



UNIVERSIDAD NACIONAL DE LA PLATA

Facultad de Ciencias Exactas

Departamento de Matemática

# Estimadores de tipo MM para el modelo lineal multivariado

Tesis presentada para optar al título de  
Doctor de la Facultad de Ciencias Exactas en el área Matemática

Lic. Nadia Laura Kudraszow

Director de tesis: Dr. Ricardo A. Maronna.

Kudraszow, Nadia Laura

Estimadores de tipo MM para el modelo lineal multivariado. - 1a ed. - La Plata :  
Universidad Nacional de La Plata, 2012.

E-Book.

ISBN 978-950-34-0914-5

1. Matemática. 2. Tesis de Doctorado. I. Título  
CDD 510.711

Fecha de catalogación: 12/11/2012



## **Estimadores de tipo MM para el modelo lineal multivariado**

### **Resumen**

En esta tesis, proponemos una clase de estimadores robustos para modelos lineales multivariados. Basados en el enfoque de la MM estimación (Yohai 1987, [36]), calculamos los coeficientes de regresión y la matriz de covarianzas de los errores de forma simultánea.

Estos estimadores tienen alto punto de ruptura y alta eficiencia asintótica bajo errores normales.

Probamos la consistencia y normalidad asintótica asumiendo que los errores tienen una distribución elíptica.

Describimos un algoritmo iterativo para el cálculo numérico de estos estimadores.

Las ventajas de los estimadores propuestos sobre sus competidores se evidencian tanto en los datos simulados como en los reales.

Finalmente, damos una aplicación de los MM-estimadores al análisis de correlación canónico mediante la definición de estimadores robustos para las coordenadas y correlaciones canónicas, y comparamos el desempeño de estos estimadores con el de otros estimadores robustos mediante un estudio de simulación.

*Palabras claves:* Métodos robustos; MM-estimador; Modelo lineal multivariado.

## Estimates of MM type for the multivariate linear model

### Abstract

In this thesis, we propose a class of robust estimates for multivariate linear models. Based on the approach of MM estimation (Yohai 1987, [36]), we estimate the regression coefficients and the covariance matrix of the errors simultaneously.

These estimates have both high breakdown point and high asymptotic efficiency under Gaussian errors.

We prove consistency and asymptotic normality assuming errors with an elliptical distribution.

We describe an iterative algorithm for the numerical calculation of these estimates.

The advantages of the proposed estimates over their competitors are demonstrated through both simulated and real data.

Finally, we give an application of the MM-estimates to the canonical correlation analysis by defining robust estimates for the canonical coordinates and correlations, and we compare the performance of these estimates with other robust estimates through a simulation study.

*Keywords:* Robust methods; MM-estimate; Multivariate linear model.

# Índice general

---

<b>Índice general</b>	<b>III</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Definición y Propiedades</b>	<b>5</b>
2.1. Preliminares . . . . .	5
2.2. Definición del MM-estimador . . . . .	7
2.3. Existencia y Propiedades . . . . .	9
2.4. Punto de Ruptura . . . . .	11
2.4.1. Punto de Ruptura Finito . . . . .	11
2.4.2. Punto de Ruptura Asintótico . . . . .	12
2.5. Función de Influencia . . . . .	13
<b>3. Teoría Asintótica</b>	<b>17</b>
3.1. Preliminares . . . . .	17
3.2. Consistencia . . . . .	18
3.3. Normalidad Asintótica . . . . .	18
<b>4. Algoritmo</b>	<b>23</b>
<b>5. Resultados Numéricos</b>	<b>25</b>
5.1. Simulación . . . . .	25
5.2. Ejemplo con Datos Reales . . . . .	32
<b>6. Aplicación al Análisis de Correlación Canónica</b>	<b>35</b>
6.1. Estudio de Simulación . . . . .	41

<b>A. Apéndice</b>	<b>51</b>
A.1. Demostración del Teorema 1 . . . . .	53
A.2. Demostración del Teorema 2 . . . . .	57
A.3. Demostración del Teorema 3 . . . . .	59
A.4. Demostración Teorema 4 . . . . .	63
A.5. Demostración del Teorema 5 . . . . .	66
A.6. Demostración del Teorema 6 . . . . .	73
A.7. Demostración de la Proposición 7 . . . . .	76
A.8. Demostración de la Proposición 8 . . . . .	77
A.9. Demostración del Teorema 9 . . . . .	79
<b>Bibliografía</b>	<b>81</b>

# Capítulo 1

## Introducción

---

Consideremos un modelo lineal multivariado (MLM) con predictores aleatorios, es decir, observamos  $n$  vectores independientes e idénticamente distribuidos (i.i.d.) de dimensión  $(p + q)$ ,  $\mathbf{z}_i = (\mathbf{y}'_i, \mathbf{x}'_i)'$  con  $1 \leq i \leq n$ , donde

$$\mathbf{y}_i = (y_{i1}, \dots, y_{iq})' \in \mathbb{R}^q, \quad \mathbf{x}_i = (x_{i1}, \dots, x_{ip})' \in \mathbb{R}^p$$

y  $'$  denota traspuesta. Los  $\mathbf{y}_i$  son los vectores respuestas y los  $\mathbf{x}_i$  son los predictores y ambos satisfacen la ecuación

$$\mathbf{y}_i = \mathbf{B}'_0 \mathbf{x}_i + \mathbf{u}_i \quad 1 \leq i \leq n, \quad (1.1)$$

donde  $\mathbf{B}_0 \in \mathbb{R}^{p \times q}$  es la matriz de los parámetros de regresión y  $\mathbf{u}_i$  es un vector de dimensión  $q$  independiente de  $\mathbf{x}_i$ . Si  $x_{ip} = 1$  para todo  $1 \leq i \leq n$ , obtenemos un modelo de regresión con intercept.

Denotamos a las distribuciones de  $\mathbf{x}_i$  y  $\mathbf{u}_i$  por  $G_0$  y  $F_0$ , respectivamente, y  $\Sigma_0$  es la matriz de covarianzas de  $\mathbf{u}_i$ . La distribución normal  $p$ -variada con vector de medias  $\boldsymbol{\mu}$  y matriz de covarianzas  $\Sigma$  es notada por  $N_p(\boldsymbol{\mu}, \Sigma)$ .

Los estimadores de máxima verosimilitud (MVE) de  $\mathbf{B}_0$  y  $\Sigma_0$  cuando la distribución de los  $\mathbf{u}_i$  es  $N_q(\mathbf{0}, \Sigma_0)$  están dados por

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (1.2)$$



y

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \widehat{\mathbf{u}}_i(\widehat{\mathbf{B}}) \widehat{\mathbf{u}}_i'(\widehat{\mathbf{B}}). \quad (1.3)$$

respectivamente, donde  $\mathbf{X}$  es la matriz de  $n \times p$  cuya  $i$ -ésima fila es  $\mathbf{x}_i'$ ,  $\mathbf{Y}$  es la matriz de  $n \times q$  cuya  $i$ -ésima fila es  $\mathbf{y}_i'$  y los residuos  $\widehat{\mathbf{u}}_i(\mathbf{B})$  están definidos por

$$\widehat{\mathbf{u}}_i(\mathbf{B}) = \mathbf{y}_i - \mathbf{B}'\mathbf{x}_i \quad 1 \leq i \leq n. \quad (1.4)$$

Como se puede observar en (1.2) y (1.3) las columnas de  $\widehat{\mathbf{B}}$  corresponden a los estimadores de mínimos cuadrados calculados para cada componente de  $\mathbf{y}$  por separado, y  $\widehat{\Sigma}$  es la matriz de covarianzas muestral de los correspondientes residuos. Es bien sabido que estos estimadores no son robustos: una pequeña proporción de datos atípicos podría causar un gran efecto sobre sus valores.

Para poder medir el grado de robustez de un estimador, Hampel [14] introdujo el concepto de punto de ruptura. Donoho [10] y Donoho y Huber [11] definieron una versión finita de este concepto que será usada en este trabajo. En términos generales, el punto de ruptura finito de un estimador es la menor proporción de datos atípicos que es necesario tener en la muestra para que el estimador tome un valor extremo.

La primer propuesta de un estimador robusto para el modelo lineal multivariado fue realizada por Koenker y Portnoy [21]. Ellos propusieron aplicar un M-estimador, basado en una función de pérdida convexa, a cada coordenada del vector respuesta. El problema de este estimador es que no es afín equivariante y tiene punto de ruptura igual a cero. Posteriormente, se definieron otros estimadores sin ese problema. Rousseeuw et al. [30] propusieron estimadores para el modelo lineal multivariado basados en un estimador robusto de la matriz de covarianzas de  $\mathbf{z} = (\mathbf{x}', \mathbf{y}')$ . Bilodeau y Duchesne [3] extendieron los S-estimadores introducidos por Davies [9] para el modelo de posición y escala multivariado, luego Van Aelst y Willems [34] estudiaron la robustez de estos estimadores. Agulló et al. [1] extendieron el estimador con determinante mínimo de la matriz de covarianzas (MCD) introducido por Rousseeuw [29] para el modelo de posición y escala multivariado y Roelandt et al. [27] extendieron al modelo lineal multivariado la definición de GS-estimadores introducida por

Croux et al. [8]. Estos estimadores tienen alto punto de ruptura pero no son altamente eficientes cuando los errores tienen distribución normal y  $q$  es pequeño. Para solucionar ese problema, Agulló et al. [1] introdujeron estimadores basados en un paso de reponderación y con esto mejoraron la eficiencia de sus estimadores manteniendo sus altos puntos de ruptura. García Ben et al. [13] extendieron los  $\tau$ -estimadores para el modelo de regresión lineal, definidos por Yohai y Zamar [38], al modelo lineal multivariado obteniendo estimadores con alto punto de ruptura y alta eficiencia normal.

En esta tesis proponemos una alternativa a los  $\tau$ -estimadores, mediante un enfoque del tipo MM, el cual fue propuesto por primera vez por Yohai [36] para el modelo lineal univariado, y luego estudiado por Lopuhaä [22], Tatsuo-ka et al. [33] y Salibián-Barrera et al. [31] para el modelo de posición y escala multivariado. Entre otras cosas mostraremos que nuestros estimadores tienen simultáneamente alto punto de ruptura y alta eficiencia normal.

A continuación, daremos una síntesis del desarrollo de esta tesis: en el Capítulo 2, introduciremos conceptos existentes en la literatura que serán utilizados con frecuencia a lo largo de este trabajo. Definiremos los MM-estimadores para el modelo lineal multivariado. Estableceremos las ecuaciones de estimación y las condiciones en las que se garantiza la existencia de solución de dichas ecuaciones. Definiremos formalmente la noción de punto de ruptura finito y punto de ruptura asintótico para estimadores del modelo lineal multivariado y encontraremos cotas inferiores para los puntos de ruptura de los MM-estimadores. Finalmente, obtendremos la expresión de la función de influencia para el estimador antes mencionado. En el Capítulo 3, estudiaremos las propiedades asintóticas (consistencia y normalidad asintótica) de la sucesión de MM-estimadores asumiendo que los predictores son aleatorios y que los errores tienen distribución elíptica unimodal. En el Capítulo 4, describiremos un algoritmo para calcular los MM-estimadores basado en un método iterativo de estimadores de máxima verosimilitud ponderados. En el Capítulo 5, presentamos los resultados de un estudio de simulación en el que comparamos a los MM-estimadores con el estimador de máxima verosimilitud y otros estimadores robustos. También mostraremos el desempeño de los estimadores propuestos mediante un ejemplo con datos reales. En el Capítulo 6, daremos una

aplicación de los MM-estimadores al Análisis Canónico. Definiremos un método robusto para estimar las coordenadas canónicas y dos métodos robustos para estimar las correlaciones canónicas. Finalmente compararemos el desempeño de los estimadores propuestos con otros estimadores robustos mediante un estudio de simulación. Las demostraciones de los teoremas y proposiciones enunciados en los Capítulos 2 al 5 pueden encontrarse en el Apéndice.

# Capítulo 2

## Definición y Propiedades

---

### 2.1. Preliminares

Antes de presentar el problema específico objeto de este Capítulo, necesitamos introducir algunos conceptos que se utilizarán sucesivamente a lo largo de la tesis.

**Definición 1.** Una *función- $\rho$*  denotará a una función  $\rho(u)$  tal que:

- $\rho(0) = 0$ .
- $\rho(-u) = \rho(u)$ .
- $0 \leq u \leq v$  implica  $\rho(u) \leq \rho(v)$ .
- $\rho$  es continua.
- $\sup_u \rho(u) = 1$ .
- Si  $\rho(u) < 1$  y  $0 \leq u < v$  entonces  $\rho(u) < \rho(v)$ .

Notar que de acuerdo a la terminología de Maronna et al. [23] esta sería una “función- $\rho$  acotada”.

**Definición 2.** Una *función- $\psi$*  denotará a una función  $\psi$  que es la derivada de una función- $\rho$ .

Una función- $\rho$  muy popular es la *función bicuadrada*:

$$\rho_B(u) = \begin{cases} 3u^2 - 3u^4 + u^6 & \text{si } |u| \leq 1 \\ 1 & \text{si } |u| > 1, \end{cases} \quad (2.1)$$

cuya derivada es

$$\psi_B(u) = \begin{cases} 6u(1 - u^2)^2 & \text{si } |u| \leq 1 \\ 0 & \text{si } |u| > 1. \end{cases}$$

A continuación se muestran las gráficas de la función- $\rho$  bicuadrada (arriba) y su correspondiente función- $\psi$  (abajo).

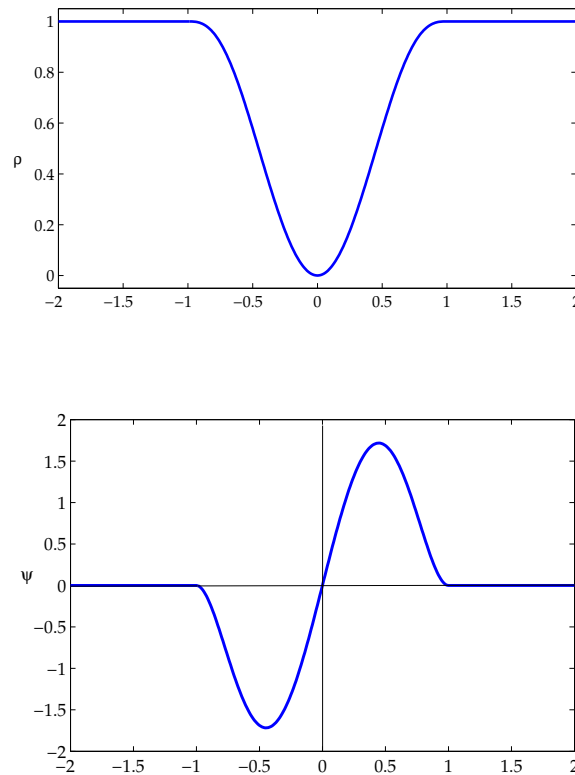


Figura 2.1: Funciones- $\rho$  y  $-\psi$  bicuadradas.

**Definición 3.** Dada  $\mathbf{v} = \{v_1, \dots, v_n\}$  una muestra de tamaño  $n$  un  $M$ -estimador de escala  $s(\mathbf{v})$  es definido como el valor de  $s$  que es solución de

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{v_i}{s} \right) = b, \quad (2.2)$$

donde  $\rho_0$  es una función- $\rho$  y  $b \in (0, 1)$ , o  $s = 0$  si  $\#(v_i = 0) \geq n(1 - b)$ .

Como ya hemos mencionado antes, el punto de ruptura es una medida del grado de robustez de un estimador. En el caso de un estimador de escala, se busca que el estimador se mantenga alejado del cero (“implosión”) y del infinito (“explosión”). Huber [17] probó que el punto de ruptura de explosión de un  $M$ -estimador de escala es  $\varepsilon_\infty^* = b$  y el punto de ruptura de implosión es  $\varepsilon_0^* = 1 - b$ . Luego, el punto de ruptura asintótico de este estimador de escala es  $\varepsilon^* = \min(b, 1 - b)$ , y tomando  $b = 0.5$  obtenemos el punto de ruptura más alto posible,  $\varepsilon^* = 0.5$ .

**Definición 4.** Dados un vector  $\mathbf{u}$  y una matriz definida positiva  $\mathbf{V}$ , la *norma de Mahalanobis de  $\mathbf{u}$  con respecto a  $\mathbf{V}$*  está definida como

$$d(\mathbf{u}, \mathbf{V}) = (\mathbf{u}' \mathbf{V}^{-1} \mathbf{u})^{1/2}.$$

En particular dadas  $\mathbf{B} \in \mathbb{R}^{p \times q}$  y  $\Sigma \in \mathbb{R}^{q \times q}$ , denotaremos por  $d_i(\mathbf{B}, \Sigma)$  ( $1 \leq i \leq n$ ) a las normas de Mahalanobis de los residuos (1.4) con respecto a la matriz  $\Sigma$ , esto es

$$d_i(\mathbf{B}, \Sigma) = (\hat{\mathbf{u}}_i(\mathbf{B})' \Sigma^{-1} \hat{\mathbf{u}}_i(\mathbf{B}))^{1/2}.$$

## 2.2. Definición del MM-estimador

Usando los conceptos definidos en la sección anterior, podemos describir un MM-estimador para el modelo lineal multivariado mediante el siguiente procedimiento:

Sea  $(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n)$  un estimador inicial de  $(\mathbf{B}_0, \Sigma_0)$ , con alto punto de ruptura y tal que  $|\tilde{\Sigma}_n| = 1$ , donde  $|\tilde{\Sigma}_n|$  es el determinante de  $\tilde{\Sigma}_n$ . Calculamos las normas de Mahalanobis de los residuos utilizando  $(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n)$ ,

$$d_i(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n) = (\hat{\mathbf{u}}'_i(\tilde{\mathbf{B}}_n) \tilde{\Sigma}_n^{-1} \hat{\mathbf{u}}_i(\tilde{\mathbf{B}}_n))^{1/2} \quad 1 \leq i \leq n. \quad (2.3)$$

Luego, calculamos el M-estimador de escala  $\hat{\sigma}_n := s(\mathbf{d}(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n))$  de dichas normas, definido por (2.2), usando una función  $\rho_0$  como la de la Definición 1 y  $b = 0.5$ .

Sean  $\rho_1$  otra función- $\rho$  tal que satisface la desigualdad

$$\rho_1 \leq \rho_0, \quad (2.4)$$

y  $\mathcal{S}_q$  el conjunto de todas las matrices de  $q \times q$  simétricas y definidas positivas.

Sea  $(\hat{\mathbf{B}}_n, \hat{\Gamma}_n)$  un mínimo local de

$$S(\mathbf{B}, \Gamma) = \sum_{i=1}^n \rho_1 \left( \frac{d_i(\mathbf{B}, \Gamma)}{\hat{\sigma}_n} \right) \quad (2.5)$$

en  $\mathbb{R}^{p \times q} \times \mathcal{S}_q$ , el cual satisface

$$S(\hat{\mathbf{B}}_n, \hat{\Gamma}_n) \leq S(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n) \quad (2.6)$$

y  $|\hat{\Gamma}_n| = 1$ . Luego el MM-estimador de  $\mathbf{B}_0$  está definido como  $\hat{\mathbf{B}}_n$ , y el respectivo estimador de  $\Sigma_0$  es

$$\hat{\Sigma}_n = \hat{\sigma}_n^2 \hat{\Gamma}_n. \quad (2.7)$$

**Observación 1.** Una forma de elegir a las funciones- $\rho$ ,  $\rho_0$  y  $\rho_1$ , de manera tal que satisfaga (2.4) es la siguiente. Sea  $\rho$  una función- $\rho$  y sean  $0 < c_0 < c_1$ : Tomamos

$$\rho_0 = \rho(u/c_0) \quad \text{y} \quad \rho_1 = \rho(u/c_1). \quad (2.8)$$

El valor  $c_0$  deberá ser elegido de manera tal que  $\hat{\sigma}_n$  sea 1 cuando los errores  $\mathbf{u}_i$ , con  $i = 1, \dots, n$ , tengan distribución  $N_q(\mathbf{0}, \mathbf{I})$ . La elección de  $c_1$  determinará la eficiencia asintótica del MM-estimador. Para más detalle ver la Observación 4.

### 2.3. Existencia y Propiedades

El siguiente teorema implica que existe el mínimo absoluto de  $S(\mathbf{B}, \Gamma/|\Gamma|^{1/q})$  en  $\mathbb{R}^{p \times q} \times \mathcal{S}_q$ . Claramente, de este mínimo absoluto podemos obtener un MM-estimador. Sin embargo, cualquier otro mínimo local  $(\mathbf{B}, \Gamma)$  que satisface (2.6), puede ser usado para obtener un MM-estimador con alto punto de ruptura y eficiente bajo errores normales.

Antes de enunciar el teorema definiremos a  $k_n$  como el máximo número de observaciones  $(\mathbf{y}'_i, \mathbf{x}'_i)$  de una muestra que están en un hiperplano, es decir,

$$k_n := \max_{\|\mathbf{v}\| + \|\mathbf{w}\| > 0} \{i : \mathbf{v}'\mathbf{x}_i + \mathbf{w}'\mathbf{y}_i = 0\}. \quad (2.9)$$

**Teorema 1.** Sea  $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$  una muestra de tamaño  $n$  que satisface el MLM (1.1), donde  $\mathbf{z}_i = (\mathbf{y}'_i, \mathbf{x}'_i)'$ . Si  $k_n/n < 0.5$  entonces existe un par  $(\hat{\mathbf{B}}_n, \hat{\Gamma}_n)$  que minimiza la función  $S(\mathbf{B}, \Gamma)$ , definida en (2.5), para todo  $(\mathbf{B}, \Gamma) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q$  tal que  $|\Gamma| = 1$ .

En el siguiente Teorema obtenemos las ecuaciones de estimación de los MM-estimadores.

**Teorema 2.** Supongamos que  $\rho_1$  es diferenciable. Luego los MM-estimadores  $(\hat{\mathbf{B}}_n, \hat{\Sigma}_n)$  satisfacen las siguientes ecuaciones:

$$\sum_{i=1}^n W(d_i(\hat{\mathbf{B}}_n, \hat{\Sigma}_n)) \hat{\mathbf{u}}_i(\hat{\mathbf{B}}_n) \mathbf{x}'_i = \mathbf{0} \quad (2.10)$$

$$\hat{\Sigma}_n = q \frac{\sum_{i=1}^n W(d_i(\hat{\mathbf{B}}_n, \hat{\Sigma}_n)) \hat{\mathbf{u}}_i(\hat{\mathbf{B}}_n) \hat{\mathbf{u}}_i(\hat{\mathbf{B}}_n)'}{\sum_{i=1}^n \psi_1(d_i(\hat{\mathbf{B}}_n, \hat{\Sigma}_n)) d_i(\hat{\mathbf{B}}_n, \hat{\Sigma}_n)} \quad (2.11)$$

donde  $\psi_1(u) = \rho'_1(u)$  y  $W(u) = \psi_1(u)/u$ .

**Observación 2.** Como podemos ver en la ecuación (2.10), la  $j$ -ésima columna de  $\hat{\mathbf{B}}_n$  es el estimador de mínimos cuadrados pesado correspondiente a la regresión univariada cuya variable dependiente es la  $j$ -ésima componente de  $\mathbf{y}$ ,



el vector de variables independientes es el mismo que en la regresión multivariada y la observación  $i$  recibe el peso  $W \left( d_i(\hat{\mathbf{B}}_n, \hat{\Sigma}_n) \right)$ . Además, por (2.11),  $\hat{\Sigma}_n$  es proporcional a la matriz de covarianzas muestral de los residuos pesados con los mismos pesos. Como estos pesos dependen de los estimadores  $\hat{\mathbf{B}}_n$  y  $\hat{\Sigma}_n$ , no podemos usar las relaciones (2.10) y (2.11) para calcular los estimadores, pero serán utilizadas para formular un algoritmo iterativo en la Sección 6.

**Observación 3.** Si  $\tilde{\mathbf{B}}_n = \tilde{\mathbf{B}}_n(\mathbf{X}, \mathbf{Y})$  es regresión-, afín- y escala-equivariante, es decir, si

$$\tilde{\mathbf{B}}_n(\mathbf{X}, \mathbf{XA} + \mathbf{Y}) = \tilde{\mathbf{B}}_n(\mathbf{X}, \mathbf{Y}) + \mathbf{A} \text{ para toda matriz } \mathbf{A} \in \mathbb{R}^{p \times q},$$

$$\tilde{\mathbf{B}}_n(\mathbf{X}, \mathbf{YW}) = \tilde{\mathbf{B}}_n(\mathbf{X}, \mathbf{Y})\mathbf{W} \text{ para toda matriz } \mathbf{W} \in \mathbb{R}^{q \times r}$$

y

$$\tilde{\mathbf{B}}_n(\mathbf{XV}, \mathbf{Y}) = \mathbf{V}^{-1}\tilde{\mathbf{B}}_n(\mathbf{X}, \mathbf{Y}) \text{ para toda matriz no singular } \mathbf{V} \in \mathbb{R}^{p \times p}.$$

Y si  $\tilde{\Sigma}_n = \tilde{\Sigma}_n(\mathbf{X}, \mathbf{Y})$  es regresión- y escala-invariante y afín-equivariante, es decir, si

$$\tilde{\Sigma}_n(\mathbf{X}, \mathbf{XA} + \mathbf{Y}) = \tilde{\Sigma}_n(\mathbf{X}, \mathbf{Y}) \text{ para toda matriz } \mathbf{A} \in \mathbb{R}^{p \times q},$$

$$\tilde{\Sigma}_n(\mathbf{XV}, \mathbf{Y}) = \tilde{\Sigma}_n(\mathbf{X}, \mathbf{Y}) \text{ para toda matriz no singular } \mathbf{V} \in \mathbb{R}^{p \times p}$$

y

$$\tilde{\Sigma}_n(\mathbf{X}, \mathbf{YW}) = \mathbf{W}'\tilde{\Sigma}_n(\mathbf{X}, \mathbf{Y})\mathbf{W} \text{ para toda matriz } \mathbf{W} \in \mathbb{R}^{q \times r}.$$

Luego  $\hat{\mathbf{B}}_n$  será regresión-, afín- y escala-equivariante y  $\hat{\Sigma}_n$  será regresión- y escala-invariante y afín-equivariante.

## 2.4. Punto de Ruptura

### 2.4.1. Punto de Ruptura Finito

Para investigar la robustez de los MM-estimadores, en esta subsección buscaremos una cota inferior de su punto de ruptura finito. El punto de ruptura finito del estimador de la matriz de coeficientes es la menor fracción de datos atípicos que hacen a dicho estimador no acotado, y el punto de ruptura finito del estimador de la matriz de covarianzas es la menor fracción de datos atípicos que hacen al estimador no acotado o singular.

Sea  $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$  un conjunto de datos de tamaño  $n$  que satisfacen el modelo lineal multivariado (1.1), donde  $\mathbf{z}_i = (\mathbf{y}'_i, \mathbf{x}'_i)'$  y sean  $\hat{\mathbf{B}}$  y  $\hat{\Sigma}$  estimadores de  $\mathbf{B}_0$  y  $\Sigma_0$  respectivamente. Definimos

$$\mathcal{Z}_m = \{\mathbf{Z}^* = \{\mathbf{z}_1^*, \dots, \mathbf{z}_n^*\} \text{ tal que } \#\{i : \mathbf{z}_i = \mathbf{z}_i^*\} \geq n - m\}, \quad (2.12)$$

$$S_m(\mathbf{Z}, \hat{\mathbf{B}}) = \sup\{\|\hat{\mathbf{B}}(\mathbf{Z}^*)\|_2 \text{ con } \mathbf{Z}^* \in \mathcal{Z}_m\},$$

$$S_m^+(\mathbf{Z}, \hat{\Sigma}) = \sup\{\lambda_1(\hat{\Sigma}(\mathbf{Z}^*)) \text{ con } \mathbf{Z}^* \in \mathcal{Z}_m\}$$

y

$$S_m^-(\mathbf{Z}, \hat{\Sigma}) = \inf\{\lambda_q(\hat{\Sigma}(\mathbf{Z}^*)) \text{ con } \mathbf{Z}^* \in \mathcal{Z}_m\},$$

donde  $\|\cdot\|_2$  es la norma Frobenius y  $\lambda_1(\hat{\Sigma}(\mathbf{Z}^*))$  y  $\lambda_q(\hat{\Sigma}(\mathbf{Z}^*))$  representan al mayor y menor autovalor de  $\hat{\Sigma}(\mathbf{Z}^*)$  respectivamente (para más detalles ver el Apéndice).

**Definición 5.** Definimos el *punto de ruptura finito* de  $\hat{\mathbf{B}}$  como  $\varepsilon^*(\mathbf{Z}, \hat{\mathbf{B}}) = m^*/n$  donde

$$m^* = \min\{m : S_m(\mathbf{Z}, \hat{\mathbf{B}}) = \infty\},$$

el *punto de ruptura finito* de  $\hat{\Sigma}$  como  $\varepsilon^*(\mathbf{Z}, \hat{\Sigma}) = m^*/n$  donde

$$m^* = \min\{m : \frac{1}{S_m^-(\mathbf{Z}, \hat{\Sigma})} + S_m^+(\mathbf{Z}, \hat{\Sigma}) = \infty\}$$

$$\text{y } \varepsilon_n^*(\mathbf{Z}, \hat{\mathbf{B}}, \hat{\Sigma}) = \min\{\varepsilon^*(\mathbf{Z}, \hat{\mathbf{B}}), \varepsilon^*(\mathbf{Z}, \hat{\Sigma})\}.$$

El siguiente teorema da una cota inferior para el punto de ruptura de un MM-estimador.

**Teorema 3.** Sean  $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ , con  $\mathbf{z}_i = (\mathbf{y}'_i, \mathbf{x}'_i)'$  que satisface el MLM (1.1), y  $k_n$  como fue definido en (2.9). Consideremos  $\rho_0$  y  $\rho_1$  dos funciones- $\rho$  que satisfacen (2.4) y supongamos que

$$k_n < n/2.$$

Luego

$$\varepsilon_n^*(\mathbf{Z}, \hat{\mathbf{B}}_n, \hat{\Sigma}_n) \geq \min\left(\varepsilon_n^*(\mathbf{Z}, \tilde{\mathbf{B}}_n, \tilde{\Sigma}_n), \frac{[n/2] - k_n}{n}\right). \quad (2.13)$$

Como  $k_n$  es siempre mayor o igual que  $p + q - 1$ , si  $\varepsilon_n^*(\mathbf{Z}, \tilde{\mathbf{B}}_n, \tilde{\Sigma}_n)$  está cerca de 0.5, la máxima cota inferior será  $([n/2] - (p + q - 1))/n$ . Es decir, cuando los puntos de la muestra están en posición general el punto de ruptura finito está cerca de 0.5 para  $n$  grande.

Si no fijamos  $b = 0.5$  y si  $k_n < n(1 - b)$ , obtendríamos la misma cota que en (2.13) pero con  $[n(1 - b)]$  en lugar de  $[n/2]$ . En este caso, el máximo punto de ruptura finito sería alcanzado en  $b = 0.5 - k_n/n$ , el cuál está muy cerca de nuestra elección de  $b = 0.5$  cuando  $k_n/n$  es chico.

### 2.4.2. Punto de Ruptura Asintótico

Sean  $H$  la distribución de  $\mathbf{z} = (\mathbf{y}', \mathbf{x}')'$  y  $\mathcal{Q}$  el conjunto de todas las distribuciones.

**Definición 6.** El *punto de ruptura asintótico* de  $\hat{\mathbf{B}}$  en  $H$ , al que denotaremos con  $\varepsilon^*(\hat{\mathbf{B}}, H)$ , es el mayor  $\varepsilon^* \in (0, 1)$  tal que para  $\varepsilon < \varepsilon^*$ , el valor asintótico de  $\hat{\mathbf{B}}$  en  $(1 - \varepsilon)H + \varepsilon Q$  como una función de  $Q \in \mathcal{Q}$  se mantiene acotado. El *punto de ruptura asintótico* de  $\hat{\Sigma}$  en  $H$ , al que denotaremos con  $\varepsilon^*(\hat{\Sigma}, H)$ , es el mayor  $\varepsilon^* \in (0, 1)$  tal que para  $\varepsilon < \varepsilon^*$ , el valor asintótico de  $\hat{\Sigma}$  en  $(1 - \varepsilon)H + \varepsilon Q$  como una función de  $Q \in \mathcal{Q}$  se mantiene acotado y no singular.

Para más detalles ver Maronna et al. [23].

También podemos encontrar una cota inferior de los puntos de ruptura asintóticos de los MM-estimadores  $\hat{\mathbf{B}}_n$  y  $\hat{\Sigma}_n$ :

Si el mínimo entre los puntos de ruptura asintóticos de los estimadores iniciales  $\tilde{\mathbf{B}}_n$  y  $\tilde{\Sigma}_n$  es igual a  $\varepsilon_0^*$ , es decir,

$$\varepsilon_0^* = \min \left( \varepsilon^*(\tilde{\mathbf{B}}_n, H), \varepsilon^*(\tilde{\Sigma}_n, H) \right)$$

y

$$\alpha := \max_{\|\mathbf{v}\| + \|\mathbf{w}\| > 0} P(\mathbf{v}'\mathbf{x} + \mathbf{w}'\mathbf{y} = 0) < 0.5, \quad (2.14)$$

entonces

$$\varepsilon^*(\hat{\mathbf{B}}_n, H) \geq \min(\varepsilon_0^*, (1 - \alpha)/2) \quad \text{y} \quad \varepsilon^*(\hat{\Sigma}_n, H) \geq \min(\varepsilon_0^*, (1 - \alpha)/2).$$

La demostración es similar a la del Teorema 3, reemplazando los promedios por esperanzas.

## 2.5. Función de Influencia

Antes de obtener una expresión para la función de influencia de los MM-estimadores de regresión lineal multivariada daremos una definición que será utilizada en esta sección y en el siguiente capítulo.

**Definición 7.** Una distribución multivariada es *elíptica simétrica* o *elíptica  $m$ -dimensional* (en adelante *elíptica* para abreviar) si su función de densidad es de la forma

$$f(\mathbf{u}) = \frac{f^*((\mathbf{u} - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{u} - \boldsymbol{\mu}))}{\sqrt{|\mathbf{V}|}}$$

donde  $\mathbf{u}$  es un vector aleatorio  $m$ -dimensional con vector de medias  $\boldsymbol{\mu}$ ,  $\mathbf{V}$  es una matriz definida positiva que es proporcional a la matriz de covarianzas de  $\mathbf{u}$  y  $f^*$  es una función no negativa que además verifica que la integral de  $f$  vale 1.

Una familia importante de distribuciones elípticas es la normal multivariada. En ese caso,

$$f^*(v) = \frac{1}{(2\pi)^{q/2}} \exp(-v/2).$$

Sea  $\hat{\theta}_n$  un estimador que depende de una muestra  $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$  de variables i.i.d. en  $\mathbb{R}^k$  con distribución  $H_{\theta}$ , donde  $\theta \in \Theta \subset \mathbb{R}^m$ . Sea  $\mathbf{T}$  un estimador funcional de  $\theta$  tal que  $\mathbf{T}(H_n) = \hat{\theta}_n$ , donde  $H_n$  es la correspondiente distribución empírica. Supongamos que  $\mathbf{T}$  es Fisher consistente, es decir  $\mathbf{T}(H_{\theta}) = \theta$ . La función de influencia de  $\mathbf{T}$ , introducida por Hampel en 1974, [15], mide el efecto sobre el funcional de una pequeña fracción de contaminación de masa puntual.

**Definición 8.** Si  $\delta_{\mathbf{z}}$  denota la distribución de probabilidad que asigna masa 1 a  $\mathbf{z}$ , la *función de influencia* está definida por

$$IF(\mathbf{z}, \mathbf{T}, \theta) = \lim_{\varepsilon \rightarrow 0} \frac{\mathbf{T}((1 - \varepsilon)H_{\theta} + \varepsilon\delta_{\mathbf{z}}) - \mathbf{T}(H_{\theta})}{\varepsilon} = \left. \frac{\partial \mathbf{T}((1 - \varepsilon)H_{\theta} + \varepsilon\delta_{\mathbf{z}})}{\partial \varepsilon} \right|_{\varepsilon=0},$$

para cada  $\mathbf{z} \in \mathbb{R}^k$ .

En nuestro caso,  $\mathbf{z} = (\mathbf{y}', \mathbf{x}')'$  satisface el modelo lineal (1.1),  $\theta = (\mathbf{B}_0, \Sigma_0)$  y  $H_{\theta} = H_0$ . Sean  $\mathbf{T}_{0,1}$  y  $\mathbf{T}_{0,2}$  los estimadores funcionales asociados a los estimadores iniciales  $\tilde{\mathbf{B}}_n$  y  $\tilde{\Sigma}_n$  y  $\mathbf{T}_1$  y  $\mathbf{T}_2$  los estimadores funcionales correspondientes a los MM-estimadores  $\hat{\mathbf{B}}_n$  y  $\hat{\Sigma}_n$ . Luego, de acuerdo a (2.10) y (2.11), dada una distribución  $H$  de  $(\mathbf{y}', \mathbf{x}')'$ , el par  $(\mathbf{T}_1(H), \mathbf{T}_2(H))$  corresponde al valor de  $(\mathbf{B}, \Sigma)$  que satisface

$$\begin{aligned} E_H W(d(\mathbf{B}, \Sigma)) \hat{\mathbf{u}}(\mathbf{B}) \mathbf{x}' &= \mathbf{0}, \\ \Sigma &= q \frac{E_H W(d(\mathbf{B}, \Sigma)) \hat{\mathbf{u}}(\mathbf{B}) \hat{\mathbf{u}}(\mathbf{B})'}{E_H \psi_1(d(\mathbf{B}, \Sigma)) d(\mathbf{B}, \Sigma)}, \\ \Sigma &= S(H)^2 \Gamma, \text{ con } |\Gamma| = 1, \end{aligned}$$

donde  $d(\mathbf{B}, \Sigma) = d(\hat{\mathbf{u}}(\mathbf{B}), \Sigma)$ ,  $\hat{\mathbf{u}}(\mathbf{B}) = \mathbf{y} - \mathbf{B}'\mathbf{x}$  y

$$E_H \rho_0 \left( \frac{d(\mathbf{T}_{0,1}(H), \mathbf{T}_{0,2}(H))}{S(H)} \right) = \mathbf{0}.$$

Notar que el M-estimador de escala,  $\hat{\sigma}_n$ , utilizado en la definición de los MM-estimadores  $(\hat{\mathbf{B}}_n, \hat{\Sigma}_n)$ , verifica  $\hat{\sigma}_n = S(H_n)$ , donde  $H_n$  es la distribución empírica de  $\mathbf{z}_1, \dots, \mathbf{z}_n$ .

A continuación hallaremos la función de influencia de los MM-estimadores para el caso en que los errores en (1.1) tienen una distribución elíptica con densidad unimodal. Para eso, necesitamos hacer las siguientes suposiciones:

(S1) La distribución  $F_0$  de  $\mathbf{u}_i$  es elíptica y tiene una densidad de la forma

$$f_0(\mathbf{u}) = \frac{f_0^*(\mathbf{u}'\Sigma_0^{-1}\mathbf{u})}{|\Sigma_0|^{1/2}} \quad (2.15)$$

donde  $f_0^*$  es no creciente y tiene al menos un punto de decrecimiento en el intervalo donde  $\rho_1$  es estrictamente creciente.

(S2) La distribución del predictor  $\mathbf{x}$ ,  $G_0$ , tiene segundos momentos y la  $E_{G_0}(\mathbf{x}\mathbf{x}')$  es no singular.

(S3)  $\rho_1$  es estrictamente creciente en  $[0, \kappa]$  y constante en  $[\kappa, +\infty)$  para cierta constante  $\kappa < \infty$ .

(S4)  $P_{G_0}(\mathbf{B}'\mathbf{x} = 0) < 0.5$  para todo  $\mathbf{B} \in \mathbb{R}^{p \times q}$ .

**Teorema 4.** *Supongamos que (S1)-(S4) son válidas y que las derivadas parciales de  $E_{H_0}W(d(\mathbf{B}_0, \Sigma_0)/S(H_0))\hat{\mathbf{u}}(\mathbf{B}_0)\mathbf{x}'$  pueden obtenerse derivando con respecto a cada parámetro dentro de la esperanza. Asumimos que los estimadores funcionales asociados a los estimadores iniciales  $\tilde{\mathbf{B}}_n$  y  $\tilde{\Sigma}_n$  son afín-equivariantes. Luego, la función de influencia para el estimador funcional  $\mathbf{T}_1$  correspondiente al MM-estimador  $\hat{\mathbf{B}}_n$  es*

$$\begin{aligned} IF(\mathbf{z}_0, \mathbf{T}_1, \mathbf{B}_0, \Sigma_0) \\ = cW \left( \frac{((\mathbf{y}_0 - \mathbf{B}_0'\mathbf{x}_0)'\Sigma_0^{-1}(\mathbf{y}_0 - \mathbf{B}_0'\mathbf{x}_0))^{1/2}}{\sigma_0} \right) \Sigma_0(\mathbf{y}_0 - \mathbf{B}_0'\mathbf{x}_0)\mathbf{x}_0'E_{G_0}(\mathbf{x}\mathbf{x}')^{-1}, \end{aligned}$$

donde  $\sigma_0 = S(H_0)$  y

$$c = \frac{E_{F_0} W' \left( (\mathbf{u}' \Sigma_0 \mathbf{u})^{1/2} / \sigma_0 \right) (\mathbf{u}' \Sigma_0 \mathbf{u})^{1/2}}{\sigma_0} + E_{F_0} W \left( \frac{(\mathbf{u}' \Sigma_0 \mathbf{u})^{1/2}}{\sigma_0} \right).$$

# Capítulo 3

## Teoría Asintótica

---

### 3.1. Preliminares

A continuación daremos algunas definiciones que serán de utilidad para obtener la distribución límite de los MM-estimadores.

**Definición 9.** Sea  $\mathfrak{F}$  una clase de funciones a valores reales sobre un conjunto  $\mathfrak{X}$ . Una *envolvente* para  $\mathfrak{F}$  es una función  $F$  tal que para toda  $f$  en  $\mathfrak{F}$  se tiene que  $|f| \leq F$ .

Si  $\mu$  es una medida sobre  $\mathfrak{X}$  para la cual  $F$  es integrable, es natural pensar a  $\mathfrak{F}$  como un subconjunto de  $\mathcal{L}^1(\mu)$ , siendo este el espacio de todas las funciones  $\mu$ -integrables. Este espacio viene dotado de una distancia definida por la norma  $\mathcal{L}^1(\mu)$ . Entonces la bola cerrada con centro  $f_0$  y radio  $R$  está formada por todas la funciones  $f$  en  $\mathcal{L}^1(\mu)$  tales que  $\int |f - f_0| d\mu \leq R$ .

**Definición 10.** Diremos que  $\mathfrak{F}$  es *Euclideana* para cierta envolvente  $F$  si existen constantes positivas  $a$  y  $r$  con la siguiente propiedad: si  $0 \leq \varepsilon \leq 1$  y si  $\mu$  es una medida tal que  $\int F d\mu < \infty$ , luego existen funciones  $f_1, \dots, f_m$  en  $\mathfrak{F}$  tales que

- (i)  $m \leq a\varepsilon^{-r}$ ,
- (ii)  $\mathfrak{F}$  está cubierto por la unión de bolas cerradas con radio  $\varepsilon \int F d\mu$  y centros  $f_1, \dots, f_m$ .

Las constantes  $a$  y  $r$  no deben depender de  $\mu$ .



### 3.2. Consistencia

A continuación estableceremos la consistencia de los MM-estimadores de regresión multivariada para el caso en que los errores en (1.1) tengan una distribución elíptica con una densidad unimodal.

**Teorema 5.** Sean  $(\mathbf{y}'_i, \mathbf{x}'_i)$ ,  $1 \leq i \leq n$ , una muestra aleatoria del modelo (1.1) con parámetros  $\mathbf{B}_0$  y  $\Sigma_0$ . Supongamos que  $\rho_0$  y  $\rho_1$  son funciones- $\rho$  que satisfacen la relación (2.4), que (S1), (S3) y (S4) son válidas y que los estimadores iniciales  $\tilde{\mathbf{B}}_n$  y  $\tilde{\Sigma}_n$  son consistentes a  $\mathbf{B}_0$  y  $\Gamma_0$  respectivamente, donde  $\Gamma_0 = \Sigma_0 |\Sigma_0|^{-1/q}$ ; entonces los MM-estimadores  $\hat{\mathbf{B}}_n$  y  $\hat{\Sigma}_n$  satisfacen

$$(a) \lim_{n \rightarrow \infty} \hat{\mathbf{B}}_n = \mathbf{B}_0 \text{ c.s..}$$

$$(b) \lim_{n \rightarrow \infty} \hat{\Sigma}_n = \sigma_0^2 \Sigma_0 \text{ c.s. con } \sigma_0 \text{ definida por}$$

$$E_{F_0} \left( \rho_0 \left( \frac{(\mathbf{u}' \Gamma_0^{-1} \mathbf{u})^{1/2}}{\sigma_0} \right) \right) = b. \quad (3.1)$$

### 3.3. Normalidad Asintótica

Antes de obtener la distribución límite de  $\hat{\mathbf{B}}_n$  necesitamos hacer algunas suposiciones adicionales:

(S5)  $\rho_1$  es diferenciable,  $\psi_1 = \rho'_1$  y  $W(u) = \psi_1(u)/u$  es continua.

(S6) Sean  $\boldsymbol{\theta} = (\mathbf{B}, \Sigma)$ ,  $\mathbf{z} = (\mathbf{x}', \mathbf{y}')$ ,  $\mathbf{b}_k$  el  $k$ -ésimo vector columna de la matriz  $\mathbf{B}$  y  $\phi_{kj}(\mathbf{z}; \boldsymbol{\theta}) = W(d(\mathbf{B}, \Sigma)) (y_k - \mathbf{b}'_k \mathbf{x}) x_j$ . Entonces  $\phi_{kj}(\mathbf{z}; \boldsymbol{\theta})$  coincide con el elemento  $h = (k-1)q + j$  de la función

$$\phi(\mathbf{z}; \boldsymbol{\theta}) = W(d(\mathbf{B}, \Sigma)) \text{vec}((\mathbf{y} - \mathbf{B}'\mathbf{x})\mathbf{x}'). \quad (3.2)$$

Sean  $\text{vec}(\mathcal{S}_q) = \{\text{vec}(\Sigma) : \Sigma \in \mathcal{S}_q\}$ ,  $\theta(\xi)$  una función que a cada  $\xi$  en  $\mathbb{R}^{qp} \times \text{vec}(\mathcal{S}_q)$  le asigna un par  $(\mathbf{B}, \Sigma)$  en  $\mathbb{R}^{q \times p} \times \mathcal{S}_q$  y  $\mathcal{C}$  un subconjunto acotado de  $\mathbb{R}^{qp} \times \text{vec}(\mathcal{S}_q)$  tal que  $(\mathbf{B}_0, \sigma_0^2 \Sigma_0) \in \theta(\mathcal{C})^\circ$ , donde  $\theta(\mathcal{C})^\circ$  es el conjunto de punto interiores de  $\theta(\mathcal{C})$ . Luego cada una de las clases de funciones

$$\mathfrak{F}_{kj} = \{\phi_{kj}(\mathbf{z}; \theta(\xi)) : \xi \in \mathcal{C}\} \quad (3.3)$$

es una clase Euclideana con envolvente  $F_{kj}$  para el cual  $E_{H_0} F_{kj}^2 < \infty$ , donde  $H_0$  es la distribución de  $\mathbf{z}$ .

(S7) La función  $\Phi(\theta) = E_{H_0} \phi(\mathbf{z}; \theta)$  tiene una derivada parcial  $\partial \Phi / \partial \text{vec}(\mathbf{B}')'$  que es continua en  $(\mathbf{B}_0, \sigma_0^2 \Sigma_0)$  y la matriz

$$\Lambda = \frac{\partial \Phi(\mathbf{B}, \Sigma)}{\partial \text{vec}(\mathbf{B}')'}(\mathbf{B}_0, \sigma_0^2 \Sigma_0) \quad (3.4)$$

es no singular.

**Teorema 6.** Sea  $\mathbf{z}_i = (\mathbf{y}'_i, \mathbf{x}'_i)'$ , con  $1 \leq i \leq n$ , una muestra aleatoria del modelo (1.1) con parámetros  $\mathbf{B}_0$  y  $\Sigma_0$ . Supongamos que la función- $\rho$   $\rho_1$  satisface (S3), que (S1)-(S2) y (S4)-(S7) son válidas y que los estimadores iniciales  $\tilde{\mathbf{B}}_n$  y  $\tilde{\Sigma}_n$  son consistentes a  $\mathbf{B}_0$  y  $\Gamma_0$  respectivamente, donde  $\Gamma_0 = \Sigma_0 |\Sigma_0|^{-1/q}$ ; entonces  $n^{1/2} \text{vec}(\hat{\mathbf{B}}'_n - \mathbf{B}'_0) \xrightarrow{d} N_{qp}(\mathbf{0}, \mathbf{V})$ , donde  $\xrightarrow{d}$  denota convergencia en distribución y

$$\mathbf{V} = \Lambda^{-1} \mathbf{M} \Lambda^{-1'} \quad (3.5)$$

donde  $\mathbf{M}$  es la matriz de covarianza de  $\phi(\mathbf{z}_1, (\mathbf{B}_0, \sigma_0^2 \Sigma_0))$ , con  $\phi$  definida en (3.2) y  $\Lambda$  está definida en (3.4).

Bajo condiciones de diferenciabilidad adecuadas puede obtenerse una expresión más detallada de la matriz de covarianzas  $\mathbf{V}$  del Teorema 6.

**Proposición 7.** Si  $\phi(\mathbf{z}; (\mathbf{B}, \sigma_0^2 \Sigma_0))$  es diferenciable con respecto a  $\mathbf{B}$  y

$$\left\| \frac{\partial \phi(\mathbf{z}; (\mathbf{B}, \sigma_0^2 \Sigma_0))}{\partial \text{vec}(\mathbf{B}')'} \right\|_2 \leq K(\mathbf{z}) \quad \text{con} \quad E_{H_0} K(\mathbf{z}) < \infty, \quad (3.6)$$

para casi todo  $\mathbf{z} \in \mathbb{R}^{p+q}$  y  $\mathbf{B}$  en un entorno de  $\mathbf{B}_0$ , entonces

$$\mathbf{V} = \left[ \frac{\sigma_0^2}{q} E_{F_0} \left( \psi_1 \left( \frac{v}{\sigma_0} \right) \right)^2 \middle/ \left( E_{F_0} W^* \left( \frac{v}{\sigma_0} \right) \right)^2 \right] (E_{G_0} \mathbf{x} \mathbf{x}')^{-1} \otimes \Sigma_0 \quad (3.7)$$

donde

$$W^* \left( \frac{v}{\sigma_0} \right) = \frac{1}{q\sigma_0} W' \left( \frac{v}{\sigma_0} \right) v + W \left( \frac{v}{\sigma_0} \right) \quad (3.8)$$

y

$$v = (\mathbf{u}' \Sigma_0^{-1} \mathbf{u})^{1/2}.$$

Se puede estimar  $\mathbf{V}$  de manera sencilla utilizando el estimador plug-in, es decir, reemplazando a las distribuciones  $F_0$  y  $G_0$  en (3.7) por las distribuciones empíricas de los residuos y de los predictores, respectivamente.

En el Teorema 6 asumimos que (S6) es válida, es decir, que las clases  $\mathfrak{F}_{kj}$  definidas en (3.3) son Euclidianas para ciertas envolventes  $F_{kj}$  con  $E_{H_0} F_{kj}^2 < \infty$ . Esta suposición es válida bajo las siguientes condiciones adicionales, las cuales son más restrictivas pero más fáciles de verificar:

(S8)  $W$  es continuamente diferenciable.

(S9)  $E_{G_0} \|\mathbf{x}\|^4 < \infty$ ,  $E_{G_0} \|\mathbf{x}\|^6 < \infty$ ,  $E_{H_0} \|\mathbf{x}\|^4 \|\mathbf{y}\|^2 < \infty$  y  $E_{H_0} \|\mathbf{x}\|^2 \|\mathbf{y}\|^4 < \infty$ .

**Proposición 8.** Si (S5), (S8) and (S9) son válidas entonces existe una función  $\theta(\xi)$  que a cada  $\xi$  en  $\mathbb{R}^{qp} \times \text{vec}(\mathcal{S}_q)$  le asigna un par  $(\mathbf{B}, \Sigma)$  en  $\mathbb{R}^{q \times p} \times \mathcal{S}_q$  y un subconjunto  $\mathcal{C}$  acotado de  $\mathbb{R}^{qp} \times \text{vec}(\mathcal{S}_q)$ , tal que  $(\mathbf{B}_0, \sigma_0^2 \Sigma_0) \in \theta(\mathcal{C})^\circ$ , para el cual cada una de las clases de funciones  $\mathfrak{F}_{kj}$  de la suposición (S6) es Euclideana para cierta envolvente  $F_{kj}$  con  $E_{H_0} F_{kj}^2 < \infty$ .

La suposición (S9), es necesaria para probar que las clases  $\mathfrak{F}_{kj}$  definidas en (3.3) son Euclidianas, pero conjeturamos que la Proposición 8 vale también bajo alguna otra hipótesis menos restrictiva.

**Observación 4.** La matriz de covarianzas del estimador de máxima verosimilitud  $\hat{\mathbf{B}}$ , dado en (1.2), está dada por

$$\mathbf{V} = (E_{F_0}(v^2)/q) (E_{G_0} \mathbf{x} \mathbf{x}')^{-1} \otimes \Sigma_0.$$

Luego, la eficiencia relativa asintótica del MM-estimador  $\hat{\mathbf{B}}_n$  con respecto al EMV  $\hat{\mathbf{B}}$  es

$$ERA(\psi_1, F_0) = E_{F_0}(v^2) \frac{\left(E_{F_0} W^* \left(\frac{v}{\sigma_0}\right)\right)^2}{\sigma_0^2 E_{F_0} \left(\psi_1 \left(\frac{v}{\sigma_0}\right)\right)^2}. \quad (3.9)$$

Como se mencionó en la Observación 1, para obtener un MM-estimador que tenga alto punto de ruptura y simultáneamente sea altamente eficiente bajo errores normales basta con elegir las constantes  $c_0$  y  $c_1$  de (2.8) adecuadamente. La constante  $c_0$  puede elegirse de modo que

$$E \left( \rho \left( \frac{(\mathbf{u}' \mathbf{\Gamma}_0^{-1} \mathbf{u})^{1/2}}{c_0} \right) \right) = b, \quad (3.10)$$

donde  $\mathbf{u}$  es  $N_q(\mathbf{0}, \Sigma_0)$ ,  $\Sigma_0 = |\Sigma_0|^{1/q} \mathbf{\Gamma}_0$  y  $b = 0.5$ , lo que garantiza un alto punto de ruptura y que la eficiencia relativa asintótica (3.9) dependa sólo de  $c_1$ . Luego,  $c_1$  puede elegirse de modo que el MM-estimador tenga la eficiencia deseada sin afectar el punto de ruptura que depende sólo de  $c_0$ .

La Tabla 3.1 da los valores de  $c_0$  que verifican (3.10) para diferentes valores de  $q$ . La Tabla 3.2 da los valores de  $c_1$  necesarios para alcanzar diferentes niveles de eficiencia asintótica. En ambos casos la función  $\rho$  de (2.8) es igual a la función bicuadrada,  $\rho_B$ , dada en (2.1).

$q$	1	2	3	4	5	10
$c_0$	1.56	2.66	3.45	4.10	4.65	6.77

Tabla 3.1: Valores de  $c_0$  para la función bicuadrada.

ERA	$q$					
	1	2	3	4	5	10
0.80	3.14	3.51	3.82	4.10	4.34	5.39
0.85	3.43	3.81	4.13	4.41	4.66	5.67
0.90	3.88	4.28	4.62	4.91	5.18	6.38
0.95	4.68	5.12	5.48	5.76	6.10	7.67

Tabla 3.2: Valores de  $c_1$  para la función bicuadrada que alcanzan valores dados de la eficiencia relativa asintótica (ARE) bajo errores normales.

# Capítulo 4

## Algoritmo

---

En este capítulo proponemos un algoritmo iterativo para calcular  $\hat{\mathbf{B}}_n$  y  $\hat{\Sigma}_n$  basado en la Observación 2. Sea  $\mathbf{z}_i = (\mathbf{y}'_i, \mathbf{x}'_i)'$  una muestra de tamaño  $n$  y supongamos que ya hemos calculado los estimadores iniciales  $\tilde{\mathbf{B}}_n$  y  $\tilde{\Sigma}_n$  con alto punto de ruptura y tal que  $|\tilde{\Sigma}_n| = 1$ .

1. Usando los valores iniciales  $\tilde{\mathbf{B}}^{(0)} = \tilde{\mathbf{B}}_n$  y  $\tilde{\Gamma}^{(0)} = \tilde{\Sigma}_n$ , calculamos el M-estimador de escala

$$\hat{\sigma}_n := s(\mathbf{d}(\tilde{\mathbf{B}}^{(0)}, \tilde{\Gamma}^{(0)})),$$

definido por (2.2), usando una función  $\rho_0$  como la de la Definición 1 y  $b = 0.5$  y la matriz  $\tilde{\Sigma}^{(0)} = \hat{\sigma}_n^2 \tilde{\Gamma}^{(0)}$ .

2. Calculamos los pesos

$$\omega_{i0} = W\left(d_i(\tilde{\mathbf{B}}^{(0)}, \tilde{\Sigma}^{(0)})\right)$$

para  $1 \leq i \leq n$ . Estos pesos son usados para calcular cada columna de  $\tilde{\mathbf{B}}^{(1)}$  por separado mediante Mínimos Cuadrados Pesados.

3. Calculamos la matriz

$$\tilde{\mathbf{C}}^{(1)} = \sum_{i=1}^n \omega_{i0} \hat{\mathbf{u}}_i(\tilde{\mathbf{B}}^{(1)}) \hat{\mathbf{u}}'_i(\tilde{\mathbf{B}}^{(1)}),$$

y con ella la matriz

$$\tilde{\Sigma}^{(1)} = \hat{\sigma}_n^2 \tilde{\mathbf{C}}^{(1)} / |\tilde{\mathbf{C}}^{(1)}|^{1/q}.$$

4. Supongamos que ya calculamos  $\tilde{\mathbf{B}}^{(k-1)}$  y  $\tilde{\Sigma}^{(k-1)}$ . Luego  $\tilde{\mathbf{B}}^{(k)}$  y  $\tilde{\Sigma}^{(k)}$  son calculados usando los pasos 2 y 3 pero empezando del  $(\tilde{\mathbf{B}}^{(k-1)}, \tilde{\Sigma}^{(k-1)})$  en lugar del  $(\tilde{\mathbf{B}}^{(0)}, \tilde{\Sigma}^{(0)})$ .
5. El procedimiento es detenido en el paso  $k$  si las diferencias absolutas relativas de todos los elementos de las matrices  $\tilde{\mathbf{B}}^{(k)}$  y  $\tilde{\mathbf{B}}^{(k-1)}$  y las diferencias absolutas relativas de todas las normas de Mahalanobis de los residuos  $\hat{\mathbf{u}}_i(\tilde{\mathbf{B}}^{(k)})$  y  $\hat{\mathbf{u}}_i(\tilde{\mathbf{B}}^{(k-1)})$  con respecto a  $\tilde{\Sigma}^{(k)}$  y  $\tilde{\Sigma}^{(k-1)}$  respectivamente son menores que cierto  $\delta$  dado.

El siguiente teorema, cuya demostración puede encontrarse en el apéndice, muestra que el procedimiento iterativo para calcular MM-estimadores produce el descenso de la función objetivo.

**Teorema 9.** *Si  $W(u)$  es no creciente en  $|u|$  entonces en cada iteración del algoritmo la función  $\sum_{i=1}^n \rho_1(d_i(\mathbf{B}, \Sigma))$  es no creciente.*

# Capítulo 5

## Resultados Numéricos

---

### 5.1. Simulación

A fin de investigar el desempeño de los estimadores propuestos realizamos un estudio de simulación.

- Consideramos el modelo lineal multivariado dado por (1.1) para dos casos:

predictores de dimensión  $p = 2$  con respuestas de dimensión  $q = 2$

y

predictores de dimensión  $p = 2$  con respuestas de dimensión  $q = 5$ .

Debido a la equivarianza de los estimadores que considerados, sin pérdida de generalidad, tomamos

$$\mathbf{B}_0 = \mathbf{0} \quad \text{y} \quad \Sigma_0 = \mathbf{I}.$$

Los errores  $\mathbf{u}_i$  son generados a partir de una distribución  $N_q(\mathbf{0}, \mathbf{I})$  y los predictores  $\mathbf{x}_i$  de una distribución  $N_p(\mathbf{0}, \mathbf{I})$ .

- Se generaron 1000 muestras de tamaño 100. Consideramos muestras no contaminadas y muestras que contienen un 10 % de datos atípicos idénticos de la



forma  $(\mathbf{x}_0, \mathbf{y}_0)$  con

$$\mathbf{x}_0 = (x_0, 0, \dots, 0) \quad \text{y} \quad \mathbf{y}_0 = (mx_0, 0, \dots, 0).$$

Los valores de  $x_0$  considerados son 1 (datos atípicos de baja influencia) y 10 (datos atípicos de alta influencia). Tomamos una malla de valores  $m$ , comenzando en el 0. La malla se eligió de manera tal que todos los estimadores robustos alcanzaran el valor máximo de su medida de error.

- Sea  $\widehat{\mathbf{B}}^{(k)}$  el estimador de  $\mathbf{B}_0$  obtenido en la  $k$ -ésima réplica. Luego, como estamos tomando  $\mathbf{B}_0 = \mathbf{0}$ , el estimador del error medio cuadrático (EMV) está dado por

$$\text{EMC} = \frac{1}{1000} \left( \sum_{k=1}^{1000} \sum_{i=1}^p \sum_{j=1}^q \left( \widetilde{\mathbf{B}}_{ij}^{(k)} \right)^2 \right).$$

Es preciso recordar que bajo distribuciones contaminadas, los estimadores robustos pueden tener colas pesadas, por lo que es prudente evaluar su desempeño mediante medidas robustas. Por esta razón, además de emplear el EMC, calculamos el error medio cuadrático podado (EMCP), el cual se obtiene del promedio con el 10 % (superior) podado de

$$\left\{ \sum_{i=1}^p \sum_{j=1}^q \left( \widetilde{\mathbf{B}}_{ij}^{(k)} \right)^2 \right\}_{k=1}^{1000}.$$

Los resultados dados a continuación corresponden al EMCP, sin embargo el EMC ordinario arroja resultados similares (en el caso no contaminado los resultados son iguales).

-Para cada caso, se calcularon cuatro estimadores: el EMV, un S-estimador, un  $\tau$ -estimador y un MM-estimador.

Para el modelo lineal multivariado, los S-estimadores están definidos por

$$(\widehat{\mathbf{B}}, \widehat{\boldsymbol{\Sigma}}) = \arg \min \{ |\boldsymbol{\Sigma}| : (\mathbf{B}, \boldsymbol{\Sigma}) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q \}$$

ERA		$q$					
		1	2	3	4	5	10
0.80	$c_2$	3.98	3.94	4.02	4.10	4.17	4.28
	$\kappa$	0.16	0.30	0.41	0.50	0.58	0.84
0.90	$c_2$	4.97	4.97	5.10	5.25	5.39	5.98
	$\kappa$	0.11	0.21	0.29	0.35	0.40	0.59
0.95	$c_2$	6.04	6.06	6.24	6.42	6.60	7.50
	$\kappa$	0.07	0.15	0.20	0.25	0.29	0.43

Tabla 5.1: Valores de  $c_2$  y  $\kappa$  para la función bicuadrada que alcanzan valores dados de la eficiencia relativa asintótica (ARE) de los  $\tau$ -estimadores bajo errores normales.

sujeto a

$$s^2(d_1(\mathbf{B}, \Sigma), \dots, d_n(\mathbf{B}, \Sigma)) = q, \quad (5.1)$$

donde  $s$  es un M-estimador de escala.

García Ben et al. [13] extendieron los  $\tau$ -estimadores para el modelo lineal multivariado de la siguiente manera:

$$(\hat{\mathbf{B}}, \hat{\Sigma}) = \arg \min\{|\Sigma| : (\mathbf{B}, \Sigma) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q\}$$

sujeto a

$$\tau^2(d_1(\mathbf{B}, \Sigma), \dots, d_n(\mathbf{B}, \Sigma)) = \kappa, \quad (5.2)$$

donde la  $\tau$ -escala está definida por

$$\tau^2(\mathbf{v}) = (s^2(\mathbf{v})/n) \sum_{i=1}^n \rho_2(|v_i|/s(\mathbf{v})),$$

donde  $\mathbf{v} = (v_1, \dots, v_n)$ ,  $\rho_2$  es una función- $\rho$  y  $s$  es un M-estimador de escala.

Los estimadores robustos están basados en funciones- $\rho$  bicuadradas. El M-estimador de escala usado en el S-estimador, está definido mediante una función

$$\rho_0(u) = \rho_B(u/c_0),$$

Estimador	$q = 2$				$q = 5$			
	EMCP	ES	ER	ERA	EMCP	ES	ER	ERA
EMV	0.041	0.001	1.00	1.00	0.103	0.002	1.00	1.00
S-estimador	0.074	0.002	0.55	0.58	0.125	0.002	0.83	0.85
$\tau$ -estimador	0.046	0.001	0.89	0.90	0.116	0.002	0.90	0.90
MM-estimador	0.046	0.001	0.89	0.90	0.116	0.002	0.90	0.90

Tabla 5.2: Simulación: error medio cuadrático podado (EMCP), error estándar del EMCP (ES), eficiencia relativa (ER) y eficiencia relativa asintótica (ERA) de los estimadores en el caso no contaminado para  $n = 100$  y  $p = 2$ .

donde  $c_0$  (ver Tabla 3.1) es la misma constante que se usa para calcular el M-estimador de escala utilizado en el procedimiento para obtener un MM-estimador, y  $b = 0.5$  con lo cual el S-estimador tiene punto de ruptura 0.5. El  $\tau$ -estimador usa la misma  $\rho_0$  y  $b$  que el S-estimador para calcular la M-escala y para la  $\tau$ -escala usa

$$\rho_2(u) = \rho_B(u/c_2),$$

donde  $c_2$  es elegida junto con la constante  $\kappa$ , de la ecuación (5.2), de modo que el  $\tau$ -estimador tenga una eficiencia relativa asintótica igual a 0.90 cuando los errores son normales (ver Tabla 5.1, también puede consultarse la Tabla 2 de [13], tener en cuenta que la funciones bicuadradas difieren en una constante y por lo tanto  $\kappa = 6\kappa_2/c_2^2$ ). Para calcular un  $\tau$ -estimador es necesario tener un estimador inicial el cual se calcula mediante submuestreo, en esta simulación usamos 2000 submuestras. El MM-estimador usa la misma  $\rho_0$  que el S-estimador, como estimador inicial para calcular el M-estimador de escala usa el S-estimador y

$$\rho_1(u) = \rho_B(u/c_1),$$

donde  $c_1$  es elegido de manera tal que el MM-estimador tenga eficiencia asintótica relativa igual a 0.90 cuando los errores son normales (ver Tabla 3.2). El valor de  $\delta$  en el paso 5 del algoritmo es tomado igual a  $10^{-4}$ .

La Tabla 5.2 muestra los errores medios cuadráticos podados, los errores estándar y las eficiencias relativas y eficiencias relativas asintóticas con res-

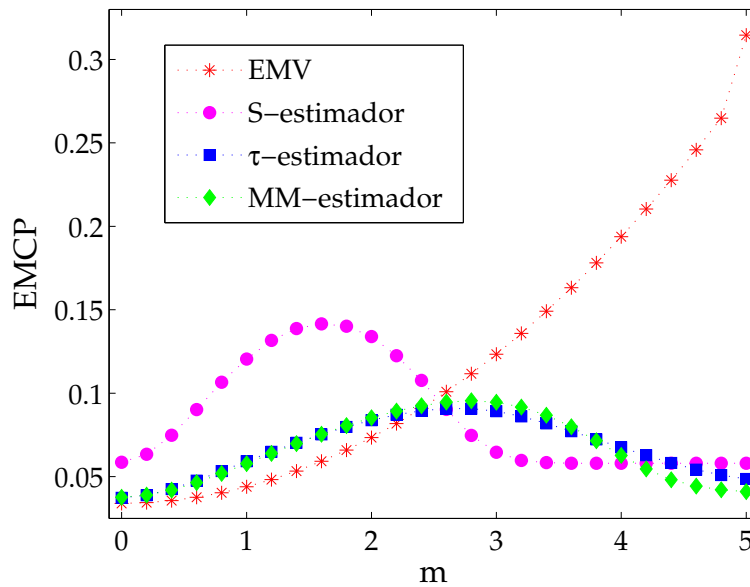


Figura 5.1: Simulación: error medio cuadrático podado excluyendo el 10 % de las observaciones más grandes para  $q = 2$  y  $x_0 = 1$ .

pecto al estimador de máxima verosimilitud para el caso no contaminado. Las eficiencias relativas de todos los estimadores robustos (calculadas mediante el cociente de sus respectivos EMCP y el EMCP del EMV) son similares a sus valores asintóticos.

En las Figuras 5.1, 5.2 y 5.3 se muestra el error medio cuadrático podado de los distintos estimadores bajo contaminación.

En la Figura 5.1, la cual corresponde al caso  $q = 2$  y  $x_0 = 1$ , observamos que el MM-estimador y el  $\tau$ -estimador tienen un EMCP menor que el del S-estimador excepto cuando  $m$  está (aproximadamente) entre 2.8 y 4, y que los MM- y  $\tau$ -estimadores se comportan de manera similar. En este caso, el S-estimador tiene el mayor EMCP máximo entre los estimadores robustos. Como se esperaba, el ECMP del estimador de máxima versomilitud crece con  $m$  alcanzando valores muy grandes. En la Figura 5.2 mostramos los resultados para  $q = 2$  y  $x_0 = 10$ . Los S-,  $\tau$ - y MM-estimadores se comportan de manera similar. En la Figura 5.3, la cual corresponde al caso  $q = 5$  y  $x_0 = 1$ , observamos que el S-estimador tiene un ECMP menor que el del MM-estimador y el

del  $\tau$ -estimador excepto cuando  $m$  está (aproximadamente) entre 0 y 4.8.

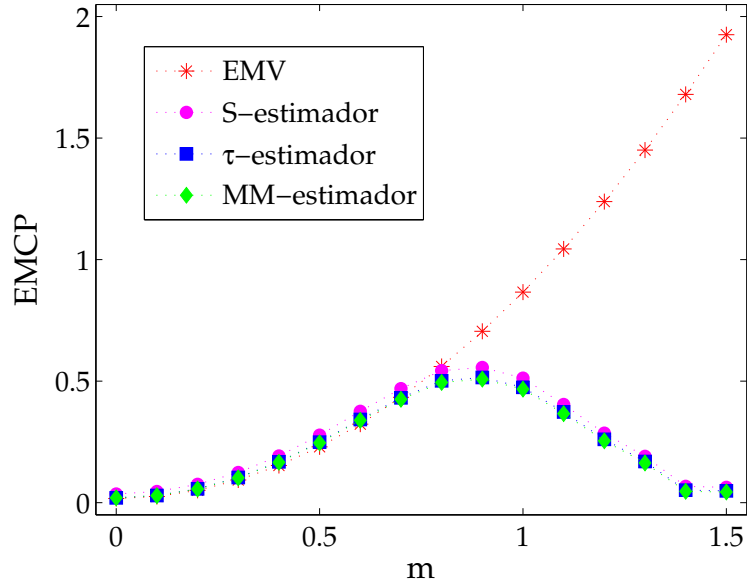


Figura 5.2: Simulación: error medio cuadrático podado excluyendo el 10 % de las observaciones más grandes para  $q = 2$  y  $x_0 = 10$ .

Los  $\tau$ - y MM-estimadores se comportan de modo similar excepto cuando  $m$  está (aproximadamente) entre 4.4 y 7 donde el MM-estimador tiene un EMCP menor que el del  $\tau$ -estimador. Como se puede ver en la Figura 5.4, para  $q = 5$  y  $x_0 = 10$  el comportamiento de los estimadores robustos es similar al ya observado para el caso  $q = 2$  y  $x_0 = 10$  (Figura 5.2).

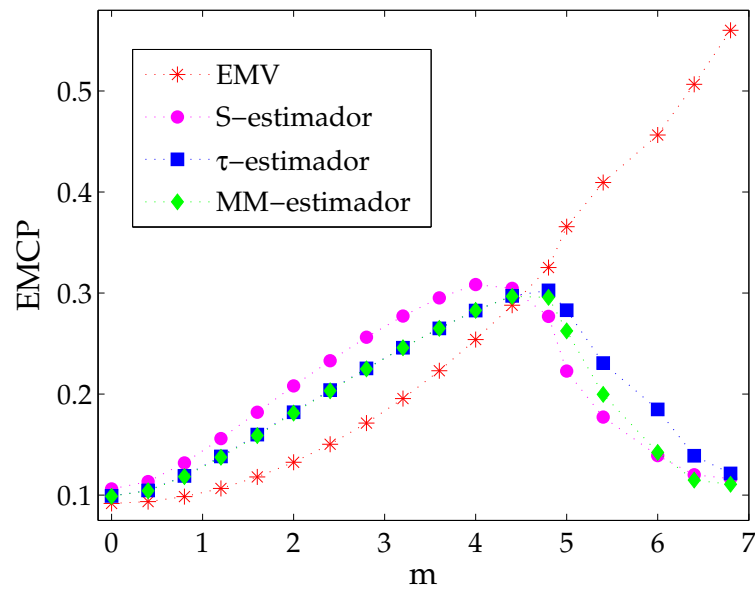


Figura 5.3: Simulación: error medio cuadrático podado excluyendo el 10 % de las observaciones más grandes para  $q = 5$  y  $x_0 = 1$ .

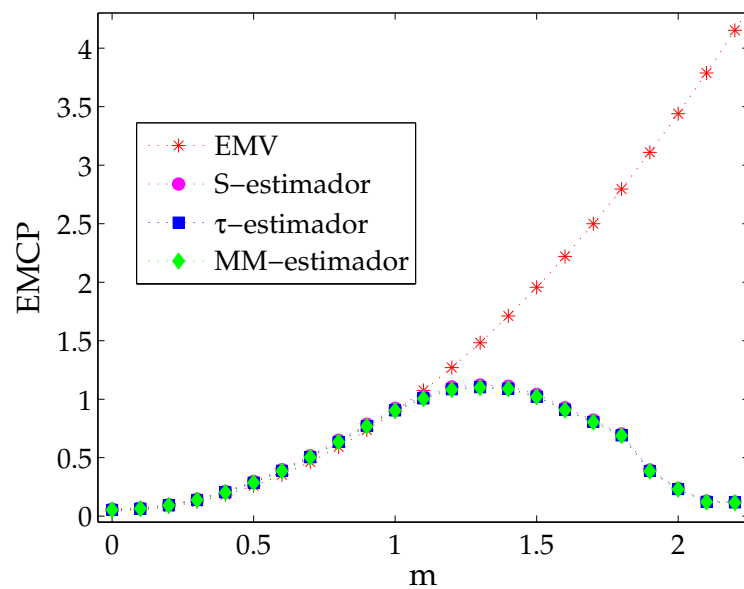


Figura 5.4: Simulación: error medio cuadrático podado excluyendo el 10 % de las observaciones más grandes para  $q = 5$  y  $x_0 = 10$ .

## 5.2. Ejemplo con Datos Reales

Analizamos un conjunto de datos correspondiente al microanálisis de rayos X de vasijas de vidrio arqueológicas (Janssens et al., [19]), donde cada una de las  $n = 180$  vasijas es representada por un espectro de 1920 frecuencias. Para cada vasija se registró el contenido de 13 compuestos químicos. Para limitar el tamaño de nuestro conjunto de datos, empleamos sólo dos componentes,  $P_2O_5$  y  $PbO$ ; y elegimos 12 frecuencias equiespaciadas entre 100 y 400. Los valores de  $x_{ij}$  son casi nulos para las frecuencias por debajo de 100 y por arriba de 400. Por lo tanto tenemos  $p = 13$  y  $q = 2$ .

Calculamos tres estimadores de los coeficientes de regresión: el EMV, un MM-estimador y un  $\tau$ -estimador. Como estimador inicial para el MM-estimador usamos un S-estimador. El  $\tau$ -estimador, el S-estimador y el MM-estimador usan funciones- $\rho$  de la familia bicuadrada con constantes tales que los MM- y  $\tau$ -estimadores tengan una eficiencia relativa asintótica igual a 0.95 cuando los errores son normales y que el S-estimador tenga punto de ruptura 0.5. En la Figura 5.5 están representados los diagramas de cuantiles (QQ-plots) de las normas de Mahalanobis de los residuos del EMV y del MM-estimador contra la raíz cuadrada de los cuantiles de la distribución chi-cuadrado con  $q$  grados de libertad. El diagrama de cuantiles del MM-estimador muestra claramente la existencia de datos atípicos.

El comportamiento predictivo de los estimadores se evaluó a través de validación cruzada en 5 partes (5-fold cross-validation); los criterios utilizados fueron el error cuadrático medio, el error cuadrático medio con el 10% superior de los datos podados y una  $\tau$ -escala de los errores de predicción (con función- $\rho$  bicuadrada con constante 4), calculados para cada componente de las respuestas por separado. La Tabla 5.3 muestra los resultados. Para los estimadores robustos, los resultados obtenidos con las medidas de desempeño robustas son muy distintos a los obtenidos con el EMC ordinario. En cambio, el EMV no muestra el mismo comportamiento. Como se esperaba, los errores medios cuadráticos podados del EMV son mayores que los de los estimadores robustos. El comportamiento predictivo del MM-estimador, medido a través del EMCP, es mejor que el del  $\tau$ -estimador para la primer componente de la

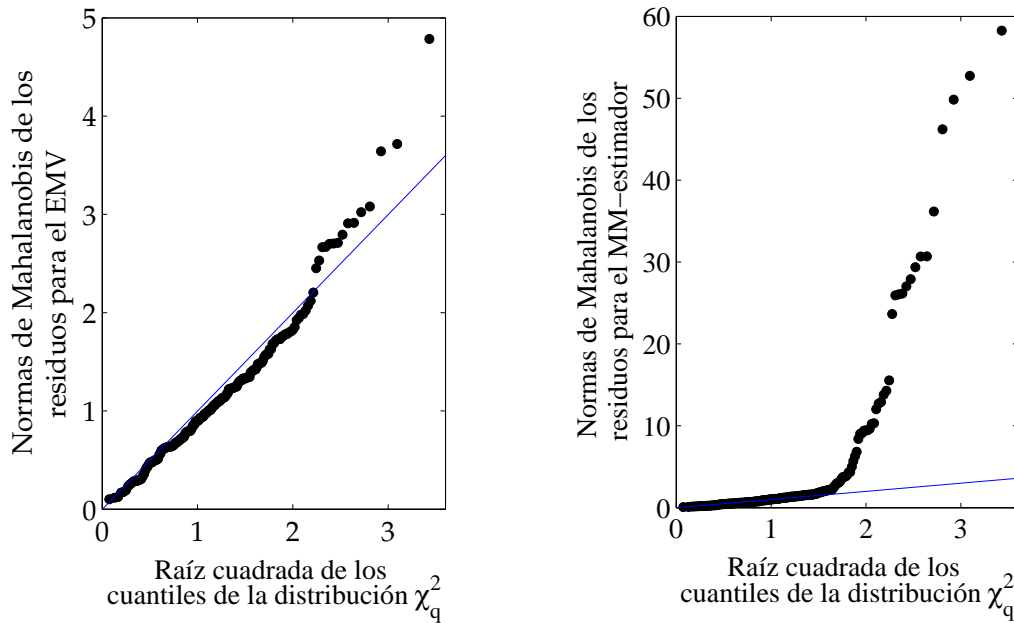


Figura 5.5: Diagramas de cuantiles de las normas de Mahalanobis de los residuos del EMV (izquierda) y del MM-estimador (derecha).

respuesta, y es similar para la segunda. Por otro lado, el  $\tau$ -estimador es ligeramente mejor que el MM-estimador para la primer componente de la respuesta, si se considera como medida de desempeño a la  $\tau$ -escala de los errores de predicción, y es similar para la segunda.

Criterio	EMV		$\tau$ -estimador		MM-estimador	
	1	2	1	2	1	2
EMC	0.0810	0.0511	0.3512	0.8062	0.3400	0.6825
EMCP	0.0405	0.0200	0.0306	0.0089	0.0242	0.0082
$\tau$ -escala	0.0437	0.0217	0.0053	0.0072	0.0085	0.0066

Tabla 5.3: Errores medios cuadráticos de predicción (EMC), errores medios cuadráticos podados de predicción (EMCP) y  $\tau$ -escala de los errores de predicción, de cada uno de los estimadores, calculados por validación cruzada.



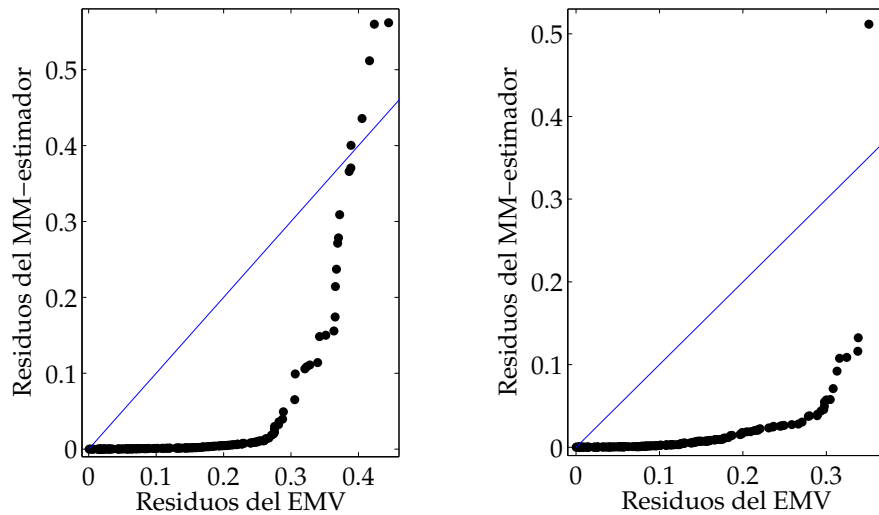


Figura 5.6: Diagramas de cuantiles de los valores absolutos de los residuos ordenados del EMV el MM-estimador para cada componente de la respuesta, calculados por validación cruzada. El izquierdo corresponde a  $P_2O_5$  y el derecho al  $PbO$ .

En la Figura 5.6 se comparan los valores absolutos de los residuos ordenados de los MM-estimadores con los obtenidos mediante el EMV para cada componente de la respuesta, calculados mediante validación cruzada. Por razones de escala en cada diagrama de cuantiles se eliminaron las observaciones correspondientes a los 12 mayores valores absolutos. Se puede ver que la mayoría de los puntos caen por debajo de la recta que representa a la función identidad, mostrando que el MM-estimador provee un mejor ajuste para la mayor parte de la muestra. Los resultados obtenidos de este gráfico, muestran que el error medio cuadrático no es una medida de desempeño confiable en presencia de datos atípicos y justifican el uso de medidas de desempeño robustas para comparar el comportamiento predictivo de los estimadores robustos.

## Capítulo 6

# Aplicación al Análisis de Correlación Canónica

---

Como una aplicación de los MM-estimadores, en este capítulo presentaremos un método robusto para el análisis de correlación canónica.

El análisis de correlación canónica (ACC) es una herramienta estándar, introducida por Hotelling [16], para descubrir y cuantificar la asociación lineal entre dos conjuntos de variables. Supongamos que tenemos dos vectores aleatorios  $\mathbf{x} = (x_1, \dots, x_p)'$  y  $\mathbf{y} = (y_1, \dots, y_q)'$  cuya distribución conjunta es  $H(\mathbf{x}, \mathbf{y})$ . Usaremos la siguiente notación para las esperanzas y covarianzas,

$$E(\mathbf{x}) = \boldsymbol{\mu}_{\mathbf{x}} \quad \text{y} \quad E(\mathbf{y}) = \boldsymbol{\mu}_{\mathbf{y}},$$

$$\text{Cov}(\mathbf{x}) = \mathbf{V}, \quad \text{Cov}(\mathbf{y}) = \mathbf{W}, \quad \text{Cov}(\mathbf{x}, \mathbf{y}) = \mathbf{C}$$

y denotamos la matriz de covarianza conjunta de  $\mathbf{z} = (\mathbf{y}', \mathbf{x}')'$  por

$$\text{Cov}(\mathbf{z}) = \begin{pmatrix} \mathbf{V} & \mathbf{C} \\ \mathbf{C}' & \mathbf{W} \end{pmatrix}. \quad (6.1)$$

El objetivo del ACC es encontrar vectores  $\mathbf{t} \in \mathbb{R}^q$  y  $\mathbf{v} \in \mathbb{R}^p$  tales que la correlación entre las combinaciones lineales  $\mathbf{t}'\mathbf{y}$  y  $\mathbf{v}'\mathbf{x}$  sea maximizada. Más generalmente, el análisis canónico busca  $K = \min(p, q)$  pares de combinaciones lineales tal que cada par sucesivo maximiza la correlación bajo la restricción

de ser incorrelacionada con los pares previos, es decir

$$(\mathbf{t}_k, \mathbf{v}_k) = \arg \max_{(\mathbf{t}, \mathbf{v})} \{\text{corr}(\mathbf{t}'\mathbf{y}, \mathbf{v}'\mathbf{x})\} \quad (6.2)$$

donde  $(\mathbf{t}, \mathbf{v})$  cumplen

$$\text{corr}(\mathbf{t}'_j\mathbf{y}, \mathbf{t}'_k\mathbf{y}) = \text{corr}(\mathbf{v}'_j\mathbf{x}, \mathbf{v}'_k\mathbf{x}) = 0, \quad j = 1, \dots, k-1 \quad (6.3)$$

Los vectores  $\{\mathbf{t}_k\}_1^K$  y  $\{\mathbf{v}_k\}_1^K$  son llamados las coordenadas canónicas de  $\mathbf{y}$  y  $\mathbf{x}$  respectivamente. Las combinaciones lineales  $\mathbf{t}'_k\mathbf{y}$  y  $\mathbf{v}'_k\mathbf{x}$  son llamadas el par de variables canónicas de orden  $k$  y sus respectivas correlaciones

$$\{c_k = \text{corr}(\mathbf{t}'_k\mathbf{y}, \mathbf{v}'_k\mathbf{x})\}_1^K \quad (6.4)$$

son conocidas como las correlaciones canónicas de  $H(\mathbf{x}, \mathbf{y})$ .

El criterio (6.2) es invariante con respecto a las escalas de las combinaciones lineales, por lo tanto hay infinitas soluciones posibles; esta ambigüedad es generalmente resuelta suponiendo que  $\text{Cov}(\mathbf{y})$  y  $\text{Cov}(\mathbf{x})$  son no singulares y luego estandarizando todas las combinaciones lineales para que tengan varianzas unitarias, es decir

$$\text{Var}(\mathbf{t}'_k\mathbf{y}) = \text{Var}(\mathbf{v}'_k\mathbf{x}) = 1 \quad k = 1, \dots, K. \quad (6.5)$$

Es bien sabido que las coordenadas canónicas  $\mathbf{t}_k$  y  $\mathbf{v}_k$ , soluciones del problema (6.2) sujeto a (6.3) y (6.4), son respectivamente los autovectores correspondientes a los autovalores  $c_1^2 \geq \dots \geq c_K^2 > 0$  de las matrices

$$\mathbf{W}^{-1}\mathbf{C}'\mathbf{V}^{-1}\mathbf{C} \quad \text{y} \quad \mathbf{V}^{-1}\mathbf{C}\mathbf{W}^{-1}\mathbf{C}'. \quad (6.6)$$

Ambas matrices tienen los mismos autovalores y son las correlaciones canónicas al cuadrado  $c_k^2$ .

Siguiendo la idea de Yohai y García Ben en [37], podemos pensar al ACC como método predictivo. Por eso nuestro enfoque no es simétrico en  $\mathbf{x}$  e  $\mathbf{y}$ . El ACC es utilizado en métodos de predicción para regresión multivariada como

Curds and Whey (C&W) (Breiman y Friedman [5]) y Regresión por Rango Reducido (Izenman [18]). Si uno quiere versiones robustas de estos métodos, tiene que usar un estimador robusto de regresión multivariada, en cuyo caso el ACC propuesto a continuación se obtiene casi sin esfuerzo adicional.

Sea  $\mathbf{A}$  la matriz de coeficientes de regresión de  $\mathbf{y}$  en  $\mathbf{x}$ , es decir  $\mathbf{A} = \mathbf{V}^{-1}\mathbf{C}$ . Sean

$$\hat{\mathbf{y}} = \boldsymbol{\mu}_{\mathbf{y}} + \mathbf{A}'(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}), \quad \mathbf{u} = \mathbf{y} - \hat{\mathbf{y}} \quad \mathbf{y} \quad \boldsymbol{\Sigma} = \text{Cov}(\mathbf{u})$$

Por lo tanto,  $\hat{\mathbf{y}}$  es el mejor predictor lineal de  $\mathbf{y}$  en  $\mathbf{x}$ . Notar que dado  $\mathbf{t}_k$ , (6.2) es equivalente a

$$\mathbf{v}_k = \arg \min_{\mathbf{v}} \{E(\mathbf{t}_k'(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}) - \mathbf{v}'(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}))^2\}. \quad (6.7)$$

Entonces, tenemos que

$$\mathbf{v}_k = \mathbf{V}^{-1}\mathbf{C}\mathbf{t} = \mathbf{A}\mathbf{t}_k \quad (6.8)$$

y el valor mínimo de (6.7) es

$$E(\mathbf{t}_k'(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}) - \mathbf{t}_k'\mathbf{A}'(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}))^2 = E(\mathbf{t}_k'(\mathbf{y} - \hat{\mathbf{y}}))^2 = E(\mathbf{t}_k'\mathbf{u})^2 = \mathbf{t}_k'\boldsymbol{\Sigma}\mathbf{t}_k.$$

Luego los vectores  $\{\mathbf{t}_k\}_1^K$  verifican

$$\mathbf{t}_k = \arg \min_{\mathbf{t}} \mathbf{t}'\boldsymbol{\Sigma}\mathbf{t} \quad (6.9)$$

donde

$$\mathbf{t}_j'\mathbf{W}\mathbf{t} = \delta_{jk} \quad \text{para todo } j = 1, \dots, k, \quad (6.10)$$

y entonces pueden obtenerse calculando los autovectores de

$$\boldsymbol{\Sigma}\mathbf{t} = \lambda\mathbf{W}\mathbf{t}, \quad (6.11)$$

tales que

$$\mathbf{t}'\mathbf{W}\mathbf{t} = 1, \quad (6.12)$$

asociados a los autovalores  $\lambda_1 \geq \dots \geq \lambda_K > 0$ . Por (6.8) y (6.5), se tiene que

$$\mathbf{v}_k = \mathbf{A}\mathbf{t}_k/b_k \quad (6.13)$$

donde

$$b_k^2 = (\mathbf{A}\mathbf{t}_k)' \mathbf{V} (\mathbf{A}\mathbf{t}_k). \quad (6.14)$$

Utilizando la igualdad

$$\mathbf{W} = \mathbf{\Sigma} + \mathbf{A}' \mathbf{V} \mathbf{A},$$

es fácil ver

$$\mathbf{v}_j' \mathbf{V} \mathbf{v}_k = 0 \quad \text{para todo } j \neq k,$$

que las coordenadas canónicas  $\mathbf{t}_k$  y  $\mathbf{v}_k$  obtenidas en (6.9)-(6.10) y (6.13) coinciden con las obtenidas por el método clásico, es decir con los autovectores de las matrices que aparecen en (6.6), y que las correlaciones canónicas verifican

$$c_k^2 = 1 - \lambda_k \quad \text{para todo } k = 1, \dots, K.$$

El método clásico para estimar las coordenadas y correlaciones canónicas consiste en tomar las matrices de covarianzas muestrales de  $\mathbf{W}$ ,  $\mathbf{V}$  y  $\mathbf{C}$  y calcular los autovectores y autovalores de las matrices en (6.6) utilizando dichas covarianzas muestrales. En nuestro enfoque se pueden estimar las coordenadas y correlaciones canónicas reemplazando las matrices de covarianzas en (6.11), (6.12) y (6.14) por sus respectivas covarianzas muestrales y reemplazando a la matriz de coeficientes de regresión  $\mathbf{A}$ , en (6.13) y (6.14), por el estimador de mínimos cuadrados. Como es bien sabido, las covarianzas muestrales y el estimador de mínimos cuadrados de la matriz de coeficientes no son resistentes a datos atípicos y por lo tanto las coordenadas y correlaciones canónicas, calculadas tanto con el método clásico como por el propuesto en esta tesis, también son sensibles a dichas observaciones.

El método robusto para el análisis de Correlación Canónica más evidente es el que se obtiene estimando todas las matrices de covarianzas que aparecen en (6.6) utilizando un método robusto y luego procediendo como en el caso clásico. Kärnel [20] utilizó como método de estimación robusto un M-estimador de posición y escala multivariado, la desventaja de esta elección es que las propiedades de robustez de los M-estimadores empeoran según aumenta la dimensión de las variables (para más detalle ver [23] pág. 186). En cambio, Croux y Dehon [6] usaron el estimador con determinante de la ma-

triz de covarianza mínimo (MCD) el cual tiene alto punto de ruptura. Oliveira y Branco presentaron un enfoque robusto distinto utilizando medidas robustas de correlación llamado método de búsqueda de proyecciones (projection pursuit approach, PPA). Este método permite obtener las variables canónicas secuencialmente. Wold en [35] presentó un método completamente diferente para calcular las variables canónicas, utilizando regresiones alternantes (alternating regressions) y evitando el uso de las matrices de covarianza. Para obtener un método robusto, Filzmoser et al. [12] utilizaron una regresión alternante robusta (RAR). Para una descripción más detallada de estos métodos ver [4].

Para robustificar el método propuesto en (6.11) - (6.14) para obtener las coordenadas canónicas, estimaremos  $\Sigma$  y  $A$  mediante los MM-estimadores definidos en el Capítulo 2 y  $V$  y  $W$  mediante los MM-estimadores de posición y escala multivariados.

**Observación 5.** Como  $\hat{y} = \mu_y + A'(x - \mu_x)$ , la matriz de coeficientes  $A$  no contiene al intercept. Por lo tanto, primero deben calcularse los MM-estimadores  $(\hat{A}_0, \hat{\Sigma}_n)$  para el modelo lineal

$$y = A'_0 x_0 + u$$

donde  $A_0 = (\mu_y - A'\mu_x, A)$  y  $x_0 = (1, x')'$ .

Para estimar las correlaciones canónicas en forma robusta proponemos dos métodos diferentes:

**Método 1:** Sean  $(y_i, x_i)$ , con  $i = 1, \dots, n$ , una muestra aleatoria de variables reales y  $s$  una escala robusta. Consideremos a los estimadores de regresión  $\hat{\beta}_0$  y  $\hat{\beta}$  definidos por

$$(\hat{\beta}_0, \hat{\beta}) = \arg \min_{\beta} s(y_i - \beta_0 - \beta x_i),$$

y al estimador de posición  $\hat{\mu}$  definido por

$$\hat{\mu} = \arg \min_{\mu} s(y_i - \mu).$$

Entonces definimos

$$\widehat{\text{Corr}}^2(y, x) = 1 - \frac{s(y_i - \hat{\beta}_0 - \hat{\beta}x_i)^2}{s(y_i - \hat{\mu})^2}. \quad (6.15)$$

Este estimador, al que denotaremos por  $R_s^2$ , fue propuesto por Croux y Dehon en [7].

Método 2: Sea  $(y_i, x_i)$ , con  $i = 1, \dots, n$ , una muestra aleatoria de variables reales. Consideremos el MM-estimador de regresión  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta})$  definido por Yohai en [36] el cual coincide con el MM-estimador definido en el Capítulo 2 tomando  $y_i = y_i$  y  $\mathbf{x}_i = (1, x_i)'$  y sea  $\hat{\sigma}_n$  un M-estimador de escala de los residuos. Sean

$$w_i = W\left(\frac{y_i - \hat{y}_i}{\hat{\sigma}_n}\right),$$

donde  $W$  es la función definida en el Teorema 2 y  $\hat{y}_i = \hat{\beta}' \mathbf{x}_i$ , y

$$\bar{\hat{y}}_w = \left(\frac{1}{\sum_{i=1}^n w_i}\right) \sum_{i=1}^n w_i \hat{y}_i.$$

Entonces el estimador propuesto por Renaud y Victoria-Feser en [26], está definido por

$$R_a^2 = \frac{\sum_{i=1}^n w_i (\hat{y}_i - \bar{\hat{y}}_w)^2}{\sum_{i=1}^n w_i (\hat{y}_i - \bar{\hat{y}}_w)^2 + a \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2}, \quad (6.16)$$

donde  $a$  es un factor de corrección que asegura la consistencia del estimador  $R_a^2$  para la eficiencia deseada del estimador  $\hat{\beta}$ .

**Teorema 10.** Sean  $(y_i, x_i)$ , con  $i = 1, \dots, n$ , una muestra aleatoria que satisface el modelo lineal

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

Asumiendo que los errores  $u_i$  tienen distribución normal con media 0 y varianza  $\sigma^2$  y que el M-estimador de escala  $\hat{\sigma}_n$  es consistente, el estimador  $R_a^2$  dado en (6.16) es consistente a  $\text{Corr}^2(y, x)$  cuando  $a = E(W(u/\sigma))/E(u\psi_1(u/\sigma)/\sigma)$  donde  $\psi_1$  es la derivada de la  $\rho$ -función,  $\rho_1$ , utilizada para calcular los MM-estimadores de  $\beta_0$  y  $\beta_1$ .

La demostración de este Teorema puede encontrarse en [26]. A continuación se presenta una tabla con los valores de  $a$  para diferentes eficiencias de los MM-estimadores calculados con una función  $\rho_1$  bicuadrada.

Eficiencia	0.95	0.90	0.85	0.80
$a$	1.2076	1.3188	1.4202	1.5222

Tabla 6.1: Valores de  $a$  para distintas elecciones de eficiencia del MM-estimador usando la función bicuadrada.

Entonces, para estimar la correlación canónica al cuadrado,  $c_k^2$ , se debe tomar  $y_i = \mathbf{t}_k' \mathbf{y}_i$  y  $x_i = \mathbf{v}_k' \mathbf{x}_i$  y aplicar (6.15) o (6.16) según el método que se desea utilizar.

## 6.1. Estudio de Simulación

Para comparar el desempeño de los estimadores de las coordenadas y correlaciones canónicas propuestos en este trabajo realizamos un estudio de simulación.

Consideramos  $\mathbf{y}_i \in \mathbb{R}^4$  y  $\mathbf{x}_i \in \mathbb{R}^5$  que satisfacen el modelo lineal multivariado dado por (1.1) con

$$B_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Los errores  $\mathbf{u}_i$  son generados a partir de una distribución  $N_4(\mathbf{0}, \mathbf{I})$  y los predictores  $\mathbf{x}_i$  de una distribución  $N_5(\mathbf{0}, \mathbf{I})$ . Debido a la equivarianza de los MM-estimadores, estas elecciones no representan una pérdida de generalidad para los estimadores propuestos.



Se generaron 300 muestras de tamaño 100. Consideramos muestras no contaminadas y muestras que contienen un 10 % de datos atípicos idénticos de la forma  $(\mathbf{x}_0, y_0)$  con

$$\mathbf{x}_0 = (0, 0, 0, 0, 1) \quad \text{y} \quad y_0 = (0, 0, 0, m).$$

Tomamos una malla de valores  $m$ , comenzando en el 0. La malla se eligió de manera tal que todos los estimadores robustos alcanzaran el valor máximo de sus medidas de desempeño.

Para cada réplica  $j$  ( $j = 1, \dots, 300$ ) se obtuvieron las estimaciones de las correlaciones canónicas, denotadas por  $\{\hat{c}_k^{(j)}\}_{k=1}^4$ , y las coordenadas canónicas de  $\mathbf{x}$  que están representadas por las columnas de la matriz  $\hat{\Delta}^{(j)}$ .

Como medida de desempeño para cada correlación canónica, se calculó el error relativo medio

$$ECM(\hat{c}_k) = \frac{1}{300} \sum_{j=1}^{300} \frac{(\hat{c}_k^{(j)} - c_k)^2}{c_k^2} \quad \text{para } k = 1, \dots, 4, \quad (6.17)$$

donde  $c_1^2 = 10^2/(1 + 10^2) \approx 0.99$ ,  $c_2^2 = 5^2/(1 + 5^2) \approx 0.96$  y  $c_3^2 = c_4^2 = 0.5$ , y para las coordenadas canónicas, se calculó

$$DE(\hat{\Delta}) = \frac{1}{300} \sum_{j=1}^{300} |E(\mathbf{y} - \mathbf{y}_{\mathbf{z}_j}^*)(\mathbf{y} - \mathbf{y}_{\mathbf{z}_j}^*)'|,$$

donde  $\mathbf{y}_{\mathbf{z}_j}^*$  es el mejor predictor lineal de  $\mathbf{y}$  basado en  $\mathbf{z}_j = \hat{\Delta}^{(j)'} \mathbf{x}$

Yohai y García Ben en [37] demostraron que el problema de elegir una matriz  $\Delta^* \in \mathbb{R}^{p \times K}$  tal que  $\mathbf{z}^* = \Delta^{*'} \mathbf{x}$  minimice

$$|E(\mathbf{y} - \mathbf{y}_{\mathbf{z}}^*)(\mathbf{y} - \mathbf{y}_{\mathbf{z}}^*)'| \quad (6.18)$$

entre todos los vectores de dimensión  $K$  de la forma  $\mathbf{z} = \Delta^{*'} \mathbf{x}$  tiene como solución a la matriz cuyas columnas son las coordenadas canónicas de  $\mathbf{x}$ .

Es fácil ver que para  $\mathbf{y} = \mathbf{B}_0' \mathbf{x} + \mathbf{u}$ , donde  $\mathbf{x} \in \mathbb{R}^p$  y  $\mathbf{u} \in \mathbb{R}^q$  tienen media cero,  $\text{Cov}(\mathbf{x}) = \mathbf{I}_p$  y  $\text{Cov}(\mathbf{u}) = \mathbf{I}_q$ , el determinante en (6.18) puede reescribirse

como

$$|\mathbf{I}_q + \mathbf{B}'_0(\mathbf{I}_p - \Delta^* \Delta^{*'})\mathbf{B}_0|. \quad (6.19)$$

Para el caso en que  $K = q = p - 1$ , como ocurre en este estudio de simulación, tenemos que

$$\Delta^* \Delta^{*'} = \mathbf{I}_p - \mathbf{v}_p^* \mathbf{v}_p^{*'} \quad (6.20)$$

donde  $\mathbf{v}_p^*$  es el vector que junto con los vectores columna de  $\Delta^*$  forman una base ortonormal de  $\mathbb{R}^p$ .

Usando (6.20) tenemos que (6.19) resulta igual a

$$|\mathbf{I}_q + \mathbf{B}'_0 \mathbf{v}_p^* (\mathbf{B}'_0 \mathbf{v}_p^*)'| = 1 + \text{tr}(\mathbf{B}'_0 \mathbf{v}_p^* (\mathbf{B}'_0 \mathbf{v}_p^*)') = 1 + \|\mathbf{B}'_0 \mathbf{v}_p^*\|^2. \quad (6.21)$$

Como se puede observar en este caso,  $\mathbf{z}^* = \Delta^{*'} \mathbf{x}$  también minimiza la expresión

$$E\|\mathbf{y} - \mathbf{y}_z^*\|^2 = \text{Tr}((\mathbf{y} - \mathbf{y}_z^*)(\mathbf{y} - \mathbf{y}_z^*)')$$

entre todos los  $\mathbf{z} = \Delta' \mathbf{x}$  con  $\Delta \in \mathbb{R}^{K \times p}$ .

De manera análoga, bajo las mismas condiciones en las que se obtuvo (6.19), se puede probar que

$$DE(\hat{\Delta}) = 1 + \frac{1}{300} \sum_{j=1}^{300} \|\mathbf{B}'_0 \mathbf{v}_p^{(j)}\|^2, \quad (6.22)$$

donde  $\mathbf{v}_p^{(j)}$  es el vector que junto con los vectores columna de  $\hat{\Delta}^{(j)} \left( \hat{\Delta}^{(j)'} \hat{\Delta}^{(j)} \right)^{-1/2}$  forman una base ortonormal de  $\mathbb{R}^p$ .

Como se mencionó en la Sección 5.1, bajo distribuciones contaminadas los estimadores robustos pueden tener colas pesadas, por lo que es prudente evaluar su desempeño mediante medidas robustas. Por esta razón, también calculamos las medidas  $ECMP(\hat{c}_k)$  y  $DEP(\hat{\Delta})$  las cuales se obtienen reemplazando los promedios en (6.17) y (6.22) por promedios con el 10 % (superior) podado. A continuación se muestran los resultados para las medidas calculadas con promedios podados. Se obtienen conclusiones similares utilizando las medi-

das calculadas con promedios ordinarios.

Se comparará el desempeño de los siguientes estimadores:

Clásico: Basado en el análisis de correlación canónica clásico se calcularon los autovalores y autovectores de las matrices en (6.6) estimadas mediante las matrices de covarianzas muestrales de  $V$ ,  $W$  y  $C$ .

MCD: Basado en el análisis de correlación canónica clásico se calcularon los autovalores y autovectores de las matrices en (6.6) estimadas mediante un estimador con Determinante de la matriz de Covarianza Mínimo  $(p + q)$ -dimensional descompuesto de acuerdo a (6.1).

PP-MCD, PP-SPM: Utilizando técnicas de búsqueda de proyecciones (PPA), se buscan las direcciones  $t_1$  y  $v_1$  que maximizan la correlación de las variables  $y$  y  $x$  proyectadas sobre esas direcciones, es decir,  $\text{Corr}(t_1'y, v_1'x)$ . La expresión a maximizar es llamada índice de proyección (projection pursuit index, IP). En el caso en que el IP sea el coeficiente de correlación de Pearson, lo que se obtiene son las primeras coordenadas canónicas clásicas y su correspondiente correlación. Repitiendo el mismo procedimiento pero agregando la restricción de que las proyecciones a las nuevas direcciones sean incorrelacionadas con las proyecciones a las direcciones ya calculadas obtenemos el resto de las correlaciones y coordenadas canónicas clásicas. Intuitivamente, la forma más simple de obtener estimadores robustos sería tomando IP robustos. En este estudio de simulación se consideraron los siguientes IP:

- Correlación proveniente de un MCD-estimador bivariado: Dada una variable aleatoria bidimensional  $z = (v, w)'$  la medida de correlación entre  $v$  y  $w$  se obtiene del MCD-estimador de la matriz de covarianzas de  $z$ .
- Coeficiente de correlación de Spearman: Esta medida de correlación  $\rho_S(v, w)$  es definida como la correlación entre los rangos de las dos variables aleatorias  $v$  y  $w$ . Si llamamos  $d_i$  a las diferencias entre los rangos de  $v_i$  y  $w_i$ , con  $i = 1, \dots, n$ , si no hay observaciones repetidas,  $\rho_S$  está dado por

$$\rho_S = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}.$$

RAR: El método de regresiones alternantes puede resumirse de la siguiente manera:

Supongamos que tenemos un valor inicial de la coordenada canónica  $t$ , luego el problema de maximización en (6.2) se reduce a

$$\mathbf{v} = \arg \max_{\mathbf{a}} \text{Corr}(\mathbf{t}'\mathbf{y}, \mathbf{a}'\mathbf{x}).$$

Usando herramientas básicas de regresión múltiple, podemos ver que  $\mathbf{v}$  es proporcional al coeficiente de regresión  $a_1$  del modelo

$$\mathbf{t}'\mathbf{y} = \mathbf{a}'_1\mathbf{x} + \gamma_1 + \varepsilon_1.$$

En cambio, si fijamos  $\mathbf{v}$ , el  $t$  óptimo se obtiene de

$$t = \arg \max_{\mathbf{b}} \text{Corr}(\mathbf{b}'\mathbf{y}, \mathbf{t}'\mathbf{x}),$$

y a su vez es proporcional al coeficiente de regresión  $b_2$  del modelo

$$\mathbf{v}'\mathbf{x} = \mathbf{b}'_2\mathbf{y} + \gamma_2 + \varepsilon_2.$$

Luego, las coordenadas canónicas se estiman mediante regresiones alternantes. En el caso clásico, las estimaciones se realizan mediante el método de mínimos cuadrados alternantes. Para el caso robusto, se utiliza una versión robusta de mínimos cuadrados alternante basada en el estimador de Mínimos Cuadrados Podado (Least Trimmed Squares) de Rousseeuw [28].

MM-S, -T, -Ra: Las coordenadas canónicas se calcularon como se explicó en la sección anterior utilizando funciones- $\rho$  de la familia bicuadrada y como estimador inicial para el MM-estimador se usó un S-estimador de regresión multivariado. Las constantes de la Observación 1 fueron elegidas como  $c_0 = 4.10$  (ver Tabla 3.1) y  $c_1 = 4.41$  (ver Tabla 3.2). Para el MM-S y MM-T las correlaciones canónicas al cuadrado se calcularon mediante (6.15) usando una M-escala y una  $\tau$ -escala, respectivamente. Para la M-escala se utilizó la función- $\rho$   $\rho_0(u) = \rho_B(u/c_0)$  con  $c_0 = 1.56$  (ver tabla 3.1). Para la  $\tau$ -escala, definida en (5.1),

Método	Clásico	MM	MCD	PP-MCD	PP-SPM	RAR
$DEP(\hat{\Delta})$	1.0382	1.0411	1.0526	1.2403	1.0538	1.0763

Tabla 6.2: Medias podadas de los determinantes de las matrices del error de estimación obtenidas en las 300 réplicas ( $DEP$ ), para cada método en el caso no contaminado.

se utilizaron  $\rho_1 = \rho_0$  para calcular la M-escala y  $\rho_2(u) = \rho_B(u/c_2)$  con  $c_2 = 4.39$  (lo que asegura una eficiencia relativa asintótica igual a 0.85). Las correlaciones canónicas para MM-Ra se calcularon mediante (6.16) utilizando  $a = 1.4202$  (ver Tabla 6.1).

Para más detalles sobre los métodos MCD, PP-MCD, PP-SPM y RAR ver [4]. Los códigos usados para calcular los estimadores antes mencionados están disponibles en <http://www.statistik.tuwien.ac.at/public/filz/programs.html>.

En las Tablas 6.2 y 6.3 se muestran las medidas de desempeño podadas  $DEP$  y la raíz cuadrada del  $EMCP$  ( $REMCP$ ), de los distintos métodos calculados en la simulación, para el caso de muestras no contaminadas.

En la Tabla 6.2 se puede observar que el método clásico tiene el menor valor de  $DEP(\hat{\Delta})$ . Los métodos MM, MCD, PP-SPM y RAR muestran desempeños aceptables comparados con el método clásico en cuanto a la estimación de las coordenadas canónicas, pues tienen valores de  $DEP(\hat{\Delta})$  próximos al del método clásico. Y entre éstos, el método MM es el que tiene menor valor de dicha medida de desempeño. En la Tabla 6.3 se muestra la raíz de los errores cuadráticos medios podados (RECMC) para cada uno de los estimadores de las correlaciones canónicas. Como era de esperarse, el método clásico es el que muestra el mejor desempeño. Entre los métodos robustos, MM- $\tau$  es el que toma valores de la RECMC más próximos a los del método clásico y salvo los métodos PP-MCD y PP-SPM, que son menos precisos al estimar las correlaciones máximas, el resto de los métodos tienen desempeños aceptables.

Método	Clásico	MM-S	MM- $\tau$	MM-Ra	MCD	PP-MCD	PP-SPM	RAR
REMCP( $\hat{c}_1$ )	0.0008	0.0012	0.0009	0.0011	0.0010	0.0106	0.0079	0.0011
REMCP( $\hat{c}_2$ )	0.0031	0.0042	0.0032	0.0042	0.0040	0.0448	0.1650	0.0062
REMCP( $\hat{c}_3$ )	0.0689	0.0997	0.0707	0.0801	0.0859	0.1010	0.0710	0.0762
REMCP( $\hat{c}_4$ )	0.0715	0.1022	0.0766	0.1332	0.0887	0.0964	0.2435	0.1077

Tabla 6.3: Raíz de los Errores cuadráticos medios podados de las correlaciones canónicas obtenidas en las 300 réplicas (RECMCP), para cada método en el caso no contaminado.

En las Figuras 6.1, 6.2, 6.3 y 6.4 se muestran los valores de las medidas de desempeño calculadas en el estudio de simulación, para los distintos métodos bajo contaminación.

En la Figura 6.1, la cual corresponde a la media podada de los determinantes de las matrices del error de estimación (DEP), observamos que el método MM tiene el menor valor máximo de la medida DEP. Además para cada valor de  $m$ , el resultado de la medida DEP obtenido con el método MM es menor que los obtenidos con los otros estimadores robusto. Los valores de DEP para el método clásico crecen con  $m$  debido a su falta de robustez.

En el gráfico superior de la Figura 6.2, correspondiente a los ECMP de la máxima correlación para todos los métodos considerados en el estudio de simulación, se puede ver que los métodos PP-MCD, PP-SPM y el clásico tienen los mayores ECMP máximos. En el gráfico inferior de la Figura 6.2 se muestra nuevamente el desempeño de los estimadores robustos y se omite el de los métodos PP-MCD y PP-SPM por razones de escala. Los estimadores MM-S,  $-\tau$ ,  $-\text{Ra}$  y RAR alcanzan su máximo ECMP para valores de  $m$  entre 3 y 5, mientras que el MCD alcanza su máximo ECMP aproximadamente en  $m = 6$  y es ligeramente más grande que el máximo ECMP de los MM y del método RAR.

En la Figura 6.3 se muestran los ECMP de la segunda correlación canónica estimada. Se omitieron los valores del ECMP( $\hat{c}_2$ ) para los estimadores PP-MCD y PP-SPM ya que varían entre los intervalos (0.0015, 0.004) y (0.022, 0.036) res-

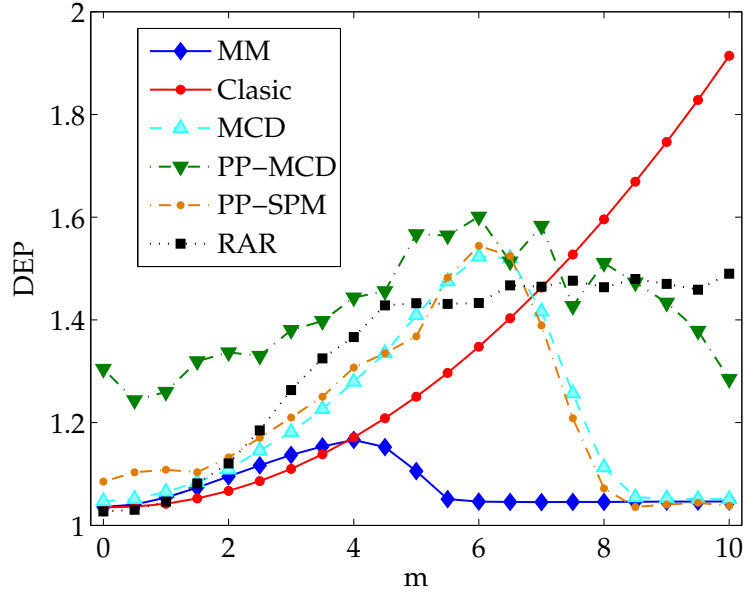


Figura 6.1: Medias podadas de los determinantes de las matrices del error de estimación (DEP) para muestras de tamaño  $n = 100$  con un 10 % de contaminación.

pectivamente, mientras que los del resto de los estimadores robustos varían en el intervalo  $(0, 0.0003)$ . En esta Figura se observa que el menor ECMP máximo lo tiene el estimador de  $MM-\tau$ .

En la Figura 6.4, se muestran los ECMP de las correlaciones canónicas estimadas  $\hat{c}_3$  (izquierda) y  $\hat{c}_4$  (derecha). En el gráfico correspondiente a la menor correlación canónica se omitieron los resultados del método PP-SPM por ser todos mayores a 0.05. Los mejores desempeños los tienen el método clásico y  $MM-\tau$ . Esto se debe a que la contaminación introducida en la simulación sólo afecta a la estimación de las dos correlaciones canónicas más grandes, ya que las coordenadas canónicas son elegidas incorrelacionadas.

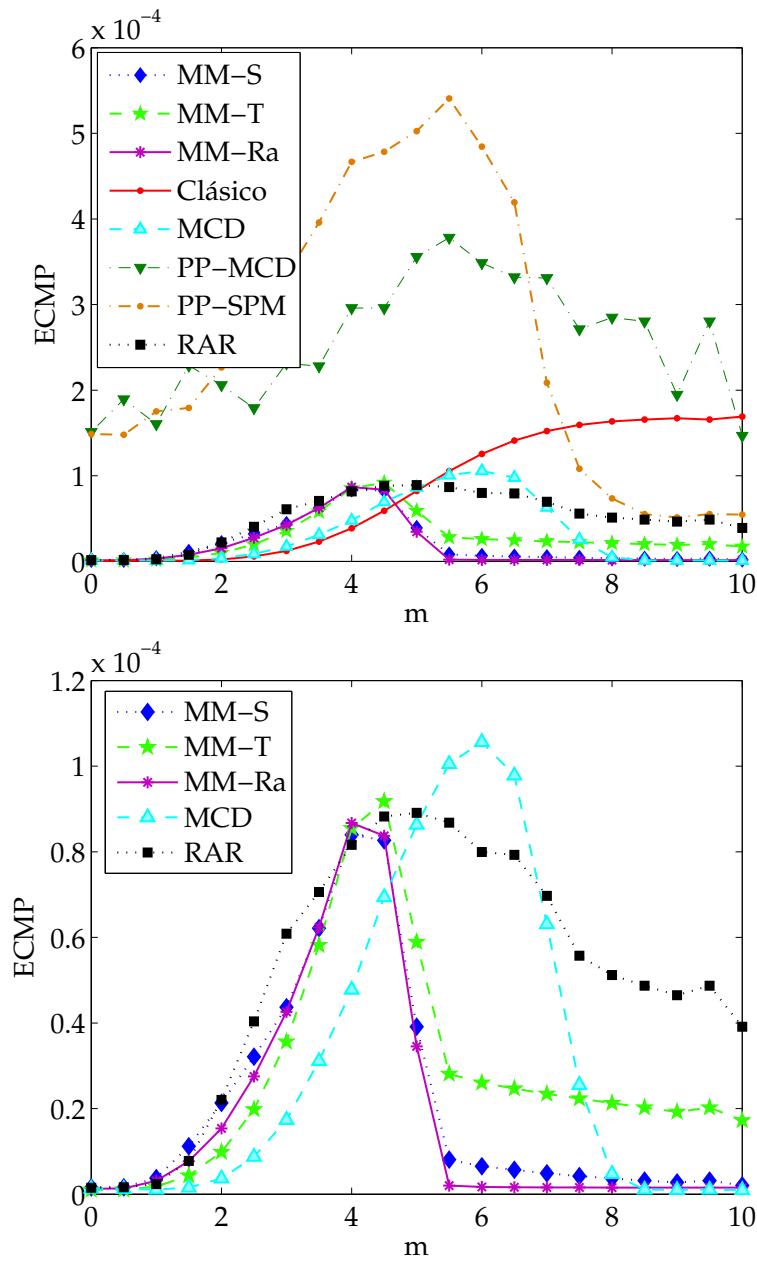


Figura 6.2: Errores cuadráticos medios podados (ECMP) de la correlación canónica máxima  $\hat{c}_1$  para muestras de tamaño  $n = 100$  con un 10 % de contaminación.



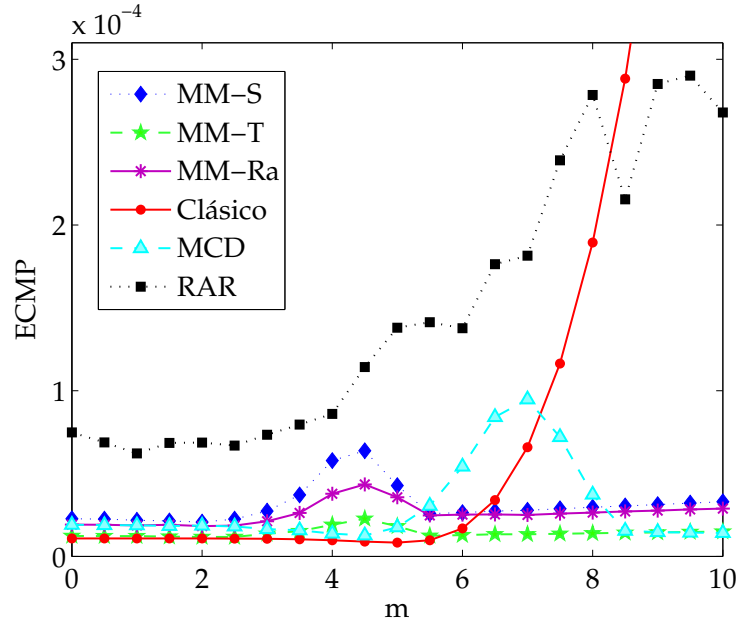


Figura 6.3: Errores cuadráticos medios podados (ECMP) de la correlación canónica  $\hat{c}_2$  para muestras de tamaño  $n = 100$  con un 10 % de contaminación.

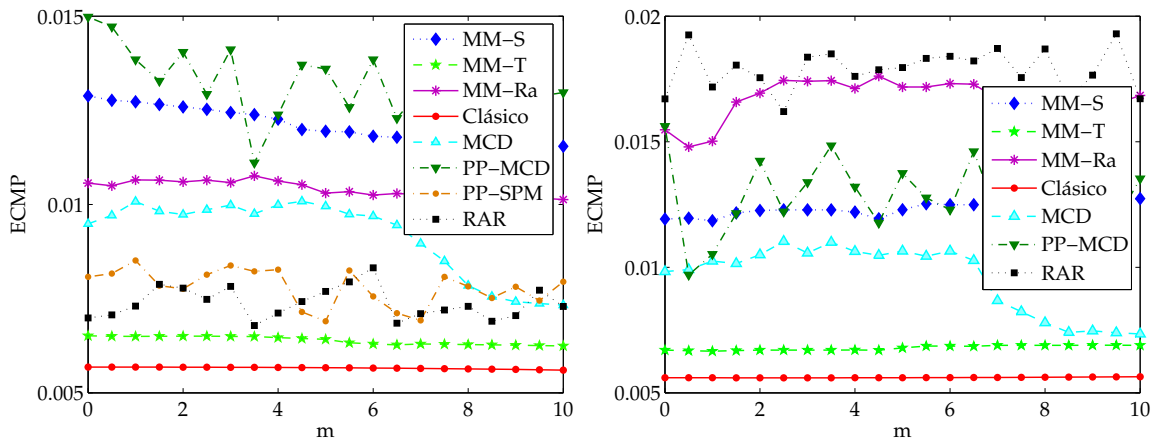


Figura 6.4: Errores cuadráticos medios podados (ECMP) de las correlaciones canónicas  $\hat{c}_3$  (izquierda) y  $\hat{c}_4$  (derecha) para muestras de tamaño  $n = 100$  con un 10 % de contaminación.

# Apéndice A

## Demostraciones

---

Antes de demostrar los teoremas y propiedades enunciados en este trabajo, fijaremos la notación para normas de vectores y matrices que usaremos a lo largo de este apéndice:

Dado  $\mathbf{v} \in \mathbb{R}^m$ , notaremos su *norma 2* o *norma Euclídea* como:

$$\|\mathbf{v}\| = \|\mathbf{v}\|_2 = \left( \sum_{i=1}^m v_i^2 \right)^{1/2},$$

donde  $v_i$  representa al  $i$ -ésimo elemento de  $\mathbf{v}$ .

Dada una matriz  $\mathbf{A} \in \mathbb{R}^{r \times m}$ , su *norma espectral*  $\|\cdot\|_{sp}$ , está definida del siguiente modo:

$$\begin{aligned} \|\mathbf{A}\| &= \|\mathbf{A}\|_{sp} = \max\{\|\mathbf{A}\mathbf{v}\| : \mathbf{v} \in \mathbb{R}^m \text{ con } \|\mathbf{v}\| = 1\} \\ &= \max\left\{ \frac{\|\mathbf{A}\mathbf{v}\|}{\|\mathbf{v}\|} : \mathbf{v} \in \mathbb{R}^m \text{ con } \mathbf{v} \neq 0 \right\}. \end{aligned} \quad (\text{A.1})$$

Su *norma 2* o *norma Frobenius*  $\|\cdot\|_2$  es su norma Euclídea si pensamos a la matriz  $\mathbf{A}$  como un vector de  $\mathbb{R}^{m \cdot r}$ , es decir

$$\|\mathbf{A}\|_2 = [\text{tr}(\mathbf{A}'\mathbf{A})]^{1/2} = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}, \quad (\text{A.2})$$

donde  $a_{ij}$  representa al  $(i, j)$ -ésimo elemento de la matriz  $\mathbf{A}$  y  $\text{tr}(\cdot)$  a la traza.

Dada  $\mathbf{V} \in \mathbb{R}^{m \times m}$  denotaremos a sus autovalores como

$$\lambda_1(\mathbf{V}) \geq \dots \geq \lambda_m(\mathbf{V}),$$

luego si  $\mathbf{V}$  es definida positiva

$$\|\mathbf{V}\| = \max_j \{\lambda_j(\mathbf{V})\} = \lambda_1(\mathbf{V}). \quad (\text{A.3})$$

**Observación 6.** Recordemos además que para dos normas cualesquiera  $\|\cdot\|_a$  y  $\|\cdot\|_b$ , tenemos que

$$\alpha \|\mathbf{A}\|_a \leq \|\mathbf{A}\|_b \leq \beta \|\mathbf{A}\|_a$$

para ciertos  $\alpha$  y  $\beta$  y para todas las matrices  $\mathbf{A} \in \mathbb{R}^{r \times m}$ . En otras palabras, son normas equivalentes, es decir inducen la misma topología en  $\mathbb{R}^{r \times m}$ . Para  $\|\cdot\|_a = \|\cdot\|$  y  $\|\cdot\|_b = \|\cdot\|_2$  se tiene  $\alpha = 1$  y  $\beta = \sqrt{k}$  siendo  $k$  el rango de la matriz  $\mathbf{A}$ .

Recordemos también que la norma espectral y la norma Frobenius son normas matriciales, es decir, para cualquier par de matrices en  $\mathbb{R}^{m \times m}$   $\mathbf{A}$  y  $\mathbf{B}$

$$\|\mathbf{AB}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2 \quad \text{y} \quad \|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|. \quad (\text{A.4})$$

Esta propiedad será utilizada posteriormente en repetidas oportunidades.

Por último enunciaremos el *Teorema de Perturbación de Weyl*, el cuál será utilizado en la siguiente sección.

**Teorema 11.** Sean  $\mathbf{A}$  y  $\mathbf{B}$  matrices simétricas en  $\mathbb{R}^{m \times m}$ . Luego

$$\max_j |\lambda_j(\mathbf{A}) - \lambda_j(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|.$$

La demostración puede encontrarse en [2], pág. 63.

## A.1. Demostración del Teorema 1

Antes de demostrar el Teorema 1 probaremos el siguiente Lema:

**Lema 1.** Sean  $\mathbf{y} \in \mathbb{R}^q$  y  $\mathbf{x} \in \mathbb{R}^p$  vectores fijos. La función

$$d^2(\mathbf{B}, \Sigma) = (\mathbf{y} - \mathbf{B}'\mathbf{x})'\Sigma^{-1}(\mathbf{y} - \mathbf{B}'\mathbf{x})$$

es continua en  $\mathbb{R}^{p \times q} \times \mathcal{S}_q$ .

**Demostración:** Sin pérdida de generalidad, debido a la Observación 6, podemos considerar en  $\mathbb{R}^{p \times q} \times \mathbb{R}^{q \times q}$  la topología inducida por la norma

$$|||(\mathbf{A}, \mathbf{V})||| = \sup\{\|\mathbf{A}\|_2, \|\mathbf{V}\|\}. \quad (\text{A.5})$$

Dados  $(\mathbf{B}_0, \Sigma_0) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q$  y  $\varepsilon > 0$ , tomemos  $(\mathbf{B}, \Sigma) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q$  tal que

$$|||(\mathbf{B}, \Sigma) - (\mathbf{B}_0, \Sigma_0)||| \leq \min\{\delta_1(\mathbf{B}_0, \Sigma_0), \delta_2(\mathbf{B}_0, \Sigma_0), 1\} \quad (\text{A.6})$$

siendo

$$\delta_1(\mathbf{B}_0, \Sigma_0) = \frac{\varepsilon \lambda_q(\Sigma_0)}{\varepsilon + 2q(\|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y} - \mathbf{B}_0'\mathbf{x}\|)}$$

y

$$\delta_2(\mathbf{B}_0, \Sigma_0) = \frac{\varepsilon \lambda_q(\Sigma_0)^2}{2\|\mathbf{y} - \mathbf{B}_0'\mathbf{x}\|^2 + \varepsilon \lambda_q(\Sigma_0)}.$$

Sumando y restando  $d^2(\mathbf{B}_0, \Sigma)$  y usando la desigualdad triangular tenemos que

$$\begin{aligned} |d^2(\mathbf{B}, \Sigma) - d^2(\mathbf{B}_0, \Sigma_0)| &\leq |d^2(\mathbf{B}, \Sigma) - d^2(\mathbf{B}_0, \Sigma)| + |d^2(\mathbf{B}_0, \Sigma) - d^2(\mathbf{B}_0, \Sigma_0)| \\ &:= T1 + T2. \end{aligned}$$

Sea  $r(\mathbf{B}) = (\mathbf{y} - \mathbf{B}'\mathbf{x})$  y  $r(\mathbf{B}_0) = (\mathbf{y} - \mathbf{B}_0'\mathbf{x})$ , consideremos  $\{\alpha_1(\mathbf{B}), \dots, \alpha_q(\mathbf{B})\}$  y  $\{\beta_1(\mathbf{B}_0), \dots, \beta_q(\mathbf{B}_0)\}$  los respectivos coeficientes de  $r(\mathbf{B})$  y  $r(\mathbf{B}_0)$  en una base ortonormal de autovectores de  $\Sigma^{-1}$  con el orden inducido por los autovalores  $\lambda_1(\Sigma^{-1}) \geq \dots \geq \lambda_q(\Sigma^{-1})$ .

Utilizando herramientas básicas de álgebra lineal tenemos que

$$\begin{aligned}
 T1 &= |r(\mathbf{B})'\Sigma^{-1}r(\mathbf{B}) - r(\mathbf{B}_0)\Sigma^{-1}r(\mathbf{B}_0)| \\
 &= \left| \sum_{i=1}^q \alpha_i^2(\mathbf{B})\lambda_i(\Sigma^{-1}) - \sum_{i=1}^q \beta_i^2(\mathbf{B}_0)\lambda_i(\Sigma^{-1}) \right| \\
 &= \left| \sum_{i=1}^q (\alpha_i^2(\mathbf{B}) - \beta_i^2(\mathbf{B}_0)) \lambda_i(\Sigma^{-1}) \right| \\
 &\leq \lambda_1(\Sigma^{-1}) \sum_{i=1}^q |\alpha_i^2(\mathbf{B}) - \beta_i^2(\mathbf{B}_0)|. \tag{A.7}
 \end{aligned}$$

Como

$$\begin{aligned}
 |\alpha_i^2(\mathbf{B}) - \beta_i^2(\mathbf{B}_0)| &= |\alpha_i(\mathbf{B}) - \beta_i(\mathbf{B}_0)| |\alpha_i(\mathbf{B}) + \beta_i(\mathbf{B}_0)| \\
 &\leq \|r(\mathbf{B}) - r(\mathbf{B}_0)\| \|r(\mathbf{B}) + r(\mathbf{B}_0)\|
 \end{aligned}$$

de (A.7), (A.6) y la definición (A.1) obtenemos

$$\begin{aligned}
 T_1 &\leq \lambda_1(\Sigma^{-1})q \|r(\mathbf{B}) - r(\mathbf{B}_0)\| \|r(\mathbf{B}) + r(\mathbf{B}_0)\| \\
 &= q\lambda_1(\Sigma^{-1}) \|(\mathbf{B} - \mathbf{B}_0)\mathbf{x}\| \|r(\mathbf{B}) - r(\mathbf{B}_0) + 2r(\mathbf{B}_0)\| \\
 &\leq q\lambda_1(\Sigma^{-1}) \|\mathbf{B} - \mathbf{B}_0\| \|\mathbf{x}\| (\|r(\mathbf{B}) - r(\mathbf{B}_0)\| + 2\|r(\mathbf{B}_0)\|) \\
 &\leq q\lambda_1(\Sigma^{-1})\delta_1(\mathbf{B}_0, \Sigma_0)\|\mathbf{x}\| (\|\mathbf{B} - \mathbf{B}_0\| \|\mathbf{x}\| + 2\|r(\mathbf{B}_0)\|) \\
 &< q\delta_1(\mathbf{B}_0, \Sigma_0)\|\mathbf{x}\| (\|\mathbf{x}\| + 2\|r(\mathbf{B}_0)\|) \lambda_1(\Sigma^{-1}).
 \end{aligned}$$

Por el Teorema 11 tenemos que

$$|\lambda_q(\Sigma) - \lambda_q(\Sigma_0)| \leq \|\Sigma - \Sigma_0\| < \|(\mathbf{B}, \Sigma) - (\mathbf{B}_0, \Sigma_0)\|. \tag{A.8}$$

Luego por (A.8) y (A.6)

$$\lambda_1(\Sigma^{-1}) = \frac{1}{\lambda_q(\Sigma)} < \frac{1}{\lambda_q(\Sigma_0) - \delta_1(\mathbf{B}_0, \Sigma_0)},$$

y por lo tanto

$$T_1 < \frac{q\delta_1(\mathbf{B}_0, \Sigma_0)\|\mathbf{x}\| (\|\mathbf{x}\| + 2\|\mathbf{y} - \mathbf{B}'_0\mathbf{x}\|)}{\lambda_q(\Sigma_0) - \delta_1(\mathbf{B}_0, \Sigma_0)} = \varepsilon/2. \tag{A.9}$$

Si  $r(\mathbf{B}_0) = \mathbf{0}$  entonces  $T2 = 0 < \varepsilon/2$  y el lema queda demostrado, en caso contrario usando la desigualdad de Cauchy-Schwarz, (A.1) y que la norma espectral es una norma matricial, tenemos que

$$\begin{aligned} T2 &= |r(\mathbf{B}_0)'(\Sigma^{-1} - \Sigma_0^{-1})r(\mathbf{B}_0)| = |r(\mathbf{B}_0)'\Sigma^{-1}(\Sigma_0 - \Sigma)\Sigma_0^{-1}r(\mathbf{B}_0)| \\ &\leq \|\Sigma^{-1}(\Sigma_0 - \Sigma)\Sigma_0^{-1}r(\mathbf{B}_0)\| \|r(\mathbf{B}_0)\| \leq \|\Sigma^{-1}(\Sigma_0 - \Sigma)\Sigma_0^{-1}\| \|r(\mathbf{B}_0)\|^2 \\ &\leq \|\Sigma^{-1}\| \|\Sigma_0 - \Sigma\| \|\Sigma_0^{-1}\| \|r(\mathbf{B}_0)\|^2 \end{aligned}$$

Luego por (A.3), (A.6) y (A.8)

$$T2 < \frac{\delta_2(\mathbf{B}_0, \Sigma_0)}{\lambda_q(\Sigma_0) - \delta_2(\mathbf{B}_0, \Sigma_0)} \lambda_1(\Sigma_0^{-1}) \|\mathbf{y} - \mathbf{B}_0' \mathbf{x}\|^2 = \varepsilon/2$$

y junto con (A.9) y (A.7) queda probado el lema. ■

**Demostración del Teorema:** Por el Lema 1, es suficiente probar que existen  $t_1$  y  $t_2$  tales que

$$\inf_{(\mathbf{B}, \Gamma) \in \mathcal{C}} \sum_{i=1}^n \rho \left( \frac{d_i(\mathbf{B}, \Gamma)}{\hat{\sigma}_n} \right) > \rho \left( \frac{d_i(\tilde{\mathbf{B}}_n, \tilde{\Gamma}_n)}{\hat{\sigma}_n} \right) \quad (\text{A.10})$$

donde

$$\mathcal{C} = \{(\mathbf{B}, \Gamma) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q \text{ con } |\Gamma| = 1 : \lambda_q(\Gamma) \leq t_1 \text{ o } \|\mathbf{B}\| \geq t_2\}. \quad (\text{A.11})$$

Por definición de  $k_n$  tenemos que para todo  $\boldsymbol{\theta} \in \mathbb{R}^{p+q}$

$$\#\{i : |\boldsymbol{\theta}' \mathbf{z}_i| > 0\} / n \geq 1 - (k_n/n).$$

Tomando  $0.5 < \delta < 1 - (k_n/n)$  y usando un argumento de compacidad podemos encontrar un  $\varepsilon > 0$  tal que

$$\inf_{\|\boldsymbol{\theta}\|=1} \#\{i : |\boldsymbol{\theta}' \mathbf{z}_i| > \varepsilon\} / n \geq \delta. \quad (\text{A.12})$$

Sean  $(\mathbf{B}, \Gamma) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q$  tal que  $|\Gamma| = 1$ ,  $\Delta$  la matriz diagonal de los autovalores de  $\Gamma$  ordenados de menor a mayor y  $\mathbf{U}$  la matriz ortonormal de autovectores

de  $\Gamma$  que verifica  $\Gamma = \mathbf{U}\Delta\mathbf{U}'$ . Luego

$$\begin{aligned} d_i^2(\mathbf{B}, \Gamma) &= (\mathbf{y}_i - \mathbf{B}'\mathbf{x}_i)' \Gamma^{-1} (\mathbf{y}_i - \mathbf{B}'\mathbf{x}_i) \\ &= (\mathbf{U}'\mathbf{y}_i - \mathbf{U}'\mathbf{B}'\mathbf{x}_i)' \Delta^{-1} (\mathbf{U}'\mathbf{y}_i - \mathbf{U}'\mathbf{B}'\mathbf{x}_i) \geq \frac{(\mathbf{e}'\mathbf{z}_i)^2}{\lambda_q(\Gamma)}, \end{aligned} \quad (\text{A.13})$$

con  $\mathbf{e} = (-\mathbf{v}_1, \mathbf{v}_1)$  donde  $\mathbf{v}_1$  y  $\mathbf{v}_1$  son la primer fila de  $\mathbf{U}'$  y de  $\mathbf{V} = \mathbf{U}'\mathbf{B}'$ , respectivamente.

Como  $\|\mathbf{e}\| \geq 1$ , por (A.12) tenemos que al menos  $n\delta$  valores de  $d_i(\mathbf{B}, \Gamma)$  son mayores o iguales que  $\varepsilon/\sqrt{\lambda_q(\Gamma)}$ . Por lo tanto

$$\sum_{i=1}^n \rho_1 \left( \frac{d_i(\mathbf{B}, \Gamma)}{\hat{\sigma}_n} \right) \geq n\delta \rho_1 \left( \frac{\varepsilon}{\hat{\sigma}_n \sqrt{\lambda_q(\Gamma)}} \right).$$

Sea  $t$  tal que  $\rho_1 \left( \frac{t}{\hat{\sigma}_n} \right) = \frac{1}{2\delta_1}$ , con  $0.5 < \delta_1 < \delta$ , y  $t_1 = \frac{\varepsilon^2}{t^2}$ , luego  $\lambda_q(\Gamma) \leq t_1$  implica que

$$\sum_{i=1}^n \rho_1 \left( \frac{d_i(\mathbf{B}, \Gamma)}{\hat{\sigma}_n} \right) \geq n\delta \rho_1 \left( \frac{t}{\hat{\sigma}_n} \right) = \frac{n}{2}. \quad (\text{A.14})$$

Si  $\lambda_q(\Gamma) > t_1$  y  $\|\mathbf{B}\|_2 \geq t_2$ , todos los autovalores de  $\Gamma$  son menores que  $1/t_1^{q-1}$  y al menos una columna de  $\mathbf{B}$  tiene norma mayor o igual que  $t_2/\sqrt{q}$ . Por (A.2), se tiene que

$$\|\mathbf{V}\|_2^2 = \|\mathbf{U}'\mathbf{B}'\|_2^2 = \text{tr}(\mathbf{B}\mathbf{U}\mathbf{U}'\mathbf{B}') = \text{tr}(\mathbf{B}\mathbf{B}') = \|\mathbf{B}'\|_2^2 = \|\mathbf{B}\|_2^2,$$

y entonces existe un  $k$  tal que  $\|\mathbf{v}_k\| \geq t_2/\sqrt{q}$  siendo  $\mathbf{v}_k$  la  $k$ -ésima fila de  $\mathbf{V}$ .

Luego procediendo como en (A.13) obtenemos que

$$d_i^2(\mathbf{B}, \Gamma) \geq (\mathbf{e}'_k \mathbf{z}_i)^2 t_1^{q-1},$$

donde  $\mathbf{e}_k = (-\mathbf{v}_k, \mathbf{v}_k)$  y  $\mathbf{v}_k$  es la  $k$ -ésima fila de  $\mathbf{U}'$ .

Por (A.12), al menos  $n\delta$  valores de  $d_i(\mathbf{B}, \Gamma)$  son mayores o iguales que  $\varepsilon\|\mathbf{e}_k\|\sqrt{t_1^{q-1}}$  y  $\|\mathbf{e}_k\|^2 = 1 + \|\mathbf{v}_k\|^2 \geq t_2^2/q$ .

Entonces si tomamos

$$t_2 = \frac{t}{\varepsilon} \sqrt{q/t_1^{q-1}}$$

obtenemos

$$\sum_{i=1}^n \rho_1 \left( \frac{d_i(\mathbf{B}, \mathbf{\Gamma})}{\hat{\sigma}_n} \right) \geq n \delta \rho_1 \left( \frac{t}{\hat{\sigma}_n} \right) = \frac{n}{2}. \quad (\text{A.15})$$

Luego por (A.14) y (A.15)

$$\inf_{(\mathbf{B}, \mathbf{\Gamma}) \in \mathcal{C}} \sum_{i=1}^n \rho_1 \left( \frac{d_i(\mathbf{B}, \mathbf{\Gamma})}{\hat{\sigma}_n} \right) \geq \frac{n}{2}$$

y por (2.4) y (2.2)

$$\sum_{i=1}^n \rho_1 \left( \frac{d_i(\tilde{\mathbf{B}}_n, \tilde{\mathbf{\Gamma}}_n)}{\hat{\sigma}_n} \right) \leq \sum_{i=1}^n \rho_0 \left( \frac{d_i(\tilde{\mathbf{B}}_n, \tilde{\mathbf{\Gamma}}_n)}{\hat{\sigma}_n} \right) = \frac{n}{2},$$

y con esto queda demostrado el Teorema. ■

## A.2. Demostración del Teorema 2

Antes de probar el Teorema daremos algunos resultados con respecto a derivadas de matrices que serán utilizados en varias de las siguientes demostraciones.

Sea  $\mathbf{b}$  un vector y  $\mathbf{V}$  una matriz simétrica,

$$\frac{\partial \mathbf{b}' \mathbf{V} \mathbf{b}}{\partial \mathbf{b}'} = 2 \mathbf{b}' \mathbf{V}, \quad (\text{A.16})$$

si  $\mathbf{V}$  es no singular,

$$\frac{\partial |\mathbf{V}|}{\partial \mathbf{V}} = |\mathbf{V}| \mathbf{V}^{-1} \quad (\text{A.17})$$

y

$$\frac{\partial \mathbf{b}' \mathbf{V}^{-1} \mathbf{b}}{\partial \mathbf{V}} = -\mathbf{V}^{-1} \mathbf{b} \mathbf{b}' \mathbf{V}^{-1}. \quad (\text{A.18})$$

Para más detalles ver Capítulo 17 de [32]. Sea  $(\mathbf{B}, \mathbf{\Sigma}) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q$ , usando  $\text{vec}(\mathbf{B}' \mathbf{A}) = (\mathbf{A} \otimes \mathbf{I}_q) \text{vec}(\mathbf{B}')$ , es fácil verificar que

$$\frac{\partial \text{vec}(\mathbf{B}' \mathbf{A})}{\partial \text{vec}(\mathbf{B}')} = (\mathbf{A}' \otimes \mathbf{I}_q). \quad (\text{A.19})$$



De (A.19) y (A.16) se obtiene

$$\frac{\partial d(\mathbf{B}, \Sigma)}{\partial \text{vec}(\mathbf{B}')'} = - \frac{(\mathbf{y} - \mathbf{B}'\mathbf{x})'\Sigma^{-1}(\mathbf{x}' \otimes \mathbf{I}_q)}{d(\mathbf{B}, \Sigma)}. \quad (\text{A.20})$$

**Demostración del Teorema:** La definición de los MM-estimadores puede reformularse, utilizando la función  $\Gamma(\Sigma) := \Sigma/|\Sigma|^{1/q}$ , de la siguiente manera:  $(\hat{\mathbf{B}}_n, \mathbf{C}_n)$  es un mínimo de  $S^*(\mathbf{B}, \Sigma) = S(\mathbf{B}, \Gamma(\Sigma))$  en  $\mathbb{R}^{p \times q} \times \mathbb{S}_q$ , que satisface

$$S^*(\hat{\mathbf{B}}_n, \mathbf{C}_n) \leq S^*(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n).$$

Finalmente, el MM-estimador  $\hat{\Sigma}_n$  es definido como

$$\hat{\Sigma}_n = \hat{\sigma}_n^2 \Gamma(\mathbf{C}_n). \quad (\text{A.21})$$

Derivando  $S^*(\mathbf{B}, \Sigma)$  con respecto a  $\mathbf{B}$  y  $\Sigma$  y evaluando en el mínimo  $(\hat{\mathbf{B}}_n, \mathbf{C}_n)$  obtenemos

$$\sum_{i=1}^n \frac{\partial \rho_1(d_i(\mathbf{B}, \Gamma(\Sigma))/\hat{\sigma}_n)}{\partial \mathbf{B}} (\hat{\mathbf{B}}_n, \mathbf{C}_n) = \mathbf{0} \quad (\text{A.22})$$

y

$$\sum_{i=1}^n \frac{\partial \rho_1(d_i(\mathbf{B}, \Gamma(\Sigma))/\hat{\sigma}_n)}{\partial \Sigma} (\hat{\mathbf{B}}_n, \mathbf{C}_n) = \mathbf{0}. \quad (\text{A.23})$$

Por (A.19), tenemos que

$$\begin{aligned} \frac{\partial \rho_1(d_i(\mathbf{B}, \Gamma(\Sigma))/\hat{\sigma}_n)}{\partial (\text{vec}(\mathbf{B}'))'} &= \frac{1}{\hat{\sigma}_n} \psi_1 \left( \frac{d_i(\mathbf{B}, \Gamma(\Sigma))}{\hat{\sigma}_n} \right) \frac{\partial d_i(\mathbf{B}, \Gamma(\Sigma))}{\partial (\text{vec}(\mathbf{B}'))'} \\ &= \frac{1}{\hat{\sigma}_n} \psi_1 \left( \frac{d_i(\mathbf{B}, \Gamma(\Sigma))}{\hat{\sigma}_n} \right) \left( \frac{\hat{\mathbf{u}}_i(\mathbf{B})'\Gamma(\Sigma)^{-1}}{d_i(\mathbf{B}, \Gamma(\Sigma))} \right) \frac{\partial \hat{\mathbf{u}}_i(\mathbf{B})}{\partial (\text{vec}(\mathbf{B}'))'} \\ &= -W \left( \frac{d_i(\mathbf{B}, \Gamma(\Sigma))}{\hat{\sigma}_n} \right) \frac{\hat{\mathbf{u}}_i(\mathbf{B})'\Gamma(\Sigma)^{-1}(\mathbf{x}'_i \otimes \mathbf{I}_q)}{\hat{\sigma}_n^2}. \end{aligned}$$

Entonces, por (A.21),

$$\frac{\partial \rho_1(d_i(\mathbf{B}, \Gamma(\Sigma))/\hat{\sigma}_n)}{\partial (\text{vec}(\mathbf{B}'))'} (\hat{\mathbf{B}}_n, \mathbf{C}_n) = -\hat{\sigma}_n W \left( d_i(\hat{\mathbf{B}}_n, \hat{\Sigma}_n) \right) \hat{\mathbf{u}}_i(\hat{\mathbf{B}}_n)' \hat{\Sigma}_n^{-1} (\mathbf{x}'_i \otimes \mathbf{I}_q). \quad (\text{A.24})$$

Usando que

$$\text{vec}(\widehat{\Sigma}_n^{-1} \widehat{\mathbf{u}}_i(\widehat{\mathbf{B}}_n) \mathbf{x}'_i)' = \widehat{\mathbf{u}}_i(\widehat{\mathbf{B}}_n)' \widehat{\Sigma}_n^{-1} (\mathbf{x}'_i \otimes \mathbf{I}_q),$$

(A.24) y (A.22), podemos ver que (2.10) es válida.

Derivando  $\rho_1(d_i(\mathbf{B}, \Gamma)/\hat{\sigma}_n)$  con respecto a  $\Sigma$ , obtenemos

$$\frac{\partial \rho_1(d_i(\mathbf{B}, \Gamma(\Sigma))/\hat{\sigma}_n)}{\partial \Sigma} = \frac{1}{\hat{\sigma}_n} \psi_1 \left( \frac{d_i(\mathbf{B}, \Gamma(\Sigma))}{\hat{\sigma}_n} \right) \frac{\partial (d_i(\mathbf{B}, \Sigma) |\Sigma|^{1/(2q)})}{\partial \Sigma}. \quad (\text{A.25})$$

De (A.17) y (A.18) obtenemos

$$\begin{aligned} \frac{\partial (d_i(\mathbf{B}, \Sigma) |\Sigma|^{1/(2q)})}{\partial \Sigma} &= \frac{\partial |\Sigma|^{1/(2q)}}{\partial \Sigma} d_i(\mathbf{B}, \Sigma) + \frac{\partial d_i(\mathbf{B}, \Sigma)}{\partial \Sigma} |\Sigma|^{1/(2q)} \\ &= \frac{|\Sigma|^{1/(2q)} \Sigma^{-1}}{2q} d_i(\mathbf{B}, \Sigma) - \frac{\Sigma^{-1} \widehat{\mathbf{u}}_i(\mathbf{B}) \widehat{\mathbf{u}}_i(\mathbf{B})' \Sigma^{-1}}{2d_i(\mathbf{B}, \Sigma)} |\Sigma|^{1/(2q)} \\ &= \frac{\Sigma^{-1}}{2q} \left( d_i(\mathbf{B}, \Gamma(\Sigma)) - q \frac{\widehat{\mathbf{u}}_i(\mathbf{B}) \widehat{\mathbf{u}}_i(\mathbf{B})' \Gamma(\Sigma)^{-1}}{d_i(\mathbf{B}, \Gamma(\Sigma))} \right). \end{aligned} \quad (\text{A.26})$$

Luego, por (A.25) y (A.26), la ecuación (A.23) resulta equivalente a

$$\sum_{i=1}^n \frac{1}{\hat{\sigma}_n} \psi_1 \left( \frac{d_i(\widehat{\mathbf{B}}_n, \Gamma(\mathbf{C}_n))}{\hat{\sigma}_n} \right) \left( d_i(\widehat{\mathbf{B}}_n, \Gamma(\mathbf{C}_n)) - q \frac{\widehat{\mathbf{u}}_i(\widehat{\mathbf{B}}_n) \widehat{\mathbf{u}}_i(\widehat{\mathbf{B}}_n)' \Gamma(\mathbf{C}_n)^{-1}}{d_i(\widehat{\mathbf{B}}_n, \Gamma(\mathbf{C}_n))} \right) = 0.$$

Reordenando y utilizando que  $W(u) = \psi(u)/u$  y que  $\widehat{\Sigma}_n = \hat{\sigma}_n^2 \Gamma(\mathbf{C}_n)$  tenemos que

$$\widehat{\Sigma}_n \left( \sum_{i=1}^n \psi_1 \left( d_i(\widehat{\mathbf{B}}_n, \widehat{\Sigma}_n) \right) d_i(\widehat{\mathbf{B}}_n, \widehat{\Sigma}_n) \right) - q \sum_{i=1}^n W \left( d_i(\widehat{\mathbf{B}}_n, \widehat{\Sigma}_n) \right) \widehat{\mathbf{u}}_i(\widehat{\mathbf{B}}_n) \widehat{\mathbf{u}}_i(\widehat{\mathbf{B}}_n)' = 0$$

y despejando  $\widehat{\Sigma}_n$  obtenemos (2.11). ■

### A.3. Demostración del Teorema 3

Antes de demostrar el Teorema 3 probaremos el siguiente Lema:

**Lema 2.** Sea  $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ , con  $\mathbf{z}_i = (\mathbf{y}'_i, \mathbf{x}'_i)'$  que satisface el MLM (1.1) y consideremos  $\rho_0$  una función- $\rho$ . Luego el punto de ruptura de explosión del M-estimador

de escala de las normas de Mahalanobis (2.3),  $\hat{\sigma}_n := s(\mathbf{d}(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n))$ , está acotado inferiormente por

$$\min(\varepsilon_n^*(\mathbf{Z}, \tilde{\mathbf{B}}_n, \tilde{\Sigma}_n), 0.5).$$

**Demostración:** Sean

$$m < \min(n\varepsilon_n^*(\mathbf{Z}, \tilde{\mathbf{B}}_n, \tilde{\Sigma}_n), n/2)$$

y

$$(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*) = (\tilde{\mathbf{B}}_n^*(\mathbf{Z}^*), \tilde{\Sigma}_n^*(\mathbf{Z}^*))$$

un estimador inicial de  $(\mathbf{B}_0, \Sigma_0)$  calculado con la muestra  $\mathbf{Z}^* \in \mathcal{Z}_m$ . Para probar el Lema basta con demostrar que  $s(\mathbf{d}(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*))$  está acotado para todo  $\mathbf{Z}^* \in \mathcal{Z}_m$ .

Como  $m < n\varepsilon_n^*(\mathbf{Z}, \tilde{\mathbf{B}}_n, \tilde{\Sigma}_n)$ , existe un conjunto compacto  $K$  tal que

$$(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*) \in K \quad \text{para todo } \mathbf{Z}^* \in \mathcal{Z}_m.$$

Luego, por el Lema 1, existe un  $t$  tal que

$$\sup_{\{i: \mathbf{z}_i = \mathbf{z}_i^*\}} d_i(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*) \leq t \quad \text{para todo } \mathbf{Z}^* \in \mathcal{Z}_m. \quad (\text{A.27})$$

Como  $m/n < 0.5$  podemos encontrar un  $\gamma > 0$  tal que  $m/n + \gamma < 0.5$ . Sea  $\delta$  el valor que verifica  $\rho_0(\delta) = \gamma$  y sea  $t_0 = t/\delta$ . Luego usando (A.27) tenemos que

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{d_i(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*)}{t_0} \right) &= \frac{1}{n} \sum_{\{i: \mathbf{z}_i = \mathbf{z}_i^*\}} \rho_0 \left( \frac{d_i(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*)}{t_0} \right) \\ &\quad + \frac{1}{n} \sum_{\{i: \mathbf{z}_i \neq \mathbf{z}_i^*\}} \rho_0 \left( \frac{d_i(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*)}{t_0} \right) \\ &\leq \frac{(n-m)}{n} \rho_0(t_2/t_0) + \frac{m}{n} \leq \rho_0(\delta) + \frac{m}{n} \\ &= \gamma + \frac{m}{n} < 0.5. \end{aligned}$$

Entonces  $s(\mathbf{d}(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*)) \leq t_0$  para todo  $\mathbf{Z}^* \in \mathcal{Z}_m$  y el lema queda demostrado. ■

**Demostración del Teorema:** Sea  $\varepsilon_n^*(\mathbf{Z}, \tilde{\mathbf{B}}_n, \tilde{\Sigma}_n)$  el punto de ruptura del estima-

dor inicial  $(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n)$  y sea

$$m < \min(n\varepsilon_n^*(\mathbf{Z}, \tilde{\mathbf{B}}_n, \tilde{\Sigma}_n), [n/2] - k_n).$$

Sean respectivamente

$$(\hat{\mathbf{B}}_n^*, \hat{\Sigma}_n^*) = (\hat{\mathbf{B}}_n^*(\mathbf{Z}^*), \hat{\Sigma}_n^*(\mathbf{Z}^*)) \quad \text{y} \quad (\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*) = (\tilde{\mathbf{B}}_n^*(\mathbf{Z}^*), \tilde{\Sigma}_n^*(\mathbf{Z}^*))$$

un MM-estimador para el modelo lineal multivariado y su estimador inicial calculados con la muestra  $\mathbf{Z}^* \in \mathcal{Z}_m$ .

Luego por (2.6), (2.4) y (2.2)

$$\begin{aligned} \frac{1}{n} \sum_{\mathbf{z}_i \in \mathbf{Z}^*} \rho_1 \left( d_i(\hat{\mathbf{B}}_n^*, \hat{\Sigma}_n^*) \right) &\leq \frac{1}{n} \sum_{\mathbf{z}_i \in \mathbf{Z}^*} \rho_1 \left( \frac{d_i(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*)}{\hat{\sigma}_n^*} \right) \leq \frac{1}{n} \sum_{\mathbf{z}_i \in \mathbf{Z}^*} \rho_0 \left( \frac{d_i(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*)}{\hat{\sigma}_n^*} \right) \\ &= 0.5. \end{aligned}$$

Además, como  $\sup_u \rho_1(u) = 1$ , obtenemos que

$$\sum_{\mathbf{z}_i \in \mathbf{Z}^*} \rho_1 \left( d_i(\hat{\mathbf{B}}_n^*, \hat{\Sigma}_n^*) \right) \leq \frac{n}{2} \rho_1(\infty).$$

Entonces existe  $c < \infty$ , que no depende de  $\mathbf{Z}^*$ , tal que  $d_i^2(\hat{\mathbf{B}}_n^*, \hat{\Sigma}_n^*) < c$  para al menos  $[n/2]$  observaciones de  $\mathbf{Z}^*$ .

Ahora, como  $m < [n/2] - k_n$ , al menos  $k_n + 1$  de estas observaciones están en  $\mathbf{Z}$ , y no están en un hiperplano. Entonces el menor autovalor de  $\hat{\Sigma}_n^*$ ,  $\lambda_q(\hat{\Sigma}_n^*)$ , está acotado inferiormente con una cota positiva (para cada  $\mathbf{x}_i \in \mathbb{R}^p$ , los ejes del elipsoide

$$\{\mathbf{y} : (\mathbf{y} - \hat{\mathbf{B}}_n^{*'} \mathbf{x}_i)' \hat{\Sigma}_n^{*-1} (\mathbf{y} - \hat{\mathbf{B}}_n^{*'} \mathbf{x}_i) \leq c\}$$

tienen longitud  $\sqrt{c\lambda_j(\hat{\Sigma}_n^*)}$ ;  $j = 1, \dots, q$ . Entonces  $\lambda_q(\hat{\Sigma}_n^*) > \alpha$ , siendo  $\alpha$  un valor positivo que no depende de  $\mathbf{Z}^*$ ).

Además, como  $|\hat{\Sigma}_n^*| = (\hat{\sigma}_n^*)^{2q} = s(\mathbf{d}(\tilde{\mathbf{B}}_n^*, \tilde{\Sigma}_n^*))^{2q}$ , por el Lema 2 el mayor autovalor de  $\hat{\Sigma}_n^*$  está acotado superiormente.

Para ver que  $\|\widehat{\mathbf{B}}_n^*\|$  está acotada consideremos el conjunto

$$\mathcal{C}(\widehat{\mathbf{B}}_n^*, \widehat{\Sigma}_n^*) = \{(\mathbf{v}, \mathbf{w}) : (\mathbf{w} - \widehat{\mathbf{B}}_n^{*'} \mathbf{v})' \widehat{\Sigma}_n^{*-1} (\mathbf{w} - \widehat{\mathbf{B}}_n^{*'} \mathbf{v}) \leq c\}$$

que, como vimos, contiene  $k_n + 1$  observaciones de  $\mathbf{Z}$  que no están en un hiperplano.

Como para matrices simétricas  $\mathbf{A}$  de dimensión  $q \times q$ , tenemos que

$$\lambda_q(\mathbf{A}) = \inf_{\mathbf{v}} \frac{\mathbf{v}' \mathbf{A} \mathbf{v}}{\mathbf{v}' \mathbf{v}}$$

y  $\lambda_q(\mathbf{A}^{-1}) = 1/\lambda_1(\mathbf{A})$  entonces

$$\|\mathbf{w} - \widehat{\mathbf{B}}_n^{*'} \mathbf{v}\|^2 \leq (\mathbf{w} - \widehat{\mathbf{B}}_n^{*'} \mathbf{v})' \widehat{\Sigma}_n^{*-1} (\mathbf{w} - \widehat{\mathbf{B}}_n^{*'} \mathbf{v}) \lambda_1(\widehat{\Sigma}_n^*) \leq \lambda_1(\widehat{\Sigma}_n^*) c,$$

en particular para  $\mathbf{w} = \mathbf{0}$

$$\|\widehat{\mathbf{B}}_n^{*'} \mathbf{v}\|^2 \leq \lambda_1(\widehat{\Sigma}_n^*) c.$$

Como  $\mathcal{C}(\widehat{\mathbf{B}}_n^*, \widehat{\Sigma}_n^*)$  contiene  $k_n + 1$  puntos que no están en un hiperplano, existe una constante  $g$ , que no depende de  $\widehat{\mathbf{B}}_n^*$  o  $\widehat{\Sigma}_n^*$ , tal que para todo  $\mathbf{v}$  con  $\|\mathbf{v}\| \leq g$  se tiene que  $(\mathbf{v}, \mathbf{0}) \in \mathcal{C}(\widehat{\mathbf{B}}_n^*, \widehat{\Sigma}_n^*)$ . Entonces

$$\sup_{\|\mathbf{v}\|=g} \|\widehat{\mathbf{B}}_n^{*'} \mathbf{v}\|^2 \leq \lambda_1(\widehat{\Sigma}_n^*) c,$$

y esto implica que

$$\|\widehat{\mathbf{B}}_n^{*'}\|^2 \leq \lambda_1(\widehat{\Sigma}_n^*) c / g^2$$

donde  $\|\cdot\|$  es la norma espectral definida en (A.1). Luego, como  $\|\cdot\|_2$  y  $\|\cdot\|$  son equivalentes, existe una constante  $\beta > 0$  tal que

$$\|\widehat{\mathbf{B}}_n^*\|_2 = \|\widehat{\mathbf{B}}_n^{*'}\|_2 \leq \beta \|\widehat{\mathbf{B}}_n^{*'}\| \leq (\beta/g) \sqrt{\lambda_1(\widehat{\Sigma}_n^*) c}$$

para todo  $\mathbf{Z}^* \in \mathcal{Z}_m$ . Esto prueba el Teorema. ■

## A.4. Demostración Teorema 4

**Lema 3.** Supongamos que observamos  $\mathbf{z} \in \mathbb{R}^k$  con función de distribución  $H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}$ , donde  $\boldsymbol{\theta}_1 \in \mathbb{R}^{m_1}$  y  $\boldsymbol{\theta}_2 \in \mathbb{R}^{m_2}$ . Consideremos un M-estimador funcional Fisher consistente de  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ ,  $\mathbf{T}(H) = (\mathbf{T}_1(H), \mathbf{T}_2(H))$  y un estimador funcional inicial de  $\mathbf{T}(H)$ ,  $\mathbf{T}_0(H) = (\mathbf{T}_{0,1}(H), \mathbf{T}_{0,2}(H))$ , tal que

$$E_H(h(\mathbf{z}, \mathbf{T}_1(H), \mathbf{T}_2(H), S(\mathbf{T}_0(H), H))) = 0,$$

donde  $h : \mathbb{R}^{k+m_1+m_2+1} \rightarrow \mathbb{R}^{m_1}$  es una función diferenciable y  $S : \mathbb{R}^{m_1+m_2} \times \mathcal{Q} \rightarrow \mathbb{R}$ , con  $\mathcal{Q}$  el espacio de las distribuciones sobre  $\mathbb{R}^{m_1+m_2}$ . Supongamos que  $\mathbf{T}$  satisface la siguiente condición fuerte de consistencia Fisher:

$$E_{H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}}(h(\mathbf{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, S)) = 0 \text{ para todo } S, \quad (\text{A.28})$$

y

$$E_{H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}}(h_3(\mathbf{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, S)) = 0 \text{ para todo } S, \quad (\text{A.29})$$

donde  $h_i$ ,  $1 \leq i \leq 4$ , es la derivada de  $h$  con respecto al  $i$ -ésimo argumento. Supongamos también que las derivadas parciales de  $E_{H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}}(h_3(\mathbf{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, S(\boldsymbol{\theta}, H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2})))$  pueden obtenerse derivando con respecto a cada parámetro dentro de la esperanza. Luego la función de influencia de  $\mathbf{T}_1$  está dada por

$$\begin{aligned} FI(\mathbf{z}_0, \mathbf{T}_1, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= - \left( E_{H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}}(h_2(\mathbf{z}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, S(\boldsymbol{\theta}, H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}))) \right)^{-1} \\ &\quad \times (h(\mathbf{z}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, S(\boldsymbol{\theta}, H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}))). \end{aligned}$$

**Demostración:** Consideremos  $H_\varepsilon = (1 - \varepsilon)H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2} + \varepsilon\delta_{\mathbf{z}_0}$ . Luego  $T(H_\varepsilon)$  satisface

$$(1 - \varepsilon)E_{H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}}(h(\mathbf{z}, \mathbf{T}_1(H_\varepsilon), \mathbf{T}_2(H_\varepsilon), S(\mathbf{T}_0(H_\varepsilon), H_\varepsilon)))$$

$$+ h(\mathbf{z}_0, \mathbf{T}_1(H_\varepsilon), \mathbf{T}_2(H_\varepsilon), S(\mathbf{T}_0(H_\varepsilon), H_\varepsilon)) = 0.$$

La prueba del Lema es inmediata derivando esta expresión con respecto a  $\varepsilon$  en  $\varepsilon = 0$  y usando (A.28) y (A.29). ■

El siguiente lema prueba para el caso  $\Sigma_0 = \mathbf{I}$  que los MM-estimadores funcionales  $\mathbf{T}_1$  y  $\mathbf{T}_2/S^2$  son Fisher consistentes a  $\mathbf{B}_0$  y  $\mathbf{I}$ , respectivamente.

**Lema 4.** Sea  $\mathbf{z} = (\mathbf{y}', \mathbf{x}')'$  un vector aleatorio que satisface el modelo lineal multivariado (1.1) por parámetros  $\mathbf{B}_0$  y  $\Sigma_0 = \mathbf{I}$ . Supongamos que  $\mathbf{x}$  satisface (S4) y que la distribución de  $\mathbf{u} = \mathbf{y} - \mathbf{B}_0\mathbf{x}$  satisface (S1). Sean  $\rho_1$  una función- $\rho$  que satisface (S3) y  $\Sigma \neq \mathbf{I}$  tal que  $|\Sigma| = 1$ . Luego

$$E_{H_0} \rho_1 \left( \frac{((\mathbf{y} - \mathbf{B}'\mathbf{x})'\Sigma^{-1}(\mathbf{y} - \mathbf{B}'\mathbf{x}))^{1/2}}{\sigma_0} \right) > E_{H_0} \rho_1 \left( \frac{((\mathbf{y} - \mathbf{B}'_0\mathbf{x})'(\mathbf{y} - \mathbf{B}'_0\mathbf{x}))^{1/2}}{\sigma_0} \right).$$

Este lema es una consecuencia directa del Lema A.10 de [13].

**Lema 5.** Consideremos las mismas suposiciones del Teorema 4 y supongamos que  $\Sigma_0 = \mathbf{I}$ . Luego, si  $H_0$  es la función de distribución de  $(\mathbf{y}', \mathbf{x}')'$ , tenemos que

$$(i) \quad E_{H_0} \left( \frac{\partial W(d(\mathbf{B}_0, \Sigma_0)/S) \text{vec}(\hat{\mathbf{u}}(\mathbf{B}_0)\mathbf{x}')}{\partial \text{vec}(\Sigma)'} \right) = 0 \text{ para todo } S,$$

$$(ii) \quad E_{H_0} (W(d(\mathbf{B}_0, \Sigma_0)/S) \text{vec}(\hat{\mathbf{u}}(\mathbf{B}_0)\mathbf{x}')) = 0 \text{ para todo } S.$$

**Demostración:** (i) Por (A.18) tenemos que

$$E_{H_0} \left( \frac{\partial W(d(\mathbf{B}_0, \Sigma_0)/S) \text{vec}(\hat{\mathbf{u}}(\mathbf{B}_0)\mathbf{x}')}{\partial \text{vec}(\Sigma)'} \right) = -E_{H_0} \left( \frac{W'(\|\mathbf{u}\|/S)}{2S\|\mathbf{u}\|} (\mathbf{x}\mathbf{u}' \otimes \mathbf{I})(\mathbf{I} \otimes \mathbf{u}\mathbf{u}') \right)$$

Como la distribución de  $\mathbf{u}$  se supuso elíptica con  $\Sigma_0 = \mathbf{I}$ , para cualquier función  $h$ ,  $E_{H_0}(h(\|\mathbf{u}\|)u_i u_j u_l x_k) = 0$ . Luego, como todos los elementos del lado derecho de la ecuación anterior tienen esta forma, la parte (i) de lema queda demostrada. (ii) es consecuencia de que  $E_{H_0}(h(\|\mathbf{u}\|)u_i x_j) = 0$  para todo  $i$  y  $j$ . ■

**Demostración del Teorema:** Supongamos que  $\mathbf{z} = (\mathbf{y}', \mathbf{x}')'$  satisface el modelo lineal (1.1) y sea  $H_0$  su función de distribución. Consideremos primero el caso en que  $\Sigma_0 = \mathbf{I}$  y denotemos  $\sigma_0 = S(H_0)$ . Usando el Lema 3 con  $\boldsymbol{\theta}_1 = \text{vec}(\mathbf{B}'_0)$ ,  $\boldsymbol{\theta}_2 = \text{vec}(\Sigma_0)$ ,  $S(\mathbf{T}_0(H), H) = S(H)$  y los lemas 4 y 5 obtenemos que

$$\begin{aligned} IF(\mathbf{z}_0, \mathbf{T}_1, (\mathbf{B}_0, \mathbf{I})) &= - \left( \frac{\partial E_{H_0} W(d(\mathbf{B}_0, \Sigma_0)/\sigma_0) \text{vec}((\mathbf{y} - \mathbf{B}'_0\mathbf{x})\mathbf{x}')}{\partial \text{vec}(\mathbf{B}')'} \right)^{-1} \\ &\quad \times W(\|\mathbf{y}_0 - \mathbf{B}'_0\mathbf{x}_0\|/\sigma_0) \text{vec}((\mathbf{y}_0 - \mathbf{B}'_0\mathbf{x}_0)\mathbf{x}'_0). \quad (\text{A.30}) \end{aligned}$$

Por (A.19) y (A.20) y la igualdad  $\text{vec}(\mathbf{u}\mathbf{x}') = (\mathbf{x} \otimes \mathbf{I}_q)\mathbf{u}$  tenemos que

$$\begin{aligned}
& \frac{\partial W(d(\mathbf{B}_0, \Sigma_0)/\sigma_0) \text{vec}((\mathbf{y} - \mathbf{B}'_0\mathbf{x})\mathbf{x}')}{\partial \text{vec}(\mathbf{B}')'} \\
&= -\frac{W'(d(\mathbf{B}_0, \sigma_0^2\mathbf{I}_q))}{d(\mathbf{B}_0, \sigma_0^2\mathbf{I}_q)\sigma_0^2} \text{vec}((\mathbf{y} - \mathbf{B}'_0\mathbf{x})\mathbf{x}')(\mathbf{y} - \mathbf{B}'_0\mathbf{x})'(\mathbf{x}' \otimes \mathbf{I}_q) \\
&\quad - W(d(\mathbf{B}_0, \sigma_0^2\mathbf{I}_q))(\mathbf{x}\mathbf{x}' \otimes \mathbf{I}_q) \\
&= -W'\left(\frac{\|\mathbf{u}\|}{\sigma_0}\right) \frac{(\mathbf{x} \otimes \mathbf{I}_q)\mathbf{u}\mathbf{u}'(\mathbf{x}' \otimes \mathbf{I}_q)}{\sigma_0\|\mathbf{u}\|} - W\left(\frac{\|\mathbf{u}\|}{\sigma_0}\right)(\mathbf{x}\mathbf{x}' \otimes \mathbf{I}_q).
\end{aligned}$$

Como la distribución de  $\mathbf{u}$  se supuso elíptica con  $\Sigma_0 = \mathbf{I}$ , para cualquier función  $h$ ,  $E_{F_0}(h(\|\mathbf{u}\|)u_i u_j) = 0$  si  $i \neq j$  y  $E_{F_0}(h(\|\mathbf{u}\|)u_i^2) = E_{F_0}(h(\|\mathbf{u}\|)\|\mathbf{u}\|^2)/q$ . Luego

$$\begin{aligned}
& \frac{\partial E_{H_0} W(d(\mathbf{B}_0, \Sigma_0)/\sigma_0) \text{vec}((\mathbf{y} - \mathbf{B}'_0\mathbf{x})\mathbf{x}')}{\partial \text{vec}(\mathbf{B}')'} \\
&= -E_{F_0} W'\left(\frac{\|\mathbf{u}\|}{\sigma_0}\right) \|\mathbf{u}\| \frac{(E_{G_0}\mathbf{x}\mathbf{x}' \otimes \mathbf{I})}{\sigma_0} - E_{F_0} W\left(\frac{\|\mathbf{u}\|}{\sigma_0}\right) (E_{G_0}\mathbf{x}\mathbf{x}' \otimes \mathbf{I}) \\
&= -\left[ \frac{E_{F_0} W'(\|\mathbf{u}\|/\sigma_0) \|\mathbf{u}\|}{\sigma_0} + E_{F_0} W\left(\frac{\|\mathbf{u}\|}{\sigma_0}\right) \right] (E_{G_0}\mathbf{x}\mathbf{x}' \otimes \mathbf{I}). \quad (\text{A.31})
\end{aligned}$$

Combinando (A.30) y (A.31) y usando que  $\text{vec}(\mathbf{C}'\mathbf{A}) = (\mathbf{A} \otimes \mathbf{I})\text{vec}(\mathbf{C}')$ , probamos el teorema para el caso  $\Sigma_0 = \mathbf{I}$ .

Para el caso general, sea  $\mathbf{R}$  una matriz tal que  $\Sigma_0 = \mathbf{R}\mathbf{R}'$  y consideremos la siguiente transformación  $\mathbf{y}^* = \mathbf{R}^{-1}\mathbf{y}$ . Luego  $\mathbf{y}^* = \mathbf{B}_0^{*'}\mathbf{x} + \mathbf{u}^*$ , con  $\mathbf{u}^* = \mathbf{R}^{-1}\mathbf{u}$  y  $\mathbf{B}_0^* = \mathbf{B}_0\mathbf{R}'^{-1}$ . Como la distribución de  $\mathbf{u}^*$  está dada por la densidad (2.15) con  $\Sigma_0 = \mathbf{I}$  y

$$\text{vec}(\mathbf{B}'_0) = \text{vec}(\mathbf{R}\mathbf{B}_0^{*'}),$$

tenemos que

$$IF(\mathbf{z}_0, \mathbf{T}_1, \mathbf{B}_0, \Sigma_0) = \mathbf{R}IF((\mathbf{R}^{-1}\mathbf{y}_0, \mathbf{x}_0), \mathbf{T}_1, \mathbf{B}_0^*, \mathbf{I}).$$

■



## A.5. Demostración del Teorema 5

Antes de probar la consistencia de los MM-estimadores de regresión para el modelo lineal multivariado necesitamos probar algunos lemas. Por simplicidad asumiremos que el estimador inicial  $\tilde{\mathbf{B}}_n$  es regresión- y afín-equivariante y que  $\tilde{\Sigma}_n$  es afín-equivariante y regresión-invariante. Luego, sin pérdida de generalidad, debido a la Observación 3, de ahora en adelante se puede suponer que  $\mathbf{B}_0 = \mathbf{0}$  y  $\Sigma_0 = \mathbf{I}$ . Estas suposiciones no son esenciales para las demostraciones de esta sección.

**Lema 6.** Sean  $(\mathbf{y}'_i, \mathbf{x}'_i)$ ,  $1 \leq i \leq n$ , una muestra aleatoria que satisface el MLM (1.1) con parámetros  $\mathbf{B}_0$  y  $\Sigma_0$ . Sean  $\Gamma_0 = \Sigma_0/|\Sigma_0|^{1/q}$  y  $\rho_0$  una función- $\rho$ . Supongamos que los estimadores iniciales  $\tilde{\mathbf{B}}_n$  y  $\tilde{\Sigma}_n$  son consistentes a  $\mathbf{B}_0$  y  $\Gamma_0$  respectivamente; luego  $\hat{\sigma}_n$  es consistente a  $\sigma_0$  definida por la ecuación (3.1).

**Demostración:** Tomemos  $\varepsilon > 0$ , luego por el Lema (1), podemos encontrar un  $\delta > 0$  tal que

$$E_{H_0} \left( \inf_{\mathcal{E}} \rho_0 \left( ((\mathbf{y} - \mathbf{B}'\mathbf{x})' \Sigma^{-1} (\mathbf{y} - \mathbf{B}'\mathbf{x}))^{1/2} / (\sigma_0 - \varepsilon) \right) \right) \geq b + \delta$$

y

$$E_{H_0} \left( \inf_{\mathcal{E}} \rho_0 \left( ((\mathbf{y} - \mathbf{B}'\mathbf{x})' \Sigma^{-1} (\mathbf{y} - \mathbf{B}'\mathbf{x}))^{1/2} / (\sigma_0 + \varepsilon) \right) \right) \leq b - \delta$$

donde  $\mathcal{E} = \{(\mathbf{B}, \Sigma) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q : \|\mathbf{B}\| \leq \delta, \|\Sigma - \mathbf{I}\| \leq \delta\}$ . Por la ley de grandes números tenemos que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \inf_{\mathcal{E}} \rho_0 \left( \frac{d_i(\mathbf{B}, \Sigma)}{(\sigma_0 - \varepsilon)} \right) \geq b + \delta \quad \text{c.s.}$$

y

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \inf_{\mathcal{E}} \rho_0 \left( \frac{d_i(\mathbf{B}, \Sigma)}{(\sigma_0 + \varepsilon)} \right) \leq b - \delta \quad \text{c.s..}$$

Luego, como  $\lim_{n \rightarrow \infty} (\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n) = (\mathbf{0}, \mathbf{I})$  c.s., tenemos que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{d_i(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n)}{(\sigma_0 - \varepsilon)} \right) \geq b + \delta \quad \text{c.s.}$$

y

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{d_i(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n)}{(\sigma_0 + \varepsilon)} \right) \leq b - \delta \quad \text{c.s..}$$

Por lo tanto debido a la monotonicidad de  $\rho_0(u)$  para los  $u > 0$ , existe  $n_0$  tal que para todo  $n \geq n_0$  tenemos que  $\sigma_0 - \varepsilon \leq \hat{\sigma}_n \leq \sigma_0 + \varepsilon$  con probabilidad 1, es decir  $\lim_{n \rightarrow \infty} \hat{\sigma}_n = \sigma_0$  c.s.. ■

**Lema 7.** Supongamos que la distribución de  $\mathbf{y}$  satisface (S1) con  $\Sigma_0 = \mathbf{I}$  y que  $\mathbf{y}$  es independiente de  $\mathbf{x}$ . Dados  $(\mathbf{B}, \Sigma) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q$  y  $\kappa$  la constante que aparece en (S3), consideremos

$$\alpha(\mathbf{B}, \Sigma; \kappa) = E_{H_0} I \left( (\mathbf{y} - \mathbf{B}'\mathbf{x})' \Sigma^{-1} (\mathbf{y} - \mathbf{B}'\mathbf{x}) \leq \kappa \right), \quad (\text{A.32})$$

donde  $H_0$  es la distribución de  $(\mathbf{y}', \mathbf{x}')$ . Luego existe una constante  $\kappa_1$  independiente de  $\mathbf{B}$  y  $\Sigma$  tal que

$$\alpha(\mathbf{B}, \Sigma; \kappa) \leq \kappa_1 \lambda_j(\Sigma)^{1/2} \text{ para todo } j, 1 \leq j \leq q.$$

**Demostración:** Notemos que  $\mathbf{V}'\Sigma\mathbf{V} = \Lambda$  donde  $\mathbf{V}$  es una matriz ortogonal de  $q \times q$  y  $\Lambda$  es una matriz diagonal que tiene como elementos no nulos a los autovalores de  $\Sigma$ . Utilizando el cambio de variables  $\mathbf{y} \rightarrow \mathbf{V}'\mathbf{y}$  y (S1) obtenemos que para cada  $j = 1, \dots, q$

$$\begin{aligned} E \left( I \left( (\mathbf{y} - \mathbf{B}'\mathbf{x})' \Sigma^{-1} (\mathbf{y} - \mathbf{B}'\mathbf{x}) \leq \kappa \right) \mid \mathbf{x} = \boldsymbol{\beta} \right) \\ &= \int_{(\mathbf{y} - (\mathbf{B}\mathbf{V})'\boldsymbol{\beta})' \Lambda^{-1} (\mathbf{y} - (\mathbf{B}\mathbf{V})'\boldsymbol{\beta}) \leq \kappa} f_0^*(\mathbf{y}'\mathbf{y}) d\mathbf{y} \\ &\leq \int_{|y_j - ((\mathbf{B}\mathbf{V})'\boldsymbol{\beta})_j| \leq \sqrt{\lambda_j(\Sigma)\kappa}} f_0^* \left( y_j^2 + \sum_{i \neq j} y_i^2 \right) dy_1 \dots dy_q \\ &\leq 2\sqrt{\lambda_j(\Sigma)\kappa} \int f_0^* \left( \sum_{i=1}^{q-1} y_i^2 \right) dy_1 \dots dy_{q-1}. \end{aligned}$$

Luego si elegimos

$$\kappa_1 = 2\sqrt{\kappa} \int f_0^* \left( \sum_{i=1}^{q-1} y_i^2 \right) dy_1 \dots dy_{q-1},$$

como  $\kappa_1$  no depende de  $\beta$  obtenemos la desigualdad deseada.  $\blacksquare$

**Lema 8.** *Bajo las suposiciones del Teorema 5, existen constantes positivas  $\varepsilon$ ,  $L_1$  y  $L_2$  tales que*

$$\overline{\lim}_{n \rightarrow \infty} \|\hat{\mathbf{B}}_n\| \leq L_2 \quad \text{c.s.} \quad (\text{A.33})$$

y

$$\delta \leq \underline{\lim}_{n \rightarrow \infty} \|\hat{\Gamma}_n\| \leq \overline{\lim}_{n \rightarrow \infty} \|\hat{\Gamma}_n\| \leq L_1 \quad \text{c.s.} \quad (\text{A.34})$$

con  $\hat{\Gamma}_n = |\hat{\Sigma}_n|^{-1/q} \hat{\Sigma}_n$ .

**Demostración:** Sea  $P$  la medida en  $\mathbb{R}^q \times \mathbb{R}^p$  cuya densidad es el producto de  $f_0(\mathbf{u})$  dada en (2.15) y la densidad de las  $\mathbf{x}_i$ , a la que denotaremos por  $g_0(\mathbf{x})$ . De acuerdo al Teorema 4.2 de [25] tenemos que

$$\lim_{n \rightarrow \infty} \sup_{C \subset \mathbb{R}^{p+q}, C \text{ convexo}} |P_n(C) - P(C)| = 0 \quad \text{c.s.} \quad (\text{A.35})$$

siendo  $P_n$  la medida empírica inducida por la muestra.

Por el Lema 6 existe un  $n_0$  y  $\sigma_1$  tal que

$$\sigma_1 > \hat{\sigma}_n \quad (\text{A.36})$$

para todo  $n \geq n_0$ . Si consideramos el conjunto

$$\mathcal{E}_n = \{(\mathbf{y}, \mathbf{x}) : \frac{(\mathbf{y} - \hat{\mathbf{B}}'_n \mathbf{x})' \hat{\Gamma}_n^{-1} (\mathbf{y} - \hat{\mathbf{B}}'_n \mathbf{x})}{\sigma_1^2} \leq \kappa\},$$

donde  $\kappa$  es la constante que aparece en (S3), por (A.35) podemos concluir que para  $n$  suficientemente grande

$$P(\mathcal{E}_n) > P_n(\mathcal{E}_n) - b/2$$

casi seguro.

Por (2.4), (A.36) y (2.6) tenemos que

$$\frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{d_i(\hat{\mathbf{B}}_n, \hat{\Gamma}_n)}{\sigma_1} \right) \leq \frac{1}{n} \sum_{i=1}^n \rho_0 \left( \frac{d_i(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n)}{\hat{\sigma}_n} \right) = b, \quad (\text{A.37})$$

entonces por (S3),

$$P_n(\mathcal{E}_n) = \frac{1}{n} \sharp \{(\mathbf{y}_i, \mathbf{x}_i) : (\mathbf{y}_i - \widehat{\mathbf{B}}'_n \mathbf{x}_i)' \widehat{\mathbf{\Gamma}}_n^{-1} (\mathbf{y}_i - \widehat{\mathbf{B}}'_n \mathbf{x}_i) / \sigma_1^2 \leq \kappa\} \geq b$$

y en consecuencia  $P(\mathcal{E}_n) > b/2$  casi seguro para  $n$  suficientemente grande.

Por el Lema 7, tenemos que para todo  $1 \leq j \leq n$

$$P(\mathcal{E}_n) \leq \lambda_j(\widehat{\mathbf{\Gamma}}_n)^{1/2} \sigma_1 \kappa_1,$$

entonces si tomamos  $\delta = b^2/(4\kappa_1^2\sigma_1^2)$  obtenemos que para  $n$  suficientemente grande

$$\lambda_j(\widehat{\mathbf{\Gamma}}_n) \geq \delta \quad \text{c.s.}$$

para todo  $j$ , en especial para  $\lambda_1(\widehat{\mathbf{\Gamma}}_n) = \|\widehat{\mathbf{\Gamma}}_n\|$ .

Luego como  $|\widehat{\mathbf{\Gamma}}_n| = 1$  tenemos que existe una constante  $L_1 > 0$  tal que para  $n$  suficientemente grande  $\|\widehat{\mathbf{\Gamma}}_n\| \leq L_1$ .

Por (A.37) y el Lema 6 para probar (A.33) es suficiente mostrar que para cualquier  $\sigma > 0$  existe  $L_2$  y  $\eta > 0$  tal que

$$\lim_{n \rightarrow \infty} \inf_{\|\mathbf{B}\| > L_2} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{d_i(\mathbf{B}, \widehat{\mathbf{\Gamma}}_n)}{\sigma} \right) \geq b + \eta \quad \text{c.s..} \quad (\text{A.38})$$

Por el Teorema de la convergencia dominada de Lebesgue, es fácil mostrar que para cualquier  $\sigma > 0$

$$\lim_{M_1 \rightarrow \infty} E_{F_0} \rho_1 \left( \frac{\|\mathbf{y}\| - M_1}{L_1^{1/2} \sigma} \right) = 1. \quad (\text{A.39})$$

Por (S4), existen  $\varphi > 0$ ,  $\gamma > 0$  y un número finito de conjuntos  $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_s$ , incluidos en  $\mathbb{R}^{p \times q}$  tal que

$$\bigcup_{i=1}^s \mathcal{C}_i \supset \mathcal{C} = \{\mathbf{B} \in \mathbb{R}^{p \times q} : \|\mathbf{B}\|_2 = 1\} \quad (\text{A.40})$$

y

$$P_{G_0} \left( \inf_{\mathbf{B} \in \mathcal{C}_i} \|\mathbf{B}' \mathbf{x}\| \geq \varphi \right) \geq b + \gamma. \quad (\text{A.41})$$

Por (A.39) podemos encontrar  $M_1$  y  $\eta > 0$  tal que

$$(b + \gamma)E_{F_0} \left( \rho_1 \left( \frac{\|\mathbf{y}\| - M_1}{L_1^{1/2}\sigma} \right) \right) > b + 2\eta. \quad (\text{A.42})$$

Luego de (A.41) y (A.42) obtenemos que

$$E \inf_{\mathbf{B} \in \mathcal{C}_i} I(\|\mathbf{B}'\mathbf{x}\| \geq \varphi) \rho_1 \left( \frac{\|\mathbf{y}\| - M_1}{L_1^{1/2}\sigma} \right) \geq b + 2\eta. \quad (\text{A.43})$$

Sea  $M_2$  tal que

$$P_{F_0}(\|\mathbf{y}\| \geq M_2) < \eta, \quad (\text{A.44})$$

y tomemos  $M = \max\{M_1, M_2\}$  y  $L_2 = M/\varphi$ , entonces por (A.34) y (A.40) tenemos que

$$\begin{aligned} \inf_{\|\mathbf{B}\|_2 > L_2} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{d_i(\mathbf{B}, \hat{\Gamma}_n)}{\sigma} \right) &\geq \inf_{\|\mathbf{B}\|_2 > L_2} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{\|\mathbf{y}_i - \mathbf{B}'\mathbf{x}\|}{L_1^{1/2}\sigma} \right) \\ &\geq \inf_{\|\mathbf{B}\|_2 = 1} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{\|\mathbf{y}_i\| - L_2\varphi}{L_1^{1/2}\sigma} \right) I(\|\mathbf{B}'\mathbf{x}\| > \varphi) I(\|\mathbf{y}_i\| < L_2\varphi) \\ &\geq \inf_{1 \leq j \leq s} \inf_{\mathbf{B} \in \mathcal{C}_j} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{\|\mathbf{y}_i\| - M}{L_1^{1/2}\sigma} \right) I(\|\mathbf{B}'\mathbf{x}\| > \varphi) I(\|\mathbf{y}_i\| < M) \\ &\geq \inf_{1 \leq j \leq s} \frac{1}{n} \sum_{i=1}^n \inf_{\mathbf{B} \in \mathcal{C}_j} \rho_1 \left( \frac{\|\mathbf{y}_i\| - M}{L_1^{1/2}\sigma} \right) I(\|\mathbf{B}'\mathbf{x}\| > \varphi) (1 - I(\|\mathbf{y}_i\| \geq M)). \end{aligned}$$

Por último, usando la Ley de Grandes Números, (A.43) y (A.44) obtenemos (A.38) y esto prueba (A.33). ■

**Lema 9.** Sea  $g : \mathbb{R}^k \times (\mathbb{R}^{m \times n} \times \mathbb{R}^{r \times t}) \longrightarrow \mathbb{R}$  continua y sea  $Q$  una distribución de probabilidad sobre  $\mathbb{R}^k$  tal que para cierto  $\delta > 0$  tenemos que

$$E_Q \left( \sup_{\|(\mathbf{A}, \mathbf{V}) - (\mathbf{A}_0, \mathbf{V}_0)\| \leq \delta} |g(\mathbf{z}, (\mathbf{A}, \mathbf{V}))| \right) < \infty,$$

donde  $\|\cdot\|$  es la norma definida en (A.5). Sea  $(\hat{\mathbf{A}}_n, \hat{\mathbf{V}}_n)$  una sucesión de estimadores en  $\mathbb{R}^{m \times n} \times \mathbb{R}^{r \times t}$  tal que  $\lim_{n \rightarrow \infty} (\hat{\mathbf{A}}_n, \hat{\mathbf{V}}_n) = (\mathbf{A}_0, \mathbf{V}_0)$  c.s.. Luego si  $\mathbf{z}_1, \dots, \mathbf{z}_n$ , son variables aleatorias i.i.d. en  $\mathbb{R}^k$  con distribución  $Q$ , tenemos que

$$\lim_{n \rightarrow \infty} (1/n) \sum_{i=1}^n g(\mathbf{z}_i, (\hat{\mathbf{A}}_n, \hat{\mathbf{V}}_n)) = E_Q g(\mathbf{z}, (\mathbf{A}_0, \mathbf{V}_0)) \quad \text{c.s..}$$

**Demostración:** Para probar el Lema es suficiente mostrar que para cualquier  $\varepsilon > 0$  existe  $\eta > 0$  tal que

$$\overline{\lim}_{n \rightarrow \infty} \sup_{\|(\mathbf{A}, \mathbf{V}) - (\mathbf{A}_0, \mathbf{V}_0)\| \leq \eta} (1/n) \sum_{i=1}^n g(\mathbf{z}_i, (\mathbf{A}, \mathbf{V})) \leq E_Q g(\mathbf{z}, (\mathbf{A}_0, \mathbf{V}_0)) + \varepsilon \quad (\text{A.45})$$

y

$$\underline{\lim}_{n \rightarrow \infty} \inf_{\|(\mathbf{A}, \mathbf{V}) - (\mathbf{A}_0, \mathbf{V}_0)\| \leq \eta} (1/n) \sum_{i=1}^n g(\mathbf{z}_i, (\mathbf{A}, \mathbf{V})) \geq E_Q g(\mathbf{z}, (\mathbf{A}_0, \mathbf{V}_0)) - \varepsilon \quad (\text{A.46})$$

Por el Teorema de la convergencia dominada de Lebesgue podemos tomar  $0 < \eta < \delta$  tal que

$$E \left( \sup_{\|(\mathbf{A}, \mathbf{V}) - (\mathbf{A}_0, \mathbf{V}_0)\| \leq \eta} g(\mathbf{z}, (\mathbf{A}, \mathbf{V})) \right) \leq E_Q g(\mathbf{z}, (\mathbf{A}_0, \mathbf{V}_0)) + \varepsilon. \quad (\text{A.47})$$

Luego usando la Ley de Grandes Números tenemos que

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sup_{\|(\mathbf{A}, \mathbf{V}) - (\mathbf{A}_0, \mathbf{V}_0)\| \leq \eta} g(\mathbf{z}_i, (\mathbf{A}, \mathbf{V})) = E \left( \sup_{\|(\mathbf{A}, \mathbf{V}) - (\mathbf{A}_0, \mathbf{V}_0)\| \leq \eta} g(\mathbf{z}, (\mathbf{A}, \mathbf{V})) \right). \quad (\text{A.48})$$

Finalmente de (A.47) y (A.48) obtenemos (A.45). Para probar (A.46) se realiza un procedimiento similar. ■

**Demostración del Teorema:** Consideremos

$$\begin{aligned}\mathcal{C}(\delta, L_1, L_2) &= \{(\mathbf{B}, \Gamma) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q : \delta \leq \|\Gamma\| \leq L_1, |\Gamma| = 1 \text{ y } \|\mathbf{B}\|_2 \leq L_2\}, \\ \mathcal{C}_1(\varepsilon) &= \{(\mathbf{B}, \Gamma) \in \mathcal{C}(\delta, L_1, L_2) : \|\mathbf{B}\|_2 \geq \varepsilon\} \text{ y} \\ \mathcal{C}_2(\varepsilon) &= \{(\mathbf{B}, \Gamma) \in \mathcal{C}(\delta, L_1, L_2) : \|\Gamma - \mathbf{I}\| \geq \varepsilon\}.\end{aligned}$$

De acuerdo a los Lemas 6 y 8 y (2.6), es suficiente mostrar que dados  $\varepsilon_1 > 0$ ,  $\varepsilon_2 > 0$  y  $L_1$  y  $L_2$  arbitrariamente grandes, existe  $\gamma > 0$  y  $\sigma_1 > \sigma_0$  tal que

$$\lim_{n \rightarrow \infty} \inf_{(\mathbf{B}, \Gamma) \in \mathcal{C}_1(\varepsilon_1)} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{d_i(\mathbf{B}, \Gamma)}{\sigma_1} \right) \geq E_{F_0} \rho_1 \left( \frac{(\mathbf{u}'\mathbf{u})^{1/2}}{\sigma_0} \right) + \gamma \quad \text{c.s.,} \quad (\text{A.49})$$

$$\lim_{n \rightarrow \infty} \inf_{(\mathbf{B}, \Gamma) \in \mathcal{C}_2(\varepsilon_2)} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{d_i(\mathbf{B}, \Gamma)}{\sigma_1} \right) \geq E_{F_0} \rho_1 \left( \frac{(\mathbf{u}'\mathbf{u})^{1/2}}{\sigma_0} \right) + \gamma \quad \text{c.s.} \quad (\text{A.50})$$

y

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{d_i(\tilde{\mathbf{B}}_n, \tilde{\Sigma}_n)}{\hat{\sigma}_n} \right) = E_{F_0} \rho_1 \left( \frac{(\mathbf{u}'\mathbf{u})^{1/2}}{\sigma_0} \right) \quad \text{c.s..} \quad (\text{A.51})$$

Como  $\mathbf{B}_0 = \mathbf{0}$ , por el Lema 4 tenemos que

$$E \rho_1 \left( \frac{d(\mathbf{B}, \Gamma)}{\sigma_0} \right) > E \rho_1 \left( \frac{(\mathbf{u}'\mathbf{u})^{1/2}}{\sigma_0} \right) \quad (\text{A.52})$$

para todo  $\mathbf{B} \in \mathbb{R}^{p \times q}$  y  $\Gamma \in \mathcal{S}_q$  con  $|\Gamma| = 1$  tal que  $\Gamma \neq \mathbf{I}$ .

Por el Lema 1, (A.52) y el Teorema de la convergencia dominada de Lebesgue, usando un argumento estándar de compacidad podemos encontrar  $\sigma_1 > \sigma_0$ ,  $\gamma > 0$  y un número finito de conjuntos,  $\mathcal{C}_1, \dots, \mathcal{C}_s$ , tal que

$$E_{H_0} \inf_{(\mathbf{B}, \Gamma) \in \mathcal{C}_j} \rho_1 \left( \frac{d(\mathbf{B}, \Gamma)}{\sigma_1} \right) > E_{F_0} \rho_1 \left( \frac{(\mathbf{u}'\mathbf{u})^{1/2}}{\sigma_0} \right) + \gamma \quad (\text{A.53})$$

y

$$\bigcup_{j=1}^s \mathcal{C}_j \supset \mathcal{C}_1(\varepsilon_1). \quad (\text{A.54})$$

Por (A.54) tenemos que

$$\begin{aligned} \lim_{n \rightarrow \infty} \inf_{(\mathbf{B}, \Gamma) \in \mathcal{C}_1(\varepsilon_1)} \frac{1}{n} \sum_{i=1}^n \rho_1 \left( \frac{d_i(\mathbf{B}, \Gamma)}{\sigma_1} \right) \\ \geq \inf_{1 \leq j \leq s} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \inf_{(\mathbf{B}, \Gamma) \in \mathcal{C}_j} \rho_1 \left( \frac{d_i(\mathbf{B}, \Gamma)}{\sigma_1} \right). \end{aligned}$$

Luego por (A.53) y la Ley de Grandes Números obtenemos (A.49). (A.50) se prueba de manera análoga a (A.49) y (A.51) es consecuencia del Lema 9. ■

## A.6. Demostración del Teorema 6

Antes de probar el Teorema 6 enunciaremos el Lema 2.16 (página 1036) de Pakes y Pollard [24] el cuál será utilizado posteriormente:

**Lema 10.** *Sea  $\mathfrak{F}$  una clase Euclideana con envolvente  $F$  tal que  $\int F^2 d\mu < \infty$ . Para cada  $\eta > 0$  y  $\varepsilon > 0$  existe un  $\delta > 0$  tal que*

$$\limsup \mathbb{P} \left\{ \sup_{[\delta]} |\nu_n(f_1) - \nu_n(f_2)| > \eta \right\} < \varepsilon,$$

donde  $[\delta]$  representa al conjunto de todos los pares de funciones en  $\mathfrak{F}$  con

$$\int (f_1 - f_2)^2 dP < \delta^2$$

$$\text{y } \nu_n(f) = n^{-1/2} \sum_{i=1}^n \left[ f(\zeta_i) - \int f dP \right].$$

La demostración del Lema 10 puede hallarse en [24].

**Demostración del Teorema:** Denotamos  $\theta_n = (\hat{\mathbf{B}}_n, \hat{\Sigma}_n)$  y  $\theta_0 = (\mathbf{B}_0, \sigma_0^2 \Sigma_0)$ .

Como asumimos que la distribución de los errores  $\mathbf{u}$  es elíptica con  $\Sigma_0 = \mathbf{I}$ , tenemos que para cualquier función  $h$ ,

$$E_{H_0} (x_j u_i h(\|\mathbf{u}\|)) = 0.$$



Esto implica que

$$E_{H_0} \frac{1}{\sigma_0^2} W(d(\mathbf{B}, \sigma_0^2 \Sigma_0)) (\mathbf{y} - \mathbf{B}' \mathbf{x}) \mathbf{x}' \quad (\text{A.55})$$

se anula en  $\mathbf{B} = \mathbf{B}_0$ . Entonces  $\boldsymbol{\theta}_0$  será un cero de la función  $\Phi(\boldsymbol{\theta}) = E_{H_0} \phi(\mathbf{z}; \boldsymbol{\theta})$ .

Como  $\boldsymbol{\theta}_0$  es un punto interior de  $\boldsymbol{\theta}(\mathcal{C})$  y  $\boldsymbol{\theta}_n \rightarrow \boldsymbol{\theta}_0$  c.s.,  $\boldsymbol{\theta}_n \in \boldsymbol{\theta}(\mathcal{C})$ , es decir,  $\phi_{kj}(\mathbf{z}; \boldsymbol{\theta}_n)$  y  $\phi_{kj}(\mathbf{z}; \boldsymbol{\theta}_0)$  pertenecen a la clase Euclidea  $\mathfrak{F}_{kj}$  a partir de un  $n$  suficientemente grande. Y además, por (S5), para cada  $\delta$ , las funciones  $\phi_{kj}(\mathbf{z}; \boldsymbol{\theta}_n)$  y  $\phi_{kj}(\mathbf{z}; \boldsymbol{\theta}_0)$  están en la clase  $[\delta]$  del Lema 10 para  $n$  suficientemente grande. Por lo tanto,

$$|\sqrt{n}\{\nu_n(\phi_{kj}(\cdot; \boldsymbol{\theta}_n)) - \nu_n(\phi_{kj}(\cdot; \boldsymbol{\theta}_0))\}| \longrightarrow 0 \quad (\text{A.56})$$

en probabilidad. Entonces como  $\nu_n(\phi_{kj}(\cdot; \boldsymbol{\theta}_n)) - \nu_n(\phi_{kj}(\cdot; \boldsymbol{\theta}_0)) = o_P(1/\sqrt{n})$  para todo  $k = 1, \dots, q$  y  $j = 1, \dots, p$ , concluimos que

$$\nu_n(\phi(\cdot; \boldsymbol{\theta}_n)) - \nu_n(\phi(\cdot; \boldsymbol{\theta}_0)) = o_P(1/\sqrt{n}). \quad (\text{A.57})$$

Como  $\partial\Phi/\partial\text{vec}(\mathbf{B}')'$  es continua en  $\boldsymbol{\theta}_0$ , tenemos que

$$\Phi(\mathbf{B}, \Sigma) = \Phi(\mathbf{B}_0, \Sigma) + \left( \frac{\partial\Phi(\mathbf{B}, \Sigma)}{\partial\text{vec}(\mathbf{B}')'}(\mathbf{B}_0, \Sigma) \right) \text{vec}(\mathbf{B}' - \mathbf{B}'_0) + r(\boldsymbol{\theta}) \text{vec}(\mathbf{B}' - \mathbf{B}'_0) \quad (\text{A.58})$$

donde  $r(\boldsymbol{\theta}) \rightarrow 0$  cuando  $\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}_0$ .

Usando un cambio de variables adecuado, para todo  $\Sigma \in \mathcal{S}_q$  tenemos que

$$\begin{aligned} E_{H_0} W(d(\mathbf{B}_0, \Sigma)) u_k x_j &= (Ex_j) E_{F_0} W(\mathbf{u}' \Sigma^{-1} \mathbf{u}) u_k \\ &= (Ex_j) \left( \int_{\{\mathbf{u}: u_k > 0\}} u_k W(\mathbf{u}' \Sigma^{-1} \mathbf{u}) f_0^*(\|\mathbf{u}\|^2) d\mathbf{u} \right. \\ &\quad \left. + \int_{\{\mathbf{u}: u_k < 0\}} W(\mathbf{u}' \Sigma^{-1} \mathbf{u}) u_k f_0^*(\|\mathbf{u}\|^2) d\mathbf{u} \right) \\ &= (Ex_j) \left( \int_{\{\mathbf{u}: u_k > 0\}} u_k W(\mathbf{u}' \Sigma^{-1} \mathbf{u}) f_0^*(\|\mathbf{u}\|^2) d\mathbf{u} \right. \\ &\quad \left. - \int_{\{\mathbf{u}: u_k > 0\}} W(\mathbf{u}' \Sigma^{-1} \mathbf{u}) u_k f_0^*(\|\mathbf{u}\|^2) d\mathbf{u} \right) = 0. \end{aligned}$$

Como esto vale para todo  $k = 1, \dots, q$  y  $j = 1, \dots, p$  tenemos que

$$\Phi(\mathbf{B}_0, \Sigma) = \mathbf{0} \quad (\text{A.59})$$

para todo  $\Sigma \in \mathcal{S}_q$ .

Por (2.10), el par  $\theta_n = (\hat{\mathbf{B}}_n, \hat{\Sigma}_n)$  es un cero de la función  $(1/n) \sum_{i=1}^n \phi(\mathbf{z}_i, \theta)$ . Utilizando esto, luego de realizar algunas simples operaciones de suma y resta y utilizar (A.57), obtenemos

$$\begin{aligned} \mathbf{0} &= (1/n) \sum_{i=1}^n \phi(\mathbf{z}_i, \theta_n) = E_{H_0} \phi(\mathbf{z}, \theta_n) + \left[ \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{z}_i, \theta_0) - E_{H_0} \phi(\mathbf{z}, \theta_0) \right] \\ &\quad + \left\{ \frac{1}{n} \sum_{i=1}^n [\phi(\mathbf{z}_i, \theta_n) - \phi(\mathbf{z}_i, \theta_0)] - E_{H_0} [\phi(\mathbf{z}, \theta_n) - \phi(\mathbf{z}, \theta_0)] \right\} \\ &= E_{H_0} \phi(\mathbf{z}, \theta_n) + \left[ \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{z}_i, \theta_0) - E_{H_0} \phi(\mathbf{z}, \theta_0) \right] \\ &\quad + [\nu_n(\phi(\cdot; \theta_n)) - \nu_n(\phi(\cdot; \theta_0))] \\ &= E_{H_0} \phi(\mathbf{z}, \theta_n) + \left[ \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{z}_i, \theta_0) - E_{H_0} \phi(\mathbf{z}, \theta_0) \right] + o_P(1/\sqrt{n}). \end{aligned}$$

Como  $E_{H_0} \phi(\mathbf{z}, \theta_n)$  es igual a  $\Phi(\hat{\mathbf{B}}_n, \hat{\Sigma}_n)$ , podemos despejar  $E_{H_0} \phi(\mathbf{z}, \theta_n)$  de la ecuación anterior y reemplazar la expresión resultante en la expansión (A.58) para  $\Phi(\hat{\mathbf{B}}_n, \hat{\Sigma}_n)$ . Utilizando además (A.59), obtenemos como resultado que

$$\begin{aligned} \mathbf{0} &= \left( \frac{\partial \Phi(\mathbf{B}, \Sigma)}{\partial \text{vec}(\mathbf{B}')'}(\mathbf{B}_0, \hat{\Sigma}_n) \right) \text{vec}(\hat{\mathbf{B}}'_n - \mathbf{B}'_0) + r(\theta_n) \text{vec}(\hat{\mathbf{B}}'_n - \mathbf{B}'_0) \\ &\quad + \left[ \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{z}_i, \theta_0) - E_{H_0} \phi(\mathbf{z}, \theta_0) \right] + o_P(1/\sqrt{n}). \end{aligned}$$

Como  $\partial \Phi / \partial \text{vec}(\mathbf{B}')'$  es continua en  $\theta_0$  y como  $r(\theta_n) = o_P(1)$ , esto se reduce a

$$\mathbf{0} = (\Lambda + o_P(1)) \text{vec}(\hat{\mathbf{B}}'_n - \mathbf{B}'_0) + \left[ \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{z}_i, \theta_0) - E_{H_0} \phi(\mathbf{z}, \theta_0) \right] + o_P(1/\sqrt{n}). \quad (\text{A.60})$$

De acuerdo al Teorema Central del Límite

$$\left[ \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{z}_i, \boldsymbol{\theta}_0) - E_{H_0} \phi(\mathbf{z}, \boldsymbol{\theta}_0) \right] \rightarrow O_P(1/\sqrt{n}),$$

y como  $\Lambda$  es no singular, de (A.60) obtenemos que

$$\text{vec}(\hat{\mathbf{B}}'_n - \mathbf{B}'_0) = O_P(1/\sqrt{n}).$$

Luego (A.60) puede reescribirse como

$$\mathbf{0} = \Lambda \text{vec}(\hat{\mathbf{B}}'_n - \mathbf{B}'_0) + \left[ \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{z}_i, \boldsymbol{\theta}_0) - E_{H_0} \phi(\mathbf{z}, \boldsymbol{\theta}_0) \right] + o_P(1/\sqrt{n}).$$

Como vimos al comienzo de la demostración,  $\boldsymbol{\theta}_0$  es un cero de  $\Phi(\boldsymbol{\theta}) = E_{H_0} \phi(\mathbf{z}; \boldsymbol{\theta})$ , y en consecuencia

$$\sqrt{n} \text{vec}(\hat{\mathbf{B}}'_n - \mathbf{B}'_0) = -\Lambda^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(\mathbf{z}_i, \boldsymbol{\theta}_0) + o_P(1).$$

Como  $\phi_{kj}(\mathbf{z}, \boldsymbol{\theta}_0)$  tiene media y covarianzas finitas para cada  $k = 1, \dots, q$  y  $j = 1, \dots, p$ , el Teorema queda demostrado después de aplicar el Teorema Central del Límite. ■

## A.7. Demostración de la Proposición 7

Consideremos primero el caso  $\Sigma_0 = \mathbf{I}$ . Como  $\phi(\mathbf{z}, (\mathbf{B}, \sigma_0^2 \Sigma_0))$  es integrable por (S6), es diferenciable con respecto a  $\mathbf{B}$  en un entorno de  $\mathbf{B}_0$  para casi todo  $\mathbf{z} \in \mathbb{R}^{p+q}$  y verifica (3.6), podemos derivar dentro de la esperanza en (3.4) y reescribir a  $\Lambda$  como

$$\Lambda = E_{H_0} \frac{\partial \phi(\mathbf{B}, \Sigma)}{\partial \text{vec}(\mathbf{B}')'}(\mathbf{B}_0, \sigma_0^2 \mathbf{I}).$$

Por (A.31) tenemos que

$$\Lambda = - \left[ \frac{E_{F_0} W'(\|\mathbf{u}\|/\sigma_0) \|\mathbf{u}\|}{q\sigma_0} + E_{F_0} W\left(\frac{\|\mathbf{u}\|}{\sigma_0}\right) \right] (E_{G_0} \mathbf{x}\mathbf{x}' \otimes \mathbf{I}).$$

Utilizando nuevamente las mismas herramientas y que  $W(u) = \psi_1(u)/u$ , se obtiene

$$\begin{aligned} M &= \left( E_{F_0} \left[ W \left( \frac{\|\mathbf{u}\|}{\sigma_0} \right) \right]^2 \frac{\|\mathbf{u}\|^2}{q} \right) E_{G_0} \mathbf{x} \mathbf{x}' \otimes \mathbf{I} \\ &= \left( E_{F_0} \left[ \psi_1 \left( \frac{\|\mathbf{u}\|}{\sigma_0} \right) \right]^2 \frac{\sigma_0^2}{q} \right) E_{G_0} \mathbf{x} \mathbf{x}' \otimes \mathbf{I}. \end{aligned}$$

Como  $(E_{G_0} \mathbf{x} \mathbf{x}' \otimes \mathbf{I})^{-1} = (E_{G_0} \mathbf{x} \mathbf{x}')^{-1} \otimes \mathbf{I}$ , la Proposición queda demostrada para el caso  $\Sigma_0 = \mathbf{I}$ .

Para el caso general, sea  $\mathbf{R}$  una matriz tal que  $\Sigma_0 = \mathbf{R} \mathbf{R}'$  y consideremos la siguiente transformación  $\mathbf{y}^* = \mathbf{R}^{-1} \mathbf{y}$ . Luego  $\mathbf{y}^* = \mathbf{B}_0^{*'} \mathbf{x} + \mathbf{u}^*$ , con  $\mathbf{u}^* = \mathbf{R}^{-1} \mathbf{u}$  y  $\mathbf{B}_0^* = \mathbf{B}_0 \mathbf{R}'^{-1}$ . Como la distribución de  $\mathbf{u}^*$  está dada por la densidad (2.15) con  $\Sigma_0 = \mathbf{I}$  y

$$\text{vec}(\mathbf{B}_0') = \text{vec}(\mathbf{R} \mathbf{B}_0^{*'}) = (\mathbf{I}_p \otimes \mathbf{R}) \text{vec}(\mathbf{B}_0^{*'})$$

tenemos que (3.7) es válida. ■

## A.8. Demostración de la Proposición 8

Para probar la Proposición 8 necesitaremos el Lemma 2.13 (página 1034) de Pakes y Pollard [24] el cual enunciamos a continuación:

**Lema 11.** Sea  $\mathfrak{F} = \{f(\cdot, \xi) : \xi \in \mathcal{C}\}$  una clase de funciones sobre  $\mathfrak{X}$  indexada por un subconjunto acotado  $\mathcal{C}$  de  $\mathbb{R}^d$ . Si existe un  $\alpha > 0$  y una función no negativa  $\varphi(\cdot)$  tal que

$$|f(x, \xi) - f(x, \xi^*)| \leq \varphi(x) \|\xi - \xi^*\|^\alpha \quad \text{para } x \in \mathfrak{X} \quad \text{y} \quad \xi, \xi^* \in \mathcal{C},$$

entonces  $\mathfrak{F}$  es Euclideana para la envolvente  $|f(\cdot, \xi_0)| + R\varphi(\cdot)$ , donde  $\xi_0$  es un punto arbitrario de  $\mathcal{C}$  y  $R = (2\sqrt{d} \sup_{\mathcal{C}} \|\xi - \xi_0\|)^\alpha$ .

La demostración del Lema 11 puede hallarse en [24].

**Demostración de la Proposición:** Para cada  $\xi$  en  $\mathbb{R}^{qp} \times \text{vec}(\mathcal{S}_q)$  existe un único par  $(\mathbf{B}, \Sigma)$  en  $\mathbb{R}^{q \times p} \times \mathcal{S}_q$  tal que  $\xi = (\text{vec}(\mathbf{B})', \text{vec}(\Sigma^{-1/2})')$ , luego definimos la

función  $\theta(\cdot)$  de la siguiente manera:  $\theta((\text{vec}(\mathbf{B})', \text{vec}(\Sigma^{-1/2})')) = (\mathbf{B}, \Sigma)$ .

Sean  $\varepsilon > 0$  y  $\delta = 2\|\sigma_0^{-1}\Sigma_0^{-1/2}\|$ , considerando la norma definida en (A.5), notamos por  $\mathcal{B}_\varepsilon$  a la bola de radio  $\varepsilon$  y centro  $(\mathbf{B}_0, \sigma_0^2\Sigma_0)$ , luego definimos

$$\mathcal{C} = \theta^{-1}(\{(\mathbf{B}, \Sigma) \in \mathbb{R}^{q \times p} \times \mathcal{S}_q : (\mathbf{B}, \Sigma) \in \mathcal{B}_\varepsilon \text{ y } \|\Sigma^{-1/2}\| \leq \delta\}).$$

Sean  $\xi$  y  $\xi^*$  dos elementos cualesquiera de  $\mathcal{C}$  tal que  $\theta(\xi) = (\mathbf{B}, \Sigma)$  y  $\theta(\xi^*) = (\mathbf{B}^*, \Sigma^*)$ , por el Teorema del Valor Medio existe un valor  $c$  entre  $d(\mathbf{B}, \Sigma)$  y  $d(\mathbf{B}^*, \Sigma^*)$  tal que

$$|W(d(\mathbf{B}^*, \Sigma^*)) - W(d(\mathbf{B}, \Sigma))| = |W'(c)| |d(\mathbf{B}, \Sigma) - d(\mathbf{B}^*, \Sigma^*)|. \quad (\text{A.61})$$

Como  $W$  y su derivada son continuas y de soporte compacto existe una constante  $M$  tal que  $|W(u)| \leq M$  y  $|W'(u)| \leq M$  para todo  $u$ , por esto y por (A.61) tenemos que

$$\begin{aligned} |\phi_{kj}(\mathbf{z}; \theta(\xi)) - \phi_{kj}(\mathbf{z}; \theta(\xi^*))| &\leq |W'(c)| |d(\mathbf{B}, \Sigma) - d(\mathbf{B}^*, \Sigma^*)| (y_k - \mathbf{b}_k^{*'} \mathbf{x}) x_j \\ &\quad + |W(d(\mathbf{B}, \Sigma))| |\mathbf{b}_k^{*'} \mathbf{x} x_j - \mathbf{b}_k' \mathbf{x} x_j| \\ &\leq M \{ |d(\mathbf{B}, \Sigma^*) - d(\mathbf{B}^*, \Sigma)| (|y_k| + \|\mathbf{b}_k^*\| \|\mathbf{x}\|) \\ &\quad + \|\mathbf{B}^{*'} - \mathbf{B}'\|_2 \|\mathbf{x}\| \} |x_j| \\ &\leq M |x_j| \{ |d(\mathbf{B}, \Sigma^*) - d(\mathbf{B}^*, \Sigma)| (|y_k| + \varepsilon \|\mathbf{x}\|) \\ &\quad + \|\xi - \xi^*\| \|\mathbf{x}\| \}. \end{aligned}$$

Aplicando desigualdades de normas de matrices tenemos que

$$\begin{aligned} |d(\mathbf{B}, \Sigma) - d(\mathbf{B}^*, \Sigma^*)| &\leq \|\Sigma^{-1/2}(\mathbf{y} - \mathbf{B}'\mathbf{x}) - \Sigma^{*-1/2}(\mathbf{y} - \mathbf{B}^{*'}\mathbf{x})\| \\ &\leq \|\Sigma^{-1/2} - \Sigma^{*-1/2}\|_2 \|\mathbf{y}\| + \|\Sigma^{-1/2}\mathbf{B}'\mathbf{x} - \Sigma^{*-1/2}\mathbf{B}^{*'}\mathbf{x}\| \\ &\leq \|\Sigma^{-1/2} - \Sigma^{*-1/2}\|_2 \|\mathbf{y}\| + \|\Sigma^{-1/2}\| \|\mathbf{B}' - \mathbf{B}^{*'}\|_2 \|\mathbf{x}\| \\ &\quad + \|\Sigma^{*-1/2} - \Sigma^{-1/2}\|_2 \|\mathbf{B}^*\|_2 \|\mathbf{x}\| \\ &\leq \|\Sigma^{-1/2} - \Sigma^{*-1/2}\|_2 \|\mathbf{y}\| + \delta \|\mathbf{B}' - \mathbf{B}^{*'}\|_2 \|\mathbf{x}\| \\ &\quad + \|\Sigma^{-1/2} - \Sigma^{*-1/2}\|_2 \varepsilon \|\mathbf{x}\| \\ &= \{(\varepsilon + \delta) \|\mathbf{x}\| + \|\mathbf{y}\|\} \|\xi - \xi^*\|. \end{aligned}$$

Luego, si definimos

$$\varphi_{kj}(\mathbf{z}) = M \{ ((\varepsilon + \delta) \|\mathbf{x}\| + \|\mathbf{y}\|) (|y_k x_j| + \varepsilon \|\mathbf{x}\| |x_j|) + \|\mathbf{x}\| |x_j| \} \quad (\text{A.62})$$

tenemos que  $|\phi_{kj}(\mathbf{z}; \boldsymbol{\theta}(\xi)) - \phi_{kj}(\mathbf{z}; \boldsymbol{\theta}(\xi^*))| \leq \varphi_{kj}(\mathbf{z}) \|\xi - \xi^*\|$ . Entonces podemos aplicar el Lema 11 y concluir que  $\mathfrak{F}_{kj}$  es euclídeana para la envolvente

$$F_{kj}(\mathbf{z}) = |\phi_{kj}(\mathbf{z}, \boldsymbol{\theta}(\xi_0))| + R \varphi_{kj}(\mathbf{z}), \quad (\text{A.63})$$

con  $\xi_0 \in \mathcal{C}$  arbitrario tal que  $\boldsymbol{\theta}(\xi_0) = (\mathbf{B}_0, \sigma_0^2 \Sigma_0)$  y

$$R = 2\sqrt{q(p+q)} \sup_{\mathcal{C}} \|\xi - \xi_0\|.$$

La demostración de  $E_{H_0} F_{kj}^2 < \infty$  sale inmediatamente usando que  $|W(u)| \leq M$  y expandiendo (A.62) como una suma de productos, y acotando sus respectivas medias mediante (A9). ■

## A.9. Demostración del Teorema 9

Notamos por  $\omega_i$  al peso

$$\omega_{ik} = W \left( d_i(\tilde{\mathbf{B}}^{(k)}, \tilde{\Sigma}^{(k)}) \right)$$

para cada  $1 \leq i \leq n$  y

$$\boldsymbol{\theta}^{(k)} = (\tilde{\mathbf{B}}^{(k)}, \tilde{\Sigma}^{(k)})$$

para cada  $k \geq 1$ . Luego, como  $W(u)$  es no creciente en  $|u|$  si y sólo si  $\rho$  es cóncava (ver pág. 326 de [23]), tenemos que

$$\begin{aligned} \sum_{i=1}^n \rho_1 \left( d_i(\boldsymbol{\theta}^{(k+1)}) \right) - \sum_{i=1}^n \rho_1 \left( d_i(\boldsymbol{\theta}^{(k)}) \right) \\ \leq \frac{1}{2\hat{\sigma}_n^2} \sum_{i=1}^n \omega_i \left[ d_i^2(\tilde{\mathbf{B}}^{(k+1)}, \tilde{\Gamma}^{(k+1)}) - d_i^2(\tilde{\mathbf{B}}^{(k)}, \tilde{\Gamma}^{(k)}) \right], \end{aligned} \quad (\text{A.64})$$

donde  $\tilde{\Gamma}^{(k+1)} = \tilde{\Sigma}^{(k+1)} / |\tilde{\Sigma}^{(k+1)}|^{1/q}$  y  $\tilde{\Gamma}^{(k)} = \tilde{\Sigma}^{(k)} / |\tilde{\Sigma}^{(k)}|^{1/q}$ .

Recordemos que para cualquier matriz definida positiva  $\mathbf{A}$ , la matriz  $\tilde{\mathbf{B}}^{(k+1)}$  minimiza

$$\sum_{i=1}^n \omega_i d_i^2(\mathbf{B}, \mathbf{A}).$$

Entonces

$$\sum_{i=1}^n \omega_i d_i^2(\tilde{\mathbf{B}}^{(k+1)}, \tilde{\mathbf{\Gamma}}^{(k)}) \leq \sum_{i=1}^n \omega_i d_i^2(\tilde{\mathbf{B}}^{(k)}, \tilde{\mathbf{\Gamma}}^{(k)})$$

y por lo tanto la suma del lado derecho de (A.64) es no mayor que

$$\sum_{i=1}^n \omega_i d_i^2(\tilde{\mathbf{B}}^{(k+1)}, \tilde{\mathbf{\Gamma}}^{(k+1)}) - \sum_{i=1}^n \omega_i d_i^2(\tilde{\mathbf{B}}^{(k+1)}, \tilde{\mathbf{\Gamma}}^{(k)}). \quad (\text{A.65})$$

Como

$$\tilde{\mathbf{\Gamma}}^{(k+1)} = \frac{\tilde{\mathbf{C}}^{(k+1)}}{|\tilde{\mathbf{C}}^{(k+1)}|^{1/q}} \text{ con } \tilde{\mathbf{C}}^{(k+1)} = \sum_{i=1}^n \omega_i \hat{\mathbf{u}}_i(\tilde{\mathbf{B}}^{(k+1)}) \hat{\mathbf{u}}_i'(\tilde{\mathbf{B}}^{(k+1)})$$

tenemos que  $\tilde{\mathbf{\Gamma}}^{(k+1)}$  es la matriz de covarianzas muestral de los residuos pesados

$\sqrt{\omega_i} \hat{\mathbf{u}}_i(\tilde{\mathbf{B}}^{(k+1)})$  normalizada a determinante uno, la cual minimiza la suma de las normas de Mahalanobis de los residuos pesados  $\sqrt{\omega_i} \hat{\mathbf{u}}_i'(\tilde{\mathbf{B}}^{(k+1)})$  al cuadrado entre las matrices con determinante uno, es decir, para cualquier matriz definida positiva  $\mathbf{V}$  con  $|\mathbf{V}| = 1$

$$\sum_{i=1}^n \omega_i d_i^2(\tilde{\mathbf{B}}^{(k+1)}, \tilde{\mathbf{\Gamma}}^{(k)}) \leq \sum_{i=1}^n \omega_i d_i^2(\tilde{\mathbf{B}}^{(k+1)}, \mathbf{V}).$$

Luego, como  $|\tilde{\mathbf{\Gamma}}^{(k+1)}| = |\tilde{\mathbf{\Gamma}}^{(k)}| = 1$ , tenemos que (A.65) es  $\leq 0$ . ■

# Bibliografía

---

- [1] Agulló J., Croux C., Van Aelst S. (2002). The multivariate least trimmed squares estimator. Research Report 0224, K.U. Leuven D.T.E.W.
- [2] Bhatia R. (1952). Matrix Analysis. Springer.
- [3] Bilodeau M., Duchesne P. (2000). Robust estimation of the SUR model. *Canad. J. Statist.*, 28, pp. 227 - 288.
- [4] Branco J.A., Croux C., Filzmoser P., Oliveira M.R. (2005). Robust canonical correlations: a comparative study. *Computational Statistics*, Vol. 20(2), pp. 203 - 229.
- [5] Breiman L., Friedman H. (1997). Predicting Multivariate Responses in Multiple Linear Regression. *J. R. Statistics Soc B*, 59, No. 1, pp. 3 - 54.
- [6] Croux C., Dehon C. (2002). Analyse canonique basée sur des estimateurs robustes de la matrice de covariance. *La Revue de Statistique Appliquée*, 2, pp. 5 - 26.
- [7] Croux C., Dehon C. (2003). Estimators of the multiple correlation coefficient: Local robustness and confidence intervals. *Statistical Papers*, 44, pp. 315 - 334.
- [8] Croux C., Rousseeuw P., Hössjer O. (1994). Generalized S-estimators. *Journal of the American Statistical Association* 89, pp. 1271 - 1281.
- [9] Davies P. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.*, 15, pp. 1269 - 1292.



- [10] Donoho D. L. (1982). Breakdown properties of multivariate location estimators. Unpublished manuscript. Harvard University. Ph. D. qualifying paper.
- [11] Donoho D. L., Huber P. J. (1983). The notion of breakdown-point. Festschrift fur Erich L. Lehmann Bickel, Doksum and Hodges, eds. Wadsworth, Belmont, CA.
- [12] Filzmoser P., Dehon C., Croux C. (2000). Outlier resistant estimators for canonical correlation analysis. J. G. Betlehem and P. G. M. van der Heijden (eds), COMPSTAT: Proceedings in Computational Statistics, Physica-Verlag, Heidelberg, pp. 301 - 306.
- [13] García Ben M. G., Martínez E., Yohai V. J. (2006). Robust estimation for the multivariate linear model based on a  $\tau$ -scale. Journal of Multivariate Analysis, 97, pp. 1600 - 1622.
- [14] Hampel, Frank R. (1971). A General Qualitative Definition of Robustness. The Annals of Mathematical Statistics. Vol. 42, No. 6, pp. 1887 - 1896.
- [15] Hampel, Frank R. (1974). The influence curve and its role in robust estimation, J. Amer. Statist Assoc., 69, pp. 383–393.
- [16] Hotelling H. (1936). Relations between two sets of variates. Biometrika, 28, 321 - 377.
- [17] Huber P. J. (1981), Robust Statistics, Wiley, New York.
- [18] Izenman A. J. (1975) Reduced-rank regression for the multivariate linear model. J. Multiv. Anal., 5, pp. 248 - 264.
- [19] Janssens K., Deraedt I., Freddy A., Veekman J. (1998). Composition of 15 – 17<sup>th</sup> century archeological glass vessels excavated in Antwerp, Belgium. Mikrochimica Acta, 15 (Suppl.), 253 - 267.

- [20] Kärnel G. (1991). Robust canonical correlation and correspondence analysis. In: *The Frontiers of Statistical Scientific and Industrial Applications*, (Volume II of the proceedings of ICOSCO-I, The First International Conference on Statistical Computing), American Sciences Press, Strassbourg, pp. 415 - 420.
- [21] Koenker P., Portnoy S. (1990). M-estimation of multivariate regressions. *J. Amer. Statist. Assoc.* 85, pp. 1060 - 1068.
- [22] Lopuhaä, H. (1992). Highly efficient estimators of multivariate location with high breakdown point. *Ann. Statist.* 20, pp. 398 - 413.
- [23] Maronna, R. A., Martin, R. D., Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*, John Wiley and Sons, New York.
- [24] Pakes A., Pollard D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*. Vol. 57, No. 5, pp. 1027 - 1057.
- [25] Ranga Rao R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* 33, pp. 659 - 680.
- [26] Renaud O., Victoria-Feser M.-P. (2010). A Robust Coefficient Of Determination For Regression. *Journal of Statistical Planning and Inference*, to appear.
- [27] Roelandt E., Van Aelst S., Croux C. (2009). Multivariate Generalized S-estimators, *J. Multivar. Anal.* 100, 5, pp. 876 - 887.
- [28] Rousseeuw, P. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, 79, pp. 871-880.
- [29] Rousseeuw P. (1985). Multivariate estimators with high breakdown point. W. Grossman, G. Pflug, I. Vincza, W. Wertz (Eds.), *Mathematical Statistics and its Applications*, vol. B, Reidel, Dordrecht, The Netherlands, pp. 283 - 297.
- [30] Rousseeuw P., Van Driessen K., Van Aelst S., Agulló J. (2004). Robust multivariate regression. *Technometrics* 46, pp. 293 - 305.

- [31] Salibián-Barrera, M., Van Aelst, S., and Willems, G. (2006). PCA based on Multivariate MM-estimators with Fast and Robust Bootstrap. *Journal of the American Statistical Association*, 101, pp. 1198 - 1211.
- [32] Seber, G.A.F. (2007). *A Matrix Handbook for Statisticians*. John Wiley & Sons, Inc.
- [33] Tatsuoka, K. S., and Tyler, D. E. (2000), The Uniqueness of S and M-Functionals Under Non-Elliptical Distributions, *The Annals of Statistics*, 28, pp. 1219 - 1243.
- [34] Van Aelst, S. and Willems, G. (2005). Multivariate regression S-estimators for robust estimation and inference. *Statist. Sinica* 15, pp. 981 - 1001.
- [35] Wold H. (1966). Nonlinear estimation by iterative least square procedures. F. N. David (ed.), *A Festschrift for J. Neyman*, Wiley, New York, pp. 411 - 444.
- [36] Yohai V. (1987). High Breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, Vol. 15, No. 2., pp. 642 - 656.
- [37] Yohai V., García Ben M. S. (1980). Canonical variables as optimal predictors. *The Annals of Statistics*, Vol. 8, No. 4, pp. 865 - 869.
- [38] Yohai V., Zamar R. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale, *Journal of the American Statistical Association* 83, pp. 406-413.