Facultad de Ciencias - Instituto de Estadística
Magíster en Estadística

Modelo mixto vector spline bajo distribuciones elípticas

Trabajo de titulación presentado por
Isaac Esteban Cortes Olmos

Para optar al grado de Ingeniero en Estadística
y al título de Magíster en Estadística

Profesor guía
Dr. Germán Ibacache Pulgar

Valparaíso, Chile

Julio, 2016

Índice general

Agradecimientos II

Resumen V

Índice de figuras VII

Índice de tablas VIII

1. Introducción **1**
- 1.1. Descripción de los objetivos 2
- 1.2. Organización del trabajo 3
- 1.3. Aspectos preliminares 3
  - 1.3.1. Distribución normal multivariada 7
  - 1.3.2. Distribución t-Student multivariada 8
  - 1.3.3. Distribución exponencial potencia 10

2. Modelo mixto vector spline bajo distribuciones elípticas **13**
- 2.1. Descripción del modelo 13
- 2.2. Representación matricial 15
- 2.3. Supuesto distribucional 16
- 2.4. Función de penalización 18

3. Estimación e Inferencia **23**
- 3.1. Proceso de estimación 23
  - 3.1.1. Discusión bibliográfica 24
  - 3.1.2. Función escore penalizada 25

3.1.3. Matriz de segundas derivadas 27
3.1.4. Matriz de información de Fisher penalizada 28
3.1.5. Maximizaciones sucesivas para obtener EMVP . . . 31
3.1.6. Estudio de la concavidad 32
3.1.7. Proceso iterativo 33
3.2. Inferencia estadística 35
3.2.1. Error estándar 35
3.2.2. Bandas de confianza 37
3.2.3. Efectos aleatorios 37
3.3. Selección de los parámetros de suavizamiento 38
3.3.1. Validación cruzada generalizada 39
3.3.2. Criterio de Akaike 40
3.4. Calidad del ajuste 40

4. Aplicación **43**
4.1. Descripción de los datos 43
4.2. Modelo no paramétrico para AFP Habitat 46
4.3. Modelo no paramétrico para AFP Planvital 49
4.4. Modelo vector spline 52

5. Conclusiones **57**
5.1. Trabajos futuros 58

Apéndices **59**

A. Derivadas de la log-verosimilitud penalizada **59**
A.1. Derivadas de primer orden 59
A.2. Derivadas de segundo orden 61

B. Datos de las AFP **64**

Referencias bibliográficas **67**

Agradecimientos

- En primer lugar, quiero agradecer a Dios por ayudarme en los momentos más díficiles de mi vida y por colocar las personas adecuadas en mi camino, antes y durante el proceso de titulación.
- Agradecer de manera especial a mi madre Elba, padre Jorge, hermanas Francheska y Sara, y hermano Jorge, por su apoyo incondicional en todas las etapas de mi vida.
- Quiero agradecer a mis tíos Héctor y Minda, y primo Elías por sus grandes consejos en todas las etapas de mi vida.
- También quiero agradecer a mis tíos Hector y Neveka, primos Yolanda, Héctor y Belén, por sus incentivos de partida y términos de estudios.
- A los profesores del Instituto de Estadística, por su formación, sugerencias y contribuciones importantes para que este trabajo pudiese culminar.
- Al profesor Germán Ibacache por su paciencia y consejos durante el transcurso de esta tesis.
- A la familia Vega Fuenzalida por su apoyo en este proceso de trabajo de tesis y personal.
- A los hermanos y hermanas de la Iglesia Cristo Latinoamericana por su apoyo y ayuda para culminar este proceso de titulación.
- A cada uno de mis compañeros y sus familias que me ayudaron en el proceso de mi formación universitaria.

Resumen

En este trabajo de tesis se estudian los modelos mixtos vector spline bajo distribuciones de contornos elípticos, que son una extensión del modelo propuesto por Ibacache et al. (2012). Este enfoque permite flexibilizar la curtosis de la distribución de los errores, posibilitando la elección de distribuciones con colas más pesada y livianas que la distribución normal. La función de verosimilitud penalizada es introducida para obtener los estimadores de máxima verosimilitud penalizada, que en el caso de las distribuciones con colas pesadas son más robustos contra observaciones aberrantes, en el sentido de la distancia de Mahalanobis. Finalmente, se presenta un ejemplo ilustrativo en que los ajustes están basados sobre errores normal y t-Student multivariados.

Índice de figuras

1.1. Gráficos de la distribución normal bivariada. 8
1.2. Gráficos de la distribución t-Student bivariada. 9
1.3. Gráficos de la distribución Cauchy bivariada. 10
1.4. Gráficos de la distribución exponencial potencia bivariada. 11
1.5. Gráficos de la distribución exponencial potencia bivariada. 11

4.1. Box-plot para las variables IPSA, AFP Habitat y Planvital. 44
4.2. Diagrama de dispersión entre las rentabilidades de la AFP Habitat y el IPSA (izquierda), y las rentabilidades de la AFP Habitat y el tiempo. 45
4.3. Diagrama de dispersión entre las rentabilidades de la AFP Planvital y el IPSA (izquierda), y las rentabilidades de la AFP Habitat y el tiempo. 45
4.4. Gráficos del ajuste no paramétrico a los datos de la AFP Habitat bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad, respectivamente. 47
4.5. Gráficos de la curva estimada y sus bandas de confianza bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad, respectivamente. 48
4.6. Gráficos normales de probabilidad para las distancias transformadas, bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad ajustados a los datos de la AFP Habitat. 48
4.7. Gráficos del ajuste no paramétrico a los datos de la AFP Habitat bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad, respectivamente. 50
4.8. Gráficos de la curva estimada y sus bandas de confianza bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad, respectivamente. 51
4.9. Gráficos normales de probabilidad para las distancias transformadas, bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad ajustados a los datos de la AFP Planvital. 51
4.10. Gráficos de los ajustes no paramétricos a los datos de las AFP Habitat y Planvital bajo el modelo normal multivariado, respectivamente. 53

4.11. Gráficos de los ajustes no paramétricos a los datos de las AFP Habitat y Planvital bajo el modelo t-Student multivariado con $\nu = 5$ grados de libertad, respectivamente. 54
4.12. Gráficos de la curva estimada y sus bandas de confianza en los datos de las AFP Habitat y Planvital bajo el modelo normal multivariado, respectivamente. 54
4.13. Gráficos de las bandas de confianza de la funciones no paramétricas a los datos AFP Habitat y Planvital, sobre el modelo t-Student con $\nu = 5$ grados de libertad. 55
4.14. Gráficos normales de probabilidad para las distancias transformadas, bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad ajustados a los datos de las AFP. 55

Índice de tablas

1.1. Distribuciones de contornos elípticos. 7

3.1. Cantidades υ_i y υ_i' para algunas distribuciones elípticas multivariadas . . . 27
3.2. Cantidades d_{g_i} y $\mathcal{F}_{g_i}$ para algunas distribuciones elípticas multivariadas . . 30

4.1. Estadísticas descriptivas para las variables IPSA, AFP Habitat y AFP Planvital. 44
4.2. Estimaciones de máxima verosimilitud penalizada, error estándar (e.e) y $L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$, para los modelos normal y t-Student con $\nu = 5$ grados de libertad ajustados a los datos de la AFP Habitat. 47
4.3. Estimaciones de máxima verosimilitud penalizada, error estándar (e.e) y $L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$, para los modelos normal y t-Student con $\nu = 5$ grados de libertad ajustados a los datos de la AFP Planvital. 50
4.4. Estimaciones de máxima verosimilitud penalizada, error estándar (e.e) y $L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$, para los modelos normal y t-Student multivariados con $\nu = 5$ grados de libertad ajustados a los datos de las AFP. 53

B.1. Rentabilidades mensuales de los fondos de pensiones de las AFP Habitat y Planvital, y el Índice de precios selectivos de acciones (IPSA) en el periodo comprendido entre Enero del 2009 a Junio del 2010. 64
B.2. Rentabilidades mensuales de los fondos de pensiones de las AFP Habitat y Planvital, y el Índice de precios selectivos de acciones (IPSA) en el periodo comprendido entre Julio del 2010 a Diciembre del 2012. 65
B.3. Rentabilidades mensuales de los fondos de pensiones de las AFP Habitat y Planvital, y el Índice de precios selectivos de acciones (IPSA) en el periodo comprendido entre Enero del 2013 a Diciembre del 2013. 66

Capítulo 1

Introducción

Los modelos estadísticos son ecuaciones matemáticas que reproducen los fenómenos que observamos de la forma más exacta posible, considerando los datos suministrados y la influencia que tiene el azar en estos datos. Estos modelos pueden ser utilizados para las siguientes acciones: describir (o explicar) los datos; y/o predecir futuras observaciones. En el caso que las observaciones predictivas están en desacuerdo con la predicciones, el modelo debe ser remplazado o refinado.

El análisis de regresión lineal normal es una de las técnicas estadísticas más utilizadas para estudiar y modelar relaciones lineales entre una variable de interés (variable de respuesta) y un conjunto de variables predictoras. Estas relaciones, se pueden identificar con claridad a través de diagramas de dispersión, donde da la impresión que los datos caen casi en línea recta. Es por esta razón, que es conveniente suponer que existe un error estadístico en la ecuación, es decir, una variable aleatoria (que sigue una distribución normal) que explica por qué el modelo no se ajusta exactamente a los datos.

Los modelos de regresión lineal normal se pueden aplicar a diferentes campos, entre ellos: economía, administración, ciencias sociales, ciencias físicas y químicas, entre otros. Sin embargo, existen situaciones en los campos de ingeniería y ciencias donde no es adecuado modelar asumiendo relaciones lineales. Esto nos conduce a utilizar modelos de regresión no lineales, es decir, la variable de respuesta y una variable predictora se relacionan a través de una función suave desconocida.

Hastie y Tibshirani (1990) presentan una extensa teoría sobre los modelos aditivos, que es una generalización de los modelos lineales y no lineales. La idea de esta clase de

modelos, es remplazar las relaciones lineales de la variables predictoras con la variable dependiente por, funciones suaves desconocidas que contribuyen de manera aditiva a la variable dependiente. Una clase particular de estos modelos son los modelos lineales semiparamétricos, que en su forma más simple, consideran relajar el supuesto de linealidad en una variable predictora y mantenerla en las otras variables predictoras. Para más detalles de estos modelos, ver por ejemplo, Green y Silverman (1994).

Otra herramienta poderosa en el modelamiento estadístico son los modelos lineales mixtos, que cuantifican la relación entre una variable dependiente continua y varias variables predictoras. El nombre se debe a que pueden incluir parámetros fijos y aleatorios asociados con uno o más covariables continuas o categóricas y efectos aleatorios asociados con uno o más factores aleatorios (Verbeke y Molenberghs, 2011). Estos se pueden utilizar en diferentes estructuras de datos, tales como: datos agrupados, medidas repetidas, datos longitudinales, entre otras. Otras ventajas que poseen son su flexibilidad para modelar estructuras de varianza y covarianza entre unidades muestrales y, capacidad de tratar con datos balanceados y desbalanceados (Ibacache-Pulgar, 2009).

Debido a la necesidad de representar fenómenos más complejos, se han construido modelos que llevan diversas características, entre ellas las mencionadas en los párrafos anteriores. Un ejemplo de esto, es el trabajo propuesto por Ibacache et al. (2012), donde extiende los modelos lineales mixtos semiparamétricos con errores normales a errores que pertenecen a la clase de distribuciones elípticas.

En este trabajo de tesis se propone una nueva clase de modelos denominados modelos mixtos vector spline bajo distribuciones elípticas, que es una extensión del modelo propuesto por Ibacache et al. (2012). Esta clase de modelos permiten explicar la dependencia entre unidades muestrales a través de efectos de las variables explicativas, que pueden contribuir de manera lineal y no lineal. La ventaja que otorgan los errores elípticos es que acomodan observaciones anómalas a través de distribuciones con colas más pesadas y livianas que la normal.

1.1. Descripción de los objetivos

El principal objetivo de este trabajo de tesis es estudiar los modelos mixtos vector spline bajo la clase de distribuciones elípticas, específicamente, el problema de estimación e

inferencia estadística. En este contexto, se han definido los siguientes objetivos específicos:

(i) Definir la función de verosimilitud penalizada para los modelos vector spline bajo distribuciones elípticas, calcular la función escore y matriz de información de Fisher penalizada para los parámetros involucrados en los modelos;

(ii) Proporcionar e implementar computacionalmente, un método para estimar los coeficientes de regresión, las funciones no paramétricas y los componentes de varianza del modelo; y

(iii) Aplicar los resultados a un conjunto de datos reales para ilustrar su aplicabilidad.

1.2. Organización del trabajo

Este trabajo de tesis se encuentra organizado en cuatro capítulos, cuya descripción se presenta a continuación. En el Capítulo 2 se describen las características fundamentales del modelo mixto vector spline bajo distribuciones elípticas. Además, se define la función log-verosimilitud penalizada. En el Capítulo 3 se obtienen la función escore y la matriz de información de Fisher basadas en la log-verosimilitud penalizada. A partir de estos resultados, se deriva un proceso iterativo para estimar los coeficientes de regresión, las funciones no paramétricas y los componentes de varianza. También, se describe un procedimiento para calcular el error estándar de los parámetros, se derivan bandas de confianza y se propone un método para estimar los efectos aleatorios. Al final del capítulo, se discute el problema de estimación de los parámetros de suavizamiento y calidad del ajuste del modelo. Finalmente, en el Capítulo 4, se ilustran los resultados de estimación e inferencia a través de una aplicación a un conjunto de datos reales.

1.3. Aspectos preliminares

Las distribuciones de contornos elípticos (también denominadas distribuciones elípticas) son una clase de distribuciones simétricas cuyos contornos de sus densidades tienen la misma forma elíptica de la normal, pero además, contienen distribuciones de colas más y menos pesadas que las de ésta. Algunos elementos particulares que pertenecen a esta clase de distribuciones son: distribución Cauchy, normal, exponencial potencia, normal contaminada, t-Student, logística, entre otras.

En el modelamiento estadístico, las distribuciones elípticas ofrecen una alternativa más viable a la otorgada por el modelamiento con errores aleatorios normales, debido a que sus colas son más realistas. Algunas aplicaciones que utilizan las distribuciones elípticas, se han realizado en diferentes campos de las ciencias, tales como: biología, genética, hidrología, física, economía, medicina, gestión de riesgos, psicología, entre otros. Un ejemplo es el modelo elíptico con efectos mixtos propuesto por Osorio (2006), que es aplicado a un conjunto de datos dentales.

Con el objetivo de introducir a los modelos mixtos vector spline bajo distribuciones elípticas, se presenta esta sección con algunas definiciones y propiedades útiles para el desarrollo de este trabajo de tesis. Un estudio detallado de las propiedades de las distribuciones elípticas (o algunos de sus elementos particulares) pueden ser encontrados, por ejemplo, en: Fang et al. (1990), Gupta y Varga (1993), Kotz y Nadaraya (2004) y Absanullah et al. (2014).

Definición 1 (Distribución de contornos elípticos)**.** *Se dice que un vector aleatorio* $\mathbf{y} \in \mathbb{R}^n$ $(n \geq 2)$ *sigue una distribución de contornos elípticos si su función característica tiene la siguiente forma*

$$\psi_{\mathbf{y}}(t) = exp\{i\ \mathbf{t}^T\ \boldsymbol{\mu}\}\varphi(\mathbf{t}^T\ \boldsymbol{\Sigma}\ \mathbf{t}), \tag{1.1}$$

donde $\boldsymbol{\mu} \in \mathbb{R}^n$ *denota el parámetro de posición,* $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ *denota el parámetro de escala (matriz simétrica y semidefinida positiva),* $\varphi : \mathbb{R}^n \to \mathbb{R}$ *es una función generadora de funciones características,* $i = \sqrt{-1}$ *y* $\mathbf{t} \in \mathbb{R}^n$*.*

Denotaremos $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ para indicar que $\mathbf{y}$ tiene distribución elíptica con la función característica dada por la expresión (1.1) y, además, que $\mathbf{y}$ pertenece a la clase de distribuciones elípticas multivariadas. En el caso particular que $\boldsymbol{\mu} = \mathbf{0}$ y $\boldsymbol{\Sigma} = \mathbf{I}_n$, la distribución de $\mathbf{y}$ es denominada esférica n-variada y la denotaremos por $\mathbf{y} \sim \mathcal{S}_n(\varphi)$. A continuación se presentan algunas de las propiedades principales que poseen las distribuciones de contornos elípticos.

Propiedad 1. *Sea* $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ *con* $\mathrm{rg}(\boldsymbol{\Sigma}) = \mathrm{r} < \mathrm{n}$*. En el caso que exista el valor esperado y varianza del vector aleatorio* $\mathbf{y}$*, entonces*

(i) $E(\mathbf{y}) = \boldsymbol{\mu}$*, y*

(ii) $Var(\mathbf{y}) = \kappa\, \boldsymbol{\Sigma}$, *donde* κ *es una constante positiva* $(\kappa > 0)$ *dada por*

$$\kappa = -2\varphi'(0) = -2\frac{d\varphi(u)}{du}\bigg|_{u=0}, \tag{1.2}$$

donde $\varphi(\cdot)$ *es la función generadora de funciones características. En el caso de la distribución t-Student,* $\kappa = \upsilon/(\upsilon - 2)$*, donde* υ *denota los grados de libertad.*

Los modelos mixtos vector spline se pueden caracterizar a través de los parámetros de escala y posición debido a la **Propiedad 1**. Además, se observa que la constante κ define la matriz varianza-covarianza del vector aleatorio, que depende directamente de la distribución que se tome en consideración.

Propiedad 2. *Suponga que* $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ *con* $rg(\boldsymbol{\Sigma}) = r < n$*. Si* $\mathbf{B}$ *es una matriz* $(n \times m)$ *y* $\boldsymbol{\delta}$ *es un vector* $(n \times 1)$*, entonces*

$$\boldsymbol{\delta} + \mathbf{B}^T\mathbf{y} \sim \mathcal{E}l_m(\delta + \mathbf{B}^T\boldsymbol{\mu}, \mathbf{B}^T\boldsymbol{\Sigma}\mathbf{B}, \varphi) \tag{1.3}$$

En un caso particular, consideramos la partición siguiente

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} y \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \tag{1.4}$$

obtenemos las siguientes distribuciones marginales

(i) $\mathbf{y}^{(1)} \sim \mathcal{E}l_m(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11}, \varphi)$ y

(ii) $\mathbf{y}^{(2)} \sim \mathcal{E}l_{(n-m)}(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22}, \varphi)$.

La **Propiedad 2** indica que la transformación lineal de un vector aleatorio con distribución elíptica, también sigue una distribución elíptica. Adicionalmente, se observa que cada elemento del vector aleatorio $\mathbf{y}$ también tiene una distribución elíptica marginal.

Propiedad 3. *Asumiendo que* $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$ *con* $\boldsymbol{\Sigma} \geq 0$*. Si consideramos una partición de* (1.4), tenemos lo siguiente

$$(\mathbf{y}^{(1)}|\mathbf{y}_0^{(2)}) \sim \mathcal{E}l_m\left(\boldsymbol{\mu}_{1.2}, \boldsymbol{\Sigma}_{11.2}, \varphi_{q(\mathbf{y}_0^{(2)})}\right) \tag{1.5}$$

en que

$$\begin{aligned}\boldsymbol{\mu}_{1.2} &= \boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_0^{(2)} - \boldsymbol{\mu}^{(2)}),\\ \boldsymbol{\Sigma}_{11.2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \text{ y}\\ q(\mathbf{y}_0^{(2)}) &= (\mathbf{y}_0^{(2)} - \boldsymbol{\mu}^{(2)})\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_0^{(2)} - \boldsymbol{\mu}^{(2)}).\end{aligned}$$

Análogamente,

$$(\mathbf{y}^{(2)}|\mathbf{y}_0^{(1)}) \sim \mathcal{E}l_m\left(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}, \varphi_{q(\mathbf{y}_0^{(1)})}\right), \tag{1.6}$$

en que

$$\begin{aligned}\boldsymbol{\mu}_{2.1} &= \boldsymbol{\mu}^{(2)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_0^{(1)} - \boldsymbol{\mu}^{(1)}),\\ \boldsymbol{\Sigma}_{22.1} &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \text{ y}\\ q(\mathbf{y}_0^{(1)}) &= (\mathbf{y}_0^{(1)} - \boldsymbol{\mu}^{(1)})\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_0^{(1)} - \boldsymbol{\mu}^{(1)}).\end{aligned}$$

La **Propiedad 3** quiere decir que las distribuciones condicionales del vector aleatorio $\mathbf{y}$ dados los valores de los subvectores de $\mathbf{y}$, son también elípticos. En este trabajo de tesis se consideran las distribuciones de contornos elípticos, cuyo $\text{rg}\,(\boldsymbol{\Sigma}) = n$ (caso no singular), esto quiere decir que existe una función de densidad en relación a la medida de Lebesgue sobre todo el espacio $\mathbb{R}^n$.

Definición 2 (Función densidad)**.** *Asumiendo que* $\text{rg}(\boldsymbol{\Sigma}) = n$*, tenemos que una función densidad de un vector aleatorio* $\mathbf{y}$ *asume la siguiente forma*

$$f_{\mathbf{y}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}}\, g(u), \tag{1.7}$$

en que $u = (\mathbf{y} - \boldsymbol{\mu})^T\ \boldsymbol{\Sigma}^{-1}\ (\mathbf{y} - \boldsymbol{\mu})$ *y* g *es una función escalar continua y diferenciable de* $\mathbb{R} \to [0, \infty]$*, llamada función generadora de densidades, que satisface la condición siguiente*

$$\int_0^\infty u^{-\frac{1}{2}}\, g(u)\, du < \infty. \tag{1.8}$$

Se denotará por $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ o simplemente $\mathbf{y} \sim \mathcal{E}l_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, en el caso que $\mathbf{y}$ tenga una distribución elíptica con densidad dada por la expresión (1.7). En la Tabla 1.1

se presentan algunas distribuciones que pertenecen a la clase de distribuciones elípticas, donde c es llamada constante de normalización.

Tabla 1.1: Distribuciones de contornos elípticos.

Distribución	$g(u)$	**Restricción**
Normal	$c\ \exp\{-\frac{u}{2}\}$	$u \geq 0$
Exponencial Potencia	$c\ \exp\{-\frac{u^s}{2}\}$	
t-Student	$c\ \{1+\frac{u}{s}\}^{-\frac{(v+m)}{2}}$	$m > 0$
Cauchy	$c\ \{1+\frac{u}{s}\}^{-\frac{v+1}{2}}$	$s > 0$
Tipo Kotz	$c\ u^{N-1}\ \exp\{-ru^s\}$	$r, s > 0$ y 2N + n > 2

Las distribuciones normal, t-Student y exponencial potencia son tres de las distribuciones elípticas más importantes en estadística y otros campos de las ciencias. Es por este motivo que se presentará una breve descripción de éstas distribuciones.

1.3.1. Distribución normal multivariada

Si un vector aleatorio $\mathbf{y} = (\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_n)^\top$ sigue una distribución normal multivariada, denotado por $\mathbf{y} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, la función densidad de probabilidad del vector está dada por:

$$g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(\sqrt{2\pi})^n} \exp\left[(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right], \quad \mathbf{y} \in \mathbb{R}^n, \tag{1.9}$$

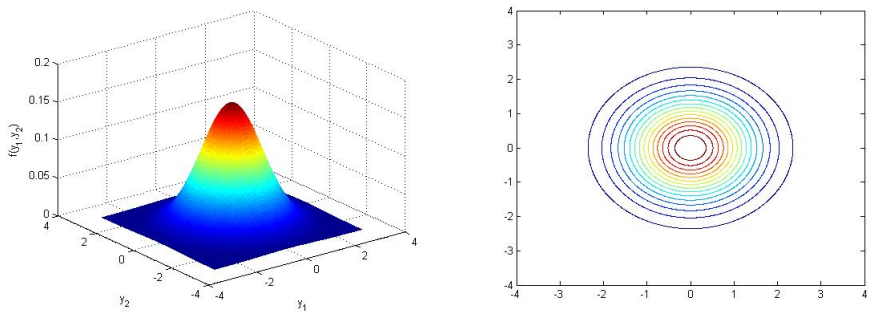
donde μ es un vector $(n \times 1)$, $\boldsymbol{\Sigma}$ es una matriz simétrica definida positiva $(n \times n)$ y $n > 1$ es el número de variables. El vector de medias y matriz de varianza-covarianza de $\mathbf{y}$ son dados, respectivamente, por:

$$\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu} \qquad \mathrm{y} \qquad \mathrm{Var}(\mathbf{y}) = \boldsymbol{\Sigma}.$$

La distribución normal multivariada es una de las distribuciones continuas más importantes, debido a que posee características bastante útiles (Absanullah, 2014), entre ellas se destacan las siguientes: los gráficos plots bivariados de datos multivariados muestran ten-

dencia lineal; la distribución puede ser completamente descrita utilizando sólo las medias, varianzas y covarianzas; si las variables no son correlacionadas, ellas son independientes; la forma conveniente de la función densidad se presta para la derivación de muchas propiedades y estadísticas de prueba; y cuando los datos no siguen una distribución normal multivariada, ésta puede ser útil como una aproximación, especialmente en inferencias que implican vectores de medias muestrales, que son aproximadamente normal por el teorema del límite central.

En la Figura 1.1, se presenta el gráfico de densidad normal bivariada ($n = 2$) con sus correspondientes contornos, donde $\boldsymbol{\mu} = \mathbf{0}$ y $\boldsymbol{\Sigma} = \mathbf{I}_{2\times 2}$.



(a) Densidad conjunta (b) Contornos de la densidad

Figura 1.1: Gráficos de la distribución normal bivariada.

1.3.2. Distribución t-Student multivariada

Si un vector aleatorio $\mathbf{y} = (\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_n)^\top$ sigue una distribución t-Student multivariada con υ grados de libertad, denotado por $\mathbf{y} \sim \mathrm{t}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \upsilon)$, la función densidad de probabilidad del vector está dada por

$$g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \upsilon) = \frac{\Gamma(\frac{\upsilon+n}{2})\ |\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{\Gamma(\frac{\upsilon}{2})\ (\upsilon\pi)^{\frac{n}{2}}\ \left[1 + \frac{1}{\upsilon}(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]}, \quad \mathbf{y} \in \mathbb{R}^n, \tag{1.10}$$

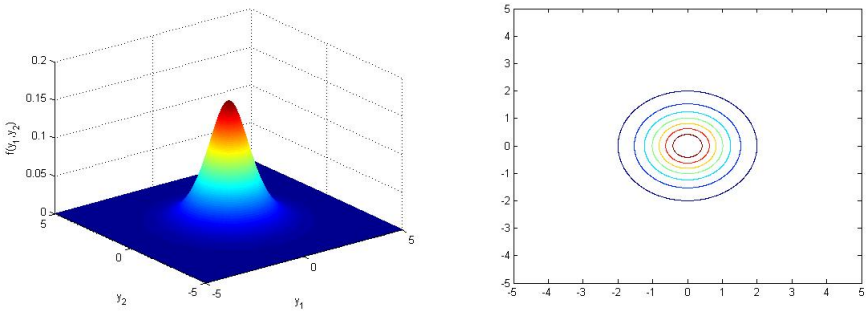
donde μ es un vector $(n \times 1)$, $\boldsymbol{\Sigma}$ es una matriz simétrica definida positiva $(n \times n)$, υ son los grados de libertad y $n > 1$ es el número de variables. El vector de medias y matriz de

varianza-covarianza de $\mathbf{y}$ son dados, respectivamente, por:

$$\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu} \qquad \text{y} \qquad \mathrm{Var}(\mathbf{y}) = \frac{\upsilon}{(\upsilon - 2)}\,\boldsymbol{\Sigma} \quad (\upsilon > 2).$$

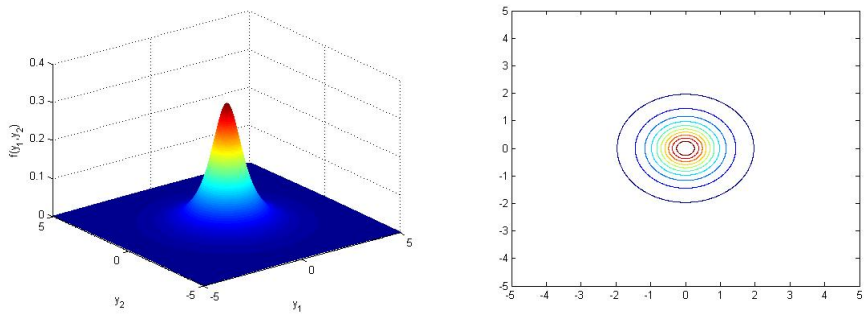
La distribución t-Student multivariada es otra de las distribuciones continuas importantes, ya que posee características bastante útiles, entre ellas se destacan las siguientes: la curtosis es controlada por sus grados de libertad, es decir, υ; tiene la propiedad de presentar colas más pesadas que la distribución normal multivariada; a medida que aumentan los grados de libertad, se aproxima a la distribución normal multivariada; cuando $\upsilon = 1$, la distribución t-Student multivariada se reduce a la distribución Cauchy multivariada; y una vez que la distribución t-Student es simétrica, su media y mediana son iguales. Para más detalles sobre esta distribución, ver por ejemplo, Borssoi (2014) y Lange et al. (1989).

En las Figuras 1.2-1.3 se presentan los gráficos de densidad t-Student y Cauchy bivariadas con sus correspondientes contornos, donde $\boldsymbol{\mu} = \mathbf{0}$ y $\boldsymbol{\Sigma} = \mathbf{I}_{2\times 2}$ para ambas distribuciones, y $\upsilon = 4$ para el caso de la t-student bivariada.



(a) Densidad conjunta (b) Contornos de la densidad

Figura 1.2: Gráficos de la distribución t-Student bivariada.



(a) Densidad conjunta (b) Contornos de la densidad

Figura 1.3: Gráficos de la distribución Cauchy bivariada.

1.3.3. Distribución exponencial potencia

Si un vector aleatorio $\mathbf{y} = (\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_n)^\top$ sigue una distribución exponencial potencia, denotado por $\mathbf{y} \sim \mathrm{PE}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathrm{s})$, la función densidad de probabilidad del vector está dada por:

$$g(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathrm{s}) = \frac{n\,\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{n}{2}}\,\sqrt{|\boldsymbol{\Sigma}|}\,\Gamma\left(1+\frac{n}{2\,\mathrm{s}}\right)\,2^{1+\frac{n}{2\,\mathrm{s}}}} \exp\left\{-\frac{1}{2}\,\left[(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]^{\mathrm{s}}\right\}, \qquad (1.11)$$

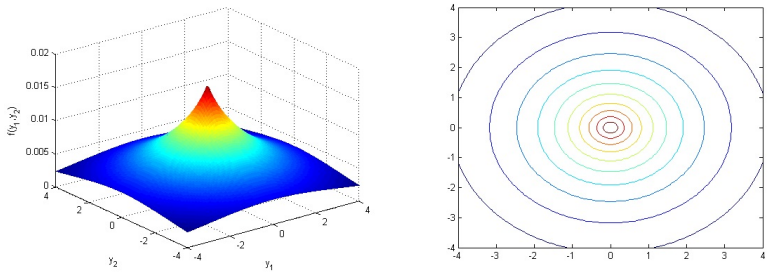
donde $\mathbf{y} \in \mathbb{R}^n$, μ es un vector $(n \times 1)$, $\boldsymbol{\Sigma}$ es una matriz simétrica definida positiva $(n \times n)$, s es el parámetro de curtosis y $n > 1$ es el número de variables. El vector de medias y matriz de varianza-covarianza de $\mathbf{y}$ son dados, respectivamente, por:

$$\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu} \qquad \text{y} \qquad \mathrm{Var}(\mathbf{y}) = \frac{2^{\frac{1}{\mathrm{s}}}\,\Gamma\left(\frac{n+2}{2\,\mathrm{s}}\right)}{n\,\Gamma\left(\frac{n}{2\,\mathrm{s}}\right)}\boldsymbol{\Sigma}.$$

La distribución exponencial potencia multivariada es un miembro de la familia de contornos elípticos multivariados, que ofrece una generalización útil de la distribución normal multivariada. Además, posee otras características interesantes, entre las cuales se destacan las siguientes: abarca las distribuciones de colas livianas y pesadas, por lo que puede ser comprobado para robustez; el parámetro de localización (media) en los modelos lineales y no lineales puede ser fácilmente construido; y la curtosis es controlada por el parámetro s.

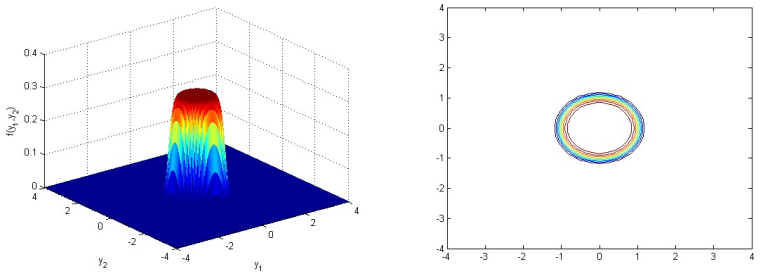
La distribución dada en (1.11) tiene algunos casos particulares, por ejemplo: cuando s = 1, se obtiene la distribución normal multivariada; cuando s = $\frac{1}{2}$, se obtiene la distribución Laplace multivariada; cuando $s \longrightarrow \infty$, se obtiene la distribución uniforme multivariada; y si $s < 1$, la distribución posee colas más pesadas que la normal, proporcionando robustez contra outliers. Para más detalles sobre esta distribución, ver por ejemplo, Gómez et al. (1998) y Lindsey (1999).

En las Figuras 1.4-1.5 se presentan los gráficos de densidad exponencial potencia bivariado ($n = 2$) con sus respectivos contornos, para los casos s = 0, 4 y s = 5, respectivamente. En ambos casos se consideran $\boldsymbol{\mu} = \mathbf{0}$ y $\boldsymbol{\Sigma} = \mathbf{I}_{2\times 2}$.



(a) Densidad conjunta (b) Contornos de la densidad

Figura 1.4: Gráficos de la distribución exponencial potencia bivariada.



(a) Densidad conjunta (b) Contornos de la densidad

Figura 1.5: Gráficos de la distribución exponencial potencia bivariada.

Capítulo 2

Modelo mixto vector spline bajo distribuciones elípticas

En este capítulo se estudian los modelos mixtos vector spline (MMVS) bajo distribuciones elípticas. En la primera sección, se realiza una breve discusión sobre algunos trabajos relacionados con esta clase de modelos, se describe el modelo mixto vector spline y mencionan algunos casos particulares de éste. Luego, en la segunda sección, se presenta el modelo mixto vector spline en su versión matricial y compacta. Posteriormente, en la tercera sección, se define el enfoque para introducir las distribuciones de contornos elípticos en los MMVS, y se mencionan algunas de sus ventajas. Finalmente, en la cuarta sección, se muestra el criterio para obtener los estimadores de los parámetros involucrados en los MMVS.

2.1. Descripción del modelo

Por muchos años, los modelos lineales mixtos normales han sido utilizados como herramientas para analizar conjuntos de datos de medidas repetidas, ver por ejemplo, Laird y Ware (1982). Sin embargo, la distribución normal ha sido criticada por proporcionar estimadores de máxima verosimilitud sensibles a observaciones aberrantes, por lo que diversos autores sugieren el uso de las distribuciones elípticas para acomodar tales observaciones potencialmente influyentes. Entre esos autores, se encuentra Savalli et al. (2006) que proponen una clase de modelos lineales mixtos elípticos, donde la distribución del vector de respuestas también sigue una distribución elíptica. Esta clase de modelos permite

desarrollar procedimientos de estimación, métodos de diagnóstico y test de componentes de varianza, flexibilizando la curtosis de la distribución de los errores; en el caso que las distribuciones de los errores tengan colas más pesadas que la normal, los estimadores de máxima verosimilitud son más robustos contra observaciones aberrantes, en el sentido de la distancia de Mahalanobis. Posteriormente, Osorio et al. (2007) derivan curvaturas normales de influencia local para el modelo lineal con efectos mixtos sobre diferentes esquemas de perturbación, y examinan su conexión con la matriz de alavancas generalizadas.

Recientemente, Ibacache et al. (2012) proponen la clase de modelos mixtos semiparamétricos elípticos, que consisten en agregar una componente no paramétrica a los efectos fijos, generalizando la clase de modelos propuesto por Savalli et al. (2006). La propuesta de este trabajo es extender el modelo propuesto por Ibacache et al. (2012), en el sentido de incluir tantas funciones no paramétricas a los efectos fijos, como unidades de tiempo que contribuyen en la media del perfil del individuo o clúster (en el contexto longitudinal).

La formulación matemática del modelo propuesto en este trabajo de tesis es la siguiente:

$$\mathrm{y}_{ij} = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + f_j(\mathrm{t}_i) + \mathbf{z}_{ij}^{\top}\mathbf{b}_i + \epsilon_{ij}, \quad i = 1, \ldots, n; j = 1, \ldots, m, \tag{2.1}$$

donde y_{ij} denota la j-ésima medida asociada con el i-ésimo clúster en el tiempo t_i, $\mathbf{x}_{ij}$ y $\mathbf{z}_{ij}$ son, respectivamente, vectores $(p \times 1)$ y $(q \times 1)$ de valores de variables explicativas, $\boldsymbol{\beta}$ es el vector $(p \times 1)$ de parámetros fijos y $\mathbf{b}_i$ denota el vector $(q \times 1)$ de efectos aleatorios, $\{f_j(\cdot)\}$ son funciones suaves arbitrarias que dependen de la variable t y ϵ_{ij} son los errores aleatorios. El modelo (2.1) es una extensión de otros modelos semiparamétricos propuestos en la literatura, por ejemplo:

(i) cuando las funciones no paramétricas no están presentes en el modelo, el MMVS se reduce al modelo mixto lineal clásico propuesto por Laird y Ware (1982), cuya aplicación ha sido ilustrada por ejemplo, en estudios epidemiológicos de efectos de la contaminación del aire sobre la salud;

(ii) cuando los efectos aleatorios no están presentes y $f_j(\cdot) = f(\cdot)$ para todo $j = 1, \ldots, m$, el MMVS se reduce al modelo lineal parcial estudiado por Heckman (1986) y Speckman (1988) en el caso univariado y He et al. (2002) en el contexto de datos longitudinales;

(iii) cuando $\boldsymbol{\beta} = \mathbf{0}$ y $f_j(\cdot) = f(\cdot)$ para todo $j = 1, \ldots, m$, el MMVS se reduce al modelo

mixto no paramétrico desarrollado por Wang (1998), quién utiliza suavizamiento spline para modelar los efectos fijos asumiendo que la función de covarianza depende sobre un número pequeño de parámetros; y,

(iv) cuando $f_j(\cdot) = f(\cdot)$ para todo $j = 1, \ldots, m$, el MMVS se reduce al modelo mixto semiparamétrico estudiado por Ibacache-Pulgar et al. (2012), en el cual extienden los modelos mixtos semiparamétricos con errores aleatorios normales a la clase de distribuciones elípticas.

En la literatura estadística, por el momento, existe un número pequeño de trabajos relacionados con los MMVS. Sin embargo, los trabajos existentes han sido de gran relevancia. Por ejemplo, Yee y Wild (1996) extienden la clase de modelos aditivos generalizados de manera natural a una clase de modelos de regresión multivariados, denominados "modelos aditivos vector generalizado (MAVG)". Esta clase de modelos abarcan muchas distribuciones y modelos estadísticos, entre los que se encuentra la familia exponencial clásica como subconjunto de esta clase. Luego, Yee y Mackenzie (2002) describen e ilustran los MAVG, con ejemplos que son de particular relevancia para ecologistas de plantas. Posteriormente, Yee y Stephenson (2007) llevan la clase de MAVG al análisis de datos de valores extremos.

2.2. Representación matricial

El modelo (2.1) puede ser expresado matricialmente de la siguiente manera:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \sum_{j=1}^{m} \mathbf{N}_i^{(j)}\mathbf{f}_{(j)} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \tag{2.2}$$

donde $\mathbf{y}_i$ es un vector aleatorio $(m \times 1)$ de respuestas observadas para el i-ésimo clúster, $\mathbf{X}_i$ es una matriz diseño $(m \times p)$ con filas $\mathbf{x}_{ij}^\top$, $\mathbf{N}_i^{(j)}$ es una matriz $(m \times n)$ de ceros con un 1 en la (j, i)-ésima posición, para $j = 1, \ldots, m$ y $i = 1, \ldots, n$, $\mathbf{f}_{(j)} = (f_j(\mathrm{t}_1), \ldots, f_j(\mathrm{t}_n))^\top$, $\mathbf{Z}_i$ es la matriz de diseño $(m \times q)$ de efectos aleatorios con filas $\mathbf{z}_{ij}^\top$, y $\boldsymbol{\epsilon}_i$ es un vector $(m \times 1)$ de errores aleatorios dentro de los clústers.

Para obtener una forma matricial más ventajosa computacionalmente, se adoptará una notación similar a la propuesta por Ibacache et al. (2012), para expresar el modelo (2.2) de la siguiente forma:

$$\mathbf{y}_i = \boldsymbol{\mu}_i + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \ \ , \tag{2.3}$$

donde $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} + \sum_{j=1}^{m} \mathbf{N}_i^{(j)}\mathbf{f}_{(j)}$, para $i = 1, \ldots, n$. Esta notación será bastante utilizada en la siguiente sección, en la que se discuten algunos aspectos teóricos sobre el modelo mixto vector spline bajo distribuciones elípticas, fundamentalmente sobre el supuesto distribucional.

Por otro lado, el modelo (2.2) también se puede escribir de una manera similar a Ibacache (2009), es decir, compactamente como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \tag{2.4}$$

donde $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_n^\top)^\top$, la i-ésima fila de $\mathbf{N}$ es $(\mathbf{N}_i^{(1)}, \ldots, \mathbf{N}_i^{(m)})$ para $i = 1, \ldots, n$, $\mathbf{f} = (\mathbf{f}_{(1)}^\top, \ldots, \mathbf{f}_{(m)}^\top)^\top$, $\mathbf{Z} = \text{diag}\{\mathbf{Z}_1, \ldots, \mathbf{Z}_n\}$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$ y $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^\top, \ldots, \boldsymbol{\epsilon}_n^\top)^\top$.

2.3. Supuesto distribucional

En los modelos mixtos vector spline bajo distribuciones elípticas se pueden adoptar dos formulaciones. La primera corresponde a una formulación jerárquica asumiendo que:

$$\begin{aligned} \mathbf{y}_i|\mathbf{b}_i &\sim \mathcal{E}l_m(\boldsymbol{\mu}_i + \mathbf{Z}_i\mathbf{b}_i\,,\, \mathbf{V}_i)\,, \\ \mathbf{b}_i &\sim \mathcal{E}l_q(\mathbf{0}\,,\, \mathbf{D}) \quad \text{y} \\ \boldsymbol{\epsilon}_i &\sim \mathcal{E}l_m(\mathbf{0}\,,\, \mathbf{V}_i)\,. \end{aligned}$$

En este caso, la distribución conjunta de $(\mathbf{y}_i^\top, \mathbf{b}_i^\top)^\top$ no necesariamente sigue una distribución elíptica y puede provocar dificultad para obtener la distribución marginal del vector de respuestas $\mathbf{y}_i$. La formulación que se adoptará en este trabajo de tesis considera la

siguiente distribución conjunta:

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{bmatrix} \sim \mathcal{E}l_{m+q}\left(\begin{bmatrix} \boldsymbol{\mu}_i \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_i & \mathbf{Z}_i\mathbf{D} \\ \mathbf{D}\mathbf{Z}_i^\top & \mathbf{D} \end{bmatrix} \right), \tag{2.5}$$

donde $\mathbf{b}_i$ y $\boldsymbol{\epsilon}_i$ no son correlacionadas, pero no necesariamente independientes, a excepción de la distribución normal. Además, las matrices $\boldsymbol{\Sigma}_i$, $\mathbf{D}$, y $\mathbf{Z}_i\mathbf{D}$ son proporcionales a las matrices de covarianza $\mathbf{y}_i$, $\mathbf{b}_i$ y $[\mathbf{y}_i, \mathbf{b}_i]$, respectivamente, por una cantidad $\xi_i > 0$ que puede ser obtenida a partir de las derivadas de las correspondientes funciones características.

Así, el modelo marginal tiene la siguente forma cerrada:

$$\mathbf{y}_i \sim \mathcal{E}l_m(\boldsymbol{\mu}_i\,, \boldsymbol{\Sigma}_i) \qquad i = 1, \ldots, n, \tag{2.6}$$

cuya función densidad de $\mathbf{y}_i$ está dada por:

$$f(\mathbf{y}_i) = |\boldsymbol{\Sigma}_i|^{-1/2}\, g(\delta_i)\,, \qquad i = 1, \ldots, n, \tag{2.7}$$

donde $\delta_i = \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i$ es la distancia de Mahalanobis, $\mathbf{r}_i = \mathbf{y}_i - \boldsymbol{\mu}_i$, y $g(\cdot)$ es la función generadora de densidad mencionada en el Capítulo 1. Para garantizar que el vector de respuestas observadas admita una densidad para todo $\mathbf{y}_i \in \mathbb{R}^m$, con respecto a la medida de Lebesgue, se asumirá que $\boldsymbol{\Sigma}_i$ es definida positiva para todo $i = 1, \ldots, n$.

Algunas de las ventajas que se obtienen al trabajar con la distribución marginal (Borssoi, 2014) son las siguientes:

(i) no se requieren métodos de integración numérica para obtener el modelo marginal;

(ii) se conserva la estructura de medias y varianza-covarianza de las variables $\mathbf{y}_i$, es decir, $\mathrm{E}(\mathbf{y}_i) = \boldsymbol{\mu}_i$ y $\mathrm{Var}(\mathbf{y}_i) = \xi_i\boldsymbol{\Sigma}_i$ se mantienen;

(iii) permiten derivar las funciones escore y las matrices Hessiana y de información de Fisher en forma cerrada, facilitando la aplicación de métodos de estimación, inferencia y diagnóstico de manera similar al caso normal; y

(iv) otorgan la posibilidad de predicción de las variables $\mathbf{b}_i$ a través del método bayesiano empírico.

2.4. Función de penalización

El método de máxima verosimilitud es la técnica más conocida y más utilizada para estimar los parámetros en un modelo estadístico. Este método consiste en obtener estimadores de los parámetros que maximicen la función de verosimilitud, requiriendo adquirir esta función como una función de los parámetros de un determinado modelo estadístico, basado en suposiciones acerca de las distribuciones de probabilidad.

En este trabajo, se asumirá que las matrices de escala $\mathbf{V}_i = \phi\mathbf{I}_m$ y $\mathbf{D} = \mathbf{D}(\boldsymbol{\lambda})$ es una matriz parametrizada por el vector $\boldsymbol{\lambda}$. Por lo tanto, el vector de parámetros que requieren ser estimados en el modelo mixto vector spline bajo distribuciones elípticas será $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \mathbf{f}_{(1)}^\top, \ldots, \mathbf{f}_{(m)}^\top, \boldsymbol{\tau}^\top)^\top$, cuyo espacio paramétrico está dado por:

$$\boldsymbol{\Theta} = \{\boldsymbol{\theta} \in \mathbb{R}^{p*} \quad | \boldsymbol{\beta} \in \boldsymbol{\Theta}_{\boldsymbol{\beta}}, \mathbf{f}_{(1)} \in \boldsymbol{\Theta}_{\mathbf{f}_{(1)}}, \ldots, \mathbf{f}_{(m)} \in \boldsymbol{\Theta}_{\mathbf{f}_{(m)}}, \boldsymbol{\tau} \in \boldsymbol{\Theta}_{\boldsymbol{\tau}}\}, \tag{2.8}$$

donde $\boldsymbol{\Theta}_{\boldsymbol{\beta}} \subseteq \mathbb{R}^p$, $\boldsymbol{\Theta}_{\mathbf{f}_{(j)}} \subseteq \mathbb{R}^n \quad (\forall\ j = 1, \ldots, m)$, $\boldsymbol{\Theta}_{\boldsymbol{\tau}} \subseteq \mathbb{R}^{d+1}$, $\boldsymbol{\tau} = (\tau_0, \tau_1, \ldots, \tau_d)^\top$ para $\tau_0 = \phi$ y $\tau_\ell = \lambda_\ell$ $(\ell = 1, \ldots, d)$, y $p^* = \text{tr}\{\mathbf{H}(\boldsymbol{\alpha})\} + d + 1$, donde $\mathbf{H}(\boldsymbol{\alpha})$ es definida en el Capítulo 3. Además, el logaritmo de la función de verosimilitud se puede expresar de la siguiente forma:

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} L_i(\boldsymbol{\theta}), \tag{2.9}$$

en que:

$$L_i(\boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{\Sigma}_i| + \log\{g(\delta_i)\}. \tag{2.10}$$

En el contexto paramétrico, los estimadores de máxima verosimilitud son un buen estimador puntual, ya que poseen algunas propiedades óptimas. Éstas son discutidas, por ejemplo, en Gourieroux y Monford (1995) y Casella y Berger (2002). Sin embargo, en el contexto semiparamétrico, si maximizamos de forma directa (2.10) sin imponer restricciones a las funciones $\mathbf{f}_{(j)}$, puede causar sobre ajuste y problemas de identificación de los parámetros $\boldsymbol{\beta}$ (Green, 1987).

Una manera de resolver este problema es aplicar un procedimiento basado en la penalización de la función de log-verosimilitud, es decir, incorporar una función de penalización

al logaritmo de la función de verosimilitud por cada función $\mathbf{f}_{(j)}$; específicamente,

$$L_p(\boldsymbol{\theta}, \alpha_1, \ldots, \alpha_m) = L(\boldsymbol{\theta}) + \sum_{j=1}^{m} \alpha_j^* \mathrm{J}(f_j), \tag{2.11}$$

donde $\mathrm{J}(f_j)$ es una función de penalización cuya elección depende de alguna aplicación específica o conocimiento a priori, y $\alpha_j^* = \alpha^*(\alpha_j)$ es un término que depende del parámetro de suavizamiento $\alpha_j > 0, \quad \forall\, j = 1, \ldots, m$. Este último, controla el equilibrio entre la bondad de ajuste, medida por valores grandes de $L(\boldsymbol{\theta})$, y la función estimada de suavizamiento, medida por pequeños valores de $\mathrm{J}(f_j)$.

En este trabajo de tesis, se propone utilizar el siguiente término de penalización:

$$\mathrm{J}(f_j) = \alpha_j^* \int_{a_i}^{b_i} [f_j^{(l)}(t_i)]^2 \mathrm{dt}_i, \tag{2.12}$$

en que $f_j^l = \frac{d^l}{\mathrm{dt}_i}$, $\mathrm{t}_i \in [a_i, b_i]$, y la función f_j pertenece al espacio de funciones de Sobolev definida por:

$$\mathrm{W}_2^{(l)}[a_i, b_i] = \left\{ f_j : f_j^{(l)} \in L^2[a_i, b_i], f_j^{(1)}, f_j^{(2)}, \ldots, f_j^{(l-1)}, abs.cont. \right\}, \tag{2.13}$$

donde $L^2[a_i, b_i] = \left\{ f : \int_{a_i}^{b_i} f^2(\mathrm{t}_i) < \infty \right\}$.

En el caso que $l = 2$, el estimador de la función corresponde a un spline cúbico natural (SCN) con nodos en los puntos t_i $(i = 1, \ldots, n)$. Según Reinsch (1967) y Silverman (1985), la curva $\widehat{f_j}$ tiene las siguientes propiedades:

(i) es un polinomio cúbico en cada intervalo $(\mathrm{t}_i, \mathrm{t}_{i+1})$;

(ii) la curva y sus primeras dos derivadas son continuas en los puntos t_i pero puede haber discontinuidad en la tercera derivada;

(iii) en cada uno de los rangos $(-\infty, \mathrm{t}_1]$ y $[\mathrm{t}_n, \infty)$, la segunda derivada es cero, de modo que $\widehat{f_j}$ es lineal fuera de los dominios dados.

De acuerdo con Green y Silverman (1994) podemos expresar el término de penalización (2.12) para SCN de la siguiente forma:

$$\int_{a_i}^{b_i} [f_j^{(2)}(\mathrm{t}_i)]\mathrm{d}\mathrm{t}_i = \mathbf{f}_{(j)}^{\top}\mathbf{K}_j\mathbf{f}_{(j)}, \tag{2.14}$$

donde $f_j^{(2)}(\mathrm{t}_i) = \frac{\mathrm{d}^2}{\mathrm{t}_i^2}f(\mathrm{t}_i)$ y $\mathbf{K}_j$ es un matriz de suavización $(n \times n)$ definida positiva para la j-ésima variable explicativa, que depende sólo de los nodos, con estructura dada por $\mathbf{K}_j = \mathbf{Q}_j\mathbf{R}_j^{-1}\mathbf{Q}_j^{\top}$, para $j = 1, \ldots, m$, donde $\mathbf{Q}_j$ es una matriz tridiagonal $n \times (n-2)$ y $\mathbf{R}_j$ es una matriz simétrica tridiagonal $(n-2)\times(n-2)$ construidas de la siguiente manera:

considerando la distancia entre los nodos consecutivos i e $i+1$ por $h_i = \mathrm{t}_{i+1} - \mathrm{t}_i$, los elementos q_{il} de $\mathbf{Q}_j$ son dados por:

$$\begin{aligned} q_{(l-1),l} &= h_{l-1}^{-1} \\ q_{l,l} &= -h_{l-1}^{-1} - h_l^{-1} \quad \text{y} \\ q_{l+1,l} &= h_l^{-1} \end{aligned} \tag{2.15}$$

para $l = 2, \ldots, n-1$ y $q_{il} = 0$ para $|i-l| \geq 2$, $i = 1, \ldots, n$, $l = 1, \ldots, n-1$. El elemento r_{il} de la matriz $\mathbf{R}_j$ es dada por:

$$\begin{aligned} r_{i,i} &= \frac{1}{3}(h_{i-1} + h_i) \quad i = 1, \ldots, n-2, \\ r_{i,(i+1)} &= r_{(i+1),i} = \frac{1}{6} \quad i = 1, \ldots, n-1 \end{aligned} \tag{2.16}$$

y $r_{il} = 0$ para $|i-l| \geq 2$.

Tomando en cuenta la función $L(\boldsymbol{\theta})$ definida por la ecuación (2.9), la función de penalización $\mathrm{J}(f_j)$ dada por (2.12), $\alpha_j^* = -\frac{\alpha_j}{2}$ y la forma cuadrática definida por la ecuación (2.14), entonces, la función de log-verosimilitud penalizada asociada al modelo mixto vector spline bajo distribuciones elípticas puede expresarse de la forma siguiente:

$$L_p(\boldsymbol{\theta}, \alpha_1, \ldots, \alpha_m) = \sum_{i=1}^{n} L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \tag{2.17}$$

donde,

$$L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = L_i(\boldsymbol{\theta}) - \frac{1}{2n} \sum_{j=1}^{m} \alpha_j \mathbf{f}_{(j)}^{\top} \mathbf{K}_j \mathbf{f}_{(j)}, \tag{2.18}$$

en que $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^{\top}$ es el vector de parámetros de suavizamiento.

A continuación se presentarán algunos ejemplos de la función de log-verosimilitud penalizada (2.17) para algunas distribuciones elípticas multivariadas. Sean $\mathbf{y}_1, \ldots, \mathbf{y}_n$ vectores aleatorios independientes y si cada una de ellos tiene la función densidad de probabilidad dada por:

(i) la ecuación (1.9), la función de log-verosimilitud penalizada es:

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \left[-\frac{1}{2} \log |\boldsymbol{\Sigma}_i| - \kappa_1 - \frac{1}{2} \delta_i - \frac{1}{2n} \sum_{j=1}^{m} \alpha_j \mathbf{f}_{(j)}^{\top} \mathbf{K}_j \mathbf{f}_{(j)} \right],$$

en que $\kappa_1 = \frac{m}{2} \log(2\,\pi)$;

(ii) la ecuación (1.10), la función de log-verosimilitud penalizada es:

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \left[-\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \kappa_2 - \kappa_3 \log\{1 + \nu^{-1} \delta_i\} - \frac{1}{2n} \sum_{j=1}^{m} \alpha_j \mathbf{f}_{(j)}^{\top} \mathbf{K}_j \mathbf{f}_{(j)} \right],$$

en que $\kappa_2 = \log \left\{ \frac{\Gamma\left(\frac{m+\nu}{2}\right)}{(\pi\nu)^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2}\right)} \right\}$ y $\kappa_3 = \left(\frac{m+v}{2}\right)$; y,

(iii) la ecuación (1.11), la función de log-verosimilitud penalizada es

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \left[-\frac{1}{2} \log |\boldsymbol{\Sigma}_i| + \kappa_4 - \frac{1}{2} \delta_i^s - \frac{1}{2n} \sum_{j=1}^{m} \alpha_j \mathbf{f}_{(j)}^{\top} \mathbf{K}_j \mathbf{f}_{(j)} \right],$$

en que $\kappa_4 = \log \left\{ \frac{m\Gamma\left(\frac{m}{2}\right)}{\pi^{\frac{m}{2}} \Gamma\left(1+\frac{m}{2s}\right) 2^{1+\frac{m}{2s}}} \right\}$.

En el próximo capítulo se abordará el proceso de estimación e inferencia estadística de los parámetros involucrados en los MMVS, basadas en la función de log-verosimilitud penalizada.

Capítulo 3

Estimación e Inferencia

En este capítulo se aborda el proceso de estimación e inferencia estadística de los parámetros involucrados en el MMVS. En la primera sección, se presenta el proceso de estimación de los parámetros basados en la función de log-verosimilitud penalizada. Luego, se discuten algunos aspectos de la inferencia estadística en el MMVS. Posteriormente, se describe el proceso de selección de los parámetros de suavizamiento. Finalmente, se sugiere un método para evaluar el ajuste del MMVS bajo errores aleatorios que siguen una distribución normal y t-Student.

3.1. Proceso de estimación

En esta sección se discuten algunos aspectos del proceso de estimación en el MMVS. Primero, se presenta una breve discusión bibliográfica sobre algunos trabajos relacionados con el proceso de estimación en los modelos semiparamétricos. En segundo lugar, se deriva la función escore penalizada, la matriz de segundas derivadas y la matriz de información de Fisher penalizada. En tercer lugar, se presenta un procedimiento de maximización para obtener los estimadores de máxima verosimilitud penalizada (EMVP) y algunas condiciones de regularidad para asegurar la existencia de los EMVP. Finalmente, de la aplicación del procedimiento de maximización, se deriva un proceso iterativo para obtener los EMVP.

3.1.1. Discusión bibliográfica

El problema de estimación en los modelos semiparamétricos o también llamados modelos lineales parciales, ha sido estudiado por diversos autores. Por ejemplo, Heckman (1986) estima los coeficientes paramétricos y la función no paramétrica basado en el criterio de mínimos cuadrados penalizados. Además, estudia la normalidad asintótica para el estimador del coeficiente de regresión. Green (1987) estudia la estimación de máxima verosimilitud penalizada, proporcionando definiciones de la devianza, grados de libertad y residuos. Speckman (1988) presenta un procedimiento de suavizamiento de kernel para estimar los coeficientes de regresión y la función no paramétrica.

Los trabajos mencionados anteriormente, han sido estudiados asumiendo normalidad para los errores aleatorios. Sin embargo, recientemente han surgido algunos estudios asumiendo otras distribuciones para los errores aleatorios. Por ejemplo, Ibacache-Pulgar y Paula (2011) extienden los modelos lineales parciales con errores normales a errores t-Student y obtienen las estimaciones del coeficiente de regresión y la función no paramétrica utilizando las ecuaciones de verosimilitud penalizada. Posteriormente, Ibacache-Pulgar y Paula (2013) proponen un algoritmo backfitting en los modelos aditivos semiparamétricos bajo distribuciones simétricas para obtener las estimaciones de máxima verosimilitud penalizada. Por otro lado, Rigby y Stasinopoulos (2005) introdujeron los modelos aditivos generalizados para localización, escala y forma. El proceso de estimación de los parámetros involucrados está basado en la función de log-verosimilitud penalizada y el algoritmo backfitting.

En el contexto de los modelos lineales mixtos semiparamétricos, Zhang et al. (1998) derivan estimadores de máxima verosimilitud penalizada de los coeficientes de regresión y la función no paramétrica, mostrando que este último conduce a un spline cúbico natural. Staudenmayer et al. (2009) desarrollan e implementan los modelos lineales mixtos utilizando un enfoque jerárquico basado en la distribución t-Student. Recientemente, Ibacache-Pulgar et al. (2012) proponen un proceso iterativo basado en el método backfitting para la estimación de los parámetros involucrados en el modelo mixto semiparamétrico bajo distribuciones elípticas.

Por otro lado, Fessler (1991 a) presenta un algoritmo eficiente para suavizamiento no paramétrico en funciones de valores vectoriales, con estructura de covarianza general. Fessler (1991 b) propone un algoritmo para una función suave de valores vectoriales en

estructuras de valores vectoriales no lineales de esa función. Además, Yee y Wild (1996) sugieren un algoritmo backfitting para estimar los parámetros involucrados en el modelo aditivo generalizado vector spline (MAVG). Finalmente, Yee (1998) presenta un algoritmo backfitting que disminuye el esfuerzo computacional del proceso de estimación en los MAVG.

3.1.2. Función escore penalizada

En esta subsección se deriva la función escore penalizada asociada al vector de parámetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \mathbf{f}_{(1)}^\top, \ldots, \mathbf{f}_{(m)}^\top, \boldsymbol{\tau}^\top)^\top$. Asumiendo que la función de log-verosimilitud penalizada $L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ es regular con respecto a los elementos de $\boldsymbol{\theta}$, la función escore penalizada del MMVS es dada por

$$\begin{aligned}\mathbf{U}_p(\boldsymbol{\theta}) &= \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} \\ &= \sum_{i=1}^{n} \frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} \\ &= \sum_{i=1}^{n} \mathbf{U}_{p_i}(\boldsymbol{\theta}),\end{aligned}$$

donde

$$\mathbf{U}_{p_i}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{U}_{p_i}^{\boldsymbol{\beta}}(\boldsymbol{\theta}) \\ \mathbf{U}_{p_i}^{\mathbf{f}_{(1)}}(\boldsymbol{\theta}) \\ \vdots \\ \mathbf{U}_{p_i}^{\mathbf{f}_{(m)}}(\boldsymbol{\theta}) \\ \mathbf{U}_{p_i}^{\boldsymbol{\tau}}(\boldsymbol{\theta}) \end{pmatrix},$$

con $\mathbf{U}_{p_i}^{\boldsymbol{\beta}}(\boldsymbol{\theta})$ y $\mathbf{U}_{p_i}^{\mathbf{f}_{(j)}}(\boldsymbol{\theta})$, para $j = 1, \ldots, m$, denotando las funciones escore parciales asociadas a los efectos fijos del modelo; y $\mathbf{U}_{p_i}^{\boldsymbol{\tau}}(\boldsymbol{\theta})$ denota la función escore asociada a la componente de varianza $\boldsymbol{\tau}$.

En este trabajo, se asumirá que la función generadora de densidades $g(\cdot)$ presentada

en el Capítulo 1 es continua y diferenciable. Defina las siguientes cantidades:

$$\upsilon_i = -2\zeta_g(\delta_i), \qquad \upsilon_i' = \frac{\partial \upsilon_i}{\partial \delta_i} \qquad \text{y} \qquad \zeta_g(\delta_i) = \frac{g'(\delta_i)}{g(\delta_i)},$$

para $i = 1, \ldots, n$. Las funciones escore parciales para $\boldsymbol{\beta}$ y $\mathbf{f}_{(j)}$, asociadas a los efectos fijos del modelo, son dadas por

$$\begin{aligned}\mathbf{U}_{p_i}^{\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \\ &= \upsilon_i \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i\end{aligned}$$

y

$$\begin{aligned}\mathbf{U}_{p_i}^{\mathbf{f}_{(j)}}(\theta) &= \frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \mathbf{f}_{(j)}} \\ &= \upsilon_i \mathbf{N}_i^{(j)\top} \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i - \frac{\alpha_j}{n} \mathbf{K}_j \mathbf{f}_{(j)} \quad (j = 1, \ldots, m).\end{aligned}$$

Por otro lado, las funciones escore parciales asociadas a las componentes de varianza están dadas por

$$\mathbf{U}_{p_i}^{\boldsymbol{\tau}}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{U}_{p_i}^{\boldsymbol{\tau}_1}(\boldsymbol{\theta}) \\ \vdots \\ \mathbf{U}_{p_i}^{\boldsymbol{\tau}_{d+1}}(\boldsymbol{\theta}) \end{pmatrix},$$

donde sus elementos se obtienen de la siguiente manera:

$$\begin{aligned}\mathbf{U}_{p_i}^{\tau_\ell}(\boldsymbol{\theta}) &= \frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \tau_\ell} \\ &= -\frac{1}{2}\left[\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathrm{i}}^{-1}\dot{\boldsymbol{\Sigma}}_{\mathrm{i}}(\ell)\} - \upsilon_{\mathrm{i}} \mathbf{r}_{\mathrm{i}}^\top \boldsymbol{\Sigma}_{\mathrm{i}}^{-1}\dot{\boldsymbol{\Sigma}}_{\mathrm{i}}(\ell)\boldsymbol{\Sigma}_{\mathrm{i}}^{-1}\mathbf{r}_{\mathrm{i}}\right],\end{aligned}$$

con $\dot{\boldsymbol{\Sigma}}_i(\ell) = \dfrac{\partial \boldsymbol{\Sigma}_i}{\partial \boldsymbol{\tau}_\ell}$ para $i = 1, \ldots, n$ y $\ell = 1, \ldots, d+1$. Los cálculos algebraicos de las componentes de varianza y efectos fijos se presentan en el Apéndice A.1.

Se observa que las cantidades positivas υ_i (excepto en las distribuciones Kotz, Kotz generalizada y doble exponencial) pueden ser interpretadas como pesos. Además, la función $g(\cdot)$ es una función positiva y decreciente para casi todas las distribuciones que pertene-

cen a la clase de distribuciones elípticas (Ibacache, 2009). En la Tabla 3.1 se muestran las expresiones de las cantidades υ_i y υ_i' para algunas distribuciones elípticas multivariadas.

Tabla 3.1: Cantidades υ_i y υ_i' para algunas distribuciones elípticas multivariadas

Distribución	υ_i	υ_i'
Normal	1	0
t-Student	$\frac{\upsilon_i+m}{(\upsilon_i+\delta_i)}$	$\frac{\upsilon_i+m}{(\upsilon_i+\delta_i)^2}$
Cauchy	$\frac{1+m}{(1+\delta_i)}$	$\frac{1+m}{(1+\delta_i)^2}$

3.1.3. Matriz de segundas derivadas

La matriz de segundas derivadas o también denominada matriz Hessiana es de suma importancia en la literatura estadística por diversas razones. Por ejemplo, la matriz Hessiana en conjunto con la función escore permiten obtener las soluciones a ecuaciones normales que no tengan soluciones explícitas, mediante un proceso iterativo. Éste proceso iterativo es llamado Newton-Raphson. Por otro lado, permite el estudio de la concavidad de la función de log-verosimilitud penalizada, propiedad necesaria para asegurar la existencia de los estimadores de máxima verosimilitud penalizada en el contexto semiparamétrico. Además, la matriz Hessiana es una herramienta esencial en el análisis de diagnóstico de influencia local, el que permite evaluar la sensibilidad de los estimadores de máxima verosimilitud penalizada, ver por ejemplo, Ibacache y Paula (2013).

Al utilizar álgebra y derivadas matriciales (ver Apéndice A.2), la matriz Hessiana está dada por la siguiente expresión:

$$\ddot{L}_p(\boldsymbol{\theta},\boldsymbol{\alpha}) = \frac{\partial^2 L_p(\boldsymbol{\theta};\boldsymbol{\alpha})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$$

$$= \begin{pmatrix} \ddot{\mathbf{L}}_p^{\boldsymbol{\beta}\boldsymbol{\beta}} & \ddot{\mathbf{L}}_p^{\boldsymbol{\beta}\mathbf{f}} & \ddot{\mathbf{L}}_p^{\boldsymbol{\beta}\boldsymbol{\tau}} \\ \ddot{\mathbf{L}}_p^{\mathbf{f}\boldsymbol{\beta}} & \ddot{\mathbf{L}}_p^{\mathbf{f}\mathbf{f}} & \ddot{\mathbf{L}}_p^{\mathbf{f}\boldsymbol{\tau}} \\ \ddot{\mathbf{L}}_p^{\boldsymbol{\tau}\boldsymbol{\beta}} & \ddot{\mathbf{L}}_p^{\boldsymbol{\tau}\mathbf{f}} & \ddot{\mathbf{L}}_p^{\boldsymbol{\tau}\boldsymbol{\tau}} \end{pmatrix}. \tag{3.1}$$

En particular, los elementos de la matriz (3.1) son los siguientes:

$$\ddot{\mathbf{L}}_p^{\boldsymbol{\beta}\boldsymbol{\beta}} = -\sum_{i=1}^{n}\left[\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right],$$

$$\ddot{\mathbf{L}}_p^{\boldsymbol{\beta}\mathbf{f}} = -\sum_{i=1}^{n}\left[\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(j)}\right],$$

$$\ddot{\mathbf{L}}_p^{\boldsymbol{\beta}\boldsymbol{\tau}} = -\sum_{i=1}^{n}\left[\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + \upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\right],$$

$$\ddot{\mathbf{L}}_p^{\mathbf{f}\boldsymbol{\beta}} = -\sum_{i=1}^{n}\left[\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + \upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right],$$

$$\ddot{\mathbf{L}}_p^{\mathbf{f}\boldsymbol{\tau}} = -\sum_{i=1}^{n}\left[\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\right],$$

$$\ddot{\mathbf{L}}_p^{\boldsymbol{\tau}\boldsymbol{\beta}} = -\sum_{i=1}^{n}\left[\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i^{-1}(\ell)\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + \upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right],$$

$$\ddot{\mathbf{L}}_p^{\boldsymbol{\tau}\mathbf{f}} = -\sum_{i=1}^{n}\left[\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i^{-1}(\ell)\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + \upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(j)}\right],$$

$$\ddot{\mathbf{L}}_p^{\mathbf{f}\mathbf{f}} = \begin{cases} -\displaystyle\sum_{i=1}^{n}\left[\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(k)}\right] & \text{si } j \neq k \\ -\displaystyle\sum_{i=1}^{n}\left[\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(j)}\right] - \dfrac{\alpha_j}{n}\mathbf{K}_j & \text{si } j = k, \end{cases}$$

y

$$\begin{aligned}\ddot{\mathbf{L}}_p^{\boldsymbol{\tau}\boldsymbol{\tau}} = &\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}_{\mathrm{i}}^{-1}\left[\dot{\boldsymbol{\Sigma}}_{\mathrm{i}}(\ell)\boldsymbol{\Sigma}_{\mathrm{i}}^{-1}\dot{\boldsymbol{\Sigma}}_{\mathrm{i}}(\jmath) - \ddot{\boldsymbol{\Sigma}}_{\mathrm{i}}(\jmath,\ell)\right]\right\} \\ &- \frac{1}{2}\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\left\{\dot{\boldsymbol{\Sigma}}_i(\ell)\boldsymbol{\Sigma}_i^{-1}\left(\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1} + 2\upsilon_i\mathbf{I}_m\right)\dot{\boldsymbol{\Sigma}}_i(\jmath) + \upsilon_i\ddot{\boldsymbol{\Sigma}}_i(\jmath,\ell)\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i,\end{aligned}$$

donde $\ddot{\boldsymbol{\Sigma}}_i(\jmath,\ell) = \dfrac{\partial^2\boldsymbol{\Sigma}_i}{\partial\tau_\jmath\partial\tau_\ell}$, para $\jmath = 1,\ldots,d+1$ y $\ell = 1,\ldots,d+1$.

3.1.4. Matriz de información de Fisher penalizada

En general, la matriz de información de Fisher tiene un rol importante en el proceso de estimación de los parámetros y el análisis inferencial paramétrico, ya que permite obtener un proceso de estimación para los parámetros del modelo y estimar de manera

aproximada el error estándar de los estimadores de máxima verosimilitud. En el contexto semiparamétrico, su aplicabilidad es similar, ya que permite derivar el algoritmo escoring de Fisher y aproximar el error estándar de los estimadores de máxima verosimilitud penalizada a partir de la matriz de información de Fisher penalizada.

Sea $b_{i(j,l)} = \mathrm{tr}\left\{\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\jmath)\right\}\mathrm{tr}\left\{\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\right\}$, $\dot{\boldsymbol{\Sigma}}_i(\jmath) = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_\jmath}$, $\dot{\boldsymbol{\Sigma}}_i(\ell) = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \tau_\ell}$, $\mathrm{d}_{g_i} = \mathrm{E}[\zeta_g^2(\delta_i)\delta_i]$ y $\mathcal{F}_{g_i} = \mathrm{E}[\zeta_g^2(\delta_i)\delta_i^2]$, con $\delta_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{r}_i^\top \mathbf{r}_i$ y $\mathbf{r}_i \sim \mathcal{E}l(\mathbf{0}, \mathbf{I}_m)$. Para el MMVS bajo distribuciones elípticas, la matriz de información de Fisher penalizada asociada al vector de parámetros $\boldsymbol{\theta}$ está dada por

$$\begin{aligned}\mathcal{I}_p(\boldsymbol{\theta}) &= -\mathrm{E}\left\{\frac{\partial^2 \mathrm{L}_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top}\right\} \\ &= \begin{pmatrix} \mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_p^{\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta}) \end{pmatrix},\end{aligned} \tag{3.2}$$

en que

$$\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathcal{I}_{p_i}^{\boldsymbol{\beta}\mathbf{f}}(\boldsymbol{\theta})$$

y

$$\mathcal{I}_p^{\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathcal{I}_{p_i}^{\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta})$$

son matrices de orden $[(p + mn) \times (p + mn)]$ y $(d^* \times d^*)$, respectivamente.

Los elementos de la matriz $\mathcal{I}_{p_i}^{\boldsymbol{\beta}\mathbf{f}}(\boldsymbol{\theta})$ están dados por

$$\mathcal{I}_{p_i}^{\boldsymbol{\beta}\mathbf{f}}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_{p_i}^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathcal{I}_{p_i}^{\boldsymbol{\beta}\mathbf{f}_{(1)}}(\boldsymbol{\theta}) & \dots & \mathcal{I}_{p_i}^{\boldsymbol{\beta}\mathbf{f}_{(m)}}(\boldsymbol{\theta}) \\ \mathcal{I}_{p_i}^{\mathbf{f}_{(1)}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathcal{I}_{p_i}^{\mathbf{f}_{(1)}\mathbf{f}_{(1)}}(\boldsymbol{\theta}) & \dots & \mathcal{I}_{p_i}^{\mathbf{f}_{(1)}\mathbf{f}_{(m)}}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{I}_{p_i}^{\mathbf{f}_{(m)}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathcal{I}_{p_i}^{\mathbf{f}_{(m)}\mathbf{f}_{(1)}}(\boldsymbol{\theta}) & \dots & \mathcal{I}_{p_i}^{\mathbf{f}_{(m)}\mathbf{f}_{(m)}}(\boldsymbol{\theta}) \end{pmatrix},$$

donde

$$\mathcal{I}_{p_i}^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{4\mathrm{d}_{g_i}}{m}\mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i,$$

$$\mathcal{I}_{p_i}^{\boldsymbol{\beta}\mathbf{f}_{(j)}}(\boldsymbol{\theta}) = \frac{4\mathrm{d}_{g_i}}{m}\mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(j)},$$

$$\mathcal{I}_{p_i}^{\mathbf{f}_{(j)}\mathbf{f}_{(j)}}(\boldsymbol{\theta}) = \frac{4\mathrm{d}_{g_i}}{m}\mathbf{N}_i^{(j)\top} \boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(j)} + \alpha_j \mathbf{K}_j$$

y

$$\mathcal{I}_{p_i}^{\mathbf{f}_{(j)}\mathbf{f}_{(k)}}(\boldsymbol{\theta}) = \frac{4\mathrm{d}_{g_i}}{m}\mathbf{N}_i^{(j)\top} \boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(k)},$$

para $i = 1, \ldots, n$ y $j = 1, \ldots, m$, con $j \neq k$.

Por otra parte, el elemento (j, l)-ésimo de la matriz $\mathcal{I}_{p_i}^{\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta})$ está dado por

$$\mathcal{I}_{p_{i(j,l)}}^{\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta}) = \left[\frac{\mathrm{b}_{i(j,l)}}{4}\left\{\frac{4\mathcal{F}_{g_i}}{m(m+2)} - 1\right\} + \frac{2\mathcal{F}_{g_i}}{m(m+2)}\mathrm{tr}\left\{\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(j)\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\right\}\right].$$

Note que las cantidades $\mathcal{F}_{g_i}$ y d_{g_i} que aparecen en las ecuaciones anteriores tienen una forma cerrada para algunas distribuciones elípticas multivariadas. Por ejemplo, para la distribución normal, t-student y Cauchy, sus expresiones son presentadas en la Tabla 3.2.

Tabla 3.2: Cantidades d_{g_i} y $\mathcal{F}_{g_i}$ para algunas distribuciones elípticas multivariadas

Distribución	d_{g_i}	$\mathcal{F}_{g_i}$
Normal	$\frac{m}{4}$	$\frac{m(m+2)}{4}$
Cauchy	$\frac{m}{4}\left(\frac{1+m}{3+m}\right)$	$\frac{m(m+2)}{4}\frac{(1+m)}{(3+m)}$
t-Student	$\frac{m}{4}\left(\frac{v_i+m}{v_i+m+2}\right)$	$\frac{m(m+2)}{4}\left(\frac{v_i+m}{v_i+m+2}\right)$

Finalmente, cabe señalar que debido a la propiedad de ortogonalidad de los elementos de $\boldsymbol{\tau}$ con respecto a $\boldsymbol{\beta}$ y $\mathbf{f}_{(1)}, \ldots, \mathbf{f}_{(m)}$, se obtiene lo siguiente:

$$\begin{aligned}\mathcal{I}_p^{\boldsymbol{\beta\tau}}(\boldsymbol{\theta}) &= -\mathrm{E}\left\{\frac{\partial^2 \mathrm{L}_p(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\tau}^\top}\right\} \\ &= \mathbf{0}\end{aligned}$$

y

$$\begin{aligned}\mathcal{I}_p^{\mathbf{f}_{(m)}\boldsymbol{\tau}}(\boldsymbol{\theta}) &= -\mathrm{E}\left\{\frac{\partial^2 \mathrm{L}_p(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\mathbf{f}_{(j)}\partial\boldsymbol{\tau}^\top}\right\} \\ &= \mathbf{0}.\end{aligned}$$

Este resultado facilita la construcción del proceso iterativo de escoring de Fisher para estimar el vector de parámetros $\boldsymbol{\theta}$.

3.1.5. Maximizaciones sucesivas para obtener EMVP

Debido a que $f_j(\cdot)$ es un parámetro de dimensión infinita, se consideran los estimadores de máxima verosimilitud penalizada (EMVP) de $\boldsymbol{\theta}$, que conduce a un estimador spline cúbico natural de $f_j(\cdot)$. Específicamente, el valor de $\boldsymbol{\theta}$ que maximiza $L_p(\boldsymbol{\theta},\boldsymbol{\alpha})$ sobre $\boldsymbol{\Theta}$, denotado por $\widehat{\boldsymbol{\theta}}$, satisface la desigualdad $L_p(\widehat{\boldsymbol{\theta}},\boldsymbol{\alpha}) \geq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_p(\boldsymbol{\theta},\boldsymbol{\alpha})$. Ahora, la determinación de $\widehat{\boldsymbol{\theta}}$ puede efectuarse considerando maximizaciones sucesivas, descritas, por ejemplo, por Gourieroux y Monford (1995). En particular, para el vector $\boldsymbol{\alpha}$ fijo y $j = 2$, la solución $\widehat{\boldsymbol{\theta}}$ al problema de maximización $L_p(\boldsymbol{\theta},\boldsymbol{\alpha}) = \text{máx}_{\boldsymbol{\beta},\mathbf{f}_{(1)},\mathbf{f}_{(2)},\boldsymbol{\tau},\boldsymbol{\alpha}} L_p(\boldsymbol{\beta},\mathbf{f}_{(1)},\mathbf{f}_{(2)},\boldsymbol{\tau},\boldsymbol{\alpha})$ se puede obtener mediante los siguientes pasos:

(i) Primero, se maximiza la función $L_p(\boldsymbol{\beta},\mathbf{f}_{(1)},\mathbf{f}_{(2)},\boldsymbol{\tau},\boldsymbol{\alpha})$ sobre $\boldsymbol{\beta}$ manteniendo los parámetros $\mathbf{f}_{(1)}$, $\mathbf{f}_{(2)}$ y $\boldsymbol{\tau}$ fijos. El valor máximo, $\widehat{\boldsymbol{\beta}}(\mathbf{f}_{(1)},\mathbf{f}_{(2)},\boldsymbol{\tau})$, es alcanzado para valores de $\boldsymbol{\beta}$ en un conjunto $\mathcal{B}(\mathbf{f}_{(1)},\mathbf{f}_{(2)},\boldsymbol{\tau})$ dependiendo de los parámetros $\mathbf{f}_{(1)}$, $\mathbf{f}_{(2)}$ y $\boldsymbol{\tau}$. Así, si $\boldsymbol{\beta} \in \mathcal{B}(\mathbf{f}_{(1)},\mathbf{f}_{(2)},\boldsymbol{\tau})$, el valor de la log-verosimilitud penalizada es $L_p^c(\mathbf{f}_{(1)},\mathbf{f}_{(2)},\boldsymbol{\tau},\boldsymbol{\alpha}) = \text{máx}_{\boldsymbol{\beta}} L_p(\boldsymbol{\beta},\mathbf{f}_{(1)},\mathbf{f}_{(2)},\boldsymbol{\tau},\boldsymbol{\alpha})$. Aquí, L_p^c es llamada la log-verosimilitud penalizada concentrada en $\boldsymbol{\beta}$.

(ii) Luego, se maximiza $L_p^c(\mathbf{f}_{(1)},\mathbf{f}_{(2)},\boldsymbol{\tau},\boldsymbol{\alpha})$ sobre $\mathbf{f}_{(1)}$ manteniendo $\mathbf{f}_{(2)}$ y $\boldsymbol{\tau}$ fijos. El máximo valor, $\widehat{\mathbf{f}}_{(1)} = \widehat{\mathbf{f}}_{(1)}(\mathbf{f}_{(2)},\boldsymbol{\tau})$ es alcanzado para valores de $\mathbf{f}_{(1)}$ en un conjunto $\mathcal{F}_1(\mathbf{f}_{(2)},\boldsymbol{\tau})$

dependiendo sobre los parámetros $\mathbf{f}_{(2)}$ y $\boldsymbol{\tau}$. Por lo tanto, si $\mathbf{f}_{(1)} \in \mathcal{F}_1(\mathbf{f}_{(2)}, \boldsymbol{\tau})$, el valor de la log-verosimilitud penalizada es $L_p^c(\mathbf{f}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha}) = \text{máx}_{\mathbf{f}_{(1)}} L_p^c(\mathbf{f}_{(1)}, \mathbf{f}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha})$. Aquí, L_p^c es llamada la log-verosimilitud concentrada en $\boldsymbol{\beta}$ y $\mathbf{f}_{(1)}$.

(iii) Ahora, se maximiza $L_p^c(\mathbf{f}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha})$ sobre $\mathbf{f}_{(2)}$ manteniendo $\boldsymbol{\tau}$ fijo. El máximo valor, $\widehat{\mathbf{f}}_{(2)} = \widehat{\mathbf{f}}_{(2)}(\boldsymbol{\tau})$ es alcanzado para valores de $\mathbf{f}_{(2)}$ en un conjunto $\mathcal{F}_2(\mathbf{f}_{(2)}, \boldsymbol{\tau})$ dependiendo sobre el parámetro $\boldsymbol{\tau}$. Por lo tanto, si $\mathbf{f}_{(2)} \in \mathcal{F}_2(\boldsymbol{\tau})$, el valor de la log-verosmilitud penalizada es $L_p^c(\boldsymbol{\tau}, \boldsymbol{\alpha}) = \text{máx}_{\mathbf{f}_{(2)}} L_p^c(\mathbf{f}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha})$. Aquí, L_p^c es llamada la log-verosimilitud concentrada en $\boldsymbol{\beta}$, $\mathbf{f}_{(1)}$ y $\mathbf{f}_{(2)}$.

(iv) Finalmente, se maximiza la función log-verosimilitud penalizada concentrada $L_p^c(\boldsymbol{\tau}, \boldsymbol{\alpha}) = L_p(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{f}}_{(1)}, \widehat{\mathbf{f}}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha})$ sobre $\boldsymbol{\tau}$. El máximo valor $\widehat{\boldsymbol{\tau}}$ es alcanzado sobre un conjunto de $\mathcal{C}$ de valores $\boldsymbol{\tau}$.

Este procedimiento permite determinar de manera aproximada los estimadores de máxima verosimilitud penalizada, considerando maximizaciones sucesivas. Algunas aplicaciones de este procedimiento en la clase de los modelos semiparamétricos pueden encontrarse, por ejemplo, en Ibacache et al. (2013).

3.1.6. Estudio de la concavidad

El estudio de la concavidad permite asegurar la existencia de los estimadores de máxima verosimilitud penalizada. Aquí se derivan las condiciones suficientes para garantizar la concavidad de la función log-verosimiltud penalizada $L_p(\boldsymbol{\beta}, \mathbf{f}_{(1)}, \mathbf{f}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha})$ en $\boldsymbol{\beta}$, $\mathbf{f}_{(1)}$, $\mathbf{f}_{(2)}$ y $\boldsymbol{\tau}$. En efecto, se debe garantizar lo siguiente:

(i') En el paso (i) (ver Subsección 3.1.5), la concavidad (en $\boldsymbol{\beta}$) de $L_p(\boldsymbol{\beta}, \mathbf{f}_{(1)}, \mathbf{f}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha})$ es garantizada si y sólo si la matriz $\ddot{\mathbf{L}}_p^{\boldsymbol{\beta}\boldsymbol{\beta}} \leq 0$ (semidefinida negativa) o, equivalentemente si y sólo si $-\ddot{\mathbf{L}}_p^{\boldsymbol{\beta}\boldsymbol{\beta}} \geq 0$ (definida no negativa). Uno tiene que $-\ddot{\mathbf{L}}_p^{\boldsymbol{\beta}\boldsymbol{\beta}} \geq 0$ si y sólo si $\{\upsilon_i \boldsymbol{\Sigma}_i + 2\upsilon_i' \mathbf{r}_i \mathbf{r}_i^\top\} \geq 0$, esto es, si $\upsilon_i \geq 0$ y $\upsilon_i' \geq 0$, para $i = 1, \ldots, n$.

(ii') Para el paso (ii) (ver Subsección 3.1.5), uno tiene concavidad (en $\mathbf{f}_{(1)}$) de $L_p^c(\mathbf{f}_{(1)}, \mathbf{f}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha})$ si y sólo si la matriz $\ddot{\mathbf{L}}_p^{\mathbf{f}_{(1)}\mathbf{f}_{(1)}} \leq 0$ o, equivalentemente, si y sólo si $-\ddot{\mathbf{L}}_p^{\mathbf{f}_{(1)}\mathbf{f}_{(1)}} \geq 0$. Consecuentemente, $-\ddot{\mathbf{L}}_p^{\mathbf{f}_{(1)}\mathbf{f}_{(1)}} \geq 0$ si y sólo si $\{\upsilon_i \boldsymbol{\Sigma}_i + 2\upsilon_i' \mathbf{r}_i \mathbf{r}_i^\top\} \geq 0$ y $\alpha_1 \mathbf{K}_1 \geq 0$. Además, tenemos que α_1 es un escalar positivo y $\mathbf{K}_1 \geq 0$, por lo tanto $\alpha_1 \mathbf{K}_1 \geq 0$. Por otra parte, $\{\upsilon_i \boldsymbol{\Sigma}_i + 2\upsilon_i' \mathbf{r}_i \mathbf{r}_i^\top\} \geq 0$ si y sólo si $\upsilon_i \geq 0$ y $\upsilon_i' \geq 0$, para $i = 1, \ldots, n$.

(iii') De manera similar, en el paso (iii) (ver subsección 3.1.5), uno tiene concavidad (en $\mathbf{f}_{(2)}$) de $L_p^c(\mathbf{f}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha})$ si y sólo si la matriz $\ddot{\mathbf{L}}_p^{\mathbf{f}_{(2)}\mathbf{f}_{(2)}} \leq 0$ o, equivalentemente si y sólo si la matriz $-\ddot{\mathbf{L}}_p^{\mathbf{f}_{(2)}\mathbf{f}_{(2)}} \geq 0$. Consecuentemente, $-\ddot{\mathbf{L}}_p^{\mathbf{f}_{(2)}\mathbf{f}_{(2)}} \geq 0$ si y sólo si la matriz $\{\upsilon_i \boldsymbol{\Sigma}_i + 2\upsilon_i' \mathbf{r}_i \mathbf{r}_i^\top\} \geq 0$ y $\alpha_2 \mathbf{K}_2 \geq 0$. Además, tenemos que α_2 es un escalar positivo y $\mathbf{K}_2 \geq 0$, por lo tanto $\alpha_2 \mathbf{K}_2 \geq 0$. Por otra parte, $\{\upsilon_i \boldsymbol{\Sigma}_i + 2\upsilon_i' \mathbf{r}_i \mathbf{r}_i^\top\}$ si y sólo si $\upsilon_i \geq 0$ y $\upsilon_i' \geq 0$, para $i = 1, \ldots, n$.

(iv') Finalmente, en el paso (iv) (ver subsección 3.1.5), la concavidad (en $\boldsymbol{\tau}$) de $L_p^c(\boldsymbol{\tau}, \boldsymbol{\alpha})$ es garantizada si y sólo si $\dfrac{\partial^2 L_p^c(\boldsymbol{\tau}, \boldsymbol{\alpha})}{\partial \boldsymbol{\tau}_\ell \partial \boldsymbol{\tau}_\ell} < 0$, para $\ell = 1, \ldots, d+1$.

Algunos autores que han realizado este tipo de estudio son: Ibacache y Paula (2011) en los modelos semiparamétricos t-Student, e Ibacache et al. (2013) para la clase de modelos aditivos semiparamétricos con errores simétricos.

3.1.7. Proceso iterativo

Cuando $j = 2$ y el vector de parámetros de suavizamiento $\boldsymbol{\alpha}$ es fijo, el procedimiento de 4 pasos descrito en la Subsección 3.1.5 es equivalente a resolver las siguientes ecuaciones:

$$\frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} = \mathbf{0}\,, \qquad \frac{\partial L_p^c(\mathbf{f}_{(1)}, \mathbf{f}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha})}{\partial \mathbf{f}_{(1)}} = \mathbf{0}\,, \qquad \frac{\partial L_p^c(\mathbf{f}_{(2)}, \boldsymbol{\tau}, \boldsymbol{\alpha})}{\partial \mathbf{f}_{(2)}} = \mathbf{0} \tag{3.3}$$

y

$$\frac{\partial L_p^c(\boldsymbol{\tau}, \boldsymbol{\alpha})}{\partial \boldsymbol{\tau}} = \mathbf{0}.$$

La solución a este conjunto de ecuaciones conlleva al siguiente proceso iterativo:

Paso 1 (Algoritmo backfitting)

Sean $\boldsymbol{\theta}^{(0)}$ y $\boldsymbol{\theta}^{(s+1)}$ los valores iniciales y actuales de $\boldsymbol{\theta}$, respectivamente ($s = 0, 1, \ldots$). En esta etapa, las ecuaciones actuales para $\boldsymbol{\beta}$, $\mathbf{f}_{(1)}$ y $\mathbf{f}_{(2)}$, que se obtienen del conjunto de ecuaciones (3.3), inducen a un algoritmo backfitting interno con ponderaciones dadas por $\mathbf{W}_i^{(s)} = \upsilon_i \boldsymbol{\Sigma}_i^{-1}\big|_{\boldsymbol{\theta}^{(s)}}$ ($\forall\ i = 1, \ldots, n$), donde la solución es el límite de la

iteración obtenida al repetir cíclicamente las siguientes tres ecuaciones:

$$\boldsymbol{\beta}^{(s+1,s^*+1)} = \mathbf{D}_0 \sum_{i=1}^{n} \mathbf{X}_i^\top \mathbf{W}_i^{(s)} (\mathbf{y}_i - \mathbf{N}_i^{(1)} \mathbf{f}_{(1)}^{(s+1,s^*)} - \mathbf{N}_i^{(2)} \mathbf{f}_{(2)}^{(s+1,s^*)}), \tag{3.4}$$

$$\mathbf{f}_{(1)}^{(s+1,s^*+1)} = \mathbf{D}_1 \sum_{i=1}^{n} \mathbf{N}_i^{(1)\top} \mathbf{W}_i^{(s)} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(s+1,s^*+1)} - \mathbf{N}_i^{(2)} \mathbf{f}_{(2)}^{(s+1,s^*)}) \tag{3.5}$$

y

$$\mathbf{f}_{(2)}^{(s+1,s^*+1)} = \mathbf{D}_2 \sum_{i=1}^{n} \mathbf{N}_i^{(2)\top} \mathbf{W}_i^{(s)} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(s+1,s^*+1)} - \mathbf{N}_i^{(1)} \mathbf{f}_{(1)}^{(s+1,s^*+1)}), \tag{3.6}$$

donde

$$\mathbf{D}_0 = \left\{ \sum_{i=1}^{n} \mathbf{X}_i^\top \mathbf{W}_i^{(s)} \mathbf{X}_i \right\}^{-1},$$

$$\mathbf{D}_1 = \left\{ \sum_{i=1}^{n} \mathbf{N}_i^{(1)\top} \mathbf{W}_i^{(s)} \mathbf{N}_i^{(1)} + \frac{\alpha_1}{n} \mathbf{K}_1 \right\}^{-1}$$

y

$$\mathbf{D}_2 = \left\{ \sum_{i=1}^{n} \mathbf{N}_i^{(2)\top} \mathbf{W}_i^{(s)} \mathbf{N}_i^{(2)} + \frac{\alpha_2}{n} \mathbf{K}_2 \right\}^{-1},$$

para $s^* = 0, 1, \ldots$. En esta etapa, se puede tomar como valores iniciales $\mathbf{f}_{(1)}^{(s+1,0)} = \mathbf{f}_{(1)}^{(0)}$ y las estimaciones de máxima verosimilitud penalizada del MMVS bajo distribución normal. El conjunto de ecuaciones (3.4)-(3.6) se iteran hasta converger.

Paso 2 (log-verosimilitud penalizada)

Sea $\boldsymbol{\beta}^{(s+1)} = \widehat{\boldsymbol{\beta}}^{(s+1,s^*+1)}$, $\mathbf{f}_{(1)}^{(s+1)} = \widehat{\mathbf{f}}_{(1)}^{(s+1,s^*+1)}$ y $\mathbf{f}_{(2)}^{(s+1)} = \widehat{\mathbf{f}}_{(2)}^{(s+1,s^*+1)}$. Actualizar los parámetros $\boldsymbol{\tau}$ maximizando la log-verosimilitud penalizada concentrada $L_p^c(\boldsymbol{\tau}, \boldsymbol{\alpha}) = L_p(\boldsymbol{\beta}^{(s+1)}, \mathbf{f}_{(1)}^{(s+1)}, \mathbf{f}_{(2)}^{(s+1)}, \boldsymbol{\tau}, \boldsymbol{\alpha})$, con respecto a $\boldsymbol{\tau}$, esto es,

$$\boldsymbol{\tau}^{(s+1)} = \max_{\boldsymbol{\tau}} L_p^c(\boldsymbol{\tau}, \boldsymbol{\alpha}). \tag{3.7}$$

De manera análoga a Ibacache et al. (2012), se sugiere utilizar el algoritmo escoring

de Fisher para resolver la ecuación (3.7), esto es,

$$\boldsymbol{\tau}^{(s+1)} = \boldsymbol{\tau}^{(s)} - \left[\mathrm{E}\left\{\ddot{\mathbf{L}}_p^{\boldsymbol{\tau}\boldsymbol{\tau}}\right\}\right]^{-1} \left.\frac{\partial L_p(\boldsymbol{\tau}, \boldsymbol{\alpha})}{\partial \boldsymbol{\tau}}\right|_{\boldsymbol{\theta}^{(s)}}, \tag{3.8}$$

donde $\ddot{\mathbf{L}}_p^{\boldsymbol{\tau}\boldsymbol{\tau}}$ es definida en la Subsección 3.1.3.

Así, alternando entre los Pasos 1 y 2, se obtienen aproximadamente las estimaciones de máxima verosimilitud penalizada de $\boldsymbol{\theta}$.

3.2. Inferencia estadística

En esta sección se discuten algunos aspectos de la inferencia estadística en el MMVS. Primero, se describe un procedimiento para calcular el error estándar aproximado para los estimadores de máxima verosimilitud penalizada de los parámetros. Luego, se derivan bandas de confianza para las funciones no paramétricas. Posteriormente, se propone un método para estimar los efectos aleatorios. Finalmente, se discute sobre los grados de libertad y la selección de los parámetros de suavizamiento.

3.2.1. Error estándar

De acuerdo a lo mencionado en la Subsección 3.1.4, el error estándar de los estimadores de máxima verosimilitud penalizada se obtiene a través de la matriz de varianza-covarianza asintótica de $\widehat{\boldsymbol{\theta}}$. Esta última, se deriva de la inversa de la matriz de información de Fisher penalizada, considerando la función de verosimilitud penalizada como una función de verosimilitud usual. Entonces, la matriz de varianza-covarianza asintótica de $\widehat{\boldsymbol{\theta}}$ para el MMVS toma una estructura diagonal en bloque como sigue:

$$\mathcal{I}_p^{-1} = \begin{pmatrix} \mathcal{I}_p^1 & \mathcal{I}_p^2 & \mathbf{0} \\ \mathcal{I}_p^3 & \mathcal{I}_p^4 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}_p^{\boldsymbol{\tau}\boldsymbol{\tau}-1} \end{pmatrix}, \tag{3.9}$$

donde

$$\mathcal{I}_p^1 = \left(\mathcal{I}_p^{\boldsymbol{\beta\beta}} - \mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^*}\mathcal{I}_p^{\mathbf{ff}-1}\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^{*\top}}\right)^{-1},$$
$$\mathcal{I}_p^2 = -\mathcal{I}_p^{\boldsymbol{\beta\beta}-1}\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^*}\left(\mathcal{I}_p^{\mathbf{ff}} - \mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^{*\top}}\mathcal{I}_p^{\boldsymbol{\beta\beta}-1}\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^*}\right)^{-1},$$
$$\mathcal{I}_p^3 = -\mathcal{I}_p^{\mathbf{ff}-1}\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^{*\top}}\left(\mathcal{I}_p^{\boldsymbol{\beta\beta}} - \mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^*}\mathcal{I}_p^{\mathbf{ff}-1}\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^{*\top}}\right)^{-1}$$

y

$$\mathcal{I}_p^4 = \left(\mathcal{I}_p^{\mathbf{ff}} - \mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^{*\top}}\mathcal{I}_p^{\boldsymbol{\beta\beta}-1}\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^*}\right)^{-1},$$

en que

$$\mathcal{I}_p^{\boldsymbol{\beta\beta}} = \sum_{i=1}^{n} \frac{4\mathbf{d}_{g_i}}{m}\mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i,$$
$$\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^*} = \left(\sum_{i=1}^{n} \frac{4\mathbf{d}_{g_i}}{m}\mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(1)} \ \ldots \ \sum_{i=1}^{n} \frac{4\mathbf{d}_{g_i}}{m}\mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(m)}\right),$$

$$\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^{*\top}} = \begin{pmatrix} \sum_{i=1}^{n} \frac{4\mathbf{d}_{g_i}}{m}\mathbf{N}_i^{(1)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i \\ \vdots \\ \sum_{i=1}^{n} \frac{4\mathbf{d}_{g_i}}{m}\mathbf{N}_i^{(m)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i \end{pmatrix}$$

y

$$\mathcal{I}_p^{\mathbf{ff}} = \begin{pmatrix} \sum_{i=1}^{n} \frac{4\mathbf{d}_{g_i}}{m}\mathbf{N}_i^{(1)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(1)} + \alpha_1\mathbf{K}_1 & \cdots & \sum_{i=1}^{n} \frac{4\mathbf{d}_{g_i}}{m}\mathbf{N}_i^{(1)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(m)} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} \frac{4\mathbf{d}_{g_i}}{m}\mathbf{N}_i^{(m)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(1)} & \cdots & \sum_{i=1}^{n} \frac{4\mathbf{d}_{g_i}}{m}\mathbf{N}_i^{(m)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(m)} + \alpha_m\mathbf{K}_m \end{pmatrix}.$$

Por lo tanto, si se desea realizar inferencia para cada uno de los elementos del vector de parámetros $\boldsymbol{\theta}$, las matrices de varianza-covarianza aproximada pueden ser estimadas de

la siguiente manera:

$$\widehat{\mathrm{Var}}_{\mathrm{aprox}}(\widehat{\boldsymbol{\beta}}) = \left\{\mathcal{I}_p^{\boldsymbol{\beta}\boldsymbol{\beta}} - \mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^*}\mathcal{I}_p^{\mathbf{f}\mathbf{f}-1}\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^*\top}\right\}^{-1}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},$$

$$\widehat{\mathrm{Var}}_{\mathrm{aprox}}(\widehat{\mathbf{f}}_{(1)},\ldots,\widehat{\mathbf{f}}_{(m)}) = \{\mathcal{I}_p^{\mathbf{f}\mathbf{f}} - \mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^*\top}\mathcal{I}_p^{\boldsymbol{\beta}\boldsymbol{\beta}-1}\mathcal{I}_p^{\boldsymbol{\beta}\mathbf{f}^*}\}^{-1}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

y

$$\widehat{\mathrm{Var}}_{\mathrm{aprox}}(\widehat{\boldsymbol{\tau}}) = \mathcal{I}_p^{\boldsymbol{\tau}\boldsymbol{\tau}-1}\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}.$$

3.2.2. Bandas de confianza

En este trabajo de tesis se propone utilizar las bandas de confianza estudiadas por Wahba (1983) y empleadas posteriormente por Ibacache et al. (2013) en el contexto semiparamétrico; esto es,

$$BC(f_j(t_i)) = \widehat{f}_j(t_i) \pm 2\sqrt{\widehat{\mathrm{Var}}_{\mathrm{aprox}}\widehat{f}_j(t_i)},$$

donde $\widehat{\mathrm{Var}}_{\mathrm{aprox}}\widehat{f}_j(t_i)$ denota la varianza aproximada estimada de $\widehat{f}_j(t_i)$, para $i = 1,\ldots,n$ y $j = 1,\ldots,m$.

3.2.3. Efectos aleatorios

Uno de los aspectos importantes de los modelos lineales mixtos tiene relación con la predicción del vector de efectos aleatorios $\boldsymbol{b}_i$, el cual se asume que sigue una distribución elíptica multivariada.

En este trabajo se propone utilizar el método de Bayes empírico para predecir los efectos aleatorios, siguiendo la misma idea empleada por Laird y Ware (1982) e Ibacahe et al. (2012) en los modelos lineales mixtos semiparamétricos elípticos. Así, se tiene que la distribución de $\mathbf{b}_i|\mathbf{y}_i$ es dada por

$$\mathbf{b}_i|\mathbf{y}_i \sim \mathcal{E}l_q(\mathbf{D}\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i, \mathbf{D} - \mathbf{D}\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_i\mathbf{D}).$$

Por lo tanto, asumiendo que $\boldsymbol{\Sigma}_i$ es fijo, el estimador empírico de Bayes de $\mathbf{b}_i$ toma la

siguiente forma:

$$\widehat{\mathbf{b}}_i = \mathrm{E}(\widehat{\mathbf{b}_i|\mathbf{y}_i}) = \mathbf{D}\mathbf{Z}_i^\top \widehat{\boldsymbol{\Sigma}}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}} - \sum_{j=1}^{m} \mathbf{N}_i^{(j)}\widehat{\mathbf{f}}_{(j)} \right), \tag{3.10}$$

donde $\widehat{\mathbf{f}}_{(j)}$, $\widehat{\boldsymbol{\beta}}$ y $\widehat{\boldsymbol{\Sigma}}_i$ son los estimadores de máxima verosimilitud penalizada. Por lo tanto, el vector de efectos aleatorios estimado es dado por $\widehat{\mathbf{b}} = (\widehat{\mathbf{b}}_1^\top, \ldots, \widehat{\mathbf{b}}_n^\top)^\top$.

Consecuentemente, el valor ajustado para el i-ésimo clúster se puede expresar de la siguiente forma:

$$\begin{aligned} \widehat{\mathbf{y}}_i &= \mathbf{X}_i\boldsymbol{\beta} + \sum_{i=1}^{n} \mathbf{N}_i^{(j)}\widehat{\mathbf{f}}_{(j)} + \mathbf{Z}_i\widehat{\mathbf{b}}_i \\ &= \widehat{\mathbf{V}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1} \left(\mathbf{X}_i\boldsymbol{\beta} + \sum_{i=1}^{n} \mathbf{N}_i^{(j)}\widehat{\mathbf{f}}_{(j)} \right) + \left(\mathbf{I}_m - \widehat{\mathbf{V}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1} \right) \mathbf{y}_i. \end{aligned} \tag{3.11}$$

Note que se puede interpretar de la ecuación (3.11), que $\widehat{\mathbf{y}}_i$ corresponde a la media entre el perfil de la población $\left(\mathbf{X}_i\widehat{\boldsymbol{\beta}} + \sum_{i=1}^{n} \mathbf{N}_i^{(j)}\widehat{\mathbf{f}}_{(j)} \right)$ y los datos observados $\mathbf{y}_i$, con pesos $\widehat{\mathbf{V}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1}$ e $\left(\mathbf{I}_m - \widehat{\mathbf{V}}_i\widehat{\boldsymbol{\Sigma}}_i^{-1} \right)$, respectivamente.

3.3. Selección de los parámetros de suavizamiento

En el proceso de estimación e inferencia se consideraron los parámetros de suavizamiento $\alpha_1, \ldots, \alpha_m$ fijos para realizar inferencia sobre las funciones $\mathbf{f}_{(1)}, \ldots, \mathbf{f}_{(m)}$. Sin embargo, en la práctica estos parámetros deben ser selecccionados a través de algún método.

En el contexto semiparamétrico Green y Silverman (1994) proponen dos maneras de escoger los parámetros de suavizamiento. La primera es considerar una forma libre de escoger los parámetros, que generalmente son escogidos después de probar varios valores para un mismo conjunto de datos y analizando las curvas obtenidas. Una segunda manera es utilizar algún método automático en el que los parámetros de suavizamiento se obtienen a través de los datos. Otra manera de seleccionar los parámetros, es aplicar el criterio de Akaike (1973), ver por ejemplo, Rigby y Stasinopoulos (2005).

En esta sección se detalla el método automático de validación cruzada generalizada

(Craven y Wahba, 1979) y el criterio de Akaike.

3.3.1. Validación cruzada generalizada

La validación cruzada generalizada es uno de los procedimientos automáticos más conocidos, y su objetivo es estimar el error de predicción de un modelo ajustado a los datos. Bajo el MMVS (cuando $j = 2$), la expresión del escore de validación cruzada generalizada es dada por

$$\text{VCG}(\boldsymbol{\alpha}) = \frac{\|\mathbf{y} - \widehat{\mathbf{y}}\|^2}{\left(1 - n^{-1}\text{tr}\left\{\mathbf{H}(\boldsymbol{\alpha})\right\}\right)^2}, \tag{3.12}$$

donde $\mathbf{H}(\boldsymbol{\alpha})$ es una matriz tal que $\widehat{\mathbf{y}} = \mathbf{H}(\boldsymbol{\alpha})\mathbf{y}$, con $\mathbf{y}$ definida en la Subsección 2.2 y $\mathbf{H}(\boldsymbol{\alpha}) = \mathbf{H}_1 + \mathbf{H}_2 + \mathbf{H}_3$, en que

$$\mathbf{H}_1 = \left\{\sum_{i=1}^{n} \mathbf{N}_i^{(1)\top}\mathbf{W}_i\mathbf{A}_1\mathbf{N}_i^{(1)} + \frac{\alpha_1}{n}\mathbf{K}_1\right\}^{-1}\left\{\sum_{i=1}^{n} \mathbf{N}_i^{(1)\top}\mathbf{W}_i\mathbf{A}_1\right\},$$
$$\mathbf{H}_2 = \left\{\sum_{i=1}^{n} \mathbf{N}_i^{(2)\top}\mathbf{W}_i\mathbf{A}_2\mathbf{N}_i^{(2)} + \frac{\alpha_2}{n}\mathbf{K}_2\right\}^{-1}\left\{\sum_{i=1}^{n} \mathbf{N}_i^{(2)\top}\mathbf{W}_i\mathbf{A}_2\right\}$$

y

$$\mathbf{H}_3 = \left\{\sum_{i=1}^{n} \mathbf{X}_i^{\top}(\mathbf{I}_m - \mathbf{B}_2)\mathbf{X}_i\right\}^{-1}\left\{\sum_{i=1}^{n} \mathbf{X}_i^{\top}(\mathbf{I}_m - \mathbf{B}_2)\right\}.$$

Además, las matrices $\mathbf{A}_1$, $\mathbf{A}_2$ y $\mathbf{B}_2$, respectivamente, están dadas por

$$\mathbf{A}_1 = \left\{\mathbf{I}_m - \mathbf{N}_i^{(2)}\mathbf{D}_2\sum_{i=1}^{n} \mathbf{N}_i^{(2)\top}\mathbf{W}_i\right\},$$
$$\mathbf{A}_2 = \left\{\mathbf{I}_m - \mathbf{N}_i^{(1)}\mathbf{D}_1\sum_{i=1}^{n} \mathbf{N}_i^{(1)\top}\mathbf{W}_i\right\},$$

y

$$\mathbf{B}_2 = \left(\mathbf{N}_i^{(1)}\ \mathbf{N}_i^{(2)}\right)\mathbf{C}^{-1}\left(\mathbf{N}_i^{(1)}\ \mathbf{N}_i^{(2)}\right)^{\top}\mathbf{W}_i,$$

en que

$$\mathbf{C} = \begin{pmatrix} \sum_{i=1}^{n} \left\{ \mathbf{N}_i^{(1)\top} \mathbf{W}_i \mathbf{N}_i^{(1)} + \frac{\alpha_1}{n} \mathbf{K}_1 \right\} & \sum_{i=1}^{n} \left\{ \mathbf{N}_i^{(1)\top} \mathbf{W}_i \mathbf{N}_i^{(2)} \right\} \\ \sum_{i=1}^{n} \left\{ \mathbf{N}_i^{(2)\top} \mathbf{W}_i \mathbf{N}_i^{(1)} \right\} & \sum_{i=1}^{n} \left\{ \mathbf{N}_i^{(2)\top} \mathbf{W}_i \mathbf{N}_i^{(2)} + \frac{\alpha_2}{n} \mathbf{K}_2 \right\} \end{pmatrix},$$

para $i = 1, \ldots, m$ y $j = 1, 2$.

La idea de la validación cruzada generalizada es escoger el vector de parámetros $\boldsymbol{\alpha}$ que minimice el escore VCG($\boldsymbol{\alpha}$). Sin embargo, no se puede garantizar que la función escore VCG($\boldsymbol{\alpha}$) tiene un único mínimo, así que se debe tener cuidado con su minimización.

3.3.2. Criterio de Akaike

El criterio de Akaike consiste en seleccionar el vector de parámetros de suavizamiento que minimicen la siguiente expresión:

$$AIC(\boldsymbol{\alpha}) = -2L_p(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha})\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} + 2\text{tr}\left\{\mathbf{H}(\boldsymbol{\alpha})\right\},$$

donde $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$, $L_p(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha})$ es el logaritmo de la función de verosimilitud penalizada y $\text{tr}\left\{\mathbf{H}(\boldsymbol{\alpha})\right\}$ son la cantidad de parámetros involucrados en el modelo.

3.4. Calidad del ajuste

En esta Sección se sugieren los gráficos de las distancias transformadas para evaluar el ajuste de los modelos normal y t-Student multivariados. Estos han sido utilizados previamente por Lange et al. (1989), Borssoi (2014) y de Castro y Galea (2010). En el caso normal, se tiene que la distancia de Mahalanobis

$$\delta_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \tag{3.13}$$

sigue una distribución chi-cuadrado con m grados de libertad, es decir, $\delta_i \sim \chi^2_m$. Análogamente, se puede probar que $\widehat{\delta}_i = (\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i)^\top \widehat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i)$ también sigue una distribución asintótica χ^2_m, para $i = 1, \ldots, n$. Luego, al utilizar la transformación de Wilson-Hilferty

se obtienen las siguientes distancias transformadas:

$$\mathrm{d}_i^{[N]} = \frac{\left\{(\widehat{\delta}_i/m)^{\frac{1}{3}} - (1 - \frac{2}{9\,m})\right\}}{(2/9)^{\frac{1}{2}}}, \tag{3.14}$$

las cuales siguen una distribución normal estándar, es decir, $d_i^{[N]} \sim \mathrm{N}(0,1)$.

Para el caso t-Student multivariado, se tiene que $\mathrm{F}_i = \frac{\delta_i}{m}$ sigue una distribución $\mathrm{F}_{(m,v)}$. Además, $\widehat{\mathrm{F}}_i$ tiene asintóticamente la misma distribución de F_i, para $i = 1, \ldots, n$. De manera análoga al caso normal, al aplicar la transformación Wilson-Hilferty, se obtiene

$$\mathrm{d}_i^{[t]} = \frac{(1 - \frac{2}{9\,m})\widehat{F}_i^{\frac{1}{3}} - (1 - \frac{2}{9\,m})}{\left\{\left(\frac{2}{9\,m}\right)\widehat{F}_i^{\frac{2}{3}} + \left(\frac{2}{9\,m}\right)\right\}^{1/2}}, \tag{3.15}$$

que sigue aproximadamente una distribución normal estándar, es decir, $d_i^{[t]} \sim \mathrm{N}(0,1)$, para $i = 1, \ldots, n$. Los gráficos normales de probabilidad de las distancias transformadas $\mathrm{d}_i^{[N]}$ y $\mathrm{d}_i^{[t]}$ se pueden utilizar para evaluar la calidad del ajuste de los modelos mixtos vector spline bajo distribuciones normal y t-Student, respectivamente.

Capítulo 4

Aplicación

En este capítulo se presenta una aplicación del modelo mixto vector spline bajo distribuciones elípticas a un conjunto de datos reales. En la primera sección, se realiza una breve descripción de los datos. Luego, en la segunda y tercera sección, se propone y ajusta un modelo no paramétrico bajo distribuciones elípticas para los datos, con el objeto de introducir el modelo vector spline. Finalmente, en la cuarta sección, se propone y ajusta un modelo vector spline.

4.1. Descripción de los datos

Para la aplicación se considera un conjunto de datos reales correspondientes a los fondos de pensiones Chilenos. Estos datos corresponden a la rentabilidad mensual de los fondos de pensiones de dos AFP, Habitat y Planvital, en el periodo comprendido entre Enero del 2009 a Diciembre del 2013. Las rentabilidades mensuales de ambas AFP corresponden al vector de respuestas, y el Índice de Precios Selectivos de Acciones (IPSA) a la variable explicativa (paramétrica) en el modelo vector spline.

En la Tabla 4.1 se presentan algunas estadísticas descriptivas básicas, tales como: media, mínimo, máximo, percentil 25, mediana, percentil 75 y coeficiente de variación para las variables IPSA, AFP's Habitat y Planvital. Se observa de esta tabla que la media y mediana de la AFP Habitat son bastante similares, indicando posible simetría en la distribución de los datos. Un comportamiento similar se observa con los datos de la AFP Planvital y la variable IPSA.

Tabla 4.1: Estadísticas descriptivas para las variables IPSA, AFP Habitat y AFP Planvital.

Estadístico descriptivo	AFP Habitat	AFP Planvital	IPSA
Media	0,8777	0,8960	0,0117
Desviación estándar	0,8598	0,8070	0,0473
Mínimo	-0,8938	-0,7121	-0,0929
Máximo	2.9761	2.9417	0,1534
Percentil 25	0,2944	0,2389	-0,0230
Mediana	0,7502	0,8612	0,0077
Percentil 75	1,4113	1,4779	0,0429
Coeficiente de Variación	0,9795	0,9007	4,0334

En la Figura 4.1 se presentan los gráficos box-plot para cada una de las variables en estudio. Se observa del gráfico izquierdo de la Figura 4.1, que las variables presentan cierto grado de simetría, confirmando lo mencionado de la Tabla 4.1. Esto mismo se presenta en el gráfico derecho de la Figura 4.1 con la variable IPSA.

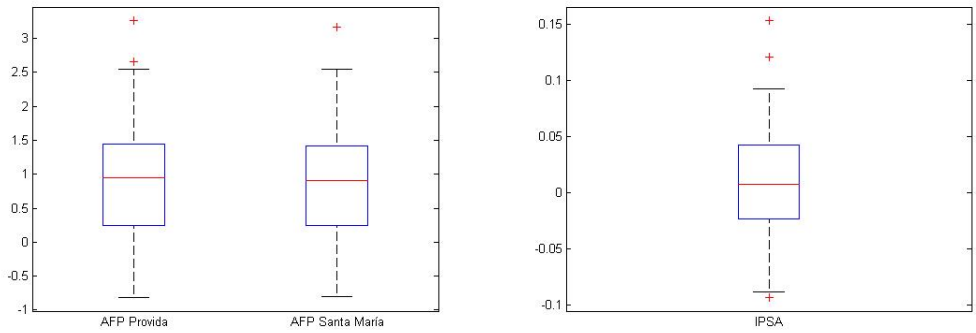


Figura 4.1: Box-plot para las variables IPSA, AFP Habitat y Planvital.

En el gráfico izquierdo de la Figura 4.2, se muestra el diagrama de dispersión entre la rentabilidad de la AFP Habitat e IPSA, donde se observa una tendencia lineal en los datos y consecuentemente, se sugiere incorporar el efecto de la variable IPSA de manera paramétrica. Por otro lado, el gráfico derecho de la Figura 4.2 muestra el comportamiento de las rentabilidades de la AFP Habitat a través del tiempo, donde se observa una ten-

dencia no lineal que sugiere incorporar el efecto del tiempo a través de una función no paramétrica.

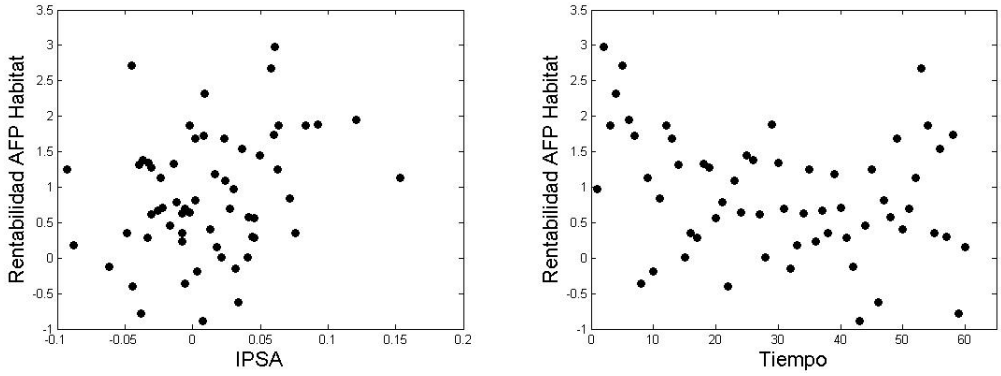


Figura 4.2: Diagrama de dispersión entre las rentabilidades de la AFP Habitat y el IPSA (izquierda), y las rentabilidades de la AFP Habitat y el tiempo.

En el gráfico izquierdo de la Figura 4.3, se muestra el diagrama de dispersión entre las rentabilidades de la AFP Planvital y el IPSA, donde se observa una tendencia lineal en los datos, lo que sugiere incorporar el efecto de la variable IPSA de manera paramétrica. Por otro lado, el gráfico derecho de la Figura 4.3 muestra el comportamiento de las rentabilidades de la AFP Planvital a través del tiempo, donde se observa una tendencia no lineal que sugiere cuantificar este efecto a través de una función no paramétrica.

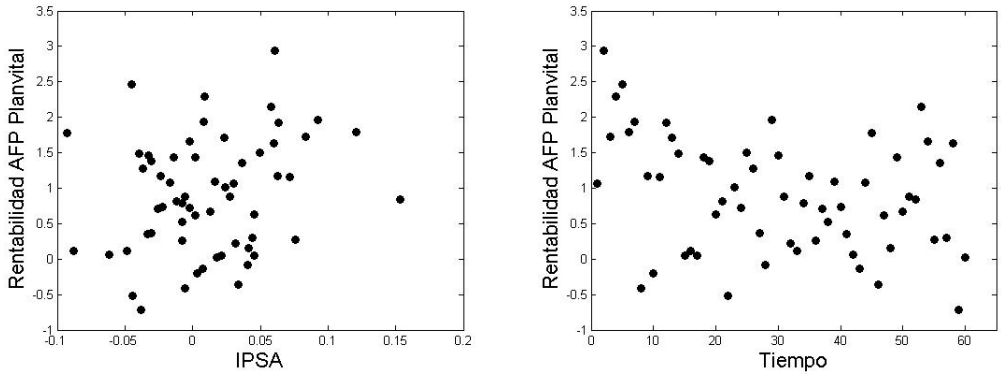


Figura 4.3: Diagrama de dispersión entre las rentabilidades de la AFP Planvital y el IPSA (izquierda), y las rentabilidades de la AFP Habitat y el tiempo.

4.2. Modelo no paramétrico para AFP Habitat

Inicialmente se propone ajustar el siguiente modelo no paramétrico a los datos de la AFP Habitat:

$$y_i = x_i\boldsymbol{\beta} + f(t_i) + \varepsilon_i, \tag{4.1}$$

donde y_i denota el valor observado de la rentabilidad de la AFP Habitat en el tiempo t_i, x_i denota la rentabilidad del IPSA en el mismo período, $\boldsymbol{\beta}$ es el parámetro asociado al coeficiente de regresión que determina el incremento en el valor de la rentabilidad AFP Habitat cuando el IPSA aumenta en una unidad, $f(\cdot)$ es una función arbitraria que depende del tiempo, y ε_i es un error aleatorio ($i = 1, \ldots, 60$). El modelo (4.1) puede ser expresado vectorialmente de la siguiente forma:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + \boldsymbol{\epsilon},$$

en que $\mathbf{y}$ es un vector (60×1) de respuestas observadas, $\mathbf{X}$ es un vector (60×1) cuyos elementos corresponden a los valores de la variable explicativa x, $\mathbf{N}$ es una matriz de incidencia (60×60) que, en este caso, es igual a la matriz identidad, $\mathbf{f}$ es un vector (60×1) cuyos componentes corresponden a la evaluación funcional de $f(\cdot)$ en los valores de la variable explicativa t que pertenecen al conjunto $\{t_1 = 1, t_2 = 2, \ldots, t_{60} = 60\}$, y $\boldsymbol{\epsilon}$ representa un vector (60×1) de errores aleatorios.

En este caso, se asume que el valor observado de la rentabilidad en el tiempo t_i ($i = 1, \ldots, 60$) sigue una distribución elíptica, es decir

$$y_i \sim \mathcal{E}l_1(x_i\boldsymbol{\beta} + f(t_i),\ \phi).$$

De esta manera, el vector de parámetros que debe ser estimado es $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{f}^\top, \phi)^\top$. Con el objeto de comparar las estimaciones de máxima verosimilitud penalizada, se consideran las distribuciones normal y t-Student.

El parámetro de suavizamiento α fue escogido a través del método de validación cruzada generalizada para ambos modelos. Se escogieron $\nu = 5$ grados de libertad para asegurar la existencia de los cuatros primeros momentos en la distribución t-Student. El error estándar de los estimadores del vector de parámetros $\boldsymbol{\theta}$ fueron estimados a través de

la matriz de información de Fisher penalizada. Estos resultados se presentan en la Tabla 4.2.

Tabla 4.2: Estimaciones de máxima verosimilitud penalizada, error estándar (e.e) y $L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$, para los modelos normal y t-Student con $\nu = 5$ grados de libertad ajustados a los datos de la AFP Habitat.

	normal			t-Student		
Parámetro	estimación	e.e	$L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$	estimación	e.e	$L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$
$\boldsymbol{\beta}$	3,3390	4,7336	-67,4754	2,9949	11,3014	-69,0308
ϕ	0,5478	0,01		0,4098	0,009	
α	1113,4			1113,4		

Se nota de la Tabla 4.2, que los parámetros de pendiente y escala son diferentes en las estimaciones de los modelos ajustados, y que el error estándar de $\widehat{\phi}$ es más pequeño en el modelo t-Student que en el normal. Lo contrario ocurre con el error estándar de $\widehat{\boldsymbol{\beta}}$.

Debido a la dimensión alta de $\mathbf{f}$ en ambos modelos, las estimaciones de máxima verosimilitud penalizada y error estándar son presentadas mediante un gráfico de ajuste y bandas de confianza, en las Figuras 4.4 y 4.5, respectivamente. Los gráficos de la Figura 4.5 exhiben la curva estimada y sus correspondientes bandas de confianza bajo ambos modelos. A través de estos gráficos, se observa que el ajuste del modelo normal parece ser más adecuado.

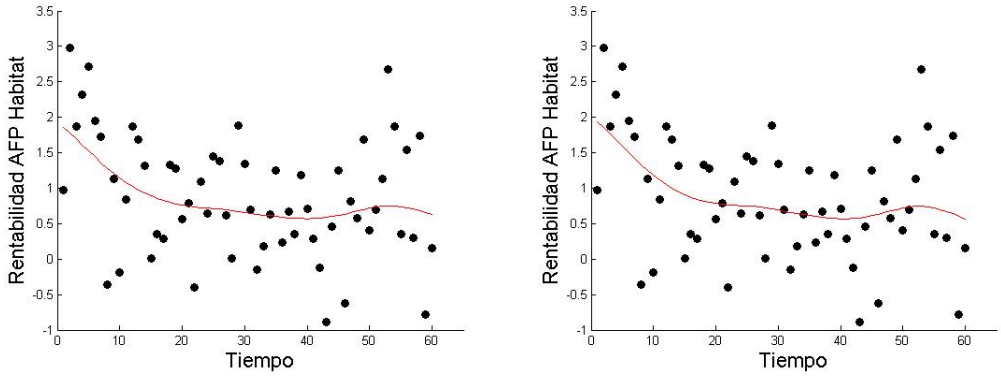


Figura 4.4: Gráficos del ajuste no paramétrico a los datos de la AFP Habitat bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad, respectivamente.

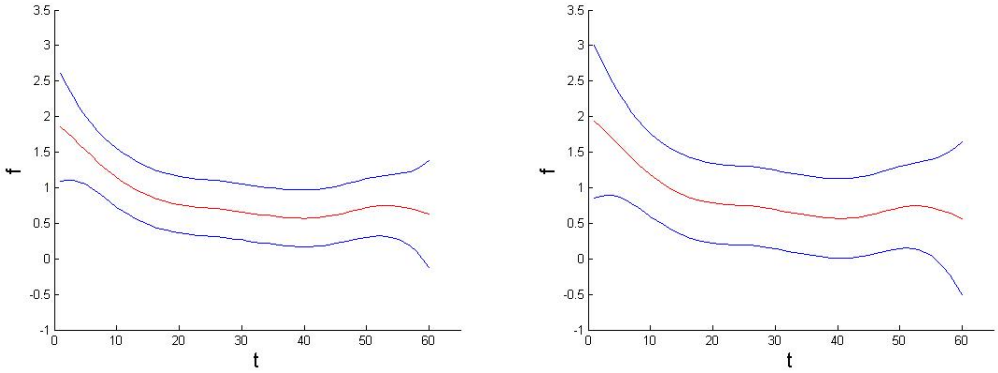


Figura 4.5: Gráficos de la curva estimada y sus bandas de confianza bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad, respectivamente.

Para evaluar el ajuste de los modelos normal y t-Student, se construyeron los gráficos de distancias transformadas sugeridas en la Sección 3.4, para el caso univariado. Basados en los valores $L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$ de la Tabla 4.2 y los gráficos normales de probabilidad de las distancias transformadas presentadas en la Figura 4.6, no hay certeza de que un modelo sea más adecuado que el otro.

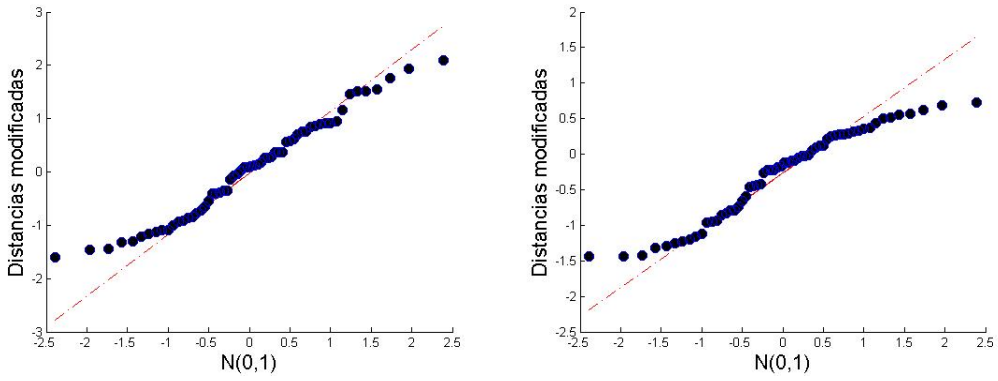


Figura 4.6: Gráficos normales de probabilidad para las distancias transformadas, bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad ajustados a los datos de la AFP Habitat.

4.3. Modelo no paramétrico para AFP Planvital

Inicialmente se propone ajustar el siguiente modelo no paramétrico a los datos de la AFP Planvital:

$$y_i = x_i\boldsymbol{\beta} + f(t_i) + \varepsilon_i, \tag{4.2}$$

donde y_i denota el valor observado de la rentabilidad de la AFP Planvital en el tiempo t_i, x_i denota la rentabilidad del IPSA en el mismo período, $\boldsymbol{\beta}$ es el parámetro asociado al coeficiente de regresión que determina el incremento en el valor de la rentabilidad AFP Planvital cuando el IPSA aumenta en una unidad, $f(\cdot)$ es una función arbitraria que depende del tiempo, y ε_i es un error aleatorio ($i = 1, \ldots, 60$). El modelo (4.2) puede ser expresado vectorialmente de la siguiente forma:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + \boldsymbol{\epsilon},$$

en que $\mathbf{y}$ es un vector (60×1) de respuestas observadas, $\mathbf{X}$ es un vector (60×1) cuyos elementos corresponden a los valores de la variable explicativa x, $\mathbf{N}$ es una matriz de incidencia (60×60) que, en este caso, es igual a la matriz identidad, $\mathbf{f}$ es un vector (60×1) cuyos componentes corresponden a la evaluación funcional de $f(\cdot)$ en los valores de la variable explicativa t que pertenecen al conjunto $\{t_1 = 1, t_2 = 2, \ldots, t_{60} = 60\}$, y $\boldsymbol{\epsilon}$ representa un vector (60×1) de errores aleatorios.

En este caso, se asume que el valor observado de la rentabilidad en el tiempo t_i ($i = 1, \ldots, 60$) sigue una distribución elíptica, es decir

$$y_i \sim \mathcal{E}l_1(x_i\boldsymbol{\beta} + f(t_i),\ \phi).$$

De manera análoga a la Sección **4.2**, el vector de parámetros que debe ser estimado es $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{f}^\top, \phi)^\top$. Con el objeto de comparar las estimaciones de máxima verosimilitud penalizada, se consideran las distribuciones normal y t-Student.

El parámetro de suavizamiento α fue escogido a través del método de validación cruzada generalizada. Se escogieron $\nu = 5$ grados de libertad para asegurar la existencia de los cuatros primeros momentos en la distribución t-Student. El error estándar de los estimadores del vector de parámetros $\boldsymbol{\theta}$ fueron estimados a través de la matriz de Información

de Fisher penalizada. Estos resultados se presentan en la Tabla 4.3.

Tabla 4.3: Estimaciones de máxima verosimilitud penalizada, error estándar (e.e) y $L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$, para los modelos normal y t-Student con $\nu = 5$ grados de libertad ajustados a los datos de la AFP Planvital.

	normal			t-Student		
Parámetro	estimación	e.e	$L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$	estimación	e.e	$L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$
$\boldsymbol{\beta}$	3,2588	4,2715	-66,3547	3,1230	10,4287	-68,3316
ϕ	0,5306	0,0094		0,4080	0,0089	
α	33449			33449		

Se nota de la Tabla 4.3, que los parámetros de pendiente y escala son diferentes en las estimaciones de los modelos ajustados, y que el error estándar de $\widehat{\phi}$ es más pequeño en el modelo t-Student que en el normal. Lo contrario ocurre con el error estándar de $\widehat{\boldsymbol{\beta}}$.

Debido a la dimensión alta de $\mathbf{f}$ en ambos modelos, las estimaciones de máxima verosimilitud penalizada y error estándar son presentadas mediante un gráfico de ajuste y bandas de confianza, en las Figuras 4.7 y 4.8, respectivamente. Los gráficos de la Figura 4.8 exhiben la curva estimada y sus correspondientes bandas de confianza bajo ambos modelos. A través de estos gráficos, se observa que el ajuste del modelo normal parece ser más adecuado.

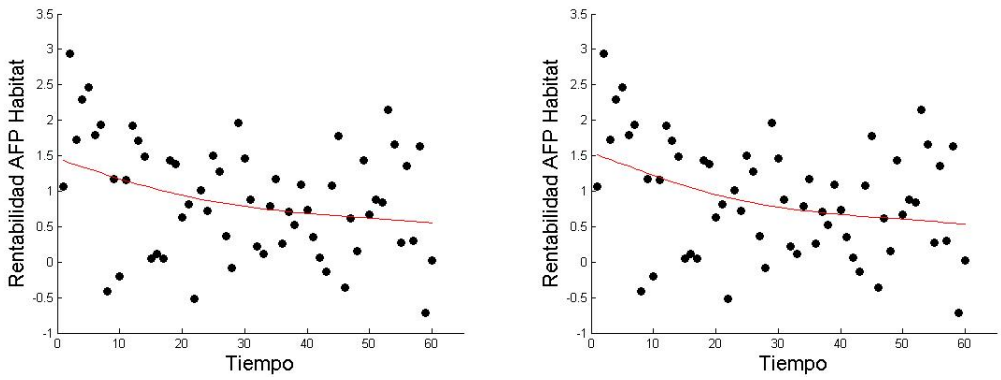


Figura 4.7: Gráficos del ajuste no paramétrico a los datos de la AFP Habitat bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad, respectivamente.

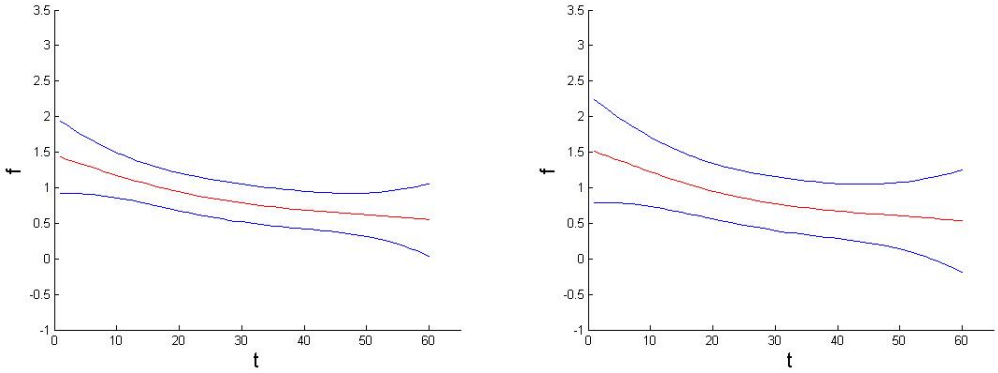


Figura 4.8: Gráficos de la curva estimada y sus bandas de confianza bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad, respectivamente.

De manera análoga a la Subsección **4.2.1**, se construyeron los gráficos de distancias transformadas para los modelos normal y t-Student. Basados en los valores $L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$ de la Tabla 4.3 y los gráficos normales de probabilidad de las distancias transformadas presentadas en la Figura 4.9, existe un grado de certeza de que el modelo normal es más adecuado.

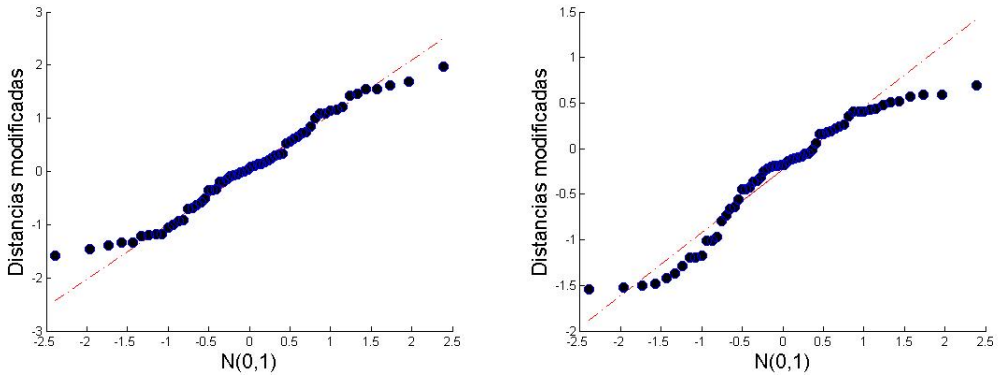


Figura 4.9: Gráficos normales de probabilidad para las distancias transformadas, bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad ajustados a los datos de la AFP Planvital.

4.4. Modelo vector spline

En esta sección, se propone un modelo que permite analizar de manera conjunta los modelos propuestos en las Secciones 4.2 y 4.3, es decir, se propone ajustar el siguiente modelo vector spline:

$$y_{ij} = x_i\boldsymbol{\beta} + f_j(t_i) + \varepsilon_{ij}, \tag{4.3}$$

donde y_{ij} denota la j-ésima rentabilidad de AFP asociada a la i-ésima rentabilidad del IPSA en el periodo t_i, x_i denota la rentabilidad del IPSA en el mismo periodo, $\boldsymbol{\beta}$ es el parámetro asociado al coeficiente de regresión que determina el incremento en el valor de rentabilidad de la AFP cuando el IPSA aumenta en una unidad, f_j es la j-ésima función arbitraria que depende del tiempo t_i, y ε_{ij} es un error aleatorio ($i = 1, \ldots, 60$ y $j = 1, 2$).

El modelo (4.3) puede ser expresado de la siguiente forma

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{N}_i^{(1)} f_{(1)} + \mathbf{N}_i^{(2)} f_{(2)} + \boldsymbol{\varepsilon}_i, \tag{4.4}$$

donde $\mathbf{y}_i$ es un vector (2×1) que contiene las respuestas observadas de la i-ésima rentabilidad de las AFP, $\mathbf{X}_i = \mathbf{1}x_i$, $\mathbf{N}_i^{(1)}$ y $\mathbf{N}_i^{(2)}$ son matrices de incidencia (2×60), $\mathbf{f}_{(1)}$ y $\mathbf{f}_{(2)}$ son vectores (60×1) cuyas componentes son la funciones evaluadas en los valores del tiempo en el conjunto $\{t_1 = 1, t_2 = 2, \ldots, t_{60} = 60\}$ y $\boldsymbol{\varepsilon}_i$ es un vector (2×1) de errores aleatorios. En este caso, se supone que el valor observado de la rentabilidad en el tiempo t_i $(i = 1, \ldots, 60)$ sigue una distribución elíptica multivariada, es decir

$$\mathbf{y}_i \sim \mathcal{E}l_2(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{N}_i^{(1)} f_{(1)} + \mathbf{N}_i^{(2)} f_{(2)},\ \phi\, \mathbf{I}_{2\times 2}).$$

De esta manera, el vector de parámetros que debe ser estimado es $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{f}_{(1)}^\top, \mathbf{f}_{(2)}^\top, \phi)^\top$. Con el objeto de comparar las estimaciones de máxima verosimilitud penalizada, se consideran las distribuciones normal y t-Student multivariadas.

Los parámetros de suavizamiento que se utilizan en el modelo vector spline, son los considerados en las Secciones 4.1 y 4.2, es decir, $\widehat{\alpha}_1 = 1113,4$ y $\widehat{\alpha}_2 = 33.449$. En tanto a los grados de libertad de la distribución t-Student multivariada, se consideró $\nu = 5$ para asegurar la existencia de los primeros cuatros momentos. El error estándar de los estimadores de coeficientes de regresión, funciones no paramétricas y del parámetro de

escala, se estimaron a través de la matriz de información de Fisher penalizada. Estos resultados se presentan en la Tabla 4.4, sin embargo, debido a la alta dimensión de las funciones no paramétricas, sus estimaciones y error estándar se muestran posteriormente.

Tabla 4.4: Estimaciones de máxima verosimilitud penalizada, error estándar (e.e) y $L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$, para los modelos normal y t-Student multivariados con $\nu = 5$ grados de libertad ajustados a los datos de las AFP.

	normal			t-Student		
Parámetro	estimación	e.e	$L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$	estimación	e.e	$L_p(\widehat{\boldsymbol{\theta}}, \widehat{\alpha})$
β	3,2976	2,2471	-134,4634	2,9329	2,2190	-118,6132
ϕ	0,5392	0,0048		0,4190	0,0051	

Se nota de la Tabla 4.4, que los parámetros de pendiente y escala son diferentes en las estimaciones de los modelos ajustados, y que el error estándar de $\widehat{\phi}$ es más pequeño en el modelo normal que en el t-Student. El caso contrario ocurre con el error estándar de $\widehat{\boldsymbol{\beta}}$.

Los gráficos del ajuste de las funciones no paramétricas para el modelo normal se presentan en la Figura 4.10. En tanto al ajuste de las funciones no paramétricas del modelo t-Student, se exhiben en la Figura 4.11. En ambos casos, se observa que el ajuste del modelo vector spline sería adecuado.

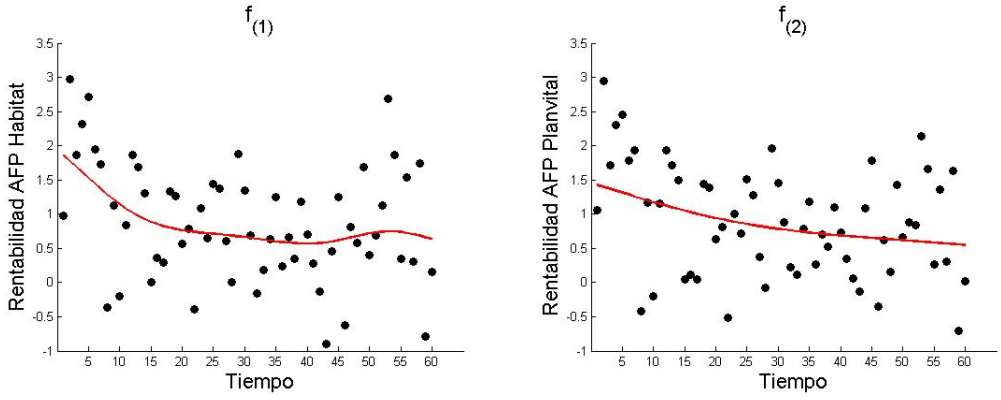


Figura 4.10: Gráficos de los ajustes no paramétricos a los datos de las AFP Habitat y Planvital bajo el modelo normal multivariado, respectivamente.

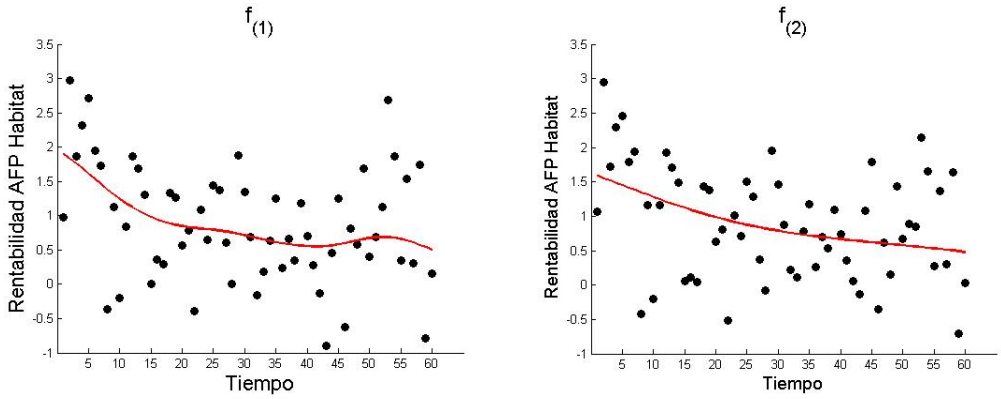


Figura 4.11: Gráficos de los ajustes no paramétricos a los datos de las AFP Habitat y Planvital bajo el modelo t-Student multivariado con $\nu = 5$ grados de libertad, respectivamente.

En la Figuras 4.12 y 4.13 se presentan los gráficos de bandas de confianza para las funciones no paramétricas correspondientes a los modelos normal y t-Student multivariados, respectivamente. Al comparar los gráficos izquierdos de ambas figuras, se observa que las amplitudes de las bandas de confianza son más pequeñas en el modelo normal que en t-Student, y en que los extremos tienen una mayor amplitud debido a la dispersión de los datos. Esto último, se repite al comparar los gráficos derechos de ambas figuras.

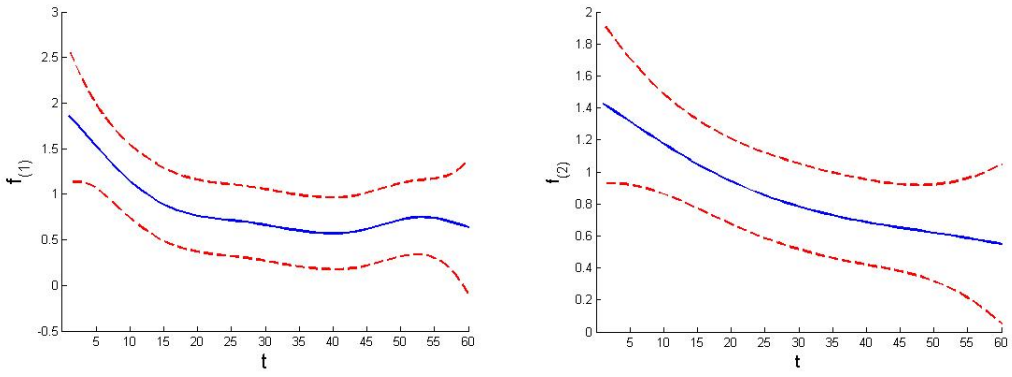


Figura 4.12: Gráficos de la curva estimada y sus bandas de confianza en los datos de las AFP Habitat y Planvital bajo el modelo normal multivariado, respectivamente.

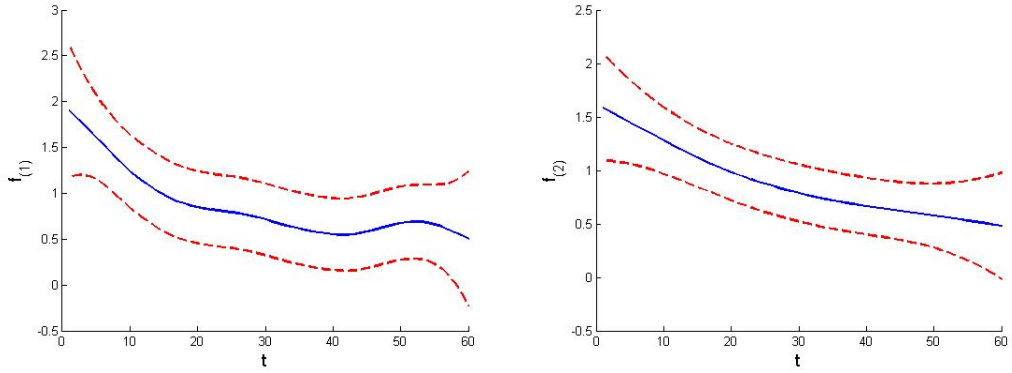


Figura 4.13: Gráficos de las bandas de confianza de la funciones no paramétricas a los datos AFP Habitat y Planvital, sobre el modelo t-Student con $\nu = 5$ grados de libertad.

Basados en los valores $L_p(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}})$ de la Tabla 4.4 y los gráficos de las distancias transformadas presentadas en la Figura 4.14, tenemos que el modelo normal parece presentar un ajuste más adecuado en relación al modelo t-Student con $\nu = 5$ grados de libertad.

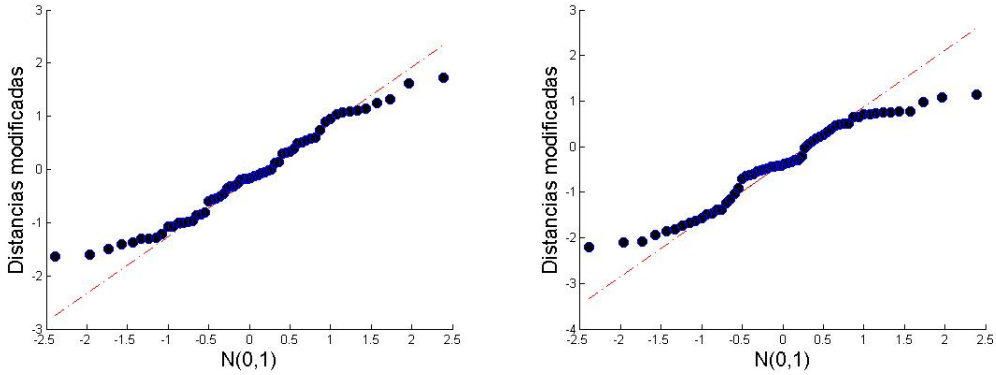


Figura 4.14: Gráficos normales de probabilidad para las distancias transformadas, bajo los modelos normal y t-Student con $\nu = 5$ grados de libertad ajustados a los datos de las AFP.

Capítulo 5

Conclusiones

En este trabajo de tesis se desarrollaron varios aspectos del proceso de estimación e inferencia estadística para el MMVS bajo distribuciones elípticas. A continuación presentamos los principales resultados y conclusiones.

- En el Capítulo, 2 se presentó el MMVS en la clase de distribuciones elípticas de contornos elípticos, asumiendo que el vector de respuestas, vector de efectos aleatorios y el vector de errores aleatorios de cada clúster sigue una distribución conjunta elíptica. Luego, se derivó la distribución marginal del vector de respuestas y su correspondiente función de densidad, permitiendo obtener una expresión analítica para la función de log-verosimilitud penalizada (ecuación 2.17).

- En el Capítulo 3, se derivó la función escore penalizada, las matrices de segundas derivadas e Información de Fisher penalizadas a partir de la función de log-verosimilitud penalizada. Además, se llevó a cabo un estudio de la concavidad de la función de log-verosimilitud penalizada y un procedimiento que a través de maximizaciones sucesivas sobre la misma función se obtienen los EMVP. A partir de este procedimiento, se derivó un proceso iterativo escoring de Fisher y backfitting para estimar los componentes de varianza y fijos (coeficientes de regresión y funciones no paramétricas) del MMVS, respectivamente. Luego, se propuso estimar el error estándar de los EMVP mediante la matriz de información de Fisher penalizada y construir bandas de confianza para las funciones no paramétricas. También, fue sugerido el estimador empírico de Bayes para la predicción de los efectos aleatorios. Finalmente, se propusieron dos métodos para escoger los parámetros de suaviza-

miento y gráficos de distancias modificadas para evaluar la calidad del ajuste del MMVS bajo errores normal y t-Student.

- En el Capítulo 4 se presentó una aplicación con datos reales para ilustrar la aplicabilidad del MMVS. Estos datos corresponden a la rentabilidad mensual de los fondos de pensiones de las AFP Habitat y Planvital, y el Índice de Precios Selectivos de Acciones (IPSA). Se ajustaron los modelos normal y t-Student con $\nu = 5$ grados de libertad y se observó que en ambos casos el ajuste es similar.

- Es importante mencionar que el proceso iterativo basado en el método de máxima verosimilitud penalizada desarrollado para estimar los coeficientes de regresión, funciones no paramétricas y componentes de varianza en el MMVS fue implementado en el software Matlab.

- Finalmente, a través del estudio del MMVS bajo distribuciones elípticas, observamos que esta clase de modelos permiten analizar datos de experimentos relacionados a fenómenos en que es preciso modelar la dependencia de medidas entre unidades muestrales, modelar los efectos de las covariables que contribuyen de manera paramétrica y no paramétrica sobre la variable de respuesta, y extender el modelado a otras distribuciones elípticas además de la normal.

5.1. Trabajos futuros

Una primera perspectiva de trabajo futuro es extender el modelo propuesto en esta tesis, en el sentido de incorporar una estructura de covarianza para los errores aleatorios que tenga en consideración la dependencia en el tiempo de las observaciones. Por ejemplo, considerar una estructura autoregresiva de orden uno, AR(1), o una estructura de equicorrelación, según sea el caso.

Otra perspectiva de trabajo futuro es extender el modelo propuesto asumiendo que los errores aleatorios siguen una distribución asimétrica, por ejemplo, las distribuciones t-skew y normal-skew.

Apéndice A

Derivadas de la log-verosimilitud penalizada

En este apéndice se presentan los detalles de los cálculos de primera y segunda orden del logaritmo de la función de verosimilitud penalizada en el modelo mixto vector spline bajo distribuciones elípticas. Estos cálculos requieren algunos resultados de álgebra y diferenciación de matrices que pueden ser encontrados, por ejemplo, en Magnus y Neudecker (1988).

A.1. Derivadas de primer orden

En el modelo mixto vector spline bajo distribuciones elípticas (2.2) el logaritmo de la función de verosimilitud penalizada es dada por

$$L_p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \tag{A.1}$$

en que

$$L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = L_i(\boldsymbol{\theta}) - \frac{1}{2\,n} \sum_{j=1}^{m} \alpha_j \mathbf{f}_{(j)}^{\top} \mathbf{K}_j \mathbf{f}_{(j)}, \tag{A.2}$$

con $L_i(\boldsymbol{\theta})$ definida por la ecuación (2.10) y $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\top}, \mathbf{f}_{(1)}^{\top}, \ldots, \mathbf{f}_{(m)}^{\top}, \boldsymbol{\tau}^{\top})^{\top}$. Al utilizar resul-

tados de diferenciación de matrices tenemos que,

$$\frac{\partial L_p(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\partial L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}}. \tag{A.3}$$

Al derivar (A.2) con relación a $\boldsymbol{\beta}$, $\mathbf{f}_{(j)}$ y $\boldsymbol{\tau}_\ell$, respectivamente, tenemos que

$$\begin{aligned}
\frac{\partial L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\
&= \frac{\partial \log g(\delta_i)}{\partial \boldsymbol{\beta}} \\
&= \frac{g'(\delta_i)}{g(\delta_i)} \frac{\partial \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i}{\partial \boldsymbol{\beta}} \\
&= \upsilon_i \mathbf{X}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i,
\end{aligned} \tag{A.4}$$

$$\begin{aligned}
\frac{\partial L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \mathbf{f}_{(j)}} &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \mathbf{f}_{(j)}} - \frac{\alpha_j}{2n} \frac{\partial [\mathbf{f}_{(j)}^\top \mathbf{K}_j \mathbf{f}_{(j)}]}{\partial \mathbf{f}_{(j)}} \\
&= \frac{\partial \log g(\delta_i)}{\partial \mathbf{f}_{(j)}} - \frac{\alpha_j}{n} \mathbf{K}_j \mathbf{f}_{(j)} \\
&= \frac{g'(\delta_i)}{g(\delta_i)} \frac{\partial \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i}{\partial \mathbf{f}_{(j)}} - \frac{\alpha_j}{n} \mathbf{K}_j \mathbf{f}_{(j)} \\
&= -2\zeta_g(\delta_i) \mathbf{N}_i^{(j)\top} \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i - \frac{\alpha_j}{n} \mathbf{K}_j \mathbf{f}_{(j)} \\
&= \upsilon_i \mathbf{N}_i^{(j)\top} \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i - \frac{\alpha_j}{n} \mathbf{K}_j \mathbf{f}_{(j)}, \quad j = 1, \ldots, m,
\end{aligned} \tag{A.5}$$

y

$$\begin{aligned}
\frac{\partial L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \boldsymbol{\tau}_\ell} &= \frac{\partial L_i(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}_\ell} \\
&= -\frac{1}{2} \frac{\partial \log |\boldsymbol{\Sigma}_i|}{\partial \boldsymbol{\tau}_\ell} + \frac{\log g(\delta_i)}{\partial \boldsymbol{\tau}_\ell} \\
&= -\frac{1}{2} \mathrm{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell)\} + \zeta_g(\delta_i) \mathbf{r}_i^\top \frac{\partial \boldsymbol{\Sigma}_i^{-1}}{\partial \boldsymbol{\tau}_\ell} \mathbf{r}_i \\
&= -\frac{1}{2} \mathrm{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell)\} - \zeta_g(\delta_i) \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \boldsymbol{\tau}_\ell} \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i \\
&= -\frac{1}{2} \left[\mathrm{tr}\{\boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell)\} - \upsilon_i \mathbf{r}_i^\top \boldsymbol{\Sigma}_i^{-1} \dot{\boldsymbol{\Sigma}}_i(\ell) \boldsymbol{\Sigma}_i^{-1} \mathbf{r}_i \right],
\end{aligned} \tag{A.6}$$

donde $\mathbf{r}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sum_{j=1}^{m}\mathbf{N}_i^{(j)}\mathbf{f}_{(j)}$, $\upsilon_i = -2\zeta_g(\delta_i)$, con $\zeta_g(\delta_i) = \frac{g'(\delta_i)}{g(\delta_i)}$, y denotando $\dot{\boldsymbol{\Sigma}}_i(\ell) = \partial\boldsymbol{\Sigma}_i/\partial\boldsymbol{\tau}_\ell$ para $\ell = 1, \ldots, d+1$. Note que los resultados (A.4), (A.5) y (A.6), son de dimensión $(p \times 1)$, $(n \times 1)$ y $(d^* \times 1)$, respectivamente.

A.2. Derivadas de segundo orden

Al utilizar resultados de diferenciación de matrices, se tiene que la matriz de segundas derivadas en relación a $\boldsymbol{\theta}$ está dada por:

$$\frac{\partial L_p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \sum_{i=1}^{n}\frac{\partial L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}. \tag{A.7}$$

Así, al utilizar la notación $\partial\boldsymbol{\Sigma}_i/\partial\boldsymbol{\tau}_\ell = \dot{\boldsymbol{\Sigma}}_i(\ell)$, y derivar (A.4) con relación a $\boldsymbol{\beta}$, $\mathbf{f}_{(j)}$ y $\boldsymbol{\tau}_\ell$, respectivamente, tenemos que las matrices de derivadas parciales son dadas por

$$\begin{aligned}
\frac{\partial^2 L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} &= \frac{\partial\left[\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\upsilon_i\right]}{\partial\boldsymbol{\beta}^\top} \\
&= \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left[\frac{\partial\mathbf{r}_i}{\partial\boldsymbol{\beta}^\top}\,\upsilon_i + \mathbf{r}_i\,\frac{\partial\upsilon_i}{\partial\boldsymbol{\beta}^\top}\right] \\
&= \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left[\frac{\partial\mathbf{r}_i}{\partial\boldsymbol{\beta}^\top}\,\upsilon_i + \mathbf{r}_i\,\frac{\partial\upsilon_i}{\partial\delta_i}\frac{\partial\delta_i}{\partial\boldsymbol{\beta}^\top}\right] \\
&= -\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left[\upsilon_i\mathbf{X}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right] \\
&= -\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\mathbf{f}_{(j)}^\top} &= \frac{\partial\left[\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\upsilon_i\right]}{\partial\mathbf{f}_{(j)}^\top} \\
&= \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left[\frac{\partial\mathbf{r}_i}{\partial\mathbf{f}_{(j)}^\top}\,\upsilon_i + \mathbf{r}_i\,\frac{\partial\upsilon_i}{\partial\mathbf{f}_{(j)}^\top}\right] \\
&= \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left[\mathbf{r}_i\upsilon_i'\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(j)} - \upsilon_i\mathbf{N}_i^{(j)}\right] \\
&= -\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left[\upsilon_i\mathbf{N}_i^{(j)} + 2\mathbf{r}_i\upsilon_i'\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(j)}\right] \\
&= -\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(j)}, \quad j = 1, \ldots, m,
\end{aligned}$$

y

$$\begin{aligned}\frac{\partial^2 L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\tau}_\ell} &= \frac{\partial\left[\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\upsilon_i\mathbf{r}_i\right]}{\partial\boldsymbol{\tau}_\ell}\\ &= -\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\boldsymbol{\Sigma}_i^{-1}\upsilon_i\mathbf{r}_i + \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\upsilon_i'\,\frac{\partial\left[\mathrm{tr}\{\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\}\right]}{\partial\boldsymbol{\tau}_\ell}\,\mathbf{r}_i\\ &= -\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\boldsymbol{\Sigma}_i^{-1}\upsilon_i\mathbf{r}_i - \mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\\ &= -\mathbf{X}_i^\top\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + \upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i,\quad \ell=1,\ldots,d+1.\end{aligned}$$

De la misma manera, al derivar (A.5) con relación a $\boldsymbol{\beta}$, $\mathbf{f}_{(j)}$ y $\boldsymbol{\tau}_\ell$, respectivamente, tenemos que las matrices de segundas derivadas parciales, para $j=1,\ldots,m$, $k=1,\ldots,m$, $\ell=1,\ldots,d+1$, son dadas por

$$\begin{aligned}\frac{\partial^2 L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\mathbf{f}_{(j)}\partial\boldsymbol{\beta}^\top} &= \frac{\partial\left[\upsilon_i\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i - \frac{\alpha_j}{n}\mathbf{K}_j\mathbf{f}_{(j)}\right]}{\partial\boldsymbol{\beta}^\top}\\ &= \frac{\partial\left[\upsilon_i\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\right]}{\partial\boldsymbol{\beta}^\top}\\ &= \frac{\partial\left[\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\upsilon_i\right]}{\partial\boldsymbol{\beta}^\top}\\ &= \mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\left[\frac{\partial\mathbf{r}_i}{\partial\boldsymbol{\beta}^\top}\upsilon_i + \mathbf{r}_i\frac{\partial\upsilon_i}{\partial\boldsymbol{\beta}^\top}\right]\\ &= -\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\left[\upsilon_i\mathbf{X}_i + 2\mathbf{r}_i\upsilon i'\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right]\\ &= -\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\left[\upsilon_i\boldsymbol{\Sigma}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right]\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i,\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\mathbf{f}_{(j)}\partial\mathbf{f}_{(j)}^\top} &= \frac{\partial\left[\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\upsilon_i\right]}{\partial\mathbf{f}_{(j)}^\top} - \frac{1}{n}\,\frac{\partial\left[\alpha_j\mathbf{K}_j\mathbf{f}_{(j)}\right]}{\partial\mathbf{f}_{(j)}^\top}\\ &= -\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(j)} - \frac{\alpha_j}{n}\mathbf{K}_j,\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 L_{p_i}(\theta,\alpha)}{\partial \mathbf{f}_{(j)}\partial \mathbf{f}_{(k)}^\top} &= \frac{\partial\left[\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\upsilon_i\right]}{\partial \mathbf{f}_{(k)}^\top}\\ &= \mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\,\frac{\mathbf{r}_i}{\partial \mathbf{f}_{(k)}^\top}\upsilon_i + \mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\frac{\partial \upsilon_i}{\partial \mathbf{f}_{(k)}^\top}\\ &= -\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(k)}\upsilon_i - 2\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(k)}\\ &= -\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + 2\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(k)}, \text{para}\quad j\neq k,\end{aligned}$$

y

$$\begin{aligned}\frac{\partial^2 L_{p_i}(\theta,\alpha)}{\partial \mathbf{f}_{(j)}\partial \boldsymbol{\tau}_\ell} &= \frac{\partial\left[\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\upsilon_i\right]}{\partial \boldsymbol{\tau}_\ell}\\ &= -\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i^{-1}(\ell)\boldsymbol{\Sigma}_i^{-1}\upsilon_i\mathbf{r}_i + \mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\upsilon_i'\quad\frac{\partial\left[\text{tr}\{\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\}\right]}{\partial\boldsymbol{\tau}_\ell}\,\mathbf{r}_i\\ &= -\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i^{-1}(\ell)\boldsymbol{\Sigma}_i^{-1}\upsilon_i\mathbf{r}_i - \mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i^{-1}(\ell)\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i\\ &= -\mathbf{N}_i^{(j)\top}\boldsymbol{\Sigma}_i^{-1}\left\{\upsilon_i\boldsymbol{\Sigma}_i + \upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\right\}\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i.\end{aligned}$$

Por último, al utilizar la notación $\upsilon_i' = \frac{\partial \upsilon_i}{\partial \delta_i}$, $\partial^2\boldsymbol{\Sigma}_i/\partial\boldsymbol{\tau}_\ell\partial\boldsymbol{\tau}_\jmath = \ddot{\boldsymbol{\Sigma}}_i(\ell,\jmath)$, y derivar (A.6) con relación a $\boldsymbol{\beta}$, $\mathbf{f}_{(j)}$ y $\boldsymbol{\tau}_\ell$, respectivamente, para $\jmath = 1,\ldots,d+1$, $\ell = 1,\ldots,d+1$, tenemos los siguientes resultados:

$$\frac{\partial^2 L_{p_i}}{\partial\boldsymbol{\tau}_\ell\partial\boldsymbol{\beta}} = -\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\boldsymbol{\Sigma}_i^{-1}\{\upsilon_i\boldsymbol{\Sigma}_i + \upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i,$$

$$\frac{\partial^2 L_{p_i}}{\partial\boldsymbol{\tau}_\ell\partial\mathbf{f}_{(j)}} = -\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\dot{\boldsymbol{\Sigma}}_i(\ell)\boldsymbol{\Sigma}_i^{-1}\{\upsilon_i\boldsymbol{\Sigma}_i + \upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i^{(j)}$$

y

$$\begin{aligned}\frac{\partial^2 L_{p_i}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial\boldsymbol{\tau}_\ell\partial\boldsymbol{\tau}_\jmath} &= \frac{1}{2}\text{tr}\left\{\boldsymbol{\Sigma}_{\text{i}}^{-1}\left[\dot{\boldsymbol{\Sigma}}_{\text{i}}(\jmath)\boldsymbol{\Sigma}_{\text{i}}^{-1}\dot{\boldsymbol{\Sigma}}_{\text{i}}(\ell) - \ddot{\boldsymbol{\Sigma}}_{\text{i}}(\ell,\jmath)\right]\right\}\\ &\quad - \frac{1}{2}\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1}\left\{\dot{\boldsymbol{\Sigma}}_i(\jmath)\boldsymbol{\Sigma}_i^{-1}\left(\upsilon_i'\mathbf{r}_i\mathbf{r}_i^\top\boldsymbol{\Sigma}_i^{-1} + 2\upsilon_i\mathbf{I}_m\right)\dot{\boldsymbol{\Sigma}}_i(\ell) + \upsilon_i\ddot{\boldsymbol{\Sigma}}_i(\ell,\jmath)\right\}\boldsymbol{\Sigma}_i^{-1}\mathbf{r}_i.\end{aligned}$$

Apéndice B

Datos de las AFP

Tabla B.1: Rentabilidades mensuales de los fondos de pensiones de las AFP Habitat y Planvital, y el Índice de precios selectivos de acciones (IPSA) en el periodo comprendido entre Enero del 2009 a Junio del 2010.

t_i	AFP Habitat	AFP Planvital	IPSA
1	0,9783	1,0614	0,0303
2	2,9761	2,9417	0,0606
3	1,8714	1,7194	0,0838
4	2,3228	2,298	0,0088
5	2,719	2,4601	-0,0449
6	1,9465	1,7866	0,1208
7	1,7242	1,9363	0,0079
8	-0,3647	-0,4174	-0,0056
9	1,1298	1,1656	-0,0239
10	-0,1936	-0,2038	0,0035
11	0,8441	1,1569	0,0716
12	1,8672	1,9278	0,0633
13	1,6862	1,7121	0,0232
14	1,3115	1,4919	-0,0392
15	0,0064	0,052	0,0213
16	0,3568	0,112	-0,0485
17	0,2866	0,0497	0,0457
18	1,326	1,4360	-0,0138

Tabla B.2: Rentabilidades mensuales de los fondos de pensiones de las AFP Habitat y Planvital, y el Índice de precios selectivos de acciones (IPSA) en el periodo comprendido entre Julio del 2010 a Diciembre del 2012.

t_i	AFP Habitat	AFP Planvital	IPSA
19	1,2703	1,3810	-0,0306
20	0,5606	0,6301	0,0452
21	0,7910	0,8117	-0,0121
22	-0,3966	-0,5123	-0,0442
23	1,0865	1,0078	0,0243
24	0,6458	0,718	-0,002
25	1,4436	1,5067	0,0498
26	1,379	1,2791	-0,0371
27	0,6119	0,3709	-0,0309
28	0,0052	-0,0804	0,0407
29	1,8803	1,9581	0,0928
30	1,3447	1,458	-0,0323
31	0,6912	0,8791	0,0274
32	-0,154	0,218	0,0317
33	0,1855	0,1125	-0,0877
34	0,6317	0,7833	-0,0076
35	1,2562	1,1753	0,0626
36	0,2335	0,2597	-0,0075
37	0,6669	0,7037	-0,0254
38	0,3516	0,5304	-0,0075
39	1,1879	1,0946	0,0167
40	0,7094	0,7343	-0,0222
41	0,2835	0,3528	-0,0335
42	-0,1248	0,0598	-0,0616
43	-0,8938	-0,1349	0,0075
44	0,4539	1,0819	-0,0165
45	1,2499	1,7841	-0,0929
46	-0,6257	-0,3561	0,0336
47	0,8133	0,6141	0,0019
48	0,5762	0,1505	0,0413

Tabla B.3: Rentabilidades mensuales de los fondos de pensiones de las AFP Habitat y Planvital, y el Índice de precios selectivos de acciones (IPSA) en el periodo comprendido entre Enero del 2013 a Diciembre del 2013.

t_i	AFP Habitat	AFP Planvital	IPSA
49	1,6901	1,4342	0,002
50	0,4069	0,6660	0,0129
51	0,6893	0,8819	-0,0057
52	1,1278	0,8433	0,1534
53	2,6802	2,1417	0,0576
54	1,868	1,6572	-0,0023
55	0,3469	0,2701	0,0759
56	1,5341	1,3618	0,0365
57	0,3022	0,3001	0,0444
58	1,7439	1,6349	0,0597
59	-0,7839	-0,7121	-0,038
60	0,1493	0,0249	0,0181

Referencias bibliográficas

[1] Absanullah, M., Golam, B. M. and Shakil, M. (2014). *Normal and Student's t Distributions and Their Applications*. ATLANTIS PRESS.

[2] Borssoi, A. J. (2014). *Modelos mistos lineares elípticos com erros de medicao*. Tese de Doutorado. IME-USP, Sao Paulo.

[3] Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury \Thompson, Pacific Grove.

[4] de Castro, M. and Galea, M. (2010). Robust inference in an heterocedastic measurement error model. *Journal of the Korean Statistical Society*, **39**, 439-447.

[5] Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerical Mathematical*, **31**, 377-403.

[6] Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symetric Multivariate and Related Distribution*. Chapman and Hall, London.

[7] Fessler, J. A. (1991 a). Nonparametric Fixed-Interval Smoothing with Vector Splines. *IEEE Trans. Signal Process.*, **39**.

[8] Fessler, J. A. (1991 b). Nonparametric Fixed-Interval Smoothing of Nonlinear Vector-Valued Measurements. *IEEE Trans. Signal Process.*, **39**.

[9] Gómez, E., Gómez-Villegas, M. A. and Marín, J. M (1998). A multivariate generalization of the power exponential family of distributions. *COMMUN. STATIST.-THEORY METH.*, **27**, 589-600.

[10] Gourieroux, C. and Monford, A. (1995). *Statistics and econometric models, Vols. 1 and 2*. Cambridge University Press.

[11] Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review*, **55**, 245-259.

[12] Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models:A Roughness Penalty Approach*. Chapman and Hall, Boca Raton.

[13] Gupta, A. K. and Varga, A. T. (1993). *Elliptically Contoured Model in Statistics*. Mathematics and Its Applications, DORDRECHT \BOSTON \LONDON.

[14] Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

[15] He, X., Zhu, Z. and Fung, W. K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, **89**, 579-590.

[16] Heckman, N. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society B*, **48**, 244-248.

[17] Ibacache-Pulgar, G., Paula, G. A. and Cysneiros, F. J. A. (2013). Semiparametric additive models under symmetric distributions. TEST 22, 103-121.

[18] Ibacache-Pulgar, G., and Paula, G. A. (2011). Local influence for Student-t partially linear models. *Computational Statistics and Data Analysis*. **55**, 1462-1478.

[19] Ibacache-Pulgar, G., Paula, G. A. and Galea, M. (2012). Influence diagnostics for elliptical semiparametric mixed models. *Statistical Modelling*, **12**, 165-193.

[20] Ibacache-Pulgar, G. (2009). *Modelos Mistos Aditivos Semiparamétricos de Contornos Elípticos*. Tese de Doutorado. IME-USP, São Paulo.

[21] Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge University Press.

[22] Laird, N. M and Ware, J. M. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.

[23] Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, **84**, 881-896.

[24] Lindsey, J. K. (1999). Multivariate elliptically contoured distributions for repeated measurements. *Biometrics*, **54**, 1277-1280.

[25] Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester.

[26] Osorio, F. (2006). *Diagnóstico de Influência em Modelos Elípticos com Efeitos Mistos*. Tese de Doutorado, Departamento de Estatística, Universidade de Sao Paulo.

[27] Osorio, F., Paula, G. A. and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computational Statistics and Data Analysis*, **51**, 4354-4368.

[28] Reinsch, C. (1967). Smoothing by spline functions. *Numerical Mathematical*, **10**, 177-183.

[29] Rigby, R. and Stasinopoulos, D. (2005). Generalized additive models for location, scale and shape. *Applied Statistical*, **54**, 507-554.

[30] Savalli, C., Paula, G. A. and Cysneiros, F. (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling*, **6**, 59-76.

[31] Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society B*, **47**,1-52.

[32] Staudenmayer J, Lake E E and Wand MP (2009). Robustness for general design mixed models using the t-distribution. *Statistical Modelling*, **9**, 235-255.

[33] Speckman, P. (1988). Kernel smoothing in partly linear models. *Journal of the Royal Statistical Society B*, **50**, 413-436.

[34] Verbeke, G. and Molenberghs, G. (2011). *Linear Mixed Models for longitudinal Data*. Springer, New York.

[35] Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *Journal of the Royal Statistical Society B*, **45**, 133-150.

[36] Wahba, G. and Wold, S. (1975). A completely automatic french curve: fitting splines functions by cross-validation. *Comunications in Statistics*, **4**, 1-17.

[37] Wang, Y. (1998). Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society B*, **60**, 159-174.

[38] Yee, T. W. (1998). On an alternative solution to the vector spline problem. *J. R. Statist. Soc. B*, **60**, 183-188.

[39] Yee, T. W. and Mackenzie, M. (2002). Vector generalized additive models in plant ecology. *Ecological Modelling*, **157**, 141-156.

[40] Yee, T. W. and Stephenson, A. G. (2007). Vector generalized linear and additive extreme value models. *Springer*, **10**, 1-19.

[41] Yee, T. W. and Wild, C. J. (1996). Vector generalized additive models. *J. R. Statist. Soc. B*, **58**, 481-493.

[42] Zhang, D. Lin, X. and Sowers, M. (1998). Semiparametric Stochastic Mixed Models for Longitudinal Data. *Journal of the American Statistical Association, Theory and Methods*, **93**, 710-719.