de selección del MMS elíptico el AIC y SIC, los cuales están dados por:

$$\text{AIC}(\hat{\boldsymbol{\theta}}) = -2l_p(\hat{\boldsymbol{\theta}}, \alpha) + 2p^* \quad (3.44)$$

$$\text{SIC}(\hat{\boldsymbol{\theta}}) = -2l_p(\hat{\boldsymbol{\theta}}, \alpha) + p^* \log(m) \quad (3.45)$$

donde $p^* = p + r + s + 2$ con $m = \sum_{i=1}^n n_i$ y $\hat{\boldsymbol{\theta}}$ corresponde al EMVP de $\boldsymbol{\theta}$. En este contexto, se han realizado investigaciones que comparan el comportamiento de ciertos criterios de selección cuando cambia el tamaño muestral, la estructura del proceso que generó los datos, del grado de colinealidad entre las variables explicativas o de la distribución del error aleatorio, ver por ejemplo: García (1996), Burnham & Anderson (2002) y Liang & Li (2009).

3.6 Comentarios del capítulo

En el presente capítulo se ha propuesto, un proceso de cálculo iterativo para los EMVP asociados a los parámetros del MMS elíptico AR(1), éstos están basados en los procesos de Backfitting y score de Fisher, ambos incorporan la función de log-verosimilitud penalizada del modelo marginal elíptico, que desde esta función, fueron derivadas las funciones de score e Información de Fisher penalizada asociadas a los parámetros del MMS elíptico.

Por otra parte, la estimación de los efectos aleatorios corresponde al estimador empírico de Bayes del MMS elíptico AR(1). Además, se describe un estimador asintótico del error estándar, en cambio, la estimación del parámetro de suavizamiento fue en base a la VCG. Por último, se utilizaron como criterios para la selección del MMS elíptico AR(1) los métodos basados en el AIC.

En el siguiente Capítulo, se discute el problema de análisis diagnóstico del MMS elíptico AR(1), en el cual sólo se abordan el análisis de la componente paramétrica, donde se proponen dos medidas de influencia global.

ANÁLISIS DIAGNÓSTICO DEL MMS ELÍPTICO PARA DATOS LONGITUDINALES

Un paso importante en la modelización estadística es la identificación de posibles desviaciones de los supuestos establecidos sobre en el modelo, en la literatura estadística, este paso se conoce como análisis de diagnóstico. En ese contexto, se han desarrollado procedimientos para detectar la presencia de valores atípicos, entre las técnicas más utilizadas esta el análisis de los residuos, el análisis de influencia local y global, este último basado en la eliminación de casos.

En la presente sección nos centraremos en la verificación de posibles alejamientos desproporcionados entre las observaciones y las estimaciones obtenidas por el MMS elíptico, sugiriendo el uso de una inspección gráfica de los residuos estandarizados, como también, se propone evaluar el impacto de cada observación en las estimaciones de la regresión mediante la eliminación de cada observación individual del conjunto de datos, para así, ver el comportamiento de las estimaciones de los parámetros del MMS elíptico bajo supresión de una observación.

4.1 Análisis de residuos del MMS elíptico

El análisis de residuos ha sido el primer procedimiento de diagnóstico sugerido para la detección de outliers y disponer de una herramienta descriptiva para estudiar la calidad del ajuste del modelo. Para esto, se definen dos tipos de residuos que son: el residuo marginal del modelo que está asociado a los efectos fijos del MMS y el residuo condicional asociado a los efectos fijos y aleatorios del MMS. Mayores detalles de este tipo de residuos en el contexto de modelos mixtos paramétricos pueden ser encontrados en Verberke & Molenberghs (2001). De esta manera, para definir los residuos marginal y condicional se considerarán las estimaciones de los parámetros de los efectos fijos vistos en la sección 3.4.1 y 3.4.4, suponiendo que α y Σ son fijos.

4.1.1 Residuo marginal del modelo

La media marginal del MMS elíptico para las n unidades de estudio, está dado por:

$$\hat{y} = \widehat{\mathbb{E}}[\mathbf{y}] = \hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{N}\hat{\mathbf{f}} \quad (4.1)$$

he incorporando (3.28) y (3.29) en la expresión anterior y después de algunas operaciones matriciales, se puede demostrar que:

$$\hat{y} = [\mathbf{X} \quad \mathbf{N}] \begin{bmatrix} \mathbf{X}^t \mathbf{W} \mathbf{X} & \mathbf{X}^t \mathbf{W} \mathbf{N} \\ \mathbf{N}^t \mathbf{W} \mathbf{X} & \mathbf{N}^t \mathbf{W} \mathbf{N} + \alpha \mathbf{K} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^t \\ \mathbf{N}^t \end{bmatrix} \mathbf{W} \mathbf{y} = \tilde{\mathbf{H}} \mathbf{y} \quad (4.2)$$

Y así, el vector de residuos marginal $\hat{\mathbf{e}}_m$ puede ser expresado como:

$$\hat{\mathbf{e}}_m = \mathbf{y} - \hat{y} = \mathbf{y} - \tilde{\mathbf{H}} \mathbf{y} = (\mathbb{I}_m - \tilde{\mathbf{H}}) \mathbf{y} \quad (4.3)$$

donde la matriz $\tilde{\mathbf{H}} \in \mathbb{R}^{m \times m}$ se comporta como una matriz de apalancamiento para el MMS elíptico. Análogamente con los modelos lineales tradicionales, los elementos de la diagonal de $\tilde{\mathbf{H}}$ son llamados puntos de apalancamiento y desempeñan un papel importante para la construcción de técnicas de diagnóstico, estos puntos están asociados a los parámetros de los efectos fijos del MMS. Algunas propiedades de $\tilde{\mathbf{H}}$ en el contexto de regresión paramétrica, no paramétrica y semiparamétrica son descritas en Rencher (2008), Eubank (1984; 1999) y Fung *et al.* (2002). Para más detalles de la demostración, puede ser visto en la primera sección del Apéndice III.

4.1.2 Residuo condicional del modelo

Asimismo, una expresión más general de la media condicional dada en (3.37), está dado por:

$$\hat{\mathbf{y}} = \widehat{\mathbb{E}}[\mathbf{y}|\mathbf{b}] = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{N}\hat{\mathbf{f}} + \mathbf{Z}\hat{\mathbf{b}} = \tilde{\mathbf{H}}\mathbf{y} + \mathbf{Z}\hat{\mathbf{b}} \quad (4.4)$$

he incorporando (2.11), (2.12) y (3.38) en la expresión anterior y después de algunas operaciones matriciales, se puede demostrar que:

$$\hat{\mathbf{y}} = \left[\mathbb{I}_m - \mathbf{Q}\boldsymbol{\Sigma}^{-1} + \mathbf{Q}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{H}} \right] \mathbf{y} = \mathbf{H}\mathbf{y} \quad (4.5)$$

donde $\mathbf{H} \in \mathbb{R}^{m \times m}$ y $\tilde{\mathbf{H}}$ fue definida en (4.2). De esta manera, el vector de residuos condicional $\hat{\mathbf{e}}_c$ puede ser expresado como:

$$\hat{\mathbf{e}}_c = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbb{I}_m - \mathbf{H})\mathbf{y} \quad (4.6)$$

Para más detalles de la demostración, puede ser visto en la primera sección del Apéndice III.

4.1.3 Residuos estandarizados

La matriz de covarianza de $\hat{\mathbf{e}}_m$ y $\hat{\mathbf{e}}_c$ son respectivamente:

$$\text{Cov}(\hat{\mathbf{e}}_m) = [\mathbb{I}_m - \tilde{\mathbf{H}}] \text{Cov}(\mathbf{y}) [\mathbb{I}_m - \tilde{\mathbf{H}}]^t \quad (4.7)$$

$$\text{Cov}(\hat{\mathbf{e}}_c) = [\mathbb{I}_m - \mathbf{H}] \text{Cov}(\mathbf{y}) [\mathbb{I}_m - \mathbf{H}]^t \quad (4.8)$$

en el cual, los residuos estandarizados de $\hat{\mathbf{e}}_m$ y $\hat{\mathbf{e}}_c$ están dados por:

$$\hat{\mathbf{e}}_{m_h}^* = \frac{\hat{\mathbf{e}}_{m_h}}{\sqrt{\widehat{\text{Var}}(\hat{\mathbf{e}}_{m_h})}} \quad (4.9)$$

$$\hat{\mathbf{e}}_{c_h}^* = \frac{\hat{\mathbf{e}}_{c_h}}{\sqrt{\widehat{\text{Var}}(\hat{\mathbf{e}}_{c_h})}} \quad (4.10)$$

donde $\widehat{\text{Var}}(\hat{\mathbf{e}}_{m_h})$ y $\widehat{\text{Var}}(\hat{\mathbf{e}}_{c_h})$ corresponden al h -ésimo elemento de la diagonal principal de las matrices $\text{Cov}(\hat{\mathbf{e}}_m)$ y $\text{Cov}(\hat{\mathbf{e}}_c)$ respectivamente. Además, $\text{Cov}(\mathbf{y}) = \mathbf{k}\boldsymbol{\Sigma}$ con $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_n)$ y \mathbf{k} es un parámetro relativo a la kurtosis de la distribución elíptica definida en (1.5).

Investigaciones relacionadas con el análisis de residuos son descritas en Rencher (2008) para los modelos de regresión lineal y Verberke & Molenberghs (2001) en el caso de modelos lineales mixtos ambos en el caso normal. Lange *et al.* (1989) sugieren medir la calidad del ajuste mediante una transformación de la distancia de Mahalanobis, no obstante, en el presente trabajo se considera estudiar el comportamiento de los residuos por medio de gráficos de probabilidad.

4.2 Diagnóstico de influencia global para los efectos fijos

Cook (1977) fue un pionero en esta línea, éste propuso un método para indicar la influencia de una observación en particular con respecto a un modelo de regresión lineal y demostró que dicho método proporcionaba información sobre el efecto de la variable respuesta sobre su predicción, indicando el grado de influencia que posee sobre ésta, el cual puede ser descrita como una función una la matriz de apalancamiento, este método es conocido como la distancia de Cook (DC). Para ver algunas discusiones y aplicaciones de esta distancia en otros tipos de modelos, pueden ser vistos en: Chatterjee & Hadi (1986), Velez (1988), Tomas & Cook (1989), Jimenez (2001), Díaz & Martínez (2004), Díaz & Gonzalez (2004), Zewotir & Galpin (2005) y Zhu *et al.* (2012).

Otra medida para el estudio del grado de influencia sobre los valores estimados bajo la eliminación de observaciones, es propuesta por Belsley *et al.*(2004), el cual describe el método de la distancia del valor estimado estandarizado¹, éste se define como el cambio en el valor estimado para un punto cuando dicho punto es extraído en el proceso de estimación de los parámetros asociados al modelo relativo a la variabilidad estimada del ajuste y fue propuesto inicialmente para el caso de un modelo de regresión lineal normal. En adelante, lo denotaremos por DFITS para el caso del MMS elíptico.

Actualmente, se han estudiado varias extensiones de este método de influencia en modelos mucho más generales que el modelo lineal tradicional bajo distribución normal, algunas de estas investigaciones son: Eubank (1984), Kim (1996), Kim *et al.*(2001), Kim *et al.*(2002), Fung *et al.*(2002),

¹también conocido como DFFITS

Sun & Wei (2004), Shi & Chen (2008), Carrasco *et al.*(2008), Jahufer & Jianbao (2009), Li *et al.* (2009) y Zare & Rasekh (2011).

En el contexto de modelos semiparamétricos, es más apropiado examinar la influencia parcial de las estimaciones de los parámetros asociados a los efectos fijos de la parte paramétrica y no paramétrica de manera separada para así poder proponer de manera análoga algunas medidas de influencia global en una observación, en particular, en la estimación de $\boldsymbol{\beta}$ y \mathbf{f} basadas en la distancia de Cook generalizada (DCG) y DFITS para el caso del MMS elíptico, para el caso del MMS normal ver Fung *et al.*(2002).

4.2.1 Diagnóstico basado en supresión de una observación

Sea $\hat{\boldsymbol{\theta}}_{(ij)} = [\hat{\boldsymbol{\beta}}_{(ij)}^t \ \hat{\mathbf{f}}_{(ij)}^t]^t$ la estimación del parámetro $\boldsymbol{\theta}$ con las $m-1$ observaciones disponibles removiendo la (i, j) -ésima observación en estudio, bajo el supuesto que los parámetros $\boldsymbol{\Sigma}$ y α son fijos, tenemos que para el cálculo de $\boldsymbol{\theta}_{(ij)} \forall (i, j)$ y comparar directamente con $\hat{\boldsymbol{\theta}}$ podría llevar a consumir muchísimo tiempo cuando el m es muy grande. Es por esto que, se describe una expresión de actualización bajo eliminación de una observación para así poder evitar la comparación directa con los parámetros del modelo de estimación de cada una de las m observaciones, lo cual es la base para los diagnósticos de supresión de casos que se describen en las secciones precedentes.

En adelante asumiremos que la posición de la (i, j) -ésima observación a suprimir está dada por el caso número $c = n_1 + n_2 + \dots + n_{i-1} + j$ y denotaremos por $\mathbf{d}_c \in \mathbb{R}^m$ a un vector cuya c -ésima posición posee el valor 1 y 0 en otro caso. Asimismo, denotaremos por \mathbf{y}_c , \mathbf{x}_c y \mathbf{n}_c como la c -ésima fila de \mathbf{y} , \mathbf{X} y \mathbf{N} como también, $\mathbf{y}_{(c)}$, $\mathbf{X}_{(c)}$ y $\mathbf{N}_{(c)}$ a las matrices correspondientes cuya c -ésima fila ha sido removida y para la construcción de la expresión de actualización, se tiene que:

$$y_{ij} - y_c = \frac{\mathbf{d}_c^t \mathbf{W} \hat{\mathbf{e}}_m}{\mathbf{d}_c^t \mathbf{W} (\mathbb{I}_m - \tilde{\mathbf{H}}) \mathbf{d}_c} \quad (4.11)$$

donde $\tilde{\mathbf{H}}$ es la matriz de apalancamiento descrita en (4.2) y $\hat{\mathbf{e}}_m$ es el residuo marginal del MMS elíptico definido en (4.3). La demostración de la expresión (4.11) está descrito en la segunda sección del Apendice III.

4.2.2 Influencia de la componente paramétrica

La medida de influencia para las estimaciones de los parámetros de la componente paramétrica del MMS elíptico se basa en la DCG que está dada por:

$$DCG_{(ij)}(\boldsymbol{\theta}) = [\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(ij)}]^t [\text{Cov}(\hat{\boldsymbol{\theta}})]^{-1} [\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(ij)}] \quad (4.12)$$

donde: $\text{Cov}(\hat{\boldsymbol{\theta}}) = \mathcal{I}^{\beta f}$ y $\mathcal{I}^{\beta f} = \mathcal{I}_{\beta f}^{-1}$ corresponden a una partición de la matriz de covarianza asintótica de los parámetros estimados la cual es una partición de la matriz de información de Fisher penalizada descritas en (3.17) y (3.40). Note que la DCG corresponde a una norma estandarizada de $\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(ij)}$ y el siguiente teorema muestra la expresión de dicha distancia para la componente paramétrica del MMS, la demostración de este teorema se encuentra en la segunda sección del Apéndice III.

Theorem 1. *Para el modelo (2.4) y bajo la expresión (4.11), la distancia de Cook generalizada de la componente paramétrica del MMS elíptico esta dado por:*

$$DCG_{(ij)}(\boldsymbol{\beta}) = \hat{\mathbf{e}}_m^t \mathbf{W} \mathbf{d}_c \mathbf{d}_c^t \mathbf{W}_n \mathbf{X} \mathbf{G} \mathbf{X}^t \mathbf{W}_n \mathbf{d}_c \mathbf{d}_c^t \mathbf{W} \hat{\mathbf{e}}_m$$

donde

$$\mathbf{G} = \frac{[\mathbf{X}^t \mathbf{W}_n \mathbf{X}]^{-1} \mathbf{X}^t \boldsymbol{\Pi}_n \mathbf{X} [\mathbf{X}^t \mathbf{W}_n \mathbf{X}]^{-1}}{\left(\mathbf{d}_c^t \mathbf{W} (\mathbb{I}_m - \tilde{\mathbf{H}}) \mathbf{d}_c \right)^2}$$

Para juzgar el tamaño de la distancia, note que $DCG_{(ij)}(\boldsymbol{\beta}) \leq \chi_{p;\alpha}^2$ es un elipsoide con un nivel de confianza del $1-\alpha$ para $\boldsymbol{\beta}$. Pero, si $DCG_{(ij)}(\boldsymbol{\beta}) = \chi_{p;\alpha}^2$ para algún α , la eliminación del (i, j) -ésimo caso podría mover la estimación desde el centro hasta el borde del nivel de confianza $1-\alpha$ del elipsoide.

4.2.3 Influencia de la componente no paramétrica

Una expresión similar a la distancia de Cook se puede definir para la componente no paramétrica \mathbf{f} , debido a la naturaleza local del ajuste Spline, esta expresión puede ser más apropiada para evaluar la influencia local de eliminación en una sola observación. Puesto que el efecto general de una estimación Spline en todos los puntos tiende a ser una medida desinflada y puede enmascarar la influencia real en una vecindad local.

Para esto, se valora la influencia global a través de la diferencia del valor estimado estandarizado, denotado por DFITS y está dado por:

$$DFITS_{(ij)}(\mathbf{f}) = \frac{|\mathbf{d}_c^t \mathbf{N}(\hat{\mathbf{f}} - \hat{\mathbf{f}}_{(ij)})|}{s_{(ij)}} \quad (4.13)$$

con $s_{(ij)}^2$ que corresponde al c -ésimo elemento de la diagonal de \mathbf{S}^2 , donde

$$\mathbf{S}^2 = \mathbf{N} \begin{bmatrix} \mathbf{0} & \mathbb{I}_r \end{bmatrix} \mathcal{I}^{\beta\mathbf{f}} \begin{bmatrix} \mathbf{0} & \mathbb{I}_r \end{bmatrix}^t \mathbf{N}^t$$

Theorem 2. *Para el modelo (2.4) y bajo la expresión (4.11), la diferencia del valor estimado estandarizado de la componente no paramétrica del MMS elíptico esta dado por:*

$$DFITS_{(ij)}(\mathbf{f}) = \frac{1}{s_{(ij)}} \left| \frac{\mathbf{d}_c^t \mathbf{N} [\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K}]^{-1} \mathbf{N}^t \mathbf{W}_x \mathbf{d}_c \mathbf{d}_c^t \mathbf{W} \hat{\mathbf{e}}_m}{\mathbf{d}_c^t \mathbf{W} (\mathbb{I}_m - \tilde{\mathbf{H}}) \mathbf{d}_c} \right|$$

La demostración del teorema anterior puede ser revisada en la segunda sección del Apéndice III. De manera análoga al caso de $DCG_{(ij)}(\boldsymbol{\beta})$, se puede emplear un nivel de confianza aproximado del $1-\alpha$ para el intervalo de $f(t_{ij})$ que puede ser obtenido desde $|\mathbf{d}_c^t \mathbf{N} \hat{\mathbf{f}} - f(t_{ij})| \leq s_{(ij)} z_{\alpha/2}$ donde $z_{\alpha/2}$ es el percentil teórico que acumula el $\alpha/2$ de una distribución normal estándar. Entonces, es posible evaluar el tamaño de $DFITS_{(ij)}(\mathbf{f})$ observando la ubicación de $\mathbf{d}_c^t \hat{\mathbf{f}}_{(ij)}$ con respecto a los límites del intervalo de confianza.

4.3 Comentarios del capítulo

Según lo descrito en este capítulo, se calcularon los residuos marginales y condicionales basados en los EMVP asociados a los parámetros del MMS elíptico, el cual se describieron expresiones para cada tipo de residuos con sus versiones estandarizadas. Debido a esto se propuso estudiar el comportamiento de ambos tipos de residuos por medios gráficos entre la respuesta y su predicción. Para así, poder verificar posibles alejamientos desproporcionados entre las observaciones y las estimaciones obtenidas por el MMS elíptico bajo supresión de una observación en estudio, de las cuales fueron consideradas de manera separada debido a su naturaleza semiparamétrica.

Las medidas utilizadas para estudiar la influencia global en las estimaciones de los parámetros del MMS elíptico asociados a las componentes paramétrica y no paramétrica, se calcularon expresiones de la distancia de Cook generalizada y la distancia del valor estimado estandarizado bajo un sistema de eliminación de una sola observación en el proceso de estimación.

En el siguiente Capítulo, se discute el problema de pruebas asintóticas para evaluar la hipótesis de presencia y homogeneidad de autocorrelación intra-unidad.

TEST DE SCORE PARA AUTOCORRELACIÓN DE PRIMER ORDEN INTRA-UNIDAD

En el contexto de pruebas asintóticas, la prueba de razón de verosimilitud (TRV), Wald y score pueden ser empleados para realizar inferencias sobre algunos parámetros del modelo en cuestión. Una revisión más detallada de estas pruebas, pueden ser consultadas en Bhattacharyya (1967), Buse (1982), Chen (1983), Tapas & Tapas (1988), Bera & Ullap (1991), Tsai (1993), Stram & Lee (1994), Radhakrishna & Mukerjee (1997), Van der Vart (2000), Bera & Biliias (2001), Bickel & Doksum (2001), Bescos (2002), Lindsay & Qu (2003), Rao (2002), Smyth (2003), Lemann & Romero (2005), Zhu & Zhang (2006), Lagos *et al.* (2010), Yi & Wang (2011), Lemonte & Ferrari (2012) y Feng *et al.*(2013).

Para el caso de observaciones longitudinales, comúnmente es utilizado el test de score para evaluar hipótesis estadísticas, varios autores han investigado modelos que, incluyen efectos aleatorios asociados a la unidad de estudio y errores aleatorios que poseen una estructura autorregresiva intra-unidad, además, se evalúa la heterocedasticidad intra-unidad y presencia

de autocorrelación en los errores del modelo, ver por ejemplo: Tsai (1986), Chi & Reinsel (1989), Chi & Reinsel (1991), Lin & Wei (2004), Lin & Wei (2007), Cao & Lin (2011; 2012).

El test de score fue propuesto inicialmente por Rao (1948) y posteriormente por Silvey (1959), este test es una prueba paramétrica asintótica general asociada a las funciones de score y la matriz de Información de Fisher por medio de una forma cuadrática, bajo el supuesto que la distribución de probabilidad de los errores subyacente del modelo es correcta, lo cual podría no comportarse de buena manera cuando el modelo asumido es mal especificado.

Algunas ventajas del uso del test de score es que, se requiere sólo la estimación de los parámetros bajo la hipótesis nula a diferencia de los test de Wald y TRV. En muchos casos, el desarrollo algebraico es más simple para el test de score, lo que facilita su incorporación en el desarrollo de programas informáticos.

De esta manera, la evaluación de pruebas estadísticas basadas en los test de score son particularmente atractivos y son utilizados para la evaluación de la estructura AR(1) del modelo, en el cual, se proponen dos pruebas basadas en el test de score que evalúan la presencia y homogeneidad de la autocorrelación intra-unidad en el MMS elíptico.

5.1 Test para la presencia de autocorrelación

Para la evaluación de la presencia de autocorrelación en los errores del MMS elíptico dado en (2.4), se emplea el test de score que evalúa la siguiente hipótesis:

$$H_0 : \rho = 0 \quad \text{contra} \quad H_1 : \rho \neq 0 \quad (5.1)$$

Al considerar que $\boldsymbol{\theta}$ pertenece a un subconjunto compacto de Θ tal que $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{p+r+s+2}$, se define $\boldsymbol{\theta}_* \in \Theta_{\boldsymbol{\theta}_*} \subset \mathbb{R}^{p+r+s+1}$ de tal manera que $\boldsymbol{\theta} = [\rho \quad \phi \quad \boldsymbol{\delta}^t \quad \boldsymbol{\beta}^t \quad \mathbf{f}^t]^t = [\rho \quad \boldsymbol{\theta}_*^t]^t$. Además, supondremos un punto interior $\boldsymbol{\theta}_0 \in \Theta$ como el verdadero valor de $\boldsymbol{\theta}$ cuando $\rho = 0$ con $\boldsymbol{\theta}_0 = [0 \quad \phi_0 \quad \boldsymbol{\delta}_0^t \quad \boldsymbol{\beta}_0^t \quad \mathbf{f}_0^t]^t = [0 \quad \boldsymbol{\theta}_{*0}^t]^t$. En este sentido, la hipótesis (5.1) puede ser reescrita como:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{contra} \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \quad (5.2)$$

Además, denotaremos por $\hat{\boldsymbol{\theta}}_0$ y $\hat{\boldsymbol{\theta}}$ a los EMVP de $\boldsymbol{\theta}_0$ y $\boldsymbol{\theta}$, que corresponden al estimador sujeto a la hipótesis nula y al estimador general sin restricciones, donde $\hat{\boldsymbol{\theta}}_0 = [0 \ \hat{\phi}_0 \ \hat{\boldsymbol{\delta}}_0^t \ \hat{\boldsymbol{\beta}}_0^t \ \hat{\mathbf{f}}_0^t]^t = [0 \ \hat{\boldsymbol{\theta}}_{*0}^t]^t$ y $\hat{\boldsymbol{\theta}} = [\hat{\rho} \ \hat{\phi} \ \hat{\boldsymbol{\delta}}^t \ \hat{\boldsymbol{\beta}}^t \ \hat{\mathbf{f}}^t]^t = [\hat{\rho} \ \hat{\boldsymbol{\theta}}_*^t]^t$.

5.1.1 Test de score para evaluar $H_0 : \rho = 0$

El test de score para evaluar la hipótesis (5.2) está dado por:

$$SC_a = [\mathbb{S}_p(\rho)]^2 \mathcal{I}^{\rho|\theta_*} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} \quad (5.3)$$

con

$$\mathbb{S}_p(\rho) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} = -\frac{1}{2} \sum_{i=1}^n \left[\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\rho} \right) - w_i \boldsymbol{\epsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\rho} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\epsilon}_i \right] \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} \quad (5.4)$$

y

$$\mathcal{I}^{\rho|\theta_*} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} = \left[\mathcal{I}_{\rho\rho} - [\mathcal{I}_{\rho\phi} \ \mathcal{I}_{\rho\delta}] \begin{bmatrix} \mathcal{I}_{\phi\phi} & \mathcal{I}_{\phi\delta} \\ \mathcal{I}_{\delta\phi} & \mathcal{I}_{\delta\delta} \end{bmatrix}^{-1} [\mathcal{I}_{\rho\phi} \ \mathcal{I}_{\rho\delta}]^t \right]^{-1} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} \quad (5.5)$$

Note que al evaluar la función de score en $\hat{\boldsymbol{\theta}}_0$, y después de realizar algunos desarrollos matriciales, se obtienen los siguientes resultados:

$$\begin{aligned} \mathbf{C}_i \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} &= \mathbb{I}_{n_i} \\ \dot{\mathbf{C}}_i \Big|_{\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_0} &= \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 \end{bmatrix} \\ \boldsymbol{\Sigma}_i \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} &= \mathbf{Z}_i \hat{\mathbf{D}}_0 \mathbf{Z}_i^t + \hat{\phi}_0 \mathbb{I}_{n_i} \\ \boldsymbol{\Sigma}_i^{-1} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} &= \frac{1}{\hat{\phi}_0} \left[\mathbb{I}_{n_i} - \mathbf{Z}_i \left(\hat{\phi}_0 \hat{\mathbf{D}}_0^{-1} + \mathbf{Z}_i^t \mathbf{Z}_i \right)^{-1} \mathbf{Z}_i^t \right] \\ \dot{\boldsymbol{\Sigma}}_{i\rho} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} &= -\dot{\phi} \left(\dot{\mathbf{C}}_i + \dot{\mathbf{C}}_i^t \right) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} \end{aligned}$$

donde \mathbf{C}_i fue descrita en (2.8), $\mathcal{I}_{p_i}(\tau_j, \tau_j)$ en (3.18), $\hat{\mathbf{D}}_0 = \mathbf{D}(\hat{\boldsymbol{\delta}}_0)$ y \mathbf{D} es la matriz de covarianzas asociada a los efectos aleatorios del MMS definida en la sección 3.1.

Note que las matrices de información de Fisher de $\boldsymbol{\beta}$ y \mathbf{f} no intervienen en el cálculo del test (5.3), esto es, porque los parámetros de los componentes de varianza son ortogonales a los parámetros asociados a los efectos fijos del MMS elíptico, $\mathcal{I}_p(\boldsymbol{\beta}, \boldsymbol{\tau}) = \mathbf{0}$ y $\mathcal{I}_p(\mathbf{f}, \boldsymbol{\tau}) = \mathbf{0}$.

Sin embargo, $\hat{\boldsymbol{\theta}}_0$ considera las estimaciones de los parámetros asociados a los efectos fijos del modelo, debido a que la función de score (5.3) depende de los residuos del MMS elíptico y de w_i , que es una función de la distancia de Mahalanobis que también depende de los efectos fijos del modelo.

Theorem 3. *Bajo algunas condiciones de regularidad, la distribución asintótica del test de score para evaluar (5.3) en el modelo (2.4) posee una distribución chi-cuadrado con un grado de libertad, esto es,*

$$SC_a = \left(\frac{1}{\sqrt{n}} \mathcal{S}_\rho(\boldsymbol{\theta}) \right) \left(n \mathcal{I}^{\rho|\theta^*}(\boldsymbol{\theta}) \right) \left(\frac{1}{\sqrt{n}} \mathcal{S}_\rho(\boldsymbol{\theta}) \right) \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} \sim \chi_1^2$$

donde $\mathcal{S}_\rho(\rho)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}$ y $\mathcal{I}^{\rho|\theta^*}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}$ definidos en (5.4) y (5.5), las condiciones de regularidad y la demostración de la distribución asintótica de SC_a están dadas en la primera sección del Apéndice IV.

5.1.2 Potencia empírica de SC_a

Un método clásico para evaluar la calidad de un test en la inferencia estadística es su potencia, el cual corresponde a la probabilidad de rechazar una hipótesis nula cuando realmente esta hipótesis no es cierta. En este contexto, la evaluación de SC_a se basó en un proceso de simulación de Montecarlo bajo diferentes distribuciones elípticas, para estudiar la proporción de veces en la cual se rechaza $H_0 : \rho = 0$, que es considerada como la potencia empírica del test. Además, todos los procedimientos de cálculo fueron realizados en el software R-cran (version 3.01), para esto, el proceso de simulación se basó en un diseño balanceado, digamos $n_i = k$, para generar un total de $m = nk$ observaciones dados por el siguiente modelo:

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{x}_{i1} & \mathbf{x}_{i2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \mathbf{N}_i \mathbf{f} + b_i \mathbf{1}_k + \boldsymbol{\varepsilon}_i$$

donde la función no paramétrica es $f(t_{ij}) = 10 \exp(-t_{ij}/2)$ y los verdaderos valores de los parámetros de efectos fijos paramétricos son $\beta = [-0,1 \quad 2,5]^t$, la varianza del efecto aleatorio fue de $\delta = 0,5$ y el parámetro constante del error aleatorio es $\phi = 0,1$, para el cual, los valores de t_{ij}^0 y \mathbf{x}_{i1} fueron generados desde una distribución uniforme en el intervalo $[0, 7]$ y $[0, 10]$ respectivamente, \mathbf{x}_2 fue generado desde una distribución Bernuolli con probabilidad $1/2$ de que cada unidad de estudio sea seleccionado.

Cuadro 5.1: Potencia empírica del test de Score para presencia de AR(1)

ρ	Distribuciones elípticas								
	Normal			t(4)			EP(-0.25)		
	tm_1	tm_2	tm_3	tm_1	tm_2	tm_3	tm_1	tm_2	tm_3
-1	,8248	,9854	,9976	,8042	,9878	,9989	,7658	,9758	,9969
-0,9	,7954	,9702	,9917	,7755	,9726	,9912	,7355	,9712	,9912
-0,8	,7453	,9492	,9842	,6985	,9504	,9856	,6967	,9438	,9835
-0,7	,6541	,9155	,9709	,6095	,8936	,9758	,6362	,9025	,9654
-0,6	,5228	,8452	,9423	,4849	,8227	,9512	,5508	,8441	,9331
-0,5	,4083	,7473	,9118	,3938	,7189	,9007	,4469	,7489	,8782
-0,4	,2974	,6168	,8563	,3125	,5465	,8215	,3427	,5976	,8113
-0,3	,2129	,4242	,7438	,2338	,4128	,6867	,2391	,4479	,6947
-0,2	,1342	,2216	,5326	,1746	,2854	,4562	,1748	,2742	,4841
-0,1	,1094	,1237	,1562	,1107	,1761	,2287	,1327	,1546	,2243
0	,0897	,0746	,0508	,0992	,0744	,0481	,1027	,0857	,0557
0,1	,1201	,1181	,1887	,1254	,1609	,2606	,1207	,1683	,2329
0,2	,1495	,2355	,5072	,1932	,3062	,4876	,1796	,2891	,4675
0,3	,1987	,3884	,6887	,2543	,4426	,6867	,2531	,4245	,6617
0,4	,2776	,5521	,8275	,3474	,5385	,8002	,3349	,5567	,7889
0,5	,3638	,6759	,8991	,4257	,6448	,8843	,4475	,6682	,8646
0,6	,4542	,7908	,9451	,5165	,7767	,9368	,5682	,7743	,9102
0,7	,5328	,8534	,9774	,6346	,8406	,9665	,6573	,8659	,9543
0,8	,6192	,8922	,9839	,7291	,8872	,9829	,7235	,9221	,9793
0,9	,6968	,9249	,9887	,7876	,9327	,9931	,7678	,9574	,9854
1	,7354	,9638	,9958	,8293	,9648	,9964	,8129	,9728	,9938

$tm_1: n = 5$ y $n_i = 5$ $tm_2: n = 15$ y $n_i = 5$ $tm_3: n = 15$ y $n_i = 10$

Entonces, para un conjunto de valores dados de ρ y manteniendo los valores de las covariables como fijas, se obtuvieron 500 réplicas de $y_{ij}; i = 1, \dots, n, j = 1, \dots, k$, y posteriormente se procedió a calcular SC_a con valores fijos de los valores de los efectos y errores aleatorios generados previamente desde las distribuciones normal, t -Student con 4 grados de libertad y Exponencial potencia con parámetro de forma de -0,25. Asimismo, todos los valores de SC_a son comparados con el valor crítico $\chi^2(1)$ con $\alpha = 0,05$ comparando distintos tamaños de muestra de unidades de estudio y de observaciones por unidad, los cuales fueron: $(n = 5; n_i = 5)$, $(n = 15; n_i = 5)$ y $(n = 15; n_i = 10)$, las proporciones de rechazos obtenidas por medio de las simulaciones realizadas se muestran en el Cuadro 5.1.

De esta manera, se pudo observar que a medida que la cantidad de unidades de estudio aumenta, la potencia empírica también aumenta, esta situación es análoga en el número de observaciones por unidad de estudio de estudio, cuya potencia empírica tiende a aumentar más rápidamente y que el número de unidades, sobre todo cuando los tamaños del test tienden a ser cercanos a 0,05 cuando ésta aumenta. Y aparentemente no se aprecia una clara diferencia en el comportamiento de la potencia empírica del test de score para la presencia de autocorrelación entre las distintas distribuciones elípticas simuladas, lo que es razonable considerar, como alguna evidencia del comportamiento robusto del test sobre la distribución subyacente de los datos, lo cual nos indica que este test es eficiente.

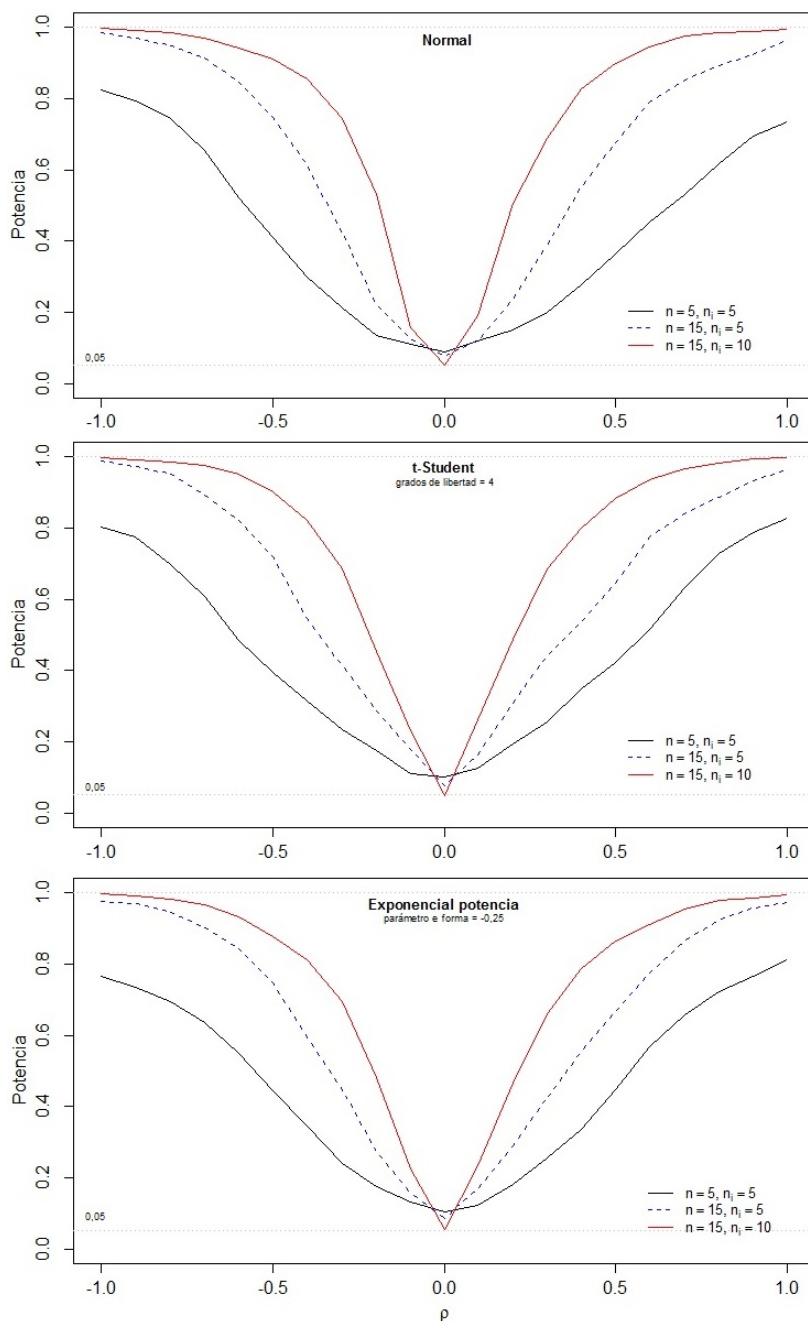


Figura 5.1: Potencia simulada de SC_a bajo distintas distribuciones elípticas

5.2 Test para la homogeneidad de autocorrelación

En el caso de asumir que la presencia de autocorrelación es una hipótesis válida, es razonable considerar probar si los coeficientes de autocorrelación entre todas las unidades de estudio son iguales, denotando por ρ_i al coeficiente de autocorrelación de los errores aleatorios correspondiente a la i -ésima unidad de estudio, la hipótesis a evaluar sería la siguiente:

$$H_0 : \rho_i = \rho \quad \text{contra} \quad H_1 : \rho_i \neq \rho \quad (5.6)$$

Para probar la homogeneidad de la autocorrelación entre las unidades, se supondrá que los coeficientes ρ_i pueden depender de algunas variables explicativas $\mathbf{v}_i \in \mathbb{R}^b$ asociadas a la i -ésima unidad de estudio a través de la siguiente forma funcional:

$$\rho_i = \rho \omega(\mathbf{v}_i, \boldsymbol{\lambda}) \quad \text{para} \quad i = 1, \dots, n \quad (5.7)$$

donde ω es una función conocida, positiva y continuamente diferenciable con respecto a $\boldsymbol{\lambda} \in \mathbb{R}^b$ que es un vector desconocido con $\boldsymbol{\lambda} = [\lambda_1 \ \dots \ \lambda_b]^t$, el cual existe un valor único $\boldsymbol{\lambda}_0$ de $\boldsymbol{\lambda}$ tal que $\omega(\mathbf{v}_i, \boldsymbol{\lambda}_0) = 1$ para $i = 1, \dots, n$. Esta metodología ha sido utilizada por varios autores, ver por ejemplo: Tsai (1986), Pourahmadi (1999), Lin & Wei (2004; 2007) y Cao & Lin (2011; 2012) emplean formas funcionales para la construcción del test de score. Considerando la hipótesis (5.6) y la forma funcional (5.7), entonces la hipótesis de homogeneidad de autocorrelación puede ser reescrita como:

$$H_0 : \boldsymbol{\lambda} = \boldsymbol{\lambda}_0 \quad \text{contra} \quad H_1 : \boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0 \quad (5.8)$$

Puesto que en la forma funcional se incorpora un vector de parámetros desconocidos, éstos deben ser estimados bajo H_0 , este nuevo vector asociado al MMS elíptico lo denotaremos por $\boldsymbol{\theta}^* = [\boldsymbol{\lambda}^t \ \boldsymbol{\theta}^t]^t$ el cual, diremos que pertenece a un subconjunto compacto de Θ^* tal que $\boldsymbol{\theta}^* \in \Theta^* \subset \mathbb{R}^{p+r+b+s+2}$, de manera que $\boldsymbol{\theta}^* = [\boldsymbol{\beta}^t \ \mathbf{f}^t \ \boldsymbol{\lambda}^t \ \boldsymbol{\delta}^t \ \phi \ \rho]^t = [\boldsymbol{\lambda} \ \boldsymbol{\theta}^t]^t$. también, supondremos un punto interior $\boldsymbol{\theta}_0^* \in \Theta^*$ como el verdadero valor de $\boldsymbol{\theta}^*$ cuando $\rho_i = \rho$, con $\boldsymbol{\theta}_0^* = [\boldsymbol{\beta}_0^t \ \mathbf{f}_0^t \ \boldsymbol{\lambda}_0^t \ \boldsymbol{\delta}_0^t \ \phi_0 \ \rho_0]^t = [\boldsymbol{\lambda}_0^t \ \boldsymbol{\theta}_0^t]^t$.

Asimismo, $\hat{\boldsymbol{\theta}}_0^*$ y $\hat{\boldsymbol{\theta}}^*$ son los EMVP de $\boldsymbol{\theta}_0^*$ y $\boldsymbol{\theta}^*$, que corresponden al estimador sujeto a la hipótesis nula y al estimador general sin restricciones, donde $\hat{\boldsymbol{\theta}}_0^* = [\hat{\boldsymbol{\beta}}_0^t \ \hat{\mathbf{f}}_0^t \ \hat{\boldsymbol{\lambda}}_0^t \ \hat{\boldsymbol{\delta}}_0^t \ \hat{\phi}_0 \ \hat{\rho}_0]^t = [\hat{\boldsymbol{\lambda}}_0^t \ \hat{\boldsymbol{\theta}}_0^t]^t$ y $\hat{\boldsymbol{\theta}}^* = [\hat{\boldsymbol{\beta}}^t \ \hat{\mathbf{f}}^t \ \hat{\boldsymbol{\lambda}}^t \ \hat{\boldsymbol{\delta}}^t \ \hat{\phi} \ \hat{\rho}]^t = [\hat{\boldsymbol{\lambda}}^t \ \hat{\boldsymbol{\theta}}^t]^t$. Además, la función de score y la matriz de información de Fisher penalizada asociada al parámetro $\boldsymbol{\lambda}$ poseen una estructura análoga a (3.13) y (3.18), pero en este caso, se debe considerar que el vector de parámetros asociados a los componentes de varianza es $\boldsymbol{\tau}^* = [\boldsymbol{\lambda}^t \ \boldsymbol{\delta}^t \ \phi \ \rho]^t = [\boldsymbol{\lambda}^t \ \boldsymbol{\tau}^t]^t$.

Cuadro 5.2: Algunas funciones de ponderación para la estructura AR(1) de tipo heterogénea, con $x, y \in \mathbb{R}$

Tipo de función y notación	Función de ponderación $\omega(x, y)$	$\frac{\partial}{\partial y}\omega(x, y)$
Logística $\omega_1(x, y)$	$2 \frac{e^{xy}}{1 + e^{xy}} = \frac{2}{1 + e^{-xy}}$	$2x \frac{e^{xy}}{[1 + e^{xy}]^2}$
Gompertz $\omega_2(x, y)$	$\exp(1 - e^{-xy})$	$x \exp[1 - xy - e^{-xy}]$

Cuando la evaluación del test para la homogeneidad de la autocorrelación a nivel intra-unidad están basados en algunas de las formas funcionales descritas en el Cuadro 5.2 es equivalente a evaluar $H_0 : \boldsymbol{\lambda} = \mathbf{0}$, puesto que para cada elemento λ_{0k} de $\boldsymbol{\lambda}_0$ se produce que $\omega_1(v_{ik}, 0) = \omega_2(v_{ik}, 0) = 1$ lo cual implica que $\rho_i = \rho$ para $i = 1, \dots, n$.

5.2.1 Test de Score para evaluar $H_0 : \rho_i = \rho$

De manera análoga al caso de la prueba de presencia de autocorrelación, el test para evaluar la homogeneidad de la autocorrelación a nivel intra-unidad en el MMS elíptico está dado por:

$$SC_h = [\mathbb{S}_p(\boldsymbol{\lambda})]^t \mathcal{I}^{\lambda|\theta} [\mathbb{S}_p(\boldsymbol{\lambda})] \Big|_{\theta^*=\hat{\theta}_0^*} \quad (5.9)$$

con

$$\mathbb{S}_p(\boldsymbol{\lambda}) \Big|_{\theta^*=\hat{\theta}_0^*} = -\frac{1}{2} \sum_{i=1}^n \left[\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\lambda} \right) - w_i \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\lambda} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right] \Big|_{\theta^*=\hat{\theta}_0^*} \quad (5.10)$$

y

$$\mathcal{I}^{\lambda|\theta} \Big|_{\theta^*=\hat{\theta}_0^*} = \left[\mathcal{I}_{\lambda\lambda} - \begin{bmatrix} \mathcal{I}_{\lambda\rho} & \mathcal{I}_{\lambda\phi} & \mathcal{I}_{\lambda\delta} \end{bmatrix} \begin{bmatrix} \mathcal{I}_{\rho\rho} & \mathcal{I}_{\rho\phi} & \mathcal{I}_{\rho\delta} \\ \mathcal{I}_{\phi\rho} & \mathcal{I}_{\phi\phi} & \mathcal{I}_{\phi\delta} \\ \mathcal{I}_{\delta\rho} & \mathcal{I}_{\delta\phi} & \mathcal{I}_{\delta\delta} \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{I}_{\rho\lambda} \\ \mathcal{I}_{\phi\lambda} \\ \mathcal{I}_{\delta\lambda} \end{bmatrix} \right]^{-1} \Big|_{\theta^*=\hat{\theta}_0^*} \quad (5.11)$$

donde $\mathcal{I}_{\tau\tau} = \sum_{i=1}^n \mathcal{I}_{i\tau\tau}$ es una matriz cuyo (j, k) -ésimo elemento fue descrito en (3.18).

Para la evaluación de $\hat{\boldsymbol{\theta}}_0^*$ en la función de score y matrices de información de Fisher penalizadas, se debe considerar maximizar la estructura de varianza del modelo $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t + \phi \mathbf{C}_i^{-1} \mathbf{C}_i^{-t}$ respecto de ρ y del nuevo parámetro $\boldsymbol{\lambda}$, que son los dos parámetros asociados a ρ_i , puesto que se dependen directamente de la estructura de la matriz \mathbf{C}_i que incorpora una relación funcional ω en la estructura de autocorrelación a nivel intra-unidad. Por lo cual, tenemos que la estructura AR(1) del MMS elíptico sujeto a (5.7) está dado por:

$$\mathbf{C}_i = \begin{bmatrix} \sqrt{1-\rho_i^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho_i & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho_i & 1 \end{bmatrix} = \begin{bmatrix} c_{11}^i & 0 & \cdots & \cdots & 0 \\ c_{21}^i & c_{22}^i & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & c_{n_i n_i - 1}^i & c_{n_i n_i}^i \end{bmatrix}$$

en el cual podemos reconocer que \mathbf{C}_i depende de la función de ponderación ω y \mathbf{v}_i , ésta posee la misma estructura funcional a través de las unidades

de estudio. También, podemos observar que los elementos de la diagonal principal de \mathbf{C}_i , digamos c_{jj}^i , son todos iguales a uno, con la excepción de $c_{11}^i = \sqrt{1 - \rho_i^2}$ y $c_{jj-1}^i = -\rho_i$. De esta manera, tenemos que $\dot{\Sigma}_{i\rho}$ está dado por:

$$\dot{\Sigma}_{i\rho} = -\phi \mathbf{C}_i^{-1} \left(\dot{\mathbf{C}}_{i\rho} \mathbf{C}_i^{-1} + \mathbf{C}_i^{-t} \dot{\mathbf{C}}_{i\rho}^t \right) \mathbf{C}_i^{-t}$$

donde $\dot{\mathbf{C}}_{i\rho}$ posee elementos que corresponden a:

$$\begin{aligned} \frac{\partial c_{11}^i}{\partial \rho} &= -\frac{\rho \omega^2(\mathbf{v}_i, \boldsymbol{\lambda})}{\sqrt{1 - [\rho \omega(\mathbf{v}_i, \boldsymbol{\lambda})]^2}} \\ \frac{\partial c_{jj-1}^i}{\partial \rho} &= -\omega(\mathbf{v}_i, \boldsymbol{\lambda}) \\ \frac{\partial c_{jj}^i}{\partial \rho} &= 0 \quad \forall i = 1, \dots, n, j = 2, \dots, n_i \end{aligned}$$

De igual manera, tenemos que $\dot{\Sigma}_{i\lambda}$ está dado por:

$$\dot{\Sigma}_{i\lambda_k} = -\phi \mathbf{C}_{i\lambda_k}^{-1} \left(\dot{\mathbf{C}}_{i\lambda_k} \mathbf{C}_{i\lambda_k}^{-1} + \mathbf{C}_{i\lambda_k}^{-t} \dot{\mathbf{C}}_{i\lambda_k}^t \right) \mathbf{C}_{i\lambda_k}^{-t} \quad \forall k = 1, \dots, b$$

donde la matriz $\dot{\mathbf{C}}_{i\lambda}$ posee los siguientes elementos:

$$\begin{aligned} \frac{\partial c_{11}^i}{\partial \lambda_k} &= -\frac{\rho^2 \omega(v_{ik}, \lambda_k)}{\sqrt{1 - [\rho \omega(v_{ik}, \lambda_k)]^2}} \frac{\partial}{\partial \lambda_k} [\omega(v_{ik}, \lambda_k)] \\ \frac{\partial c_{jj-1}^i}{\partial \lambda_k} &= -\frac{\partial \rho_i}{\partial \lambda_k} = -\rho \frac{\partial}{\partial \lambda_k} [\omega(v_{ik}, \lambda_k)] \\ \frac{\partial c_{jj}^i}{\partial \lambda_k} &= 0 \quad \forall i = 1, \dots, n, j = 2, \dots, n_i, k = 1, \dots, b \end{aligned}$$

Theorem 4. *Bajo algunas condiciones de regularidad, la distribución asintótica del test de score para evaluar (5.9) en el modelo (2.4) posee una distribución chi-cuadrado con b grados de libertad, esto es,*

$$SC_h = \left(\frac{1}{\sqrt{n}} \mathcal{S}_\lambda(\boldsymbol{\theta}^*) \right) \left(n \mathcal{I}^{\lambda|\boldsymbol{\theta}^*}(\boldsymbol{\theta}^*) \right) \left(\frac{1}{\sqrt{n}} \mathcal{S}_\lambda(\boldsymbol{\theta}^*) \right) \Big|_{\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}_0^*} \sim \chi_b^2$$

donde $\mathbb{S}_p(\lambda)|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0}$ y $\mathcal{I}^{\lambda|\boldsymbol{\theta}^*}|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_0}$ definidos en (5.10) y (5.11), las condiciones de regularidad y la demostración de la distribución asintótica de SC_h están dadas en la segunda sección del Apéndice IV.

5.2.2 Potencia empírica de SC_h

De manera análoga a la sección 5.1.2, para la evaluación de SC_h se basó en un proceso de simulación de Montecarlo bajo diferentes distribuciones elípticas, para estudiar la proporción de veces en la cual se rechaza $H_0 : \rho_i = \rho$, que es la potencia empírica del test. Además, todos los procedimientos de cálculo fueron realizados en el software R-cran (version 3.01), para esto, el proceso de simulación se basó en un diseño balanceado, para generar un total de $m = nk$ observaciones dados por el siguiente modelo:

$$\mathbf{y}_i = [\mathbf{x}_{i1} \quad \mathbf{x}_{i2}] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \mathbf{N}_i \mathbf{f} + b_i \mathbf{1}_k + \boldsymbol{\varepsilon}_i$$

donde la función no paramétrica es $f(t_{ij}) = 10 \exp(-t_{ij}/2)$ y los verdaderos valores de los parámetros de efectos fijos paramétricos son $\boldsymbol{\beta} = [-0,1 \quad 2,5]^t$, la varianza del efecto aleatorio fue de $\delta = 0,5$, el parámetro constante del error aleatorio fue de $\phi = 0,1$ y con parámetro de autocorrelación común $\rho = 0,25$, para el cual, los valores de t_{ij}, v_i y \mathbf{x}_{i1} fueron generados desde una distribución uniforme en los intervalos $[0, 5]$, $[0, 7]$ y $[0, 10]$ respectivamente, en cambio, \mathbf{x}_2 fue generado desde una distribución Bernuolli con probabilidad $1/2$ de que cada unidad de estudio sea seleccionado.

Entonces, para un conjunto de valores dados de $\lambda \in [0, 0,1]$ y manteniendo los valores de las covariables como fijas, se obtuvieron 500 réplicas de $y_{ij}; i = 1, \dots, n, j = 1, \dots, k$, y posteriormente se procedió a calcular SC_h con valores fijos de los valores de los efectos y errores aleatorios generados previamente desde las distribuciones normal, t -Student con 4 grados de libertad y Exponencial potencia con parámetro de forma de $-0,25$. Asimismo, todos los valores de SC_h son comparados con el valor crítico $\chi^2(1)$ con $\alpha = 0,05$ comparando distintos tamaños de muestra de unidades de estudio y de observaciones por unidad, los cuales fueron: $(n = 30; n_i = 20)$, $(n = 30; n_i = 30)$ y $(n = 50; n_i = 30)$, las proporciones de rechazos obtenidas por medio de las simulaciones realizadas se muestran en los Cuadros 5.3 y 5.4 bajo estructuras de autocorrelación con formas funcionales logística y gompertz.

De esta manera, se pudo observar que a medida que la cantidad de unidades de estudio aumenta, la potencia empírica también aumenta, esta situación es análoga en el número de observaciones por unidad de estudio,

cuya potencia empírica tiende a aumentar más rápidamente y que los tamaños del test tienden a ser cercanos a 0,05 cuando la cantidad de observaciones aumenta. Y aparentemente no se aprecia una clara diferencia en el comportamiento de la potencia empírica del test de score para la homogeneidad de autocorrelación entre las distintas distribuciones elípticas simuladas así como también, entre la forma funcional utilizada para la estructura de autocorrelación, lo cual se que podría considerarse como una evidencia sobre el comportamiento robusto del test de score sobre la distribución subyacente de los datos y de la estructura de autocorrelación intra-unidad.

5.3 Comentarios del capítulo

En el presente capítulo se determinó el test de score para evaluar las hipótesis de presencia de autocorrelación y su homogeneidad a través de las unidades de estudio, de los cuales, se describieron las expresiones de dichas pruebas y sus distribuciones asintóticas respectivas, mostrando las funciones de score asociadas a los nuevos parámetros incorporados al modelo para evaluar el test.

Además, se realizó un estudio de simulación de la potencia del test a través de una simulación Monte-Carlo, del cual se muestra que el test de score para evaluar la presencia de autocorrelación posee una buena potencia empírica para una cantidad -relativamente no muy alta- de unidades de estudio. En cambio, la potencia empírica del test de homogeneidad de autocorrelación depende de una mayor cantidad de observaciones, no obstante, el comportamiento de la potencia empírica es robusta a la distribución subyacente de los datos y de la forma funcional de la estructura de autocorrelación de tipo no homogénea, por lo que el test de score se comporta de manera muy eficiente.

En el siguiente Capítulo, se describen dos aplicaciones de la metodología expuesta en este trabajo, ambos conjuntos de datos provienen del área de salud.

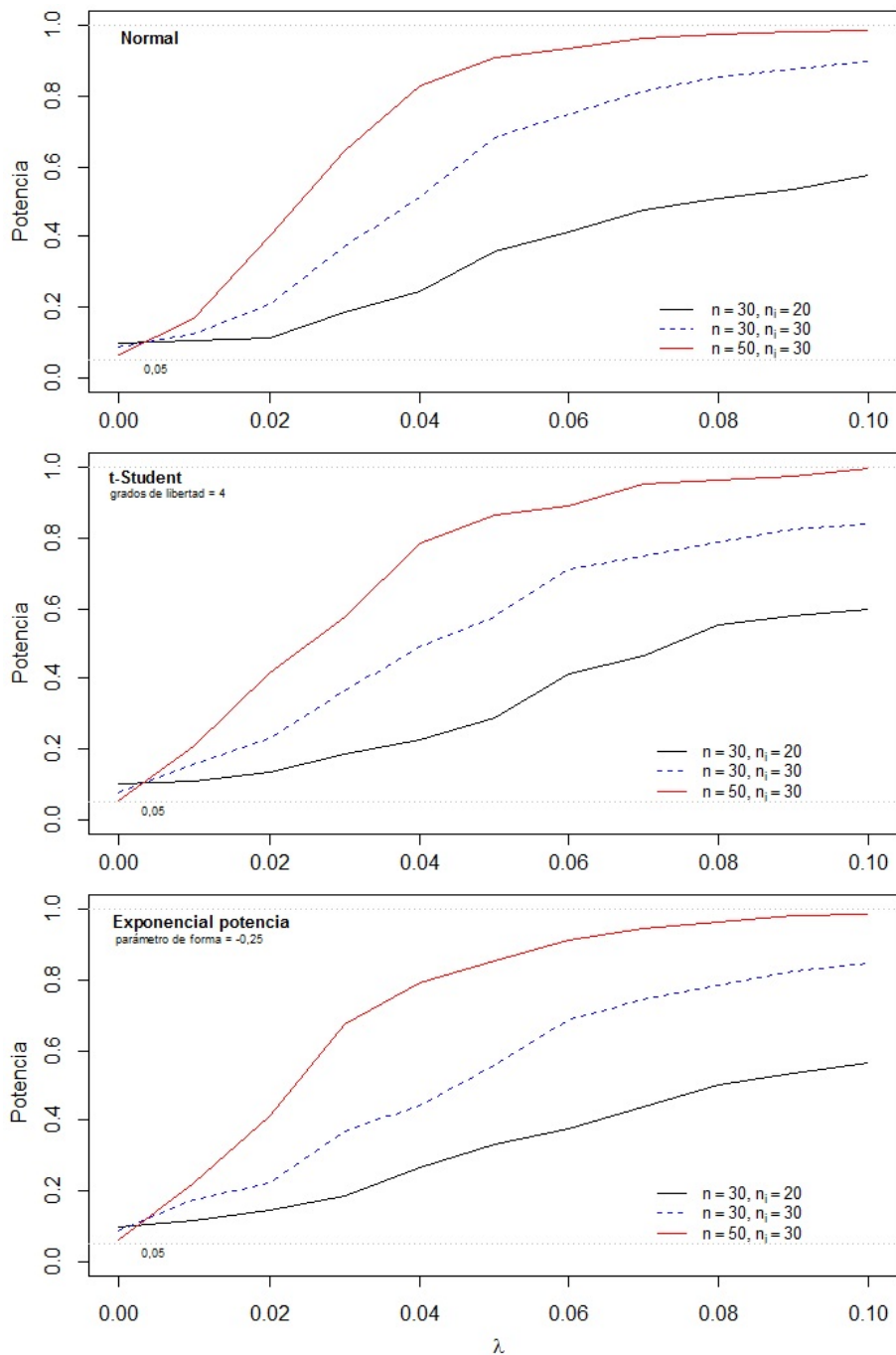


Figura 5.2: Potencia empírica del test SC_h bajo una forma funcional de tipo logística según algunas distribuciones elípticas

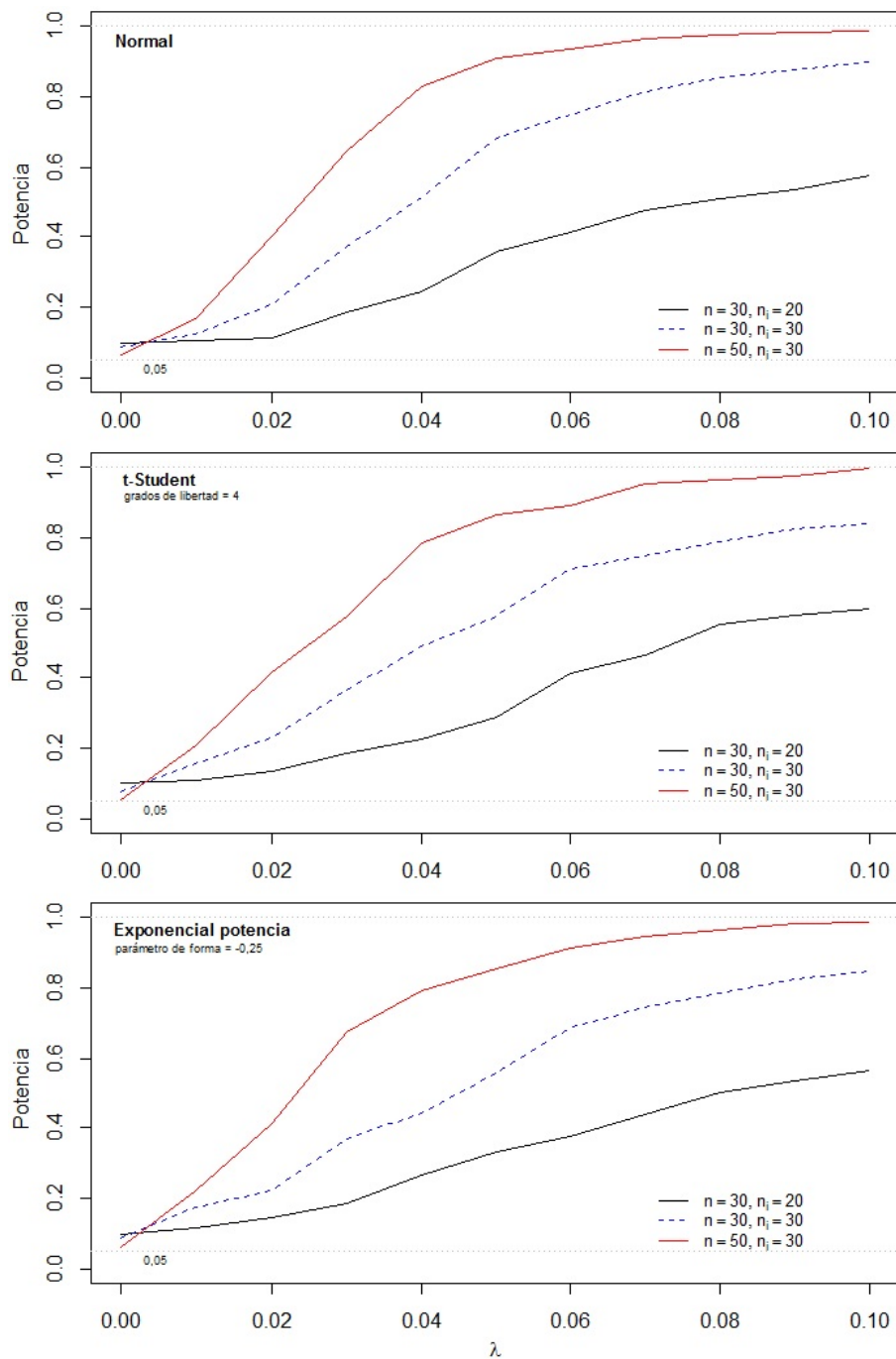


Figura 5.3: Potencia empírica del test SC_h bajo una forma funcional de tipo Gompertz según algunas distribuciones elípticas

Cuadro 5.3: Potencia empírica del test de Score para homogeneidad de autocorrelación sujeto a una forma funcional logística

λ	Distribuciones elípticas								
	Normal			t(4)			EP(-0.25)		
	tm_1	tm_2	tm_3	tm_1	tm_2	tm_3	tm_1	tm_2	tm_3
0	,0982	,0847	,0649	,1013	,0757	,0548	,0978	,0882	,0601
,01	,1042	,1239	,1679	,1087	,1566	,2078	,1142	,1736	,2207
,02	,1138	,2066	,4018	,1329	,2279	,4174	,1432	,2235	,4108
,03	,1859	,3711	,6477	,1864	,3642	,5768	,1853	,3689	,6762
,04	,2448	,5142	,8291	,2245	,4934	,7835	,2658	,4457	,7933
,05	,3573	,6819	,9093	,2864	,5751	,8655	,3301	,5564	,8546
,06	,4123	,7494	,9348	,4163	,7109	,8905	,3739	,6879	,9115
,07	,4774	,8128	,9652	,4678	,7496	,9531	,4428	,7434	,9453
,08	,5104	,8549	,9741	,5561	,7882	,9638	,5037	,7865	,9657
,09	,5358	,8768	,9826	,5798	,8258	,9762	,5361	,8244	,9834
,10	,5781	,8991	,9875	,5967	,8394	,9981	,5638	,8456	,9871

Cuadro 5.4: Potencia empírica del test de Score para homogeneidad de autocorrelación sujeto a una forma funcional gompertz

λ	Distribuciones elípticas								
	Normal			t(4)			EP(-0.25)		
	tm_1	tm_2	tm_3	tm_1	tm_2	tm_3	tm_1	tm_2	tm_3
0	,0947	,089	,0545	,0962	,0756	,0652	,0724	,0526	,0501
,01	,1053	,1263	,1578	,1048	,1353	,1761	,0921	,1743	,2207
,02	,1347	,2079	,3879	,1294	,2234	,4572	,1142	,2375	,4375
,03	,2253	,3447	,6243	,2078	,3328	,6647	,1895	,3869	,6936
,04	,2867	,4933	,7658	,2448	,4758	,7872	,2378	,4755	,8121
,05	,3656	,6018	,8844	,2876	,5856	,8544	,2664	,5482	,8742
,06	,4185	,6852	,9237	,4239	,6894	,8689	,3529	,6671	,9089
,07	,4972	,7782	,9569	,4682	,7651	,9002	,4085	,7438	,9326
,08	,5562	,8453	,9741	,5457	,7944	,9506	,4958	,7645	,9543
,09	,5694	,8909	,9782	,5872	,8175	,9795	,5287	,8193	,9758
,10	,6128	,9137	,9896	,6261	,8265	,9867	,5498	,8546	,9927

tm_1 : $n = 30$ y $n_i = 20$ tm_2 : $n = 30$ y $n_i = 30$ tm_3 : $n = 50$ y $n_i = 30$

APLICACIONES

En este último capítulo, se describen dos aplicaciones del MMS elíptico para datos longitudinales con estructura AR(1), el primer conjunto de datos fue reportado por Zerbe (1979) y estos datos fueron analizados posteriormente en Ibacache *et al.* (2011; 2014). En cambio, el segundo conjunto de datos corresponden a un diseño experimental realizado por la Dra. Jessica Rojas¹ y Dr. Jorge Godoy O.² en el contexto de Odontología pediodontal.

Para ambos conjuntos de datos, describiremos las estimaciones de los parámetros del MMS elíptico y sus errores estándar asintóticos, como también, las distancias de Cook asociadas al modelo y los test de presencia y homogeneidad de autocorrelación a nivel intra-unidad, los cálculos y gráficos fueron hechos en el software R-cran (versión 3.0.1).

¹Facultad de Odontología de la Universidad Andrés Bello, Viña del Mar

²Facultad de Odontología de la Universidad de Valparaíso, Valparaíso

Cuadro 6.1: Medidas descriptivas para los niveles de fosfato inorgánico

mg/dl	Grupo		
	Control	Obesos	Ambos
n	13	20	33
Mínimo	1,500	1,900	1,500
Q1	2,875	3,275	3,100
Mediana	3,400	3,800	3,700
Q3	3,825	4,300	4,100
Máximo	5,000	6,600	6,600
Promedio	3,319	3,777	3,597
Desviación estándar	0,712	0,831	0,816

6.1 Conjunto de datos N°1: fosfato inorgánico

En un estudio médico se aplicaron diferentes pruebas estándar de tolerancia a la glucosa hechas a 20 pacientes obesos y 13 pacientes de un grupo control, los datos corresponden a mediciones de fosfato inorgánico de plasma en muestras de sangre obtenidas de cada paciente, cuya unidad de medida es miligramos por decilitro (mg/dl). Además, los datos fueron registrados a las 0, 0,5, 1, 1,5, 2, 3, 4 y 5 horas después de la aplicación oral de una dosis estándar de glucosa.

Las mediciones de fosfato inorgánico de plasma son presentadas en el Apéndice V, éstas fueron reportadas por Zerbe (1979) y posteriormente discutidos por Chi & Reinsel (1989) e Ibacache et al. (2014), donde proponen modelar la respuesta media de cada paciente en el tiempo a través de una función de tipo no paramétrica.

Las medidas descriptivas del nivel de fosfato se muestran en el Cuadro 6.1, donde, las dos primeras columnas muestran las medidas descriptivas para el grupo de pacientes control y obesos, en cambio, la última columna representa al total de pacientes en estudio. En la Figura 6.1, se observa el nivel de fosfato inorgánico de plasma a través del tiempo, donde tiende ser mayor los niveles de fosfato en pacientes obesos, pero el nivel de fosfato promedio en en ambos grupos de estudio y en general, se comporta de manera no lineal a través de las horas de observación del paciente.

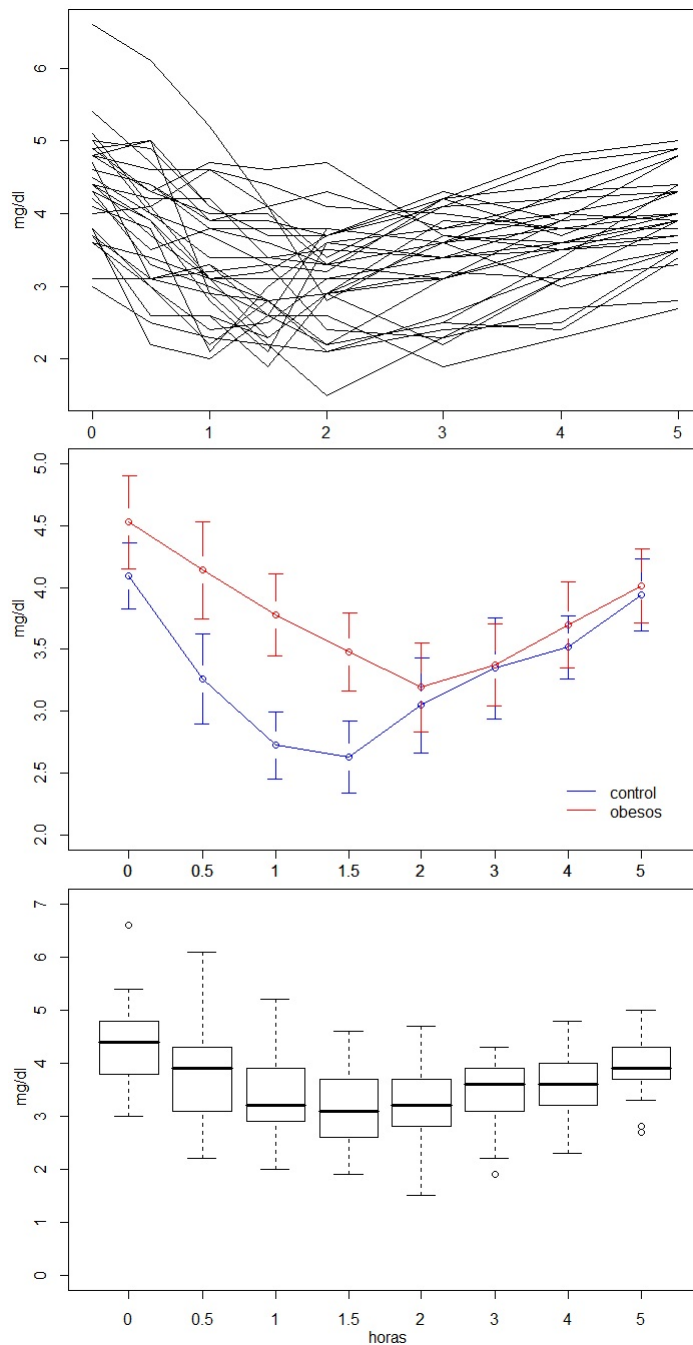


Figura 6.1: Gráficos descriptivos: perfiles de mediciones de fosfato inorgánico por paciente, IC del 95% para la media y boxplot a través de los tiempos de medición

6.1.1 Modelo propuesto

Es razonable considerar que el nivel de fosfato inorgánico medio se comporta de manera no lineal a través del tiempo, para esto, se incorpora un efecto del tiempo por medio de una función no paramétrica sobre la respuesta, y así se propone el siguiente modelo:

$$y_{ij} = \beta x_i + f(t_{ij}) + b_i + \varepsilon_{ij} \quad (6.1)$$

donde y_{ij} es la j -ésima medición de fosfato inorgánico de plasma del i -ésimo paciente obtenida al tiempo t_{ij} , la función f es una función arbitraria definida en la sección 3.1.1. Además, la variable explicativa x_i es una variable indicadora, que corresponde a $x_i = 1$ cuando el paciente es obeso y $x_i = 0$ en caso contrario, $\beta \in \mathbb{R}$ es un parámetro desconocido, $b_i \in \mathbb{R}$ es el efecto aleatorio asociado al i -ésimo paciente en estudio y ε_{ij} corresponde al error aleatorio asociado a la j -ésima observación medida al i -ésimo paciente, para $i = 1, \dots, 33$ y $j = 1, \dots, 8$. El modelo (6.1) puede ser reescrito de forma matricial de la siguiente manera:

$$\mathbf{y}_i = \mathbf{X}_i \beta + \mathbf{N}_i \mathbf{f} + \mathbf{Z}_i b_i + \boldsymbol{\varepsilon}_i \quad (6.2)$$

donde $\mathbf{y}_i, \boldsymbol{\varepsilon}_i \in \mathbb{R}^8$ son los vectores de respuestas y errores aleatorios asociados al i -ésimo paciente, $\mathbf{X}_i \in \mathbb{R}^{8 \times 1}$ y $\mathbf{Z}_i \in \mathbb{R}^{8 \times 1}$ son los vectores asociados a la componentes paramétrica y aleatoria del MMS, el vector de la parte aleatoria contiene sólo valores iguales a uno, esto es, $\mathbf{Z}_i = \mathbf{1}_8$. Además, $\mathbf{N}_i \in \mathbb{R}^{8 \times 8}$ es una matriz de incidencia con $\mathbf{N}_i = \mathbb{I}_8$ y \mathbf{f} corresponde a un vector cuyos elementos son los 8 tiempos diferentes evaluados en la función f desconocida con nodos en $t_1^0 = 0, t_2^0 = 0,5, \dots, t_8^0 = 5$. Asimismo, al considerar el total de observaciones realizadas a los 33 pacientes ($m = 264$), podemos reescribir el MMS de la siguiente manera:

$$\mathbf{y} = \mathbf{X} \beta + \mathbf{N} \mathbf{f} + \mathbf{Z} \mathbf{b} + \boldsymbol{\varepsilon}$$

donde $\mathbf{X} \in \mathbb{R}^{264 \times 1}$, $\mathbf{N} \in \mathbb{R}^{264 \times 8}$, $\mathbf{b} \in \mathbb{R}^{33 \times 1}$ y $\mathbf{Z} = \text{diag}(\mathbf{1}_8, \mathbf{1}_8, \dots, \mathbf{1}_8)$. Y al asumir que la variable respuesta y los efectos aleatorios provienen de una distribución elíptica conjunta dada por:

$$\begin{bmatrix} \mathbf{y}_i \\ b_i \end{bmatrix} \sim \text{EC}_9 \left(\begin{bmatrix} \boldsymbol{\mu}_i \\ 0 \end{bmatrix}, \begin{bmatrix} \delta \mathbf{Z}_i \mathbf{Z}_i^t + \phi \mathbf{C}_i^{-1} \mathbf{C}_i^{-t} & \delta \mathbf{Z}_i \\ \delta \mathbf{Z}_i^t & \delta \end{bmatrix} \right) \quad (6.3)$$

Cuadro 6.2: Valores estimados de la log-verosimilitud penalizada, AIC y SIC

Distribución	$l_p(\hat{\boldsymbol{\theta}}, \hat{\alpha})$	SIC	AIC
Normal	-177,672	422,256	379,345
t -Student(21)	-176,952	420,816	377,904
Exponencial potencia(0,45)	-176,741	420,394	377,482

se obtiene que, el vector de respuestas sigue una distribución marginal elíptica de la forma:

$$\mathbf{y}_i \sim \text{EC}_8 (\mathbf{X}_i \boldsymbol{\beta} + \mathbf{N}_i \mathbf{f}, \delta \mathbf{Z}_i \mathbf{Z}_i^t + \phi \mathbf{C}_i^{-1} \mathbf{C}_i^{-t}) \quad (6.4)$$

Bajo este supuesto, el vector de parámetros a estimar está dado por $\boldsymbol{\theta} = [\beta \ \mathbf{f}^t \ \delta \ \phi \ \rho]^t$, con el propósito de comparar el ajuste del modelo para algunas distribuciones de la familia elíptica, se consideraron específicamente las distribuciones normal, t -Student y Exponencial potencia (Expp) multivariadas.

6.1.2 Ajuste del MMS elíptico

En esta sección se describen los resultados de los ajustes obtenidos de los MMS con errores AR(1) a nivel intra-unidad utilizando el criterio de máxima verosimilitud penalizada bajo las distribuciones normal, t -Student y Expp, que en el Cuadro 6.2 se describen los valores de SIC, AIC y las función de log-verosimilitud penalizada de los tres MMS elípticos estimados, presentándose la menor pérdida de información en el MMS Expp, además, los grados de libertad y el parámetro de forma fueron escogidos mediante el criterio de información de Schwartz obteniendo 21 grados de libertad y 0,45 como parámetro de forma, cuyo comportamiento de la función de log-verosimilitud penalizada puede ser visualizado en la Figura 6.2.

Por otra parte, el parámetro de suavizamiento fue estimado a partir de un MMS normal, obteniendo $\hat{\alpha} = 25,7$, cuyo comportamiento de la función de log-verosimilitud penalizada para distintos valores de α se muestra en la Figura 6.3 y los errores estándar (EE) asintóticos de los parámetros fueron estimados a partir de la matriz de información de Fisher penalizada.

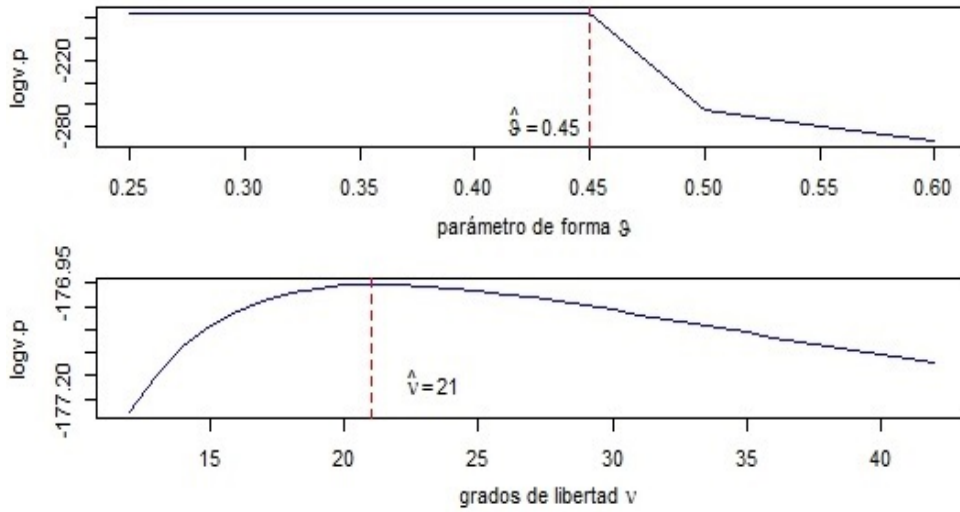


Figura 6.2: Función de log-verosimilitud penalizada estimada según parámetro de forma del MMS Exponencial potencia y grados de libertad del MMS t -Student

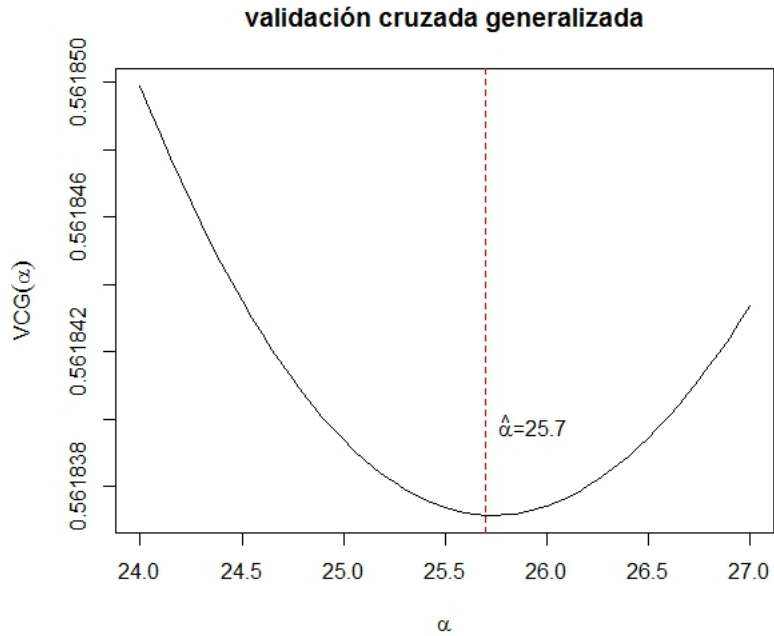


Figura 6.3: Valores estimados de validación cruzada generalizada para el MMS normal según valores predeterminados del parámetro de suavizamiento

Asimismo, las estimaciones de los parámetros asociados a los tres MMS elípticos se muestran en el Cuadro 6.3, donde las estimaciones de los parámetros de la parte paramétrica y no paramétrica son similares entre los tres MMS, no así con las estimaciones de los componentes de varianza, ya que las estimaciones $\hat{\delta}$ y $\hat{\phi}$ son menores en el MMS Expp. Respecto a los errores estándar asintóticos de $\hat{\mathbf{f}}$, éstos tienden a ser mucho más bajos en el MMS Expp, así como también, el error estándar asintótico asociado al efecto del grupo es considerablemente menor en el MMS Expp en comparación de los demás modelos. De esta manera, los resultados nos indican que la pérdida de información es menor bajo el MMS Expp y de este modelo se obtienen en general estimaciones mucho más precisas de los parámetros.

Cuadro 6.3: EMVP de los parámetros del MMS según distribución normal, $t(21)$ y $\text{Expp}(0,45)$

θ	Normal		t -Student		Exp. potencia	
	$\hat{\theta}$	$\text{EE}(\hat{\theta})$	$\hat{\theta}$	$\text{EE}(\hat{\theta})$	$\hat{\theta}$	$\text{EE}(\hat{\theta})$
β	0,4053	0,1871	0,3956	0,1905	0,3977	0,0055
$f(t_1^0)$	4,0339	0,1616	4,0311	0,1636	4,0353	4,778e-03
$f(t_2^0)$	3,5837	0,1548	3,5881	0,1571	3,5927	4,777e-03
$f(t_3^0)$	3,2198	0,1544	3,2291	0,1567	3,2330	4,776e-03
$f(t_4^0)$	2,9989	0,1550	3,0064	0,1573	3,0107	4,776e-03
$f(t_5^0)$	2,9285	0,1568	2,9313	0,1591	2,9358	4,777e-03
$f(t_6^0)$	3,0852	0,1590	3,0928	0,1612	3,0955	4,778e-03
$f(t_7^0)$	3,3710	0,1593	3,3917	0,1614	3,3948	4,778e-03
$f(t_8^0)$	3,7217	0,1625	3,7399	0,1645	3,7397	4,778e-03
δ	0,1809	0,0752	0,1779	0,0759	0,0423	0,0180
ϕ	0,1768	0,0169	0,1598	0,0182	0,0382	0,0043
ρ	0,5855	0,0812	0,5962	0,0840	0,5951	0,0839

En la Figura 6.4 se muestran los perfiles de fosfato inorgánico de plasma, cuya estimación es $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{N}\hat{\mathbf{f}} + \mathbf{Z}\hat{\mathbf{b}}$ con $\hat{\mathbf{b}}$ es el estimador empírico de bayes evaluado en el EMVP de $\boldsymbol{\theta}$. Además, se aprecia que el nivel de fosfato inorgánico de plasma estimado se comporta de manera razonablemente similar al nivel de fosfato medido a los pacientes en estudio, capturando su comportamiento no lineal a través de las horas de observación.

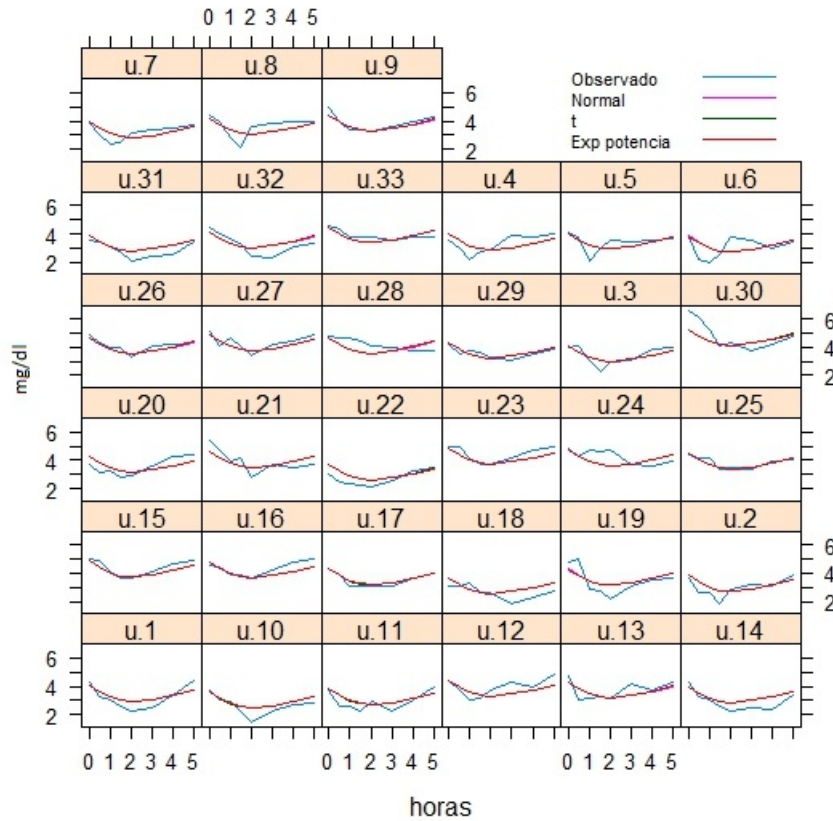


Figura 6.4: Perfiles de fosfato inorgánico de plasma observado y estimado a través de los tiempos de observación

En comparación con reportes anteriores de este conjunto de datos, los autores Chi & Reinsel (1989) e Ibacache *et al.*(2012) obtienen resultados similares en la EMVP del MMS elíptico, no obstante, Chi & Reinsel (1989) consideran una estructura AR(1) a nivel intra-unidad pero no consideran la parte no paramétrica, puesto que el tiempo es modelado por medio de una función lineal por partes con un punto de cambio de dos horas, por otra parte, Ibacache *et al.*(2012) si considera la parte no paramétrica en función del tiempo, pero no se incorporó una estructura de autocorrelación en el tiempo de medición a nivel intra-paciente.

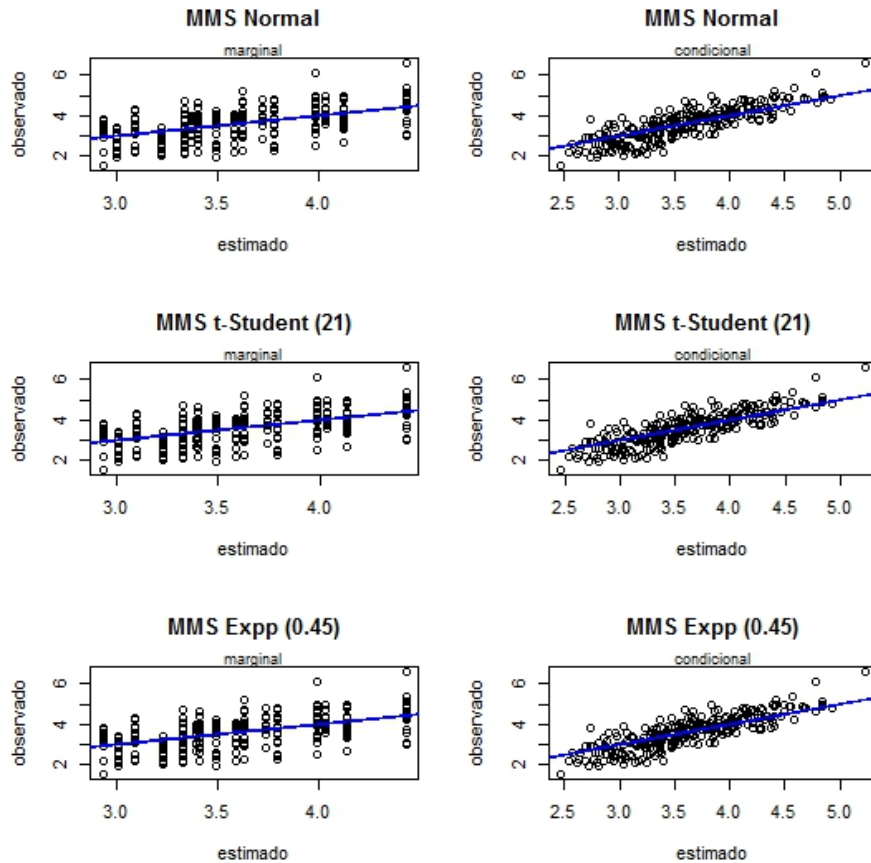


Figura 6.5: Mediciones de fosfato inorgánico de plasma observado vs estimado

Por último, en la Figura 6.5, se aprecia que el comportamiento de las estimaciones vs observaciones son semejantes entre los distintos MMS elípticos estimados, sin embargo, note que que las estimaciones del MMS elíptico tienden a acercarse a la línea de referencia de ajuste perfecto cuando se incorpora un efecto aleatorio asociado a las observaciones medidas al mismo paciente. A continuación se describen los resultados del análisis de diagnóstico de influencia global del MMS por medio de la distancia de Cook generalizada y las distancias del valor estimado estandarizado de las mediciones de fosfato inorgánico de plasma.

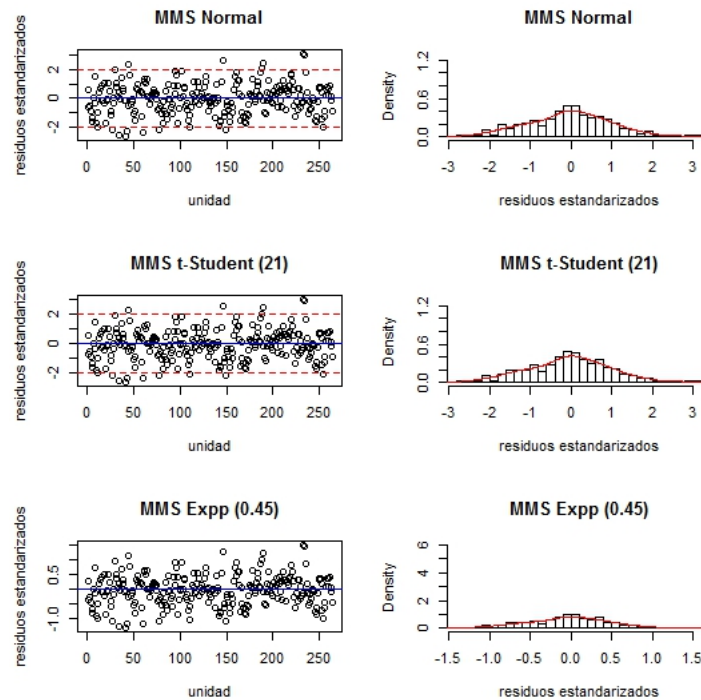


Figura 6.6: Gráficos de residuos estandarizados de los MMS con distribución elíptica normal, t -Student y Exponencial potencia

6.1.3 Análisis diagnóstico del modelo

En esta sección, el análisis diagnóstico del MMS estará enfocado en el estudio de la distancia de Cook generalizada (DCG), la distancia del valor estimado estandarizado (DFITS) y del comportamiento de los residuos estandarizados por medio de gráficos de Cuartil-Cuartil³ que están basados en los residuos condicionales dados por la expresión (4.4).

Respecto al comportamiento de los residuos estandarizados, se puede ver de la Figura 6.6 que dichos residuos mayormente fluctúan entre -2 y 2 cuando la distribución subyacente de los errores del MMS es normal y t -Student, pero en el caso de que la distribución subyacente de los errores del MMS es exponencial potencia, los residuos varían entre -1 y 1, por lo cual, se

³del inglés QQplot

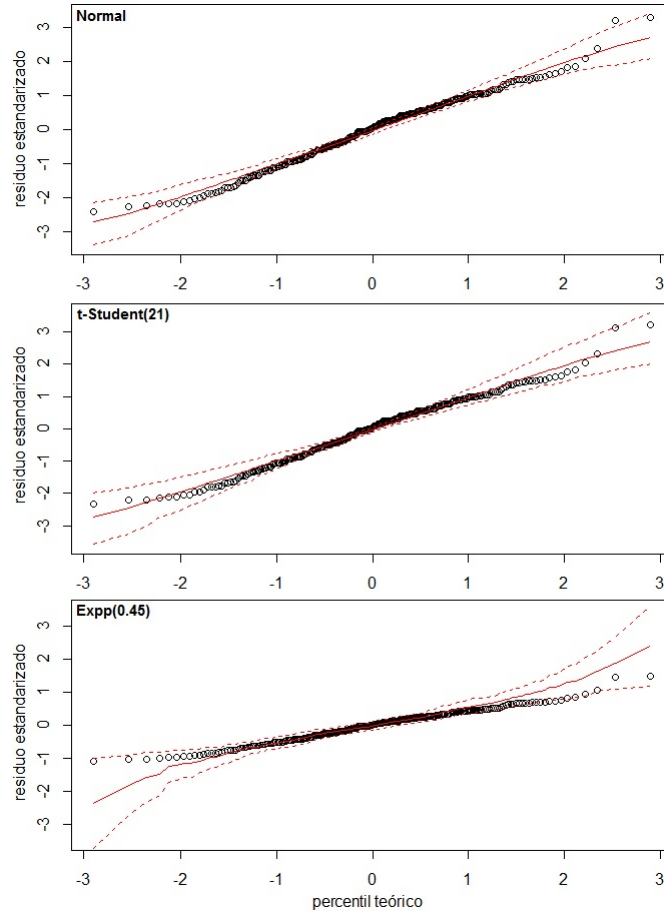


Figura 6.7: Gráfico QQplot para los residuos del MMS bajo distribución: Normal, t -Student y Exponencial potencia

comportan de mejor manera en este último MMS. De la Figura 6.7, se observa que los residuos estandarizados se comportan de mejor manera para el MMS t -Student, puesto que se ajusta mucho mejor en las colas de la distribución normal y exponencial potencia.

Para el estudio del grado de influencia global sobre las estimaciones de los parámetros asociados a los efectos fijos del MMS elíptico fueron utilizados la distancia de Cook generalizada para la componente paramétrica, basado en la expresión (4.12), y la distancia del valor estimado estandarizado para la componente no paramétrica, basado en la expresión (4.13).

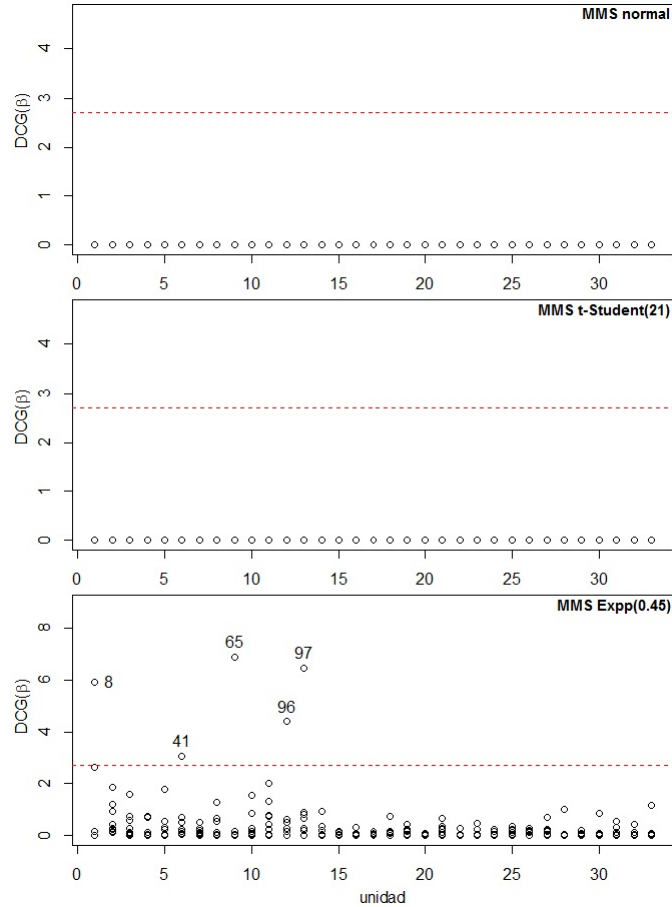


Figura 6.8: Gráfico de distancia de Cook generalizada del MMS bajo distribución: Normal, t -Student y Exponencial potencia

Respecto a la influencia global de la componente paramétrica, en la figura 6.8 se muestran los valores estimados de la $DCG(\beta)$, donde se aprecia que el comportamiento de la DCG en los MMS normal y t -Student es estable bajo el límite de confianza, sin embargo, se observan cambios sobre la estimación del parámetro de la parte paramétrica al extraer algunas observaciones en el MMS Expp y dichas distancias superan el valor de referencia $\chi_{0,9}^2(1) = 2.705543$, donde estas observaciones corresponden a mediciones realizadas a los pacientes n°1, n°6, n°9, n°12 y n°13.

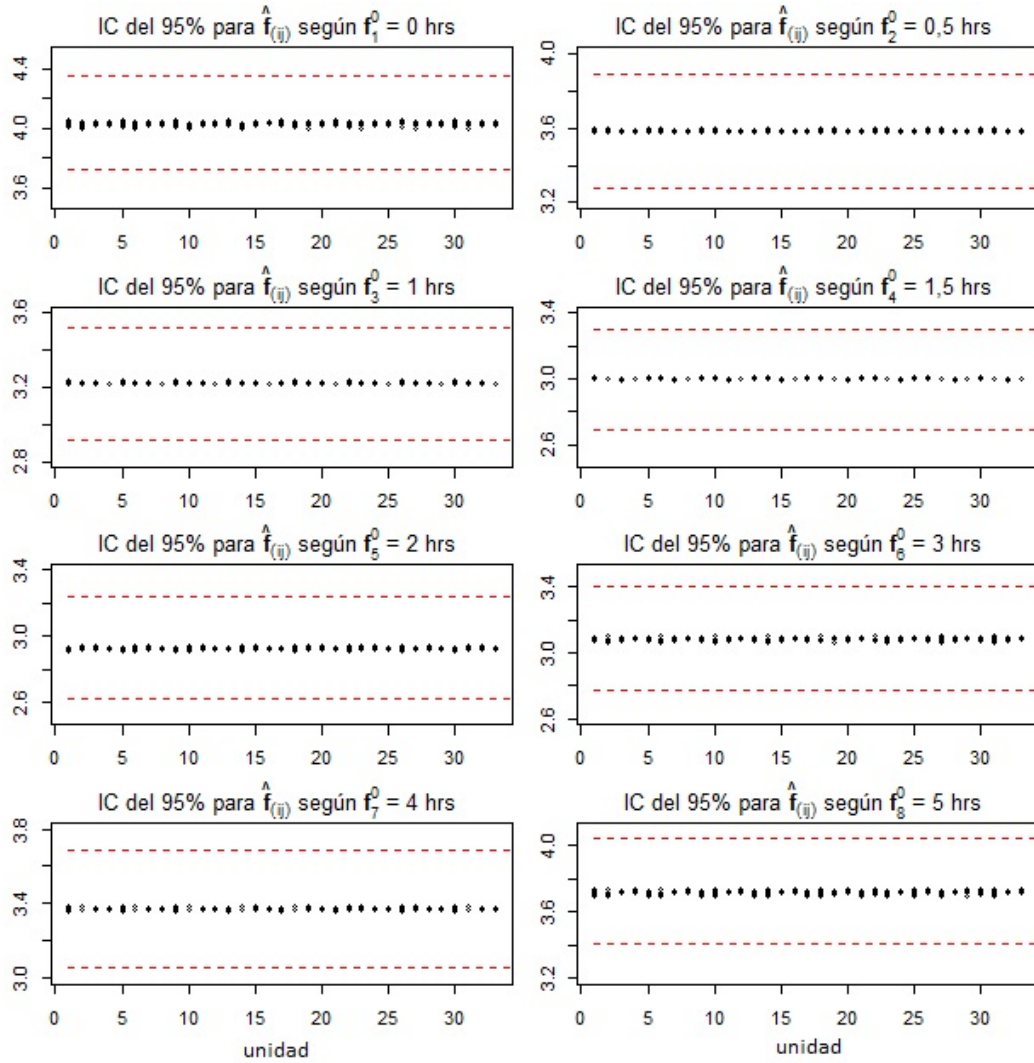


Figura 6.9: Gráfico de IC del 95% para la estimación de la componente no paramétrica del MMS normal

Respecto a la influencia global de la componente no paramétrica, en las Figuras 6.9 y 6.10 se muestran los parámetros estimados $\hat{f}_{(ij)}$ para el MMS normal y t -Student, donde se observa que las estimaciones de los parámetros de la componente no paramétrica sin la (i, j) -ésima observación en todos los tiempos se comportan de manera estable dentro de los límites de confianza.

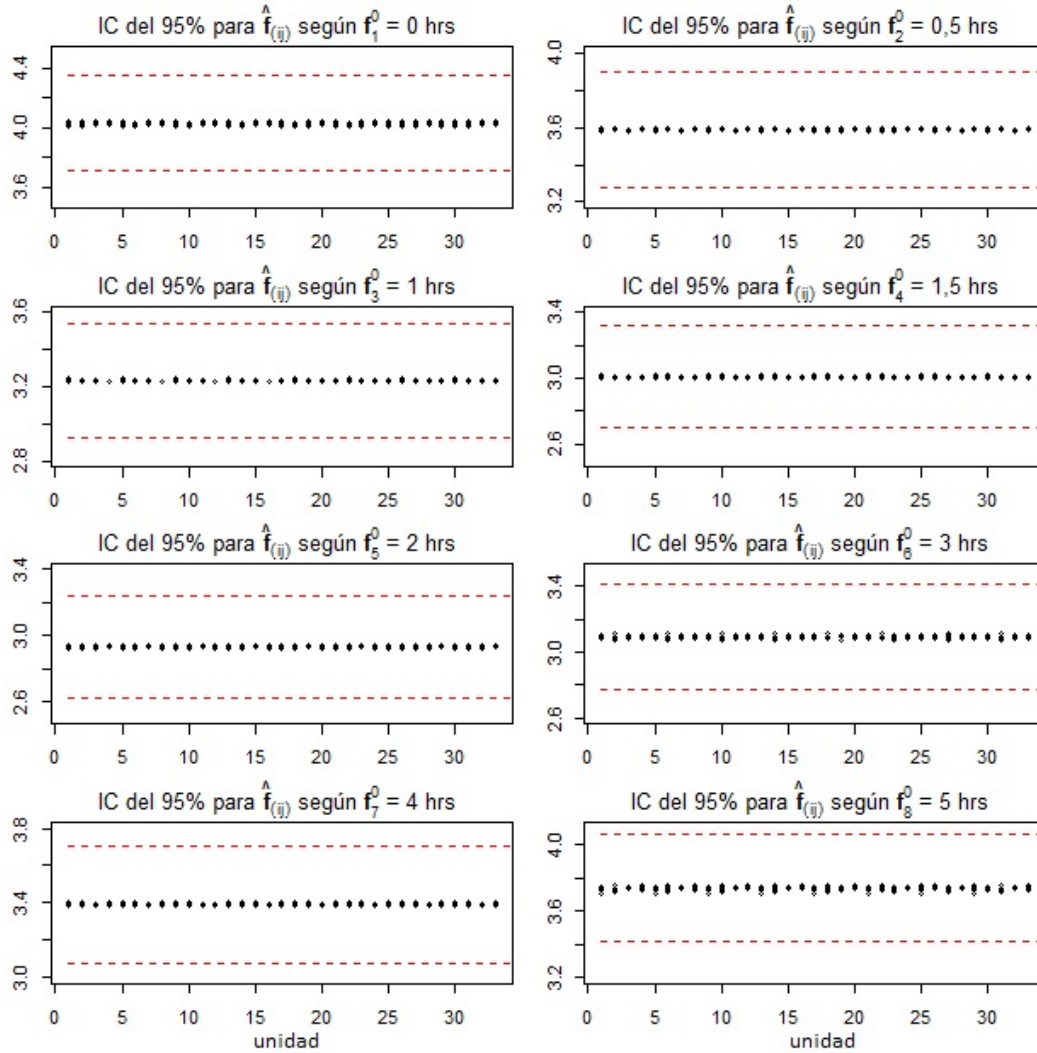


Figura 6.10: Gráfico de IC del 95% para la estimación de la componente no paramétrica del MMS t -Student

Sin embargo, en la Figura 6.11 se muestra los parametros estimados $\hat{f}_{(ij)}$ para el MMS Expp, el cual no se observa en general un comportamiento estable dentro de los límites de confianza, aparentemente las estimaciones de los parámetros de la componente no paramétrica sin la (i, j) -ésima observación a las 0,5 horas se comporta de manera estable, pero también se presentan estimaciones sobre los límites de confianza superior.

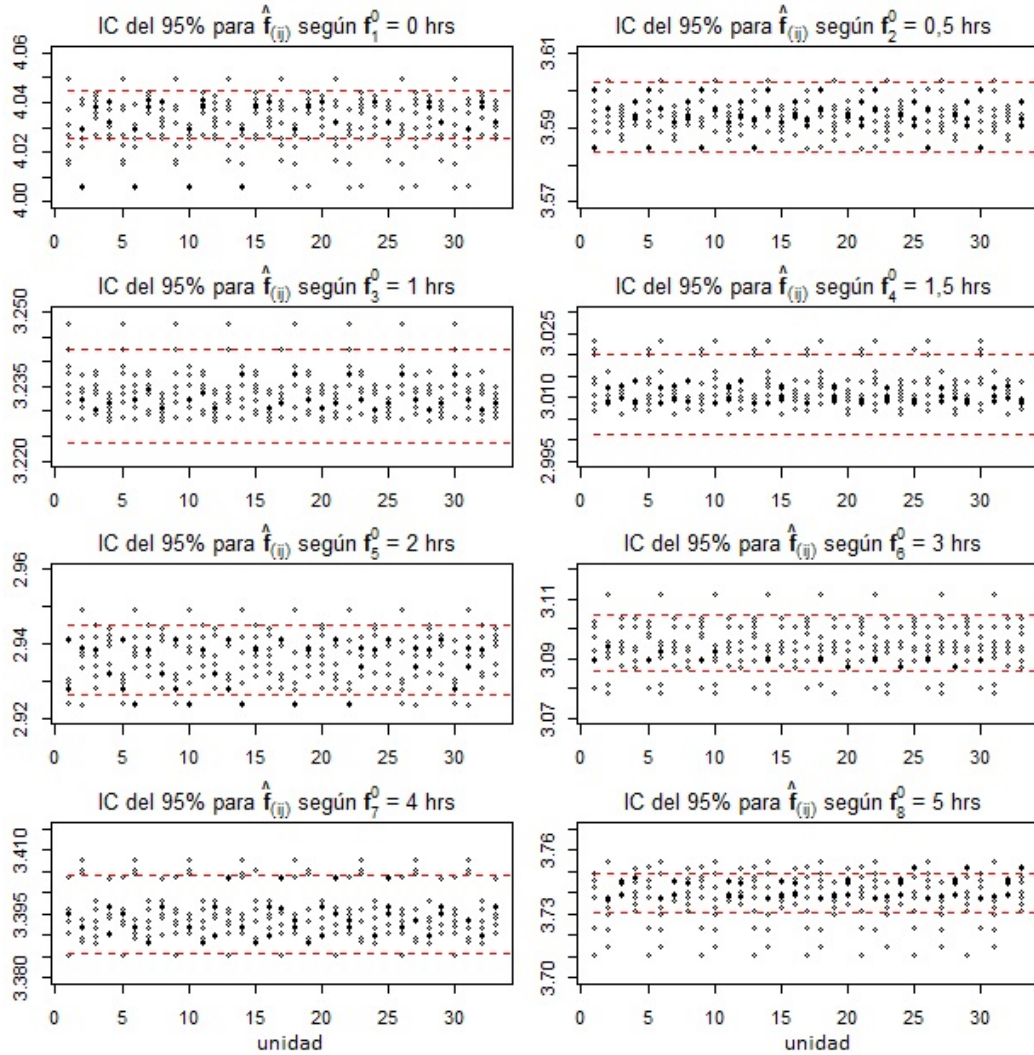


Figura 6.11: Gráfico de IC del 95% para la estimación de la componente no paramétrica del MMS Exponencial potencia

En cambio, las estimaciones de los parámetros de la componente no paramétrica en las demas horas de observación no se comportan de manera estable dentro de los límites de confianza, siendo que a las 1 y 1,5 horas sólo se presentan estimaciones que están sobre el limite de confianza, el cual nos indica que al extraer algunas observaciones, el valor de la función no paramétrica en dichos tiempo tienden a ser sobreestimados.

De esta manera, se pudo observar que las estimaciones de los parámetros de la componente paramétrica y no paramétrica de los MMS normal y t -Student no se vieron afectados significativamente bajo la eliminación de una observación, pero en el MMS Expp si se observaron cambios significativos en la estimación de los parámetros de ambas componentes, presentándose mayores cambios en la estimación de la componente no paramétrica. Sin embargo, no podemos indicar que estamos bajo la presencia de observaciones influyentes, solo podemos concluir que las estimaciones de dichos parámetros en el MMS Expp son afectados significativamente cuando se suprime una observación en el proceso de estimación máximo verosímil penalizado.

6.1.4 Análisis de autocorrelación de primer orden

Para la inferencia de la presencia y homogeneidad de autocorrelación en los errores del MMS elíptico a nivel intra-unidad, se basaron en los resultados obtenidos por los test de score descritos en el Capítulo 5, el parámetro ρ está asociado a la autocorrelación de primer orden del nivel de fosfato inorgánico de plasma en la sangre y ρ_i es el parámetro de la estructura AR(1) del nivel de fosfato inorgánico de plasma en la sangre perteneciente al i -ésimo paciente.

Cuadro 6.4: Test de Score para $H_0 : \rho = 0$ y EMVP de los componentes de varianza para los MMS normal, $t(21)$ y Expp(0,45)

MMS elíptico	Componentes de varianza		$l_p(\hat{\theta}, \hat{\alpha})$	Test Score bajo H_0	
	$\hat{\delta}$	$\hat{\phi}$		SC _a	p-valor
Normal	0,2217	0,2332	-204,5658	55,1936	1,09e-13
t -Student	0,2132	0,2065	-202,2627	56,9074	4,57e-14
Exp. potencia	0,0514	0,0500	-202,1206	55,6583	8,62e-14

Se puede apreciar del Cuadro 6.4, las estimaciones de los parámetros asociados a los componentes de varianza entre los MMS normal y t -Student son similares, pero en el caso del MMS Expp, las estimaciones de los parámetros de la variabilidad del efecto y error aleatorio son menores que las estimaciones obtenidas en los MMS normal y t -Student, además, la función de log-verosimilitud penalizada es maximizada en el MMS Expp.

Considerando los resultados de la función de log-verosimilitud estimada en el Cuadro 6.2, note que las estimaciones de los parámetros bajo $H_0 : \rho = 0$ son mucho más bajas que los valores de dicha función cuando no se tiene alguna restricción sobre el parámetro de autocorrelación, lo cual nos indica una mayor probabilidad de la existencia de autocorrelación según los tres MMS evaluados. Asimismo, el valor del test de Score para la presencia de autocorrelación es mayor bajo el MMS Expp, y así se obtiene evidencia significativa de presencia de autocorrelación, incluso utilizando los otros dos MMS elípticos se obtiene la misma conclusión, en otras palabras, se detectó un efecto de autocorrelación de primer orden estadísticamente significativo en las mediciones de fosfato inorgánico de plasma a través de los tiempos de observación.

Cuadro 6.5: Test de Score para $H_0 : \lambda = 0$ y EMVP de los componentes de varianza para los MMS normal, $t(21)$ y $\text{Expp}(0,45)$

MMS elíptico	Componentes de varianza			$l_p(\hat{\theta}, \hat{\alpha})$	Test Score bajo H_0	
	$\hat{\delta}$	$\hat{\phi}$	$\hat{\rho}$		SC_h	p-valor
Normal	0,1808	0,1769	0,5102	-171,6053	15,73	7,31e-05
t -Student	0,1761	0,1602	0,5153	-171,0545	16,225	5,63e-05
Exp. potencia	0,0421	0,0384	0,5259	-170,4793	14,367	1,5e-04

En cambio, se puede observar del Cuadro 6.5 que las estimaciones de los parámetros de los componentes de varianza entre los MMS normal y t -Student son similares, no así en el MMS Expp, además, la función de log-verosimilitud penalizada es maximizada en dicho MMS, la cual es mayor que la función de log-verosimilitud penalizada sin restricciones. Respecto al test de score para la homogeneidad de la autocorrelación a nivel intra-unidad, éste es menor bajo el MMS Expp, esto significa que el nivel de autocorrelación de primer orden se comporta de manera estadísticamente diferente en las mediciones de fosfato inorgánico de plasma a través del tiempo de seguimiento. Cabe de señalar que las estimaciones de los parámetros del MMS elíptico utilizando una forma funcional logística para la modelación de la estructura AR(1) intra-unidad, coinciden con las estimaciones obtenidas del MMS elíptico empleando una forma funcional Gompertz, es por esto que sólo se muestra una sola tabla con dichos resultados.

Note que para cuantificar el efecto de la obesidad sobre el nivel de fosfato inorgánico de plasma, la presencia de obesidad en el paciente fue considerada como una covariable para la componente paramétrica del MMS elíptico y como covariable para la modelar la estructura AR(1) a nivel intra-unidad.

6.1.5 Conclusiones del análisis de datos

De los resultados obtenidos, se observó que los niveles de fosfato inorgánico de plasma en la sangre de los pacientes tienen un comportamiento a través del tiempo de seguimiento de forma no lineal, la cual es modelada por la componente no paramétrica del MMS elíptico. Además, de los tres MMS elípticos estimados, las observaciones se ajustan de mejor manera al MMS Expp en el sentido de la maximización de la función de log-verosimilitud penalizada, minimización del SIC y AIC, comportamiento de los residuos estandarizados y estimaciones con errores estándar mas bajos, sobre todo en la estimación del error estándar de la componente no paramétrica.

Respecto al análisis de influencia global, en los MMS normal y t -Student no se detectan cambios significativos en la estimación de la curva no paramétrica bajo la supresión de una observación, pero en el caso del MMS Expp si se detectan cambios significativos, aunque es razonable considerar que al sustraer una observación para la estimación de \mathbf{f} se produzcan cambios en la estructura de varianza a nivel intra-unidad, además los errores estándar estimados de la componente no paramétrica son muy pequeños. Sin embargo, se recomienda para futuros análisis de este conjunto de datos, es realizar un análisis de influencia local introduciendo perturbaciones sobre la componente no paramétrica bajo un MMS Expp.

Por último, utilizando los tres MMS elípticos evaluados se obtiene evidencia significativa de presencia de autocorrelación de primer orden sobre los niveles de fosfato inorgánico de plasma y que dicha autocorrelación depende de la condición de obesidad del paciente, ya sea como un efecto aditivo asociado a la parte paramétrica, como también del comportamiento de la estructura de AR(1) a nivel intra-unidad.

6.2 Conjunto de datos N°2: índices Odontológicos

En el área de Odontología periodontal, en el año 2010 fue diseñado un experimento dirigido por la Dra. Jessica Rojas⁴, Dr. Jorge Godoy O.⁵ y el presente autor, posteriormente, en el año 2011 se realizó dicho experimento cuyo objetivo fue cuantificar el nivel de crecimiento de un conjunto de características en la encía de los pacientes durante los primeros 21 días después de una limpieza periodontal. Este conjunto de características se estudia a través de tres índices odontológicos periodontales dados por:

- Índice de placa bacteriana de O'Leary (IP): Este índice mide el porcentaje de biofilm bacteriano de todos los dientes en boca.
- Profundidad de sondaje (PS): Esta corresponde a la profundidad máxima -en cada uno de los dientes- del surco gingival definido por el espacio circundante del diente que forman la superficie dental.
- Índice de sangrado gingival de Ainamo y Bay [1975] (IH): Este indicador determina el porcentaje de hemorragia al sondaje periodontal de todos los dientes en boca.

Asimismo, se consideraron el género y la edad de cada uno de los 41 pacientes que participaron en el estudio de manera voluntaria, los cuales corresponden a 27 mujeres y 14 varones, donde previamente todos firmaron un consentimiento informado. Además, se definieron los siguientes grupos de estudio que son:

- Grupo CHX: Los pacientes de este grupo se sometieron a una limpieza periodontal inicial y se enjuagaron la boca con un colutorio en base a digluconato de Clorhexidina (CHX) al 0,06%.
- Grupo CPC: Los pacientes de este grupo se sometieron a una limpieza periodontal inicial y se enjuagaron la boca con un colutorio en base a cloruro de cetilpiridinio (CPC) al 0,05%.
- Grupo Control: Los pacientes de este grupo se sometieron a una limpieza periodontal inicial y se enjuagaron la boca con un colutorio placebo.

⁴Facultad de Odontología de la Universidad Andrés Bello, Viña del Mar

⁵Facultad de Odontología de la Universidad de Valparaíso, Valparaíso

Cuadro 6.6: Medidas descriptivas para la profundidad de sondaje según grupo y sexo de los pacientes

PS	Sexo		Grupo de estudio			Total
	Femenino	Masculino	CHX	CPC	Control	
Min	1,50	1,25	1,50	1,50	1,25	1,25
Q1	2,00	2,00	2,00	2,00	2,00	2,00
Mediana	2,25	2,25	2,25	2,25	2,00	2,25
Promedio	2,28	2,24	2,33	2,29	2,18	2,27
Q3	2,50	2,50	2,50	2,50	2,50	2,50
Máx	3,00	3,00	3,00	3,00	3,00	3,00
Desv. Est.	0,35	0,38	0,33	0,34	0,39	0,36

donde se procedió a seleccionar pacientes de manera aleatoria para la conformación de cada uno de estos grupos. Cabe de señalar que, el principal objetivo es modelar la PS en función del IP e IH, cuyas unidades de medida son porcentuales para IP e IH, en cambio, la PS fue medida en milímetros (mm) y éstas son presentadas en el Apéndice VI.

Para el estudio del comportamiento general de la profundidad de sondaje según género y grupo de estudio se describen las medidas descriptivas en el Cuadro 6.6. Además, de la Figura 6.7 se aprecia que, la PS promedio a través de los tiempos de observación posee un comportamiento no lineal, cuyos tiempos corresponden a los 0, 7, 14 y 21 días después de la intervención. Debido a esto, se propone modelar la respuesta media de cada paciente en el tiempo a través de una función no paramétrica.

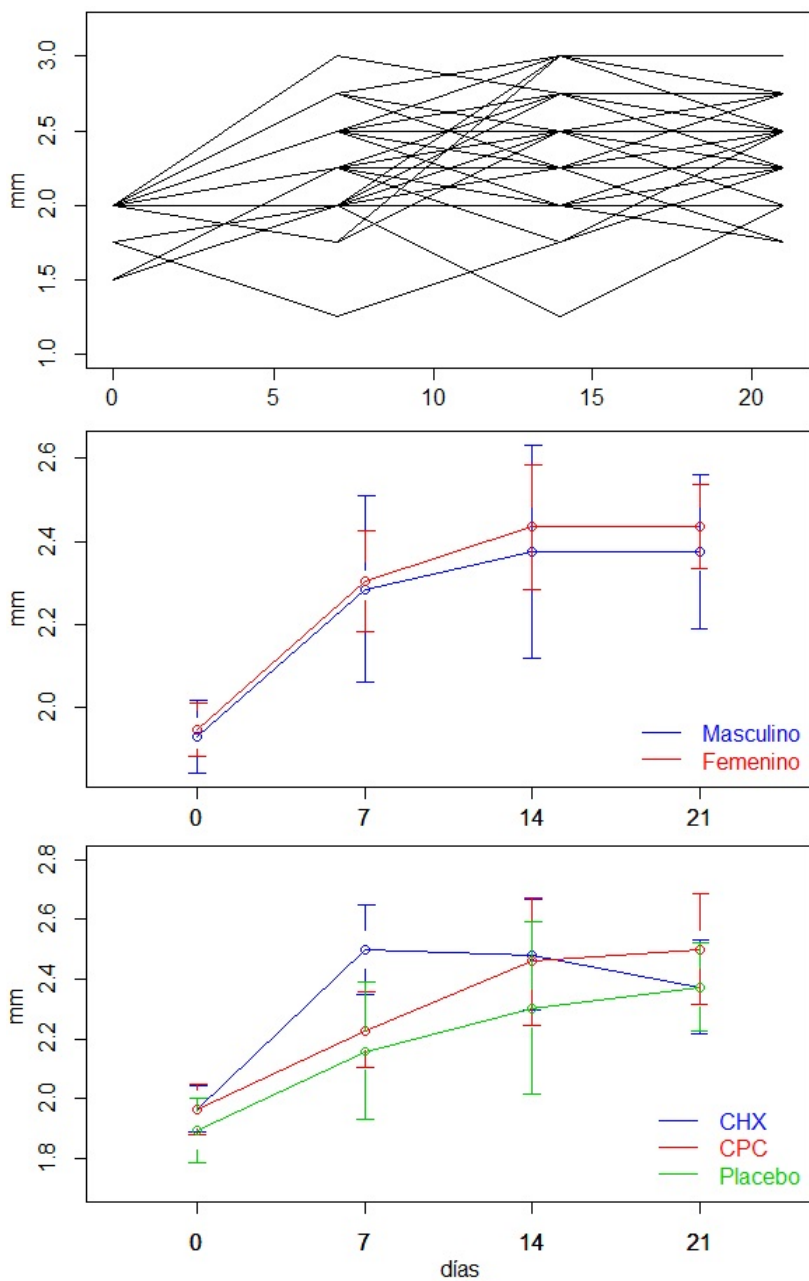


Figura 6.12: Gráficos descriptivos: perfiles de mediciones de profundidad de sondaje para cada paciente a través del tiempo, IC del 95% para la media de la PS según sexo y grupo de estudio

6.2.1 Modelo propuesto

Puesto que la variable respuesta PS se comporta de manera no lineal a través de los tiempos de observación, se propone el siguiente modelo:

$$y_{ij} = \sum_{k=1}^6 x_{ijk} \beta_k + f(t_{ij}) + b_i + \varepsilon_{ij} \quad (6.5)$$

donde y_{ij} es la j -ésima observación de la PS del i -ésimo paciente obtenida al tiempo t_{ij} , y la función f es una función arbitraria que sólo depende de los tiempos de medición y dicha función se supone continua y diferenciable, $x_{ijk} \in \mathbb{R}$ es la k -ésima variable explicativa asociada a la PS del i -ésimo paciente en el tiempo t_{ij} , $b_i \in \mathbb{R}$ es el efecto aleatorio asociado al i -ésimo paciente en estudio y ε_{ij} corresponde al error aleatorio asociado a la j -ésima observación medida al i -ésimo paciente, para $i = 1, \dots, 41$ y $j = 1, 2, 3, 4$. El modelo (6.5) puede ser reescrito de forma matricial de la siguiente manera:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{N}_i \mathbf{f} + \mathbf{Z}_i b_i + \boldsymbol{\varepsilon}_i \quad (6.6)$$

donde $\mathbf{y}_i, \boldsymbol{\varepsilon}_i \in \mathbb{R}^8$ son los vectores de respuestas y errores aleatorios asociados al i -ésimo paciente, $\mathbf{Z}_i \in \mathbb{R}^4$ es un vector de diseño de la parte aleatoria el cual contiene sólo valores iguales a uno, esto es: $\mathbf{Z}_i = \mathbf{1}_4$. Asimismo, $\mathbf{N}_i \in \mathbb{R}^{4 \times 4}$ es una matriz de incidencia con $\mathbf{N}_i = \mathbb{I}_4$ y \mathbf{f} corresponde a un vector cuyos elementos son los 4 tiempos diferentes evaluados en la función f con $t_1^0 = 0, t_2^0 = 7, t_3^0 = 14, t_4^0 = 21$. Además, la matriz de diseño de la parte paramétrica $\mathbf{X}_i \in \mathbb{R}^{4 \times 6}$ corresponde a $\mathbf{X}_i = [\mathbf{x}_{i1} \ \mathbf{x}_{i2} \ \cdots \ \mathbf{x}_{i6}]$, la cual esta conformada por las siguientes variables:

- $\mathbf{x}_{i1} \in \mathbb{R}^4$ es un vector de variables indicadoras que corresponden a $x_{ij1} = 1$ cuando el i -ésimo paciente es mujer y $x_{ij1} = 0$ en caso contrario.
- $\mathbf{x}_{i2} \in \mathbb{R}^4$ corresponde a la edad del i -ésimo paciente.
- $\mathbf{x}_{i3} \in \mathbb{R}^4$ es un vector de variables indicadoras que corresponden a $x_{ij3} = 1$ cuando el i -ésimo paciente pertenece al grupo CHX y $x_{ij3} = 0$ en caso contrario.
- $\mathbf{x}_{i4} \in \mathbb{R}^4$ es un vector de variables indicadoras que corresponden a $x_{ij4} = 1$ cuando el i -ésimo paciente pertenece al grupo CPC y $x_{ij4} = 0$ en caso contrario.

- \mathbf{x}_{i5} es un vector que corresponde a las mediciones del IH del i -ésimo paciente a través del tiempo.
- \mathbf{x}_{i6} es un vector que corresponde a las mediciones del IP del i -ésimo paciente a través del tiempo.

Asimismo, al considerar las 164 observaciones realizadas a los 34 pacientes ($m = 164$), tenemos que el MMS puede ser escrito como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \quad (6.7)$$

donde $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_6]^t$, $\mathbf{N} = [\mathbb{I}_4 \ \mathbb{I}_4 \ \dots \ \mathbb{I}_4]^t$, $\mathbf{b} \in \mathbb{R}^{41 \times 1}$ y $\mathbf{Z} = \text{diag}(\mathbf{1}_4, \mathbf{1}_4, \dots, \mathbf{1}_4)$. De esta manera, al asumir que la variable respuesta y los efectos aleatorios provienen de una distribución elíptica conjunta dada por:

$$\begin{bmatrix} \mathbf{y}_i \\ b_i \end{bmatrix} \sim \text{EC}_5 \left(\begin{bmatrix} \boldsymbol{\mu}_i \\ 0 \end{bmatrix}, \begin{bmatrix} \delta \mathbf{Z}_i \mathbf{Z}_i^t + \phi \mathbf{C}_i^{-1} \mathbf{C}_i^{-t} & \delta \mathbf{Z}_i \\ \delta \mathbf{Z}_i^t & \delta \end{bmatrix} \right) \quad (6.8)$$

por lo cual tenemos que el vector de respuestas sigue una distribución marginal elíptica de la forma:

$$\mathbf{y}_i \sim \text{EC}_4 (\mathbf{X}_i \boldsymbol{\beta} + \mathbf{N}_i \mathbf{f}, \delta \mathbf{Z}_i \mathbf{Z}_i^t + \phi \mathbf{C}_i^{-1} \mathbf{C}_i^{-t}) \quad (6.9)$$

Bajo este supuesto, el vector de parámetros a estimar está dado por $\boldsymbol{\theta} = [\boldsymbol{\beta} \ \mathbf{f}^t \ \delta \ \phi \ \rho]^t$ donde $\boldsymbol{\beta} \in \mathbb{R}^{6 \times 1}$, $\mathbf{f} \in \mathbb{R}^{4 \times 1}$, $\delta \in \mathbb{R}$, $\phi \in \mathbb{R}$ y $\rho \in \mathbb{R}$. Con el propósito de comparar el ajuste del modelo para algunas distribuciones de la familia elíptica, se consideraron específicamente las distribuciones normal, t -Student y Exponencial potencia multivariadas.

6.2.2 Ajuste del MMS elíptico

De esta manera, fueron ajustados los MMS elípticos con errores AR(1) a nivel intra-unidad utilizando el criterio de máxima verosimilitud penalizada, para esto se estimó el parámetro de suavizamiento del modelo a partir de un MMS normal, donde no se obtuvo convergencia, por lo cual se fijó dicho parámetro en $\hat{\alpha} = 0,05$. Los grados de libertad y el parámetro de forma de las ditribuciones t -Student y Exponencial potencia fueron escogidos mediante el criterio de información de Schwartz, obteniendo 17 grados de libertad y

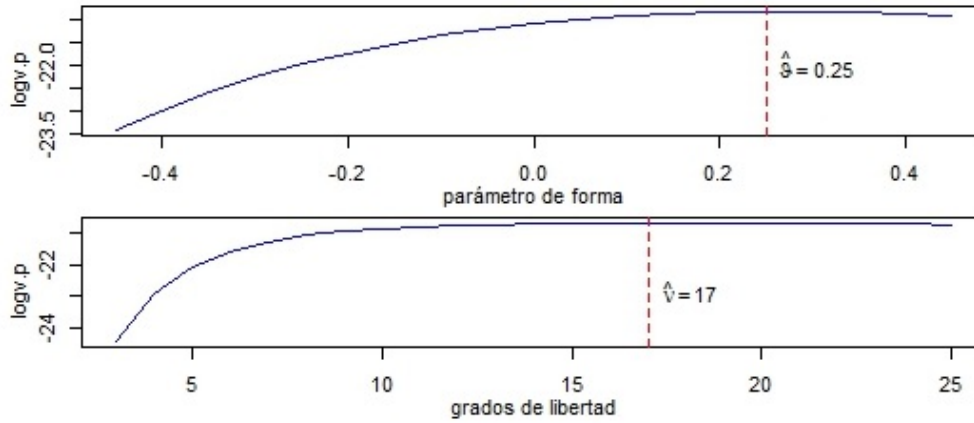


Figura 6.13: Función de log-verosimilitud penalizada estimada según parámetro de forma del MMS Exponencial potencia y grados de libertad del MMS t -Student

0,25 como parámetro de forma respectivamente, ver Figura 6.13. Asimismo, podemos observar del Cuadro 6.7 que la función de log-verosimilitud penalizada estimada es mayor bajo un MMS t -Student, cuyos valores de AIC y SIC son mínimos en comparación con los MMS estimados bajo las distribuciones normal y Expp, siendo claramente menor la pérdida de información en el MMS t -Student.

Cuadro 6.7: Valores estimados de la log-verosimilitud penalizada, AIC y SIC

Distribución	$l_p(\hat{\theta}, \hat{\alpha})$	SIC	AIC
Normal	-21,104	108,507	68,209
t -Student	-20,668	107,633	67,335
Exponencial potencia	-20,848	107,994	67,696

De la estimación máximo verosimil penalizada de los parámetros obtenidos, tanto en la parte paramétrica como la parte no paramétrica, se aprecia que dichas estimaciones poseen valores similares entre los tres MMS evaluados, pero obteniendo errores estándar (EE) asintóticos de $\hat{\beta}$ y \hat{f} mas bajos en el MMS Exponencial potencia, esto puede ser visto del Cuadro 6.8. Además, la curva no paramétrica estimada posee un comportamiento creciente a través del tiempo, y esto es válido en los tres MMS elípticos estimados. Respecto

Cuadro 6.8: EMVP de los parámetros asociados a los MMS normal, $t(21)$ y $\text{Expp}(0,25)$

θ	Normal		t -Student		Exp. potencia	
	$\hat{\theta}$	$EE(\hat{\theta})$	$\hat{\theta}$	$EE(\hat{\theta})$	$\hat{\theta}$	$EE(\hat{\theta})$
β_1	0,04595	0,06308	0,07153	0,06210	0,06422	0,00576
β_2	-0,06422	0,02344	-0,06040	0,02307	-0,06122	0,00214
β_3	0,08654	0,06839	0,07870	0,06731	0,08374	0,00625
β_4	0,04494	0,06963	0,03661	0,06855	0,04952	0,00636
β_5	0,00278	0,00302	0,00420	0,00298	0,00365	0,00028
β_6	-0,00305	0,00223	-0,00344	0,00219	-0,00328	0,00020
$f(t_1^0)$	2,77650	0,33186	2,69143	0,32665	2,70814	0,03033
$f(t_2^0)$	3,16810	0,33576	3,09139	0,33045	3,10955	0,03068
$f(t_3^0)$	3,28055	0,33482	3,19470	0,32954	3,21138	0,03060
$f(t_4^0)$	3,28607	0,33550	3,19338	0,33020	3,21378	0,03066
δ	0,00548	0,01036	0,00574	0,00925	0,00318	0,00529
ϕ	0,07134	0,01023	0,06270	0,01004	0,03587	0,00558
ρ	0,17913	0,14056	0,15781	0,14633	0,16431	0,14501

a los componentes de varianza, se presentaron estimaciones similares de la varianza del efecto aleatorio intra-unidad entre los MMS normal y t -Student, pero en el MMS Expp se obtuvo una varianza estimada menor.

En cambio, se obtuvo estimaciones de la varianza del error aleatorio y del parámetro de autocorrelación de primer orden mayores en el MMS normal, pero con mayores errores estándar estimados que los otros MMS, note que el parámetro de autocorrelación $\text{AR}(1)$ estimado en los tres modelos bordea entre 0,15 y 0,18 aproximadamente, lo cual, puede ser indicio de que las mediciones de profundidad de sondaje medidos en un tiempo específico dependa de la profundidad de sondaje observada en el tiempo de medición anterior. Cabe de señalar que los errores estándar asintóticos de las estimaciones de los parámetros asociados al MMS elíptico fueron calculados a partir de la inversa de la matriz de información de Fisher penalizada.

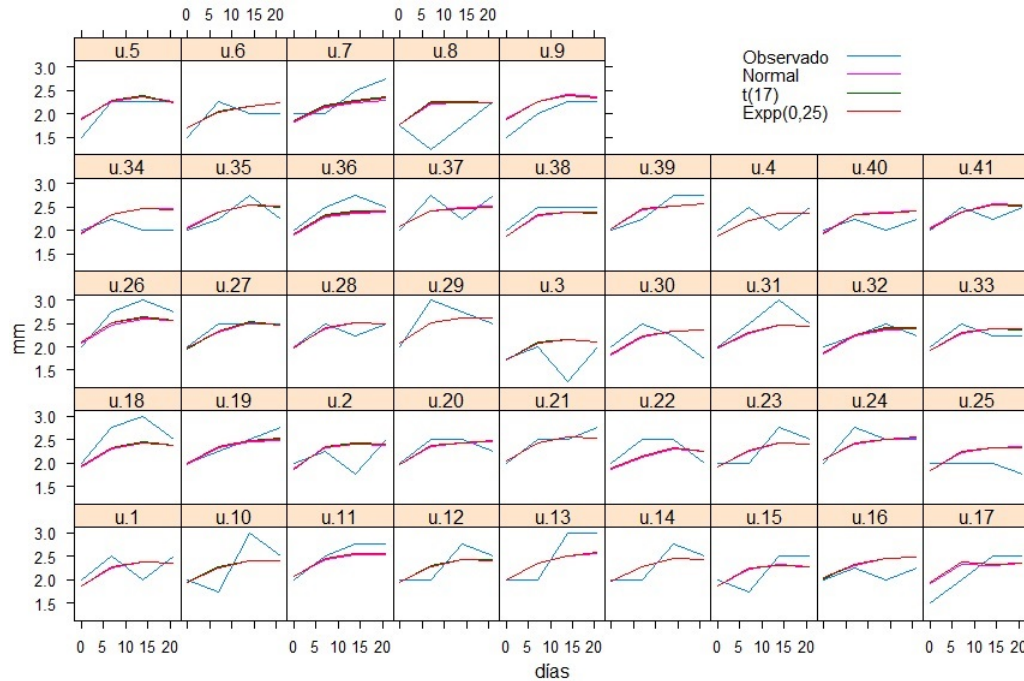


Figura 6.14: Perfiles de las profundidad de sondaje (mm) observado y estimado a través de los días de medición

De la Figura 6.14 se puede observar que las mediciones de PS estimados a través de los días de observación son similares entre los MMS elípticos evaluados, donde \hat{y} es la media condicional del MMS y \hat{b} es el estimador empírico de bayes evaluado en los EMVP bajo el MMS elíptico en cuestión. Cabe de mencionar que los pacientes n°2, 3, 8 10, 13, 15 y n°25 no siguieron el protocolo de estudio hasta los siete días después de iniciado el tratamiento, lo cual pudo afectar el comportamiento de la PS a través de los días de medición, pero dichos pacientes declararon haber seguido el protocolo posteriormente.

Por último, de la Figura 6.15 se aprecia que el comportamiento de las predicciones vs observaciones entre los distintos MMS elípticos estimados son semajante entre sí. No obstante, note que en general las estimaciones obtenidas tienden a la subestimación cuando se presenta valores de PS bajos alejandose de la línea de referencia de ajuste perfecto, además, las estima-

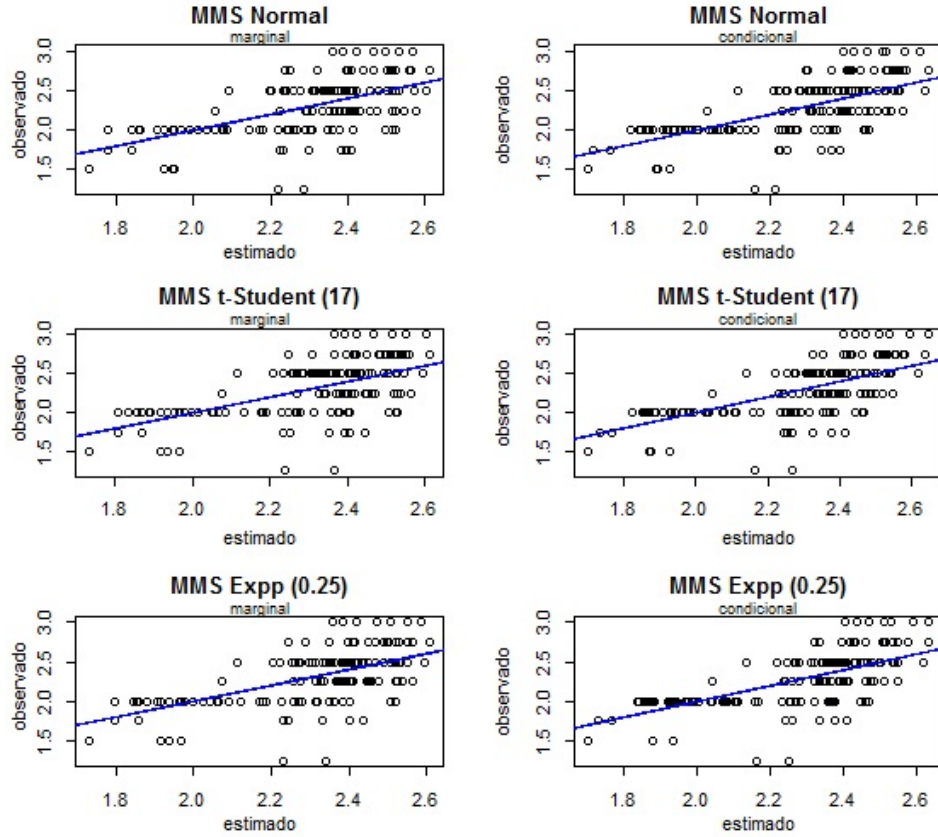


Figura 6.15: Mediciones de profundidad de sondaje observado vs estimado

ciones de la media marginal y condicional no cambian significativamente y esto puede ser debido a la baja variabilidad estimada del efecto aleatorio a nivel intra-unidad.

De esta manera, es razonable realizar un análisis de diagnóstico del MMS por medio de la distancia de Cook generalizada y las distancias del valor estimado estandarizado para el estudio del nivel de influencia global de las mediciones de profundidad de sondaje.

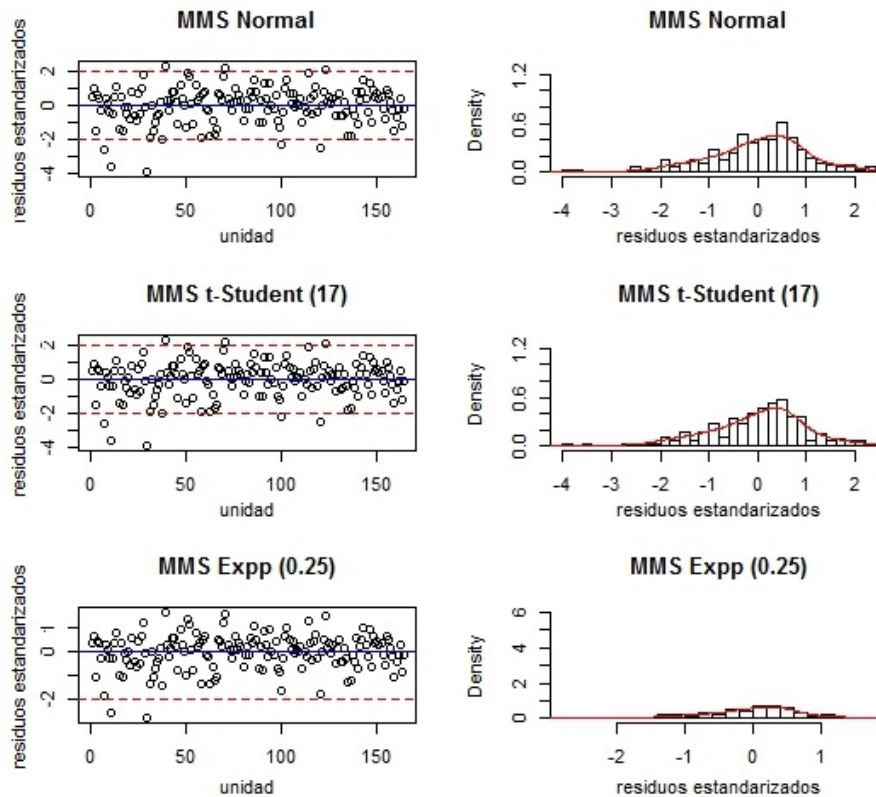


Figura 6.16: Gráficos de residuos estandarizados de los MMS con distribución normal, t -Student y Exponencial potencia

6.2.3 Análisis diagnóstico del modelo

El análisis diagnóstico del MMS está enfocado de manera análoga a la Sección 6.1.3, esto es, enfocado en el estudio de la distancia de Cook generalizada (DCG), la distancia del valor estimado estandarizado (DFITS) y del comportamiento de los residuos estandarizados por medio de gráficos de Cuartil-Cuartil que están basados en los residuos condicionales dados por la expresión (4.6) evaluados en los MMS normal, t -Student y Exponencial potencia.

Respecto al comportamiento de los residuos estandarizados, se puede ver de la Figura 6.16 que dichos residuos mayormente fluctúan entre -3 y 2 cuando la distribución subyacente de los errores del MMS es normal y t -Student, pero en el caso de que la distribución subyacente de los errores del MMS Expp, los

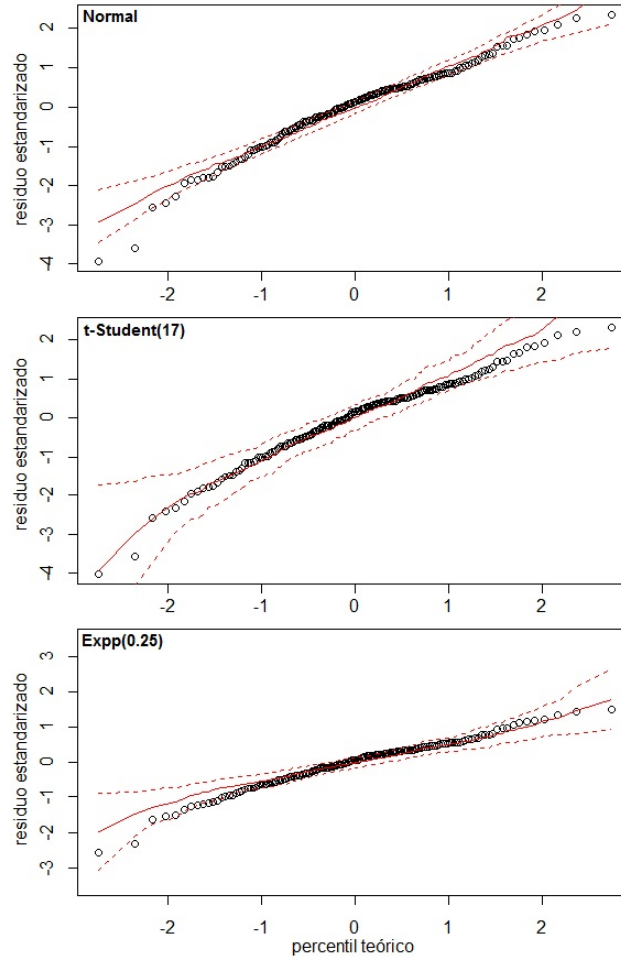


Figura 6.17: Gráfico QQplot para los residuos del MMS bajo distribución: Normal, t -Student y Exponencial potencia

residuos varían entre -2 y 1 aproximadamente, por lo cual, en estos términos se comporta de mejor manera en este último MMS. No obstante, en términos de la evaluación de la distribución de probabilidad subyacente de los errores del MMS, los residuos estandarizados del MMS t -Student se comportan de mejor manera, ver Figura 6.17, puesto que dicho comportamiento se ajusta mucho mejor en la cola inferior que las distribuciones normal y Exponencial potencia.

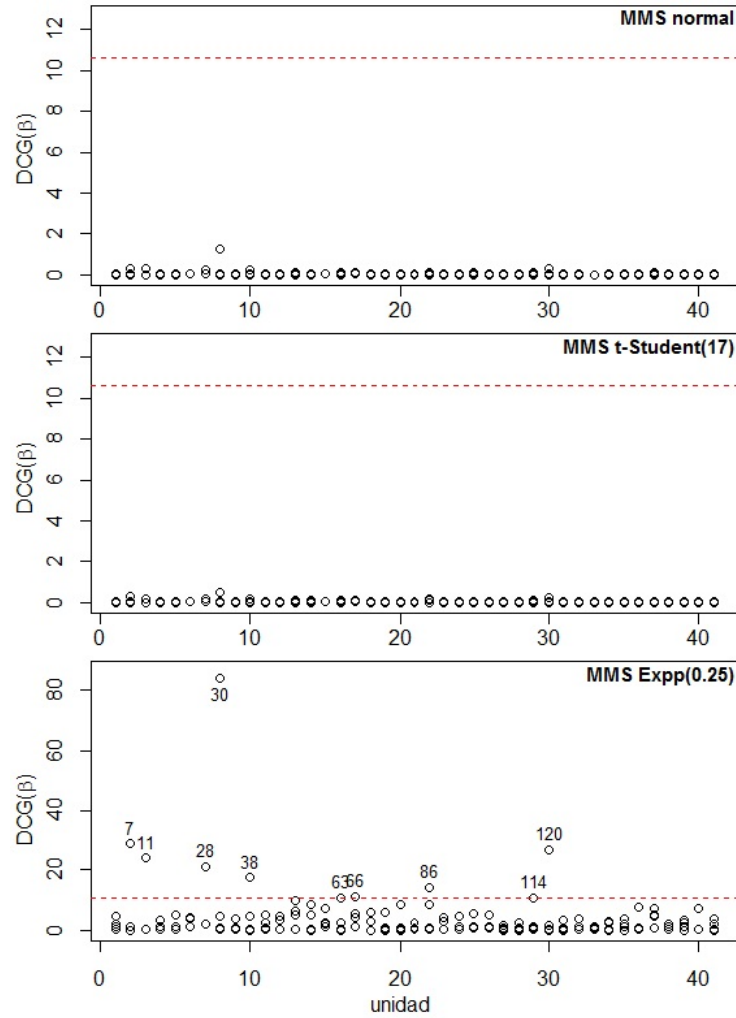


Figura 6.18: Gráfico de distancia de Cook generalizada del MMS bajo distribución: Normal, t -Student y Exponencial potencia

Para el estudio del grado de influencia global sobre las estimaciones de los parámetros asociados a los efectos fijos del MMS elíptico fueron utilizados la distancia de Cook generalizada para la componente paramétrica, basado en la expresión (4.12), y la distancia del valor estimado estandarizado para la componente no paramétrica, basado en la expresión (4.13).

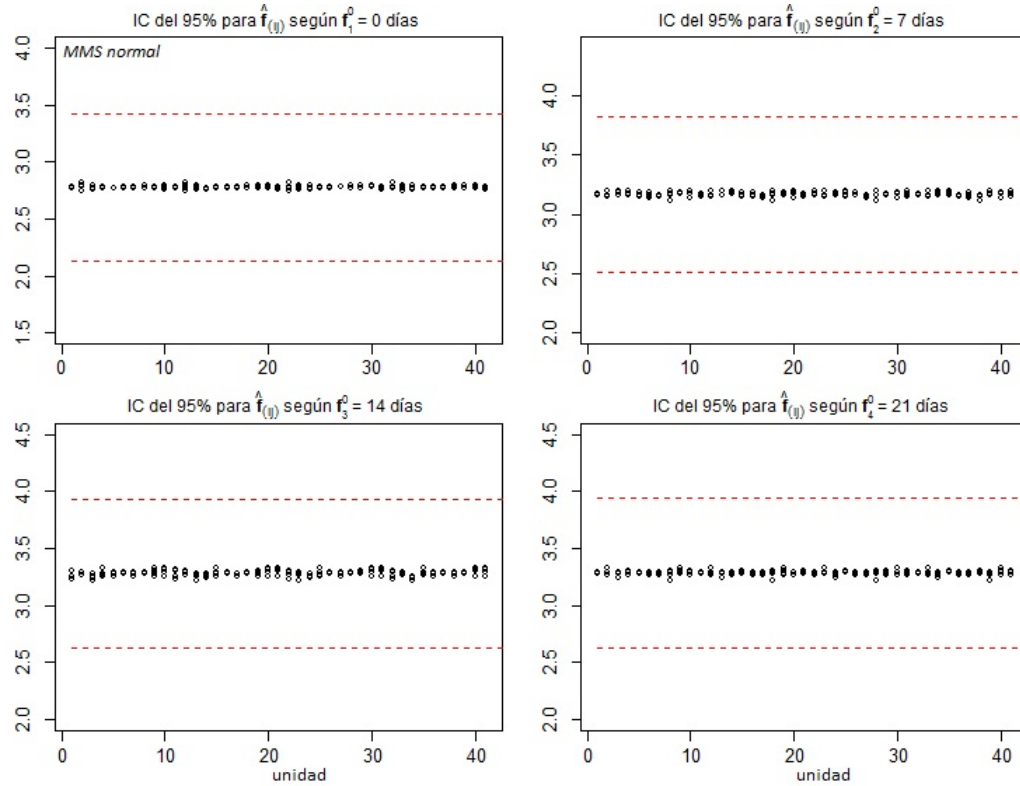


Figura 6.19: Gráfico de IC del 95% para la estimación de la componente no paramétrica del MMS normal

Respecto a la influencia global de la componente paramétrica, en la Figura 6.18 se muestran los valores estimados de la DCG (β), donde se aprecia que el comportamiento de la DCG en los MMS normal y t -Student están dentro del límite de confianza, sin embargo, se observan cambios sobre la estimación del parámetro de la parte paramétrica al extraer algunas observaciones en el MMS Expp y dichas distancias superan el valor de referencia $\chi_{0,9}^2(6) = 10.645$, donde estas observaciones corresponden a mediciones realizadas a los pacientes n°2, 3, 7, 8, 10, 16, 17, 22, n°29 y n°30.

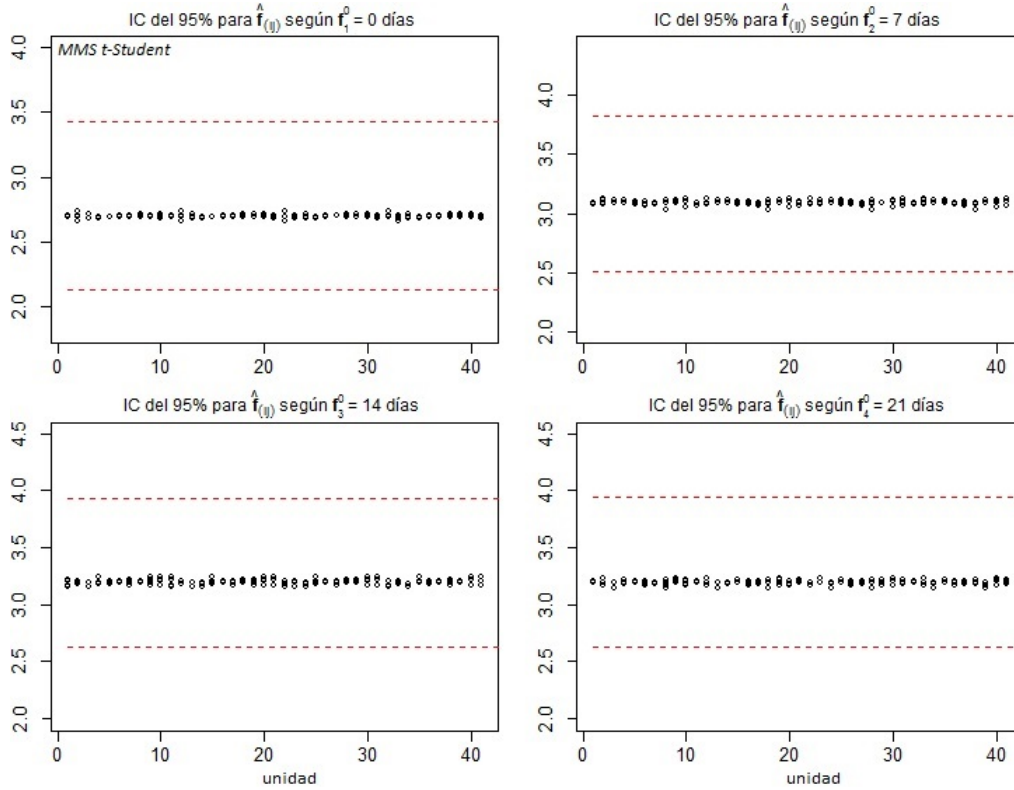


Figura 6.20: Gráfico de IC del 95% para la estimación de la componente no paramétrica del MMS t-Student(17)

Respecto a la influencia global de la componente no paramétrica, en las Figuras 6.19, 6.20 y 6.21 se muestran los parámetros estimados $\hat{f}_{(ij)}$ para los tres MMS evaluados, donde se observa que las todas las estimaciones de los parámetros de la componente no paramétrica sin la (i, j) -ésima observación en todos los tiempos se comportan se manera estable dentro de los límites de confianza.

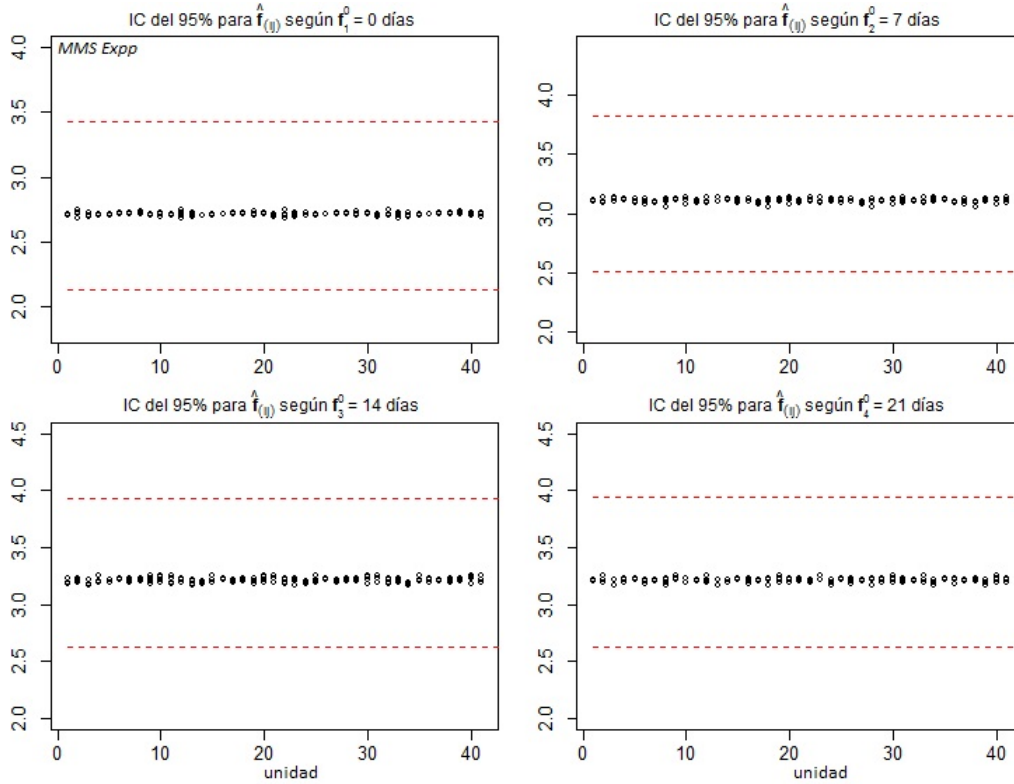


Figura 6.21: Gráfico de IC del 95% para la estimación de la componente no paramétrica del MMS Expp(0,25)

De esta manera, se aprecia que existe un grado de influencia global significativa sobre la estimación de la componente paramétrica pero sólo en el MMS Expp, sin embargo, no se observó algún grado de influencia global sobre la estimación de la componente no paramétrica en los tres MMS elípticos, el cual está asociado al efecto del tiempo de seguimiento sobre la profundidad de sondaje del paciente, siendo que se presentaron estimaciones muy diferentes a los tiempos de observación de algunos pacientes, tales como los pacientes n°3 y n°8. Por lo que, las distancias de Cook generalizadas y las distancias del valor estimado estandarizado, poseen un mejor comportamiento de las distancias bajo el MMS t-Student.

6.2.4 Análisis de autocorrelación de primer orden

Para evaluar las hipótesis de presencia de autocorrelación de primer orden y de homogeneidad en la estructura AR(1) de los errores aleatorios del MMS elíptico a nivel intra-unidad, los análisis se basaron en los resultados obtenidos en los test de score, que fueron descritos en el Capítulo 5, donde ρ es el parámetro asociado a la estructura AR(1) de la profundidad de sondaje y ρ_i es el parámetro de autocorrelación de primer orden heterogénea de la profundidad de sondaje perteneciente al i -ésimo paciente en estudio.

Cuadro 6.9: EMVP de los componentes de varianza y test de Score para presencia de autocorrelación para los MMS normal, $t(17)$ y $\text{Expp}(0,25)$

MMS elíptico	Componentes de varianza		$l_p(\hat{\theta}, \hat{\alpha})$	Test Score bajo H_0	
	$\hat{\delta}$	$\hat{\phi}$		SC_a	p-valor
Normal	0,0080	0,0720	-22,6567	5,9954	0,0143
t -Student	0,0071	0,0633	-21,9044	4,7987	0,0285
Exp. potencia	0,0041	0,0363	-22,1517	5,0308	0,0249

En el contexto de la evaluación de $H_0 : \rho = 0$, se presentó que el menor valor de SC_a fue bajo el MMS t -Student, pero con los tres modelos evaluados se obtiene la misma conclusión, esto es, se presentó evidencia significativa de presencia de autocorrelación de primer orden en las mediciones de la profundidad de sondaje a través de los tiempos de observación. Además, del Cuadro 6.9 se observa las estimaciones de los parámetros asociados a los componentes de varianza entre los MMS normal y t -Student son similares, aunque es levemente mayor en el MMS normal y en el MMS Expp las estimaciones son menores que los demás MMS elípticos.

Dado que se encontró evidencia de presencia de autocorrelación, se procedió a evaluar la hipótesis de homogeneidad de estructura AR(1), esto es probar $H_0 : \lambda = 0$, utilizando como covariable para modelación de la estructura de AR(1) el género y la edad del paciente de manera separada, cuyos resultados se describen en los Cuadros 6.10 y 6.11. Sin embargo, las estimaciones de los parámetros del MMS elíptico bajo una forma funcional logística para la estructura AR(1) intra-unidad, coinciden con las estimaciones obtenidas bajo una forma funcional Gompertz, es por esto que sólo se muestra una tabla con los resultados para la edad y el género del paciente.

Cuadro 6.10: EMVP de los componentes de varianza y test de Score para presencia de autocorrelación para los MMS normal, $t(17)$ y Expp(0,25), con edad como variable de la estructura AR(1) heterogénea

MMS elíptico	Componentes de varianza			$l_p(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})$	Test Score bajo H_0	
	$\hat{\delta}$	$\hat{\phi}$	$\hat{\rho}$		SC _h	p-valor
Normal	0,0055	0,0713	-0,5094	-51,6053	5635,09	0,1e-95
$t(17)$	0,0055	0,0631	-0,4690	-43,2291	4571,96	0,1e-95
Expp(0,25)	0,0035	0,0356	-0,1605	-26,6898	1841,46	0,1e-95

Cuadro 6.11: EMVP de los componentes de varianza y test de Score para presencia de autocorrelación para los MMS normal, $t(17)$ y Expp(0,25), con género como variable de la estructura AR(1) heterogénea

MMS elíptico	Componentes de varianza			$l_p(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})$	Test Score bajo H_0	
	$\hat{\delta}$	$\hat{\phi}$	$\hat{\rho}$		SC _h	p-valor
Normal	0,0055	0,0713	0,3052	-22,2648	20,2275	6,88e-06
$t(17)$	0,0055	0,0631	0,2957	-21,8401	20,8230	5,04e-06
Expp(0,25)	0,0037	0,0359	0,2797	-21,8120	19,2965	1,12e-05

Cuando la estructura de autocorrelación es función de la edad del paciente, se puede observar que las estimaciones de los parámetros $\hat{\delta}$ y $\hat{\phi}$ entre los MMS normal y t -Student son similares, no así en el MMS Expp donde son mas bajas dichas estimaciones, ver Cuadro 6.10, pero note que los parámetros de autocorrelación estimados son negativos, lo cual cambia la dirección de la correlación a nivel intra-unidad respecto de los resultados obtenidos en la subsección 6.2.2, donde dichas estimaciones fueron positivas. Además, la función de log-verosimilitud penalizada es maximizada en el MMS Expp, no obstante, todas las funciones de log-verosimilitud penalizada bajo H_0 son menores que las funciones de log-verosimilitud penalizada de los parámetros estimados sin restricciones.

En cambio, cuando la estructura de autocorrelación es función del género del paciente, se puede observar que las estimaciones de los parámetros $\hat{\delta}$ y $\hat{\phi}$ entre los MMS normal y t -Student son similares, no así en el MMS Expp donde son mas bajas dichas estimaciones (ver Cuadro 6.11), pero

en este caso los parámetros de autocorrelación estimados son positivos, lo cual son consistentes con los resultados obtenidos en la subsección 6.2.2, donde dichas estimaciones también fueron positivas. Además, la función de log-verosimilitud penalizada es maximizada en el MMS Expp, donde todas las funciones de log-verosimilitud penalizada bajo H_0 son cercanas a las funciones de log-verosimilitud penalizada de los parámetros estimados sin restricciones pero siguen siendo menores que estas últimas.

Respecto a los resultados del test de score para la homogeneidad de la autocorrelación a nivel intra-unidad, para cuando se emplean el género y edad del paciente como covariable para modelar la estructura AR(1) se obtienen conclusiones similares, esto significa que el nivel de autocorrelación de primer orden se comporta de manera estadísticamente diferente en las mediciones de profundidad de sondaje a través del tiempo de seguimiento y esta correlación depende de cada paciente en estudio. Cabe de señalar que, para cuantificar el efecto de la autocorrelación intra-paciente sobre la profundidad de sondaje, la edad y género fueron consideradas como una covariable para la componente paramétrica del MMS elíptico y como covariable para la modelar la estructura AR(1) a nivel intra-unidad.

6.2.5 Conclusiones del análisis de datos

De los resultados obtenidos, se presentó que la profundidad de sondaje de los pacientes tienen un comportamiento no lineal a través de los tiempos de observación, la cual es modelada por la componente no paramétrica del MMS elíptico. Además, de los tres MMS elípticos estimados, las observaciones se ajustan de mejor manera al MMS t -Student en el sentido de la minimización del SIC y AIC, como de la maximización de la función de log-verosimilitud penalizada y del comportamiento de los residuos estandarizados.

Respecto al análisis de influencia global, en los tres MMS evaluados no se detectan cambios significativos en la estimación de la curva no paramétrica bajo la supresión de una observación, pero si se detectaron cambios significativos en la estimación de la parte paramétrica del MMS bajo supresión de una observación sólo en el MMS Exponencial potencia, es así que se recomienda para futuros análisis de este conjunto de datos, es realizar un análisis de influencia local introduciendo perturbaciones sobre la componente paramétrica del MMS pero bajo supresión de observaciones

de un paciente cualquiera, debido a que algunos pacientes no siguieron el protocolo de investigación entre el inicio y el día siete después de iniciado el tratamiento.

Por último, utilizando los tres MMS elípticos evaluados se obtiene evidencia significativa de presencia de autocorrelación de primer orden sobre la profundidad de sondaje a nivel intra-paciente y que dicha autocorrelación depende del género o la edad del paciente, ya sea como un efecto aditivo aditivo asociado a la parte paramétrica, como también del comportamiento de la estructura de AR(1) a nivel intra-unidad.

CONSIDERACIONES FINALES

7.1 Discusión

En el presente trabajo se ha descrito el MMS elíptico para datos de tipo longitudinal que incorpora una estructura AR(1) a nivel intra-unidad, es decir, una estructura en el modelo que permite modelar los niveles de autocorrelación a través del tiempo de seguimiento de cada unidad de estudio. Además, se ha propuesto un proceso de cálculo iterativo para los EMVP asociados a los parámetros del MMS elíptico AR(1), que están basados en los procesos de Backfitting y score de Fisher, donde ambos incorporan la función de log-verosimilitud penalizada del modelo marginal elíptico, del cual, fueron derivadas las funciones de score e Información de Fisher penalizada asociadas a los parámetros del MMS elíptico.

El modelo presentado es un modelo semiparamétrico con una estructura sistemática más general que el modelo lineal parcial propuesto por Speckman (1988) y Robinson (1988). En el contexto de modelos semiparamétricos, es más apropiado examinar la influencia parcial de las estimaciones de los parámetros asociados a los efectos fijos de la componente paramétrica y no paramétrica de manera separada. Para esto, se propuso dos medidas diagnósticas de la influencia global en dichas componentes del MMS

elíptico, basada en la distancia de Cook generalizada y el valor estimado estandarizado, que corresponden a una generalización de la medida diagnóstica propuesta por Fung et al. (2002), pero la diferencia radica en estos autores, incorporan una variable aleatoria adicional al modelo, que depende del tiempo de medición t_{ij} y con efectos aleatorios \mathbf{b}_i en cada unidad de estudio, que miden de manera conjunta la estructura de correlación intra-unidad.

No obstante, la evaluación de la presencia y homogeneidad de autocorrelación AR(1) intra-unidad, el test de score fue utilizado, debido a las ventajas asociadas al esfuerzo del proceso de cálculo computacional y que sólo se requiere la estimación de los parámetros bajo la hipótesis nula. El test de score propuesto se evalúa un modelo más general que en Tsai (1986), Chi & Reinsel (1989; 1991), Lin & Wei (2004), Cao & Lin (2011; 2012) e IbacachePulgar *et al.* (2012).

Finalmente, los resultados de la simulación Monte-Carlo de la potencia del test de score, para evaluar la presencia y homogeneidad de autocorrelación dentro de las unidades de estudio, se observaron comportamientos similares entre la potencia empírica de los MMS normal, MMS t -Studnet y MMS exponencial potencia. Además, para evaluar $H_0 \rho = 0$ no requiere de una elevada cantidad de unidades de estudio para obtener una buena potencia, en cambio, para evaluar $H_0 \rho_i = \rho$ depende de una mayor cantidad de observaciones, aunque, el comportamiento de la potencia empírica no cambia significativamente según la forma funcional de la estructura de autocorrelación no homogénea. De esta manera, es razonable considerar que el test de score tiene un comportamiento robusto sobre la distribución subyacente de los datos y la forma funcional de la estructura de autocorrelación intra-unidad, por lo que el test de score, se comporta de manera muy eficiente.

7.2 Implementación computacional

Para los procesos de cálculo iterativo que fueron basados en el algoritmo Backfitting y score de Fisher para estimar los efectos fijos y los componentes de varianza del MMS elíptico AR(1) y fue implementado en el software R-Cran (versión 3.0.1 x64), utilizando los paquetes estadísticos *car*, *haplo.stats*, *MASS*, *Matrix*, *matrixcalc* y *nlme*.

Debido a la estructura de bloque diagonal de la matriz de información de Fisher, la programación se facilita en el desarrollo de este método iterativo, al permitir que los parámetros de estimación asociados para modelar los parámetros de efectos fijos y parámetros asociados los componentes de varianza utilizando procesos iterativos independientes.

En el análisis diagnóstico del MMS elíptico AR(1) también fue realizado en R-Cran, para esto, se desarrollaron funciones para el cálculo de las las distancias de Cook generalizadas y los DFITS, se estudia el comportamiento distribucional de los residuos del MMS elíptico, por medio de la inspección de gráficos cuartil-cuartil construídas por el procedimiento *envelope*, las cuales son réplicas bootstrap (se emplearon 500 simulaciones) de la curva evaluada en un número de puntos fijos. En esta etapa de la programación fue necesario utilizar los paquetes estadísticos *boot*, *gdata*, *gplots* y *lattice*.

Para el proceso de simulación Monte-Carlo de un MMS elíptico AR(1) y la potencia empírica de los test de score, para la evaluación de presencia y homogeneidad de estructura AR(1), la programación también fue realizada en el software R-Cran, empleando algunas funciones de generación de vectores y variables aleatorias del paquete estadístico *LaplacesDemon*. No obstante, se tuvieron que reprogramar funciones para la generación de variables aleatorias univariadas con distribución exponencial potencia, debido a la diferencia en la parametrización de la función ofrecida por el paquete y la parametrización de la distribución exponencial potencia utilizada para el MMS elíptico.

7.3 Consideraciones de trabajo futuro

En el análisis diagnóstico del MMS elíptico, se propusieron expresiones de la distancia de Cook generalizada para la componente paramétrica y otra de la distancia del valor estimado estandarizado para la componente no paramétrica. Sin embargo, estos indicadores miden el nivel de influencia de cada observación respecto de los EMVP cuando se extrae una observación, es decir, un nivel de influencia global, lo cual no permite asegurar si un punto es considerado como influyente o no, es por esto que, se recomienda realizar investigaciones que aborden el estudio de influencia local, y así introducir pequeñas perturbaciones en el MMS elíptico que afecten las componentes de los efectos fijos.

Respecto a las componentes del MMS elíptico, se sugiere estudiar el comportamiento del test de score para evaluar hipótesis respecto a los parámetros de la componente paramétrica, es decir, evaluar $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ y $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$. De igual manera, considerar evaluar la hipótesis de significancia de la componente no paramétrica en el modelo, esto es, $H_0 : \mathbf{N}\mathbf{f} = \mathbf{0}$, como también, evaluar la linealidad de la componente no paramétrica.

Para finalizar, en la actualidad ha aumentado la necesidad estudiar datos que poseen una estructura de correlación diferente a la temporal, tales como los datos espaciales. El uso de este tipo de datos se ha diversificado en varias áreas, como por ejemplo: Minería, investigación pesquera, Epidemiología, Economía y en el estudio de fenómenos naturales. Es por esto que se recomienda estudiar, un MMS elíptico para datos con estructura de correlación espacial.

APÉNDICE I: GRÁFICOS CAPÍTULO 1

En este anexo se describen las rutinas de programación de las figuras mostradas del capítulo 1, la sintaxis de programación está realizada en R-Cran que depende de los paquetes *sn*, *graphics* y *scatterplot3d*.

I.1. Figura 1.1

```
rm(list=ls())
library(graphics);library(scatterplot3d)

### Parametrico ###
b1<-0.5;b2<-0.35;x1<-seq(-5,5,.2);x2<-seq(-5,5,.2)
f<-function(x1,x2){
5-x1*b1+x2*b2
};z<-outer(x1,x2,f);persp(x1,x2,z,theta=35,phi=20,col=cm.colors(0)
,xlab="x1",ylab="x2",zlab="y")

### No Parametrico ###
b1<-0.4;b2<-0;x1<-seq(-2,2,.1);x2<-seq(-2,2,.1)
f<-function(x1,x2){
.5*(cos(x1/b1)/(1+(x2+b2)^2))
};z<-outer(x1,x2,f);persp(x1,x2,z,theta=35,phi=20,col=cm.colors(0)
,xlab="x1",ylab="x2",zlab="y")
```

```

### lineal parcial ###
b1<- .8;b2<- .1;x1<-seq(-2,2,.1);x2<-seq(-2,2,.1)
f<-function(x1,x2){
  (b1/2)/(1+((x1*b1)^2))+ (1+(x2*b2)) }
z<-outer(x1,x2,f);persp(x1,x2,z,theta=35,phi=20,col=cm.colors(0)
,xlab="x1",ylab="x2",zlab="y")

### parcialmente paramétrico ###
b1<- .8;b2<- .1;x1<-seq(-5,5,.3);x2<-seq(-5,5,.3)
f<-function(x1,x2){
  (b1/2)/(1+((x1*b1)^2))+ (1+(x2*b2)^2)
};z<-outer(x1,x2,f);persp(x1,x2,z,theta=35,phi=20,col=cm.colors(0)
,xlab="x1",ylab="x2",zlab="y")

### aditivo separable ###
x1<-seq(-2.5*pi/2,2.5*pi/2,.2);x2<-seq(-2.5*pi/2,2.5*pi/2,.2)
b1<-1;b2<- .7;f<-function(x1,x2){
  cos(x1*b1)+cos(x2*b2)
};z<-outer(x1,x2,f);persp(x1,x2,z,theta=35,phi=20,col=cm.colors(0)
,xlab="x1",ylab="x2",zlab="y")

### modelo indexado ###
b1<-1;b2<-1;x1<-seq(-2,1,.15);x2<-seq(-2,1,.15)
f<-function(x1,x2){
  sin((x1*b1+x2*b2))
};z<-outer(x1,x2,f);persp(x1,x2,z,theta=35,phi=20,col=cm.colors(0)
,xlab="x1",ylab="x2",zlab="y")

```

I.2. Figura 1.2

```

rm(list=ls())
library(sn);library(graphics)
x<-seq(-4,4,length=100)

##### función de distribución simétricas #####
f.dist<-function(x,mu=0,sigma=1,gl=4,pf=-.25,type="normal"){

```

```

switch(type,
normal = { ## Normal
  k<-1/sqrt(pi*2*(sigma^2))
  Maha<-((x-mu)/sigma)^2
  y<-k*exp((-1/2)*Maha)
},
t = { ## t-student(gl)
  k<-gamma((1+gl)/2)/(gamma(gl/2)*sqrt(pi*gl*(sigma^2)))
  Maha<-((x-mu)/sigma)^2
  y<-k*(1+(Maha/gl))^-((1+gl)/2)
},
Expp = { ## Exponencial potencia(pf)
  k<-1/(gamma(1.5+(pf/2))*2^(1.5+(pf/2)))
  Maha<-((x-mu)/sigma)^2
  y<-k*exp((-1/2)*(Maha^(1/(1+pf))))
}) # fin switch #
### resultados ###
return(y)
} ##### fin función #####

BD<-data.frame(x,y1=f.dist(x,type="normal"),y2=f.dist(x,type="t")
,y3=f.dist(x,type="t",gl=10),y4=f.dist(x,type="Expp"),pf=-.5
,y5=f.dist(x,type="Expp",pf=.5))

### Gráfico de distribuciones ###
plot(BD$x,BD$y1,xlab="y",ylab=expression(italic(f)[Y]~"(y)")
,xlim=c(-4,4),ylim=c(0,.5),main="Algunas distribuciones de la
clase simétrica",type="l")
lines(BD$x,BD$y2,col=2,lty=1);lines(BD$x,BD$y3,col=2,lty=2)
lines(BD$x,BD$y4,col=4,lty=1);lines(BD$x,BD$y5,col=4,lty=2)
legend("topleft",c("normal","t(4)","t(10)","expp(-0,5)","expp(0,5)")
,cex=.8,col=c(1,2,2,4,4),lty=c(1,1,2,1,2),bty="n")
mtext(expression(mu==0~";"~sigma^2==1))

plot(BD$x,BD$y1,xlab="y",ylab=expression(italic(f)[Y]~"(y)")
,xlim=c(0,4),ylim=c(0,.3),main="Algunas distribuciones de la
clase simétrica",type="l")
lines(BD$x,BD$y2,col=2,lty=1);lines(BD$x,BD$y3,col=2,lty=2)

```

```

lines(BD$x,BD$y4,col=4,lty=1);lines(BD$x,BD$y5,col=4,lty=2)
legend("topright",c("normal","t(4)","t(14)","exp(-0,5)","exp(0,5)")
,cex=.8,col=c(1,2,2,4,4),lty=c(1,1,2,1,2),bty="n")
mtext(expression(mu==0~";"~sigma^2==1))

```

I.3. Figura 1.3

```

rm(list=ls())
library(graphics);library(scatterplot3d)
x1<-x2<-seq(-4,4,length=80)

##### función de distribución elípticas #####
f2.dist<-function(x1,x2,mu1=0,mu2=0,rho=.5,gl=4,pf=-.25,type="normal"){
  switch(type,
    normal = { ## Normal
      sigma<-matrix(c(1,rho,rho,1),2,2)
      Is<-solve(sigma)
      k<-1/(pi*2)
      Maha<-(Is[1,1]*(x1-mu1)^2)+(2*Is[2,1]*(x1-mu1)*(x2-mu2))+(Is[2,2]
      *(x2-mu2)^2);y<-k*(det(sigma)^(-.5))*exp((-1/2)*Maha)
    },
    t = { ## t-student(gl)
      sigma<-matrix(c(1,rho,rho,1),2,2)
      Is<-solve(sigma)
      k<-gamma((2+gl)/2)/(gl*pi*gamma(gl/2))
      Maha<-(Is[1,1]*(x1-mu1)^2)+(2*Is[2,1]*(x1-mu1)*(x2-mu2))+(Is[2,2]
      *(x2-mu2)^2);y<-k*(det(sigma)^(-.5))*(1+(Maha/gl))^(-(2+gl)/2)
    },
    Expp = { ## Exponencial potencia(pf)
      sigma<-matrix(c(1,rho,rho,1),2,2)
      Is<-solve(sigma)
      k<-(2^(-1-pf))/(pi*gamma(2+pf))
      Maha<-(Is[1,1]*(x1-mu1)^2)+(2*Is[2,1]*(x1-mu1)*(x2-mu2))+(Is[2,2]
      *(x2-mu2)^2);y<-k*(det(sigma)^(-.5))*exp((-1/2)*(Maha)^(1/(1+pf)))
    }) # fin switch #
  return(y) ### resultados ###
} ##### fin función #####

```

```

z1<-outer(x1,x2,f2.dist,type="normal",rho=0)
z2<-outer(x1,x2,f2.dist,type="normal",rho=0.6)
z3<-outer(x1,x2,f2.dist,type="t",rho=0,gl=3)
z4<-outer(x1,x2,f2.dist,type="t",rho=0.6,gl=15)
z5<-outer(x1,x2,f2.dist,type="Expp",rho=0,pf=-.3)
z6<-outer(x1,x2,f2.dist,type="Expp",rho=0.6,pf=.3)

##### Gráficos de distribuciones #####

### Normal ###
contour(x1,x2,z1,nlevels=8,xlab="x",ylab="y",main="Distribución
normal esférica",xlim=c(-4,4),ylim=c(-4,4))
text(-4,3.7,expression(mu[x]==~mu[y]==0~";~sigma[x]==~sigma[y]==
1~";~rho[xy]==0),adj = c(0,0))
persp(expand=.9,x1,x2,z1,theta=45,phi=20,col=cm.colors(0)
,xlab="x",ylab="y",zlab=expression(f(x,y)),main="Distribución
normal esférica");mtext(expression(rho[xy]==0))

contour(x1,x2,z2,nlevels=8,xlab="x",ylab="y",main="Distribución
normal elíptica",xlim=c(-4,4),ylim=c(-4,4))
text(-4,3.7,expression(mu[x]==~mu[y]==0~";~sigma[x]==~sigma[y]==
1~";~rho[xy]==0.6),adj = c(0,0))
persp(expand=.9,x1,x2,z2,theta=45,phi=20,col=cm.colors(0)
,xlab="x",ylab="y",zlab=expression(f(x,y)),main="Distribución
normal elíptica");mtext(expression(rho[xy]==0.6))

### t-Student ###
contour(x1,x2,z3,nlevels=8,xlab="x",ylab="y",main="Distribución
t-student esférica",xlim=c(-4,4),ylim=c(-4,4))
text(-4,3.7,expression(mu[x]==~mu[y]==0~";~sigma[x]==~sigma[y]==
1~";~rho[xy]==0~";~upsilon==3),adj = c(0,0))
persp(expand=.9,x1,x2,z3,theta=45,phi=20,col=cm.colors(0)
,xlab="x",ylab="y",zlab=expression(f(x,y)),main="Distribución
t-student esférica");mtext(expression(rho[xy]==0~";~upsilon==3))

contour(x1,x2,z4,nlevels=8,xlab="x",ylab="y",main="Distribución
t-student elíptica",xlim=c(-4,4),ylim=c(-4,4))

```

```

text(-4,3.7,expression(mu[x]==~mu[y]==0~";"~sigma[x]==~sigma[y]==
  1~";"~rho[xy]==0.6~";"~upsilon==15),adj = c(0,0))
persp(expand=.9,x1,x2,z4,theta=45,phi=20,col=cm.colors(0)
  ,xlab="x",ylab="y",zlab=expression(f(x,y)),main="Distribución
  t-student elíptica");mtext(expression(rho[xy]==0.6~";"~nu==15))

### Exponencial potencia ###
contour(x1,x2,z5,nlevels=8,xlab="x",ylab="y",main="Distribución
  Expp esférica",xlim=c(-4,4),ylim=c(-4,4))
text(-4,3.7,expression(mu[x]==~mu[y]==0~";"~sigma[x]==~sigma[y]==
  1~";"~rho[xy]==0~";"~vartheta==~"-~"~0.3),adj = c(0,0))
persp(expand=.9,x1,x2,z5,theta=45,phi=20,col=cm.colors(0)
  ,xlab="x",ylab="y",zlab=expression(f(x,y)),main="Distribución
  Expp esférica")
mtext(expression(rho[xy]==0~";"~vartheta==~"-~"~0.3))

contour(x1,x2,z6,nlevels=8,xlab="x",ylab="y",main="Distribución
  Expp elíptica",xlim=c(-4,4),ylim=c(-4,4))
text(-4,3.7,expression(mu[x]==~mu[y]==0~";"~sigma[x]==~sigma[y]==
  1~";"~rho[xy]==0.6~";"~vartheta==0.3),adj = c(0,0))
persp(expand=.9,x1,x2,z6,theta=45,phi=20,col=cm.colors(0)
  ,xlab="x",ylab="y",zlab=expression(f(x,y)),main="Distribución
  Expp elíptica")
mtext(expression(rho[xy]==0.6~";"~vartheta==0.3))

```

APÉNDICE II: DEMOSTRACIONES

CAPITULO 3

II.1. Funciones de score penalizadas

II.1.1. Componente paramétrica

La función de score penalizada de la parte paramétrica de la i -ésima unidad de estudio, según (3.7), está dado por:

$$\mathbb{S}_{p_i}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} l_{p_i}(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) = \frac{\partial}{\partial \boldsymbol{\beta}} l_i(\boldsymbol{\theta}; \mathbf{y}_i) = \mathcal{W}_g(u_i) \frac{\partial u_i}{\partial \boldsymbol{\beta}}$$

y desarrollando un poco de algebra de matrices, tenemos:

$$\begin{aligned} \frac{\partial u_i}{\partial \boldsymbol{\beta}} &= \left(\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\varepsilon}_i^t \right) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i + \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \left(\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\varepsilon}_i \right) = \left[\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\varepsilon}_i \right]^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i + \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \left[\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{\varepsilon}_i \right] \\ &= -\mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i = -\mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i = -2\mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \end{aligned}$$

Por lo cual, obtenemos:

$$\mathbb{S}_{p_i}(\boldsymbol{\beta}) = \mathcal{W}_g(u_i) \frac{\partial u_i}{\partial \boldsymbol{\beta}} = w_i \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i$$

Y por tanto, la función de score penalizada de las m observaciones de la parte paramétrica del MMS esta dado por:

$$\begin{aligned}
\mathbb{S}_p(\boldsymbol{\beta}) &= \sum_{i=1}^n w_i \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i = w_1 \mathbf{X}_1^t \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\varepsilon}_1 + \cdots + w_n \mathbf{X}_n^t \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\varepsilon}_n \\
&= [\mathbf{X}_1^t \quad \cdots \quad \mathbf{X}_n^t] \begin{bmatrix} w_1 \boldsymbol{\Sigma}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & w_2 \boldsymbol{\Sigma}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & w_n \boldsymbol{\Sigma}_n^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix} \\
&= \mathbf{X}^t \mathbf{W} \boldsymbol{\varepsilon},
\end{aligned}$$

donde $\mathbf{W} = \text{diag}(w_1 \boldsymbol{\Sigma}_1^{-1}, \dots, w_n \boldsymbol{\Sigma}_n^{-1})$ y $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1^t \cdots \boldsymbol{\varepsilon}_n^t]^t$, con $\mathbf{W} \in \mathbb{R}^{m \times m}$ y $\boldsymbol{\varepsilon} \in \mathbb{R}^m$.

II.1.2. Componente no paramétrica

La función de score penalizada de la parte no paramétrica de la i -ésima unidad de estudio, según (3.7), está dado por:

$$\begin{aligned}
\mathbb{S}_{p_i}(\mathbf{f}) &= \frac{\partial}{\partial \mathbf{f}} l_{p_i}(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) = \frac{\partial}{\partial \mathbf{f}} l_i(\boldsymbol{\theta}; \mathbf{y}_i) - \frac{\alpha}{2n} \frac{\partial}{\partial \mathbf{f}} (\mathbf{f}^t \mathbf{K} \mathbf{f}) \\
&= \mathcal{W}_g(u_i) \frac{\partial u_i}{\partial \mathbf{f}} - \frac{\alpha}{n} \mathbf{K} \mathbf{f}
\end{aligned}$$

al desarrollar lo anterior, tenemos que:

$$\begin{aligned}
\frac{\partial u_i}{\partial \mathbf{f}} &= \left(\frac{\partial}{\partial \mathbf{f}} \boldsymbol{\varepsilon}_i^t \right) \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i + \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \left(\frac{\partial}{\partial \mathbf{f}} \boldsymbol{\varepsilon}_i \right) = \left[\frac{\partial}{\partial \mathbf{f}} \boldsymbol{\varepsilon}_i \right]^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i + \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \left[\frac{\partial}{\partial \mathbf{f}} \boldsymbol{\varepsilon}_i \right] \\
&= -\mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i = -\mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i = -2\mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i
\end{aligned}$$

Y así obtenemos:

$$\mathbb{S}_{p_i}(\mathbf{f}) = \mathcal{W}_g(u_i) \frac{\partial u_i}{\partial \mathbf{f}} - \frac{\alpha}{n} \mathbf{K} \mathbf{f} = w_i \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \frac{\alpha}{n} \mathbf{K} \mathbf{f}$$

Y por tanto, la función de Score penalizada de las n unidades de estudio está dado por:

$$\begin{aligned} \mathbb{S}_p(\mathbf{f}) &= \sum_{i=1}^n \left(w_i \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \frac{\alpha}{n} \mathbf{Kf} \right) = w_1 \mathbf{N}_1^t \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\varepsilon}_1 + \cdots + w_n \mathbf{N}_n^t \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\varepsilon}_n - \alpha \mathbf{Kf} \\ &= [\mathbf{N}_1^t \ \cdots \ \mathbf{N}_n^t] \begin{bmatrix} w_1 \boldsymbol{\Sigma}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & w_2 \boldsymbol{\Sigma}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & w_n \boldsymbol{\Sigma}_n^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix} - \alpha \mathbf{Kf} \\ &= \mathbf{N}^t \mathbf{W} \boldsymbol{\varepsilon} - \alpha \mathbf{Kf}, \end{aligned}$$

donde $\mathbf{N} = [\mathbf{N}_1^t \ \cdots \ \mathbf{N}_n^t]^t$, $\mathbf{W} \in \mathbb{R}^{m \times m}$ y $\boldsymbol{\varepsilon} \in \mathbb{R}^m$.

II.1.3. Componentes de varianza

La función de score penalizada de los parámetros asociados a los componentes de varianza para la i -ésima unidad de estudio, según (3.7), está dado por:

$$\mathbb{S}_{p_i}(\boldsymbol{\tau}) = \frac{\partial}{\partial \boldsymbol{\tau}} l_{p_i}(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) = -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \right) + \mathcal{W}_g(u_i) \frac{\partial u_i}{\partial \boldsymbol{\tau}},$$

con

$$\frac{\partial u_i}{\partial \boldsymbol{\tau}} = \boldsymbol{\varepsilon}_i^t \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \boldsymbol{\tau}} \boldsymbol{\varepsilon}_i = -\boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i = -\boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i$$

Por lo tanto:

$$\mathbb{S}_{p_i}(\boldsymbol{\tau}) = -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \right) - \mathcal{W}_g(u_i) \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i$$

Y la función de score total para los componentes de varianza $\boldsymbol{\tau}$ de las unidades de estudio es:

$$\begin{aligned} \mathbb{S}_p(\boldsymbol{\tau}) &= \sum_{i=1}^n \mathbb{S}_{p_i}(\boldsymbol{\tau}) = \sum_{i=1}^n \left[-\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \right) - \mathcal{W}_g(u_i) \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right] \\ &= -\frac{1}{2} \sum_{i=1}^n \left[\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \right) - w_i \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right] \end{aligned}$$

Y el vector de parámetros $\boldsymbol{\tau} = [\boldsymbol{\delta}^t \boldsymbol{\eta}^t]^t$ a estimar es $\boldsymbol{\tau} = [\delta_1 \ \cdots \ \delta_s \ \phi \ \rho]^t = [\tau_1 \ \cdots \ \tau_s \ \tau_{s+1} \ \tau_{s+2}]^t$ y $\boldsymbol{\Sigma}_i(\boldsymbol{\tau})$. Además, $\dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}}$ posee diferentes expresiones según el parámetro de la componente de varianza.

Parámetros de varianza intra-unidad

Para τ_k con $k = 1, \dots, s$ corresponde al vector $\boldsymbol{\delta} = [\delta_1 \cdots \delta_s]$ que es compuesta por los elementos de la matriz triangular superior de la matriz $\mathbf{D} \in \mathbb{R}^{q \times q}$ con $s = q(q+1)/2$ y considerando que la matriz de varianzas-covarianzas de los efectos aleatorios es simétrica, entonces la matriz \mathbf{D} puede ser escrita de la siguiente manera:

$$\begin{aligned} \mathbf{D} &= \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1q} \\ d_{21} & d_{22} & \cdots & d_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ d_{q1} & d_{q2} & \cdots & d_{qq} \end{bmatrix} = \begin{bmatrix} \delta_1 & \delta_2 & \cdots & \delta_q \\ \delta_2 & \delta_{q+1} & \cdots & \delta_{2q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_q & \delta_{2q-1} & \cdots & \delta_s \end{bmatrix} \\ &= \begin{bmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_{q+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_s \end{bmatrix} + \begin{bmatrix} 0 & \delta_2 & \cdots & \delta_q \\ 0 & 0 & \cdots & \delta_{2q-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \delta_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_q & \delta_{2q-1} & \cdots & 0 \end{bmatrix} \\ &= \text{diag}(\mathbf{D}) + \tilde{\mathbf{D}} + \tilde{\mathbf{D}}^t \end{aligned}$$

donde $\tilde{\mathbf{D}}$ denota la matriz estrictamente triangular superior¹ de la matriz \mathbf{D} .

De esta manera, tenemos:

$$\dot{\Sigma}_{i\tau_k} = \frac{\partial \Sigma_i}{\partial \tau_k} = \frac{\partial \Sigma_i}{\partial \delta_k} = \frac{\partial}{\partial \delta_k} (\mathbf{z}_i \mathbf{D} \mathbf{z}_i^t + \mathbf{Q}_i) = \mathbf{z}_i \frac{\partial \mathbf{D}}{\partial \delta_k} \mathbf{z}_i^t = \mathbf{z}_i \dot{\mathbf{D}}_k \mathbf{z}_i^t$$

en el cual, $\dot{\mathbf{D}}_k \in \mathbb{R}^{q \times q}$ que depende de una matriz $\mathbf{E}_{lj} \in \mathbb{R}^{q \times q}$ cuyo (l, j) -ésimo elemento es igual a $e_{lj} = \mathbf{I}(\delta_k = d_{lj})$ con $l, j = 1, \dots, q$ y $k = 1, \dots, s$ donde:

$$\dot{\mathbf{D}}_k = \begin{cases} \mathbf{E}_{lj} + \mathbf{E}_{lj}^t & \text{para } l \neq j \\ \mathbf{E}_{jj} & \text{para } l = j \end{cases}$$

Estimación del parámetro constante

Para τ_k con $k = s+1$, corresponde al parámetro ϕ , donde:

$$\dot{\Sigma}_{i\tau_k} = \frac{\partial \Sigma_i}{\partial \tau_{s+1}} = \frac{\partial \Sigma_i}{\partial \phi} = \frac{\partial}{\partial \phi} (\mathbf{z}_i \mathbf{D} \mathbf{z}_i^t + \mathbf{Q}_i) = \frac{\partial \mathbf{Q}_i}{\partial \phi} = \mathbf{C}_i^{-1} \mathbf{C}_i^{-t}$$

¹Una matriz \mathbf{A} se dice ser estrictamente triangular superior si su elemento $a_{ij} = 0$ para $i \geq j$

Estimación del parámetro de autocorrelación

Para τ_k con $k = s + 2$, corresponde al parámetro ρ , donde:

$$\begin{aligned}\dot{\Sigma}_{i\tau_k} &= \frac{\partial \Sigma_i}{\partial \tau_{s+2}} = \frac{\partial \Sigma_i}{\partial \rho} = \frac{\partial}{\partial \rho} (\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t + \mathbf{Q}_i) = \frac{\partial \mathbf{Q}_i}{\partial \rho} = \phi \frac{\partial}{\partial \rho} (\mathbf{C}_i^{-1} \mathbf{C}_i^{-t}) \\ &= \phi \frac{\partial \mathbf{C}_i^{-1}}{\partial \rho} \mathbf{C}_i^{-t} + \phi \mathbf{C}_i^{-1} \frac{\partial \mathbf{C}_i^{-t}}{\partial \rho} = -\phi \mathbf{C}_i^{-1} \frac{\partial \mathbf{C}_i}{\partial \rho} \mathbf{C}_i^{-1} \mathbf{C}_i^{-t} - \phi \mathbf{C}_i^{-1} \mathbf{C}_i^{-t} \frac{\partial \mathbf{C}_i^t}{\partial \rho} \mathbf{C}_i^{-t}\end{aligned}$$

Al definir la matriz $\dot{\mathbf{C}}_i \in \mathbb{R}^{n_i \times n_i}$ que corresponde a:

$$\dot{\mathbf{C}}_i = \begin{bmatrix} \frac{\partial}{\partial \rho} \sqrt{1 - \rho^2} & 0 & \cdots & 0 \\ -\frac{\partial}{\partial \rho} \rho & \frac{\partial}{\partial \rho} 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial}{\partial \rho} 1 \end{bmatrix} = \begin{bmatrix} -\rho & 0 & \cdots & 0 \\ \sqrt{1 - \rho^2} & 0 & \cdots & 0 \\ -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Por lo tanto:

$$\begin{aligned}\frac{\partial \Sigma_i}{\partial \rho} &= -\phi \mathbf{C}_i^{-1} \dot{\mathbf{C}}_i \mathbf{C}_i^{-1} \mathbf{C}_i^{-t} - \phi \mathbf{C}_i^{-1} \mathbf{C}_i^{-t} \dot{\mathbf{C}}_i^t \mathbf{C}_i^{-t} \\ &= -\phi \mathbf{C}_i^{-1} \left(\dot{\mathbf{C}}_i \mathbf{C}_i^{-1} + \mathbf{C}_i^{-t} \dot{\mathbf{C}}_i^t \right) \mathbf{C}_i^{-t}\end{aligned}$$

II.2. Matriz Hessiana penalizada

La matriz Hessiana del MMS elíptico es definida como la segunda derivada de la función de log-verosimilitud penalizada dada en (3.4) respecto del parámetro $\boldsymbol{\theta}$, por lo tanto:

$$\mathbb{H}(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} l_p(\boldsymbol{\theta}, \alpha; \mathbf{y}_1, \dots, \mathbf{y}_n) = -\sum_{i=1}^n \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} l_i^p(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) = -\sum_{i=1}^n \mathbb{H}_i(\boldsymbol{\theta})$$

, donde

$$\mathbb{H}_i(\boldsymbol{\theta}) = \begin{bmatrix} \mathbb{H}_i(\boldsymbol{\beta}, \boldsymbol{\beta}) & \mathbb{H}_i(\boldsymbol{\beta}, \mathbf{f}) & \mathbb{H}_i(\boldsymbol{\beta}, \boldsymbol{\tau}) \\ \mathbb{H}_i(\mathbf{f}, \boldsymbol{\beta}) & \mathbb{H}_i(\mathbf{f}, \mathbf{f}) & \mathbb{H}_i(\mathbf{f}, \boldsymbol{\tau}) \\ \mathbb{H}_i(\boldsymbol{\tau}, \boldsymbol{\beta}) & \mathbb{H}_i(\boldsymbol{\tau}, \mathbf{f}) & \mathbb{H}_i(\boldsymbol{\tau}, \boldsymbol{\tau}) \end{bmatrix}$$

Además, note que podemos escribir lo siguiente:

$$\mathbb{H}_i(\boldsymbol{\theta}) = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} l_i^p(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) = \frac{\partial}{\partial \boldsymbol{\theta}^t} \left[\frac{\partial}{\partial \boldsymbol{\theta}} l_i^p(\boldsymbol{\theta}, \alpha; \mathbf{y}_i) \right] = \frac{\partial}{\partial \boldsymbol{\theta}^t} \mathbb{S}_{p_i}(\boldsymbol{\theta})$$

Y según con lo anterior, en adelante se usa la siguiente notación:

$$\dot{w}_i = \frac{\partial w_i}{\partial u_i}, \quad \ddot{\Sigma}_{i\theta\theta} = \frac{\partial^2 \Sigma_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t}, \quad \mathbb{H}_i(\theta_j, \theta_k) = \frac{\partial^2}{\partial \theta_j \partial \theta_k} l_{p_i}(\boldsymbol{\theta}, \alpha; \mathbf{y}_i)$$

para $j, k = 1, \dots, p^*$ con $p^* = p + r + s + 2$ y en las siguientes secciones se describen las expresiones de $\mathbb{H}_i(\theta_j, \theta_k)$.

II.2.1. Componente paramétrica

La matriz Hessiana de la componente paramétrica del MMS elíptico de la i -ésima unidad de estudio puede escribirse de la siguiente manera:

$$\begin{aligned} \mathbb{H}_i(\boldsymbol{\beta}, \boldsymbol{\beta}) &= \frac{\partial}{\partial \boldsymbol{\beta}^t} \mathbb{S}_{p_i}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}^t} [w_i \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i] = \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \frac{\partial [w_i \boldsymbol{\varepsilon}_i]}{\partial \boldsymbol{\beta}^t} \\ &= \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \left[\boldsymbol{\varepsilon}_i \frac{\partial w_i}{\partial \boldsymbol{\beta}^t} + w_i \frac{\partial \boldsymbol{\varepsilon}_i}{\partial \boldsymbol{\beta}^t} \right] = \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \left[\boldsymbol{\varepsilon}_i \left(\frac{\partial w_i}{\partial u_i} \frac{\partial u_i}{\partial \boldsymbol{\beta}^t} \right) - w_i \mathbf{X}_i \right] \\ &= \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \left[\boldsymbol{\varepsilon}_i \dot{w}_i \left[\frac{\partial u_i}{\partial \boldsymbol{\beta}} \right]^t - w_i \mathbf{X}_i \right], \end{aligned}$$

note que $\mathbb{H}_i(\boldsymbol{\beta}, \boldsymbol{\beta}) \in \mathbb{R}^{p \times p}$, y al desarrollar:

$$\left[\frac{\partial u_i}{\partial \boldsymbol{\beta}} \right]^t = [-2\mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i]^t = -2\boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i,$$

obtenemos:

$$\begin{aligned} \mathbb{H}_i(\boldsymbol{\beta}, \boldsymbol{\beta}) &= \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} [-2\dot{w}_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i - w_i \mathbf{X}_i] \\ &= -\mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} [2\dot{w}_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t + w_i \boldsymbol{\Sigma}_i] \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \end{aligned}$$

II.2.2. Componente no paramétrica

La matriz Hessiana de la componente no paramétrica del MMS elíptico de la i -ésima unidad de estudio puede escribirse de la siguiente manera:

$$\mathbb{H}_i(\mathbf{f}, \mathbf{f}) = \frac{\partial}{\partial \mathbf{f}^t} \mathbb{S}_{p_i}(\mathbf{f}) = \frac{\partial}{\partial \mathbf{f}^t} \left[w_i \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \frac{\alpha}{n} \mathbf{K} \mathbf{f} \right] = \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \frac{\partial [w_i \boldsymbol{\varepsilon}_i]}{\partial \mathbf{f}^t} - \frac{\alpha}{n} \mathbf{K}$$

note que $\mathbb{H}_i(\mathbf{f}, \mathbf{f}) \in \mathbb{R}^{r \times r}$, y al desarrollar:

$$\begin{aligned} \frac{\partial [w_i \boldsymbol{\varepsilon}_i]}{\partial \mathbf{f}^t} &= \boldsymbol{\varepsilon}_i \frac{\partial w_i}{\partial \mathbf{f}^t} + w_i \frac{\partial \boldsymbol{\varepsilon}_i}{\partial \mathbf{f}^t} = \boldsymbol{\varepsilon}_i \frac{\partial w_i}{\partial u_i} \frac{\partial u_i}{\partial \mathbf{f}^t} + w_i \frac{\partial \boldsymbol{\varepsilon}_i}{\partial \mathbf{f}} \\ &= \boldsymbol{\varepsilon}_i \dot{w}_i \left[\frac{\partial u_i}{\partial \mathbf{f}} \right]^t - w_i \mathbf{N}_i = \boldsymbol{\varepsilon}_i \dot{w}_i [-2\mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i]^t - w_i \mathbf{N}_i \\ &= - [2\dot{w}_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t + w_i \boldsymbol{\Sigma}_i] \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i \end{aligned}$$

obtenemos:

$$\mathbb{H}_i(\mathbf{f}, \mathbf{f}) = -\mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} [2\dot{w}_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t + w_i \boldsymbol{\Sigma}_i] \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i - \frac{\alpha}{n} \mathbf{K}$$

II.2.3. Componentes paramétrico y no paramétrico

La matriz Hessiana de las componentes paramétrica y no paramétrica del MMS elíptico de la i -ésima unidad de estudio puede escribirse de la siguiente manera:

$$\begin{aligned} \mathbb{H}_i(\boldsymbol{\beta}, \mathbf{f}) &= \frac{\partial}{\partial \mathbf{f}^t} \mathbb{S}_{p_i}(\boldsymbol{\beta}) = \frac{\partial}{\partial \mathbf{f}^t} [w_i \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i] = \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \frac{\partial [w_i \boldsymbol{\varepsilon}_i]}{\partial \mathbf{f}^t} \\ &= -\mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} [2\dot{w}_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t + w_i \boldsymbol{\Sigma}_i] \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i \end{aligned}$$

donde $\mathbb{H}_i(\boldsymbol{\beta}, \mathbf{f}) \in \mathbb{R}^{p \times r}$.

II.2.4. Componentes de varianza y paramétrico

La matriz Hessiana de los componentes de varianza y paramétrico del MMS elíptico de la i -ésima unidad de estudio puede escribirse de la siguiente manera:

$$\mathbb{H}_i(\boldsymbol{\beta}, \boldsymbol{\tau}) = \frac{\partial}{\partial \boldsymbol{\tau}^t} \mathbb{S}_{p_i}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\tau}^t} [w_i \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i] = \mathbf{X}_i^t \frac{\partial [w_i \boldsymbol{\Sigma}_i^{-1}]}{\partial \boldsymbol{\tau}^t} \boldsymbol{\varepsilon}_i$$

note que $\mathbb{H}_i(\boldsymbol{\beta}, \boldsymbol{\tau}) \in \mathbb{R}^{p \times (s+2)}$, y al desarrollar:

$$\begin{aligned} \frac{\partial [w_i \boldsymbol{\Sigma}_i^{-1}]}{\partial \boldsymbol{\tau}^t} &= \boldsymbol{\Sigma}_i^{-1} \frac{\partial w_i}{\partial \boldsymbol{\tau}^t} + w_i \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \boldsymbol{\tau}^t} = \boldsymbol{\Sigma}_i^{-1} \frac{\partial w_i}{\partial u_i} \frac{\partial u_i}{\partial \boldsymbol{\tau}^t} - w_i \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \boldsymbol{\tau}^t} \boldsymbol{\Sigma}_i^{-1} \\ &= -\dot{w}_i \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} - w_i \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \\ &= -\dot{w}_i \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} - w_i \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \\ &= -\boldsymbol{\Sigma}_i^{-1} [\dot{w}_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t + w_i \boldsymbol{\Sigma}_i] \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \end{aligned}$$

obtenemos:

$$\mathbb{H}_i(\boldsymbol{\beta}, \boldsymbol{\tau}) = -\mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} [\dot{w}_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t + w_i \boldsymbol{\Sigma}_i] \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i$$

II.2.5. Componentes de varianza y no paramétrico

La matriz Hessiana de la parte no paramétrica y los componentes de varianza del MMS elíptico de la i -ésima unidad de estudio puede escribirse de la siguiente manera:

$$\mathbb{H}_i(\mathbf{f}, \boldsymbol{\tau}) = \frac{\partial}{\partial \boldsymbol{\tau}} \mathbb{S}_{p_i}(\mathbf{f}) = \frac{\partial}{\partial \boldsymbol{\tau}} \left[w_i \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \frac{\alpha}{n} \mathbf{K} \mathbf{f} \right] = \mathbf{N}_i^t \frac{\partial [w_i \boldsymbol{\Sigma}_i^{-1}]}{\partial \boldsymbol{\tau}} \boldsymbol{\varepsilon}_i$$

note que $\mathbb{H}_i(\mathbf{f}, \boldsymbol{\tau}) \in \mathbb{R}^{r \times (s+2)}$, y por lo tanto, la matriz Hessiana está dada por:

$$\mathbb{H}_i(\mathbf{f}, \boldsymbol{\tau}) = -\mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} [\dot{w}_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t + w_i \boldsymbol{\Sigma}_i] \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i$$

II.2.6. Componentes de varianza

La matriz Hessiana de los componentes de varianza del MMS elíptico de la i -ésima unidad de estudio puede escribirse de la siguiente manera:

$$\mathbb{H}_i(\boldsymbol{\tau}, \boldsymbol{\tau}) = \frac{\partial}{\partial \boldsymbol{\tau}} \mathbb{S}_{p_i}(\boldsymbol{\tau}) = \frac{\partial}{\partial \boldsymbol{\tau}} \left\{ -\frac{1}{2} \left[\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau} \right) - w_i \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right] \right\}$$

Correspondientemente, para el (j, k) -ésimo elemento de la matriz $\mathbb{H}_i(\boldsymbol{\tau}, \boldsymbol{\tau})$ con $j, k = 1, \dots, s, s+1, s+2$, podemos escribir lo siguiente:

$$\mathbb{H}_i(\tau_j, \tau_k) = \frac{\partial}{\partial \tau_j} \left\{ -\frac{1}{2} \left[\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) - w_i \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right] \right\},$$

y al desarrollar la derivada matricial, tenemos que:

$$\begin{aligned} \mathbb{H}_i(\tau_j, \tau_k) &= -\frac{1}{2} \frac{\partial}{\partial \tau_j} \left[\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_k} \right) - w_i \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_k} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right] \\ &= -\frac{1}{2} \left\{ \text{tr} \left(\frac{\partial}{\partial \tau_j} \left[\boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_k} \right] \right) - \boldsymbol{\varepsilon}_i^t \frac{\partial}{\partial \tau_j} \left[w_i \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_k} \boldsymbol{\Sigma}_i^{-1} \right] \boldsymbol{\varepsilon}_i \right\} \\ &= -\frac{1}{2} \left\{ \text{tr} \left(\frac{\partial}{\partial \tau_j} \left[\boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_k} \right] \right) - \boldsymbol{\varepsilon}_i^t \frac{\partial \psi}{\partial \tau_j} \boldsymbol{\varepsilon}_i \right\}, \end{aligned}$$

pero la derivada matricial dentro de la traza es:

$$\begin{aligned} \frac{\partial}{\partial \tau_j} \left[\Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_k} \right] &= \frac{\partial \Sigma_i^{-1}}{\partial \tau_j} \frac{\partial \Sigma_i}{\partial \tau_k} + \Sigma_i^{-1} \frac{\partial^2 \Sigma_i}{\partial \tau_j \partial \tau_k} = -\Sigma_i^{-1} \dot{\Sigma}_{i\tau_j} \Sigma_i^{-1} \dot{\Sigma}_{i\tau_k} + \Sigma_i^{-1} \ddot{\Sigma}_{i\tau_j \tau_k} \\ &= -\Sigma_i^{-1} \left[\dot{\Sigma}_{i\tau_j} \Sigma_i^{-1} \dot{\Sigma}_{i\tau_k} - \ddot{\Sigma}_{i\tau_j \tau_k} \right], \end{aligned}$$

y la derivada de ψ queda dado por:

$$\begin{aligned} \frac{\partial \psi}{\partial \tau_j} &= \frac{\partial w_i}{\partial \tau_j} \Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_k} \Sigma_i^{-1} + w_i \frac{\partial \Sigma_i^{-1}}{\partial \tau_j} \frac{\partial \Sigma_i}{\partial \tau_k} \Sigma_i^{-1} + w_i \Sigma_i^{-1} \frac{\partial^2 \Sigma_i}{\partial \tau_j \tau_k} \Sigma_i^{-1} \\ &+ w_i \Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_k} \frac{\partial \Sigma_i^{-1}}{\partial \tau_j} = \psi_1 + \psi_2 + \psi_3 + \psi_4 = \sum_{i=1}^4 \psi_i, \end{aligned}$$

al desarrollar por separado cada una de las expresiones de ψ_i , obtenemos:

$$\begin{aligned} \psi_1 &= \frac{\partial w_i}{\partial \tau_j} \Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_k} \Sigma_i^{-1} = \frac{\partial w_i}{\partial u_i} \Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_k} \Sigma_i^{-1} \frac{\partial w_i}{\partial u_i} \\ &= \dot{w}_i \Sigma_i^{-1} \dot{\Sigma}_{i\tau_k} \Sigma_i^{-1} \left[-\epsilon_i \epsilon_i^t \Sigma_i^{-1} \dot{\Sigma}_{i\tau_j} \Sigma_i^{-1} \right] \\ &= -\dot{w}_i \Sigma_i^{-1} \dot{\Sigma}_{i\tau_k} \Sigma_i^{-1} \epsilon_i \epsilon_i^t \Sigma_i^{-1} \dot{\Sigma}_{i\tau_j} \Sigma_i^{-1} \\ \psi_2 &= w_i \frac{\partial \Sigma_i^{-1}}{\partial \tau_j} \frac{\partial \Sigma_i}{\partial \tau_k} \Sigma_i^{-1} = -w_i \Sigma_i^{-1} \dot{\Sigma}_{i\tau_j} \Sigma_i^{-1} \Sigma_i^{-1} \dot{\Sigma}_{i\tau_k} \Sigma_i^{-1} \\ \psi_3 &= w_i \Sigma_i^{-1} \frac{\partial^2 \Sigma_i}{\partial \tau_j \tau_k} \Sigma_i^{-1} = w_i \Sigma_i^{-1} \ddot{\Sigma}_{i\tau_j \tau_k} \Sigma_i^{-1} \\ \psi_4 &= w_i \Sigma_i^{-1} \frac{\partial \Sigma_i}{\partial \tau_k} \frac{\partial \Sigma_i^{-1}}{\partial \tau_j} = -w_i \Sigma_i^{-1} \dot{\Sigma}_{i\tau_k} \Sigma_i^{-1} \Sigma_i^{-1} \dot{\Sigma}_{i\tau_j} \Sigma_i^{-1} \end{aligned}$$

Además, note que $\psi_2 = [\psi_4]^t$ y al sumar las cuatro expresiones anteriores:

$$\begin{aligned} \frac{\partial \psi}{\partial \tau_j} &= -\dot{w}_i \Sigma_i^{-1} \dot{\Sigma}_{i\tau_k} \Sigma_i^{-1} \epsilon_i \epsilon_i^t \Sigma_i^{-1} \dot{\Sigma}_{i\tau_j} \Sigma_i^{-1} - 2w_i \Sigma_i^{-1} \dot{\Sigma}_{i\tau_k} \Sigma_i^{-1} \Sigma_i^{-1} \dot{\Sigma}_{i\tau_j} \Sigma_i^{-1} \\ &+ w_i \Sigma_i^{-1} \ddot{\Sigma}_{i\tau_j \tau_k} \Sigma_i^{-1} \\ &= -\Sigma_i^{-1} \left[\dot{\Sigma}_{i\tau_k} \Sigma_i^{-1} (\dot{w}_i \epsilon_i \epsilon_i^t + 2w_i \Sigma_i) \Sigma_i^{-1} \dot{\Sigma}_{i\tau_j} \right] \Sigma_i^{-1} \end{aligned}$$

De esta manera, obtenemos:

$$\begin{aligned} \mathbb{H}_i(\tau_j, \tau_k) = & -\frac{1}{2} \left\{ \text{tr} \left(-\Sigma_i^{-1} \left[\dot{\Sigma}_{i\tau_j} \Sigma_i^{-1} \dot{\Sigma}_{i\tau_k} - \ddot{\Sigma}_{i\tau_j\tau_k} \right] \right) \right. \\ & \left. + \boldsymbol{\varepsilon}_i^t \Sigma_i^{-1} \left[\dot{\Sigma}_{i\tau_k} \Sigma_i^{-1} (\dot{w}_i \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^t + 2w_i \Sigma_i) \Sigma_i^{-1} \dot{\Sigma}_{i\tau_j} - w_i \ddot{\Sigma}_{i\tau_j\tau_k} \right] \Sigma_i^{-1} \boldsymbol{\varepsilon}_i \right\} \end{aligned}$$

II.3. Información de Fisher penalizada

En adelante para el cálculo de $\mathcal{I}_p(\boldsymbol{\theta})$ nos basaremos en la expresión (3.16) en conjunto con las propiedades (1.13)-(1.16) expuestas en la sección 4 del capítulo 1, con $\mathbf{P}_i = \Sigma_i^{-1/2} \boldsymbol{\varepsilon}_i$.

II.3.1. Componente paramétrica

La función de score penalizada de la parte paramétrica asociada a la i -ésima unidad de estudio está dada por:

$$\mathbb{S}_{p_i}(\boldsymbol{\beta}) = w_i \mathbf{X}_i^t \Sigma_i^{-1} \boldsymbol{\varepsilon}_i = w_i \mathbf{X}_i^t \Sigma_i^{-1/2} \Sigma_i^{-1/2} \boldsymbol{\varepsilon}_i = w_i \mathbf{X}_i^t \Sigma_i^{-1/2} \mathbf{P}_i$$

Consecuentemente, la función de score penalizada asociada a la i -ésima unidad respecto del j -ésimo elemento del vector $\boldsymbol{\beta}$ es:

$$\mathbb{S}_{p_i}(\beta_j) = w_i \mathbf{x}_{ij}^t \Sigma_i^{-1} \boldsymbol{\varepsilon}_i = w_i \mathbf{x}_{ij}^t \Sigma_i^{-1/2} \Sigma_i^{-1/2} \boldsymbol{\varepsilon}_i = w_i \mathbf{x}_{ij}^t \Sigma_i^{-1/2} \mathbf{P}_i$$

donde \mathbf{x}_{ij} corresponde a la j -ésima columna de la matriz de diseño de efectos fijos \mathbf{X}_i y $\mathbf{P} \sim S_{n_i}(g)$. Entonces, el producto entre las funciones de scores asociada a la i -ésima unidad respecto del j -ésimo y k -ésimo elemento del vector $\boldsymbol{\beta}$, con $j \neq k$ es:

$$\begin{aligned} \mathbb{S}_{p_i}(\beta_j) \mathbb{S}_{p_i}(\beta_k) &= \left[w_i \mathbf{x}_{ij}^t \Sigma_i^{-1/2} \mathbf{P}_i \right]^t \left[w_i \mathbf{x}_{ik}^t \Sigma_i^{-1/2} \mathbf{P}_i \right] \\ &= w_i^2 \mathbf{P}_i^t \Sigma_i^{-1/2} \mathbf{x}_{ij} \mathbf{x}_{ik}^t \Sigma_i^{-1/2} \mathbf{P}_i \\ &= w_i^2 \|\mathbf{P}_i\|^2 \frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \Sigma_i^{-1/2} \mathbf{x}_{ij} \mathbf{x}_{ik}^t \Sigma_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \end{aligned}$$

Pero sabemos que $\mathcal{I}_{p_i}(\beta_j, \beta_k) = \mathbb{E} [\mathbb{S}_{p_i}(\beta_j)\mathbb{S}_{p_i}(\beta_k)]$, y así obtenemos:

$$\begin{aligned}
\mathcal{I}_{p_i}(\beta_j, \beta_k) &= \mathbb{E} \left[\mathbb{E} \left(w_i^2 \|\mathbf{P}_i\|^2 \frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \Sigma_i^{-1/2} \mathbf{x}_{ij} \mathbf{x}_{ij}^t \Sigma_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\| \right) \right] \\
&= \mathbb{E} \left[w_i^2 \|\mathbf{P}_i\|^2 \mathbb{E} \left(\frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \Sigma_i^{-1/2} \mathbf{x}_{ij} \mathbf{x}_{ij}^t \Sigma_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\| \right) \right] \\
&= \mathbb{E} \left[w_i^2 \|\mathbf{P}_i\|^2 \frac{1}{n_i} \text{tr} \left(\Sigma_i^{-1/2} \mathbf{x}_{ij} \mathbf{x}_{ij}^t \Sigma_i^{-1/2} \right) \right] \\
&= \frac{1}{n_i} \text{tr} \left(\mathbf{x}_{ij}^t \Sigma_i^{-1/2} \Sigma_i^{-1/2} \mathbf{x}_{ij} \right) \mathbb{E} [w_i^2 \|\mathbf{P}_i\|^2] = \frac{4d_{g_i}}{n_i} \mathbf{x}_{ij}^t \Sigma_i^{-1} \mathbf{x}_{ij}
\end{aligned}$$

Por lo tanto, tenemos que la matriz de información de Fisher de la parte paramétrica asociada a la i -ésima unidad es:

$$\mathcal{I}_{p_i}(\boldsymbol{\beta}, \boldsymbol{\beta}) = \frac{4d_{g_i}}{n_i} \mathbf{X}_i^t \Sigma_i^{-1} \mathbf{X}_i$$

Finalmente, la matriz de información de Fisher penalizada de la parte paramétrica del MMS elíptico de las n unidades de estudio está dada por:

$$\begin{aligned}
\mathcal{I}_p(\boldsymbol{\beta}, \boldsymbol{\beta}) &= \sum_{i=1}^n \mathcal{I}_{p_i}(\boldsymbol{\beta}, \boldsymbol{\beta}) = \sum_{i=1}^n \frac{4d_{g_i}}{n_i} \mathbf{X}_i^t \Sigma_i^{-1} \mathbf{X}_i \\
&= \frac{4d_{g_1}}{n_1} \mathbf{X}_1^t \Sigma_1^{-1} \mathbf{X}_1 + \dots + \frac{4d_{g_n}}{n_n} \mathbf{X}_n^t \Sigma_n^{-1} \mathbf{X}_n \\
&= [\mathbf{X}_1^t \quad \dots \quad \mathbf{X}_n^t] \begin{bmatrix} \frac{4d_{g_1}}{n_1} \Sigma_1^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{4d_{g_2}}{n_2} \Sigma_2^{-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \frac{4d_{g_n}}{n_n} \Sigma_n^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \\
&= \mathbf{X}^t \boldsymbol{\Pi} \mathbf{X}
\end{aligned}$$

donde $\mathbf{X} \in \mathbb{R}^{m \times p}$ y $\boldsymbol{\Pi} \in \mathbb{R}^{m \times m}$ con $\boldsymbol{\Pi} = \text{diag} \left(\frac{4d_{g_1}}{n_1} \Sigma_1^{-1}, \dots, \frac{4d_{g_n}}{n_n} \Sigma_n^{-1} \right)$.

II.3.2. Componente no paramétrica

La función de score penalizada de la parte no paramétrica asociada a la i -ésima unidad de estudio está dada por:

$$\begin{aligned}\mathbb{S}_{p_i}(\mathbf{f}) &= w_i \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i - \frac{\alpha}{n} \mathbf{K} \mathbf{f} = w_i \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\varepsilon}_i - \frac{\alpha}{n} \mathbf{K} \mathbf{f} \\ &= w_i \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i - \frac{\alpha}{n} \mathbf{K} \mathbf{f}\end{aligned}$$

Además, sabemos que $\mathbb{S}_{p_i}(\mathbf{f}) = \mathbb{S}_i(\mathbf{f}) - \frac{\alpha}{n} \mathbf{K} \mathbf{f}$ y consecuentemente tenemos que la función de score penalizada asociada a la i -ésima unidad respecto del j -ésimo elemento del vector \mathbf{f} es:

$$\mathbb{S}_{p_i}(f_j^0) = \mathbb{S}_i(f_j^0) - \frac{\alpha}{n} \mathbf{K} \mathbf{f} = w_i \mathbf{n}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \mathbf{P}_i - \frac{\alpha}{n} \mathbf{K} \mathbf{f}$$

donde \mathbf{n}_{ij} corresponde a la j -ésima columna de \mathbf{N}_i y $\mathbf{P}_i \sim S_{n_i}(g)$. Entonces, al desarrollar las expresiones de las funciones de scores asociada a la i -ésima unidad respecto del j -ésimo y k -ésimo elemento del vector \mathbf{f} , con $j \neq k$ es:

$$\begin{aligned}\frac{\partial^2 l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial f_j^0 \partial f_k^0} &= \frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial f_j^0 \partial f_k^0} - \frac{\alpha}{n} \mathbf{K} \\ -\frac{\partial^2 l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial f_j^0 \partial f_k^0} &= -\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial f_j^0 \partial f_k^0} + \frac{\alpha}{n} \mathbf{K} \\ \mathbb{S}_{p_i}(f_j^0) \mathbb{S}_{p_i}(f_k^0) &= \mathbb{S}_i(f_j^0) \mathbb{S}_i(f_k^0) + \frac{\alpha}{n} \mathbf{K} \\ \mathbb{E} [\mathbb{S}_{p_i}(f_j^0) \mathbb{S}_{p_i}(f_k^0)] &= \mathbb{E} [\mathbb{S}_i(f_j^0) \mathbb{S}_i(f_k^0)] + \frac{\alpha}{n} \mathbf{K} \\ \mathcal{I}_{p_i}(f_j^0, f_k^0) &= \mathbb{E} \{ [\mathbb{S}_i(f_j^0) \mathbb{S}_i(f_k^0) \mid \|\mathbf{P}_i\|] \} + \frac{\alpha}{n} \mathbf{K}\end{aligned}$$

pero antes, tenemos:

$$\begin{aligned}\mathbb{E} [\mathbb{S}_i(f_j^0) \mathbb{S}_i(f_k^0)] &= \mathbb{E} \left[\mathbb{E} \left(w_i^2 \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \mathbf{n}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \mid \|\mathbf{P}_i\| \right) \right] \\ &= \mathbb{E} \left[\mathbb{E} \left(w_i^2 \|\mathbf{P}_i\|^2 \frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \mathbf{n}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \mid \|\mathbf{P}_i\| \right) \right]\end{aligned}$$

$$\begin{aligned}
\mathbb{E} [\mathbb{S}_i(\mathbf{f}_j^0)\mathbb{S}_i(\mathbf{f}_k^0)] &= \mathbb{E} \left[w_i^2 \|\mathbf{P}_i\|^2 \mathbb{E} \left(\frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \mathbf{n}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\| \right) \right] \\
&= \mathbb{E} \left[w_i^2 \|\mathbf{P}_i\|^2 \frac{1}{n_i} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \mathbf{n}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \right) \right] \\
&= \frac{1}{n_i} \text{tr} \left(\mathbf{n}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \right) \mathbb{E} [w_i^2 \|\mathbf{P}_i\|^2] = \frac{4d_{g_i}}{n_i} \mathbf{n}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \mathbf{n}_{ij}
\end{aligned}$$

Por lo tanto, tenemos que la matriz de información de Fisher penalizada asociada a la i -ésima unidad es:

$$\mathcal{I}_{p_i}(\mathbf{f}, \mathbf{f}) = \frac{4d_{g_i}}{n_i} \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i + \frac{\alpha}{n} \mathbf{K}$$

Finalmente, la matriz de información de Fisher penalizada de la parte no paramétrica del MMS elíptico de las n unidades de estudio está dada por:

$$\begin{aligned}
\mathcal{I}_p(\mathbf{f}, \mathbf{f}) &= \sum_{i=1}^n \mathcal{I}_{p_i}(\mathbf{f}, \mathbf{f}) = \sum_{i=1}^n \left(\frac{4d_{g_i}}{n_i} \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i + \frac{\alpha}{n} \mathbf{K} \right) \\
&= \frac{4d_{g_1}}{n_1} \mathbf{N}_1^t \boldsymbol{\Sigma}_1^{-1} \mathbf{N}_1 + \dots + \frac{4d_{g_n}}{n_n} \mathbf{N}_n^t \boldsymbol{\Sigma}_n^{-1} \mathbf{N}_n + \alpha \mathbf{K} \\
&= [\mathbf{N}_1^t \quad \dots \quad \mathbf{N}_n^t] \begin{bmatrix} \frac{4d_{g_1}}{n_1} \boldsymbol{\Sigma}_1^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{4d_{g_2}}{n_2} \boldsymbol{\Sigma}_2^{-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \frac{4d_{g_n}}{n_n} \boldsymbol{\Sigma}_n^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{N}_1 \\ \vdots \\ \mathbf{N}_n \end{bmatrix} + \alpha \mathbf{K} \\
&= \mathbf{N}^t \boldsymbol{\Pi} \mathbf{N} + \alpha \mathbf{K}
\end{aligned}$$

donde $\boldsymbol{\Pi} \in \mathbb{R}^{m \times m}$ y $\mathbf{N} \in \mathbb{R}^{m \times r}$.

II.3.3. Componentes paramétrica y no paramétrica

Las funciones de score penalizada para la parte paramétrica y no paramétrica asociada a la i -ésima unidad de estudio respecto del j -ésimo elemento del vector $\boldsymbol{\beta}$ y del k -ésimo elemento del vector \mathbf{f} está dada por:

$$\begin{aligned}
\mathbb{S}_{p_i}(\boldsymbol{\beta}) &= \mathbb{S}_i(\boldsymbol{\beta}) = w_i \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \implies \mathbb{S}_i(\beta_j) = w_i \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \\
\mathbb{S}_{p_i}(\mathbf{f}) &= w_i \mathbf{N}_i^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i - \frac{\alpha}{n} \mathbf{K} \mathbf{f} \implies \mathbb{S}_i(\mathbf{f}_j^0) = w_i \mathbf{n}_{ik}^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i
\end{aligned}$$

Además, sabemos que $\mathbb{S}_{p_i}(\mathbf{f}) = \mathbb{S}_i(\mathbf{f}) - \frac{\alpha}{n}\mathbf{K}\mathbf{f}$ y consecuentemente tenemos que la función de score penalizada asociada a la i -ésima unidad respecto del k -ésimo elemento del vector \mathbf{f} es:

$$\mathbb{S}_{p_i}(\mathbf{f}_k^0) = \mathbb{S}_i(\mathbf{f}_k^0) - \frac{\alpha}{n}\mathbf{K}\mathbf{f} = w_i \mathbf{n}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \mathbf{P}_i - \frac{\alpha}{n}\mathbf{K}\mathbf{f}$$

donde \mathbf{n}_{ij} corresponde a la j -ésima columna de la matriz \mathbf{N}_i . Entonces, al desarrollar el (j, k) -ésimo de la matriz de información de Fisher penalizada asociada a la i -ésima unidad respecto del j -ésimo elemento de $\boldsymbol{\beta}$ y el k -ésimo elemento del vector \mathbf{f} está dado por:

$$\begin{aligned} \mathcal{I}_{p_i}(\beta_j, \mathbf{f}_k^0) &= \mathbb{E} \left[-\frac{\partial^2 l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \beta_j \partial \mathbf{f}_k^0} \right] = \mathbb{E} \left[-\frac{\partial}{\partial \beta_j} \left\{ \mathbb{S}_i(\mathbf{f}_k^0) + \frac{\alpha}{n}\mathbf{K}\mathbf{f} \right\} \right] = \mathbb{E} \left[-\frac{\partial^2 l_i(\boldsymbol{\theta})}{\partial \beta_j \partial \mathbf{f}_k^0} \right] \\ &= \mathbb{E} [\mathbb{S}_i(\beta_j) \mathbb{S}_i(\mathbf{f}_k^0)] = \mathbb{E} \left\{ \mathbb{E} [\mathbb{S}_i(\beta_j) \mathbb{S}_i(\mathbf{f}_k^0) \mid \|\mathbf{P}_i\|] \right\} \end{aligned}$$

Según lo anterior, al desarrollar el producto de las funciones de Score no penalizadas obtenemos:

$$\begin{aligned} \mathbb{S}_i(\beta_j) \mathbb{S}_i(\mathbf{f}_k^0) &= \left[w_i \mathbf{n}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right]^t \left[w_i \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right] = w_i^2 \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \\ &= w_i^2 \|\mathbf{P}_i\|^2 \frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \end{aligned}$$

Luego, desarrollando tenemos que:

$$\begin{aligned} \mathcal{I}_{p_i}(\beta_j, \mathbf{f}_k^0) &= \mathbb{E} \left[\mathbb{E} \left(w_i^2 \|\mathbf{P}_i\|^2 \frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \mid \|\mathbf{P}_i\| \right) \right] \\ &= \mathbb{E} \left[w_i^2 \|\mathbf{P}_i\|^2 \mathbb{E} \left(\frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \mid \|\mathbf{P}_i\| \right) \right] \\ &= \mathbb{E} \left[w_i^2 \|\mathbf{P}_i\|^2 \frac{1}{n_i} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \right) \right] \\ &= \frac{1}{n_i} \text{tr} \left(\mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{n}_{ij} \right) \mathbb{E} [w_i^2 \|\mathbf{P}_i\|^2] = \frac{4d_{g_i}}{n_i} \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \mathbf{n}_{ij} \end{aligned}$$

Por lo tanto, la matriz de información de Fisher penalizada asociada a la i -ésima unidad es:

$$\mathcal{I}_{p_i}(\boldsymbol{\beta}, \mathbf{f}) = \frac{4d_{g_i}}{n_i} \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i$$

Finalmente, la matriz de información de Fisher penalizada de la parte paramétrica y no paramétrica del MMS elíptico de las n unidades de estudio está dada por:

$$\begin{aligned}
\mathcal{I}_p(\boldsymbol{\beta}, \mathbf{f}) &= \sum_{i=1}^n \mathcal{I}_{p_i}(\boldsymbol{\beta}, \mathbf{f}) = \sum_{i=1}^n \left(\frac{4d_{g_i}}{n_i} \mathbf{X}_i^t \boldsymbol{\Sigma}_i^{-1} \mathbf{N}_i \right) \\
&= \frac{4d_{g_1}}{n_1} \mathbf{X}_1^t \boldsymbol{\Sigma}_1^{-1} \mathbf{N}_1 + \cdots + \frac{4d_{g_n}}{n_n} \mathbf{X}_n^t \boldsymbol{\Sigma}_n^{-1} \mathbf{N}_n \\
&= [\mathbf{X}_1^t \quad \cdots \quad \mathbf{X}_n^t] \begin{bmatrix} \frac{4d_{g_1}}{n_1} \boldsymbol{\Sigma}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{4d_{g_2}}{n_2} \boldsymbol{\Sigma}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{4d_{g_n}}{n_n} \boldsymbol{\Sigma}_n^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{N}_1 \\ \vdots \\ \mathbf{N}_n \end{bmatrix} \\
&= \mathbf{X}^t \mathbf{\Pi} \mathbf{N} \tag{7.1}
\end{aligned}$$

donde $\mathbf{\Pi} \in \mathbb{R}^{m \times m}$, $\mathbf{X} \in \mathbb{R}^{m \times p}$ y $\mathbf{N} \in \mathbb{R}^{m \times r}$.

II.3.4. Componentes de varianza

La función de score penalizada del vector de componentes de varianza asociada a la i -ésima unidad de estudio está dada por:

$$\mathbb{S}_{p_i}(\boldsymbol{\tau}) = -\frac{1}{2} \left[\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \right) - w_i \boldsymbol{\epsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\boldsymbol{\tau}} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\epsilon}_i \right]$$

Consecuentemente, la función de score penalizada asociada a la i -ésima unidad respecto del j -ésimo elemento del vector $\boldsymbol{\tau}$ es:

$$\begin{aligned}
\frac{\partial l_{p_i}(\boldsymbol{\theta})}{\partial \tau_j} &= -\frac{1}{2} \left[\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) - w_i \boldsymbol{\epsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\epsilon}_i \right] \\
&= -\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) + \frac{w_i}{2} \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i
\end{aligned}$$

Al desarrollar las expresiones de las funciones de scores de la i -ésima unidad respecto del j -ésimo y k -ésimo elemento del vector $\boldsymbol{\tau}$, con $j \neq k$, tenemos:

$$\begin{aligned} \frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \tau_j} \frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \tau_k} &= \left[-\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) + \frac{w_i}{2} \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right]^t \\ &\quad \left[-\frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) + \frac{w_i}{2} \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \right] \\ &= \frac{1}{4} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) \\ &\quad - \frac{w_i}{4} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \\ &\quad - \frac{w_i}{4} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \\ &\quad + \frac{w_i^2}{4} \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \end{aligned}$$

donde $b_{jk}^i = \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right)$ y entonces:

$$\begin{aligned} \mathcal{I}_{p_i}(\tau_j, \tau_k) &= \mathbb{E} \left[\frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \tau_j} \frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \tau_k} \right] = \mathbb{E} \left[\mathbb{E} \left(\frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \tau_j} \frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \tau_k} \middle| \|\mathbf{P}_i\| \right) \right] \\ &= \sum_{h=1}^4 \mathbb{E} \left[\mathbb{E} (\mathbf{A}_h | \|\mathbf{P}_i\|) \right] = \sum_{h=1}^4 \varphi_h \end{aligned}$$

Al desarrollar las esperanzas de la sumatoria por separado, tenemos que:

$$\begin{aligned} \varphi_1 &= \mathbb{E} \left[\mathbb{E} (\mathbf{A}_1 | \|\mathbf{P}_i\|) \right] = \mathbb{E} \left[\mathbb{E} \left(\frac{b_{ij}^i}{4} \middle| \|\mathbf{P}_i\| \right) \right] = \frac{b_{ij}^i}{4} \\ \varphi_2 &= \mathbb{E} \left[\mathbb{E} \left(-\frac{w_i}{4} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \middle| \|\mathbf{P}_i\| \right) \right] \\ &= \mathbb{E} \left[-\frac{w_i}{4} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) \|\mathbf{P}_i\|^2 \mathbb{E} \left(\frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\| \right) \right] \\ &= \mathbb{E} \left[-\frac{w_i}{4} \|\mathbf{P}_i\|^2 \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) \frac{1}{n_i} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \right) \right] \end{aligned}$$

$$\begin{aligned}\varphi_2 &= -\frac{1}{4n_i} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) \text{tr} \left(\boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) [w_i \|\mathbf{P}_i\|^2] \\ &= -\frac{1}{4} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) = -\frac{b_{jk}^i}{4}\end{aligned}$$

De manera análoga, se puede demostrar que $\varphi_3 = \mathbb{E} [\mathbb{E} (\mathbf{A}_3 | \|\mathbf{P}_i\|)] = -\frac{b_{jk}^i}{4}$.

$$\begin{aligned}\varphi_4 &= \mathbb{E} \left[\mathbb{E} \left(\frac{w_i^2}{4} \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \mathbf{P}_i^t \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i \middle| \|\mathbf{P}_i\| \right) \right] \\ &= \mathbb{E} \left[\frac{w_i^2}{4} \|\mathbf{P}_i\|^4 \mathbb{E} \left(\frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \frac{\mathbf{P}_i^t}{\|\mathbf{P}_i\|} \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\| \right) \right] \\ &= \mathbb{E} \left\{ \frac{w_i^2 \|\mathbf{P}_i\|^4}{4n_i(n_i+2)} \mathbb{E} \left[2\text{tr} \left(\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \right) \right. \right. \\ &\quad \left. \left. + \text{tr} \left(\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1/2} \right) \text{tr} \left(\boldsymbol{\Sigma}_i^{-1/2} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \right) \right] \right\}\end{aligned}$$

$$\begin{aligned}\varphi_4 &= \mathbb{E} \left\{ \frac{w_i^2 \|\mathbf{P}_i\|^4}{4n_i(n_i+2)} \mathbb{E} \left[2\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) + \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \right) \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) \right] \right\} \\ &= \frac{f_{g_i}}{n_i(n_i+2)} \mathbb{E} \left[2\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) + b_{jk}^i \right]\end{aligned}$$

Por lo tanto, el (j, k) -ésimo elemento de la matriz de información de Fisher penalizada respecto del vector de parámetros $\boldsymbol{\tau}$ asociada a la i -ésima unidad de estudio está dado por:

$$\begin{aligned}\mathcal{I}_{p_i}(\tau_j, \tau_k) &= \frac{b_{jk}^i}{4} + \left(-\frac{b_{jk}^i}{2} \right) + \frac{f_{g_i}}{n_i(n_i+2)} \mathbb{E} \left[2\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) + b_{jk}^i \right] \\ &= \frac{b_{jk}^i}{4} \left(\frac{4f_{g_i}}{n_i(n_i+2)} - 1 \right) + \frac{2f_{g_i}}{n_i(n_i+2)} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_j} \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right)\end{aligned}$$

II.3.5. Componentes paramétrica y de varianza

El (j, k) -ésimo elemento de la matriz de información de Fisher penalizada $\mathcal{I}_{p_i}(\boldsymbol{\beta}, \boldsymbol{\tau})$ entre la parte paramétrica y los componentes de varianza asociada a la i -ésima unidad de estudio está dado por:

$$\mathcal{I}_{p_i}(\beta_j, \tau_k) = \mathbb{E} \left[-\frac{\partial^2 l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \beta_j \partial \tau_k} \right] = \mathbb{E} \left[\mathbb{E} \left(-\frac{\partial^2 l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \beta_j \partial \tau_k} \middle| \|\mathbf{P}_i\| \right) \right]$$

Al desarrollar la derivada doble, obtenemos:

$$\begin{aligned} -\frac{\partial^2 l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \beta_j \partial \tau_k} &= \frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \beta_j} \frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \tau_k} = \frac{\partial}{\partial \beta_j} \left(\frac{\partial l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial \tau_k} \right) \\ &= \frac{\partial}{\partial \beta_j} \left\{ -\frac{1}{2} \left[\text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \right) - w_i \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right] \right\} \\ &= \frac{\partial}{\partial \beta_j} \left[\frac{w_i}{2} \boldsymbol{\varepsilon}_i^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \right] = -\dot{w}_i \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i \\ &= -\dot{w}_i \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \mathbf{P}_i = -\dot{w}_i \|\mathbf{P}_i\| \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \end{aligned}$$

De esta manera, tenemos que:

$$\begin{aligned} \mathcal{I}_{p_i}(\beta_j, \tau_k) &= \mathbb{E} \left[\mathbb{E} \left(-\dot{w}_i \|\mathbf{P}_i\| \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\| \right) \right] \\ &= \mathbb{E} \left[-\dot{w}_i \|\mathbf{P}_i\| \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \mathbb{E} \left(\frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\| \right) \right] \end{aligned}$$

Sin embargo, puesto que es \dot{w}_i es función de $\|\mathbf{P}_i\|$, entonces podemos escribir lo siguiente:

$$\begin{aligned} \mathcal{I}_{p_i}(\beta_j, \tau_k) &= \mathbb{E} [-\dot{w}_i \|\mathbf{P}_i\|] \mathbb{E} \left[\mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \mathbb{E} \left(\frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\| \right) \right] \\ &= \mathbb{E} [-\dot{w}_i \|\mathbf{P}_i\|] \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \mathbb{E} \left[\mathbb{E} \left(\frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \middle| \|\mathbf{P}_i\| \right) \right] \\ &= \mathbb{E} [-\dot{w}_i \|\mathbf{P}_i\|] \mathbf{x}_{ij}^t \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_{i\tau_k} \boldsymbol{\Sigma}_i^{-1/2} \mathbb{E} \left[\frac{\mathbf{P}_i}{\|\mathbf{P}_i\|} \right] = 0 \end{aligned}$$

Por lo tanto, la matriz de información de Fisher penalizada entre la parte paramétrica y los componentes de varianza asociada a la i -ésima unidad de estudio es:

$$\mathcal{I}_{p_i}(\boldsymbol{\beta}, \boldsymbol{\tau}) = \mathbf{0}$$

II.3.6. Componentes no paramétrica y de varianza

El (j, k) -ésimo elemento de la matriz de información de Fisher penalizada $\mathcal{I}_{p_i}(\mathbf{f}, \boldsymbol{\tau})$ entre la parte no paramétrica y los componentes de varianza asociada a la i -ésima unidad de estudio está dado por:

$$\mathcal{I}_{p_i}(f_j^0, \tau_k) = \mathbb{E} \left[-\frac{\partial^2 l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial f_j^0 \partial \tau_k} \right] = \mathbb{E} \left[\mathbb{E} \left(-\frac{\partial^2 l_{p_i}(\boldsymbol{\theta}, \alpha)}{\partial f_j^0 \partial \tau_k} \middle| \|\mathbf{P}_i\| \right) \right]$$

Después de realizar un algebra matricial análogo al realizado en la sección anterior, se puede demostrar que la matriz de información de Fisher penalizada entre la parte no paramétrica y los componentes de varianza asociada a la i -ésima unidad de estudio es:

$$\mathcal{I}_{p_i}(\mathbf{f}, \boldsymbol{\tau}) = \mathbf{0}$$

II.4. Proceso de estimación

En esta sección se describen las demostraciones para la estimación de los efectos fijos y aleatorios del MMS elíptico.

II.4.1. Estimación de los efectos fijos

Para encontrar la estimación del vector de parámetros de la componente paramétrica debemos considerar (3.9) y resolver (3.23) asumiendo α , $\boldsymbol{\Sigma}_i$ y \mathbf{f} son fijos, se tiene:

$$\begin{aligned} \mathbf{X}^t \mathbf{W} \boldsymbol{\varepsilon} &= \mathbf{0} \\ \mathbf{X}^t \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}\mathbf{f}) &= \mathbf{0} \\ \mathbf{X}^t \mathbf{W} \mathbf{X} \boldsymbol{\beta} &= \mathbf{X}^t \mathbf{W} (\mathbf{y} - \mathbf{N}\mathbf{f}) \\ \boldsymbol{\beta} &= (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} (\mathbf{y} - \mathbf{N}\mathbf{f}) \end{aligned}$$

Además, sabemos que:

$$\begin{aligned}\mathbf{X}^t\mathbf{W}\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^t\mathbf{W}(\mathbf{y} - \mathbf{N}\mathbf{f}) \\ \mathbf{X}^t\mathbf{W}\mathbf{y} - \mathbf{X}^t\mathbf{W}\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^t\mathbf{W}\mathbf{N}\mathbf{f}\end{aligned}$$

y al incorporar (3.27) en la expresión anterior, tenemos:

$$\mathbf{X}^t\mathbf{W}\mathbf{y} - \mathbf{X}^t\mathbf{W}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^t\mathbf{W}\mathbf{N} \left[(\mathbf{N}^t\mathbf{W}\mathbf{N} + \alpha\mathbf{K})^{-1} \mathbf{N}^t\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right]$$

y después de algunas operaciones matriciales, se obtiene:

$$\begin{aligned}\mathbf{X}^t\mathbf{W}_n\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^t\mathbf{W}_n\mathbf{y} \\ \boldsymbol{\beta} &= (\mathbf{X}^t\mathbf{W}_n\mathbf{X})^{-1} \mathbf{X}^t\mathbf{W}_n\mathbf{y}\end{aligned}$$

donde $\mathbf{W}_n = \mathbf{W} - \mathbf{W}\mathbf{N}(\mathbf{N}^t\mathbf{W}\mathbf{N} + \alpha\mathbf{K})^{-1} \mathbf{N}^t\mathbf{W}$.

Para el caso de la componente no paramétrica, debemos considerar (3.11) y resolver (3.24) asumiendo que los parámetros α , $\boldsymbol{\Sigma}_i$ y $\boldsymbol{\beta}$ son fijos, se tiene:

$$\begin{aligned}\mathbf{N}^t\mathbf{W}\boldsymbol{\varepsilon} - \alpha\mathbf{K}\mathbf{f} &= \mathbf{0} \\ \mathbf{N}^t\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{N}\mathbf{f}) - \alpha\mathbf{K}\mathbf{f} &= \mathbf{0} \\ (\mathbf{N}^t\mathbf{W}\mathbf{N} + \alpha\mathbf{K})\mathbf{f} &= \mathbf{N}^t\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ \mathbf{f} &= (\mathbf{N}^t\mathbf{W}\mathbf{N} + \alpha\mathbf{K})^{-1} \mathbf{N}^t\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\end{aligned}$$

Además, sabemos que:

$$\begin{aligned}(\mathbf{N}^t\mathbf{W}\mathbf{N} + \alpha\mathbf{K})\mathbf{f} &= \mathbf{N}^t\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ (\mathbf{N}^t\mathbf{W}\mathbf{N} + \alpha\mathbf{K})\mathbf{f} &= \mathbf{N}^t\mathbf{W}\mathbf{y} - \mathbf{N}^t\mathbf{W}\mathbf{X}\boldsymbol{\beta}\end{aligned}$$

y al incorporar (3.26) en la expresión anterior, tenemos:

$$(\mathbf{N}^t\mathbf{W}\mathbf{N} + \alpha\mathbf{K})\mathbf{f} = \mathbf{N}^t\mathbf{W}\mathbf{y} - \mathbf{N}^t\mathbf{W}\mathbf{X} \left[(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1} \mathbf{X}^t\mathbf{W}(\mathbf{y} - \mathbf{N}\mathbf{f}) \right]$$

y después de algunas operaciones matriciales, se obtiene:

$$\begin{aligned}\mathbf{N}^t\mathbf{W}_x\mathbf{y} - \mathbf{N}^t\mathbf{W}_x\mathbf{N}\mathbf{f} &= \alpha\mathbf{K}\mathbf{f} \\ (\mathbf{N}^t\mathbf{W}_x\mathbf{N} + \alpha\mathbf{K})\mathbf{f} &= \mathbf{N}^t\mathbf{W}_x\mathbf{y} \\ \mathbf{f} &= (\mathbf{N}^t\mathbf{W}_x\mathbf{N} + \alpha\mathbf{K})^{-1} \mathbf{N}^t\mathbf{W}_x\mathbf{y}\end{aligned}$$

donde $\mathbf{W}_x = \mathbf{W} - \mathbf{W}\mathbf{X}(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1} \mathbf{X}^t\mathbf{W}$.

II.4.2. Estimación de los efectos aleatorios

Como el estimador empírico de Bayes de los efectos aleatorios asociados a la i -ésima unidad de estudio esta dado por:

$$\hat{\mathbf{b}}_i = \widehat{\mathbb{E}}[\mathbf{b}_i | \mathbf{y}_i] = \mathbf{D}\mathbf{Z}_i^t \boldsymbol{\Sigma}_i^{-1} [\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i]$$

donde $\hat{\boldsymbol{\mu}}_i = \mathbf{X}_i \hat{\boldsymbol{\beta}} - \mathbf{N}_i \hat{\mathbf{f}}$ entonces el vector de estimaciones del MMS elíptico asociados a la i -ésima unidad de estudio, denotado por $\hat{\mathbf{y}}_i$, puede ser escrito de la siguiente forma:

$$\begin{aligned} \hat{\mathbf{y}}_i &= \mathbf{X}_i \hat{\boldsymbol{\beta}} + \mathbf{N}_i \hat{\mathbf{f}} + \mathbf{Z}_i \hat{\mathbf{b}}_i \\ &= \hat{\boldsymbol{\mu}}_i + \mathbf{Z}_i \widehat{\mathbf{D}} \mathbf{Z}_i^t \widehat{\boldsymbol{\Sigma}}_i^{-1} [\mathbf{y}_i - \hat{\boldsymbol{\mu}}] \\ &= \left[\mathbb{I}_{n_i} - \mathbf{Z}_i \widehat{\mathbf{D}} \mathbf{Z}_i^t \widehat{\boldsymbol{\Sigma}}_i^{-1} \right] \hat{\boldsymbol{\mu}}_i + \mathbf{Z}_i \widehat{\mathbf{D}} \mathbf{Z}_i^t \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{y}_i \\ &= \left[\widehat{\boldsymbol{\Sigma}}_i - \mathbf{Z}_i \widehat{\mathbf{D}} \mathbf{Z}_i^t \right] \widehat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\mu}}_i + \left[\widehat{\boldsymbol{\Sigma}}_i - \widehat{\mathbf{Q}}_i \right] \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{y}_i \end{aligned}$$

puesto que $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^t + \mathbf{Q}_i$, entonces después de algunas operaciones matriciales se obtiene:

$$\hat{\mathbf{y}}_i = \widehat{\mathbf{Q}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1} \hat{\boldsymbol{\mu}}_i + \left[\mathbb{I}_{n_i} - \widehat{\mathbf{Q}}_i \widehat{\boldsymbol{\Sigma}}_i^{-1} \right] \mathbf{y}_i$$

Y al considerar el vector $\mathbf{b} \in \mathbb{R}^{nq}$ con $\mathbf{b} = [\mathbf{b}_1^t, \dots, \mathbf{b}_n^t]^t$, y según (7.2) $\hat{\mathbf{b}}$ puede ser escrito de la siguiente manera:

$$\begin{aligned} \hat{\mathbf{b}} &= \begin{bmatrix} \hat{\mathbf{b}}_1 \\ \hat{\mathbf{b}}_2 \\ \vdots \\ \hat{\mathbf{b}}_n \end{bmatrix} = \begin{bmatrix} \widehat{\mathbb{E}}[\mathbf{b}_1 | \mathbf{y}_1] \\ \widehat{\mathbb{E}}[\mathbf{b}_2 | \mathbf{y}_2] \\ \vdots \\ \widehat{\mathbb{E}}[\mathbf{b}_n | \mathbf{y}_n] \end{bmatrix} = \begin{bmatrix} \mathbf{D}\mathbf{Z}_1^t \boldsymbol{\Sigma}_1^{-1} [\mathbf{y}_1 - \hat{\boldsymbol{\mu}}_1] \\ \mathbf{D}\mathbf{Z}_2^t \boldsymbol{\Sigma}_2^{-1} [\mathbf{y}_2 - \hat{\boldsymbol{\mu}}_2] \\ \vdots \\ \mathbf{D}\mathbf{Z}_n^t \boldsymbol{\Sigma}_n^{-1} [\mathbf{y}_n - \hat{\boldsymbol{\mu}}_n] \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{D}\mathbf{Z}_1^t \boldsymbol{\Sigma}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}\mathbf{Z}_2^t \boldsymbol{\Sigma}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}\mathbf{Z}_n^t \boldsymbol{\Sigma}_n^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 - \hat{\boldsymbol{\mu}}_1 \\ \mathbf{y}_2 - \hat{\boldsymbol{\mu}}_2 \\ \vdots \\ \mathbf{y}_n - \hat{\boldsymbol{\mu}}_n \end{bmatrix} \end{aligned}$$

$$\begin{aligned}
\hat{\mathbf{b}} &= \begin{bmatrix} \mathbf{D}\mathbf{Z}_1^t & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}\mathbf{Z}_2^t & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}\mathbf{Z}_n^t \end{bmatrix} \begin{bmatrix} \Sigma_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Sigma_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Sigma_n^{-1} \end{bmatrix} [\mathbf{y} - \hat{\boldsymbol{\mu}}] \\
&= \begin{bmatrix} \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1^t & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2^t & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_n^t \end{bmatrix} \Sigma^{-1} [\mathbf{y} - \hat{\boldsymbol{\mu}}] \\
&= (\mathbb{I}_n \otimes \mathbf{D}) \mathbf{Z}^t \Sigma^{-1} [\mathbf{y} - \hat{\boldsymbol{\mu}}]
\end{aligned}$$

donde $\mathbf{Z} \in \mathbb{R}^{m \times nq}$, $\mathbf{D} \in \mathbb{R}^{q \times q}$ y $\Sigma \in \mathbb{R}^{m \times m}$.

APÉNDICE III: DEMOSTRACIONES

CAPITULO 4

III.1. Análisis de residuos

En esta sección se describen las medias marginal y condicional del MMS elíptico, el cual son la base para el cálculo de los residuos marginal y condicional respectivamente.

III.1.1. Media marginal del modelo

La media marginal del MMS elíptico esta dado por:

$$\hat{\mathbf{y}} = \widehat{\mathbb{E}}[\mathbf{y}] = \widehat{\boldsymbol{\mu}} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{N}\hat{\mathbf{f}}$$

he incorporando (3.28) y (3.29) obtenemos:

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{X} \left[(\mathbf{X}^t \mathbf{W}_n \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}_n \mathbf{y} \right] + \mathbf{N} \left[(\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^t \mathbf{W}_x \mathbf{y} \right] \\ &= [\mathbf{X} \quad \mathbf{N}] \begin{bmatrix} (\mathbf{X}^t \mathbf{W}_n \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}_n \\ (\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^t \mathbf{W}_x \end{bmatrix} \mathbf{y} \\ &= [\mathbf{X} \quad \mathbf{N}] \begin{bmatrix} (\mathbf{X}^t \mathbf{W}_n \mathbf{X})^{-1} \mathbf{X}^t [\mathbf{W} - \mathbf{W} \mathbf{N} (\mathbf{N}^t \mathbf{W} \mathbf{N} + \alpha \mathbf{K} \mathbf{f})^{-1} \mathbf{N}^t \mathbf{W}] \\ (\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^t [\mathbf{W} - \mathbf{W} \mathbf{X} (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}] \end{bmatrix} \mathbf{y} \\ &= [\mathbf{X} \quad \mathbf{N}] \begin{bmatrix} (\mathbf{X}^t \mathbf{W}_n \mathbf{X})^{-1} \mathbf{X}^t [\mathbb{I}_m - \mathbf{W} \mathbf{N} (\mathbf{N}^t \mathbf{W} \mathbf{N} + \alpha \mathbf{K} \mathbf{f})^{-1} \mathbf{N}^t] \\ (\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^t [\mathbb{I}_m - \mathbf{W} \mathbf{X} (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t] \end{bmatrix} \mathbf{W} \mathbf{y} \\ &= [\mathbf{X} \quad \mathbf{N}] \tilde{\mathbf{A}} \mathbf{W} \mathbf{y} \end{aligned}$$

Sin embargo, la matriz $\tilde{\mathbf{A}}$ puede ser reescrita como sigue:

$$\begin{aligned}\tilde{\mathbf{A}} &= \begin{bmatrix} (\mathbf{X}^t \mathbf{W}_n \mathbf{X})^{-1} \mathbf{X}^t - (\mathbf{X}^t \mathbf{W}_n \mathbf{X})^{-1} \mathbf{W} \mathbf{N} (\mathbf{N}^t \mathbf{W} \mathbf{N} + \alpha \mathbf{K} \mathbf{f})^{-1} \mathbf{N}^t \\ (\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{N}^t - (\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{W} \mathbf{X} (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{X}^t \mathbf{W}_n \mathbf{X})^{-1} & -\tilde{\mathbf{A}}_1 \\ -\tilde{\mathbf{A}}_2 & (\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{X}^t \\ \mathbf{N}^t \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{X}^t \mathbf{W} \mathbf{X} & \mathbf{X}^t \mathbf{W} \mathbf{N} \\ \mathbf{N}^t \mathbf{W} \mathbf{X} & \mathbf{N}^t \mathbf{W} \mathbf{N} + \alpha \mathbf{K} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^t \\ \mathbf{N}^t \end{bmatrix}\end{aligned}$$

donde:

$$\begin{aligned}\tilde{\mathbf{A}}_1 &= (\mathbf{X}^t \mathbf{W}_n \mathbf{X})^{-1} \mathbf{W} \mathbf{N} (\mathbf{N}^t \mathbf{W} \mathbf{N} + \alpha \mathbf{K} \mathbf{f})^{-1} \\ \tilde{\mathbf{A}}_2 &= (\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K})^{-1} \mathbf{W} \mathbf{X} (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1}\end{aligned}$$

Y después de algunas operaciones algebraicas matriciales que son utilizadas de manera análoga para el cálculo de la matriz de información de Fisher penalizada vista en la expresión (3.40), se puede escribir que el vector estimado para las n observaciones es:

$$\begin{aligned}\hat{\mathbf{y}} &= [\mathbf{X} \quad \mathbf{N}] \begin{bmatrix} \mathbf{X}^t \mathbf{W} \mathbf{X} & \mathbf{X}^t \mathbf{W} \mathbf{N} \\ \mathbf{N}^t \mathbf{W} \mathbf{X} & \mathbf{N}^t \mathbf{W} \mathbf{N} + \alpha \mathbf{K} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^t \\ \mathbf{N}^t \end{bmatrix} \mathbf{W} \mathbf{y} \\ &= \tilde{\mathbf{H}} \mathbf{y}\end{aligned}$$

III.1.2. Media condicional del modelo

La media condicional del MMS elíptico esta dado por:

$$\hat{\mathbf{y}} = \hat{\mathbb{E}}[\mathbf{y}|\mathbf{b}] = \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{N} \hat{\mathbf{f}} + \mathbf{Z} \hat{\mathbf{b}} = \tilde{\mathbf{H}} \mathbf{y} + \mathbf{Z} \hat{\mathbf{b}}$$

he incorporando (2.11), (2.12) y (3.38), obtenemos:

$$\begin{aligned}\hat{\mathbf{y}} &= \tilde{\mathbf{H}} \mathbf{y} + \mathbf{Z} (\mathbb{I}_n \otimes \mathbf{D}) \mathbf{Z}^t \boldsymbol{\Sigma}^{-1} [\mathbf{y} - \hat{\boldsymbol{\mu}}] \\ &= \tilde{\mathbf{H}} \mathbf{y} + [\boldsymbol{\Sigma} - \mathbf{Q}] \boldsymbol{\Sigma}^{-1} [\mathbf{y} - \tilde{\mathbf{H}} \mathbf{y}] \\ &= \tilde{\mathbf{H}} \mathbf{y} + [\mathbb{I}_m - \mathbf{Q} \boldsymbol{\Sigma}^{-1}] [\mathbb{I}_m - \tilde{\mathbf{H}}] \mathbf{y} \\ &= \left[\tilde{\mathbf{H}} + \mathbb{I}_m - \tilde{\mathbf{H}} - \mathbf{Q} \boldsymbol{\Sigma}^{-1} + \mathbf{Q} \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{H}} \right] \mathbf{y} \\ &= \left[\mathbb{I}_m - \mathbf{Q} \boldsymbol{\Sigma}^{-1} + \mathbf{Q} \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{H}} \right] \mathbf{y} = \mathbf{H} \mathbf{y}\end{aligned}$$

III.2. Diagnóstico de influencia global para los efectos fijos

III.2.1. Diagnóstico basado en supresión de una observación

Para la construcción de la expresión (4.11), definimos lo siguiente:

$$\begin{aligned} y_c &= y_{ij}^* = \mathbf{x}_{ij}^t \widehat{\boldsymbol{\beta}}_{(ij)} + \hat{f}_{(ij)}(t_{ij}) - w_{cc}^{-1} \mathbf{w}_{(c)}^t [\mathbf{y}_{(c)} - \mathbf{X}_{(c)} \widehat{\boldsymbol{\beta}}_{(ij)} - \mathbf{N}_{(c)} \hat{\mathbf{f}}_{(ij)}] \\ &= \mathbf{x}_c^t \widehat{\boldsymbol{\beta}}_{(ij)} + \mathbf{n}_c^t \hat{\mathbf{f}}_{(ij)} - w_{cc}^{-1} \mathbf{w}_{(c)}^t [\mathbf{y}_{(c)} - \mathbf{X}_{(c)} \widehat{\boldsymbol{\beta}}_{(ij)} - \mathbf{N}_{(c)} \hat{\mathbf{f}}_{(ij)}] \end{aligned} \quad (7.2)$$

$$\mathbf{y}_{(c)} = \mathbf{y}_i^* = [y_{i1} \quad \cdots \quad y_{ij-1} \quad y_{ij}^* \quad y_{ij+1} \quad \cdots \quad y_{in_i}]^t \quad (7.3)$$

$$\mathbf{y}^* = [\mathbf{y}_1^t \quad \cdots \quad \mathbf{y}_{i-1}^t \quad \mathbf{y}_i^{*t} \quad \mathbf{y}_{i+1}^t \quad \cdots \quad \mathbf{y}_m^t]^t \quad (7.4)$$

donde $\mathbf{y} - \mathbf{y}^* = \mathbf{d}_c^t (y_{ij} - y_c)$ y asumiendo que las matrices $\boldsymbol{\Sigma}$ y \mathbf{W} pueden ser particionadas como:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{cc} & \boldsymbol{\sigma}_{(c)}^t \\ \boldsymbol{\sigma}_{(c)} & \boldsymbol{\Sigma}_{(c)} \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} w_{cc} & \mathbf{w}_{(c)}^t \\ \mathbf{w}_{(c)} & \mathbf{W}_{(c)} \end{bmatrix} \quad (7.5)$$

donde $\sigma_{cc}, w_{cc} \in \mathbb{R}$, $\boldsymbol{\sigma}_{(c)}, \mathbf{w}_{(c)} \in \mathbb{R}^{(m-1)}$ y $\boldsymbol{\Sigma}_{(c)}, \mathbf{W}_{(c)} \in \mathbb{R}^{(m-1) \times (m-1)}$ y considerando por $\sigma_* = \sigma_{cc} - \boldsymbol{\sigma}_{(c)}^t \boldsymbol{\Sigma}_{(c)}^{-1} \boldsymbol{\sigma}_{(c)}$ bajo el supuesto de $\mathbf{y} \sim \text{EC}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ donde la estructura de $\boldsymbol{\Sigma}$ fue dada en (2.13), es posible describir que para cualquier parámetro $\boldsymbol{\beta}$ y \mathbf{f} , que por definición de $\widehat{\boldsymbol{\beta}}_{(ij)}$ y $\hat{\mathbf{f}}_{(ij)}$, la función de log-verosimilitud penalizada evaluada en (7.4) está dada por:

$$\begin{aligned} l_p(\boldsymbol{\theta}; \mathbf{y}^*) &= l_p(\boldsymbol{\theta}; y_1, \dots, y_c, \dots, y_m) = \ln [g(u^*)] - \frac{\alpha}{2} \mathbf{f}^t \mathbf{K} \mathbf{f} \\ &= \ln \left\{ g \left([\mathbf{y}^* - \boldsymbol{\mu}]^t \boldsymbol{\Sigma}^{-1} [\mathbf{y}^* - \boldsymbol{\mu}] \right) \right\} - \frac{\alpha}{2} \mathbf{f}^t \mathbf{K} \mathbf{f} \end{aligned}$$

No obstante, note que:

$$\begin{aligned} g(u^*) &= g \left\{ \left([\mathbf{y}_{(c)} - \mathbf{X}_{(c)} \boldsymbol{\beta} - \mathbf{N}_{(c)} \mathbf{f}]^t \boldsymbol{\Sigma}_{(c)}^{-1} [\mathbf{y}_{(c)} - \mathbf{X}_{(c)} \boldsymbol{\beta} - \mathbf{N}_{(c)} \mathbf{f}] \right) \right. \\ &\quad \left. + \left(y_c - \mathbf{x}_c^t \boldsymbol{\beta} - \mathbf{n}_c^t \mathbf{f} - [\mathbf{y}_{(c)} - \mathbf{X}_{(c)} \boldsymbol{\beta} - \mathbf{N}_{(c)} \mathbf{f}]^t \boldsymbol{\Sigma}_{(c)}^{-1} \boldsymbol{\sigma}_{(c)} \right)^2 \sigma_*^{-1} \right\} \\ &\geq g \left([\mathbf{y}_{(c)} - \mathbf{X}_{(c)} \boldsymbol{\beta} - \mathbf{N}_{(c)} \mathbf{f}]^t \boldsymbol{\Sigma}_{(c)}^{-1} [\mathbf{y}_{(c)} - \mathbf{X}_{(c)} \boldsymbol{\beta} - \mathbf{N}_{(c)} \mathbf{f}] \right) \\ &\geq g \left([\mathbf{y}^* - \mathbf{X} \widehat{\boldsymbol{\beta}}_{(ij)} - \mathbf{N} \hat{\mathbf{f}}_{(ij)}]^t \boldsymbol{\Sigma}^{-1} [\mathbf{y}_{(c)} - \mathbf{X} \widehat{\boldsymbol{\beta}}_{(ij)} - \mathbf{N}_{(c)} \hat{\mathbf{f}}_{(ij)}] \right) \\ &\geq g(u_{(ij)}^*) \end{aligned}$$

y así, obtenemos:

$$l_p(\boldsymbol{\theta}; \mathbf{y}^*) \geq l_p(\widehat{\boldsymbol{\theta}}_{(ij)}; \mathbf{y}^*) = \ln \{g(u_{(ij)}^*)\} - \frac{\alpha}{2} \widehat{\mathbf{f}}_{(ij)}^t \mathbf{K} \widehat{\mathbf{f}}_{(ij)} \quad (7.6)$$

el cual $\widehat{\boldsymbol{\beta}}_{(ij)}$ y $\widehat{\mathbf{f}}_{(ij)}$ minimizan $l_p(\widehat{\boldsymbol{\theta}}_{(ij)}; \mathbf{y}^*)$. De esta manera, los EMVP de $\widehat{\boldsymbol{\theta}}_{(ij)}$ están dados por:

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{(ij)} &= [\mathbf{X}^t \mathbf{W}_n \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W}_n \mathbf{y}^* \\ &= \widehat{\boldsymbol{\beta}} - [\mathbf{X}^t \mathbf{W}_n \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W}_n \mathbf{d}_c (y_{ij} - y_c) \end{aligned} \quad (7.7)$$

$$\begin{aligned} \widehat{\mathbf{f}}_{(ij)} &= [\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K}]^{-1} \mathbf{N}^t \mathbf{W}_x \mathbf{y}^* \\ &= \widehat{\mathbf{f}} - [\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K}]^{-1} \mathbf{N}^t \mathbf{W}_x \mathbf{d}_c (y_{ij} - y_c) \end{aligned} \quad (7.8)$$

Sin embargo, según con la expresión (7.2) podemos escribir $y_{ij} - y_c$ como sigue:

$$\begin{aligned} y_{ij} - y_c &= y_{ij} - \left(\mathbf{x}_c^t \widehat{\boldsymbol{\beta}}_{(ij)} + \mathbf{n}_c^t \widehat{\mathbf{f}}_{(ij)} - w_{cc}^{-1} \mathbf{w}_{(c)}^t [\mathbf{y}_{(c)} - \mathbf{X}_{(c)} \widehat{\boldsymbol{\beta}}_{(ij)} - \mathbf{N}_{(c)} \widehat{\mathbf{f}}_{(ij)}] \right) \\ &= w_{cc}^{-1} \left(w_{cc} y_{ij} - w_{cc} \mathbf{x}_c^t \widehat{\boldsymbol{\beta}}_{(ij)} - w_{cc} \mathbf{n}_c^t \widehat{\mathbf{f}}_{(ij)} + \mathbf{w}_{(c)}^t [\mathbf{y}_{(c)} - \mathbf{X}_{(c)} \widehat{\boldsymbol{\beta}}_{(ij)} \right. \\ &\quad \left. - \mathbf{N}_{(c)} \widehat{\mathbf{f}}_{(ij)}] \right) = w_{cc}^{-1} \left(w_{cc} y_c + \mathbf{w}_{(c)}^t \mathbf{y}_{(c)} - w_{cc} \mathbf{x}_c^t \widehat{\boldsymbol{\beta}}_{(ij)} - \mathbf{w}_{(c)}^t \mathbf{X}_{(c)} \widehat{\boldsymbol{\beta}}_{(ij)} \right. \\ &\quad \left. - w_{cc} \mathbf{n}_c^t \widehat{\mathbf{f}}_{(ij)} - \mathbf{w}_{(c)}^t \mathbf{N}_{(c)} \widehat{\mathbf{f}}_{(ij)} \right) \\ &= w_{cc}^{-1} [w_{cc} \quad \mathbf{W}_{(c)}^t] \left[\mathbf{y} - \mathbf{X} \widehat{\boldsymbol{\beta}}_{(ij)} - \mathbf{N} \widehat{\mathbf{f}}_{(ij)} \right] \\ &= w_{cc}^{-1} \mathbf{d}_c^t \mathbf{W} [\mathbf{y} - \widetilde{\mathbf{H}} \mathbf{y}^*] = w_{cc}^{-1} \mathbf{d}_c^t \mathbf{W} [\mathbf{y} - \widetilde{\mathbf{H}} \mathbf{y} + \widetilde{\mathbf{H}} (\mathbf{y} - \mathbf{y}^*)] \\ &= w_{cc}^{-1} \mathbf{d}_c^t \mathbf{W} [\mathbf{y} - \widetilde{\mathbf{H}} \mathbf{y}] + w_{cc}^{-1} \mathbf{d}_c^t \mathbf{W} [\widetilde{\mathbf{H}} (\mathbf{y} - \mathbf{y}^*)] \\ &= w_{cc}^{-1} \mathbf{d}_c^t \mathbf{W} [\mathbf{y} - \widetilde{\mathbf{H}} \mathbf{y}] + w_{cc}^{-1} \mathbf{d}_c^t \mathbf{W} \widetilde{\mathbf{H}} \mathbf{d}_c^t (y_{ij} - y_c) \end{aligned}$$

donde $\widetilde{\mathbf{H}}$ es la matriz de apalancamiento descrita en (4.2) y despues de algunas operaciones algebraicas, se obtiene:

$$\begin{aligned} y_{ij} - y_c &= \frac{w_{cc}^{-1} \mathbf{d}_c^t \mathbf{W} (\mathbf{y} - \widetilde{\mathbf{H}} \mathbf{y})}{1 - w_{cc}^{-1} \mathbf{d}_c^t \mathbf{W} \widetilde{\mathbf{H}} \mathbf{d}_c} = \frac{w_{cc}^{-1} \mathbf{d}_c^t \mathbf{W} (\mathbf{y} - \widetilde{\mathbf{H}} \mathbf{y})}{w_{cc}^{-1} \mathbf{d}_c^t \mathbf{W} (\mathbb{I}_m - \widetilde{\mathbf{H}}) \mathbf{d}_c} \\ &= \frac{\mathbf{d}_c^t \mathbf{W} \widehat{\mathbf{e}}_m}{\mathbf{d}_c^t \mathbf{W} (\mathbb{I}_m - \widetilde{\mathbf{H}}) \mathbf{d}_c} \end{aligned} \quad (7.9)$$

III.2.2. Influencia de la componente paramétrica

De la componente paramétrica del MMS elíptico, ver expresión (4.12), es definida como:

$$\begin{aligned} \text{DCG}_{(ij)}(\boldsymbol{\beta}) &= [\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(ij)}]^t [\text{Cov}(\hat{\boldsymbol{\beta}})]^{-1} [\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(ij)}] \\ &= [\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(ij)}]^t \left[\begin{bmatrix} \mathbb{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathcal{I}^{\beta f} \begin{bmatrix} \mathbb{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}^t \right]^{-1} [\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(ij)}] \end{aligned} \quad (7.10)$$

y considerando las expresiones (7.7) y (4.11) tenemos que:

$$\begin{aligned} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(ij)} &= [\mathbf{X}^t \mathbf{W}_n \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W}_n \mathbf{d}_c (y_{ij} - y_c) \\ &= \frac{[\mathbf{X}^t \mathbf{W}_n \mathbf{X}]^{-1} \mathbf{X}^t \mathbf{W}_n \mathbf{d}_c \mathbf{d}_c^t \mathbf{W} \hat{\mathbf{e}}_m}{\mathbf{d}_c^t \mathbf{W} (\mathbb{I}_m - \tilde{\mathbf{H}}) \mathbf{d}_c} \end{aligned}$$

Además, al desarrollar $[\text{Cov}(\hat{\boldsymbol{\beta}})]^{-1}$ obtenemos:

$$\begin{aligned} \left[\begin{bmatrix} \mathbb{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathcal{I}^{\beta f} \begin{bmatrix} \mathbb{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}^t \right]^{-1} &= \begin{bmatrix} \mathbb{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} (\mathbf{X}^t \boldsymbol{\Pi}_n \mathbf{X})^{-1} & -\mathbf{L} \\ -\mathbf{L}^t & (\mathbf{N}^t \boldsymbol{\Pi}_x \mathbf{N} + \alpha \mathbf{K})^{-1} \end{bmatrix} \begin{bmatrix} \mathbb{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}^t \right]^{-1} \\ &= \left[[(\mathbf{X}^t \boldsymbol{\Pi}_n \mathbf{X})^{-1} \quad -\mathbf{L}] \begin{bmatrix} \mathbb{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}^t \right]^{-1} \\ &= [(\mathbf{X}^t \boldsymbol{\Pi}_n \mathbf{X})^{-1}]^{-1} = (\mathbf{X}^t \boldsymbol{\Pi}_n \mathbf{X}) \end{aligned}$$

Y reemplazando posteriormente en (7.10), se demuestra la DCG descrita en el Teorema 1.

III.2.3. Influencia de la componente no paramétrica

Puesto que el DFITS de la c -ésima observación removida corresponde a:

$$\text{DFITS}_{(ij)}(\mathbf{f}) = \frac{\left| \mathbf{d}_c^t \mathbf{N} (\hat{\mathbf{f}} - \hat{\mathbf{f}}_{(ij)}) \right|}{s_{(ij)}} \quad (7.11)$$

donde $s_{(ij)}^2$ corresponde al c -ésimo elemento de la diagonal de \mathbf{S}^2 , y según con las expresiones (4.11) y (7.8) se tiene que:

$$\begin{aligned} \hat{\mathbf{f}} - \hat{\mathbf{f}}_{(ij)} &= [\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K}]^{-1} \mathbf{N}^t \mathbf{W}_x \mathbf{d}_c (y_{ij} - y_c) \\ &= \frac{[\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K}]^{-1} \mathbf{N}^t \mathbf{W}_x \mathbf{d}_c \mathbf{d}_c^t \mathbf{W} \hat{\mathbf{e}}_m}{\mathbf{d}_c^t \mathbf{W} (\mathbb{I}_m - \tilde{\mathbf{H}}) \mathbf{d}_c} \end{aligned}$$

Por lo tanto, evaluando en (7.11) obtenemos:

$$\text{DFITS}_{(ij)}(\mathbf{f}) = \frac{1}{s_{(ij)}} \left| \frac{\mathbf{d}_c^t \mathbf{N} [\mathbf{N}^t \mathbf{W}_x \mathbf{N} + \alpha \mathbf{K}]^{-1} \mathbf{N}^t \mathbf{W}_x \mathbf{d}_c \mathbf{d}_c^t \mathbf{W} \hat{\mathbf{e}}_m}{\mathbf{d}_c^t \mathbf{W} (\mathbb{I}_m - \tilde{\mathbf{H}}) \mathbf{d}_c} \right|$$

APÉNDICE IV: DEMOSTRACIONES

CAPITULO 5

En adelante y según con las expresiones descritas en (3.17) y (3.40), para el MMS elíptico utilizamos la siguiente notación para describir las particiones de la matrices Hessiana, información de Fisher penalizada y sus inversas como:

$$\mathbb{H}(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{H}_{\beta\beta} & \mathcal{H}_{\beta f} & \mathcal{H}_{\beta\tau} \\ \mathcal{H}_{f\beta} & \mathcal{H}_{ff} & \mathcal{H}_{f\tau} \\ \mathcal{H}_{\tau\beta} & \mathcal{H}_{\tau f} & \mathcal{H}_{\tau\tau} \end{bmatrix} \quad \mathbb{H}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{H}^{\beta\beta} & \mathcal{H}^{\beta f} & \mathcal{H}^{\beta\tau} \\ \mathcal{H}^{f\beta} & \mathcal{H}^{ff} & \mathcal{H}^{f\tau} \\ \mathcal{H}^{\tau\beta} & \mathcal{H}^{\tau f} & \mathcal{H}^{\tau\tau} \end{bmatrix}$$

$$\mathcal{I}_p(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{I}_{[\beta, f]} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{\tau\tau} \end{bmatrix} = \begin{bmatrix} \mathcal{I}_{\beta\beta} & \mathcal{I}_{\beta f} & \mathbf{0} \\ \mathcal{I}_{f\beta} & \mathcal{I}_{ff} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}_{\tau\tau} \end{bmatrix}$$

$$\mathcal{I}_p^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{I}^{[\beta, f]} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}^{\tau\tau} \end{bmatrix} = \begin{bmatrix} \mathcal{I}^{\beta\beta} & \mathcal{I}^{\beta f} & \mathbf{0} \\ \mathcal{I}^{f\beta} & \mathcal{I}^{ff} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}^{\tau\tau} \end{bmatrix}$$

Y las matrices particionadas asociadas a los componentes de varianza, pueden ser escritas de la siguiente manera:

$$\mathcal{H}_{\tau\tau} = \begin{bmatrix} \mathcal{H}_{\rho\rho} & \mathcal{H}_{\rho\phi} & \mathcal{H}_{\rho\delta} \\ \mathcal{H}_{\phi\rho} & \mathcal{H}_{\phi\phi} & \mathcal{H}_{\phi\delta} \\ \mathcal{H}_{\delta\rho} & \mathcal{H}_{\delta\phi} & \mathcal{H}_{\delta\delta} \end{bmatrix} \quad \mathcal{H}^{\tau\tau} = \begin{bmatrix} \mathcal{H}^{\rho\rho} & \mathcal{H}^{\rho\phi} & \mathcal{H}^{\rho\delta} \\ \mathcal{H}^{\phi\rho} & \mathcal{H}^{\phi\phi} & \mathcal{H}^{\phi\delta} \\ \mathcal{H}^{\delta\rho} & \mathcal{H}^{\delta\phi} & \mathcal{H}^{\delta\delta} \end{bmatrix}$$

$$\mathcal{I}_{\tau\tau} = \begin{bmatrix} \mathcal{I}_{\rho\rho} & \mathcal{I}_{\rho\phi} & \mathcal{I}_{\rho\delta} \\ \mathcal{I}_{\phi\rho} & \mathcal{I}_{\phi\phi} & \mathcal{I}_{\phi\delta} \\ \mathcal{I}_{\delta\rho} & \mathcal{I}_{\delta\phi} & \mathcal{I}_{\delta\delta} \end{bmatrix} \quad \mathcal{I}^{\tau\tau} = \begin{bmatrix} \mathcal{I}^{\rho\rho} & \mathcal{I}^{\rho\phi} & \mathcal{I}^{\rho\delta} \\ \mathcal{I}^{\phi\rho} & \mathcal{I}^{\phi\phi} & \mathcal{I}^{\phi\delta} \\ \mathcal{I}^{\delta\rho} & \mathcal{I}^{\delta\phi} & \mathcal{I}^{\delta\delta} \end{bmatrix}$$

Sin pérdida de generalidad, el test de score depende del parámetro $\boldsymbol{\theta}$ con $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^t \ \boldsymbol{\theta}_2^t]^t$, que evalúa la hipótesis $H_0 : \boldsymbol{\theta}_1 = \mathbf{0}$ contra $H_1 : \boldsymbol{\theta}_1 \neq \mathbf{0}$, o equivalentemente puede ser escrito por $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ contra $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, donde $\boldsymbol{\theta}_0 = [\mathbf{0}^t \ \boldsymbol{\theta}_2^t]^t$, bajo la hipótesis nula las funciones de score de $\boldsymbol{\theta}_1 \in \mathbb{R}^{p_1}$ y $\boldsymbol{\theta}_2 \in \mathbb{R}^{p_2}$, son $\mathbb{S}(\boldsymbol{\theta}_1)$ y $\mathbb{S}(\boldsymbol{\theta}_2)$. Suponiendo que $\boldsymbol{\theta}_2$ es desconocido, entonces en el estadístico de prueba del test score se incorpora el EMV de $\boldsymbol{\theta}_2$ sujeto a la hipótesis nula, digamos $\hat{\boldsymbol{\theta}}_{02}$, por tanto, la prueba está dada por:

$$SC(\boldsymbol{\theta}_1) = [\mathbb{S}(\boldsymbol{\theta}_1)]^t \mathcal{I}^{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2} [\mathbb{S}(\boldsymbol{\theta}_1)] \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} \quad (7.12)$$

donde $\hat{\boldsymbol{\theta}}_0 = [\mathbf{0}^t \ \hat{\boldsymbol{\theta}}_{02}^t]^t$, $\mathcal{I}^{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2} = [\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} - \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}\mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}^{-1}\mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_1}]^{-1}$ y $SC(\boldsymbol{\theta}_1) \sim \chi_{p_1}^2$.

Note que si $\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2} = \mathbf{0}$ entonces $\boldsymbol{\theta}_1$ y $\boldsymbol{\theta}_2$ se dicen ser ortogonales y $\mathcal{I}_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_2} = \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}$, lo que implica que la matriz de información de Fisher $\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}$ no debe ser ajustada para la estimación de $\boldsymbol{\theta}_2$ bajo la hipótesis nula.

IV.1. Distribución asintótica SC_a

En el contexto asintótico del test de score, para encontrar la distribución de SC_a realizaremos los siguientes supuestos adicionales sobre la matriz de información de Fisher asociada al modelo, para algún punto arbitrario $\boldsymbol{\theta}$ dentro de una vecindad de $\boldsymbol{\theta}_0$, supondremos que:

- 1) La matriz $\mathcal{I}_{\boldsymbol{\theta}_*\boldsymbol{\theta}_*}(\boldsymbol{\theta})$ es invertible, para algún punto arbitrario de n .
- 2) $n\mathcal{I}^{\rho|\boldsymbol{\theta}_*}(\boldsymbol{\theta}) \rightarrow \mathcal{J}^{\rho|\boldsymbol{\theta}_*}(\boldsymbol{\theta}) > 0$, cuando $n \rightarrow \infty$ uniformemente.

y así se obtiene que:

$$\frac{1}{n} [\mathcal{I}^{\rho|\boldsymbol{\theta}_*}(\boldsymbol{\theta}_0)]^{-1} \rightarrow [\mathcal{J}^{\rho|\boldsymbol{\theta}_*}(\boldsymbol{\theta}_0)]^{-1} > 0$$

Además, utilizaremos la siguiente notación:

$$\begin{aligned} \mathbb{S}_p(\boldsymbol{\theta}) &= \frac{\partial l_p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = [\mathbb{S}_p(\rho) \ [\mathbb{S}_p(\boldsymbol{\theta}_*)^t]^t] = [\mathcal{S}_\rho(\boldsymbol{\theta}) \ [\mathcal{S}_{\boldsymbol{\theta}_*}(\boldsymbol{\theta})^t]^t] \\ \mathbb{H}(\boldsymbol{\theta}) &= \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} = \begin{bmatrix} \mathcal{H}_{\rho\rho} & \mathcal{H}_{\rho\boldsymbol{\theta}_*} \\ \mathcal{H}_{\boldsymbol{\theta}_*\rho} & \mathcal{H}_{\boldsymbol{\theta}_*\boldsymbol{\theta}_*} \end{bmatrix} \\ \mathbb{H}^{-1}(\boldsymbol{\theta}) &= \begin{bmatrix} \mathcal{H}^{\rho\rho} & \mathcal{H}^{\rho\boldsymbol{\theta}_*} \\ \mathcal{H}^{\boldsymbol{\theta}_*\rho} & \mathcal{H}^{\boldsymbol{\theta}_*\boldsymbol{\theta}_*} \end{bmatrix} \end{aligned}$$

Entonces, por medio de series de Taylor, tenemos que la aproximación de primer orden del $\mathbb{S}_p(\boldsymbol{\theta})$ evaluada en $\hat{\boldsymbol{\theta}}$ alrededor de $\hat{\boldsymbol{\theta}}_0$ esta dado por:

$$\begin{aligned}\mathbb{S}_p(\hat{\boldsymbol{\theta}}) &= \mathbb{S}_p(\hat{\boldsymbol{\theta}}_0) + \frac{\partial}{\partial \boldsymbol{\theta}^t} \mathbb{S}_p(\hat{\boldsymbol{\theta}}_0) [\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0] + \mathcal{O}(1) \\ \mathbb{S}_p(\hat{\boldsymbol{\theta}}) &= \mathbb{S}_p(\hat{\boldsymbol{\theta}}_0) + \mathbb{H}(\hat{\boldsymbol{\theta}}_0) [\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0] + \mathcal{O}(1) \\ \mathbb{S}_p(\hat{\boldsymbol{\theta}}) - \mathbb{S}_p(\hat{\boldsymbol{\theta}}_0) &= \mathbb{H}(\hat{\boldsymbol{\theta}}_0) \Delta \hat{\boldsymbol{\theta}} + \mathcal{O}(1)\end{aligned}$$

donde $\Delta \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0 = [\hat{\rho} \quad \Delta \hat{\boldsymbol{\theta}}_*^t]^t$. como también sabemos que $\mathbb{S}_p(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ y $\mathcal{S}_{\theta_*}(\hat{\boldsymbol{\theta}}_0) = \mathbf{0}$, luego, al evaluarlas obtenemos:

$$\begin{aligned}\mathbf{0} - \mathbb{S}_p(\hat{\boldsymbol{\theta}}_0) &= \mathbb{H}(\hat{\boldsymbol{\theta}}_0) \Delta \hat{\boldsymbol{\theta}} + \mathcal{O}(1) \\ \begin{bmatrix} \mathcal{S}_\rho(\hat{\boldsymbol{\theta}}_0) \\ \mathcal{S}_{\theta_*}(\hat{\boldsymbol{\theta}}_0) \end{bmatrix} &= - \begin{bmatrix} \mathcal{H}_{\rho\rho}(\hat{\boldsymbol{\theta}}_0) & \mathcal{H}_{\rho\theta_*}(\hat{\boldsymbol{\theta}}_0) \\ \mathcal{H}_{\theta_*\rho}(\hat{\boldsymbol{\theta}}_0) & \mathcal{H}_{\theta_*\theta_*}(\hat{\boldsymbol{\theta}}_0) \end{bmatrix} \begin{bmatrix} \hat{\rho} \\ \Delta \hat{\boldsymbol{\theta}}_* \end{bmatrix} + \mathcal{O}(1) \\ \begin{bmatrix} \mathcal{S}_\rho(\hat{\boldsymbol{\theta}}_0) \\ \mathbf{0} \end{bmatrix} &= - \begin{bmatrix} \hat{\rho} \mathcal{H}_{\rho\rho}(\hat{\boldsymbol{\theta}}_0) + \mathcal{H}_{\rho\theta_*}(\hat{\boldsymbol{\theta}}_0) \Delta \hat{\boldsymbol{\theta}}_* \\ \hat{\rho} \mathcal{H}_{\theta_*\rho}(\hat{\boldsymbol{\theta}}_0) + \mathcal{H}_{\theta_*\theta_*}(\hat{\boldsymbol{\theta}}_0) \Delta \hat{\boldsymbol{\theta}}_* \end{bmatrix} + \mathcal{O}(1)\end{aligned}$$

pero,

$$\begin{aligned}\mathbf{0} &= \hat{\rho} \mathcal{H}_{\theta_*\rho}(\hat{\boldsymbol{\theta}}_0) + \mathcal{H}_{\theta_*\theta_*}(\hat{\boldsymbol{\theta}}_0) \Delta \hat{\boldsymbol{\theta}}_* \\ \mathcal{H}_{\theta_*\theta_*}(\hat{\boldsymbol{\theta}}_0) \Delta \hat{\boldsymbol{\theta}}_* &= -\hat{\rho} \mathcal{H}_{\theta_*\rho}(\hat{\boldsymbol{\theta}}_0) \\ \Delta \hat{\boldsymbol{\theta}}_* &= -\hat{\rho} \mathcal{H}_{\theta_*\theta_*}^{-1}(\hat{\boldsymbol{\theta}}_0) \mathcal{H}_{\theta_*\rho}(\hat{\boldsymbol{\theta}}_0)\end{aligned}$$

y así, obtenemos:

$$\begin{aligned}\mathcal{S}_\rho(\hat{\boldsymbol{\theta}}_0) &= - \left(\hat{\rho} \mathcal{H}_{\rho\rho}(\hat{\boldsymbol{\theta}}_0) + \mathcal{H}_{\rho\theta_*}(\hat{\boldsymbol{\theta}}_0) \Delta \hat{\boldsymbol{\theta}}_* \right) + \mathcal{O}(1) \\ &= - \left(\hat{\rho} \mathcal{H}_{\rho\rho}(\hat{\boldsymbol{\theta}}_0) - \hat{\rho} \mathcal{H}_{\rho\theta_*}(\hat{\boldsymbol{\theta}}_0) \mathcal{H}_{\theta_*\theta_*}^{-1}(\hat{\boldsymbol{\theta}}_0) \mathcal{H}_{\theta_*\rho}(\hat{\boldsymbol{\theta}}_0) \right) + \mathcal{O}(1) \\ &= -\hat{\rho} \left[\mathcal{H}^{\rho|\theta_*}(\hat{\boldsymbol{\theta}}_0) \right]^{-1} + \mathcal{O}(1) = -\hat{\rho} \left[\mathcal{H}^{\rho|\theta_*}(\hat{\boldsymbol{\theta}}_0) \right]^{-1} \frac{\sqrt{n}\sqrt{n}}{n} + \mathcal{O}(1) \\ \frac{1}{\sqrt{n}} \mathcal{S}_\rho(\hat{\boldsymbol{\theta}}_0) &= \sqrt{n} \hat{\rho} \left[-n \mathcal{H}^{\rho|\theta_*}(\hat{\boldsymbol{\theta}}_0) \right]^{-1} + \mathcal{O}(n^{-1/2})\end{aligned}$$

y debido algunos resultados asintóticos estándar, se puede demostrar que $\sqrt{n} \hat{\rho} \sim N(\mathbf{0}, \mathcal{J}^{\rho|\theta_*}(\boldsymbol{\theta}_0))$, puesto que $\hat{\boldsymbol{\theta}}_0 \xrightarrow{\text{Pr}} \boldsymbol{\theta}_0$, $-n \mathcal{H}^{\rho|\theta_*}(\hat{\boldsymbol{\theta}}_0) = -n \mathcal{I}^{\rho|\theta_*}(\hat{\boldsymbol{\theta}}_0) + \mathcal{O}(n^{-1/2}) \xrightarrow{\text{Pr}} \mathcal{J}^{\rho|\theta_*}(\boldsymbol{\theta}_0)$, entonces se obtenemos:

$$\frac{1}{\sqrt{n}} \mathcal{S}_\rho(\hat{\boldsymbol{\theta}}_0) \sim N\left(\mathbf{0}, [\mathcal{J}^{\rho|\theta_*}(\boldsymbol{\theta}_0)]^{-1}\right)$$

Por lo tanto, cuando $n \rightarrow \infty$, el estadístico de prueba del test de score está dado por:

$$SC_a = \left(\frac{1}{\sqrt{n}} \mathcal{S}_\rho(\boldsymbol{\theta}) \right) (n\mathcal{I}^{\rho|\theta^*}(\boldsymbol{\theta})) \left(\frac{1}{\sqrt{n}} \mathcal{S}_\rho(\boldsymbol{\theta}) \right) \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}$$

la cual converge a una distribución chi-cuadrado con un grado de libertad cuando H_0 es cierta, lo cual denotaremos por $SC_a \sim \chi_1^2$. Y así, para evaluar la hipótesis (5.2), se calcula el el p-valor asociado al estadístico de Score bajo el supuesto distribucional elíptico.

IV.2. Distribución asintótica SC_h

Para encontrar la distribución de SC_h realizaremos los siguientes supuestos adicionales sobre la matriz de información de Fisher asociada al modelo, para algún punto arbitrario $\boldsymbol{\theta}^*$ dentro de una vecindad de $\boldsymbol{\theta}_0^*$, supondremos que:

- 1) La matriz $\mathcal{I}_{\theta\theta}(\boldsymbol{\theta}^*)$ es invertible, para algún punto arbitrario de n .
- 2) $n\mathcal{I}^{\lambda|\theta}(\boldsymbol{\theta}^*) \rightarrow \mathcal{J}^{\lambda|\theta}(\boldsymbol{\theta}^*) > 0$, cuando $n \rightarrow \infty$ uniformemente.

y así se obtiene que:

$$\frac{1}{n} [\mathcal{I}^{\lambda|\theta}(\boldsymbol{\theta}_0^*)]^{-1} \rightarrow [\mathcal{J}^{\lambda|\theta}(\boldsymbol{\theta}_0^*)]^{-1} > 0$$

Además, utilizaremos la siguiente notación:

$$\begin{aligned} \mathbb{S}_p(\boldsymbol{\theta}^*) &= \frac{\partial l_p(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}^*} = [\mathbb{S}_p(\boldsymbol{\lambda}) \quad [\mathbb{S}_p(\boldsymbol{\theta})]^t]^t = [\mathcal{S}_\lambda(\boldsymbol{\theta}^*) \quad [\mathcal{S}_\theta(\boldsymbol{\theta}^*)]^t]^t \\ \mathbb{H}(\boldsymbol{\theta}^*) &= \frac{\partial^2 l(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}^* \partial \boldsymbol{\theta}^{*t}} = \begin{bmatrix} \mathcal{H}_{\lambda\lambda} & \mathcal{H}_{\lambda\theta} \\ \mathcal{H}_{\theta\lambda} & \mathcal{H}_{\theta\theta} \end{bmatrix} \\ \mathbb{H}^{-1}(\boldsymbol{\theta}^*) &= \begin{bmatrix} \mathcal{H}^{\lambda\lambda} & \mathcal{H}^{\lambda\theta} \\ \mathcal{H}^{\theta\lambda} & \mathcal{H}^{\theta\theta} \end{bmatrix} \end{aligned}$$

Entonces, por medio de series de Taylor, tenemos que la aproximación de primer orden del $\mathbb{S}_p(\boldsymbol{\theta}^*)$ evaluada en $\hat{\boldsymbol{\theta}}^*$ alrededor de $\hat{\boldsymbol{\theta}}_0^*$ esta dado por:

$$\begin{aligned} \mathbb{S}_p(\hat{\boldsymbol{\theta}}^*) &= \mathbb{S}_p(\hat{\boldsymbol{\theta}}_0^*) + \frac{\partial}{\partial \boldsymbol{\theta}^{*t}} \mathbb{S}_p(\hat{\boldsymbol{\theta}}_0^*) [\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}_0^*] + \mathcal{O}(1) \\ \mathbb{S}_p(\hat{\boldsymbol{\theta}}^*) &= \mathbb{S}_p(\hat{\boldsymbol{\theta}}_0^*) + \mathbb{H}(\hat{\boldsymbol{\theta}}_0^*) [\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}_0^*] + \mathcal{O}(1) \\ \mathbb{S}_p(\hat{\boldsymbol{\theta}}^*) - \mathbb{S}_p(\hat{\boldsymbol{\theta}}_0^*) &= \mathbb{H}(\hat{\boldsymbol{\theta}}_0^*) \Delta \hat{\boldsymbol{\theta}}^* + \mathcal{O}(1) \end{aligned}$$

donde $\Delta\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}}_0^* = [\hat{\boldsymbol{\lambda}} \quad \Delta\hat{\boldsymbol{\theta}}^t]^t$. Además, sabemos que $\mathbb{S}_p(\hat{\boldsymbol{\theta}}^*) = \mathbf{0}$ y $\mathcal{S}_\theta(\hat{\boldsymbol{\theta}}_0^*) = \mathbf{0}$ que al evaluarlas obtenemos:

$$\begin{aligned} \mathbf{0} - \mathbb{S}_p(\hat{\boldsymbol{\theta}}_0^*) &= \mathbb{H}(\hat{\boldsymbol{\theta}}_0^*)\Delta\hat{\boldsymbol{\theta}}^* + \mathcal{O}(1) \\ \begin{bmatrix} \mathcal{S}_\lambda(\hat{\boldsymbol{\theta}}_0^*) \\ \mathcal{S}_\theta(\hat{\boldsymbol{\theta}}_0^*) \end{bmatrix} &= - \begin{bmatrix} \mathcal{H}_{\lambda\lambda}(\hat{\boldsymbol{\theta}}_0^*) & \mathcal{H}_{\lambda\theta}(\hat{\boldsymbol{\theta}}_0^*) \\ \mathcal{H}_{\theta\lambda}(\hat{\boldsymbol{\theta}}_0^*) & \mathcal{H}_{\theta\theta}(\hat{\boldsymbol{\theta}}_0^*) \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\lambda}} \\ \Delta\hat{\boldsymbol{\theta}} \end{bmatrix} + \mathcal{O}(1) \\ \begin{bmatrix} \mathcal{S}_\lambda(\hat{\boldsymbol{\theta}}_0^*) \\ \mathbf{0} \end{bmatrix} &= - \begin{bmatrix} \mathcal{H}_{\lambda\lambda}(\hat{\boldsymbol{\theta}}_0^*)\hat{\boldsymbol{\lambda}} + \mathcal{H}_{\lambda\theta}(\hat{\boldsymbol{\theta}}_0^*)\Delta\hat{\boldsymbol{\theta}} \\ \mathcal{H}_{\theta\lambda}(\hat{\boldsymbol{\theta}}_0^*)\hat{\boldsymbol{\lambda}} + \mathcal{H}_{\theta\theta}(\hat{\boldsymbol{\theta}}_0^*)\Delta\hat{\boldsymbol{\theta}} \end{bmatrix} + \mathcal{O}(1) \end{aligned}$$

pero:

$$\begin{aligned} \mathbf{0} &= \mathcal{H}_{\theta\lambda}(\hat{\boldsymbol{\theta}}_0^*)\hat{\boldsymbol{\lambda}} + \mathcal{H}_{\theta\theta}(\hat{\boldsymbol{\theta}}_0^*)\Delta\hat{\boldsymbol{\theta}} \\ \mathcal{H}_{\theta\theta}(\hat{\boldsymbol{\theta}}_0^*)\Delta\hat{\boldsymbol{\theta}} &= -\mathcal{H}_{\theta\lambda}(\hat{\boldsymbol{\theta}}_0^*)\hat{\boldsymbol{\lambda}} \\ \Delta\hat{\boldsymbol{\theta}} &= -\mathcal{H}_{\theta\theta}^{-1}(\hat{\boldsymbol{\theta}}_0^*)\mathcal{H}_{\theta\lambda}(\hat{\boldsymbol{\theta}}_0^*)\hat{\boldsymbol{\lambda}} \end{aligned}$$

y así, tenemos que:

$$\begin{aligned} \mathcal{S}_\lambda(\hat{\boldsymbol{\theta}}_0^*) &= - \left(\mathcal{H}_{\lambda\lambda}(\hat{\boldsymbol{\theta}}_0^*)\hat{\boldsymbol{\lambda}} + \mathcal{H}_{\lambda\theta}(\hat{\boldsymbol{\theta}}_0^*)\Delta\hat{\boldsymbol{\theta}} \right) + \mathcal{O}(1) \\ &= - \left(\mathcal{H}_{\lambda\lambda}(\hat{\boldsymbol{\theta}}_0^*)\hat{\boldsymbol{\lambda}} - \mathcal{H}_{\lambda\theta}(\hat{\boldsymbol{\theta}}_0^*)\mathcal{H}_{\theta\theta}^{-1}(\hat{\boldsymbol{\theta}}_0^*)\mathcal{H}_{\theta\lambda}(\hat{\boldsymbol{\theta}}_0^*)\hat{\boldsymbol{\lambda}} \right) + \mathcal{O}(1) \\ &= - \left[\mathcal{H}^{\lambda|\theta}(\hat{\boldsymbol{\theta}}_0^*) \right]^{-1} \hat{\boldsymbol{\lambda}} + \mathcal{O}(1) = - \left[\mathcal{H}^{\lambda|\theta}(\hat{\boldsymbol{\theta}}_0^*) \right]^{-1} \hat{\boldsymbol{\lambda}} \frac{\sqrt{n}\sqrt{n}}{n} + \mathcal{O}(1) \\ \frac{1}{\sqrt{n}}\mathcal{S}_\lambda(\hat{\boldsymbol{\theta}}_0^*) &= \left[-n\mathcal{H}^{\lambda|\theta}(\hat{\boldsymbol{\theta}}_0^*) \right]^{-1} \sqrt{n}\hat{\boldsymbol{\lambda}} + \mathcal{O}(n^{-1/2}) \end{aligned}$$

y debido algunos resultados asintóticos estándar, para más detalles pueden verse en Cox & Hinkley (1974) y Van der Vart (2000), por lo cual, se puede demostrar que: $\sqrt{n}\hat{\boldsymbol{\lambda}} \sim N(\mathbf{0}, \mathcal{J}^{\lambda|\theta}(\boldsymbol{\theta}_0^*))$, puesto que $\hat{\boldsymbol{\theta}}_0^* \xrightarrow{\text{Pr}} \boldsymbol{\theta}_0^*$, $-n\mathcal{H}^{\lambda|\theta}(\hat{\boldsymbol{\theta}}_0^*) = -n\mathcal{I}^{\lambda|\theta}(\hat{\boldsymbol{\theta}}_0^*) + \mathcal{O}(n^{-1/2}) \xrightarrow{\text{Pr}} \mathcal{J}^{\lambda|\theta}(\boldsymbol{\theta}_0^*)$. Por lo cual, se obtiene que:

$$\frac{1}{\sqrt{n}}\mathcal{S}_\lambda(\hat{\boldsymbol{\theta}}_0^*) \sim N\left(\mathbf{0}, [\mathcal{J}^{\lambda|\theta}(\boldsymbol{\theta}_0^*)]^{-1}\right)$$

Por lo tanto, cuando $n \rightarrow \infty$, el estadístico de prueba del test de Score está dado por:

$$\text{SC}_h = \left(\frac{1}{\sqrt{n}}\mathcal{S}_\lambda(\boldsymbol{\theta}^*) \right) \left(n\mathcal{I}^{\lambda|\theta}(\boldsymbol{\theta}^*) \right) \left(\frac{1}{\sqrt{n}}\mathcal{S}_\lambda(\boldsymbol{\theta}^*) \right) \Bigg|_{\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}_0^*}$$

la cual converge a una distribución chi-cuadrado con b grados de libertad cuando H_0 es cierta, lo cual denotaremos por $SC_h \sim \chi_b^2$. Y así, para evaluar la hipótesis (5.8), se calcula el el p-valor asociado al estadístico de Score bajo este supuesto distribucional.

APÉNDICE V: DATOS FOSFATO
INORGÁNICO DE PLASMA

Paciente	Grupo	Tiempo de observación (horas)							
		0	0,5	1	1,5	2	3	4	5
1	Control	4,3	3,3	3	2,6	2,2	2,5	3,4	4,4
2	Control	3,7	2,6	2,6	1,9	2,9	3,2	3,1	3,9
3	Control	4	4,1	3,1	2,3	2,9	3,1	3,9	4
4	Control	3,6	3	2,2	2,8	2,9	3,9	3,8	4
5	Control	4,1	3,8	2,1	3	3,6	3,4	3,6	3,7
6	Control	3,8	2,2	2	2,6	3,8	3,6	3	3,5
7	Control	3,8	3	2,4	2,5	3,1	3,4	3,5	3,7
8	Control	4,4	3,9	2,8	2,1	3,6	3,8	4	3,9
9	Control	5	4	3,4	3,4	3,3	3,6	4	4,3
10	Control	3,7	3,1	2,9	2,2	1,5	2,3	2,7	2,8
11	Control	3,7	2,6	2,6	2,3	2,9	2,2	3,1	3,9
12	Control	4,4	3,7	3,1	3,2	3,7	4,3	3,9	4,8
13	Control	4,7	3,1	3,2	3,3	3,2	4,2	3,7	4,3
14	Obeso	4,3	3,3	3	2,6	2,2	2,5	2,4	3,4
15	Obeso	5	4,9	4,1	3,7	3,7	4,1	4,7	4,9
16	Obeso	4,6	4,4	3,9	3,9	3,7	4,2	4,8	5
17	Obeso	4,3	3,9	3,1	3,1	3,1	3,1	3,6	4
18	Obeso	3,1	3,1	3,3	2,6	2,6	1,9	2,3	2,7
19	Obeso	4,8	5	2,9	2,8	2,2	3,1	3,5	3,6
20	Obeso	3,7	3,1	3,3	2,8	2,9	3,6	4,3	4,4
21	Obeso	5,4	4,7	3,9	4,1	2,8	3,7	3,5	3,7
22	Obeso	3	2,5	2,3	2,2	2,1	2,6	3,2	3,5
23	Obeso	4,9	5	4,1	3,7	3,7	4,1	4,7	4,9
24	Obeso	4,8	4,3	4,7	4,6	4,7	3,7	3,6	3,9
25	Obeso	4,4	4,2	4,2	3,4	3,5	3,4	3,9	4
26	Obeso	4,9	4,3	4	4	3,3	4,1	4,2	4,3
27	Obeso	5,1	4,1	4,6	4,1	3,4	4,2	4,4	4,9
28	Obeso	4,8	4,6	4,6	4,4	4,1	4	3,8	3,8
29	Obeso	4,2	3,5	3,8	3,6	3,3	3,1	3,5	3,9
30	Obeso	6,6	6,1	5,2	4,1	4,3	3,8	4,2	4,8
31	Obeso	3,6	3,4	3,1	2,8	2,1	2,4	2,5	3,5
32	Obeso	4,5	4	3,7	3,3	2,4	2,3	3,1	3,3
33	Obeso	4,6	4,4	3,8	3,8	3,8	3,6	3,8	3,8

APÉNDICE VI: MEDICIONES DE
ÍNDICES ODONTOLÓGICOS

Paciente	Grupo	Género	Edad	Día	Índices Odontológicos		
					IH	IP	PS
1	Control	Femenino	15	0	0	0,00	2,00
1	Control	Femenino	15	7	24,75	29,00	2,50
1	Control	Femenino	15	14	16,5	16,50	2,00
1	Control	Femenino	15	21	16,5	24,50	2,50
2	CPC	Femenino	15	0	0	0,00	2,00
2	CPC	Femenino	15	7	54	37,25	2,25
2	CPC	Femenino	15	14	45,75	41,50	1,75
2	CPC	Femenino	15	21	33	32,75	2,50
3	Control	Masculino	17	0	0	0,00	1,75
3	Control	Masculino	17	7	0	8,25	2,00
3	Control	Masculino	17	14	0	20,75	1,25
3	Control	Masculino	17	21	0	37,25	2,00
4	Control	Femenino	15	0	0	0,00	2,00
4	Control	Femenino	15	7	8,25	24,75	2,50
4	Control	Femenino	15	14	16,5	20,75	2,00
4	Control	Femenino	15	21	20,25	24,75	2,50
5	Control	Femenino	14	0	0	0,00	1,50
5	Control	Femenino	14	7	33	37,25	2,25
5	Control	Femenino	14	14	33	41,50	2,25
5	Control	Femenino	14	21	12,25	58,00	2,25
6	Control	Femenino	17	0	0	0,00	1,50
6	Control	Femenino	17	7	16,5	37,25	2,25
6	Control	Femenino	17	14	12,25	28,75	2,00
6	Control	Femenino	17	21	20,25	16,00	2,00
7	Control	Masculino	17	0	0	0,00	2,00
7	Control	Masculino	17	7	16,5	41,50	2,00
7	Control	Masculino	17	14	12	37,25	2,50
7	Control	Masculino	17	21	37,25	45,75	2,75
8	Control	Masculino	16	0	0	0,00	1,75
8	Control	Masculino	16	7	37,25	16,50	1,25
8	Control	Masculino	16	14	33	45,50	1,75
8	Control	Masculino	16	21	20,5	33,00	2,25

Paciente	Grupo	Género	Edad	Día	Índices Odontológicos		
					IH	IP	PS
9	CHX	Femenino	15	0	0	0,00	1,50
9	CHX	Femenino	15	7	8,25	16,50	2,00
9	CHX	Femenino	15	14	12,25	8,00	2,25
9	CHX	Femenino	15	21	4	16,25	2,25
10	Control	Femenino	14	0	0	0,00	2,00
10	Control	Femenino	14	7	24,75	49,75	1,75
10	Control	Femenino	14	14	16,5	28,75	3,00
10	Control	Femenino	14	21	12	24,75	2,50
11	Control	Masculino	13	0	0	0,00	2,00
11	Control	Masculino	13	7	16,5	24,75	2,50
11	Control	Masculino	13	14	8	16,25	2,75
11	Control	Masculino	13	21	4	12,25	2,75
12	Control	Femenino	14	0	0	0,00	2,00
12	Control	Femenino	14	7	24,75	41,50	2,00
12	Control	Femenino	14	14	20,5	24,50	2,75
12	Control	Femenino	14	21	41,25	53,75	2,50
13	CPC	Femenino	14	0	0	0,00	2,00
13	CPC	Femenino	14	7	20,75	37,25	2,00
13	CPC	Femenino	14	14	12	12,25	3,00
13	CPC	Femenino	14	21	33	12,25	3,00
14	CPC	Femenino	14	0	0	0,00	2,00
14	CPC	Femenino	14	7	16,5	45,50	2,00
14	CPC	Femenino	14	14	28,75	37,25	2,75
14	CPC	Femenino	14	21	16	33,00	2,50
15	Control	Femenino	15	0	0	0,00	2,00
15	Control	Femenino	15	7	8,25	16,50	1,75
15	Control	Femenino	15	14	0	16,25	2,50
15	Control	Femenino	15	21	12,25	41,25	2,50
16	CPC	Masculino	13	0	0	0,00	2,00
16	CPC	Masculino	13	7	16,5	54,00	2,25
16	CPC	Masculino	13	14	24,5	49,50	2,00
16	CPC	Masculino	13	21	28,75	49,75	2,25
17	CPC	Masculino	15	0	0	0,00	1,50
17	CPC	Masculino	15	7	16,5	8,25	2,00
17	CPC	Masculino	15	14	16,25	53,75	2,50
17	CPC	Masculino	15	21	24,5	53,75	2,50

Paciente	Grupo	Género	Edad	Día	Índices Odontológicos		
					IH	IP	PS
18	Control	Femenino	15	0	0	0,00	2,00
18	Control	Femenino	15	7	8	12,00	2,75
18	Control	Femenino	15	14	8	8,00	3,00
18	Control	Femenino	15	21	4	20,50	2,50
19	CPC	Masculino	15	0	0	0,00	2,00
19	CPC	Masculino	15	7	24,5	37,25	2,25
19	CPC	Masculino	15	14	20,25	28,75	2,50
19	CPC	Masculino	15	21	28,75	24,50	2,75
20	CPC	Masculino	15	0	0	0,00	2,00
20	CPC	Masculino	15	7	24,5	24,50	2,50
20	CPC	Masculino	15	14	16,25	28,75	2,50
20	CPC	Masculino	15	21	8	8,00	2,25
21	CPC	Femenino	13	0	0	0,00	2,00
21	CPC	Femenino	13	7	4	8,25	2,50
21	CPC	Femenino	13	14	8	8,00	2,50
21	CPC	Femenino	13	21	12	16,25	2,75
22	Control	Masculino	16	0	0	0,00	2,00
22	Control	Masculino	16	7	8	53,75	2,50
22	Control	Masculino	16	14	16,25	37,25	2,50
22	Control	Masculino	16	21	8	50,00	2,00
23	CPC	Femenino	15	0	0	0,00	2,00
23	CPC	Femenino	15	7	16	33,00	2,00
23	CPC	Femenino	15	14	24,5	24,50	2,75
23	CPC	Femenino	15	21	24,5	33,00	2,50
24	CHX	Masculino	14	0	0	0,00	2,00
24	CHX	Masculino	14	7	8	24,50	2,75
24	CHX	Masculino	14	14	8,25	33,00	2,50
24	CHX	Masculino	14	21	4	16,50	2,50
25	CPC	Femenino	15	0	0	0,00	2,00
25	CPC	Femenino	15	7	28,75	28,75	2,00
25	CPC	Femenino	15	14	16,5	24,75	2,00
25	CPC	Femenino	15	21	16,25	20,50	1,75
26	CHX	Masculino	14	0	0	0,00	2,00
26	CHX	Masculino	14	7	16	16,50	2,75
26	CHX	Masculino	14	14	20,5	16,00	3,00
26	CHX	Masculino	14	21	12	24,75	2,75

Paciente	Grupo	Género	Edad	Día	Índices Odontológicos		
					IH	IP	PS
27	CPC	Femenino	14	0	0	0,00	2,00
27	CPC	Femenino	14	7	24,5	37,25	2,50
27	CPC	Femenino	14	14	33	20,25	2,50
27	CPC	Femenino	14	21	28,75	33,00	2,50
28	CHX	Femenino	14	0	0	0,00	2,00
28	CHX	Femenino	14	7	16,25	12,00	2,50
28	CHX	Femenino	14	14	16,25	12,00	2,25
28	CHX	Femenino	14	21	8	8,00	2,50
29	CHX	Femenino	13	0	0	0,00	2,00
29	CHX	Femenino	13	7	20,25	8,00	3,00
29	CHX	Femenino	13	14	12,25	0,00	2,75
29	CHX	Femenino	13	21	12,25	4,00	2,50
30	CHX	Masculino	17	0	0	0,00	2,00
30	CHX	Masculino	17	7	12,25	16,00	2,50
30	CHX	Masculino	17	14	4	8,00	2,25
30	CHX	Masculino	17	21	4	0,00	1,75
31	CHX	Femenino	15	0	0	0,00	2,00
31	CHX	Femenino	15	7	8	33,00	2,50
31	CHX	Femenino	15	14	12,25	20,25	3,00
31	CHX	Femenino	15	21	16,25	28,75	2,50
32	CHX	Masculino	17	0	0	0,00	2,00
32	CHX	Masculino	17	7	16,25	24,75	2,25
32	CHX	Masculino	17	14	20,5	16,50	2,50
32	CHX	Masculino	17	21	20,5	16,25	2,25
33	CHX	Femenino	15	0	0	0,00	2,00
33	CHX	Femenino	15	7	12	20,50	2,50
33	CHX	Femenino	15	14	4	16,25	2,25
33	CHX	Femenino	15	21	0	16,50	2,25
34	CHX	Femenino	14	0	0	0,00	2,00
34	CHX	Femenino	14	7	16,25	16,25	2,25
34	CHX	Femenino	14	14	16,25	12,00	2,00
34	CHX	Femenino	14	21	12	12	2,00

Paciente	Grupo	Género	Edad	Día	Índices Odontológicos		
					IH	IP	PS
35	CHX	Femenino	13	0	0	0	2
35	CHX	Femenino	13	7	4	20,5	2,25
35	CHX	Femenino	13	14	8	12,25	2,75
35	CHX	Femenino	13	21	4	16,5	2,25
36	CHX	Masculino	17	0	0	0	2
36	CHX	Masculino	17	7	20,25	24,5	2,5
36	CHX	Masculino	17	14	24,5	37	2,75
36	CHX	Masculino	17	21	16	24,75	2,5
37	CHX	Femenino	13	0	0	0	2
37	CHX	Femenino	13	7	16,25	37,25	2,75
37	CHX	Femenino	13	14	4	37,25	2,25
37	CHX	Femenino	13	21	16,25	41,5	2,75
38	CPC	Femenino	16	0	0	0	2
38	CPC	Femenino	16	7	24,5	8	2,5
38	CPC	Femenino	16	14	12,25	8	2,5
38	CPC	Femenino	16	21	8	8	2,5
39	CPC	Femenino	13	0	0	0	2
39	CPC	Femenino	13	7	20,5	16,25	2,25
39	CPC	Femenino	13	14	8	16,25	2,75
39	CPC	Femenino	13	21	20,25	12,25	2,75
40	Control	Femenino	13	0	0	0	2
40	Control	Femenino	13	7	8,25	12	2,25
40	Control	Femenino	13	14	4	28,75	2
40	Control	Femenino	13	21	12,25	24,5	2,25
41	CHX	Femenino	13	0	0	0	2
41	CHX	Femenino	13	7	0	16,5	2,5
41	CHX	Femenino	13	14	4	4	2,25
41	CHX	Femenino	13	21	0	4	2,5

BIBLIOGRAFÍA

- Akaike, H. (1974) A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Arnau, J., Balluerka, N. (2004). Análisis de datos longitudinales y de curvas de crecimiento. Enfoque clásico y propuestas actuales, *Psicothema*, 16, 156-162.
- Bastos, Pinho, L. G., Nobre, J. S., De Freitas, S. M. (2012). On linear mixed models and their influence diagnostics applied to an actuarial problem, *Chilean Journal of Statistics (ChJS)*, 3(1), 57-73.
- Bates, D. M., Watts, D. G. (1988). Nonlinear regression analysis and its applications, *John Wiley & Sons*.
- Bera, Anil K., Biliyas, Yannis (2001) Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: an essay on historical developments and some new results, *Journal of Statistical planning and inference*, 97, 9-44.
- Bera, A., Ullah, A. (1991). Rao's Score Test in Econometrics (Discussion Paper No. 1991-43), *Tilburg University, Center for Economic Research*, 9143.
- Belsley, David A., Edwin Kuh, Roy E. Welsch (2005) Regression diagnostics - identifying influential data and sources of collinearity, *Wiley Series in Probability and Statistics*.
- Bhattacharyya, G. K. (1967). Asymptotic Efficiency of Multivariate Normal Score Test, *The Annals of Mathematical Statistics*, 38(6), 1753-1758.

- Bickel, P. J. (2001). Mathematical statistics: basic ideas and selected topics (2nd ed.), *TUpper Saddle River, N.J: Prentice Hall*.
- Bolfarine, H., Montenegro, L. C., Lachos, V. H. (2007). Influence Diagnostics for Skew-Normal Linear Mixed Models, *Sankhyā: The Indian Journal of Statistics (2003-2007)*, 69(4), 648-670.
- Buja, A., Hastie, T., Tibshirani, R. (1989). Linear Smoothers and Additive-Models, *Annals of Statistics*, 17(2), 453-510.
- Burnham, K. P. (2002). Model selection and multimodel inference: a practical information-theoretic approach (2nd ed.), *New York, Springer*.
- Buse, A. (1982). The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note, *The American Statistician*, 36(3), 153-157.
- Cao, C. Z., Lin, J. G. (2011) Diagnostics for elliptical linear mixed models with first-order autoregressive errors, *Journal of Statistical Computation and Simulation*, 81(10), 1281-1296.
- Cao, C. Z., Lin, J. G. (2012) Heteroscedasticity and/or autocorrelation checks in longitudinal nonlinear models with elliptical and AR(1) errors. *Acta Mathematicae Applicatae Sinica-English Series*, 28(1), 49-62.
- Carrasco, J. M. R., Ortega, E. M. M., Paula, G. A. (2008). Log-modified Weibull regression models with censored data: Sensitivity and residual analysis, *Computational Statistics & Data Analysis*, 52(8), 4021-4039.
- Castaño, E. (1988). Observaciones Influenciales, *Revista Colombiana de Estadística*, (17-18), 38-74.
- Chandra, T. K., Samanta, T. (1988). On the second order local comparison between perturbed maximum likelihood estimators and Rao's statistic as test statistics, *Journal of Multivariate Analysis*, 25(2), 201-222.
- Chatterjee, S., Hadi, A. S. (1986). Influential Observations, High Leverage Points, and Outliers in Linear Regression, *Statistical Science*, 1(3), 379-393.
- Chen, C. F. (1983) Score Tests for Regression Models. *Journal of the American Statistical Association*, 78(381), 158-161.

- Chen, Y. Q., Wang, M. C., Huang, Y. J. (2004) Semiparametric regression analysis on longitudinal pattern of recurrent gap times. *Biostatistics*, 5(2), 277-290.
- Chi, E., Reinsel, G. (1989) Models for Longitudinal Data with Random Effects and AR(1) Errors. *Journal of the American Statistical Association*, 84(406), 452-459.
- Chi, E., Reinsel, G. (1991) Asymptotic Properties of the Score Test for Autocorrelation in a Random Effects with AR(1) Errors Model. *Statistics & Probability Letters*, 11(5), 453-457.
- Cordeiro, G. M. (2004) Corrected likelihood ratio tests in symmetric nonlinear regression models. *Journal of Statistical Computation and Simulation*, 74(8), 609-620.
- Cook, R. D. (1977). Detection of Influential Observation in Linear Regression, *Technometrics*, 19(1), 15.
- Cook, D. (1986). Assessment of Local Influence, *Journal of the Royal Statistical Society*, 48(2), 133-169.
- Cox, D.R., Hinkley, D.V. (1974) theoretical Statistics. *Chapman & Hall*, London.
- Cysneiros, Francisco José A., Paula, G. A. (2005) Restricted methods in symmetrical linear regression models. *Comput. Stat. Data Anal.*, 49(3), 689-708.
- Cysneiros, Francisco José A., Paula, G. A., Galea, M. (2007) Heteroscedastic symmetrical linear models. *Statistics & Probability Letters*, 77(11), 1084-1090.
- DasGupta, A. (2008) Asymptotic Theory of Statistics and Probability. Springer-Verlag New York Inc.
- Diaz-Garcia, J. A., Gonzalez-Farias, G. (2004). A note on the Cook's distance, *Journal of Statistical Planning and Inference*, 120(1-2), 119-136.
- Diggle, P., Heagarty, P., Liang, K. Y., Zeger, S. (2002) Analysis of Longitudinal Data (*second edition*) (2nd ed., Vol. 3). Gran Bretaña: Oxford University Press.

- Efromovich, S. (1999). Nonparametric Curve Estimation - Methods, Theory, and Applications, *USA, Springer*.
- Eggermont, P. P., LaRiccia, V. N. (2009) Maximum Penalized Likelihood Estimation: Volume II: Regression (1st ed.). *Springer*.
- Eubank, R. L. (1984). The hat matrix for smoothing splines, *Statistics & Probability Letters*, 2, 9-14.
- Eubank, R. L. (1999). Nonparametric regression and spline smoothing (2nd ed.), *New York, Marcel Dekker*.
- Fang, K. T., Zhang, Y. T. (1990) Generalized multivariate analysis. *Science Press*.
- Fang, K.T., Kotz, S., Ng, K. W. (1989) Symmetric Multivariate and Related Distributions. *Chapman and Hall/CRC*.
- Feng, L., Zou, C., Wang, Z., Chen, B. (2013). Rank-based score tests for high-dimensional regression coefficients, *Electronic Journal of Statistics*, 7, 2131-2149.
- Fengler, M. (2005) Semiparametric Modeling of Implied Volatility. *Springer*.
- Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G. (2008). Longitudinal Data Analysis, *CRC Press*.
- Fitzmaurice, G. M., Laird, N. M., Rotnitzky, A. G. (1993). Regression Models for Discrete Longitudinal Responses, *Statistical Science*, 8(3), 284-299.
- Freedman, D. A. (2007). How Can the Score Test Be Inconsistent?, *The American Statistician*, 61(November), 291-295.
- Frieden, B. R. (1998). Physics from Fisher information: a unification, Cambridge, U.K.; New York, N.Y: *Cambridge University Press*.
- Frieden, B. R., Gatenby, R. A. (Eds.). (2007). Exploratory data analysis using Fisher information, London, *Springer*.
- Fung, W.K., Zhu, Z.Y., Wei, B.C., He X. (2002). Influence diagnostics and outlier tests for semiparametric mixed models. *Journal of the Royal Statistical Society B-Statistical Methodology*, 64(3), 565-579.

- Galea, M., Paula, G. A., Bolfarine, H. (1997) Local influence in elliptical linear regression models. *Statistician*, 46(1), 71-79.
- Galea, M., Paula, G. A., Cysneiros, F. J. A. (2005) On diagnostics in symmetrical nonlinear models. *Elsevier BV*, 73(4), 459-467.
- Galea, M., Paula, G. A., Uribe-Opazo, M. (2003) On influence diagnostic in univariate elliptical linear regression models. *Statistical Papers*, 44(1), 23-45.
- Galea, M., Riquelme, M., Paula, G. A. (2000) Diagnostic Methods in Elliptical Linear Regression Models. *Brazilian Journal of Probability and Statistics*, 14, 167-184.
- García, J. A. D., Jaime, O. A. M. (2004). Una modificación a la distancia de Cook, *Acta Universitaria*, 14(3), 21-31.
- Govindarajulu, Z. (2007). Nonparametric Inference, *New Jersey, World Scientific Publishing Company*.
- Green, P. J. (1987). Penalized Likelihood for General Semi-Parametric Regression Models, *International Statistical Review / Revue Internationale de Statistique*, 55(3), 245-259.
- Green, P. J. (1996). Penalized likelihood, *University of Bristol*.
- Green, P. J., Silverman, B. W. (1993). Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, *Chapman & Hall/CRC*.
- Gu, C. (2013). Smoothing spline ANOVA models (2nd ed.), *New York, Springer*.
- Gupta, A. K., Varga, T. (1993) Elliptically Contoured Models in Statistics (1st ed., Vol. 240). *Kluwic Academic Publishers*.
- Györfi, L., Kohler, M., Krzyzak, A., Walk, H. (2002). A Distribution-Free Theory of Nonparametric Regression, *New York, Springer*.
- Hamilton, S. A., Truong, Y. K. (1997) Local Linear Estimation in Partly Linear Models. *Journal of Multivariate Analysis*, 60(1), 1-19.

- Härdle, W. K., Müller, M., Sperlich, S., Werwatz, A. (2004) Nonparametric and Semiparametric Models. Springer.
- Harville, D. A. (2008) Matrix Algebra From a Statistician's Perspective, Springer.
- Hastie, T. (1990). Generalized additive models (1st ed.), London, New York: Chapman and Hall.
- Hastie, T., Buja, A., Tibshirani, R. (1995). Penalized Discriminant Analysis, *The Annals of Statistics*, 23(1), 73-102.
- Hastie, T., Tibshirani, R. (1986). Generalized Additive Models, *Statistical Science*, 1(3), 297-310.
- Hastie, T., Tibshirani, R. (2000). Bayesian backfitting, *Statistical Science*, 15(3), 196-213.
- Hedeker, D. R. (2006). Longitudinal data analysis. Hoboken, N.J, Wiley-Interscience.
- Horowitz, J. L. (2009) Semiparametric and Nonparametric Methods in Econometrics, Springer.
- Huber, J. L. (1981) Robust Statistics, Wiley.
- Hulin, W., Jin-Ting, Z. (2006) Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches. Springer. <http://onlinelibrary.wiley.com/book/10.1002/0470009675>
- Ibacache-Pulgar, Germán, Paula, G. A. (2011) Local influence for Student- t partially linear models. *Computational Statistics & Data Analysis*, 55(3), 1462-1478.
- Ibacache-Pulgar, German, Paula, G. A., Galea, M. (2012) Influence diagnostics for elliptical semiparametric mixed models. *Statistical Modelling*, 12(2), 165-193.
- Ibacache-Pulgar, Germán, Paula, G. A., Cysneiros, F. J. A. (2013) Semiparametric additive models under symmetric distributions. *TEST*, 22(1), 103-121.

- Ibacache-Pulgar, German, Paula, G. A., Galea, M. (2014) On influence diagnostics in elliptical multivariate regression models with equicorrelated random errors, *Statistical Methodology*, 16, 14-31.
- Jahufer, A., Chen, J. (2009). Assessing global influential observations in modified ridge regression, *Statistics & Probability Letters*, 79(4), 513-518.
- Jamalizadeh A., N. B.(2010) Distributions of order statistics and linear combinations of order statistics from an elliptical distribution as mixtures of unified skew-elliptical distributions, *Journal of Multivariate Analysis*, 101(6), 1412-1427.
- Jiménez, J. (2001) Una generalización de la distancia de Cook, *Revista Colombiana de Estadística*, 24(2), 111-120.
- Jin-Guan Lin, B.C.W. (2004). Testing for Heteroscedasticity and/or Correlation in Nonlinear Models with Correlated Errors, *Communications in Statistics-theory and Methods - COMMUN STATIST-THEOR METHOD*, 33(2), 251-275.
- Jureckova, J., Picek, J. (2005). Robust Statistical Methods with R, *Boca Raton: Chapman and Hall/CRC*.
- Kauermann, G. (2005). A note on smoothing parameter selection for penalized spline smoothing, *Journal of Statistical Planning and Inference*, 127(1-2), 53-69.
- Keele, L. J. (2008) Semiparametric Regression for the Social Sciences (1st ed.). Wiley.
- Khuri, A. I. (2003). Advanced calculus with applications in statistics, 2nd ed., rev. and expanded, *Hoboken, N.J, Wiley-Interscience*.
- Kim, C., Lee, Y., Park, B. U. (2001). Cook's distance in local polynomial regression, *Statistics & Probability Letters*, 54(1), 33-40.
- Choongrak Kim, Byeong U. Park., Woochul, Kim. (2002) Influence diagnostics in semiparametric regression models. *Statistics & Probability Letters*,(1), 49-58.
- Kim, C. R. (1996). Cook's distance in spline smoothing, *Statistics & Probability Letters*, 31(2), 139-144.

- Klein, A., Melard, G., Spreij, P. (2005). On the resultant property of the Fisher information matrix of a vector ARMA process, *Linear Algebra and Its Applications*, 403, 291-313.
- Klonias, V. K. (1982) Consistency of Two Nonparametric Maximum Penalized Likelihood Estimators of the Probability Density Function, *The Annals of Statistics*, 10(3), 811-824.
- Kollo, T. (2005). Advanced multivariate statistics with matrices, Dordrecht, New York; Springer.
- Kosorok, M. R. (2008) Introduction to Empirical Processes and Semiparametric Inference. Springer.
- Lagos, B. M., Morettin, P. A., Barroso, L. P. (2010) Some corrections of the score test statistic for Gaussian ARMA models, *Brazilian Journal of Probability and Statistics*, 24(3), 434-456.
- Laird, N. M., Ware, J. H. (1982). Random-Effects Models for Longitudinal Data, *Biometrics*, 38(4), 963.
- Lange, K.L., Little, R.J.A., Taylor, J.M.G. (1989) Robust Statistical modeling the t distribution, *Journal of the American Statistical Association*, 84(408), 881-896.
- Lee, T. C. M. (2003) Smoothing parameter selection for smoothing splines: a simulation study, *Computational Statistics & Data Analysis*, 42(1-2), 139-148.
- Lee, Y. K., Mammen, E., Park, B. U. (2010) Backfitting and smooth backfitting for additive quantile models, *Annals of Statistics*, 38(5), 2857-2883.
- Lehmann, E. L. (2005). Testing statistical hypotheses, 3rd ed., New York: Springer.
- Lei Shi, G. C. (2008). Detection of outliers in multilevel models, *Journal of Statistical Planning and Inference*,(10), 3189-3199.
- Leiva, Victor, Díaz G., José A. (2001) Distribuciones elípticas multivariadas singulares y no singulares: teoría y aplicaciones, *Universidad Autónoma Agraria Antonio Narro*. http://staff.deuv.cl/leiva/archivos/books/leiva_diaz_2011.pdf

- Lemonte, A. J., Ferrari, S. L. P. (2012). A note on the local power of the LR, Wald, score and gradient tests, *Electronic Journal of Statistics*, 6, 421-434.
- Li, Z., Xu, W., Zhu, L. (2009). Influence diagnostics and outlier tests for varying coefficient mixed models, *Journal of Multivariate Analysis*, 100(9), 2002-2017.
- Liang, H. (2009) Generalized partially linear mixed-effects models incorporating mismeasured covariates, *Annals of the Institute of Statistical Mathematics*, 61(1), 27-46.
- Liang, H., Li, R. (2009) Variable Selection for Partially Linear Models with Measurement Errors. *Journal of the American Statistical Association*, 104(485), 234-248.
- Lin, D. Y., Ying, Z. L. (2003) Semiparametric regression analysis of longitudinal data with informative drop-outs. *Biostatistics*, 4(3), 385-398.
- Lin, J. G., Wei, B. C. (2004) Testing for Heteroscedasticity and/or Correlation in Nonlinear Models with Correlated Errors. *Communications in Statistics - Theory and Methods*, 33(2), 251-275.
- Lin, J. G., Wei, B. C. (2007) Testing for Heteroscedasticity and/or Autocorrelation in Longitudinal Mixed Effect Nonlinear Models with AR(1) Errors. *Communications in Statistics - Theory and Methods*, 36(3), 567-586.
- Lin, T.I. (2008). Longitudinal data analysis using t linear mixed models with autoregressive dependence structures, *Journal of Data Science*, 6, 333-355.
- Lindsay, B. G., Qu, A. (2003) Inference Functions and Quadratic Score Tests. *Statistical Science*, 18(3), 394-410.
- Liu, S. (2000) On local influence for elliptical linear models. *Statistical Papers*, 41(2), 211-224.
- Liu, S. (2002) Local influence in multivariate elliptical linear regression models. *Linear Algebra and its Applications*, 354(1-3), 159-174.
- Liu, M., Bozdogan, H. (2008) Multivariate Regression Models with Power Exponential Random Errors and Subset Selection Using Genetic Algorithms With Information Complexity, *European Journal of Pure and Applied Mathematics*, 1(1), 4-37.

- Lu, T.T., Shiou S.H. (2002). Inverses of 2×2 block matrices. *Computers and mathematics with applications* 43, 119-129.
- Luo, Z. (1998). Backfitting in smoothing spline ANOVA, *Annals of Statistics* 26(5), 1733-1759.
- Magnus, J. R., Neudecker, H. (1999) Matrix Differential Calculus with Applications in Statistics and Econometrics, 2nd Edition, Wiley.
- Mammen, E., Park, B. U. (2005) Bandwidth selection for smooth backfitting in additive models. *Annals of Statistics*, *Annals of Statistics*, 33(3), 1260-1294.
- Mammen, E., Park, B. U. (2006) A simple smooth backfitting method for additive models, *Annals of Statistics*, 34(5), 2252-2271.
- Manor, O., Zucker, D. M. (2004) Small sample inference for the fixed effects in the mixed linear model, *Computational Statistics & Data Analysis*, 46(4), 801-817.
- Mardia, K. V., Kent, J. T., Bibby, J. M. (1979) Multivariate Analysis , 1st edition, Academic Press.
- Maronna, R. A. (2006). Robust Statistics: Theory and Methods, 1st edition, Chichester, England; Wiley.
- Maruyama, Y., Seo, T. (2003) Estimation of Moment Parameter in Elliptical Distributions, *Journal Japan Statistical Society*, 33(2), 215-229.
- Molenberghs, G. (2005) Models for discrete longitudinal data, New York, London, Springer.
- Moneta, G. B. (1991). Implicit Construction of McCulloch's G Matrix for the Numerical Evaluation of Fisher Information Matrixes, *Comput. Stat. Data Anal.*, 11(3), 333-344.
- Muller, K. E., Stewart, P. W. (2006) Linear model theory: univariate, multivariate and mixed models, John Wiley & Sons, Inc.
- Olaverri, C. G. (1996) Estabilidad de algunos criterios de selección de modelos. : quaderns d'estadística i investigació operativa, *Qüestió*, 20(2). Recuperado a partir de <http://www.raco.cat/index.php/Questiio/article/view/26766>

- Osorio, F., Paula, G. A., Galea, M. (2007) Assessment of local influence in elliptical linear models with longitudinal structure, *Computational Statistics & Data Analysis*, 51(9), 4354-4368.
- Paula, G. A., Medeiros, M., Vilca-Labra, F. E. (2009) Influence diagnostics for linear models with first-order autoregressive elliptical errors. *Statistics & Probability Letters*, 79(3), 339–346.
- Pinheiro, J., Bates, D. (2009) Mixed-Effects Models in S and S-PLUS, 1st edition 2000, 2nd printing 2009 edition, New York, Springer.
- Poltronieri Vargas, J., Vargas, J. P. (2009) Productos de Kronecker, *Revista de Matemática: Teoría y Aplicaciones*, 3(1), 45-60.
- Pourahmadi, M. (1999) Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation, *Biometrika*, 86 (3), 677-690.
- Qin, G., Zhu, Z. (2008) Robust estimation in partial linear mixed model for longitudinal data, *Acta Mathematica Scientia*, 28(2), 333-347.
- Qin, G., Zhu, Z., Fung, W. K. (2009) Robust estimation of covariance parameters in partial linear model for longitudinal data, *Journal of Statistical Planning and Inference*, 139(2), 558-570.
- Radhakrishna Rao, C. (1973) Linear Statistical Inference and Its Applications, 2nd ed., Wiley-Interscience.
- Radhakrishna Rao, C. (2002) Linear Statistical Inference and Its Applications, 2nd ed., Wiley-Interscience.
- Rao, C.R., (1948a). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Cambridge Philos. Soc.*, 44, 50–57.
- Rao, C.R., (1948b). Tests of significance in multivariate analysis. *Biometrika*, 35, 58–79.
- Rao, C. R., Mukerjee, R. (1997). Comparison of LR, Score, and Wald Tests in a Non-IID Setting, *Journal of Multivariate Analysis*, 60(1), 99-110.

- Rencher, A. C., Christensen, W. F. (2012) *Methods of Multivariate Analysis*, 3rd Edition. John Wiley & Sons, Inc.
- Rencher, A. C., Schaalje, G. B. (2002) *Linear models in statistics*, Wiley-Interscience.
- Riquelme, M., Leiva, V., Galea, M., Sanhueza, A. (2011) Influence diagnostics on the coefficient of variation of elliptically contoured distributions. *Journal of Applied Statistics*, 38(3), 513-532.
- Ritz, C., Streibig, J. C. (2008). *Nonlinear Regression with R*, Springer.
- Robinson, P. M. (1988) Root -N-Consistent Semiparametric Regression. *Econometrica*, 56(4), 931-54.
- Ruppert, D., Wand, M. P., Carroll, R. J. (2003) *Semiparametric Regression*, USA, Cambridge University Press.
- Ruppert, D., Wand, M. P., Carroll, R. J. (2009) Semiparametric regression during 2003-2007, *Electronic Journal of Statistics*, 3, 1193-1256.
- Savalli, C., Paula, G. A., Cysneiros, F. J. A. (2006) Assessment of variance components in elliptical linear mixed models, *Statistical Modelling*, 6(1), 59-76.
- Schwarz, G. (1978) Estimating the Dimension of a Model, *The Annals of Statistics*, 6(2), 461-464.
- Seber, G. A. F., Wild, C. J. (2003) *Nonlinear Regression*, John Wiley & Sons.
- Seber, G. A. F. (2004) *Multivariate Observations* (New edition), Wiley.
- Seber, George A. F. (2008) *A Matrix Handbook for Statisticians*, 1st ed., Wiley-Interscience.
- Shi, L., Chen, G. (2008) Case deletion diagnostics in multilevel models, *Journal of Multivariate Analysis*, 99(9), 1860-1877.
- Silverman, B. W. (1982) On the Estimation of a Probability Density Function by the Maximum Penalized Likelihood Method, *The Annals of Statistics*, 10(3), 795-810.

- Silverman, B. W. (1985) Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting, *Journal of the Royal Statistical Society, Series B (Methodological)*, 47(1), 1-52.
- Silvey, S. D. (1959) The Lagrangian Multiplier Test, *The Annals of Mathematical Statistics*, 30(2), 389-407.
- Sinde, J. B. (2002) El paradigma de la verosimilitud, *Estadística española*, 44(149), 113-127.
- Smyth, G. K. (2003) Pearson's goodness of fit statistic as a score test statistic, *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, 115-126. <http://projecteuclid.org/euclid.lnms/1215091138>
- Solaro, N. (2004) Random variate generation from Multivariate Exponential Power distribution, *Statistica & Applicazioni*, 2(2).
- Speckman, P. (1988) Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society*, 50(3), 413-436.
- Sperlich, S., Härdle, W. K., Aydinli, G. (2006) The Art of Semiparametrics. *Physica*.
- Stein, M. L. (1993) Spline Smoothing with an Estimated Order Parameter, *The Annals of Statistics*, 21(3), 1522-1544.
- Stram, D. O., Lee, J. W. (1994) Variance Components Testing in the Longitudinal Mixed Effects Model, *Biometrics*, 50(4), 1171-1177.
- Sun, R. B., Wei, B. C. (2004) On influence assessment for LAD regression, *Statistics & Probability Letters*, 67(2), 97-110.
- Takezawa, K. (2005) Introduction to Nonparametric Regression, 1st edition, Hoboken, N.J:Wiley-Interscience.
- Thomas, W., Cook, R. D. (1989) Assessing influence on regression coefficients in generalized linear-models, *Biometrika*, 76(4), 741-749.
- Tsai, C. L. (1986) Score Test for the First-Order Autoregressive Model with Heteroscedasticity, *Biometrika*, 73(2), 455-460.

- Tsai, M. T. M. (1993) UI Score Tests for Some Restricted Alternatives in Exponential Families, *Journal of Multivariate Analysis*, 45(2), 305-323.
- Tsiatis, A. (2006) Semiparametric Theory and Missing Data. **Springer**.
- Tsybakov, A. B. (2008) Introduction to Nonparametric Estimation, *Springer Science & Business Media*.
- Turrero Nogues, A. (1989a) Dos funcionales reales de la matriz de Fisher, convexidad y pérdida de información debida al agrupamiento de observaciones, *Estadística Española*, 30(119), 435-444.
- Turrero Nogues, A. (1989b) Nuevas medidas de información paramétricas reales basadas en la matriz de Fisher, *Trabajos de Estadística*, 4(2), 89-106.
- Vaart, A. W. van der. (2000) Asymptotic Statistics, **Cambridge University Press**.
- Verbeke, G., Molenberghs, G. (2009) Linear Mixed Models for Longitudinal Data, 2nd edition, 1st edition 2000, **Springer**.
- Wahba, G. (1983) Bayesian "Confidence Intervals" for the Cross-Validated Smoothing Spline, *Journal of the Royal Statistical Society, Series B (Methodological)*, 45(1), 133-150.
- Wahba, G. (1985) A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem, *The Annals of Statistics*, 13(4), 1378-1402.
- Wang, Q. H., Jing, B. Y. (2003) Empirical likelihood for partial linear models, *Annals of the Institute of Statistical Mathematics*, 55(3), 585-595.
- Wong, C. S. (1985) On the use of differentials in statistics, *Linear Algebra and its Applications*, 70, 285-299.
- Wood, S. N. (2006) Generalized additive models: an introduction with R, **Chapman & Hall/CRC**.
- Wu, Hulin, Zhang, Jin-Ting (2006) Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches. **Wiley-Interscience**.

- Xia, Y. (2009) A note on the backfitting estimation of additive models, *Bernoulli*, 15(4), 1148-1153.
- Xia, Y. (2009) A note on the backfitting estimation of additive models, *Bernoulli*, 15(4), 1148-1153.
- Yatchew, A. (2003) Semiparametric Regression for the Applied Econometrician. Cambridge University Press.
- Yi, Y., Wang, X. (2011) Comparison of Wald, Score, and Likelihood Ratio Tests for Response Adaptive Designs, *Journal of Statistical Theory and Applications*, 10(4), 553-569.
- Yu, K., Park, B. U., Mammen, E. (2008) Smooth backfitting in generalized additive models, *Annals of Statistics*, 36(1), 228-260.
- Zare, K., Rasekh, A. (2011) Diagnostic measures for linear mixed measurement error models, *Sort: Statistics and Operations Research Transactions*, 35(2), 125-144.
- Zeng, D., Lin, D. Y. (2007) Maximum likelihood estimation in semiparametric regression models with censored data, *Journal Royal Statistical Society B*, 69(4), 507-564.
- Zerbe, G.O. (1979) Randomization analysis of the completely randomized design extended to growth and response curves. *Journal of the American Statistical Association*, 74, 215-221.
- Zewotir, T., Galpin, J. S. (2005). Influence Diagnostics for Linear Mixed Models, *Journal of Data Science*, 3, 153-177.
- Zhang, D., Lin, X. H., Raz, J., Sowers, M. F. (1998) Semiparametric stochastic mixed models for longitudinal data, *Journal of the American Statistical Association*, 93(442), 710-719.
- Zhang, D., Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics* (Oxford, England), 4(1), 57-74.
- Zhu, H., Zhang, H. (2006) Generalized score test of homogeneity for mixed effects models, *The Annals of Statistics*, 34(3), 1545-1569.

- Zhu, H., Ibrahim, J. G., Cho, H. (2012) Perturbation and scaled cook's distance, *Annals of Statistics*, 40(2), 785-811.
- Zimmerman, D. L., Núñez-Antón, V. (2001) Modelización de datos longitudinales con estructuras de covarianza no estacionarias: modelos de coeficientes aleatorios frente a modelos alternativos, *Qüestiió: quaderns d'estadística i investigació operativa*, 25(2), 225-262.
- Zografos, K. (2000) Measures of multivariate dependence based on a distance between Fisher information matrices, *Journal of Statistical Planning and Inference*, 89(1-2), 91-107.