

Tesis de Posgrado

Métodos robustos para el modelo de análisis factorial

Villar, Ana Julia

2000

Tesis presentada para obtener el grado de Doctor en Ciencias Matemáticas de la Universidad de Buenos Aires

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Villar, Ana Julia. (2000). Métodos robustos para el modelo de análisis factorial. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.
http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_3276_Villar.pdf

Cita tipo Chicago:

Villar, Ana Julia. "Métodos robustos para el modelo de análisis factorial". Tesis de Doctor. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2000.
http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_3276_Villar.pdf

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires

Trabajo para optar al título de Doctor de la Universidad de Buenos Aires en Ciencias Matemáticas

Métodos robustos para el modelo de análisis factorial

Director: Dr. Víctor J. Yohai

Autora: Lic. Ana Julia Villar

Lugar de trabajo: Lab. de Estadística – Depto. de Matemáticas – F.C.E.N.

Fecha: junio de 2000



Métodos robustos para el modelo de análisis factorial

Palabras clave: análisis factorial, estimadores robustos de posición multivariada y covarianza, consistencia, distribución asintótica

Resumen:

Los estimadores usuales de los parámetros del modelo de análisis factorial son los estimadores de máxima verosimilitud que corresponden a factores y errores normales. Estos estimadores dependen de la media y la covarianza muestrales, las cuales son muy sensibles a observaciones atípicas. En este trabajo se obtienen estimadores robustos para el modelo de análisis factorial reemplazando en las ecuaciones de verosimilitud la media y la covarianza muestral por estimadores robustos de posición multivariada y covarianza. Se prueba la consistencia de estos estimadores cuando éstos están basados en estimadores multivariados consistentes. También se prueba la normalidad asintótica de los mismos y se obtiene la matriz de covarianza asintótica cuando los estimadores en los que se basan son asintóticamente normales. Un estudio de Monte Carlo muestra que las familias propuestas contienen estimadores que combinan alta eficiencia bajo el modelo normal con alta robustez cuando la muestra contiene observaciones atípicas

Robust methods for the factor analysis model

Keywords: factor analysis, robust multivariate location and scatter estimators, consistency, asymptotic distribution

Abstract:

The usual estimates of the parameters of the factor analysis model are maximum likelihood estimates corresponding to gaussian factors and errors. These estimators depend on the sample mean and covariance matrix, which are very sensitive to few outliers. In this work, robust estimators for the factor analysis model are obtained replacing into maximum likelihood equations sample mean and covariance by robust multivariate location and scatter estimates. Consistency of the proposed estimates is proved when consistent multivariate location and scatter matrix are used. The asymptotic normality of the proposed robust estimates is proved and their asymptotic covariance matrix obtained, provided the location and scatter matrix are asymptotically normal. A Monte Carlo study compares the behavior of the different robust proposals with maximum likelihood estimators, showing that it is possible to find estimates which combine highly efficient estimates under a normal model with a highly robust behavior when the sample contains outliers.

A mis padres,
quienes me enseñaron que el mayor tesoro que uno puede tener es la sabiduría.

A todos aquellos que me ayudaron a llegar a la meta.

Labor omnia vincit improbus
- Virgilio

Índice General

1	Introducción	3
2	El modelo de análisis factorial	5
2.1	Hipótesis adicionales debidas a indeterminaciones en el modelo	6
3	Estimación de los parámetros del modelo de análisis factorial por máxima verosimilitud	9
4	Métodos robustos	11
4.1	Estimador MVE	17
4.2	Estimador de Donoho-Stahel	19
4.3	S-estimador	20
4.4	τ -estimador	23
4.5	Estimadores ponderados de un paso para la media y la matriz de covarianza	27
5	Consistencia de los estimadores de los parámetros del modelo de análisis factorial	29
5.1	Consistencia de los estimadores de Donoho-Stahel para el modelo de análisis factorial	30
5.2	Consistencia de los S-estimadores para el modelo de análisis factorial	32
5.3	Consistencia de los τ -estimadores para el modelo de análisis factorial	33
6	Distribución asintótica de los estimadores de los parámetros del modelo de análisis factorial	35
6.1	Distribución asintótica de los S-estimadores para el modelo de análisis factorial	38
6.2	Distribución asintótica de los τ -estimadores para el modelo de análisis factorial	41
7	Simulaciones	44
8	APÉNDICES	52
A	Unicidad de los parámetros	52

B	Resultados previos	56
B.1	Resultados de álgebra lineal	56
B.2	Resultados sobre derivadas	60
C	Demostración de los resultados de consistencia y normalidad asintótica	63
C.1	Propiedades de la matriz de covarianza del modelo de análisis factorial	63
C.2	Propiedades de la función F_k	64
C.3	Obtención de los estimadores de los parámetros del modelo de análisis factorial	67
C.4	Demostración de la consistencia de los estimadores de los parámetros del modelo de análisis factorial	76
C.5	Demostración de la normalidad asintótica de los estimadores de los parámetros del modelo de análisis factorial	79
D	Métodos numéricos para calcular los estimadores	116
D.1	El procedimiento iterativo para encontrar los estimadores de los parámetros del modelo	116
D.2	Algoritmos para el cálculo de los estimadores robustos de la matriz de covarianza muestral	117
D.2.1	Estimador MVE	117
D.2.2	Estimador de Donoho-Stahel	119
D.2.3	S-estimador	119
D.2.4	τ -estimador	121

1 Introducción

Supongamos que se tiene un conjunto de variables definidas sobre cada elemento de la población. El modelo de análisis factorial postula que la relación existente entre dichas variables puede explicarse como combinaciones lineales de un número pequeño de otras variables, que denominaremos los factores, más un error aleatorio independiente de los factores. Así, bajo este modelo, la matriz de covarianza de las observaciones tendrá una estructura particular, ya que puede descomponerse como la suma de la covarianza de los factores más la covarianza de los errores.

Dada una muestra de la población se quiere identificar los factores y estimar los parámetros que conforman la estructura de la matriz de covarianza. Los estimadores que habitualmente se utilizan para este modelo son los de máxima verosimilitud correspondientes a factores y errores con distribución normal, basados en algoritmos desarrollados por Jöreskog ([16] (1963) y [17] (1967)), Clarke [6] (1970), Lawley y Maxwell [18] (1971). Sin embargo, estos estimadores son muy sensibles a unas pocas observaciones atípicas o a pequeños alejamientos en la hipótesis de normalidad.

Bajo la hipótesis de normalidad de las observaciones, los estimadores de máxima verosimilitud de los parámetros del modelo dependen de la muestra sólo a través de la matriz de covarianza muestral. Entonces, la propuesta realizada en esta tesis para obtener estimadores robustos de los parámetros del modelo de análisis factorial, fue reemplazar la matriz de covarianza muestral por un estimador robusto, consistente y asintóticamente normal de la covarianza en las ecuaciones de verosimilitud.

Los estimadores robustos estudiados están basados en los estimadores de posición multivariada y covarianza de Donoho-Stahel (Maronna y Yohai [26] (1995)), estimadores de mínimo volumen elipsoidal (MVE) de Rousseeuw (Rousseeuw y Leroy [32] (1987)) y S-estimadores (Davies [8] (1987)). Con el fin de obtener estimadores que combinen simultáneamente las propiedades de alto punto de ruptura y alta eficiencia bajo normalidad, también se desarrollaron estimadores para los parámetros de este modelo a partir de los τ -estimadores (Lopuhaä [21] y [22] (1991)) y de versiones pesadas de la matriz de covarianza muestral, con pesos basados en los estimadores MVE, S- y τ -estimadores.

Mediante un estudio de Monte Carlo, en el Capítulo 7 se muestra que los estimadores robustos así obtenidos no se ven seriamente afectados por los valores atípicos que pueden aparecer en la muestra, y además son altamente eficientes en un entorno del modelo central.

En los Capítulos 5 y 6 se obtienen resultados de consistencia y norma-

lidad asintótica de los estimadores de los parámetros del modelo de análisis factorial obtenidos a partir de un estimador consistente y asintóticamente normal de la matriz de covarianza de las observaciones. Como corolario, a partir de los resultados de Maronna y Yohai [26] (1995), Davies [8] (1987) y Lopuhaä [22] (1991) sobre la consistencia de estimadores de posición multivariada y covarianza generales, se obtiene la consistencia de los estimadores de los parámetros del análisis factorial basados en los estimadores de Donoho-Stahel y los S - y τ -estimadores. También a partir de lo demostrado por Davies [8] (1987) y Lopuhaä [20] (1989), [21] y [22] (1991) sobre la distribución asintótica de los S - y τ -estimadores multivariados, se obtiene la normalidad asintótica de los S - y τ -estimadores de análisis factorial y la expresión de la matriz de covarianza asintótica para dichos estimadores.

2 El modelo de análisis factorial

El modelo de análisis factorial se puede escribir como:

$$\mathbf{x} = \boldsymbol{\mu}_0 + \Lambda_0 \mathbf{f} + \boldsymbol{\varepsilon}, \quad (1)$$

donde $\mathbf{x} \in \mathcal{R}^p$ es el vector que contiene las p variables observadas, $\boldsymbol{\mu}_0 \in \mathcal{R}^p$, $\Lambda_0 \in \mathcal{R}^{p \times k}$ es la matriz de las ponderaciones, $\mathbf{f} \in \mathcal{R}^k$ representan las k variables latentes o factores comunes, $\boldsymbol{\varepsilon} \in \mathcal{R}^p$ son los términos residuales que representan los efectos asociados de los factores específicos y el error aleatorio.

Asumiremos que los errores tienen media cero y son no correlacionados, es decir

$$E(\boldsymbol{\varepsilon}) = \mathbf{0};$$

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \Psi_0,$$

siendo $\Psi_0 \in \mathcal{R}^{p \times p}$ una matriz diagonal con elementos diagonales positivos.

Supondremos que \mathbf{f} es un vector aleatorio. Sin pérdida de generalidad se puede asumir que

$$E(\mathbf{f}) = \mathbf{0},$$

pues si esto no ocurre, como $E(\mathbf{x}) = \Lambda_0 E(\mathbf{f}) + \boldsymbol{\mu}_0$, $\boldsymbol{\mu}_0$ puede redefinirse para absorber a $\Lambda_0 E(\mathbf{f})$.

Denotaremos con Φ_0 a la matriz de covarianza de \mathbf{f} , es decir

$$E(\mathbf{f}\mathbf{f}') = \Phi_0.$$

Asumiremos además que los errores son no correlacionados con los factores, es decir

$$E(\mathbf{f}\boldsymbol{\varepsilon}') = \mathbf{0}.$$

Observemos que las hipótesis del modelo de análisis factorial implican que la matriz de covarianza de las variables observadas \mathbf{x} , Σ_0 , está dada por:

$$\begin{aligned} \Sigma_0 &= \text{cov}(\mathbf{x}) = \text{cov}(\Lambda_0 \mathbf{f} + \boldsymbol{\varepsilon}) = \text{cov}(\Lambda_0 \mathbf{f}) + \text{cov}(\boldsymbol{\varepsilon}) \\ &= \Lambda_0 E(\mathbf{f}\mathbf{f}') \Lambda_0' + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \Lambda_0 \Phi_0 \Lambda_0' + \Psi_0. \end{aligned} \quad (2)$$

Dada una muestra aleatoria de observaciones $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{R}^p$, el problema básico es decidir si Σ_0 puede expresarse de la forma (2) para un valor determinado de $k < p$ de modo que $\text{rg}(\Lambda_0) = k$, y estimar los elementos de Λ_0 , Φ_0 y Ψ_0 .

2.1 Hipótesis adicionales debidas a indeterminaciones en el modelo

Las hipótesis hechas en la Sección anterior no son suficientes para obtener un modelo identificable, ya que el modelo presenta varias indeterminaciones.

Una indeterminación fundamental en el modelo es la siguiente. Si reemplazamos a \mathbf{f} y a Λ_0 en (1) por $\tilde{\mathbf{f}} = M\mathbf{f}$ y $\tilde{\Lambda} = \Lambda_0 M^{-1}$, donde $M \in \mathfrak{R}^{k \times k}$ es una matriz no singular, como

$$\tilde{\Lambda}\tilde{\mathbf{f}} = \Lambda_0 M^{-1} M\mathbf{f} = \Lambda_0 \mathbf{f},$$

entonces la ecuación (1) se sigue satisfaciendo cuando se cambia \mathbf{f} y Λ_0 por $\tilde{\mathbf{f}}$ y $\tilde{\Lambda}$. Además

$$E(\tilde{\mathbf{f}}) = ME(\mathbf{f}) = \mathbf{0}.$$

Luego el modelo con Λ_0 y \mathbf{f} es equivalente al modelo con $\tilde{\Lambda}$ y $\tilde{\mathbf{f}}$, es decir, aún conociendo la distribución de \mathbf{x} no se puede distinguir entre ambos modelos.

En el caso en que $M \in \mathfrak{R}^{k \times k}$ sea una matriz ortogonal, tal transformación corresponde a una rotación de los factores.

Notación: Notaremos con I_k a la matriz identidad de $k \times k$.

Parte de la indeterminación en el modelo puede eliminarse pidiendo que

$$E(\mathbf{f}\mathbf{f}') = I_k,$$

ya que si esto no ocurre, por lo visto anteriormente, si $M = \Phi_0^{-1/2}$, el modelo con $\tilde{\Lambda} = \Lambda_0 \Phi_0^{1/2}$ y $\tilde{\mathbf{f}} = \Phi_0^{-1/2} \mathbf{f}$ es equivalente al modelo con Λ_0 y \mathbf{f} , y además se verifica que

$$E(\tilde{\mathbf{f}}\tilde{\mathbf{f}}') = \Phi_0^{-1/2} E(\mathbf{f}\mathbf{f}') \Phi_0^{-1/2} = I_k.$$

Entonces podemos suponer sin pérdida de generalidad que los factores son no correlacionados y con varianzas unitarias, esto es $\Phi_0 = I_k$. En este caso, los factores se dirán que son *ortogonales*.

Sin embargo, para cualquier matriz $M \in \mathfrak{R}^{k \times k}$ ortogonal, se cumple que $E(\tilde{\mathbf{f}}\tilde{\mathbf{f}}') = I_k$, y por lo tanto, queda aún por resolver el problema de eliminar la indeterminación en la rotación. Para esto, Anderson y Rubin [5] (1956) proponen requerir la siguiente condición

$$\Lambda_0' \Psi_0^{-1} \Lambda_0 = \Delta, \quad (3)$$

donde Δ es una matriz diagonal, tal que sus elementos diagonales son las raíces no nulas de

$$|\Sigma_0 - \Psi_0 - \gamma\Psi_0| = 0. \quad (4)$$

La condición (3) no implica una pérdida de generalidad en el modelo, ya que si no se satisface, como Δ es una matriz simétrica definida positiva, existen matrices $T \in \mathfrak{R}^{k \times k}$ ortogonal y $D \in \mathfrak{R}^{k \times k}$ diagonal tales que $\Delta = TDT'$ y por lo visto anteriormente, si $M = T'$ el modelo con $\tilde{\Lambda} = \Lambda_0 T$ y $\tilde{\mathbf{f}} = T' \mathbf{f}$ es equivalente al modelo con Λ_0 y \mathbf{f} , y además se verifica que

$$E(\tilde{\mathbf{f}}\tilde{\mathbf{f}}') = T' I_k T = I_k.$$

La condición (3) determina Λ_0 unívocamente si las raíces de (4) son distintas y los elementos diagonales de Δ están ordenados en forma decreciente.

Notación: Notaremos con $\mathcal{D}(k)$ al conjunto de las matrices diagonales de $k \times k$ de rango k , con $\mathcal{D}_+(k)$ al conjunto de matrices en $\mathcal{D}(k)$ tales que sus elementos diagonales son positivos, y con $\mathcal{D}^*(k)$ al conjunto de matrices en $\mathcal{D}(k)$ tales que sus elementos diagonales son distintos y están ordenados en forma decreciente.

Resumiendo lo visto, las hipótesis que haremos sobre el modelo son:

$$E(\varepsilon) = \mathbf{0}; \quad (5)$$

$$E(\mathbf{f}) = \mathbf{0}; \quad (6)$$

$$E(\varepsilon\varepsilon') = \Psi_0; \quad (7)$$

$$E(\mathbf{f}\varepsilon') = \mathbf{0}; \quad (8)$$

$$E(\mathbf{f}\mathbf{f}') = I; \quad (9)$$

$$\Lambda_0' \Psi_0^{-1} \Lambda_0 = \Delta, \quad (10)$$

siendo $\Psi_0 \in \mathcal{D}_+(p)$ y $\Delta \in \mathcal{D}^*(k)$.

Supondremos además que

$$\text{rg}(\Lambda_0) = k. \quad (11)$$

Luego, bajo estas hipótesis, los errores son no correlacionados y los factores son ortogonales, no correlacionados entre sí ni con los errores, y el modelo es, en general, identificable (para una discusión más detallada sobre el problema de identificabilidad, ver Sección A del Apéndice).

Además, por (2)

$$\Sigma_0 = \text{cov}(\mathbf{x}) = E(\mathbf{x}\mathbf{x}') = \Lambda_0\Lambda_0' + \Psi_0. \quad (12)$$

En la Sección A del Apéndice se enuncian algunas condiciones necesarias y suficientes para la existencia del modelo.

Notación: Dada $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathfrak{R}^p$ una muestra aleatoria, llamemos

$$\mathcal{X}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}.$$

Notación: Notaremos con $\Xi = \mathfrak{R}^p \times \text{PDS}(p)$ al conjunto de pares $\xi = (\mathbf{t}, C)$, donde $\text{PDS}(p)$ es la clase de matrices en $\mathfrak{R}^{p \times p}$ que son simétricas y definidas positivas. De este modo Ξ es un subconjunto abierto de $\mathfrak{R}^{p \times p(p+1)/2}$

Definición 1 Dada una muestra aleatoria \mathcal{X}_n en \mathfrak{R}^p diremos que tiene distribución elíptica $F_{\mu, \Sigma}$ si su densidad común es de la forma

$$|\Sigma|^{-1/2} f[(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})]. \quad (13)$$

donde $f : \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$ es una función fija y $(\boldsymbol{\mu}, \Sigma) \in \Xi$.

Si $\boldsymbol{\mu} = \mathbf{0}$ y $\Sigma = I$, se dice que \mathcal{X}_n tiene distribución esférica.

Observación: Sean $\mathbf{f} \in \mathfrak{R}^k$ y $\boldsymbol{\varepsilon} \in \mathfrak{R}^p$ vectores aleatorios tales que

$$(\mathbf{f}, \boldsymbol{\varepsilon}) \text{ tiene distribución conjunta elíptica } F_{\mathbf{0}, V} \quad (14)$$

con $V = \begin{bmatrix} I_k & 0 \\ 0 & \Psi_0 \end{bmatrix}$ y $\Psi_0 \in \mathcal{D}_+(p)$.

Entonces, si existen los momentos de $(\mathbf{f}, \boldsymbol{\varepsilon})$, se satisfacen las condiciones (5) a (9).

En la Sección 6 veremos que para obtener la normalidad asintótica de los estimadores robustos de los parámetros del modelo de análisis factorial estudiados, basta pedir que se cumplan las hipótesis (10), (11) y (14), ya que si $(\mathbf{f}, \boldsymbol{\varepsilon})$ satisface (14), entonces, por (1) y (12), $\mathbf{x} \sim F_{\boldsymbol{\mu}_0, \Sigma_0}$.

Si, además de (14), \mathbf{f} y $\boldsymbol{\varepsilon}$ son independientes, entonces $\mathbf{f} \sim \mathcal{N}_k(\mathbf{0}, I_k)$ y $\boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \Psi_0)$, y por (1) y (12), $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}_0, \Sigma_0)$. Como veremos en la próxima Sección, bajo estas hipótesis (normalidad de los factores y de los errores), se obtienen los estimadores de máxima verosimilitud.

3 Estimación de los parámetros del modelo de análisis factorial por máxima verosimilitud

Supongamos que $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ es una muestra aleatoria tal que $\mathbf{x}_i \sim \mathcal{N}_p(\boldsymbol{\mu}_0, \Sigma_0)$ para $1 \leq i \leq n$, donde $\boldsymbol{\mu}_0 \in \mathbb{R}^p$, $\Sigma_0 \in \mathbb{R}^{p \times p}$ es una matriz simétrica definida positiva. Entonces, los estimadores de máxima verosimilitud del vector $\boldsymbol{\mu}_0$ y de la matriz Σ_0 están dados respectivamente por:

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i; \quad (15)$$

$$S_n = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'. \quad (16)$$

Ahora consideremos el modelo (1). Supongamos que, además de las hipótesis (5) a (11), los vectores aleatorios \mathbf{f} y $\boldsymbol{\varepsilon}$ tienen distribuciones normales multivariadas independientes, es decir:

$$\left. \begin{array}{l} \mathbf{f} \sim \mathcal{N}_k(\mathbf{0}, I_k) \\ \boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \Psi_0) \end{array} \right\} \text{independientes,} \quad (17)$$

siendo $\Psi_0 \in \mathcal{D}_+(p)$.

Entonces, las observaciones muestrales \mathbf{x}_i ($i = 1, \dots, n$) verificarán que

$$\mathbf{x}_i \sim \mathcal{N}_p(\boldsymbol{\mu}_0, \Sigma_0), \quad (18)$$

donde, por (12), Σ_0 está dada por la ecuación

$$\Sigma_0 = \Sigma(\Lambda_0, \Psi_0) = \Lambda_0 \Lambda_0' + \Psi_0. \quad (19)$$

Por (18) y (19), la densidad conjunta de $\mathbf{x}_1, \dots, \mathbf{x}_n$ está dada por:

$$\begin{aligned} L(\Lambda_0, \Psi_0, \boldsymbol{\mu}_0) &= \\ &= \prod_{i=1}^n \left\{ (2\pi)^{-p/2} |\Sigma(\Lambda_0, \Psi_0)|^{-1/2} \times \right. \\ &\quad \left. \times \exp \left[-\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_0)' \Sigma(\Lambda_0, \Psi_0)^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_0) \right] \right\} \\ &= (2\pi)^{-np/2} |\Sigma(\Lambda_0, \Psi_0)|^{-n/2} \times \\ &\quad \times \exp \left[-\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu}_0)' \Sigma(\Lambda_0, \Psi_0)^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_0) \right]. \end{aligned}$$

Entonces, los estimadores de máxima verosimilitud de $(\Lambda_0, \Psi_0, \mu_0)$ son tales que:

$$\left(\widehat{\Lambda}, \widehat{\Psi}, \widehat{\mu}\right) = \arg \max_{(\Lambda, \Psi, \mu)} L(\Lambda, \Psi, \mu).$$

Por (15)

$$\widehat{\mu} = \arg \max_{\mu} L(\Lambda, \Psi, \mu) = \bar{x}.$$

Luego

$$\left(\widehat{\Lambda}, \widehat{\Psi}\right) = \arg \max_{(\Lambda, \Psi)} L(\Lambda, \Psi, \widehat{\mu}) = \arg \max_{(\Lambda, \Psi)} L(\Lambda, \Psi, \bar{x}).$$

Sea $\mathcal{L} = \ln L(\Lambda, \Psi, \bar{x})$ el logaritmo de la función de verosimilitud evaluada en $\mu = \bar{x}$. Entonces

$$\begin{aligned} \mathcal{L} &= -\frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma(\Lambda, \Psi)| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' \Sigma(\Lambda, \Psi)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \\ &= c - \frac{n}{2} \ln |\Sigma(\Lambda, \Psi)| - \frac{1}{2} \text{tr} \left(n \Sigma(\Lambda, \Psi)^{-1} S_n \right) \\ &= c - \frac{n}{2} \left\{ \ln |\Lambda \Lambda' + \Psi| + \text{tr} \left[(\Lambda \Lambda' + \Psi)^{-1} S_n \right] \right\}, \end{aligned} \tag{20}$$

donde $c = -(np/2) \ln(2\pi)$.

De esta manera, los valores de Λ y Ψ que maximicen \mathcal{L} son los estimadores de máxima verosimilitud de Λ_0 y Ψ_0 y los denotaremos por $\widehat{\Lambda}$ y $\widehat{\Psi}$.

Para obtener los estimadores, vamos a minimizar sobre Λ y Ψ la función

$$\begin{aligned} F_k(\Lambda, \Psi, S_n) &= \ln |\Sigma(\Lambda, \Psi)| + \text{tr} \left[S_n \Sigma(\Lambda, \Psi)^{-1} \right] - \ln |S_n| \\ &= \text{tr} \left[S_n \Sigma(\Lambda, \Psi)^{-1} \right] - \ln \left| S_n \Sigma(\Lambda, \Psi)^{-1} \right|, \end{aligned} \tag{21}$$

que es equivalente a maximizar \mathcal{L} sobre Λ y Ψ , pues

$$\mathcal{L} = \ln L = -\frac{n}{2} F_k + g,$$

donde g es una constante.

4 Métodos robustos

Bajo la hipótesis de normalidad de las observaciones, hemos visto que los estimadores de máxima verosimilitud de los parámetros del modelo de análisis factorial se obtenían minimizando sobre Λ y Ψ la función

$$F_k(\Lambda, \Psi, S_n) = \ln |\Lambda\Lambda' + \Psi| + \text{tr} \left[S_n (\Lambda\Lambda' + \Psi)^{-1} \right] - \ln |S_n|.$$

Esta función depende de la muestra sólo a través de la matriz de covarianza muestral S_n . Pero S_n es un estimador de la covarianza muy sensible a unas pocas observaciones atípicas o a pequeños alejamientos en la hipótesis de normalidad.

Por lo tanto, en esta tesis, para obtener estimadores robustos de los parámetros del modelo de análisis factorial, se propone minimizar la función F_k pero reemplazando la matriz de covarianza muestral por una matriz de covarianza robusta.

Las siguientes Proposiciones, cuyas demostraciones se dan en la Sección C.3 del Apéndice, establecen qué ecuaciones deben satisfacer los estimadores de los parámetros del modelo de análisis factorial obtenidos a partir de un estimador consistente de la matriz de covarianza y proporcionan un método para desarrollar un algoritmo para calcularlos.

Observemos que si se considera la covarianza muestral S_n dada por (16), se obtienen los estimadores de máxima verosimilitud de los parámetros del modelo de análisis factorial, que satisfacen los resultados de las proposiciones 1 y 3 tal como había sido demostrado por Lawley y Maxwell [18] (1971: Cap. 4).

Notación: Notaremos con \mathcal{R} al conjunto dado por:

$$\mathcal{R} = \left\{ \Lambda \in \mathfrak{R}^{p \times k} : \text{rg}(\Lambda) = k \right\}.$$

Notación: Sean $\Lambda \in \mathcal{R}$, $\Psi \in \mathcal{D}_+(p)$ y $S \in \text{PDS}(p)$. Notaremos con $F_k(\Lambda, \Psi, S)$ a la función dada por:

$$F_k(\Lambda, \Psi, S) = \ln |\Lambda\Lambda' + \Psi| + \text{tr} \left[S (\Lambda\Lambda' + \Psi)^{-1} \right] - \ln |S|. \quad (22)$$

Proposición 1 Sea $\tilde{\Sigma} \in PSD(p)$ y F_k la función dada por (22).
Entonces,

$$\left(\hat{\Lambda}, \hat{\Psi}\right) = \arg \min_{(\Lambda, \Psi) \in \mathcal{R} \times \mathcal{D}_-(p)} F_k \left(\Lambda, \Psi, \tilde{\Sigma}\right)$$

si y sólo si $\hat{\Lambda}$ y $\hat{\Psi}$ satisfacen las ecuaciones:

$$\tilde{\Sigma} \hat{\Psi}^{-1} \hat{\Lambda} = \hat{\Lambda} \left(I + \hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda} \right) \quad (23)$$

$$\text{diag} \left(\hat{\Psi} \right) = \text{diag} \left(\tilde{\Sigma} - \hat{\Lambda} \hat{\Lambda}' \right) \quad (24)$$

Observación: Notemos que si reemplazamos en F_k a Λ por ΛM donde $M \in \mathfrak{R}^{k \times k}$ es una matriz ortogonal, no se producen cambios en la función. De este modo, si encontramos $\hat{\Lambda}$ y $\hat{\Psi}$ que minimicen F_k , $\hat{\Lambda} M$ y $\hat{\Psi}$ serán también minimizadores de F_k . Por lo tanto, para obtener unicidad, impondremos la restricción de que $\Lambda' \Psi^{-1} \Lambda \in \mathcal{D}^*(k)$.

Más formalmente, podemos probar el siguiente resultado:

Proposición 2 Sea $\tilde{\Sigma} \in PSD(p)$ y F_k la función dada por (22).
Sean $\hat{\Lambda}$ y $\hat{\Psi}$ tales que

$$\left(\hat{\Lambda}, \hat{\Psi}\right) = \arg \min_{(\Lambda, \Psi) \in \mathcal{R} \times \mathcal{D}_+(p)} F_k \left(\Lambda, \Psi, \tilde{\Sigma}\right).$$

Sea $M \in \mathfrak{R}^{k \times k}$ una matriz ortogonal tal que $\text{rg}(M) = k$, y sea $\Lambda^* = \hat{\Lambda} M$.
Entonces

$$\left(\Lambda^*, \hat{\Psi}\right) = \arg \min_{(\Lambda, \Psi) \in \mathcal{R} \times \mathcal{D}_+(p)} F_k \left(\Lambda, \Psi, \tilde{\Sigma}\right).$$

Notación: Sean $\Psi \in \mathcal{D}_+(p)$ y $\Lambda \in \mathcal{R}$. Notaremos con $\mathcal{A}(\Psi)$ y $\mathcal{B}(\Lambda, \Psi)$ a los conjuntos

$$\begin{aligned} \mathcal{A}(\Psi) &= \{ \Lambda : \Lambda' \Psi^{-1} \Lambda \in \mathcal{D}^*(k) \}; \\ \mathcal{B}(\Lambda, \Psi) &= \{ (\Lambda, \Psi) : \Lambda' \Psi^{-1} \Lambda \in \mathcal{D}^*(k) \}. \end{aligned}$$

Corolario 1 Sea $\tilde{\Sigma} \in PSD(p)$ y F_k la función dada por (22).
Entonces, si

$$\left(\hat{\Lambda}, \hat{\Psi}\right) = \arg \min_{(\Lambda, \Psi) \in \mathcal{B}(\Lambda, \Psi)} F_k \left(\Lambda, \Psi, \tilde{\Sigma}\right),$$

$\widehat{\Lambda}$ y $\widehat{\Psi}$ satisfacen las ecuaciones (23) y (24) de la Proposición 1, y además

$$\widehat{\Lambda}'\widehat{\Psi}^{-1}\widehat{\Lambda} = \Delta, \quad (25)$$

donde $\Delta \in \mathcal{D}^*(k)$.

Notación: Sean $\Psi \in \mathcal{D}_+(p)$ y $(\widetilde{\Sigma}_n)_{n \geq 1} \in \mathfrak{R}^{p \times p}$ una sucesión de matrices tales que $\widetilde{\Sigma}_n \in \text{PSD}(p) \forall n \in \mathbf{N}$.

Definamos

$$\widetilde{\Sigma}_n^*(\Psi) = \Psi^{-1/2}\widetilde{\Sigma}_n\Psi^{-1/2}. \quad (26)$$

Notaremos con $\widetilde{\theta}_1^n > \dots > \widetilde{\theta}_k^n \geq \dots \geq \widetilde{\theta}_p^n$ a los autovalores de $\widetilde{\Sigma}_n^*(\Psi)$ y con $\widetilde{\mathbf{w}}_1^n, \dots, \widetilde{\mathbf{w}}_p^n$ a sus correspondientes autovectores normalizados.

Sean

$$\widetilde{\Theta}_n(\Psi) = \begin{bmatrix} \widetilde{\theta}_1^n & 0 \\ & \vdots \\ 0 & \widetilde{\theta}_k^n \end{bmatrix}; \quad \widetilde{\Omega}_n(\Psi) = [\widetilde{\mathbf{w}}_1^n \cdots \widetilde{\mathbf{w}}_k^n]. \quad (27)$$

En el caso de considerar $\widetilde{\Sigma}^*(\Psi) = \Psi^{-1/2}\widetilde{\Sigma}\Psi^{-1/2}$, notaremos en forma análoga con $\widetilde{\Theta}(\Psi)$ y $\widetilde{\Omega}(\Psi)$ a las matrices correspondientes a los autovalores y autovectores de $\widetilde{\Sigma}^*(\Psi)$.

Proposición 3 Sea $\widetilde{\Sigma} \in \text{PSD}(p)$ y F_k la función dada por (22).

Entonces:

i) dada $\Psi \in \mathcal{D}_+(p)$, si

$$\widetilde{\Lambda} = \arg \min_{\Lambda \in \mathcal{A}(\Psi)} F_k(\Lambda, \Psi, \widetilde{\Sigma}),$$

se tiene que

$$\widetilde{\Lambda} = \Psi^{1/2}\widetilde{\Omega}(\widetilde{\Theta} - I)^{1/2},$$

siendo $\widetilde{\Theta} = \widetilde{\Theta}(\Psi)$ y $\widetilde{\Omega} = \widetilde{\Omega}(\Psi)$ dadas por (27).

ii) si

$$(\widehat{\Lambda}, \widehat{\Psi}) = \arg \min_{(\Lambda, \Psi) \in \mathcal{B}(\Lambda, \Psi)} F_k(\Lambda, \Psi, \widetilde{\Sigma}),$$

$\widehat{\Lambda}$ y $\widehat{\Psi}$ deben satisfacer las ecuaciones:

$$\begin{aligned}\widehat{\Lambda} &= \widehat{\Psi}^{1/2} \widetilde{\Omega}(\widehat{\Psi}) \left[\widetilde{\Theta}(\widehat{\Psi}) - I \right]^{1/2}; \\ \text{diag}(\widehat{\Psi}) &= \text{diag}(\widetilde{\Sigma} - \widehat{\Lambda} \widehat{\Lambda}'); \\ \widehat{\Lambda}' \widehat{\Psi}^{-1} \widehat{\Lambda} &= \widetilde{\Theta}(\widehat{\Psi}) - I.\end{aligned}\tag{28}$$

Definición 2 Sea $A \in \mathfrak{R}^{m \times n}$. Definiremos la norma Frobenius de A como

$$\|A\| = [\text{tr}(AA')]^{1/2}.$$

Corolario 2 Sean $\Psi_0 \in \mathcal{D}_+(p)$ y $\Lambda_0 \in \mathcal{R}$. Sea $\Sigma_0 = \Lambda_0 \Lambda_0' + \Psi_0$.

Sea $(\widetilde{\Sigma}_n)_{n \geq 1} \in \mathfrak{R}^{p \times p}$ una sucesión de matrices tales que $\widetilde{\Sigma}_n \in \text{PSD}(p) \forall n$ y

$$\widetilde{\Sigma}_n \xrightarrow{c.s.} \Sigma_0.$$

Dado $\varepsilon > 0$, sean

$$C = \{\Psi \in \mathcal{D}_+(p) : \|\Psi - \Psi_0\| < \varepsilon\};$$

$$D = \{\Lambda \in \mathcal{R} : \|\Lambda - \Lambda_0\| < \varepsilon\}.$$

Entonces, dada $\Psi \in C$, si

$$f_k(\Psi) = \min_{\Lambda \in \mathcal{A}(\Psi) \cap D} F_k(\Lambda, \Psi, \widetilde{\Sigma}_n),$$

se tiene que:

$$f_k(\Psi) = \sum_{m=k+1}^p (\widetilde{\theta}_m^n - \ln \widetilde{\theta}_m^n) - (p - k),\tag{29}$$

siendo $\widetilde{\Theta}_n = \widetilde{\Theta}_n(\Psi)$ y $\widetilde{\Omega}_n = \widetilde{\Omega}_n(\Psi)$ dadas por (27).

Los estimadores robustos de los parámetros del modelo de análisis factorial estudiados fueron obtenidos a partir de los siguientes estimadores robustos de posición multivariada y covarianza:

- estimadores MVE
- estimadores de Donoho-Stahel

- S-estimadores
- τ -estimadores

Veremos que los estimadores de los parámetros del modelo de análisis factorial así obtenidos “heredan” las propiedades de robustez de los estimadores anteriores, y que los basados en los S- y τ -estimadores son consistentes y asintóticamente normales.

Una propiedad que deseáramos que cumpliesen los estimadores de Σ es que, si se hace una reparametrización en el espacio de las observaciones mediante una transformación afín, el estimador sea equivariante.

Es decir, si se reemplaza \mathbf{x} por $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, como $\Sigma_{\mathbf{y}} = A\Sigma_{\mathbf{x}}A'$, es natural que el estimador $\widehat{\Sigma}_{\mathbf{y}}$ satisfaga la misma ecuación, o sea $\widehat{\Sigma}_{\mathbf{y}} = A\widehat{\Sigma}_{\mathbf{x}}A'$. De este modo, dilaciones, contracciones, rotaciones o traslaciones en los vectores observados no afectarán al estimador.

Más formalmente daremos la siguiente definición:

Definición 3 Diremos que un estimador de covarianza $C_n(\mathbf{x}_1, \dots, \mathbf{x}_n) = C_n$ es equivariante afín si y sólo si

$$C_n(A\mathbf{x}_1 + \mathbf{b}, \dots, A\mathbf{x}_n + \mathbf{b}) = AC_n(\mathbf{x}_1, \dots, \mathbf{x}_n)A',$$

para cualquier vector \mathbf{b} y cualquier matriz A no singular.

La matriz de covarianza muestral S_n , lo mismo que los estimadores robustos estudiados tienen esta propiedad.

Una medida de la robustez de un estimador es el punto de ruptura. Informalmente, el punto de ruptura es la menor fracción de contaminación en la muestra que puede causar que el estimador se vaya más allá de cualquier límite. En el caso de matrices de covarianza, el estimador puede explotar hacia infinito o hacia una matriz singular.

Seguidamente definiremos los puntos de ruptura para estimadores de posición multivariados y de matriz de covarianza.

Notación: Dada $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathfrak{R}^p$ una muestra aleatoria, llamemos

$$\mathcal{X}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\};$$

$$\mathcal{D}(\mathcal{X}_n, m) = \{\mathcal{X}_n^* = \{\mathbf{x}_1^*, \dots, \mathbf{x}_n^*\} \text{ tales que } \#(\mathcal{X}_n^* \cap \mathcal{X}_n^C) = m\}.$$

Definición 4 El punto de ruptura de un estimador de posición multivariado \mathbf{t}_n basado en \mathcal{X}_n se define como la menor fracción de contaminación m/n con outliers que puede tomarse de modo que el estimador se haga infinito.

$$\varepsilon_n^*(\mathbf{t}_n, \mathcal{X}_n) = \min_{1 \leq m \leq n} \left\{ \frac{m}{n} : \sup_{\mathcal{X}_n^* \in \mathcal{D}(\mathcal{X}_{n,m})} \|\mathbf{t}_n(\mathcal{X}_n) - \mathbf{t}_n(\mathcal{X}_n^*)\| = \infty \right\}.$$

Definición 5 El punto de ruptura de un estimador de covarianza C_n basado en \mathcal{X}_n se define como la menor fracción de contaminación m/n de outliers que puede tomarse de modo que el mayor autovalor $\gamma_1(C_n)$ se vaya a infinito o que el menor autovalor $\gamma_p(C_n)$ se acerque arbitrariamente a cero.

$$\varepsilon_n^*(C_n, \mathcal{X}_n) = \min_{1 \leq m \leq n} \left\{ \frac{m}{n} : \sup_{\mathcal{X}_n^* \in \mathcal{D}(\mathcal{X}_{n,m})} D(C_n(\mathcal{X}_n), C_n(\mathcal{X}_n^*)) = \infty \right\},$$

donde $D(A, B) = \max\{|\gamma_1(A) - \gamma_1(B)|, |\gamma_p^{-1}(A) - \gamma_p^{-1}(B)|\}$, con $\gamma_1(A) \geq \dots \geq \gamma_p(A)$ los autovalores ordenados de la matriz A .

Davies [8] (1987) dio la siguiente definición de punto de ruptura para estimadores de posición multivariados y matriz de covarianza simultáneamente.

Definición 6 Una muestra \mathcal{X}_n de dimensión p y tamaño n está en posición general si la cantidad de puntos que yacen en cualquier subespacio afín de dimensión $p - 1$ no es mayor que p .

Definición 7 Sea \mathcal{X}_n una muestra aleatoria de dimensión p y tamaño n tal que $n/2 \geq p+1$ y que está en posición general. Para \mathbf{t}_n y C_n los estimadores de posición multivariada y covarianza respectivamente basados en \mathcal{X}_n , sea

$$T(\mathcal{X}_n) = \|\mathbf{t}_n\| + \sum_{j=1}^p \left(\gamma_j^{(n)} + \frac{1}{\gamma_j^{(n)}} \right),$$

donde $\gamma_1^{(n)}, \dots, \gamma_p^{(n)}$ son los autovalores de C_n .

El punto de ruptura muestral finito ε_n^* para (\mathbf{t}_n, C_n) se define por:

$$\varepsilon_n^*(T_n, \mathcal{X}_n) = \min_{1 \leq m \leq n} \left\{ \frac{m}{n} : \sup_{\mathcal{X}_n^* \in \mathcal{D}(\mathcal{X}_{n,m})} T(\mathcal{X}_n^*) = \infty \right\}.$$

En esta definición se incluyen simultáneamente el estimador de posición multivariado y de covarianza; este último usualmente suele tener un punto de ruptura menor que el que se obtiene restringiendo la definición al punto de ruptura del estimador de posición multivariado.

Acorde con la intuición, se puede demostrar que ningún estimador equivariante afín puede tener punto de ruptura mayor que $1/2$.

Notación: Notaremos con $\lfloor z \rfloor$ al mayor entero menor o igual que z (parte entera de z) y con $\lceil z \rceil$ al menor entero mayor o igual que z .

Davies [8] (1987: Teorema 6) encuentra una cota superior para el punto de ruptura de los estimadores equivariantes afines multivariados.

Teorema 1 *Supongamos que $n \geq p + 1$ y que la muestra está en posición general.*

Entonces, una cota superior para el punto de ruptura de cualquier estimador equivariante afín de posición multivariado y covarianza es

$$\left\lfloor \frac{(n - p + 1)}{2} \right\rfloor / n.$$

4.1 Estimador MVE

Un estimador equivariante afín y con punto de ruptura maximal fue propuesto por Rousseeuw y Leroy [32] (1987). Este estimador se denomina *estimador de mínimo volumen elipsoidal* (estimador MVE) y se define de la siguiente manera: dada una muestra aleatoria \mathcal{X}_n en \mathfrak{R}^p , sea:

$$\mathbf{t}(\mathcal{X}_n) = \text{centro del elipsoide de mínimo volumen que cubre (al menos) } h \text{ puntos de } \mathcal{X}_n$$

donde se suele tomar $h = \lfloor (n + p + 1)/2 \rfloor$.

El estimador de la matriz de covarianza está dado por la matriz de la forma cuadrática que define al elipsoide, multiplicado por un factor apropiado para obtener consistencia en el caso normal.

Más formalmente daremos las siguientes definiciones.

Definición 8 *Dados $\mathbf{x} \in \mathfrak{R}^p$, $\boldsymbol{\mu} \in \mathfrak{R}^p$ y $\Sigma \in PDS(p)$, se define la distancia de Mahalanobis por la ecuación*

$$d^2(\mathbf{x}, \boldsymbol{\mu}, \Sigma) = (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}).$$

Definición 9 Sean $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$. Se define el MVE estimador multivariado de posición y covarianza como el vector \mathbf{t}_n y la matriz $C_n = \kappa C_n^*$, donde (\mathbf{t}_n, C_n^*) es la solución del problema de encontrar $(\mathbf{t}, C) \in \Xi$ tal que minimicen

$$|C|$$

sujeto a

$$\#\{i : d^2(\mathbf{x}_i, \mathbf{t}, C) \leq 1\} \geq \left\lfloor \frac{n+p+1}{2} \right\rfloor.$$

y κ es una constante apropiada para obtener consistencia bajo normalidad.

Rousseeuw y Leroy [32] (1987: pág. 258) demuestran que los estimadores MVE son equivariantes afines.

El siguiente resultado, obtenido por Lopuhaä y Rousseeuw [23] (1991: Teorema 3.1) determina los puntos de ruptura de los estimadores MVE según las Definiciones 4 y 5.

Teorema 2 Sea \mathcal{X}_n una muestra aleatoria de dimensión p y tamaño n en posición general tal que $n \geq p+1$.

Sean \mathbf{t}_n y C_n los estimadores MVE de posición multivariada y covarianza. Si $p = 1$, entonces

$$\begin{aligned} \varepsilon_n^*(\mathbf{t}_n, \mathcal{X}_n) &= \frac{1}{n} \left\lfloor \frac{n+1}{2} \right\rfloor; \\ \varepsilon_n^*(C_n, \mathcal{X}_n) &= \frac{1}{n} \left\lfloor \frac{n}{2} \right\rfloor. \end{aligned}$$

Si $p \geq 2$, entonces

$$\varepsilon_n^*(\mathbf{t}_n, \mathcal{X}_n) = \varepsilon_n^*(C_n, \mathcal{X}_n) = \frac{1}{n} \left\lfloor \frac{n-p+1}{2} \right\rfloor.$$

Como la velocidad de convergencia de los estimadores MVE es $n^{1/3}$ (Davies [8] (1987)), como era de esperar, al realizar las simulaciones se observó que la estimación de los parámetros del modelo de análisis factorial era muy ineficiente. Sin embargo, como veremos en la Sección 4.5, el estimador MVE puede ser utilizado como una solución inicial para un estimador pesado de un paso, mejorando de esta manera su eficiencia.

4.2 Estimador de Donoho-Stahel

Otra propuesta, con la que se han obtenido resultados más satisfactorios, en el sentido de menor sesgo en la estimación, es el estimador propuesto por Donoho (1982) y Stahel (1981). Este estimador robusto multivariado de posición y escala es simultáneamente equivariante afín y tiene alto punto de ruptura para cualquier dimensión.

La idea de los estimadores de Donoho-Stahel es considerar medias y covarianzas muestrales ponderadas donde el peso de cada observación es una función decreciente de una medida de atipicidad.

Esta medida está basada en la idea de que si un punto es una observación atípica multivariada, debe existir una proyección unidimensional de los datos para la cual el punto es una observación atípica univariada.

Definición 10 Sea \mathcal{X}_n una muestra aleatoria de dimensión p y tamaño n . Sean $\mu(\cdot)$ y $\sigma(\cdot)$ estadísticos univariados de posición y escala equivariantes por traslaciones y escala respectivamente.

Definamos para cualquier $\mathbf{y} \in \mathbb{R}^p$ la “atipicidad” r :

$$r(\mathbf{y}, \mathcal{X}_n) = \sup_{\mathbf{a} \in S_p} r_1(\mathbf{y}, \mathbf{a}, \mathcal{X}_n), \quad (30)$$

donde

$$r_1(\mathbf{y}, \mathbf{a}, \mathcal{X}_n) = \frac{|\mathbf{a}'\mathbf{y} - \mu(\mathbf{a}'\mathcal{X}_n)|}{\sigma(\mathbf{a}'\mathcal{X}_n)}, \quad (31)$$

y $S_p = \{\mathbf{a} \in \mathbb{R}^p : \|\mathbf{a}\| = 1\}$.

Definición 11 Sea \mathcal{X}_n una muestra aleatoria de dimensión p y tamaño n . Sea $w : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ una función, a la que llamaremos “función de peso”, tal que $w(u)$ es estrictamente decreciente y $w(u)$ y $u^2w(u)$ son acotadas para $u \geq 0$.

El estimador de Donoho-Stahel de locación y escala $(\mathbf{t}(\mathcal{X}_n), C(\mathcal{X}_n))$ se define como

$$\mathbf{t}(\mathcal{X}_n) = \frac{\sum_{i=1}^n w_i \mathbf{x}_i}{\sum_{i=1}^n w_i};$$

$$C(\mathcal{X}_n) = \kappa \left\{ \frac{\sum_{i=1}^n w_i [\mathbf{x}_i - \mathbf{t}(\mathcal{X}_n)] [\mathbf{x}_i - \mathbf{t}(\mathcal{X}_n)]'}{\sum_{i=1}^n w_i} \right\}$$

con $w_i = w(r(\mathbf{x}_i, \mathcal{X}_n))$, r la medida de atipicidad dada en (30) y κ es una constante apropiada para obtener consistencia bajo normalidad.

Un caso particular del estimador de Donoho-Stahel, que proporciona estimadores robustos (Donoho [9] (1982)), se obtiene considerando en (30) y (31) μ y σ como la mediana y la MAD respectivamente. Es decir,

$$r(\mathbf{x}_i, \mathcal{X}_n) = \sup_{\mathbf{a} \in \mathbb{S}_p} \left[\frac{|\mathbf{x}'_i \mathbf{a} - \text{med}(\mathbf{x}'_1 \mathbf{a}, \dots, \mathbf{x}'_n \mathbf{a})|}{\text{MAD}(\mathbf{x}'_1 \mathbf{a}, \dots, \mathbf{x}'_n \mathbf{a})} \right]. \quad (32)$$

El valor de la constante κ está dado por:

$$\kappa = \frac{1}{\chi_{p,0.5}^2} \left\{ \text{med}_{1 \leq i \leq n} [\mathbf{x}_i - \mathbf{t}(\mathcal{X}_n)]' C(\mathcal{X}_n)^{-1} [\mathbf{x}_i - \mathbf{t}(\mathcal{X}_n)] \right\} \quad (33)$$

donde $\chi_{p,0.5}^2$ es la mediana de la distribución χ^2 .

Si en (32) se reemplaza la mediana por la media y la MAD por la desviación standard muestral, se obtiene la raíz cuadrada de la distancia de Mahalanobis respecto de $\bar{\mathbf{x}}$ y S_n , es decir, si $z_{a_i} = \mathbf{x}'_i \mathbf{a}$

$$d(\mathbf{x}_i, \bar{\mathbf{x}}, S_n) = \sup_{\mathbf{a} \in \mathbb{S}_p} \left[\frac{|z_{a_i} - \bar{z}_a|}{\text{SD}(z_{a_1}, \dots, z_{a_n})} \right].$$

La utilización de la mediana y la MAD hace que esta medida sea más estable a la presencia de outliers en la muestra.

Donoho [9] (1982) ha probado que estos estimadores, en el caso en que μ y σ sean la mediana y la MAD, son equivariantes afines y que tienen punto de ruptura, según la Definición 7, (suponiendo $n > 2p + 1$) $\varepsilon_n^*(T_n, \mathcal{X}_n) = \lfloor (n - 2p + 2)/2 \rfloor / n$.

Tyler [42] (1994) obtuvo que, si se consideran en (31) μ como la mediana y σ como la MAD, si las observaciones no son repetidas, y como la $p - 1 + \lfloor (n + 1)/2 \rfloor$ -ésima menor desviación absoluta de la mediana en caso de haber observaciones repetidas, se obtiene el punto de ruptura maximal de los estimadores equivariantes afines de posición y covarianza, es decir $\varepsilon_n^*(T_n, \mathcal{X}_n) = \lfloor (n - p + 1)/2 \rfloor / n$ (ver Teorema 1).

4.3 S-estimador

Los S-estimadores fueron introducidos inicialmente en el contexto de regresión por Rousseeuw y Yohai [34] (1984), y posteriormente fueron generalizados por Davies [8] (1987) al caso de posición y covarianza multivariado.

Definición 12 Sean $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$. Sean $\rho : \mathbb{R} \rightarrow \mathbb{R}_+$ una función no decreciente y b una constante tal que $0 < b < \sup \rho$. Se define el S-estimador

de posición y covarianza multivariado como la solución (\mathbf{t}_n, C_n) del problema de encontrar $(\mathbf{t}, C) \in \Xi$ tal que minimicen

$$|C| \quad (34)$$

sujeto a

$$\frac{1}{n} \sum_{i=1}^n \rho[d(\mathbf{x}_i, \mathbf{t}, C)] = b. \quad (35)$$

Denotemos a este problema de minimización con \mathcal{P}_n .

Si $\rho(x) = I_{[0,1]}^{(x)}$ con $nb = \lfloor (n+p+1)/2 \rfloor$ obtenemos el MVE estimador. Si $\rho(x) = x^2$ obtenemos el estimador de mínimos cuadrados, que es equivalente al estimador de máxima verosimilitud. Por lo tanto, ambos estimadores que se han estudiado previamente, pueden ser vistos como casos particulares de los S-estimadores.

Rousseeuw y Yohai [34] (1984), con el objetivo de lograr simultáneamente un alto punto de ruptura y normalidad asintótica, sugirieron restringir la elección de las funciones ρ a la clase que satisfacen las siguientes condiciones:

- (H1) ρ es simétrica, con derivada continua ψ y tal que $\rho(0) = 0$.
- (H2) Existe una constante c_0 tal que $0 < c_0 < \infty$ y ρ es estrictamente creciente en $[0, c_0]$ y constante en $[c_0, \infty)$. Sea $a = \sup \rho = \rho(c_0)$.

De esta manera, los S-estimadores pueden verse como versiones “suaves” del MVE estimador.

Un ejemplo de una función ρ dentro de esta clase es

$$\rho_c(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4} & \text{si } |x| \leq c \\ \frac{c^2}{6} & \text{si } |x| > c \end{cases} \quad (36)$$

donde la constante $c > 0$ y por supuesto, b dependerá de c . Su derivada es una función redescendiente, conocida como la función bicuadrada de Tukey

$$\psi_c(x) = x \left[1 - \left(\frac{x}{c} \right)^2 \right]^2 I_{[-c,c]}^{(x)}$$

En el estudio de Monte Carlo de la Sección 7 se ha considerado esta familia de funciones ρ_c .

Davies [8] (1987) define los S-estimadores en forma análoga a la Definición 12, sólo que en vez de (35), considera la restricción

$$\frac{1}{n} \sum_{i=1}^n \rho[d(\mathbf{x}_i, \mathbf{t}, C)] \leq b.$$

Además, bajo hipótesis más débiles que las consideradas en (H1) y (H2), de modo que la función ρ sea simétrica, no decreciente, continua en 0, y continua a izquierda en $(0, \infty)$, demuestra la existencia y la consistencia de estos estimadores.

Las soluciones del problema de minimización \mathcal{P}_n dependen obviamente de la elección de las constantes b y c . Como veremos en el siguiente Teorema, que fue probado por Davies [8] (1987: Teorema 5), $b/\rho(c)$ determina el punto de ruptura del S-estimador.

Teorema 3 Sea $\rho : \mathfrak{R} \rightarrow \mathfrak{R}_+$ una función no decreciente y b una constante tal que $0 < b < a = \sup \rho$.

Supongamos que exista $\eta > 0$ tal que $\rho(x) = 0$ para $0 \leq x < \eta$.

Sea \mathcal{X}_n una muestra aleatoria en \mathfrak{R}^p tal que $n(1 - 2b/a) \geq p + 1$ y que está en posición general.

Para los S-estimadores (\mathbf{t}_n, C_n) basados en \mathcal{X}_n , sea

$$T_n = T(\mathcal{X}_n) = \|\mathbf{t}_n\| + \sum_{j=1}^p \left(\gamma_j^{(n)} + \frac{1}{\gamma_j^{(n)}} \right),$$

donde $\gamma_1^{(n)}, \dots, \gamma_p^{(n)}$ son los autovalores de C_n .

Entonces, el punto de ruptura muestral finito ϵ_n^* para (\mathbf{t}_n, C_n) (según Definición 7) está dado por:

$$\epsilon_n^*(T_n, \mathcal{X}_n) = \frac{1}{n} \left(\left\lfloor \frac{nb}{a} \right\rfloor + 1 \right).$$

Corolario 3

$$\lim_{n \rightarrow \infty} \epsilon_n^*(T_n, \mathcal{X}_n) = \frac{b}{a}.$$

Corolario 4 Si $2(b/a) = [1 - (p + 1)/n]$ entonces

$$\varepsilon_n^*(T_n, \mathcal{X}_n) = \frac{1}{n} \left\lfloor \frac{n - p + 1}{2} \right\rfloor.$$

Sea $r = b/a$. Del Teorema 1 y de los resultados anteriores, se deduce que si $r = (n - p)/2n$, se obtiene el punto de ruptura maximal. Sin embargo, la constante c determina al mismo tiempo la varianza asintótica y Lopuhaä [20] (1989) señala que no es posible lograr simultáneamente alta eficiencia asintótica respecto al estimador de máxima verosimilitud y un punto de ruptura del 50%.

A partir de los resultados anteriores, veamos cómo determinar las constantes b y c para obtener estimadores con alto punto de ruptura y consistentes. La constante $0 < b < a$ puede elegirse de acuerdo con la distribución subyacente de modo de obtener estimadores consistentes. Por ejemplo si $\mathbf{x}_1, \dots, \mathbf{x}_n$ son vectores aleatorios en \mathfrak{R}^p con distribución elíptica $F_{\mu, \Sigma}$, entonces la constante b , que depende de c , se elige de manera que

$$b = \mathbf{E}[\rho(d(\mathbf{x}, \mu, \Sigma))] = \int \rho(\|\mathbf{x}\|) f(\|\mathbf{x}\|) d\mathbf{x}, \quad (37)$$

y la constante c se determina de modo que

$$0 < r = \frac{b}{a} \leq \frac{n - p}{2n},$$

lo que lleva a un punto de ruptura para muestras finitas de $\varepsilon_n^*(T_n, \mathcal{X}_n) = \lfloor nr \rfloor / n$.

Davies [8] (1987) demuestra que, si $p + 1 \leq n(1 - b/a)$ entonces el problema \mathcal{P}_n tiene al menos una solución con probabilidad 1, es decir el S-estimador está bien definido. Además, demuestra que estos estimadores son equivariantes afines.

4.4 τ -estimador

Los τ -estimadores fueron inicialmente introducidos en el contexto de regresión por Yohai y Zamar [44] (1988) como una extensión de los S-estimadores, con el fin de obtener estimadores que tengan simultáneamente alta eficiencia bajo normalidad y alto punto de ruptura. Posteriormente, fueron generalizados por Lopuhaä [22] (1991) para el caso de posición multivariada y covarianza.

Asintóticamente, los τ -estimadores son equivalentes a los M-estimadores definidos mediante una función ψ que es un promedio pesado de dos funciones ψ redescendientes, una correspondiente a un estimador robusto con alto punto de ruptura, y la otra a un estimador altamente eficiente.

Definición 13 Sean $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$. Dadas dos funciones ρ_i ($i = 1, 2$) no decrecientes de \mathbb{R}_+ en \mathbb{R}_+ , se define el τ -estimador multivariado de posición y covarianza como el vector \mathbf{t}_n y la matriz C_n , donde (\mathbf{t}_n, C_n^*) es la solución del problema de encontrar $(\mathbf{t}, C) \in \Xi$ tal que minimicen

$$|C| \sum_{i=1}^n \rho_2 [d(\mathbf{x}_i, \mathbf{t}, C)] \quad (38)$$

sujeito a

$$\frac{1}{n} \sum_{i=1}^n \rho_1 [d(\mathbf{x}_i, \mathbf{t}, C)] = b_1 \quad (39)$$

y

$$C_n = \frac{C_n^*}{b_2} \left\{ \frac{1}{n} \sum_{i=1}^n \rho_2 [d(\mathbf{x}_i, \mathbf{t}_n, C_n^*)] \right\}$$

Denotemos a este problema de minimización con \mathcal{P}_n^τ .

Notemos que los τ -estimadores son una extensión de los S-estimadores, que se definen como $\min |C|$ sujeto a (39). Además, si se eligen $\rho_1 = \rho_2$ y $b_1 = b_2$, entonces \mathbf{t}_n y $C_n = C_n^*$ son los S-estimadores. El estimador de máxima verosimilitud puede obtenerse como un caso particular de los τ -estimadores haciendo $\rho_1(y) = \rho_2(y) = y^2$ y $b_1 = b_2 = p$; y el estimador MVE haciendo $\rho_1(y) = \rho_2(y) = I_{[0,1]}^{(y)}$ con $b_1 = b_2 = (1 - \lfloor (n+p+1)/2 \rfloor)/n$. Para combinar las buenas propiedades de robustez del estimador MVE con el buen comportamiento asintótico de los estimadores de máxima verosimilitud, se eligen las funciones ρ_i de modo que sean “algo intermedio” entre ambos casos.

Lopuhaã consideró las siguientes condiciones para las funciones ρ_i ($i = 1, 2$):

(R1) $\rho_i(0) = 0$, ρ_i simétrica y dos veces continuamente diferenciable. Denotaremos con ψ_i a la derivada de ρ_i .

(R2) Existen constantes c_i tales que $0 < c_i < \infty$ y ρ_i es estrictamente creciente en $[0, c_i]$ y constante en $[c_i, \infty)$. Sea $a_i = \rho_i(c_i)$.

(R3) $2\rho_2(y) - y\psi_2(y) > 0$ para $y > 0$.

Las condiciones (R1) y (R2) pedidas sobre las funciones ρ_i son análogas a las condiciones (H1) y (H2) sobre ρ de los S-estimadores. La condición (R3) junto con la condición en (R2) de que $a_i = \sup \rho_i < \infty$ asegura que el τ -estimador tenga alto punto de ruptura.

Una posible elección para las funciones ρ_i sería considerar la función bicuadrática (36). Hennig [14] (1995) compara distintas elecciones para las funciones ρ_i en el caso de τ -estimadores de regresión y recomienda usar las siguientes funciones:

$$\begin{aligned}\rho_1(y) &= h_1(y^2); \\ \rho_2(y) &= k^2 h_2\left(\frac{y^2}{k}\right),\end{aligned}\tag{40}$$

donde k es una constante que se elige adecuadamente, y

$$h_1(y) = \left[-0.1 \left(\frac{y}{c}\right)^4 - 1.7 \left(\frac{y}{c}\right)^3 + 2.7 \left(\frac{y}{c}\right)^2 + 0.1 \left(\frac{y}{c}\right) \right] I_{(|y| \leq c)} + I_{(|y| > c)}$$

$$h_2(y) = \begin{cases} \frac{y}{c} & \text{si } \frac{y}{c} < 1 - \frac{t}{2} \\ 1 - \frac{t}{2} + tg \left(\frac{\frac{y}{c} - 1 + \frac{t}{2}}{t} \right) & \text{si } 1 - \frac{t}{2} \leq \frac{y}{c} \leq 1 + \frac{t}{2} \\ 1 & \text{si } 1 + \frac{t}{2} < \frac{y}{c} \end{cases}$$

siendo $t = 0.5$ y g una función de suavizado dada por

$$g(y) = \frac{y^4}{2} - y^3 + y.$$

En el estudio de Monte Carlo realizado en la Sección 7 se ha comparado la propuesta de Hennig con la función bicuadrática.

La constante b_1 cumple el mismo rol que b en (35) en el caso de los S-estimadores y determina el punto de ruptura. Para garantizar la existencia del τ -estimador se debe cumplir que

$$0 < b_1 < \rho_1(c_1).$$

La idea es elegir la constante c_1 de modo que t_n y C_n tengan alto punto de ruptura y sean consistentes. La alta eficiencia se obtiene variando la constante de ajuste c_2 . De esta manera se puede evitar el “compromiso” entre alto punto de ruptura y eficiencia asintótica hecho en el caso de los S-estimadores.

Lopuhaä [22] (1991: Teorema 5.1) prueba la siguiente propiedad sobre el punto de ruptura de los τ -estimadores.

Teorema 4 *Sea \mathcal{X}_n una muestra aleatoria de $n \geq p + 1$ puntos en \mathbb{R}^p que está en posición general.*

Si $b_1/a_1 \leq (n - p)/(2n)$, entonces los τ -estimadores (t_n, C_n) tienen punto de ruptura

$$\varepsilon_n^*(t_n, \mathbf{x}) = \varepsilon_n^*(C_n, \mathbf{x}) = \frac{1}{n} \left\lceil \frac{nb_1}{a_1} \right\rceil.$$

El valor óptimo del punto de ruptura, $\lfloor (n - p + 1)/2 \rfloor / n$ se obtiene eligiendo $b_1/a_1 = (n - p)/(2n)$.

Si se supone que la muestra tiene distribución elíptica, para obtener estimadores consistentes, las constantes b_i deben definirse de modo tal que:

$$b_i = E \left\{ \rho_i \left\{ \left[(\mathbf{x}_1 - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}) \right]^{1/2} \right\} \right\} = \int \rho_i(\|\mathbf{x}\|) f(\|\mathbf{x}\|) d\mathbf{x}. \quad (41)$$

Notemos que el punto de ruptura de los τ -estimadores dependen sólo de la constante b_1/a_1 , o sólo de la constante c_1 si b_1 está dado por (41); valores pequeños de c_1 se corresponden con un alto punto de ruptura y viceversa. Esto significa que puede variarse ρ_2 sin cambiar el valor del punto de ruptura.

Las condiciones de “suavidad” de las ρ_i son necesarias para obtener normalidad asintótica y función de influencia acotada.

La constante $b_2 > 0$ es una constante de normalización para obtener la consistencia de C_n al valor verdadero Σ . La alta eficiencia se consigue a través de una elección adecuada de la constante c_2 de la función ρ_2 . Esto permite controlar la eficiencia sin modificar el punto de ruptura.

Lopuhaä [22] (1991) demuestra que si $F_{\boldsymbol{\mu}, \Sigma}$ es una función de distribución elíptica tal que la densidad f es no creciente y existe un punto $c_3 \in [0, \min(c_1, c_2)]$ tal que f es estrictamente decreciente en $[0, c_3]$, entonces el problema \mathcal{P}_n^τ tiene una única solución con probabilidad 1, es decir el τ -estimador está bien definido.

4.5 Estimadores ponderados de un paso para la media y la matriz de covarianza

Supongamos que se tienen estimadores (\mathbf{t}_n, C_n) robustos pero no muy eficientes. Rousseeuw y Leroy [32] (1987) propusieron utilizar estos estimadores para definir pesos para cada una de las observaciones. Luego, estimadores más eficientes y con buenas propiedades de robustez se obtendrán tomando medias y covarianzas ponderadas utilizando estos pesos.

Más precisamente, dado un estimador robusto (\mathbf{t}_n, C_n) de $(\boldsymbol{\mu}, \Sigma)$ consideremos la distancia de Mahalanobis $d_i = d^2(\mathbf{x}_i, \mathbf{t}_n, C_n)$, luego el peso otorgado a la observación i se define por $w_i = W(d_i)$, donde W es una función no creciente. Se define (\mathbf{t}_n^*, C_n^*) como la media pesada y la matriz de covarianza muestral ponderada dadas por:

$$\mathbf{t}_n^* = \frac{\sum_{i=1}^n w_i \mathbf{x}_i}{\sum_{i=1}^n w_i} \quad (42)$$

$$C_n^* = \kappa \frac{\sum_{i=1}^n w_i (\mathbf{x}_i - \mathbf{t}_n^*) (\mathbf{x}_i - \mathbf{t}_n^*)'}{\sum_{i=1}^n w_i}, \quad (43)$$

donde κ es una constante apropiada para obtener consistencia bajo normalidad

Se pueden considerar distintas funciones de peso W . Maronna y Yohai [26] (1995) recomiendan usar la función “hard rejection”, dada por

$$W(d) = I_{(d \leq \tilde{c}_n)} \quad (44)$$

Esta función W hace que se le asigne peso 1 a las observaciones que caen dentro del elipsoide determinado mediante el estimador robusto inicial (\mathbf{t}_n, C_n) , que debería cubrir los datos con probabilidad β en el caso normal, y peso 0 en caso contrario.

El punto de corte \tilde{c}_n está dado por

$$\tilde{c}_n = c_{p,n} \text{med}(\mathbf{d}) \left(\frac{\chi_{p,\beta}^2}{\chi_{p,0.5}^2} \right), \quad (45)$$

donde la constante

$$c_{p,n} = \left(1 + \frac{15}{n-p} \right)^2$$

es una “corrección por muestra finita” sugerida por Rousseeuw y van Zomeren [33] (1990). En este caso, la constante κ en (43) valdrá 1.

El estimador (\mathbf{t}_n^*, C_n^*) así definido conservará las propiedades de robustez del estimador inicial y además será más eficiente que éste.

Lopuhaä y Rousseeuw [23] (1991: Teorema 5.1) demuestran lo siguiente:

Teorema 5 *Sea \mathcal{X}_n una muestra aleatoria de dimensión p y tamaño n en posición general tal que $n \geq p + 1$.*

Sean $(\mathbf{t}_n, C_n) \in \Xi$ estimadores equivariantes afines de posición multivariado y covarianza que satisfacen

$$\# \{i : d^2(\mathbf{x}_i, \mathbf{t}_n, C_n) \leq c_0^2\} \geq \left\lfloor \frac{n+p+1}{2} \right\rfloor.$$

Sea $W : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ una función de peso que satisface:

(W1) *W es no creciente y acotada.*

(W2) *$W(y) > 0$ para $y \in [0, c_0]$ y existe una constante $c_1 \geq c_0$ tal que $W(y) = 0$ para $y \in (c_1, \infty)$.*

Entonces, si (\mathbf{t}_n^, C_n^*) son los estimadores definidos en (42) y (43)*

$$\begin{aligned} \varepsilon_n^*(\mathbf{t}_n^*, \mathcal{X}_n) &\geq \min \{ \varepsilon_n^*(\mathbf{t}_n, \mathcal{X}_n), \varepsilon_n^*(C_n, \mathcal{X}_n) \} \\ \varepsilon_n^*(C_n^*, \mathcal{X}_n) &\geq \min \{ \varepsilon_n^*(\mathbf{t}_n, \mathcal{X}_n), \varepsilon_n^*(C_n, \mathcal{X}_n) \}. \end{aligned}$$

Para nuestro estudio de análisis factorial hemos considerado tres versiones de estos estimadores, en las cuales los pesos están basados en el estimador MVE, el S-estimador y el τ -estimador.

En todos los casos, en el estudio de Monte Carlo del Capítulo 7 se tomó el valor de $\beta = 0.975$, como se sugiere en Rousseeuw y Leroy [32] (1987).

5 Consistencia de los estimadores de los parámetros del modelo de análisis factorial

La consistencia de los estimadores de máxima verosimilitud de los parámetros del modelo de análisis factorial ha sido estudiada por diversos autores (ver por ej. Anderson y Rubin [5] (1956), Amemiya, Fuller y Pantula [1] (1987)).

Siguiendo las ideas de Swain [41] (1975), quien plantea un modelo general de estructura de covarianza, $\Sigma = \Sigma(\nu)$, donde $\Sigma \in \mathfrak{R}^{p \times p}$, $\nu \in \mathfrak{R}^q$ está contenido en algún espacio paramétrico real \mathcal{Q} , hemos demostrado la consistencia de los estimadores de los parámetros del modelo de análisis factorial obtenidos a partir de un estimador consistente de la matriz de covarianza.

Como Corolario de este resultado, como los estimadores robustos de la matriz de covarianza que hemos estudiado son consistentes, se obtiene la consistencia de los estimadores de los parámetros del modelo de análisis factorial obtenidos a partir de estos estimadores robustos.

Dada \mathcal{X}_n una muestra aleatoria en \mathfrak{R}^p que satisface el modelo (1) bajo las hipótesis (10), (11) y (14), los estimadores robustos $(\widehat{\Lambda}_n, \widehat{\Psi}_n)$ de los parámetros del modelo de análisis factorial que hemos propuesto se obtienen como:

$$(\widehat{\Lambda}_n, \widehat{\Psi}_n) = \arg \min_{(\Lambda, \Psi) \in \mathcal{B}(\Lambda, \Psi)} F_k(\Lambda, \Psi, \widetilde{\Sigma}_n),$$

donde

$$F_k(\Lambda, \Psi, \widetilde{\Sigma}_n) = \ln |\Lambda \Lambda' + \Psi| + \text{tr} \left[\widetilde{\Sigma}_n (\Lambda \Lambda' + \Psi)^{-1} \right] - \ln \left| \widetilde{\Sigma}_n \right|, \quad (46)$$

y

$$\mathcal{B}(\Lambda, \Psi) = \{(\Lambda, \Psi) \in \mathcal{R} \times \mathcal{D}_+(p) : \Lambda' \Psi^{-1} \Lambda \in \mathcal{D}^*(k)\}, \quad (47)$$

siendo $\Lambda \Lambda' + \Psi = \Sigma$ y $\widetilde{\Sigma}_n \in \text{PSD}(p)$ un estimador robusto de la matriz de covarianza muestral, por ejemplo, el τ -estimador.

El Teorema 6 muestra que si $\widetilde{\Sigma}_n$ es consistente y el modelo es identificable, $\widehat{\Lambda}_n$ y $\widehat{\Psi}_n$ también lo son. Antes de enunciarlo, veamos qué significa que el modelo de análisis factorial sea identificable.

Dada $\Sigma_0 \in \text{PSD}(p)$, si existen una única matriz diagonal $\Psi_0 \in \mathcal{D}_+(p)$ y una única matriz $\Lambda_0 \in \mathcal{R}$ de modo que se satisfaga la ecuación $\Sigma_0 = \Lambda_0 \Lambda_0' + \Psi_0$ sujeto a la restricción $\Lambda_0' \Psi_0^{-1} \Lambda_0 = \Delta$, donde $\Delta \in \mathcal{D}^*(k)$, diremos que el modelo es identificable.

Más formalmente daremos las siguientes definiciones:

Definición 14 Se dirá que el modelo de análisis factorial

$$\Sigma_0 = \Lambda_0 \Lambda_0' + \Psi_0 \quad (48)$$

es globalmente identificable en (Λ_0, Ψ_0) si no existe una matriz Ω diagonal definida no negativa, distinta de Ψ_0 , tal que $\Sigma_0 - \Omega$ es definida no negativa y $\text{rg}(\Sigma_0 - \Omega) \leq \text{rg}(\Sigma_0 - \Psi_0)$.

Definición 15 Se dirá que el modelo (48) es localmente identificable si tal matriz Ω no existe en un entorno de la matriz Ψ_0 .

El siguiente Teorema, cuya demostración está dada en la Sección C.4 del Apéndice, muestra que los estimadores de los parámetros del modelo de análisis factorial obtenidos a partir de un estimador robusto y consistente de la matriz de covarianza, son consistentes.

Teorema 6 Sea \mathcal{X}_n una muestra aleatoria en \mathbb{R}^p que satisface el modelo (1) bajo las hipótesis (10), (11) y (14).

Sea $\Sigma_0 = \Lambda_0 \Lambda_0' + \Psi_0$ y sea $\tilde{\Sigma}_n$ un estimador fuertemente consistente de Σ_0 . Supongamos que:

1. Σ_0 es definida positiva.
2. los parámetros del modelo son localmente identificables en un entorno de (Λ_0, Ψ_0) .

Sea

$$(\hat{\Lambda}_n, \hat{\Psi}_n) = \arg \min_{(\Lambda, \Psi) \in \mathcal{B}(\Lambda, \Psi)} F_k(\Lambda, \Psi, \tilde{\Sigma}_n)$$

donde F_k es la función dada en (46) y $\mathcal{B}(\Lambda, \Psi)$ es el conjunto dado en (47). Entonces:

$$\begin{aligned} \hat{\Psi}_n &\xrightarrow{\text{c.s.}} \Psi_0, \\ \hat{\Lambda}_n &\xrightarrow{\text{c.s.}} \Lambda_0. \end{aligned}$$

5.1 Consistencia de los estimadores de Donoho-Stahel para el modelo de análisis factorial

A partir de la consistencia del estimador de Donoho-Stahel de la matriz de covarianza demostrada por Maronna y Yohai [26] (1995: Teoremas 1 y 2), como Corolario del Teorema 6 se ha probado la consistencia de los estimadores de Donoho-Stahel de los parámetros del modelo de análisis factorial.

Notación: Sean $\mathbf{x}_1, \dots, \mathbf{x}_n$ vectores aleatorios en \mathbb{R}^p con distribución F y sea $\mathbf{a} \in \mathbb{R}^p$. Denotaremos a la distribución empírica de $\mathbf{x}'_1 \mathbf{a}, \dots, \mathbf{x}'_n \mathbf{a}$ con $F_{n\mathbf{a}}$ y a la distribución de $\mathbf{a}'\mathbf{x}$ bajo F con $F_{\mathbf{a}}$.

Definición 16 Sean μ y σ funcionales definidos para cualquier distribución. Definamos:

$$\begin{aligned} \mu_{\mathbf{a}} &= \mu(F_{\mathbf{a}}); & \sigma_{\mathbf{a}} &= \sigma(F_{\mathbf{a}}); \\ \mu_{n\mathbf{a}} &= \mu(F_{n\mathbf{a}}); & \sigma_{n\mathbf{a}} &= \sigma(F_{n\mathbf{a}}). \end{aligned}$$

Teorema 7 Supongamos que:

$$\sqrt{n} \sup_{\|\mathbf{a}\|=1} |\mu_{n\mathbf{a}} - \mu_{\mathbf{a}}| = O_p(1); \quad (49)$$

$$\sqrt{n} \sup_{\|\mathbf{a}\|=1} |\sigma_{n\mathbf{a}} - \sigma_{\mathbf{a}}| = O_p(1); \quad (50)$$

$$\sup_{\|\mathbf{a}\|=1} \mu_{\mathbf{a}} = c < \infty; \quad (51)$$

$$\inf_{\|\mathbf{a}\|=1} \sigma_{\mathbf{a}} > 0; \quad (52)$$

$$\sup_{\|\mathbf{a}\|=1} \sigma_{\mathbf{a}} = E < \infty, \quad (53)$$

y la función de peso w es tal que existen constantes γ y η tales que

$$|w(y) - w(y')| \leq \gamma |y - y'| \quad \forall y, y' \in \mathbb{R}$$

$$|w(y) - w(y')| \leq \frac{\eta |y - y'|}{\min(y, y')^3} \quad \forall y, y' \in \mathbb{R}$$

Entonces

$$\sqrt{n} \{[t(F_n), C(F_n)] - [t(F), C(F)]\} = O_p(1).$$

Teorema 8 Sean μ y σ la mediana y el MAD.

Supongamos que existen constantes $c > 0$ y $\delta > 0$ tales que

$$|F_{\mathbf{a}}(\mu_{\mathbf{a}} + \nu) - F_{\mathbf{a}}(\mu_{\mathbf{a}})| \geq c|\nu| \quad \forall \mathbf{a} \in S_p, \forall |\nu| \leq \delta$$

$$|F_{\mathbf{a}}(\mu_{\mathbf{a}} + \sigma_{\mathbf{a}} + \nu) - F_{\mathbf{a}}(\mu_{\mathbf{a}} + \sigma_{\mathbf{a}})| \geq c|\nu| \quad \forall \mathbf{a} \in S_p, \forall |\nu| \leq \delta$$

y que

$$P_F(\mathbf{x}'\mathbf{a} \neq 0) > 0.5 \quad \forall \mathbf{a} \neq 0$$

entonces μ y σ satisfacen (49) a (53) del Teorema 7.

Observación: Mediante el Teorema 8 se obtiene que (\mathbf{t}_n, C_n) es un estimador fuertemente consistente a $(\boldsymbol{\mu}, \Sigma)$ con velocidad de convergencia $n^{1/2}$.

Para estar bajo las condiciones del Teorema 6 y obtener la consistencia de los estimadores de Donoho-Stahel de los parámetros del modelo de análisis factorial, bastaba con tener la convergencia en casi todo punto del estimador de la covarianza. Para asegurarnos que el estimador converge al valor verdadero de la matriz de covarianza, y no a un múltiplo de la misma, consideraremos el valor de la constante κ de normalización dado por (33).

Corolario 5 *Bajo las hipótesis de los Teoremas 6 y 8 los estimadores de Donoho-Stahel $(\widehat{\Lambda}_n, \widehat{\Psi}_n)$ de los parámetros del modelo de análisis factorial son fuertemente consistentes, es decir*

$$\begin{aligned}\widehat{\Psi}_n &\xrightarrow{\text{c.s.}} \Psi_0; \\ \widehat{\Lambda}_n &\xrightarrow{\text{c.s.}} \Lambda_0.\end{aligned}$$

5.2 Consistencia de los S-estimadores para el modelo de análisis factorial

A partir de la consistencia del S-estimador de la matriz de covarianza demostrada por Davies [8] (1987: Teorema 3), como corolario del Teorema 6 se ha probado la consistencia de los S-estimadores de los parámetros del modelo de análisis factorial.

Definición 17 *Diremos que dos funciones f y g reales, no decrecientes, definidas en un intervalo común I no degenerado tienen un punto común de decrecimiento d , si:*

1. d es un punto interior de I .
2. $f(u) > f(d) > f(v)$ y $g(u) > g(d) > g(v) \quad \forall u, v \in I$ con $u < d < v$.

Teorema 9 *Sean \mathcal{X}_n una muestra aleatoria en \mathbb{R}^p con distribución elíptica F_{μ_0, Σ_0} con densidad $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$.*

Sea $\rho : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ una función no decreciente que satisface (H1), (H2) y además:

(H3) κ y f tienen al menos un punto común de decrecimiento d_0 , siendo $\kappa(x) = 1 - \rho(x^2)/a$.

Sea $\xi_n = (\mathbf{t}_n, C_n)$ una solución del problema \mathcal{P}_n definido en la Sección 4.3. Entonces,

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathbf{t}_n &= \boldsymbol{\mu}_0; \\ \lim_{n \rightarrow \infty} C_n &= \Sigma_0\end{aligned}$$

casi seguramente.

Observación: Mediante el Teorema 9 se obtiene que (\mathbf{t}_n, C_n) es un estimador fuertemente consistente a $(\boldsymbol{\mu}_0, \Sigma_0)$.

Como Corolario de los Teoremas 6 y 9, se obtiene la consistencia de los S-estimadores de los parámetros del modelo de análisis factorial.

Corolario 6 *Bajo las hipótesis de los Teoremas 6 y 9 los S-estimadores $(\widehat{\Lambda}_n, \widehat{\Psi}_n)$ de los parámetros del modelo de análisis factorial son fuertemente consistentes, es decir*

$$\begin{aligned}\widehat{\Psi}_n &\xrightarrow{\text{c.s.}} \Psi_0; \\ \widehat{\Lambda}_n &\xrightarrow{\text{c.s.}} \Lambda_0.\end{aligned}$$

5.3 Consistencia de los τ -estimadores para el modelo de análisis factorial

En forma análoga al caso del S-estimador, teniendo en cuenta los resultados obtenidos por Lopuhaä [22] (1991: Teorema 3.1 y Corolarios 3.1 y 3.2) se puede demostrar la consistencia de los τ -estimadores de los parámetros del modelo de análisis factorial.

Teorema 10 *Sean \mathcal{X}_n una muestra aleatoria en \mathfrak{R}^p con distribución elíptica $F_{\boldsymbol{\mu}_0, \Sigma_0}$ con densidad $f : \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$.*

Sean $\rho_i : \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$ funciones que satisfacen (R1), (R2), (R3) y b_i están dadas por (41) ($i = 1, 2$).

Supongamos además que

(R4) *f es no creciente y tiene al menos un punto de decrecimiento en $[0, \min(c_1, c_2)]$.*

Sea $\xi_n = (\mathbf{t}_n, C_n)$ una solución del problema \mathcal{P}_n^τ definido en la Sección 4.4. Entonces,

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathbf{t}_n &= \boldsymbol{\mu}_0; \\ \lim_{n \rightarrow \infty} C_n &= \Sigma_0\end{aligned}$$

casi seguramente.

Observación: Mediante el Teorema 10 se obtiene que (\mathbf{t}_n, V_n) es un estimador fuertemente consistente a $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$.

Como Corolario de los Teorema 6 y 10, se obtiene la consistencia de los τ -estimadores de los parámetros del modelo de análisis factorial.

Corolario 7 *Bajo las hipótesis de los Teoremas 6 y 10 los τ -estimadores $(\widehat{\Lambda}_n, \widehat{\Psi}_n)$ de los parámetros del modelo de análisis factorial son fuertemente consistentes, es decir*

$$\begin{aligned}\widehat{\Psi}_n &\xrightarrow{\text{c.s.}} \Psi_0; \\ \widehat{\Lambda}_n &\xrightarrow{\text{c.s.}} \Lambda_0.\end{aligned}$$

6 Distribución asintótica de los estimadores de los parámetros del modelo de análisis factorial

Bajo condiciones generales, Anderson y Rubin [5] (1956) demostraron la normalidad asintótica de los estimadores de máxima verosimilitud de los parámetros del modelo de análisis factorial. Lawley y Maxwell [18] (1971) encuentran expresiones para las matrices de covarianza asintóticas de $\widehat{\Lambda}_n$ y $\widehat{\Psi}_n$ para el caso de los estimadores de máxima verosimilitud bajo la hipótesis de normalidad de los factores y los errores. Amemiya, Fuller y Pantula [1] (1987), y posteriormente Anderson y Amemiya [4] (1988), bajo las hipótesis de normalidad de los errores y de que la matriz de derivadas parciales de Σ_0 con respecto a los parámetros estimados es de rango completo, obtienen la expresión de la matriz de covarianza asintótica tanto para el caso de factores fijos como aleatorios. Sin embargo, no encuentran la expresión explícita de dicha matriz de covarianza, sino que la expresan como función de la matriz de información.

En esta tesis, se ha obtenido la distribución asintótica de los estimadores de los parámetros del modelo de análisis factorial, Λ_0 y Ψ_0 , resultantes de reemplazar la matriz de covarianza muestral por un estimador consistente y asintóticamente normal de la covarianza en las ecuaciones de verosimilitud. Estos resultados, que demuestran la normalidad asintótica de los estimadores y dan una expresión explícita de las matrices de covarianza asintóticas, están expuestos en el Teorema 11.

Como Corolario de este Teorema, puede obtenerse la expresión explícita de la matriz de covarianza asintótica de los estimadores de máxima verosimilitud de los parámetros del modelo de análisis factorial. Además, aplicando los resultados de Davies [8] (1987) y Lopuhaä [20] (1989) sobre la distribución asintótica de los S-estimadores para posición multivariada y covarianza y de Lopuhaä [21] y [22] (1991) de los τ -estimadores para posición multivariada y covarianza, se obtiene la normalidad asintótica de los S- y τ -estimadores para el modelo de análisis factorial (Corolarios 8 y 9).

Veamos ahora los resultados que permiten obtener la normalidad asintótica de los estimadores robustos de los parámetros del modelo de análisis factorial. Previamente introduciremos alguna notación.

Notación: Sea $A \in \mathfrak{R}^{n \times m}$. Notaremos con $\text{vec}(A)$ al vector de dimensión nm cuyas componentes son las columnas de A listadas consecutivamente, es decir

$$\text{vec}(A)' = (a_{11}, \dots, a_{n1}, \dots, a_{1m}, \dots, a_{nm}).$$

Notación: Sea $A \in \mathfrak{R}^{n \times n}$ una matriz simétrica. Notaremos con $\text{vec}_r(A)$ al

vector de dimensión $n(n+1)/2$ cuyas componentes son los elementos del triángulo inferior de A incluyendo la diagonal (i.e. los a_{ij} tales que $i \geq j$) listados consecutivamente por columnas, es decir

$$\text{vec}_r(A)' = (a_{11}, \dots, a_{n1}, a_{22}, \dots, a_{n2}, \dots, a_{n-1,n-1}, a_{n-1,n}, a_{nn}).$$

Observación: Dada la matriz de covarianza Σ , que es simétrica, al considerar $\text{vec}_r(\Sigma)$ en vez de $\text{vec}(\Sigma)$ estamos trabajando sólo con los elementos no repetidos de Σ . De esta manera, se evitan los problemas que aparecerían al tomar perturbaciones no simétricas al calcular las derivadas de funciones de Λ y Ψ con respecto a Σ , necesarias en las demostraciones de los teoremas de normalidad asintótica de los estimadores, y que no serían consistentes con las hipótesis hechas sobre el modelo de análisis factorial.

Notación: (delta de Kronecker)

Sea

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{en otro caso} \end{cases}$$

Notación: Sean $E \in \mathfrak{R}^{p \times p}$, $D \in \mathfrak{R}^{p \times p(p+1)/2}$, $H \in \mathfrak{R}^{pk \times p}$, $L \in \mathfrak{R}^{pk \times p(p+1)/2}$ las matrices cuyos elementos están dados por:

$$E_{ij} = \frac{\psi_{ii}}{\psi_{jj}} \sum_{h=1}^k w_{ih} w_{jh} \left\{ -\theta_h w_{ih} w_{jh} + (\theta_h - 1) \left[\delta_{ij} + \sum_{r=1, r \neq h}^p \left(\frac{\theta_r + \theta_h}{\theta_r - \theta_h} \right) w_{ir} w_{jr} \right] \right\};$$

$$D_{i, \alpha + (\beta-1)(p-\beta/2)} = \frac{-2\psi_{ii}}{\sqrt{\psi_{\alpha\alpha}\psi_{\beta\beta}}} \sum_{h=1}^k w_{ih} \left\{ w_{ih} w_{\alpha h} w_{\beta h} + (\theta_h - 1) \left[\sum_{r=1, r \neq h}^p \frac{w_{ir}}{\theta_h - \theta_r} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) \right] \right\};$$

$$D_{i, \alpha + (\alpha-1)(p-\alpha/2)} = \delta_{i\alpha} - \frac{\psi_{ii}}{\psi_{\alpha\alpha}} \sum_{h=1}^k w_{ih} w_{\alpha h} \left\{ w_{ih} w_{\alpha h} + 2(\theta_h - 1) \left(\sum_{r=1, r \neq h}^p \frac{w_{\alpha r} w_{ir}}{\theta_h - \theta_r} \right) \right\};$$

$$H_{i+(h-1)p,j} = \frac{w_{jh}\sqrt{\psi_{ii}(\theta_h-1)}}{2\psi_{jj}} \left[\delta_{ij} + \sum_{r=1, r \neq h}^p \left(\frac{\theta_r + \theta_h}{\theta_r - \theta_h} \right) w_{ir}w_{jr} - \frac{\theta_h w_{ih}w_{jh}}{\theta_h - 1} \right];$$

$$L_{i+(h-1)p,\alpha+(\beta-1)(p-\beta/2)} = \sqrt{\frac{\psi_{ii}(\theta_h-1)}{\psi_{\alpha\alpha}\psi_{\beta\beta}}} \left[w_{\beta h} \sum_{r=1, r \neq h}^p \left(\frac{w_{\alpha r}w_{ir}}{\theta_h - \theta_r} \right) + w_{\alpha h} \sum_{r=1, r \neq h}^p \left(\frac{w_{\beta r}w_{ir}}{\theta_h - \theta_r} \right) + \frac{w_{\alpha h}w_{ih}w_{\beta h}}{\theta_h - 1} \right];$$

$$L_{i+(h-1)p,\alpha+(\alpha-1)(p-\alpha/2)} = \frac{w_{\alpha h}\sqrt{\psi_{ii}(\theta_h-1)}}{\psi_{\alpha\alpha}} \left[\sum_{r=1, r \neq h}^p \frac{w_{\alpha r}w_{ir}}{\theta_h - \theta_r} + \frac{w_{\alpha h}w_{ih}}{2(\theta_h - 1)} \right],$$

para $1 \leq i \leq j \leq p$, $1 \leq h \leq k$, $1 \leq \alpha \leq p$, $1 \leq \beta \leq p$, $\alpha \neq \beta$, donde $\theta_1 \geq \dots \geq \theta_p$ son los autovalores de $\Sigma^* = \Psi^{-1/2}\Sigma\Psi^{-1/2}$ y $\mathbf{w}_1, \dots, \mathbf{w}_p$ sus correspondientes autovectores normalizados.

El siguiente teorema, cuya demostración está dada en la Sección C.5 del Apéndice, permite obtener la distribución asintótica de los estimadores de los parámetros del modelo de análisis factorial.

Teorema 11 *Sea $\mathcal{X}_n = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ una muestra aleatoria en \mathbb{R}^p que satisface el modelo (1) bajo las hipótesis (10), (11) y (14) de modo que el modelo sea identificable, es decir, para $1 \leq i \leq n$*

$$\mathbf{x}_i = \boldsymbol{\mu}_0 + \Lambda_0 \mathbf{f}_i + \boldsymbol{\varepsilon}_i,$$

con

$$(\mathbf{f}_i, \boldsymbol{\varepsilon}_i) \sim F_{0,V}, \quad (54)$$

siendo $V = \begin{bmatrix} I_k & 0 \\ 0 & \Psi_0 \end{bmatrix}$ y además

$$\Lambda_0' \Psi_0^{-1} \Lambda_0 = \Delta,$$

donde $\mu_0 \in \mathfrak{R}^p$, $\Lambda_0 \in \mathcal{R}$, $\Psi_0 \in \mathcal{D}_+(p)$ y $\Delta \in \mathcal{D}^*(k)$.

Sea $\tilde{\Sigma}_n$ un estimador fuertemente consistente de $\Sigma_0 = \Sigma(\Lambda_0, \Psi_0) = \Lambda_0 \Lambda_0' + \Psi_0$ tal que

$$\sqrt{n} \text{vec}_r \left(\tilde{\Sigma}_n - \Sigma_0 \right) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, N). \quad (55)$$

Sean

$$\left(\hat{\Psi}_n, \hat{\Lambda}_n \right) = \arg \min_{(\hat{\Lambda}, \hat{\Psi}) \in \mathcal{B}(\Psi, \Lambda)} F_k \left(\Psi, \Lambda, \tilde{\Sigma}_n \right),$$

donde

$$F_k \left(\Psi, \Lambda, \tilde{\Sigma}_n \right) = \ln |\Sigma| + \text{tr} \left(\tilde{\Sigma}_n \Sigma^{-1} \right) - \ln \left| \tilde{\Sigma}_n \right|,$$

y

$$\mathcal{B}(\Psi, \Lambda) = \{ (\Psi, \Lambda) \in \mathcal{D}_+(p) \times \mathcal{R} : \Lambda' \Psi^{-1} \Lambda \in \mathcal{D}^*(k) \},$$

siendo $\Sigma = \Lambda \Lambda' + \Psi$.

Entonces, si $\hat{\nu}_n = \left(\hat{\Psi}_n, \hat{\Lambda}_n \right)$ y $\nu_0 = (\Psi_0, \Lambda_0)$

$$\sqrt{n} (\hat{\nu}_n - \nu_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, C).$$

siendo C la matriz tal que

$$C = \begin{bmatrix} C_1 & C_2 \\ C_2' & C_3 \end{bmatrix},$$

donde

$$\begin{aligned} C_1 &= (I + E)^{-1} D N D' (I + E)^{-1}; \\ C_2 &= (I + E)^{-1} D N [H(I + E)^{-1} + L]'; \\ C_3 &= [H(I + E)^{-1} + L] N [H(I + E)^{-1} + L]', \end{aligned}$$

y $E \in \mathfrak{R}^{p \times p}$, $D \in \mathfrak{R}^{p \times p(p+1)/2}$, $H \in \mathfrak{R}^{pk \times p}$, $L \in \mathfrak{R}^{pk \times p(p+1)/2}$ las matrices dadas anteriormente.

6.1 Distribución asintótica de los S-estimadores para el modelo de análisis factorial

El siguiente resultado, obtenido por Lopuhaä [20] (1989: Corolario 5.1), prueba la normalidad asintótica de los S-estimadores de posición y covarianza para modelos multivariados generales. Lo usaremos para probar la normalidad asintótica de los parámetros del modelo de análisis factorial basados en el S-estimador.

Definición 18 Sean $A \in \mathfrak{R}^{n \times m}$, $B \in \mathfrak{R}^{p \times q}$ matrices tales que $A_{ij} = a_{ij}$. El producto de Kronecker de estas matrices $A \otimes B \in \mathfrak{R}^{np \times mq}$ es la matriz en bloques tal que el bloque (i,j) será $\{a_{ij}B\}$.

Notación: $K_{nn} \in \mathfrak{R}^{n^2 \times n^2}$ es la matriz en bloques tal que el bloque (i,j) será Δ_{ji} , donde $\Delta_{ji} \in \mathfrak{R}^{n \times n}$ y $(\Delta_{ji})_{hk} = \delta_{jh}\delta_{ik}$.

Para obtener la normalidad asintótica de los S-estimadores de posición multivariada y covarianza, Lopuhaä [20] (1989) hace la siguiente hipótesis adicional sobre la función ρ :

(H4) Sea $u(y) = \psi(y)/y$. Si ψ' es la derivada segunda de ρ , ψ' y u son funciones acotadas y continuas.

Teorema 12 Sea \mathcal{X}_n una muestra aleatoria en \mathfrak{R}^p con distribución elíptica F_{μ_0, Σ_0} .

Sea $\rho: \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$ una función no decreciente que satisface (H1), (H2), (H4) y b está dada por (37).

Supongamos además que:

$$\begin{aligned} \mathbf{E} [\psi'(\|\mathbf{x}_1\|)] &> 0, \\ \mathbf{E} [\psi'(\|\mathbf{x}_1\|)\|\mathbf{x}_1\|^2 + (p+1)\psi(\|\mathbf{x}_1\|)\|\mathbf{x}_1\|] &> 0. \end{aligned}$$

Entonces, si $\xi_n = (\mathbf{t}_n, C_n)$ es el S-estimador basado en la función ρ , se tiene que $\sqrt{n}(\mathbf{t}_n - \mu_0)$ y $\sqrt{n}(C_n - \Sigma_0)$ son asintóticamente normales e independientes con media $\mathbf{0}$.

La matriz de covarianza asintótica de $\sqrt{n}(\mathbf{t}_n - \mu_0)$ está dada por

$$(\alpha/\beta^2)\Sigma_0,$$

donde

$$\begin{aligned} \alpha &= \frac{1}{p}\mathbf{E} [\psi^2(\|\mathbf{x}_1\|)]; \\ \beta &= \mathbf{E} \left[\left(1 - \frac{1}{p}\right)u(\|\mathbf{x}_1\|) + \frac{1}{p}\psi'(\|\mathbf{x}_1\|) \right]. \end{aligned}$$

La matriz de covarianza asintótica de $\sqrt{n}(C_n - \Sigma)$ está dada por

$$M^S = \sigma_1(I + K_{pp})(\Sigma_0 \otimes \Sigma_0) + \sigma_2 \text{vec}(\Sigma_0) \text{vec}(\Sigma_0)',$$

donde

$$\sigma_1 = \frac{p(p+2)\mathbf{E}[\psi^2(\|\mathbf{x}_1\|)\|\mathbf{x}_1\|^2]}{\left\{\mathbf{E}\left[\psi'(\|\mathbf{x}_1\|)\|\mathbf{x}_1\|^2 + (p+1)\psi(\|\mathbf{x}_1\|)\|\mathbf{x}_1\|\right]\right\}^2};$$

$$\sigma_2 = -\frac{2}{p}\sigma_1 + \frac{4\mathbf{E}\left\{[\rho(\|\mathbf{x}_1\|) - b]^2\right\}}{\left\{\mathbf{E}[\psi(\|\mathbf{x}_1\|)\|\mathbf{x}_1\|]\right\}^2}.$$

Observación: Dada la distribución asintótica de C_n , el S-estimador de la matriz de covarianza, queremos encontrar la distribución asintótica del vector formado por los elementos no repetidos de dicha matriz, el cual está dado por $\text{vec}_r(C_n)$.

Sea $M^S \in \mathfrak{R}^{p^2 \times p^2}$ la matriz de covarianza asintótica de $\sqrt{n}(C_n - \Sigma_0)$ dada en el Teorema 12 ($(M^S)_{ij} = m_{ij}$).

Sea $N^S \in \mathfrak{R}^{p(p+1)/2 \times p(p+1)/2}$ la matriz dada por

$$N^S = \begin{bmatrix} N^{11} & N^{1p} \\ \vdots & \vdots \\ N^{p1} & N^{pp} \end{bmatrix}$$

donde el bloque $N^{ij} \in \mathfrak{R}^{(p-i+1) \times (p-j+1)}$, $1 \leq i \leq p$, $1 \leq j \leq p$ está dado por:

$$N^{ij} = \begin{bmatrix} m_{(i-1)p+i, (j-1)p+j} & m_{(i-1)p+i, jp} \\ \vdots & \vdots \\ m_{ip, (j-1)p+j} & m_{ip, jp} \end{bmatrix}.$$

Entonces

$$\sqrt{n} \text{vec}_r(C_n - \Sigma_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, N^S).$$

De esta manera, como por el Teorema 12 se obtiene que C_n es asintóticamente normal, entonces aplicando el Teorema 11 se obtiene la normalidad asintótica de los S-estimadores de parámetros del modelo de análisis factorial y la expresión de la matriz de covarianza asintótica para dichos estimadores.

Corolario 8 *Bajo las hipótesis de los Teoremas 11 y 12, si $(\widehat{\Psi}_n, \widehat{\Lambda}_n)$ son los S-estimadores de los parámetros del modelo de análisis factorial se cumple que*

$$\sqrt{n}(\widehat{\Psi}_n - \Psi_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, (I + E)^{-1}DN^SD'(I + E)^{-1});$$

$$\sqrt{n} (\widehat{\Lambda}_n - \Lambda_0) \xrightarrow{D} \mathcal{N} \left(\mathbf{0}, [H(I + E)^{-1} + L] N^S [H(I + E)^{-1} + L]' \right).$$

siendo N^S la matriz de covarianza asintótica de $\sqrt{n} \text{vec}_\tau (C_n - \Sigma_0)$ dada en la observación anterior y E , D , H y L las matrices dadas en el Teorema 11.

6.2 Distribución asintótica de los τ -estimadores para el modelo de análisis factorial

Siguiendo un razonamiento análogo al caso de los S-estimadores, se puede demostrar la normalidad asintótica de los τ -estimadores de los parámetros del modelo de análisis factorial. Veremos que la matriz de covarianza asintótica de estos estimadores es similar a la obtenida en el Corolario 8, excepto por las constantes α_0 y β_0 .

Para esto consideremos previamente los siguientes resultados obtenidos por Lopuhaä [22] (1991) para τ -estimadores de posición y covarianza para modelos multivariados generales. Análogamente al caso de los S-estimadores, haremos la siguiente hipótesis adicional sobre las funciones ρ_i :

(R5) Sean $u_i(y) = \psi_i(y)/y$ ($i = 1, 2$). Si ψ'_i es la derivada segunda de ρ_i , ψ'_i y u_i son funciones acotadas y continuas.

Teorema 13 Sea \mathcal{X}_n una muestra aleatoria en \mathbb{R}^p con distribución elíptica F_{μ_0, Σ_0} .

Sean $\rho_i : \mathbb{R} \rightarrow \mathbb{R}_+$ funciones que satisfacen (R1), (R2), (R3), (R5) y b_i están dada por (41) ($i = 1, 2$).

Supongamos además que:

$$\begin{aligned} \mathbf{E} [\psi'_i (\|\mathbf{x}_1\|)] &> 0; \\ \mathbf{E} [\psi'_i (\|\mathbf{x}_1\|) \|\mathbf{x}_1\|^2 + (p+1)\psi_i (\|\mathbf{x}_1\|) \|\mathbf{x}_1\|] &> 0. \end{aligned}$$

Entonces, si $\xi_n = (\mathbf{t}_n, C_n)$ es el τ -estimador basado en las funciones ρ_i , se tiene que $\sqrt{n}(\mathbf{t}_n - \mu_0)$ y $\sqrt{n}(C_n - \Sigma_0)$ son asintóticamente normales e independientes con media 0.

La matriz de covarianza asintótica de $\sqrt{n}(\mathbf{t}_n - \mu_0)$ está dada por

$$(\alpha_0 / \beta_0^2) \Sigma_0,$$

donde

$$\begin{aligned}\alpha_0 &= \frac{1}{p} \mathbf{E} [\psi_0^2 (\|\mathbf{x}_1\|)]; \\ \beta_0 &= \mathbf{E} \left[\left(1 - \frac{1}{p}\right) u_0 (\|\mathbf{x}_1\|) + \frac{1}{p} \psi_0' (\|\mathbf{x}_1\|) \right].\end{aligned}$$

La matriz de covarianza asintótica de $\sqrt{n}(C_n - \Sigma_0)$ está dada por

$$M^T = \sigma_{01} (I + K_{pp}) (\Sigma_0 \otimes \Sigma_0) + \sigma_{02} \text{vec}(\Sigma_0) \text{vec}(\Sigma_0)',$$

donde

$$\begin{aligned}\sigma_{01} &= \frac{p(p+2) \mathbf{E} [\psi_0^2 (\|\mathbf{x}_1\|) \|\mathbf{x}_1\|^2]}{\left\{ \mathbf{E} \left[\psi_0' (\|\mathbf{x}_1\|) \|\mathbf{x}_1\|^2 + (p+1) \psi_0 (\|\mathbf{x}_1\|) \|\mathbf{x}_1\| \right] \right\}^2}; \\ \sigma_{02} &= -\frac{2}{p} \sigma_{01} + \frac{4 \mathbf{E} \left\{ [\rho_0 (\|\mathbf{x}_1\|) - b_0]^2 \right\}}{\left\{ \mathbf{E} [\psi_0 (\|\mathbf{x}_1\|) \|\mathbf{x}_1\|] \right\}^2},\end{aligned}$$

siendo

$$\begin{aligned}\rho_0(y) &= (2b_2 - \omega_2) \rho_1(y) + \omega_1 \rho_2(y); \\ \psi_0(y) &= (2b_2 - \omega_2) \psi_1(y) + \omega_1 \psi_2(y); \\ b_0 &= \mathbf{E} [\rho_0 (\|\mathbf{x}_1\|)]; \\ \omega_i &\stackrel{\text{def}}{=} \mathbf{E} [\psi_i (\|\mathbf{x}_1\|) \|\mathbf{x}_1\|] \quad \text{para } i = 0, 1, 2.\end{aligned}$$

Observación: En forma análoga a lo hecho para los S-estimadores, dada la distribución asintótica de C_n , el S-estimador de la matriz de covarianza, queremos encontrar la distribución asintótica del vector formado por los elementos no repetidos de dicha matriz, el cual está dado por $\text{vec}_r(C_n)$.

Sea ahora $M^T \in \mathfrak{R}^{p^2 \times p^2}$ la matriz de covarianza asintótica de $\sqrt{n}(C_n - \Sigma_0)$ dada en el Teorema 13 ($(M^T)_{ij} = m_{ij}$).

Sea $N^T \in \mathfrak{R}^{p(p+1)/2 \times p(p+1)/2}$ la matriz dada por

$$N^T = \begin{bmatrix} N^{11} & N^{1p} \\ \vdots & \vdots \\ N^{p1} & N^{pp} \end{bmatrix},$$

donde el bloque $N^{ij} \in \mathfrak{R}^{(p-i+1) \times (p-j+1)}$, $1 \leq i \leq p$, $1 \leq j \leq p$ está dado por:

$$N^{ij} = \begin{bmatrix} m_{(i-1)p+i, (j-1)p+i} & m_{(i-1)p+i, jp} \\ & \vdots \\ m_{ip, (j-1)p+i} & m_{ip, jp} \end{bmatrix}.$$

Entonces

$$\sqrt{n} \operatorname{vec}_\tau (C_n - \Sigma_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, N^T).$$

De esta manera, como por el Teorema 13 se obtiene que C_n es asintóticamente normal y su matriz de covarianza asintótica es M^T , entonces aplicando el Teorema 11 se obtiene la normalidad asintótica de los τ -estimadores de parámetros del modelo de análisis factorial y la expresión de la matriz de covarianza asintótica para dichos estimadores.

Corolario 9 *Bajo las hipótesis de los Teoremas 11 y 13, si $(\widehat{\Psi}_n, \widehat{\Lambda}_n)$ son los τ -estimadores de los parámetros del modelo de análisis factorial se cumple que*

$$\begin{aligned} \sqrt{n} (\widehat{\Psi}_n - \Psi_0) &\xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, (I + E)^{-1} D N^T D' (I + E)^{-1}); \\ \sqrt{n} (\widehat{\Lambda}_n - \Lambda_0) &\xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, [H(I + E)^{-1} + L] N^T [H(I + E)^{-1} + L]'), \end{aligned}$$

siendo N^T la matriz de covarianza asintótica de $\sqrt{n} \operatorname{vec}_\tau (C_n - \Sigma_0)$ dada en la observación anterior y E , D , H y L las matrices dadas en el Teorema 11.

7 Simulaciones

Dado el modelo

$$\mathbf{x} = \boldsymbol{\mu}_0 + \Lambda_0 \mathbf{f} + \boldsymbol{\varepsilon}, \quad (56)$$

donde $\mathbf{x} \in \mathbb{R}^p$, $\boldsymbol{\mu}_0 \in \mathbb{R}^p$, $\Lambda_0 \in \mathcal{R}$, $\mathbf{f} \in \mathbb{R}^k$ y $\boldsymbol{\varepsilon} \in \mathbb{R}^p$ de modo que se cumplen las hipótesis (10), (11) y (14), y por lo tanto

$$\Sigma_0 = \text{cov}(\mathbf{x}) = \Lambda_0 \Lambda_0' + \Psi_0,$$

$$\Lambda_0' \Psi_0^{-1} \Lambda_0 = \Delta,$$

siendo $\Psi_0 \in \mathcal{D}_+(p)$ y $\Delta \in \mathcal{D}^*(k)$, se realizaron simulaciones para comparar el comportamiento de los distintos estimadores robustos propuestos de los parámetros del modelo de análisis factorial, Λ_0 y Ψ_0 , con los estimadores de máxima verosimilitud.

Para esto se estimó la matriz de covarianza de las observaciones Σ_0 mediante el estimador de máxima verosimilitud, el estimador de Donoho-Stahel, el S-estimador, el τ -estimador y los estimadores ponderados de un paso basados en el estimador MVE, y en el S- y τ -estimador. Posteriormente, se estimaron los parámetros que conforman la estructura de Σ_0 mediante el algoritmo propuesto en la Sección D.1 del Apéndice.

Definición 19 Sea $A \in \mathbb{R}^{n \times n}$ una matriz simétrica con autovalores $\lambda_1 \geq \dots \geq \lambda_n$. Definiremos el número de condición de A como

$$\text{cond}(A) = \frac{\lambda_1}{\lambda_n}.$$

Como una medida del error que se cometía al realizar la estimación se calculó el sesgo con respecto a Σ_0 , dado por

$$\text{sesgo}(\Sigma_0, \widehat{\Sigma}_n) = \ln [\text{cond}(A_n \Sigma_0 A_n')],$$

donde $\widehat{\Sigma}_n = \widehat{\Lambda}_n \widehat{\Lambda}_n' + \widehat{\Psi}_n$, $\widehat{\Sigma}_n^{-1} = A_n' A_n$. Esta es una medida de error en la estimación de la “forma” de la matriz Σ_0 , es decir $\text{sesgo}(\Sigma_0, \widehat{\Sigma}_n) = 0$ si y sólo si $\widehat{\Sigma}_n = \lambda \Sigma_0$.

Para comparar el comportamiento de los distintos estimadores con respecto al sesgo, se calcularon el promedio de sesgo y la mediana de sesgo; que se denotarán respectivamente por “pro” y “me”; esta última medida es menos sensible a unas pocas muestras con comportamiento atípico.

En las Tablas y Gráficos siguientes, denotaremos los resultados al calcular $\widehat{\Sigma}_n$ mediante máxima verosimilitud con MV, para los estimadores de Donoho-Stahel usaremos DS. Para los S-estimadores usando la función ρ_c de Tukey dada en (36), usaremos la notación S; para los τ -estimadores basados en las funciones sugeridas por Hennig dadas en (40) tau H, y para los basados en la función bicuadrática de Tukey dada en (36), tau B. Para los estimadores ponderados de un paso, usaremos las notaciones WR, WS, WTh y WTb para los basados en el estimador MVE de Rousseeuw, el S-estimador y los τ -estimadores basados en las funciones de Hennig y bicuadrática respectivamente.

Para el modelo (56) se eligieron como valores de los parámetros $p = 5$, $k = 2$,

$$\mu_0 = [0, 0, 0, 0, 0]';$$

$$\Psi_0 = I_5,$$

y se generó aleatoriamente una matriz de 5×2 , la cual se transformó de modo tal que se cumpliera la condición de diagonalidad (10). La matriz resultante fue

$$\Lambda_0 = \begin{bmatrix} -0.8696 & -0.3933 \\ -0.9293 & -0.5526 \\ -0.3634 & 0.9001 \\ -1.1079 & 0.3965 \\ 0.7941 & -0.1123 \end{bmatrix},$$

de modo tal que

$$\Lambda_0' \Psi_0^{-1} \Lambda_0 = \begin{bmatrix} 3.61 & 0 \\ 0 & 1.44 \end{bmatrix}.$$

Posteriormente se hicieron 100 replicaciones de muestras de tamaño $n = 100$ donde $\mathbf{f} \sim \mathcal{N}_2(0, \mathbf{I})$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_5(0, \Psi_0)$ independientes. En este caso, en el que se supone la normalidad de los errores y los factores, se calculó la eficiencia esférica para comparar los métodos robustos con respecto al de máxima verosimilitud, que es el óptimo en el sentido de mínimo sesgo para este caso. Esta medida está dada por el cociente de las medianas de los sesgos obtenidos empleando el método MV con el obtenido mediante el método robusto y la denotaremos por "efic. esfer."

Para el método de Donoho-Stahel se consideraron distintos valores de c , siendo c el punto donde los pesos comienzan a decrecer. Por resultados obtenidos por Maronna y Yohai [26] (1995), $c = 1.06$ es la constante asintóticamente óptima en el sentido de mínimo sesgo, en el caso de utilizar la

función de peso de Huber para $p = 5$ y una contaminación del 10% en la muestra. También se consideraron algunos otros valores de c con el fin de aumentar la eficiencia esférica.

Los resultados se muestran en la Tabla 1

Tabla 1: errores y factores normales

estimador	me	pro	efic. esfer.
MV	0,6071	0,6273	1,0000
DS(1.06)	0,7447	0,7384	0,8152
DS(2)	0,7085	0,7076	0,8569
DS(2.5)	0,6752	0,6778	0,8991
DS(3)	0,6335	0,8003	0,9583
DS(4)	0,6181	0,6350	0,9822

En las simulaciones posteriores, la constante del estimador de Donoho-Stahel se fijó en el valor $c = 3$ para obtener un compromiso entre minimizar el sesgo y aumentar la eficiencia.

Para el S-estimador, la constante c de la función ρ_c se fijó en $c = 4.83$ de modo de obtener punto de ruptura maximal de acuerdo a los resultados expuestos en la Sección 4.3. Se calculó la eficiencia teórica de este estimador bajo el modelo de observaciones normales obteniéndose un valor del 86.75%.

Las constantes del τ -estimador se fijaron en los valores de $c_1 = 9.84$ y $c_2 = 13.33$, en el caso de usar las funciones de Hennig, y de $c_1 = 4.83$ y $c_2 = 6.51$, en el caso de usar la función bicuadrática, con el fin de obtener punto de ruptura maximal y eficiencia del 95%.

En el caso de los estimadores ponderados de un paso para la media y la matriz de covarianza, como hemos visto en la Sección 4.5, la constante de corte para asignar peso 1 a una observación es $\chi_{p,\alpha}^2$, en las simulaciones se fijó el valor de α en $\alpha = 0.975$.

En la Tabla 2 se observa, como era de esperar que el método de máxima verosimilitud (MV) produce los menores valores de las medianas de los sesgos; obteniéndose también valores pequeños y alta eficiencia esférica para los estimadores de Donoho-Stahel y los τ -estimadores. De esta Tabla también se deduce que los estimadores ponderados son más eficientes que su versión sin ponderar. Como hemos visto en la Sección 4.4, se podría aumentar la eficiencia de los τ -estimadores modificando adecuadamente la constante c_2 .

Debido a la estructura particular del modelo de análisis factorial, en el que hay dos elementos aleatorios, los factores y los errores, se podrían considerar tres tipos de contaminación:

Tabla 2: errores y factores normales

estimador	me	pro	efic. esfer.
MV	0,6071	0,6273	1,0000
WR	0,7701	0,6273	0,7883
DS	0,6335	0,8003	0,9583
S	0,7073	0,7050	0,8583
WS	0,6714	0,6603	0,9042
Tau H	0,6752	0,6655	0,8991
WT H	0,6714	0,6597	0,9042
Tau B	0,6502	0,6508	0,9337
WT B	0,6541	0,6531	0,9281

1. sólo en los factores:

$n - m$ observaciones son tales que

$$\mathbf{x} = \boldsymbol{\mu}_0 + \Lambda_0 \mathbf{f} + \boldsymbol{\varepsilon},$$

siendo $\mathbf{f} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$

m observaciones son tales que

$$\mathbf{x} = \boldsymbol{\mu}_0 + k_1 \Lambda_0 \mathbf{b} + \boldsymbol{\varepsilon},$$

siendo $\mathbf{b}' = (1, 0, \dots, 0)$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$ y $m = \lfloor n\epsilon \rfloor$ es el porcentaje de contaminación.

2. sólo en los errores

$n - m$ observaciones son tales que

$$\mathbf{x} = \boldsymbol{\mu}_0 + \Lambda_0 \mathbf{f} + \boldsymbol{\varepsilon},$$

siendo $\mathbf{f} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$

m observaciones son tales que

$$\mathbf{x} = \boldsymbol{\mu}_0 + \Lambda_0 \mathbf{f} + k_2 \mathbf{b} + \boldsymbol{\varepsilon},$$

siendo $\mathbf{b}' = (1, 0, \dots, 0)$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$ y $m = \lfloor n\epsilon \rfloor$ es el porcentaje de contaminación.

3. en los factores y los errores simultáneamente

$n - m$ observaciones son tales que

$$\mathbf{x} = \boldsymbol{\mu}_0 + \Lambda_0 \mathbf{f} + \boldsymbol{\varepsilon},$$

siendo $\mathbf{f} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$

m observaciones son tales que

$$\mathbf{x} = \boldsymbol{\mu}_0 + k_1 \Lambda_0 \mathbf{b}_1 + k_2 \mathbf{b}_2 + 0.01 \boldsymbol{\varepsilon},$$

siendo $\mathbf{b}'_1 = (1, 0, \dots, 0)$, $\mathbf{b}'_2 = (0, 1, 0, \dots, 0)$, $\boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$ y $m = \lfloor n\epsilon \rfloor$ es el porcentaje de contaminación (el término $0.01\boldsymbol{\varepsilon}$ se agrega para evitar los problemas de colinealidad).

Para comparar las distintas formas de contaminación se realizaron algunas simulaciones exploratorias en el caso de muestras normales con un 10% de contaminación, obteniéndose los peores resultados en el caso de contaminar los factores y errores simultáneamente. Por lo tanto, en lo siguiente estudiaremos el caso de muestras normales contaminadas de esta forma, a las cuales llamaremos $NC(\epsilon, k_1, k_2)$.

Para el modelo (56) con los mismos valores de los parámetros elegidos anteriormente, se hicieron 100 replicaciones de muestras de tamaño $n = 100$ para los casos $NC(0.1, k_1, k_2)$ y $NC(0.2, k_1, k_2)$, es decir se estudió el comportamiento de los estimadores para contaminaciones del 10 y 20%.

Se realizó una grilla para k_1 y k_2 con el fin de encontrar con mayor precisión para qué valores de estos parámetros se obtenían los máximos valores de las medianas de los sesgos para cada uno de los estimadores. Para las muestras $NC(0.1, k_1, k_2)$ los valores de k_1 se tomaron en el conjunto $\{1, 2, \dots, 16\}$ y los de k_2 en $\{1, 2, \dots, 20\}$; para las muestras $NC(0.2, k_1, k_2)$, k_1 en $\{1, 2, \dots, 24\}$ y k_2 en $\{1, 2, \dots, 29, 30, 32, \dots, 88, 90\}$ con el fin de encontrar a partir de qué valores de (k_1, k_2) los sesgos de cada uno de los estimadores robustos comenzaban a decrecer. Para cada par (k_1, k_2) se generaron las mismas muestras con el fin de realizar una mejor comparación de los resultados.

Para las muestras normales contaminadas se graficaron para los distintos valores de k_1 y k_2 las medianas de los sesgos con el fin de comparar el comportamiento de los distintos estimadores frente a datos atípicos.

Los máximos valores de las medianas de los sesgos están dadas en la Tabla 3

Observando los Gráficos, se ve que la “forma” de las curvas de las medianas de los sesgos no presenta diferencias al variar los valores de k_1 (que

Tabla 3: Máximos valores de las medianas de los sesgos para muestras contaminadas

estimador	NC(0.1, k_1, k_2)	NC(0.2, k_1, k_2)
MV	∞	∞
WR	2.1736	4.8008
DS	1.3115	2.6334
S	1.7254	5.1627
WS	1.7095	4.6949
Tau H	3.1007	6.4110
WT H	3.1027	6.6595
Tau B	2.3354	7.1640
WT B	2.6365	6.9963

determina la contaminación en los factores), salvo para los valores de $k_1 \geq 7$ para los S-estimadores y su versión ponderada y para $k_1 \geq 13$ para el resto de los estimadores robustos en el caso de una contaminación del 10%, y para $k_1 \geq 21$ en el caso de una contaminación del 20%.

También se observa que los estimadores MV se ven sensiblemente afectado por la presencia de observaciones atípicas, y a medida que crecen los valores de k_1 y k_2 , aumentan los valores de los sesgos. No se observan diferencias significativas entre los valores de los sesgos de los estimadores robustos y sus versiones ponderadas, aunque en el caso de una contaminación del 20%, las diferencias son mayores, en particular para los estimadores ponderados basados en los τ -estimadores usando las funciones de Hennig, siendo en este caso menores.

En el caso de una contaminación del 10%, se ve que de los métodos robustos estudiados, el estimador de Donoho-Stahel es el que produce los mejores resultados, en el sentido de que produce el menor valor del máximo de las medianas de los sesgos siendo $\max[\text{me}(\text{DS})] = 1.3115$. En los gráficos se observa que los valores de los sesgos para este estimador permanecen casi "constantes" al variar los valores de k_1 y k_2 .

Comparando el resto de los estimadores robustos para valores grandes de k_2 , se ve que para los S-estimadores y su versión ponderada se obtienen menores sesgos que para los demás, y que los sesgos de los τ -estimadores basados en la función bicuadrática son inferiores que los de los basados en las funciones de Hennig. Para valores pequeños de k_1 y k_2 los estimadores ponderados basados en el MVE producen sesgos mayores que el resto de los estimadores.

El siguiente paso fue estudiar el comportamiento de estos estimadores para el caso de muestras con una contaminación del 20%. Los resultados obtenidos por Maronna y Yohai [26] (1995) para el caso de estimadores de covarianza generales demuestran que al aumentar el porcentaje de contaminación del 10% al 20% los valores de las medianas de los sesgos aumentan considerablemente, aún para el caso de estimadores robustos como el MVE, el S-estimador, el τ -estimador y los estimadores ponderados de la matriz de covarianza; en cambio para el estimador de Donoho-Stahel este incremento no es tan significativo. De los resultados obtenidos en las simulaciones realizadas se deduce que ocurre lo mismo al obtener los estimadores de los parámetros del modelo de análisis factorial donde se impone una estructura a la matriz de covarianza.

Si comparamos los resultados obtenidos en la Tabla 1 de Maronna y Yohai [26] (1995) con los de los estimadores de Donoho-Stahel de los parámetros del modelo de análisis factorial, se observa que los sesgos son mayores en nuestras simulaciones. Esto se debe, teniendo en cuenta el algoritmo dado en la Sección D.1, a que en nuestro caso estamos estimando además la estructura en la covarianza impuesta por el modelo, lo cual aumenta el error que ya tenía la estimación de la matriz de covarianza sin tener en cuenta una estructura.

Observando la Tabla 3 y las Figuras se deduce que el estimador de Donoho-Stahel sigue produciendo los mejores resultados, en el sentido de menor valor de las medianas de los sesgos (como en el caso de una contaminación del 10%), ya que el menor valor del máximo de las medianas de los sesgos es de $\max[\text{me}(\text{DS})] = 2.6334$. Como se deduce de la Tabla 1, si se elige adecuadamente el valor de c , se puede mejorar la eficiencia bajo el modelo con observaciones normales. Además se advierte una diferencia significativa con respecto a los valores de las medianas de los sesgos obtenidos mediante los otros métodos robustos para valores pequeños de k_2 .

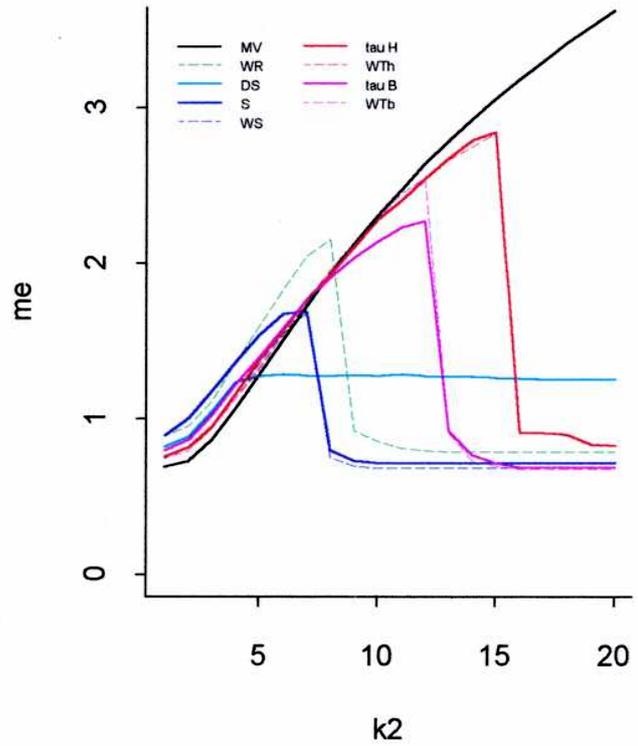
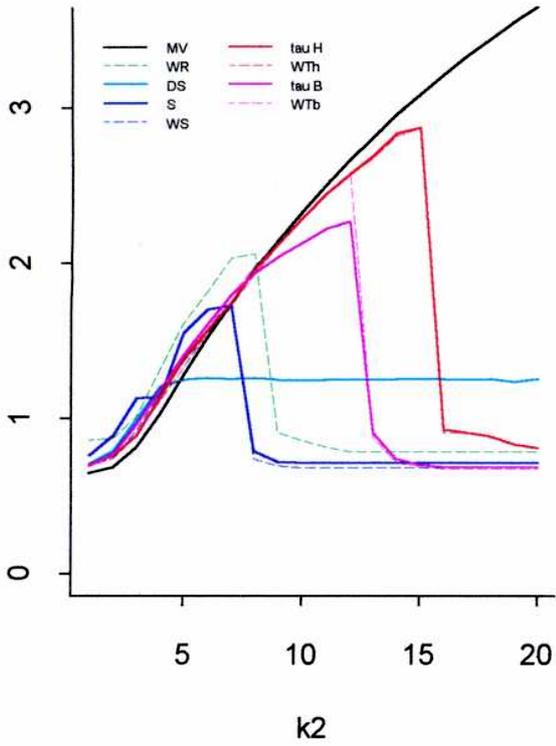
Con respecto a otras propuestas robustas, se observa que para los S-estimadores y los estimadores ponderados basados en el MVE y el S-estimador, las medianas de los sesgos comienzan a decrecer para valores de k_2 menores que para los demás estimadores (excepto Donoho-Stahel). A pesar de ser menos eficientes bajo el modelo normal que los τ -estimadores, los S-estimadores son más eficaces que éstos en la detección de outliers.

A diferencia del caso de una contaminación del 10%, los sesgos obtenidos τ -estimadores basados en las funciones de Hennig y su versión ponderada son menores que los obtenidos con los estimadores usando la función bicuadrática y comienzan a decrecer para valores menores de k_2 .

De las Figuras se deduce que el estimador de Donoho-Stahel es eficaz en la detección de outliers, aún ante la presencia de una contaminación del 20%, hecho que no ocurre con los demás estimadores robustos considerados. También se observa que para los estimadores de máxima verosimilitud (que son los que habitualmente se utilizan en los paquetes estadísticos) los sesgos aumentan a medida que aumentan los valores de la contaminación en los factores y en los errores, lo que evidencia la falta de robustez de este estimador.

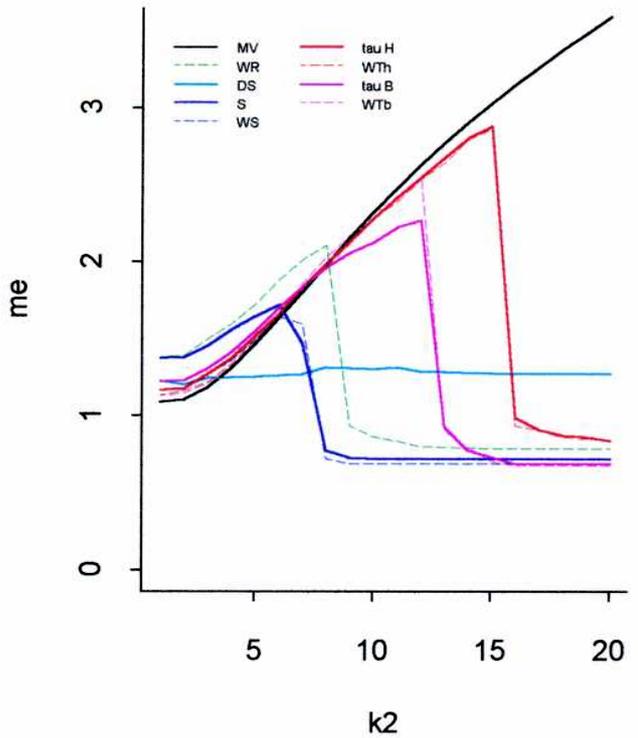
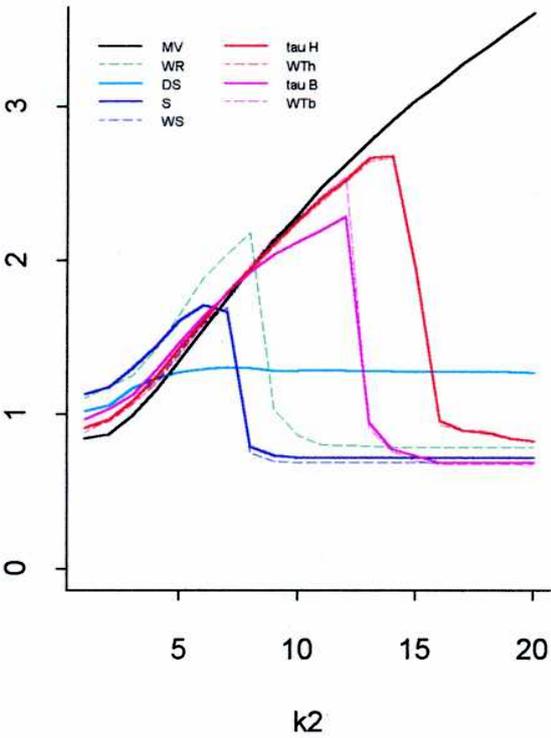
$k_1 = 1$ Contaminación 10%

$k_1 = 2$



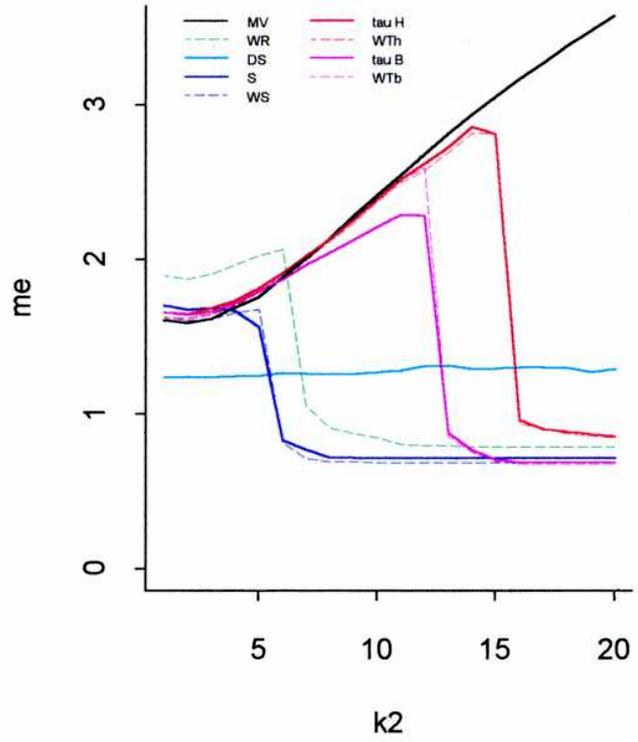
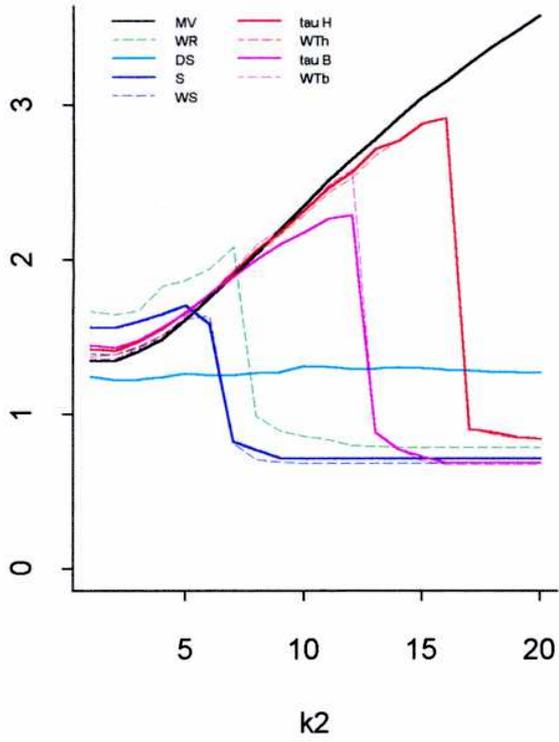
$k_1 = 3$

$k_1 = 4$



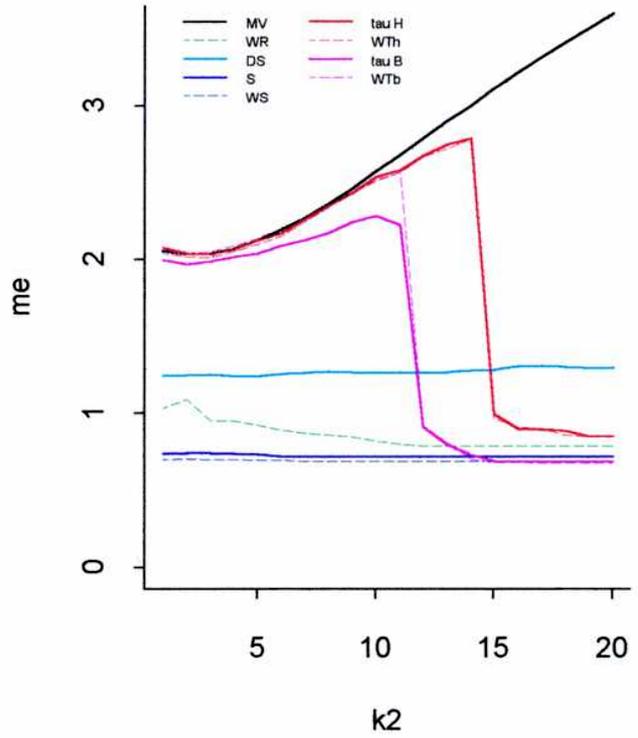
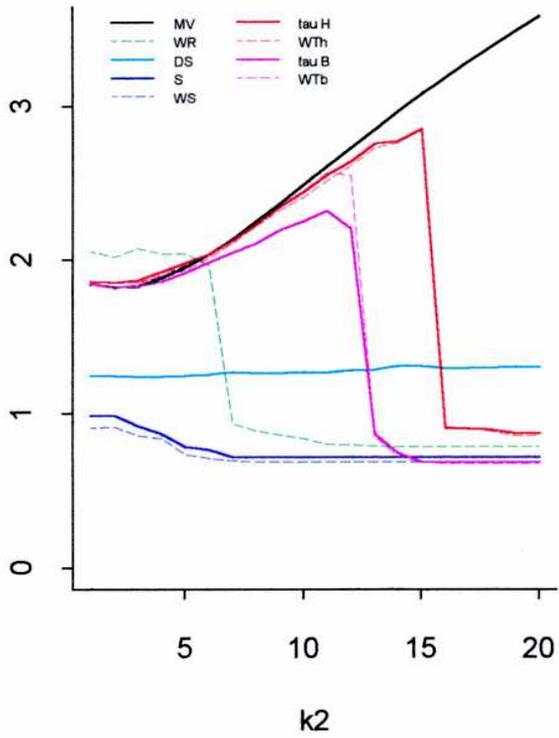
$k_1 = 5$ Contaminación 10%

$k_1 = 6$



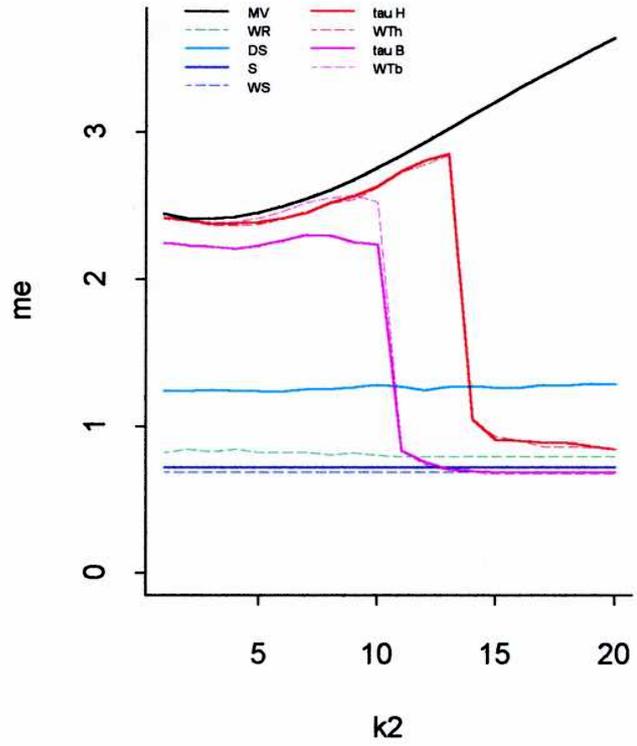
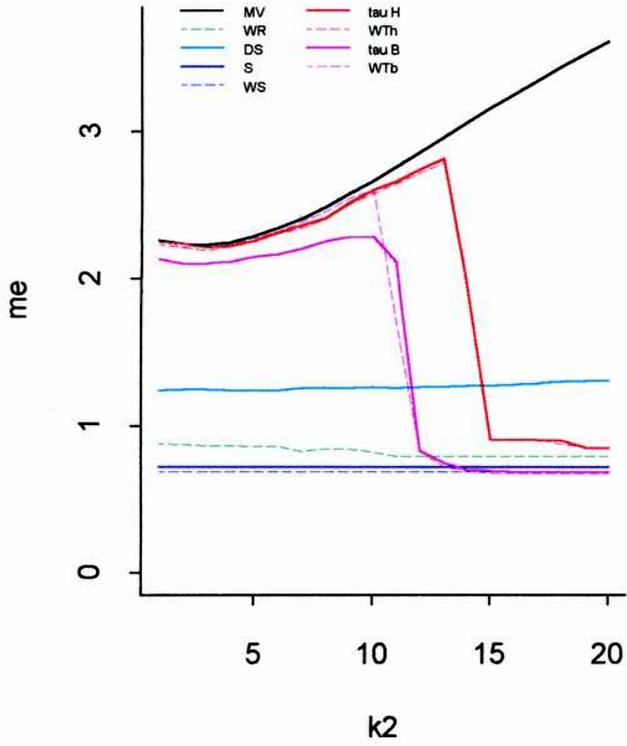
$k_1 = 7$

$k_1 = 8$



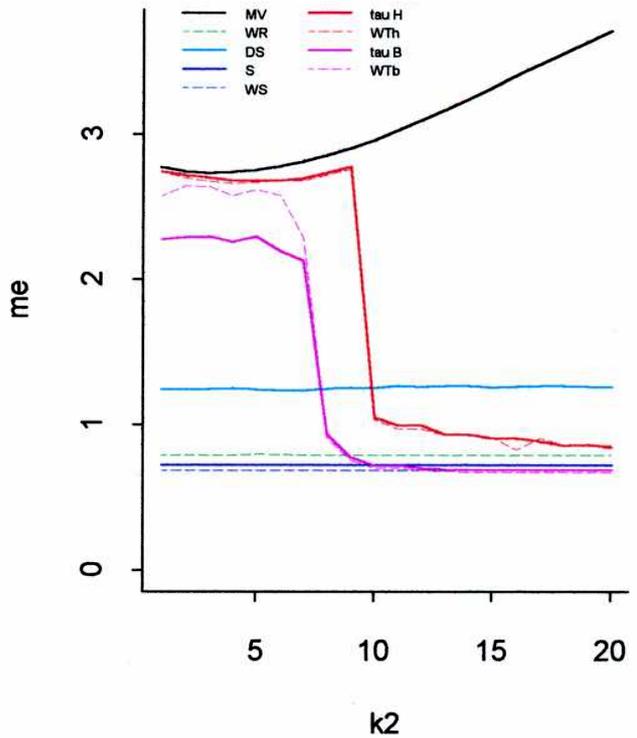
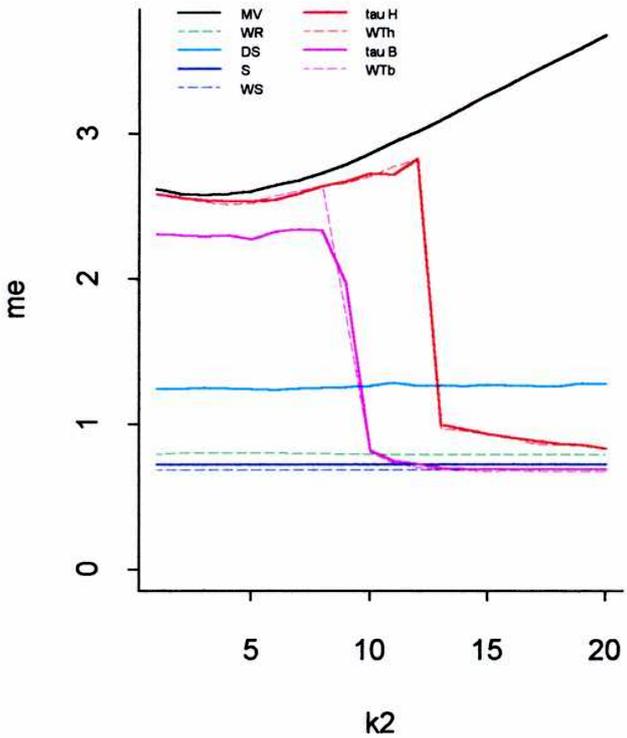
$k_1 = 9$ Contaminación 10%

$k_1 = 10$



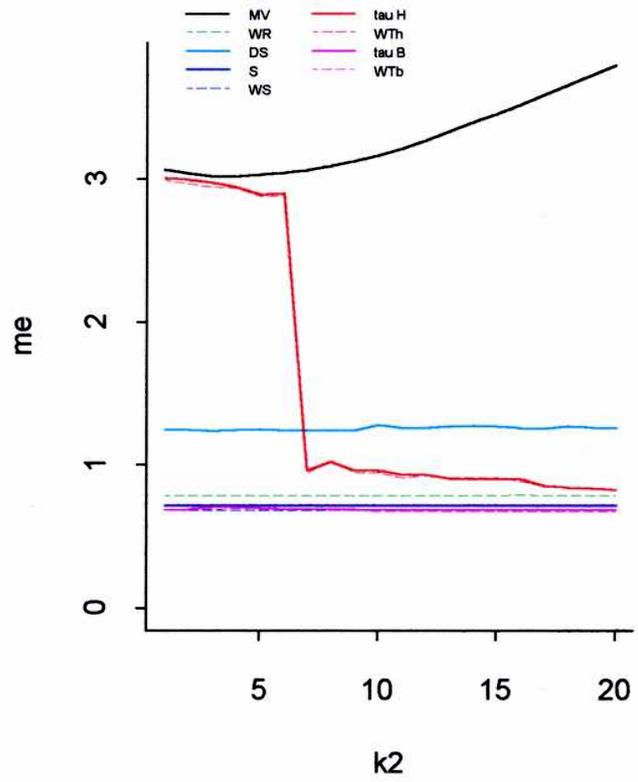
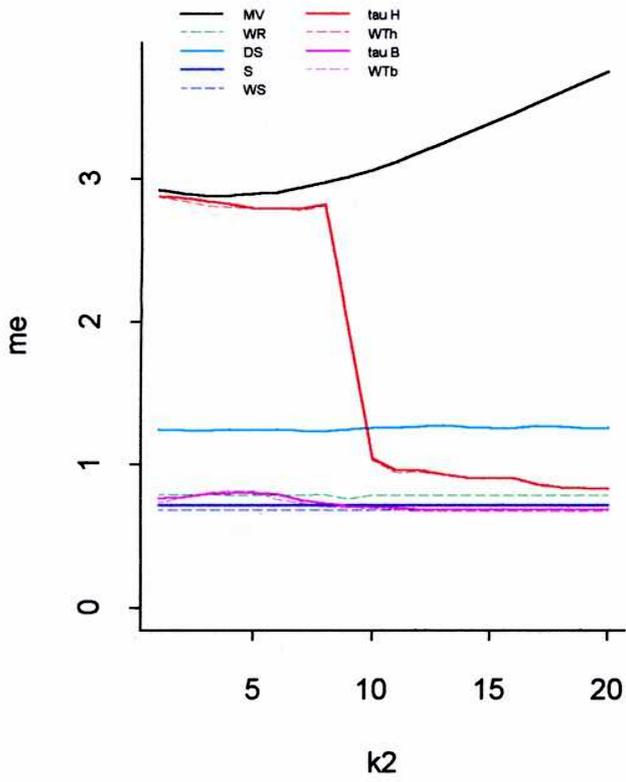
$k_1 = 11$

$k_1 = 12$



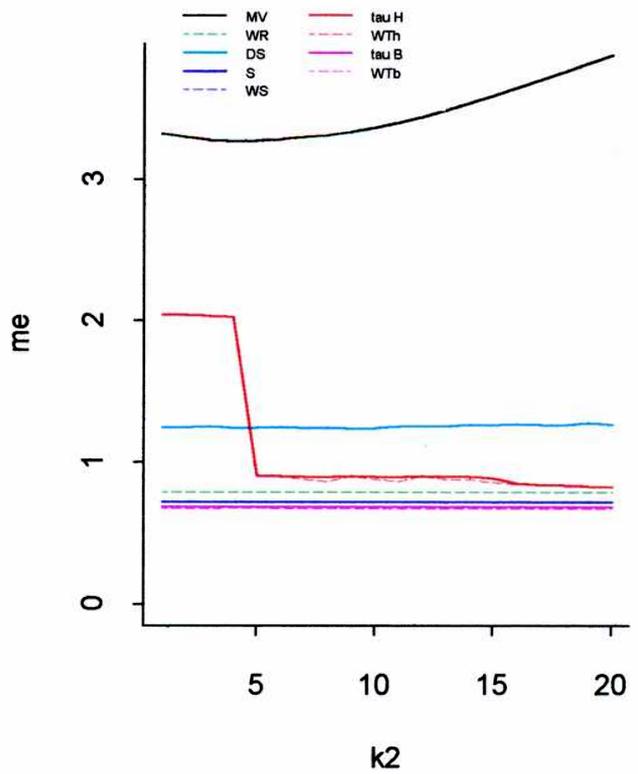
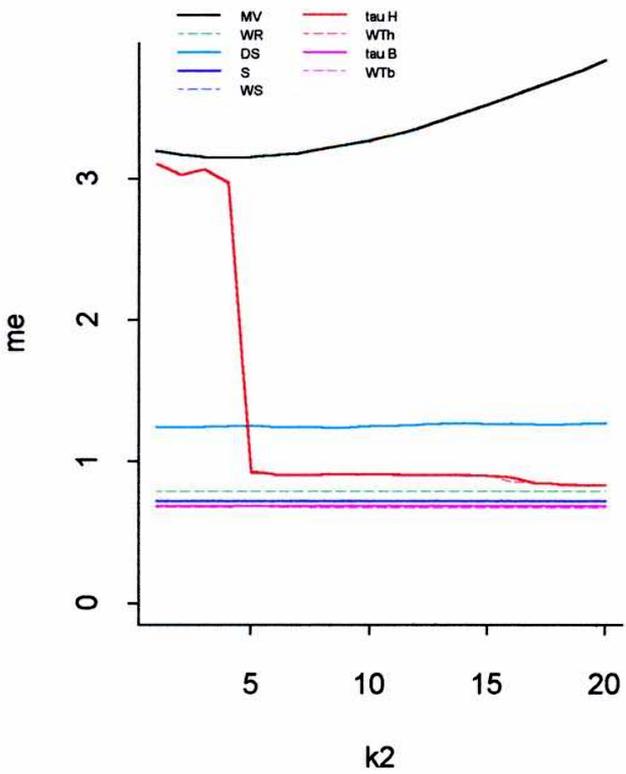
$k_1 = 13$ Contaminación 10%

$k_1 = 14$



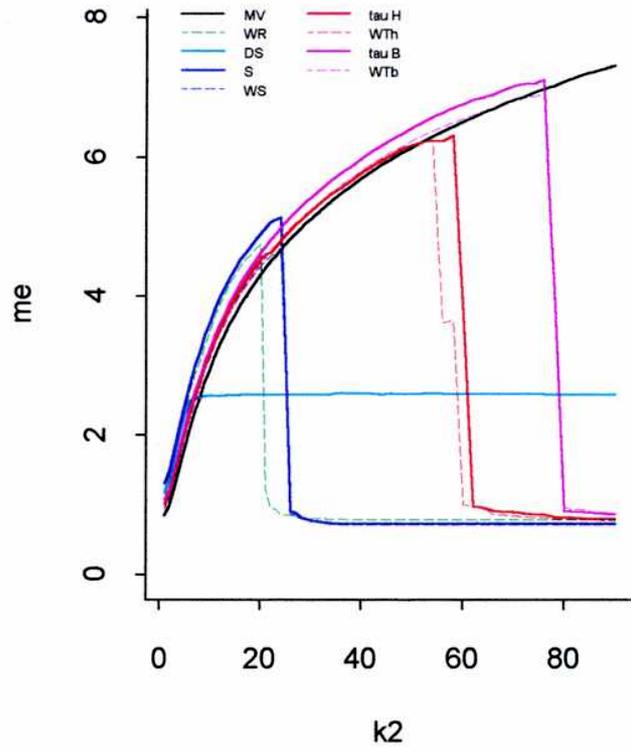
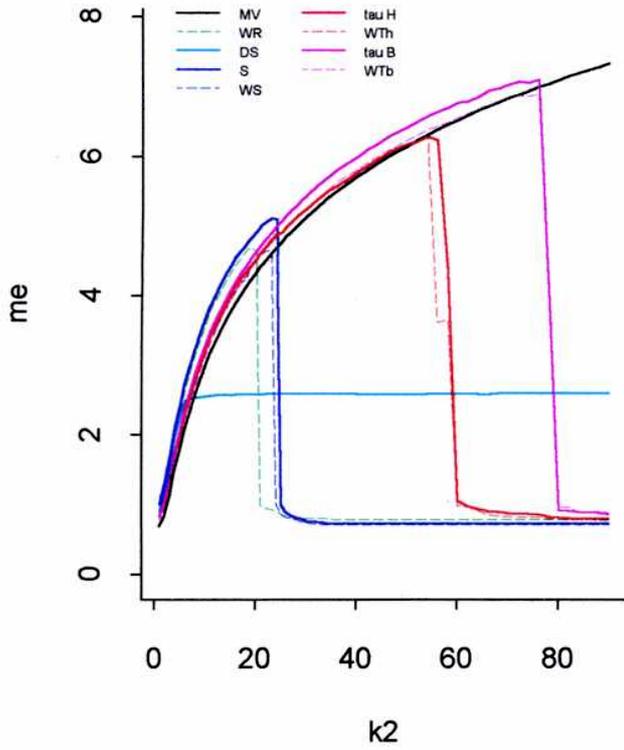
$k_1 = 15$

$k_1 = 16$



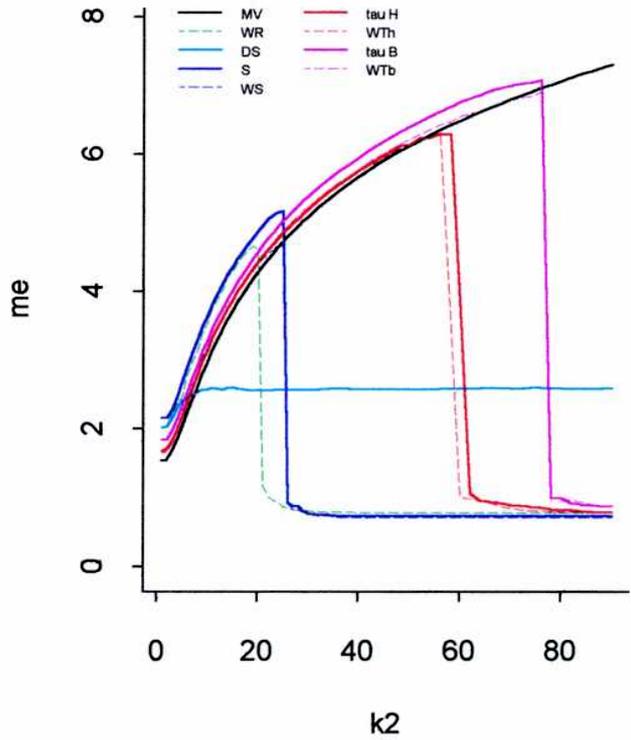
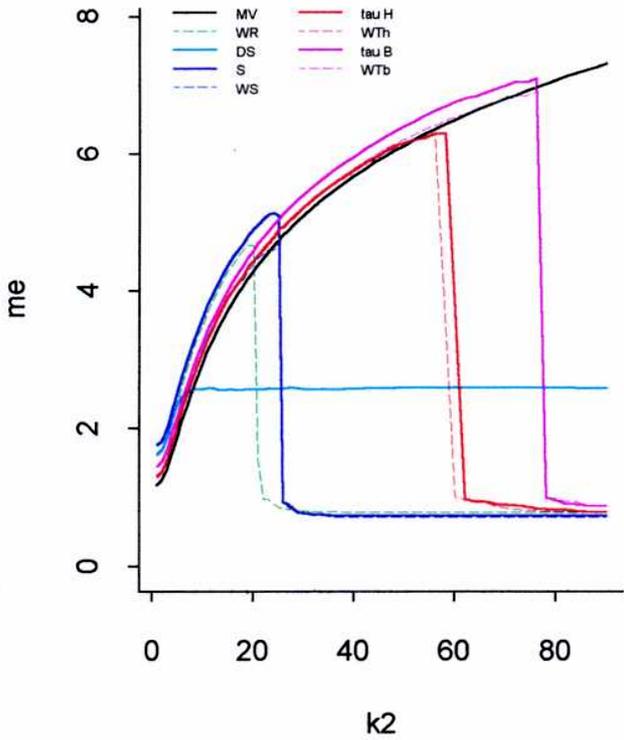
$k_1 = 1$ Contaminación 20%

$k_1 = 2$



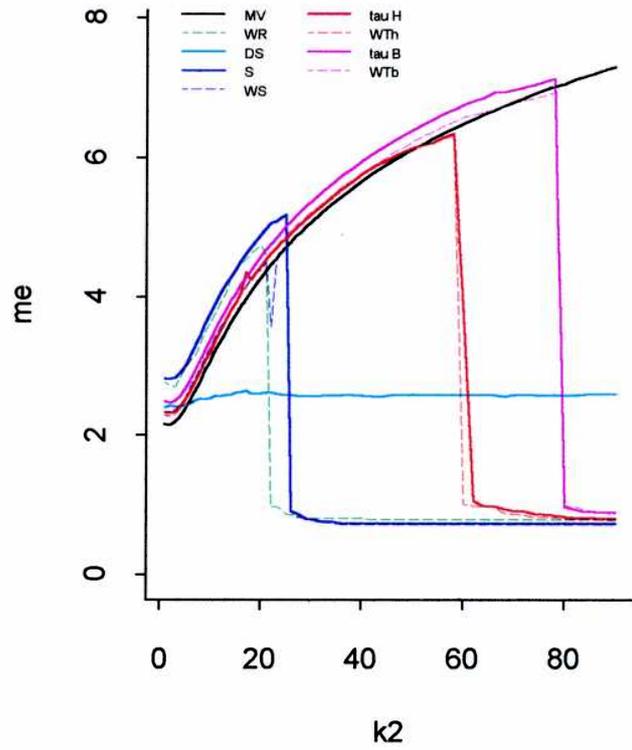
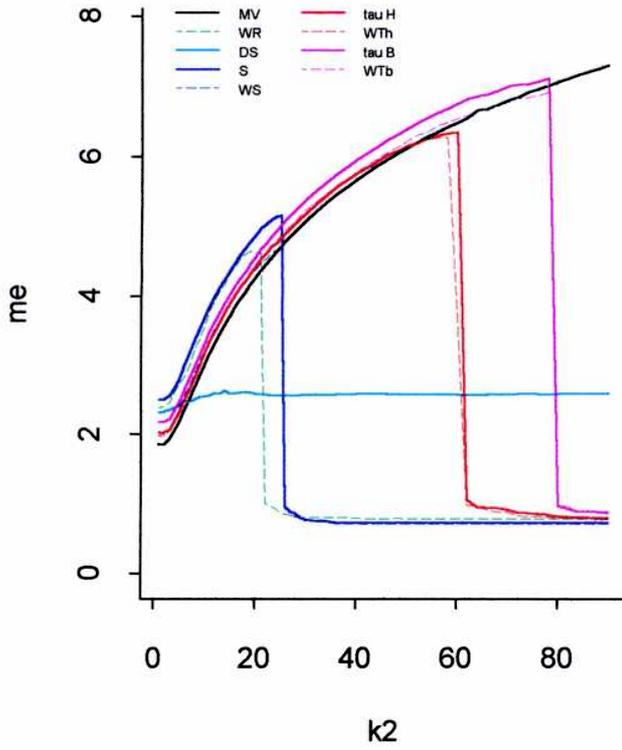
$k_1 = 3$

$k_1 = 4$



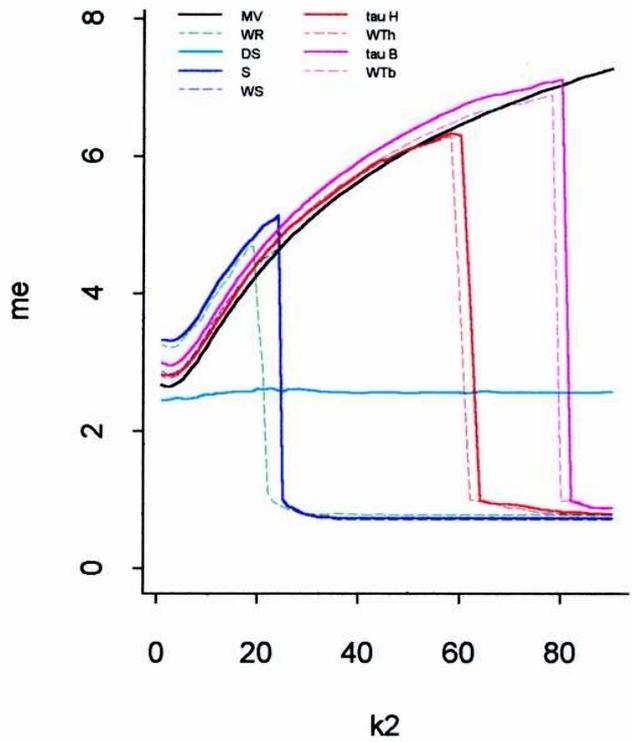
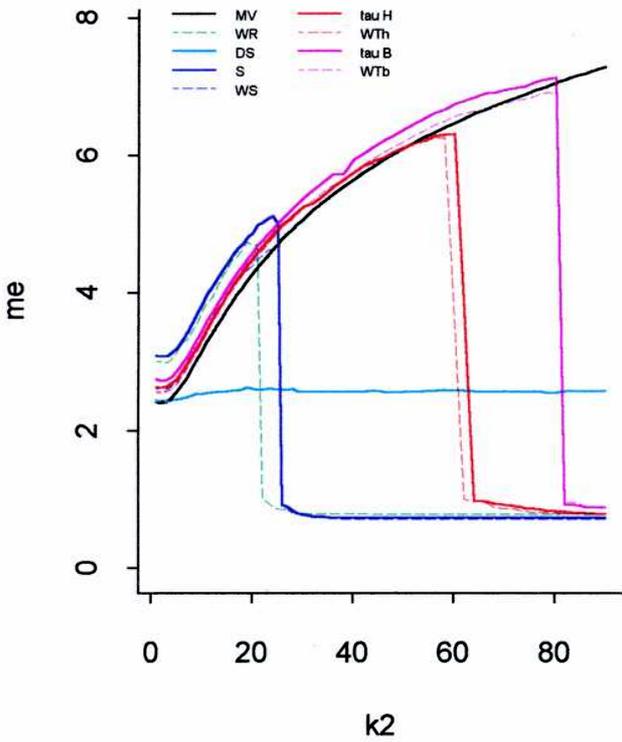
$k_1 = 5$ Contaminación 20%

$k_1 = 6$



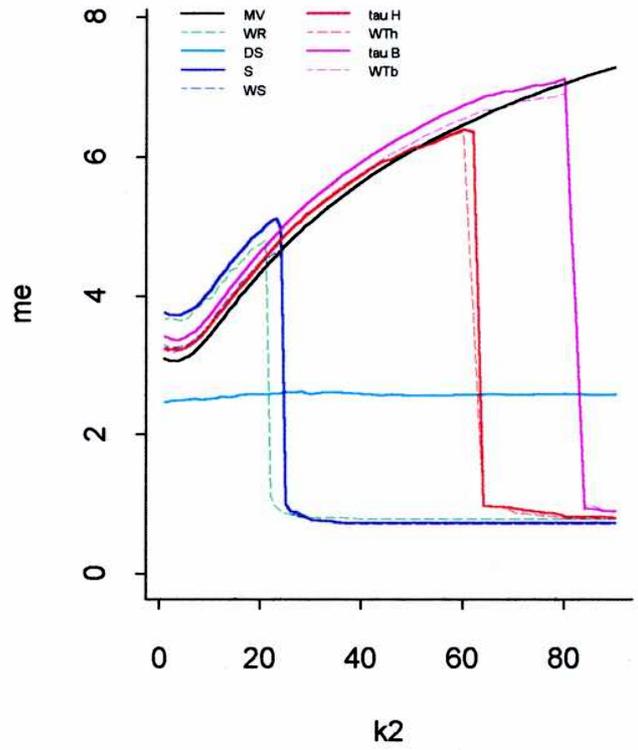
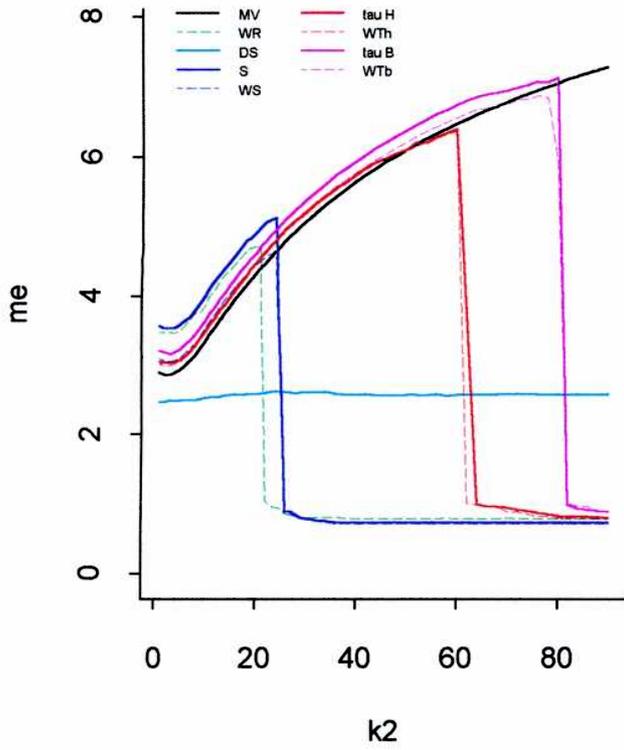
$k_1 = 7$

$k_1 = 8$



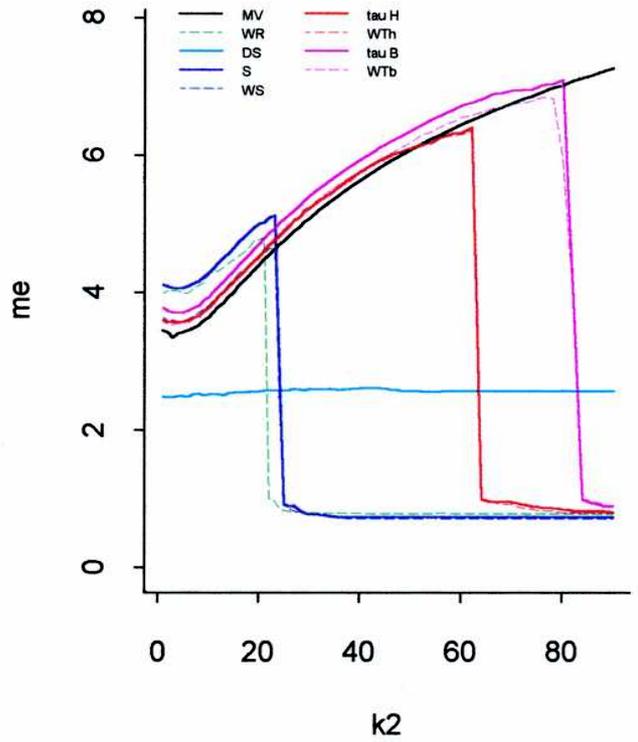
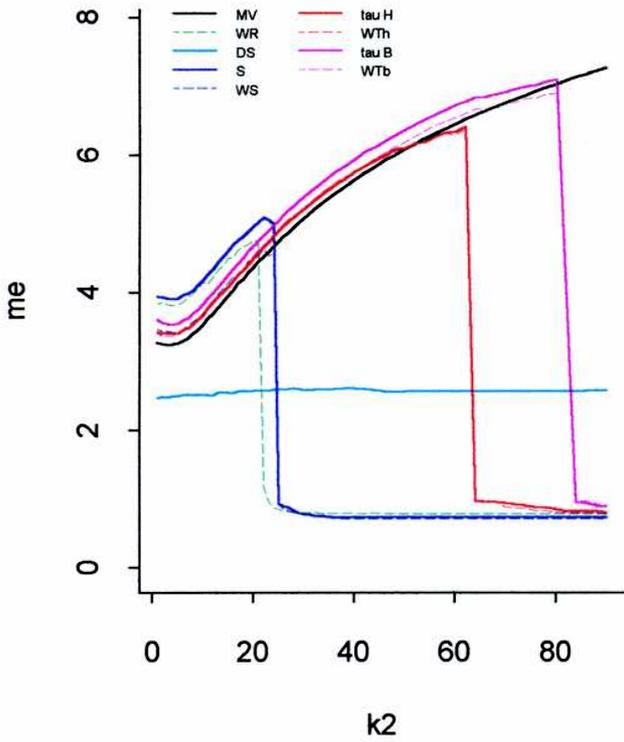
$k_1 = 9$ Contaminación 20%

$k_1 = 10$



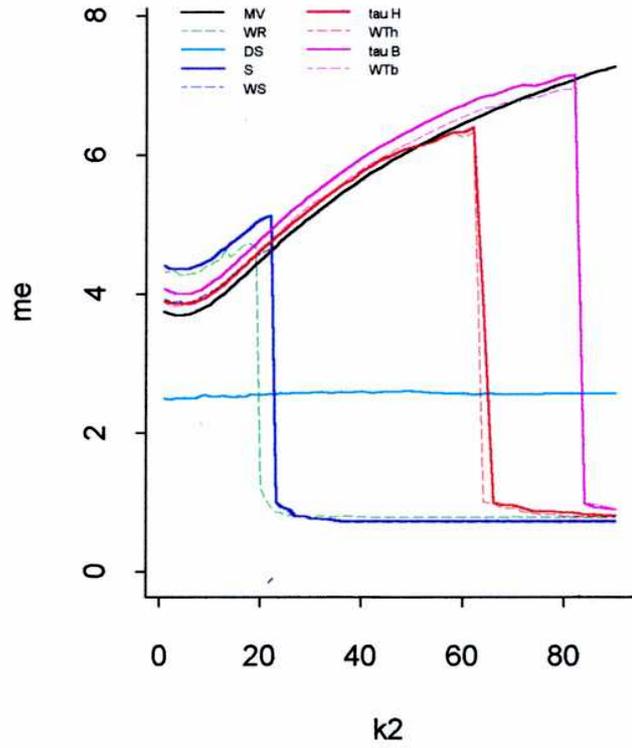
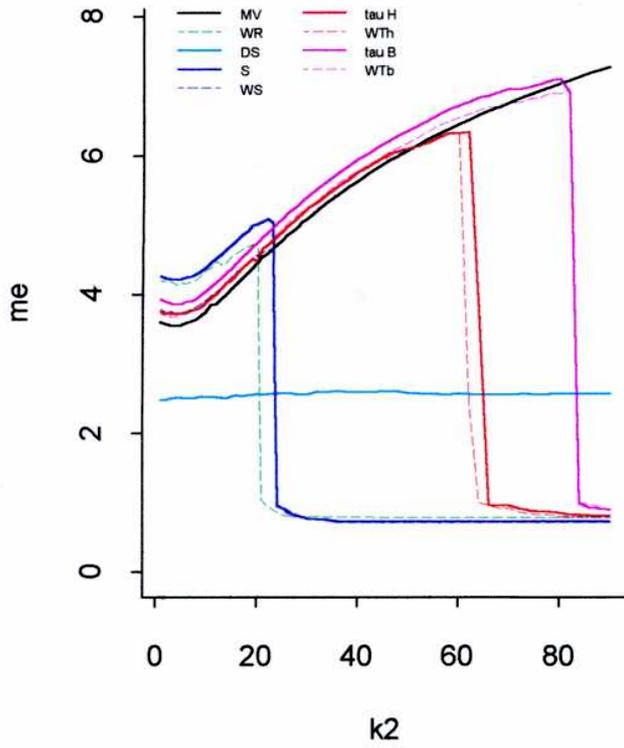
$k_1 = 11$

$k_1 = 12$



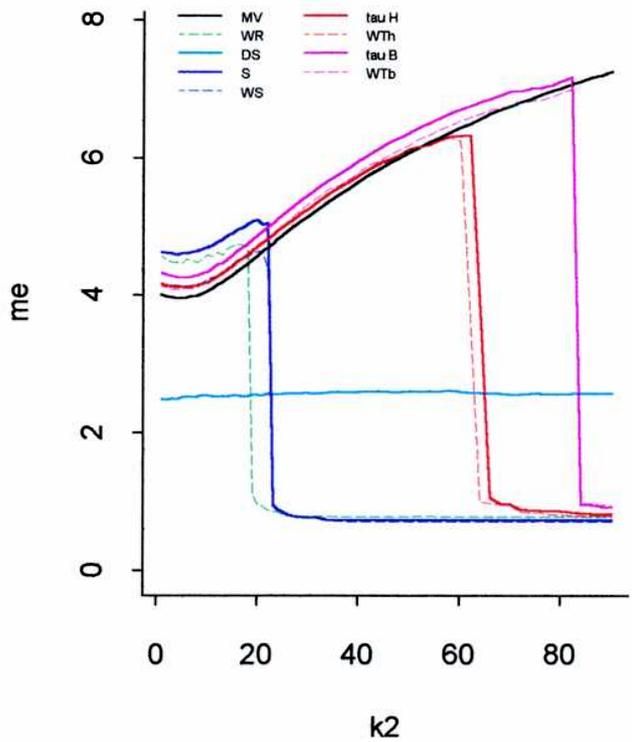
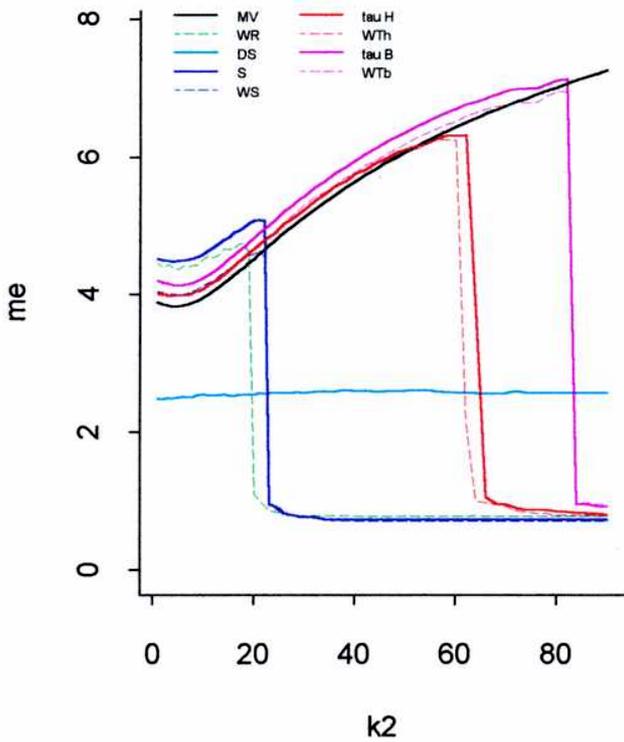
$k_1 = 13$ Contaminación 20%

$k_1 = 14$



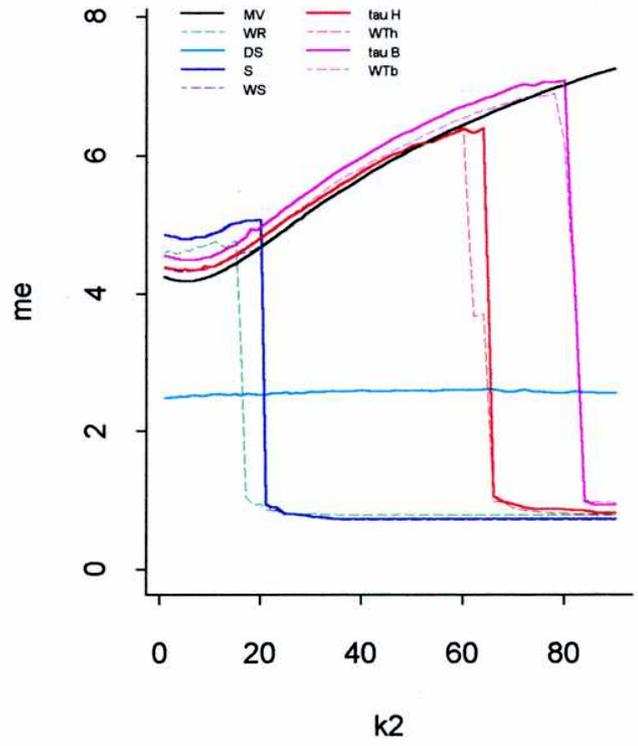
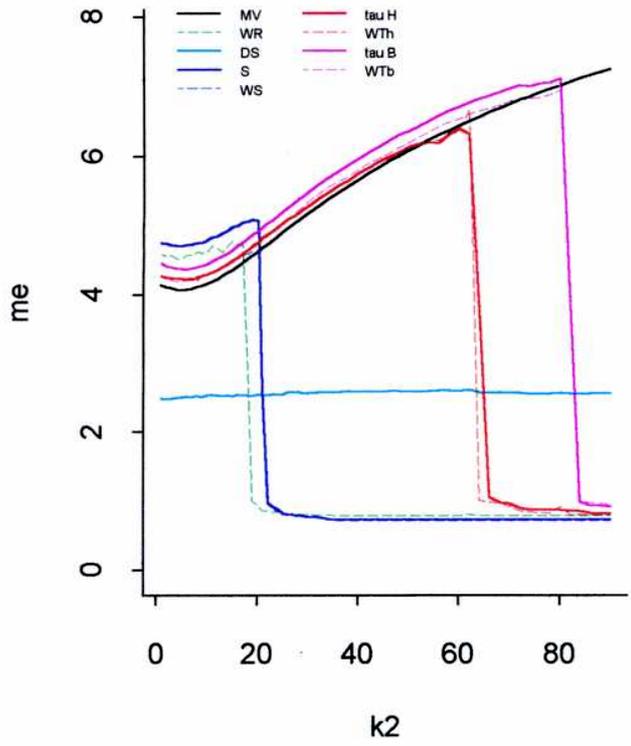
$k_1 = 15$

$k_1 = 16$



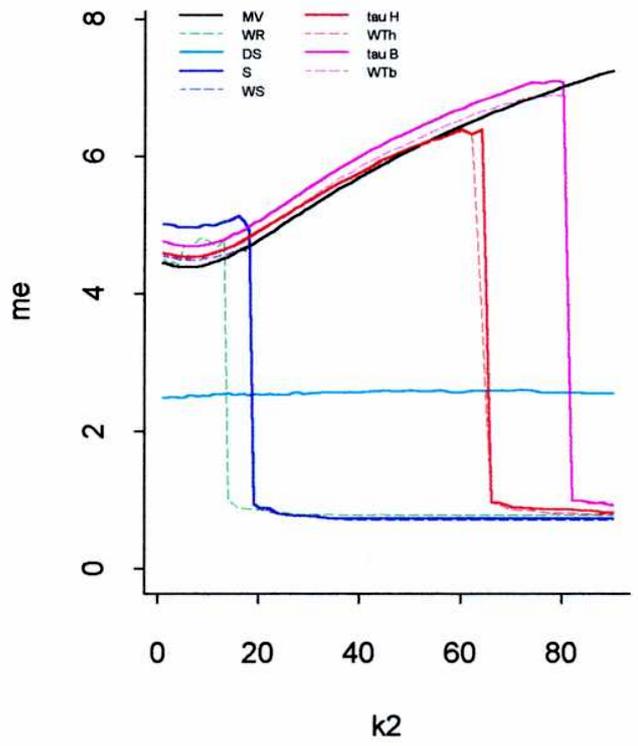
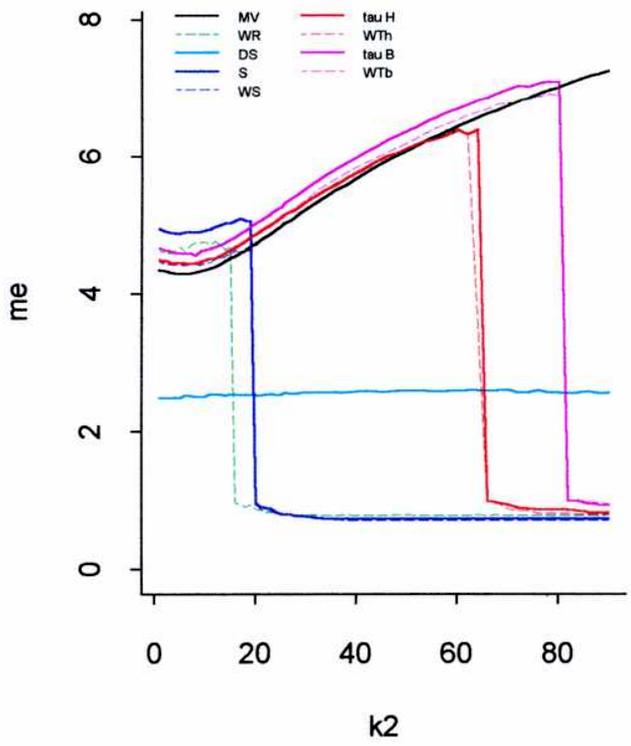
$k_1 = 17$ Contaminación 20%

$k_1 = 18$



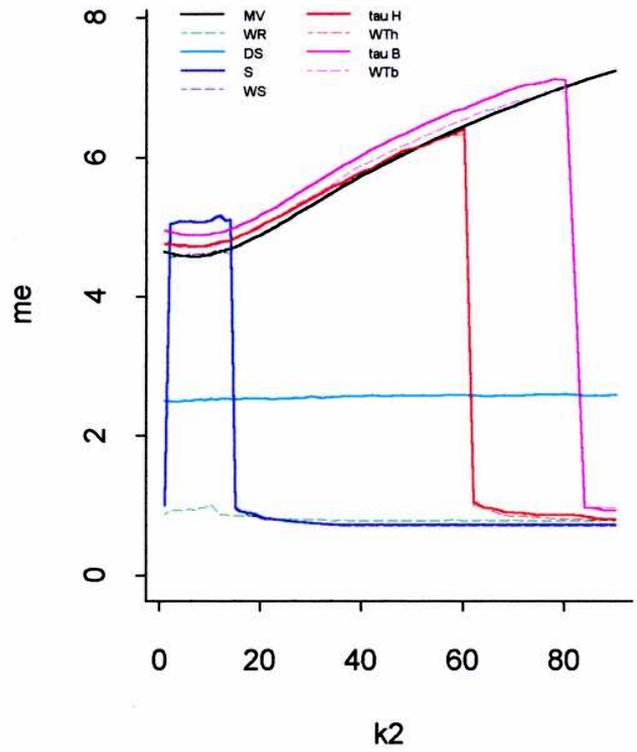
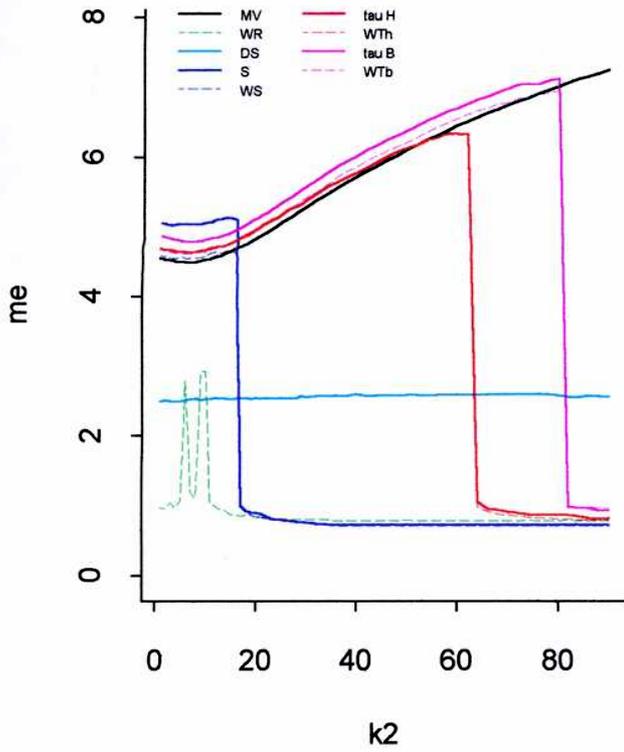
$k_1 = 19$

$k_1 = 20$



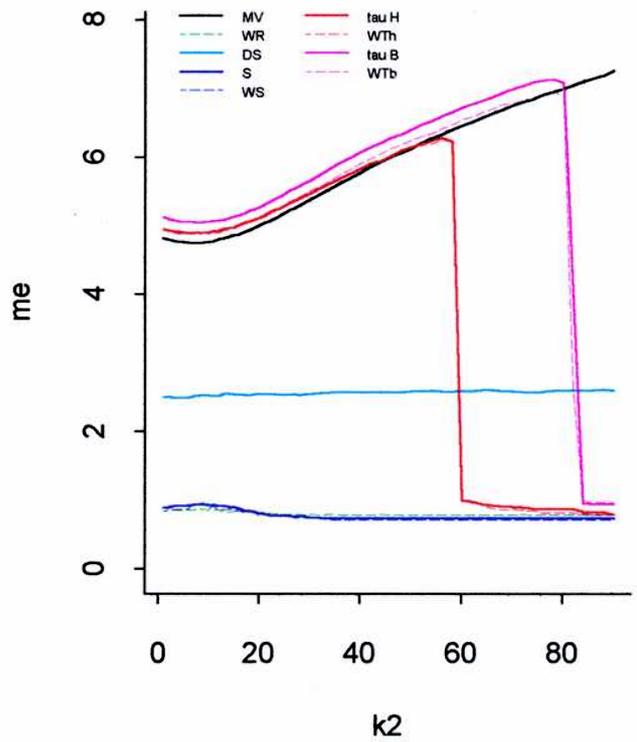
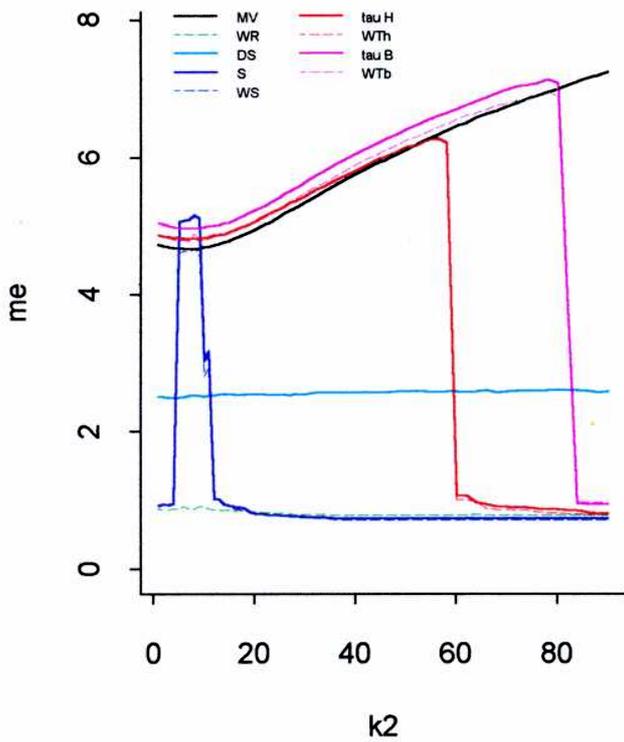
$k_1 = 21$ Contaminación 20%

$k_1 = 22$



$k_1 = 23$

$k_1 = 24$



8 APÉNDICES

A Unicidad de los parámetros

Como hemos visto en los Capítulos 5 y 6, una condición necesaria para obtener la consistencia y la normalidad asintótica de los estimadores de los parámetros del modelo de análisis factorial es la identificabilidad del modelo. Hemos definido los conceptos de identificabilidad global y local (Definiciones 14 y 15). El objetivo es encontrar condiciones para identificación global, las cuales por supuesto, implican identificación local.

Dada $\Sigma_0 \in \text{PSD}(p)$ queremos ver qué condición necesaria se debe cumplir para que existan una única matriz $\Psi_0 \in \mathcal{D}(p)$ con elementos diagonales positivos y una única matriz $\Lambda_0 \in \mathfrak{R}^{p \times k}$ con $\text{rg}(\Lambda_0) = k$ de modo que se satisfaga la ecuación (12)

$$\Sigma_0 = \Lambda_0 \Lambda_0' + \Psi_0 \quad (57)$$

sujeta a la restricción

$$\Lambda_0' \Psi_0^{-1} \Lambda_0 = \Delta, \quad (58)$$

siendo $\Delta \in \mathcal{D}^*(k)$, y cuál es el mínimo k ($k < p$) de modo que esto ocurra.

Una manera de determinar si Σ_0 puede expresarse en la forma deseada, es ver si se puede resolver el conjunto de ecuaciones

$$\begin{aligned} \sigma_{ii} &= \psi_{ii} + \sum_{a=1}^k \lambda_{ia}^2 & 1 \leq i \leq p; \\ \sigma_{ij} &= \sum_{a=1}^k \lambda_{ia} \lambda_{ja} & 1 \leq i < j \leq p. \end{aligned}$$

Si hay una solución algebraica, se debe verificar además que los λ_{ia} sean reales y que los ψ_{ii} sean reales y no negativos.

Una respuesta parcial a la pregunta anterior puede obtenerse comparando el número de ecuaciones con el número de incógnitas. Como en cualquier solución de (57) Λ_0 puede reemplazarse por $\Lambda_0 M$ donde $M \in \mathfrak{R}^{k \times k}$ es una matriz ortogonal, se deben imponer condiciones adicionales para determinar unívocamente la solución de (57). Como hemos visto en la Sección 2.1, una forma de eliminar la indeterminación debida a M es imponer la condición (58), donde Δ es una matriz diagonal, tal que sus elementos son las raíces no nulas de $|\Sigma_0 - \Psi_0 - \gamma \Psi_0| = 0$.

De esta manera, en Σ_0 hay $p(p+1)/2$ elementos; en la diagonal de Ψ_0 hay p elementos y pk elementos en Λ_0 ; además con la condición (58) se están imponiendo $k(k-1)/2$ restricciones adicionales. Por lo tanto el número de ecuaciones y condiciones menos el número de incógnitas está dado por:

$$\begin{aligned}\varsigma &= \frac{p(p+1)}{2} + \frac{k(k-1)}{2} - p - pk \\ &= \frac{(p-k)^2 - (p+k)}{2}.\end{aligned}\tag{59}$$

Podemos esperar que si $\varsigma \geq 0$, entonces se puede encontrar una solución algebraica. La desigualdad $\varsigma \geq 0$ puede escribirse como:

$$k \geq \frac{2p+1 - \sqrt{8p+1}}{2} = p - \frac{\sqrt{8p+1} - 1}{2}.\tag{60}$$

Si $\varsigma > 0$, para encontrar una única solución algebraica de (57) sujeta a (58), y por lo tanto obtener identificabilidad en el modelo, deberíamos agregar $-\varsigma$ restricciones adicionales.

Por otra parte, como hemos visto en la Sección 2.1, para Σ_0 y Ψ_0 dados, la matriz de ponderaciones Λ_0 está definida unívocamente salvo por una post-multiplicación por una matriz ortogonal. Por lo tanto nos concentraremos en la unicidad (identificabilidad) de Ψ_0 . Notemos que la matriz $\Sigma_0 - \Psi_0 = \Lambda_0 \Lambda_0'$ es definida no negativa y tiene el mismo rango que Λ_0 .

Anderson y Rubin [5] (1956) mostraron que las siguientes condiciones son suficientes para identificación global.

1. Si se elimina cualquier fila de Λ_0 , entonces las dos submatrices disjuntas remanentes de Λ_0 tienen rango k .
2. Si se satisface la condición de que toda submatriz de $q \times q$ de Λ_0 es no singular, entonces cualquier solución de rango menor que $p - q$ es una extensión trivial de (Λ_0, Ψ_0) .

La condición 1 se relaciona con soluciones de rango q menor que $p/2$. Más aún, se puede ver que si $q < p/2$ el modelo de análisis factorial es *genéricamente* globalmente identificable.

De hecho, para $q < p/2$ resultados más fuertes ocurren genéricamente. No existe una extensión trivial de la solución (Λ_0, Ψ_0) .

Se debe notar que la condición 2 es más fuerte que 1.

Podríamos preguntarnos, cuando suponemos que hay una Λ_0 y una Ψ_0 que satisfacen (57) bajo la restricción (58), si hay otro par de tales matrices que cumplan estas condiciones en un entorno de (Λ_0, Ψ_0) . Si esto no ocurre, diremos que Λ_0 y Ψ_0 son localmente identificables.

Una condición suficiente para esto está dada en el siguiente Teorema debido a Anderson y Rubin [5] (1956: Teorema 5.9)

Definición 20 Sean $A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{n \times m}$. El producto de Hadamard de estas matrices es una matriz de $n \times m$ tal que:

$$(A * B)_{ij} = a_{ij}b_{ij}.$$

$A * A$ será el cuadrado de Hadamard de la matriz A .

Teorema 14 Sea $R = \Psi_0 - \Lambda_0 (\Lambda_0' \Psi_0^{-1} \Lambda_0)^{-1} \Lambda_0'$.

Si $|R * R| \neq 0$, entonces Ψ_0 y Λ_0 son localmente identificables bajo la restricción de que $\Lambda_0' \Psi_0^{-1} \Lambda_0 \in \mathcal{D}^*(k)$.

Definición 21 Sea

$$\phi(p) = \frac{2p + 1 - \sqrt{8p + 1}}{2}.$$

La función $\phi(p)$ se denomina cota de Ledermann.

Definición 22 Consideremos a (Λ_0, Ψ_0) como un punto del espacio vectorial $\mathbb{R}^{p \times k + p}$. Diremos que una condición acerca de la identificabilidad de (Λ_0, Ψ_0) se satisface para casi todo (Λ_0, Ψ_0) si dicha condición se satisface para todo (Λ_0, Ψ_0) excepto en un conjunto de medida de Lebesgue nula.

El siguiente Teorema, dado por Shapiro [38] (1985: Teorema 3.2) proporciona una regla simple para determinar la identificabilidad local del modelo (48) a partir del conteo de ecuaciones.

Teorema 15 Si $k \leq \phi(p)$ el modelo (48) es localmente identificable para casi todo (Λ_0, Ψ_0) .

Para $k > \phi(p)$ el modelo de análisis factorial es localmente (y por lo tanto globalmente) no identificable para casi todo (Λ_0, Ψ_0) .

El problema de la identificabilidad global del modelo permanece aún sin resolver. Shapiro [37] (1982) postula las siguientes conjeturas.

Conjetura 1 *Si $k < \phi(p)$ entonces el modelo de análisis factorial es globalmente identificable para casi todo (Λ_0, Ψ_0) .*

La condición suficiente del Teorema 14 pueden ocurrir sólo si $k \leq \phi(p)$. Sin embargo, esto no prueba que (48) sea no identificable para casi todo (Λ_0, Ψ_0) si $k > \phi(p)$. Por supuesto para $k > \phi(p)$ la no identificabilidad ocurre genéricamente como se señaló en la segunda parte del Teorema 15

Conjetura 2 *Si $k > \phi(p)$, entonces el modelo de análisis factorial es (localmente) no identificable para casi todo (Λ_0, Ψ_0) .*

B Resultados previos

B.1 Resultados de álgebra lineal

Recordaremos algunos resultados de álgebra lineal que usaremos para obtener la descomposición de la matriz Σ (ver por ej, Muirhead [27] (1982) o Seber [36] (1984)).

Teorema 16 (*descomposición espectral*)

Sea $A \in \mathbb{R}^{n \times n}$ una matriz simétrica. Entonces existe una matriz $T \in \mathbb{R}^{n \times n}$ ortogonal, $T = [\mathbf{t}_1, \dots, \mathbf{t}_n]$ tal que

$$T'AT = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = D,$$

donde $\lambda_1 \geq \dots \geq \lambda_n$ son los autovalores ordenados de A y $\mathbf{t}_1, \dots, \mathbf{t}_n$ sus correspondientes autovectores.

Con este orden, D es única y T también, salvo un factor de post-multiplicación

$$M = \begin{bmatrix} M_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & M_k \end{bmatrix},$$

siendo $M_i \in \mathcal{O}(m_i)$, donde k es el número de autovalores distintos de A , m_1, \dots, m_k sus multiplicidades y $\mathcal{O}(m_i)$ corresponde al conjunto de matrices ortogonales de $m_i \times m_i$.

Corolario 10 Si $A \in \mathbb{R}^{n \times n}$ es una matriz simétrica tal que sus autovalores son distintos, $m_i = 1$ para $1 \leq i \leq k$, M se reduce a una matriz diagonal cuyos elementos no nulos son ± 1 .

En este caso las columnas \mathbf{t}_i de T son únicas excepto por sus signos.

Lema 1 Sea $A \in \mathbb{R}^{n \times n}$ una matriz no singular.

Entonces, si λ es autovalor de A con autovector \mathbf{t} , $1/\lambda$ es autovalor de A^{-1} con autovector \mathbf{t} .

Lema 2 Sea $A \in \mathbb{R}^{m \times n}$. Entonces AA' es una matriz semidefinida positiva

Lema 3 Una matriz simétrica es semidefinida positiva si y sólo si sus autovalores son no negativos.

Lema 4 Sea $A \in \mathbb{R}^{n \times n}$ una matriz simétrica semidefinida positiva. Entonces:

1. $\text{tr}(A) \geq 0$.
2. Sea $C \in \mathbb{R}^{p \times n} \Rightarrow C'AC$ es simétrica semidefinida positiva.
3. Existe una matriz simétrica semidefinida positiva $A^{1/2}$ tal que $A = (A^{1/2})^2$. Los autovalores de $A^{1/2}$ son las raíces cuadradas de los autovalores de A .

Lema 5 Una matriz simétrica A es definida positiva si y sólo si existe una matriz no singular R tal que $A = R'R$.

Lema 6 Sea $A \in \mathbb{R}^{n \times n}$ una matriz simétrica definida positiva. Entonces:

1. Existe una matriz simétrica $A^{1/2}$ tal que $A = (A^{1/2})^2$ y $A^{1/2}$ es definida positiva.
2. A^{-1} es definida positiva.
3. Sea $C \in \mathbb{R}^{p \times n}$, $\text{rg}(C) = p \Rightarrow C'AC$ es definida positiva.

Lema 7 Sean $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times m}$.

Si $m > n$, entonces los autovalores de AB son los de BA junto con el autovalor 0 con multiplicidad $m - n$.

Además, los autovalores de $I_m + AB$ son los de $I_n + BA$ junto con el autovalor 1 con multiplicidad $m - n$.

Por lo tanto

$$|I_m + AB| = |I_n + BA|.$$

Lema 8 Sea $A \in \mathbb{R}^{n \times n}$ una matriz simétrica con autovalores $\lambda_1 \geq \dots \geq \lambda_n$ y autovectores respectivos ortonormales $\mathbf{t}_1, \dots, \mathbf{t}_n$.

Sea $C = \{\mathbf{x} \in \mathbb{R}^n / \mathbf{x} \neq \mathbf{0}\}$.

Entonces:

1. $\sup_C \left\{ \frac{\mathbf{x}'A\mathbf{x}}{\mathbf{x}'\mathbf{x}} \right\} = \lambda_1$ y el supremo se alcanza en $\mathbf{x} = \mathbf{t}_1$.
2. $\inf_C \left\{ \frac{\mathbf{x}'A\mathbf{x}}{\mathbf{x}'\mathbf{x}} \right\} = \lambda_n$ y el ínfimo se alcanza en $\mathbf{x} = \mathbf{t}_n$.

Corolario 11 Sea $A \in \mathfrak{R}^{n \times n}$ una matriz simétrica con autovalores $\lambda_1 \geq \dots \geq \lambda_n$ y autovectores respectivos ortonormales $\mathbf{t}_1, \dots, \mathbf{t}_n$.

Sea $C^* = \{\mathbf{x} \in \mathfrak{R}^n : \|\mathbf{x}\| = 1\}$.

Entonces:

1. $\sup_{C^*} \|A\mathbf{x}\| = \lambda_1$ y el supremo se alcanza en $\mathbf{x} = \mathbf{t}_1$.

2. $\inf_{C^*} \|A\mathbf{x}\| = \lambda_n$ y el ínfimo se alcanza en $\mathbf{x} = \mathbf{t}_n$.

Lema 9 Sea $A \in \mathfrak{R}^{n \times n}$ una matriz simétrica con autovalores $\lambda_1 \geq \dots \geq \lambda_n$.

Entonces

$$\|A\|^2 = \sum_{i=1}^n \lambda_i^2.$$

Lema 10 Sean $A \in \mathfrak{R}^{n \times n}$ y $B \in \mathfrak{R}^{n \times n}$ matrices simétricas tales que $A - B$ es una matriz semidefinida positiva.

Entonces:

1. $\text{tr}(A) \geq \text{tr}(B)$.

2. $\|A\| \geq \|B\|$.

Lema 11 Sean $(A_n)_{n \geq 1} \in \mathfrak{R}^{p \times p}$ y $(B_n)_{n \geq 1} \in \mathfrak{R}^{p \times p}$ sucesiones de matrices simétricas no singulares tales que $\|A_n\| \xrightarrow{n \rightarrow \infty} \infty$, B_n es definida positiva $\forall n$ y $B_n \xrightarrow{n \rightarrow \infty} B_0$, siendo B_0 una matriz simétrica definida positiva.

Sea θ_n el mínimo autovalor de $A_n^{-1}B_n$.

Entonces $\theta_n \xrightarrow{n \rightarrow \infty} 0$.

Dem. Sean $\lambda_1^n \geq \dots \geq \lambda_p^n$ los autovalores de A_n con autovectores respectivos ortonormales $\mathbf{x}_1^n, \dots, \mathbf{x}_p^n$, donde $\lambda_i^n \neq 0$ para $1 \leq i \leq p$ por ser A_n no singular.

Por Lema 1, $\mu_p^n = 1/\lambda_1^n \leq \dots \leq \mu_1^n = 1/\lambda_p^n$ son los autovalores de A_n^{-1} con autovectores respectivos ortonormales $\mathbf{x}_1^n, \dots, \mathbf{x}_p^n$.

Por Lema 9, $\lambda_1^n \xrightarrow{n \rightarrow \infty} \infty$ entonces

$$\mu_p^n \xrightarrow{n \rightarrow \infty} 0. \quad (61)$$

Además

$$\|A_n^{-1}\mathbf{x}_1^n\| = \mu_p^n. \quad (62)$$

Sean $\delta_1^n \geq \dots \geq \delta_p^n$ los autovalores de B_n , donde $\delta_i^n \neq 0$ para $1 \leq i \leq p$ por ser B_n no singular. Entonces $\gamma_p^n = 1/\delta_1^n \leq \dots \leq \gamma_1^n = 1/\delta_p^n$ son los autovalores de B_n^{-1} .

Como $B_n \xrightarrow{n \rightarrow \infty} B_0$, entonces $\delta_i^n \xrightarrow{n \rightarrow \infty} \delta_i$ para $1 \leq i \leq p$, siendo $\delta_1 \geq \dots \geq \delta_p > 0$ los autovalores de B_0 .

Además

$$\gamma_i^n = \frac{1}{\delta_{p-i+1}^n} \xrightarrow{n \rightarrow \infty} \frac{1}{\delta_{p-i+1}} = \gamma_i,$$

siendo $\gamma_1 \geq \dots \geq \gamma_p > 0$ los autovalores de B_0^{-1} .

Definamos

$$\mathbf{y}_n = B_n^{-1} \mathbf{x}_1^n.$$

Luego

$$\|\mathbf{y}_n\| = \|B_n^{-1} \mathbf{x}_1^n\| \geq \gamma_p^n \xrightarrow{n \rightarrow \infty} \gamma_p,$$

es decir,

$$\liminf_{n \rightarrow \infty} \|\mathbf{y}_n\| > 0. \quad (63)$$

Además

$$\|\mathbf{y}_n\| \leq \|B_n^{-1}\| \|\mathbf{x}_1^n\| = \|B_n^{-1}\| = \sum_{i=1}^p (\gamma_i^n)^2 \xrightarrow{n \rightarrow \infty} \sum_{i=1}^p \gamma_i^2 < \infty,$$

es decir

$$\limsup_{n \rightarrow \infty} \|\mathbf{y}_n\| < \infty. \quad (64)$$

Sea $\mathbf{z}_n = \frac{\mathbf{y}_n}{\|\mathbf{y}_n\|}$. Entonces $\|\mathbf{z}_n\| = 1$.

Pero

$$A_n^{-1} B_n \mathbf{z}_n = \frac{A_n^{-1} B_n B_n^{-1} \mathbf{x}_1^n}{\|\mathbf{y}_n\|} = \frac{A_n^{-1} \mathbf{x}_1^n}{\|\mathbf{y}_n\|},$$

entonces, por (62)

$$\|A_n^{-1} B_n \mathbf{z}_n\| = \frac{\|A_n^{-1} \mathbf{x}_1^n\|}{\|\mathbf{y}_n\|} = \frac{\mu_p^n}{\|\mathbf{y}_n\|}.$$

Luego, por (61), (63) y (64)

$$\|A_n^{-1} B_n \mathbf{z}_n\| \xrightarrow{n \rightarrow \infty} 0.$$

Sea $\mathcal{C}^* = \{\mathbf{z} \in \mathfrak{R}^p : \|\mathbf{z}\| = 1\}$. Entonces

$$0 \leq \theta_n = \inf_{\mathcal{C}^*} \|A_n^{-1} B_n \mathbf{z}\| \leq \|A_n^{-1} B_n \mathbf{z}_n\| \xrightarrow{n \rightarrow \infty} 0,$$

y por lo tanto

$$\theta_n \xrightarrow{n \rightarrow \infty} 0.$$

B.2 Resultados sobre derivadas

En la demostración de varios teoremas hemos tenido que calcular derivadas matriciales. Magnus y Neudecker [25] (1985) dan la siguiente definición para las derivadas de una función matricial $F(X)$ con respecto a su matriz argumento X .

Definición 23 Sea $X \in \mathfrak{R}^{m \times n}$. Sea $y = y(X)$ una función real diferenciable.

Entonces $\partial y(X)/\partial X$ es la matriz de $m \times n$ tal que

$$\frac{\partial y(X)}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{1n}} \\ \vdots & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{mn}} \end{bmatrix},$$

o equivalentemente

$$\left(\frac{\partial y(X)}{\partial X} \right)_{\alpha\beta} = \frac{\partial y}{\partial x_{\alpha\beta}}$$

con $1 \leq \alpha \leq m$, $1 \leq \beta \leq n$.

Algunos resultados sobre derivación matricial (ver por ej. Dwyer [10] (1967)) que usaremos en nuestras demostraciones son los siguientes:

Notación: Sea $A \in \mathfrak{R}^{n \times n}$ una matriz no singular. Notaremos con A^{-t} a la traspuesta de la inversa de A , ie $A^{-t} = (A^{-1})'$.

Lema 12 Sea $X \in \mathfrak{R}^{m \times n}$

1. Si X es no singular, entonces

$$\frac{\partial |X|}{\partial X} = |X|X^{-t}.$$

2. Si $n = m$, entonces

$$\frac{\partial(\text{tr } X)}{\partial X} = \frac{\partial(\text{tr } X')}{\partial X} = I_m. \quad (65)$$

Lema 13 (Regla de la cadena)

Sean $X \in \mathfrak{R}^{m \times n}$ e $y = y(X)$. Si $v = v(y)$, entonces

$$\frac{\partial v}{\partial X} = \frac{\partial v}{\partial y} \frac{\partial y}{\partial X}.$$

Corolario 12 Si $X \in \mathfrak{R}^{m \times n}$ es no singular, entonces

$$\frac{\partial \ln |X|}{\partial X} = X^{-t}. \quad (66)$$

Notación: Sea $K^{ij} \in \mathfrak{R}^{n \times n}$ la matriz tal que:

$$(K^{ij})_{\gamma\delta} = \begin{cases} 1 & \text{si } \gamma = i, \delta = j \\ 0 & \text{en otro caso} \end{cases}$$

Observación: Si $A \in \mathfrak{R}^{n \times n}$ vale que:

$$\sum_{\gamma} \sum_{\delta} a_{\gamma\delta} K^{\gamma\delta} = A; \quad (67)$$

$$\sum_{\gamma} \sum_{\delta} a_{\gamma\delta} (K^{\gamma\delta})' = A'. \quad (68)$$

Lema 14 Sea $X \in \mathfrak{R}^{m \times n}$. Entonces:

$$\frac{\partial X_{\gamma\delta}}{\partial X} = K^{\gamma\delta}. \quad (69)$$

Lema 15 Sean $X \in \mathfrak{R}^{m \times n}$ y $A \in \mathfrak{R}^{p \times m}$, donde A es una matriz constante.

1. Entonces:

$$\frac{\partial (XAX')_{\gamma\delta}}{\partial X} = K^{\gamma\delta} XA' + (K^{\gamma\delta})' XA. \quad (70)$$

2. Sea además $B \in \mathfrak{R}^{n \times q}$, donde B es una matriz constante. Entonces:

$$\frac{\partial (AX^{-1}B)_{\gamma\delta}}{\partial X} = -X^{-t} A' K^{\gamma\delta} B' X^{-t}. \quad (71)$$

Lema 16 (Regla de la cadena)

Sean $X \in \mathfrak{R}^{n \times m}$ e $Y \in \mathfrak{R}^{p \times q}$, tal que $Y = Y(X)$.

Sea w una función escalar de Y , i.e. $w = w(Y)$.

Entonces:

$$\frac{\partial w}{\partial x_{\alpha\beta}} = \sum_{\gamma} \sum_{\delta} \frac{\partial w}{\partial y_{\gamma\delta}} \frac{\partial y_{\gamma\delta}}{\partial x_{\alpha\beta}},$$

o equivalentemente en forma matricial

$$\frac{\partial w}{\partial X} = \sum_{\gamma} \sum_{\delta} \frac{\partial w}{\partial y_{\gamma\delta}} \frac{\partial y_{\gamma\delta}}{\partial X} = \sum_{\gamma} \sum_{\delta} \left(\frac{\partial w}{\partial Y} \right)_{\gamma\delta} \frac{\partial Y_{\gamma\delta}}{\partial X}. \quad (72)$$

Lema 17 Sean $X \in \mathbb{R}^{n \times m}$ e $Y \in \mathbb{R}^{p \times q}$, tal que $Y = Y(X)$. Entonces:

$$\frac{\partial Y}{\partial X} = \sum_{\gamma} \sum_{\delta} |Y| (Y^{-t})_{\gamma\delta} \frac{\partial Y_{\gamma\delta}}{\partial X}. \quad (73)$$

Lema 18 Sea $\Sigma \in \mathbb{R}^{p \times p}$ una matriz simétrica y $x \in \mathbb{R}^p$. Entonces

$$\frac{\partial}{\partial \Sigma} (x \Sigma^{-1} x) = -\Sigma^{-1} x x' \Sigma^{-1}. \quad (74)$$

Definición 24 Sea $A \in \mathbb{R}^{n \times m}$. Se denomina inversa de Moore-Penrose de A a la matriz A^+ que satisface:

1. $AA^+ A = A$.
2. $A^+ AA^+ = A^+$.
3. $(A^+ A)' = A^+ A$.
4. $(AA^+)' = AA^+$.

Se puede demostrar que A^+ existe y es única (ver, por ej., Rao [29] (1967)).

El siguiente resultado, obtenido por Magnus [24] (1985: Teorema 1), permite calcular las derivadas de una matriz simétrica respecto a sus autovalores y autovectores.

Teorema 17 Sea $A_0 \in \mathbb{R}^{p \times p}$ una matriz simétrica. Sea \mathbf{x}_0 un autovector normalizado asociado al autovalor simple λ_0 de A_0 .

Entonces existen una función real λ y una función vectorial \mathbf{x} definidas para toda A en algún entorno $B(A_0) \subset \mathbb{R}^{p \times p}$ de A_0 , tales que

$$\begin{aligned} \lambda(A_0) &= \lambda_0; \\ \mathbf{x}(A_0) &= \mathbf{x}_0; \\ A\mathbf{x} &= \lambda\mathbf{x}; \\ \mathbf{x}'\mathbf{x} &= 1, \end{aligned}$$

para $A \in B(A_0)$.

Además las funciones λ y \mathbf{x} son $C^\infty(B(A_0))$, y las derivadas en A_0 están dadas por

$$\frac{\partial \lambda}{\partial (\text{vec } A)'} = \mathbf{x}'_0 \otimes \mathbf{x}_0; \quad (75)$$

$$\frac{\partial \mathbf{x}}{\partial (\text{vec } A)'} = \mathbf{x}'_0 \otimes (\lambda_0 I - A_0)^+. \quad (76)$$

C Demostración de los resultados de consistencia y normalidad asintótica

C.1 Propiedades de la matriz de covarianza del modelo de análisis factorial

Supongamos que $\mathbf{x} \in \mathfrak{R}^p$ satisface el modelo (1) bajo las hipótesis (10), (11) y (14). Luego, si Σ_0 es su matriz de covarianza, se tiene que

$$\Sigma_0 = \Lambda_0 \Lambda_0' + \Psi_0, \quad (77)$$

con $\Lambda_0 \in \mathcal{R}$ y $\Psi_0 \in \mathcal{D}_+(p)$.

Laywley y Maxwell [18] (1971: pág. 89) demuestran que la matriz Σ_0 satisface las siguientes propiedades:

Proposición 4 Sea $A_0 = I + \Lambda_0' \Psi_0^{-1} \Lambda_0$. Entonces:

1.
$$\Sigma_0^{-1} = \Psi_0^{-1} - \Psi_0^{-1} \Lambda_0 A_0^{-1} \Lambda_0' \Psi_0^{-1}. \quad (78)$$

2.
$$\Lambda_0' \Sigma_0^{-1} = A_0^{-1} \Lambda_0' \Psi_0^{-1}. \quad (79)$$

3.
$$|\Sigma_0| = |\Psi_0| |A_0|.$$

Proposición 5 Sea $\Sigma^* = \Psi_0^{-1/2} \Sigma_0 \Psi_0^{-1/2}$ y sean $\theta_1 \geq \dots \geq \theta_p$ sus autovalores. Entonces:

1. $\theta_1 > \dots > \theta_k > 1.$
2. $\theta_{k+1} = \dots = \theta_p = 1.$

Dem. Sean $\mathbf{w}_1, \dots, \mathbf{w}_p$ los autovectores normalizados de Σ^* y sean

$$T = \begin{bmatrix} \theta_1 & 0 \\ \vdots & \vdots \\ 0 & \theta_p \end{bmatrix}; \quad W = [\mathbf{w}_1 \cdots \mathbf{w}_p].$$

Entonces:

$$WT = \Sigma^* W = \left(\Psi_0^{-1/2} \Lambda_0 \Lambda_0' \Psi_0^{-1/2} + I \right) W.$$

Luego

$$\left(\Psi_0^{-1/2} \Lambda_0 \right) \left(\Lambda_0' \Psi_0^{-1/2} \right) W = W (T - I),$$

es decir los autovalores de $\Psi_0^{-1/2} \Lambda_0 \Lambda_0' \Psi_0^{-1/2}$ son $\theta_1 - 1 \geq \dots \geq \theta_p - 1$.

Pero, por Lema 7, como $p > k$, deben ser:

$$\theta_{k+1} = \dots = \theta_p = 1.$$

Además, por Lema 2, como $\text{rg}(\Lambda_0) = k$ y $\Psi_0 \in \mathcal{D}_+(p)$, deben ser:

$$\theta_1 > \dots > \theta_k > 1.$$

C.2 Propiedades de la función F_k

Habíamos visto en los Capítulos 3 y 4 que los estimadores propuestos de los parámetros del modelo de análisis factorial se obtenían minimizando sobre Λ y Ψ la función

$$F_k(\Lambda, \Psi, \tilde{\Sigma}_n) = \ln |\Lambda \Lambda' + \Psi| + \text{tr} \left[\tilde{\Sigma}_n (\Lambda \Lambda' + \Psi)^{-1} \right] - \ln |\tilde{\Sigma}_n|,$$

donde $\tilde{\Sigma}_n$ es un estimador consistente de la matriz de covarianza de las observaciones.

Si hacemos $\Lambda \Lambda' + \Psi = \Sigma$, esta función es equivalente a:

$$F(\Sigma, \tilde{\Sigma}_n) = \text{tr} \left(\tilde{\Sigma}_n \Sigma^{-1} \right) - \ln |\tilde{\Sigma}_n \Sigma^{-1}|, \quad (80)$$

donde $\Sigma \in \mathfrak{R}^{p \times p}$ y $\tilde{\Sigma}_n \in \mathfrak{R}^{p \times p}$ son matrices simétricas definidas positivas.

Veamos algunas propiedades que satisface la función F dada por (80).

Lema 19 Sean $\Sigma \in \text{PSD}(p)$ y $\Sigma^* \in \text{PSD}(p)$.

Sea

$$F(\Sigma, \Sigma^*) = \text{tr}(\Sigma^* \Sigma^{-1}) - \ln |\Sigma^* \Sigma^{-1}|.$$

Entonces:

1. F es continua en Σ y Σ^* .
2. $F(\Sigma^*, \Sigma^*) = p$.

3. Si $\Sigma \neq \Sigma^*$, $F(\Sigma, \Sigma^*) > p$.

4. $F(\Sigma, \Sigma^*)$ se minimiza con respecto a Σ únicamente en $\Sigma = \Sigma^*$, es decir

$$\min_{\Sigma \in PSD(p)} F(\Sigma, \Sigma^*) = F(\Sigma^*, \Sigma^*) = p.$$

Dem.

1. Se deduce de la continuidad de las funciones traza, logaritmo y determinante.

2.

$$\begin{aligned} F(\Sigma^*, \Sigma^*) &= \text{tr}(\Sigma^* \Sigma^{*-1}) - \ln |\Sigma^* \Sigma^{*-1}| \\ &= \text{tr}(I) - \ln |I| = p. \end{aligned} \quad (81)$$

3. Sean $\mu_1 \geq \dots \geq \mu_p$ los autovalores de $\Sigma^{-1/2} \Sigma^* \Sigma^{-1/2}$ y $\mathbf{v}_1, \dots, \mathbf{v}_p$ los correspondientes autovectores normalizados.

Como Σ y Σ^* son matrices definidas positivas, entonces $\mu_i > 0 \quad \forall i$.

Definamos

$$M = \begin{bmatrix} \mu_1 & 0 \\ \vdots & \\ 0 & \mu_p \end{bmatrix}; \quad V = [\mathbf{v}_1, \dots, \mathbf{v}_p],$$

entonces

$$\Sigma^{-1/2} \Sigma^* \Sigma^{-1/2} = VMV', \quad (82)$$

con $VV' = I$.

Por el Lema 6, como Σ es una matriz simétrica definida positiva, $\Sigma^{-1} = \Sigma^{-1/2} \Sigma^{-1/2}$ y por el Lema 7, los autovalores de $\Sigma^* \Sigma^{-1}$ son los mismos que los de $\Sigma^{-1/2} \Sigma^* \Sigma^{-1/2}$.

Luego, por (81)

$$F(\Sigma, \Sigma^*) - F(\Sigma^*, \Sigma^*) = \text{tr}(\Sigma^* \Sigma^{-1}) - \ln |\Sigma^* \Sigma^{-1}| - p.$$

Pero, por (82)

$$|\Sigma^* \Sigma^{-1}| = |\Sigma^{-1/2} \Sigma^* \Sigma^{-1/2}| = |VMV'| = |M| = \prod_{i=1}^p \mu_i;$$

$$\text{tr}(\Sigma^* \Sigma^{-1}) = \text{tr}(\Sigma^{-1/2} \Sigma^* \Sigma^{-1/2}) = \text{tr}(VMV') = \text{tr}(M) = \sum_{i=1}^p \mu_i,$$

y por lo tanto

$$\begin{aligned} F(\Sigma, \Sigma^*) - F(\Sigma^*, \Sigma^*) &= \sum_{i=1}^p \mu_i - \ln \left(\prod_{i=1}^p \mu_i \right) - p \\ &= \sum_{i=1}^p (\mu_i - \ln \mu_i - 1). \end{aligned}$$

Sea

$$g(x) = x - \ln x - 1,$$

luego

$$g'(x) = 1 - \frac{1}{x};$$

$$g''(x) = \frac{1}{x^2}.$$

Además, dicha función g tiene derivada segunda continua en $x > 0$ y satisface:

$$1. \quad g(1) = 0, \quad g'(1) = 0, \quad g''(1) = 1.$$

$$2. \quad g'(x) \begin{cases} < 0 & \text{si } 0 < x < 1 \\ > 0 & \text{si } 1 < x \end{cases}$$

luego $g(x)$ tiene un mínimo absoluto en $x = 1$.

Por lo tanto

$$F(\Sigma, \Sigma^*) - F(\Sigma^*, \Sigma^*) \geq 0.$$

La igualdad ocurre si $\mu_i = 1 \quad \forall i \implies M = I \implies \Sigma^{-1/2} \Sigma^* \Sigma^{-1/2} = VV' = I \implies \Sigma^* = \Sigma$.

4. Se deduce de 2. y 3.

Lema 20 Si $(\Sigma_n)_{n \geq 1} \in \mathfrak{R}^{p \times p}$ y $(\Sigma_n^*)_{n \geq 1} \in \mathfrak{R}^{p \times p}$ son sucesiones de matrices simétricas no singulares tales que $\|\Sigma_n\| \xrightarrow{n \rightarrow \infty} \infty$, Σ_n^* es definida positiva $\forall n$ y $\Sigma_n^* \xrightarrow{n \rightarrow \infty} \Sigma_0$, siendo Σ_0 una matriz simétrica definida positiva, entonces

$$F(\Sigma_n, \Sigma_n^*) \xrightarrow{n \rightarrow \infty} \infty,$$

donde F es la función dada en (80).

Dem. Sean $\mu_1^n \geq \dots \geq \mu_p^n$ los autovalores de $\Sigma_n^* \Sigma_n^{-1}$. Entonces haciendo un desarrollo análogo al de la demostración del Lema 19 se obtiene que:

$$F(\Sigma_n, \Sigma_n^*) = \sum_{i=1}^p (\mu_i^n - \ln \mu_i^n).$$

Por Lema 7 los autovalores de $\Sigma_n^* \Sigma_n^{-1}$ son los mismos que los de $\Sigma_n^{-1} \Sigma_n^*$, entonces por el Lema 11, $\mu_p^n \xrightarrow{n \rightarrow \infty} 0$.

Luego

$$F(\Sigma_n, \Sigma_n^*) \geq p (\mu_p^n - \ln \mu_p^n) \xrightarrow{n \rightarrow \infty} \infty.$$

C.3 Obtención de los estimadores de los parámetros del modelo de análisis factorial

Como una extensión de las Demostraciones de Lawley y Maxwell [18] (1971: Cap. 4) y Jöreskog [17] (1967) obtendremos los resultados de las Proposiciones 1 a 3 y sus Corolarios.

Demostración de la Proposición 1

Como $(\hat{\Lambda}, \hat{\Psi}) = \arg \min_{(\Lambda, \Psi) \in \mathcal{R} \times \mathcal{D}_+(p)} F_k(\Lambda, \Psi, \tilde{\Sigma})$, entonces $\hat{\Lambda}$ y $\hat{\Psi}$ deben satisfacer:

$$\left. \frac{\partial F_k}{\partial \Lambda} \right|_{(\Lambda, \Psi) = (\hat{\Lambda}, \hat{\Psi})} = \mathbf{0}; \quad (83)$$

$$\left. \frac{\partial F_k}{\partial \Psi} \right|_{(\Lambda, \Psi) = (\hat{\Lambda}, \hat{\Psi})} = \mathbf{0}. \quad (84)$$

Por (72) con $w = w(\Sigma)$ y $\Sigma = \Sigma(\Lambda, \Psi) = \Lambda \Lambda' + \Psi$, resulta:

$$\frac{\partial F_k}{\partial \Lambda} = \sum_{\gamma} \sum_{\zeta} \left(\frac{\partial F_k}{\partial \Sigma} \right)_{\gamma \zeta} \frac{\partial \sigma_{\gamma \zeta}}{\partial \Lambda}; \quad (85)$$

$$\frac{\partial F_k}{\partial \psi_{ii}} = \sum_{\gamma} \sum_{\zeta} \left(\frac{\partial F_k}{\partial \Sigma} \right)_{\gamma \zeta} \frac{\partial \sigma_{\gamma \zeta}}{\partial \psi_{ii}}. \quad (86)$$

Usando (70) con $A = I$, $X = \Lambda$,

$$\frac{\partial \sigma_{\gamma \zeta}}{\partial \Lambda} = K^{\gamma \zeta} \Lambda + (K^{\gamma \zeta})' \Lambda = [K^{\gamma \zeta} + (K^{\gamma \zeta})'] \Lambda. \quad (87)$$

Por (69)

$$\frac{\partial \sigma_{\gamma\zeta}}{\partial \Psi} = K^{\gamma\zeta} \implies \frac{\partial \sigma_{\gamma\zeta}}{\partial \Psi_{ii}} = \left(K^{\gamma\zeta} \right)_{ii}. \quad (88)$$

Por (66)

$$\frac{\partial \ln |\Sigma|}{\partial \Sigma} = \Sigma^{-t} = \Sigma^{-1}. \quad (89)$$

Por (71) con $A = \tilde{\Sigma}$, $B = I$

$$\frac{\partial \tilde{\Sigma} \Sigma^{-1}}{\partial \Sigma} = -\Sigma^{-t} \tilde{\Sigma} K I \Sigma^{-t} = -\Sigma^{-1} \tilde{\Sigma} K \Sigma^{-1}. \quad (90)$$

Luego, por (72), con $w(Y) = \text{tr } Y$, $Y(\Sigma) = \tilde{\Sigma} \Sigma^{-1}$ y (65)

$$\frac{\partial \text{tr} \left(\tilde{\Sigma} \Sigma^{-1} \right)}{\partial \Sigma} = - \sum_{\gamma} \sum_{\zeta} \delta_{\gamma\zeta} \Sigma^{-1} \tilde{\Sigma} K^{\gamma\zeta} \Sigma^{-1} = -\Sigma^{-1} \tilde{\Sigma} \Sigma^{-1}. \quad (91)$$

Por lo tanto, por (89) y (91)

$$\begin{aligned} \frac{\partial F_k}{\partial \Sigma} &= \frac{\partial \ln |\Sigma|}{\partial \Sigma} + \frac{\partial \text{tr} \left(\tilde{\Sigma} \Sigma^{-1} \right)}{\partial \Sigma} \\ &= \Sigma^{-1} - \Sigma^{-1} \tilde{\Sigma} \Sigma^{-1} \\ &= \Sigma^{-1} \left(\Sigma - \tilde{\Sigma} \right) \Sigma^{-1}. \end{aligned} \quad (92)$$

Reemplazando (92) y (87) en (85), y por (67) y (68) resulta:

$$\begin{aligned} \frac{\partial F_k}{\partial \Lambda} &= \sum_{\gamma} \sum_{\zeta} \left(\Sigma^{-1} \left(\Sigma - \tilde{\Sigma} \right) \Sigma^{-1} \right)_{\gamma\zeta} \left[K^{\gamma\zeta} + (K^{\gamma\zeta})' \right] \Lambda \\ &= \left\{ \Sigma^{-1} \left(\Sigma - \tilde{\Sigma} \right) \Sigma^{-1} + \left[\Sigma^{-1} \left(\Sigma - \tilde{\Sigma} \right) \Sigma^{-1} \right]' \right\} \Lambda \\ &= 2 \Sigma^{-1} \left(\Sigma - \tilde{\Sigma} \right) \Sigma^{-1} \Lambda. \end{aligned}$$

Reemplazando (92) y (88) en (86), y por (67) y (68) resulta:

$$\frac{\partial F_k}{\partial \psi_{ii}} = \sum_{\gamma} \sum_{\zeta} \left(\Sigma^{-1} \left(\Sigma - \tilde{\Sigma} \right) \Sigma^{-1} \right)_{\gamma\zeta} \left(K^{\gamma\zeta} \right)_{ii} = \left(\Sigma^{-1} \left(\Sigma - \tilde{\Sigma} \right) \Sigma^{-1} \right)_{ii}.$$

Resumiendo, por (83) y (84) $\hat{\Lambda}$ y $\hat{\Psi}$ deben satisfacer

$$2 \Sigma^{-1} \left(\Sigma - \tilde{\Sigma} \right) \Sigma^{-1} \Lambda \Big|_{(\Lambda, \Psi) = (\hat{\Lambda}, \hat{\Psi})} = \mathbf{0}; \quad (93)$$

$$\text{diag} \left[\Sigma^{-1} (\Sigma - \tilde{\Sigma}) \Sigma^{-1} \right] \Big|_{(\Lambda, \Psi) = (\hat{\Lambda}, \hat{\Psi})} = \mathbf{0}. \quad (94)$$

Sea

$$\hat{\Sigma} = \hat{\Lambda} \hat{\Lambda}' + \hat{\Psi}. \quad (95)$$

Por (93),

$$\hat{\Sigma}^{-1} (\hat{\Sigma} - \tilde{\Sigma}) \hat{\Sigma}^{-1} \hat{\Lambda} = \mathbf{0}$$

y premultiplicando por $\hat{\Sigma}$

$$(\hat{\Sigma} - \tilde{\Sigma}) \hat{\Sigma}^{-1} \hat{\Lambda} = \mathbf{0}. \quad (96)$$

Sea

$$\hat{A} = I + \hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda}. \quad (97)$$

Haciendo uso de la identidad (78) de la Proposición 4

$$\begin{aligned} \hat{\Sigma}^{-1} \hat{\Lambda} &= \hat{\Psi}^{-1} \hat{\Lambda} - \hat{\Psi}^{-1} \hat{\Lambda} \hat{A}^{-1} \hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda} \\ &= \hat{\Psi}^{-1} \hat{\Lambda} (I - \hat{A}^{-1} \hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda}) \\ &= \hat{\Psi}^{-1} \hat{\Lambda} [I - \hat{A}^{-1} (\hat{A} - I)] = \hat{\Psi}^{-1} \hat{\Lambda} \hat{A}^{-1}. \end{aligned}$$

Usando esta ecuación, (96) puede escribirse en la forma

$$(\hat{\Sigma} - \tilde{\Sigma}) \hat{\Psi}^{-1} \hat{\Lambda} \hat{A}^{-1} = \mathbf{0}. \quad (98)$$

Postmultiplicando por \hat{A} y usando (19)

$$\begin{aligned} (\hat{\Lambda} \hat{\Lambda}' + \hat{\Psi} - \tilde{\Sigma}) \hat{\Psi}^{-1} \hat{\Lambda} = \mathbf{0} &\implies \hat{\Lambda} \hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda} + \hat{\Lambda} - \tilde{\Sigma} \hat{\Psi}^{-1} \hat{\Lambda} = \mathbf{0} \\ &\implies \hat{\Lambda} (\hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda} + I) = \tilde{\Sigma} \hat{\Psi}^{-1} \hat{\Lambda}, \end{aligned}$$

que se simplifica a

$$\tilde{\Sigma} \hat{\Psi}^{-1} \hat{\Lambda} = \hat{\Lambda} \hat{A} = \hat{\Lambda} (I + \hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda}),$$

con lo que obtenemos la ecuación (23).

Por (78)

$$\begin{aligned} \hat{\Sigma}^{-1} (\hat{\Sigma} - \tilde{\Sigma}) \hat{\Sigma}^{-1} &= \\ &= \hat{\Psi}^{-1} (I - \hat{\Lambda} \hat{A}^{-1} \hat{\Lambda}') \hat{\Psi}^{-1} (\hat{\Sigma} - \tilde{\Sigma}) \hat{\Psi}^{-1} (I - \hat{\Lambda} \hat{A}^{-1} \hat{\Lambda}') \hat{\Psi}^{-1}, \end{aligned}$$

y por (98)

$$\begin{aligned}
& \widehat{\Sigma}^{-1} (\widehat{\Sigma} - \widetilde{\Sigma}) \widehat{\Sigma}^{-1} = \\
& = \widehat{\Psi}^{-1} (I - \widehat{\Lambda} \widehat{A}^{-1} \widehat{\Lambda}') \widehat{\Psi}^{-1} [(\widehat{\Sigma} - \widetilde{\Sigma}) \widehat{\Psi}^{-1} - (\widehat{\Sigma} - \widetilde{\Sigma}) \widehat{\Psi}^{-1} \widehat{\Lambda} \widehat{A}^{-1} \widehat{\Lambda}' \widehat{\Psi}^{-1}] \\
& = \widehat{\Psi}^{-1} (\widehat{\Sigma} - \widetilde{\Sigma}) \widehat{\Psi}^{-1} - \widehat{\Psi}^{-1} \widehat{\Lambda} \widehat{A}^{-1} \widehat{\Lambda}' \widehat{\Psi}^{-1} (\widehat{\Sigma} - \widetilde{\Sigma}) \widehat{\Psi}^{-1} \\
& = \widehat{\Psi}^{-1} (\widehat{\Sigma} - \widetilde{\Sigma}) \widehat{\Psi}^{-1}.
\end{aligned}$$

Luego,

$$\text{diag} [\widehat{\Sigma}^{-1} (\widehat{\Sigma} - \widetilde{\Sigma}) \widehat{\Sigma}^{-1}] = \text{diag} [\widehat{\Psi}^{-1} (\widehat{\Sigma} - \widetilde{\Sigma}) \widehat{\Psi}^{-1}].$$

Como $\widehat{\Psi} \in \mathcal{D}(p)$ tal que sus elementos diagonales son no nulos,

$$(\widehat{\Psi}^{-1} (\widehat{\Sigma} - \widetilde{\Sigma}) \widehat{\Psi}^{-1})_{ii} = (\widehat{\Psi}^{-1})_{ii}^2 (\widehat{\Sigma} - \widetilde{\Sigma})_{ii},$$

y por lo tanto,

$$\text{diag} [\widehat{\Sigma}^{-1} (\widehat{\Sigma} - \widetilde{\Sigma}) \widehat{\Sigma}^{-1}] = 0 \implies \text{diag} (\widehat{\Sigma} - \widetilde{\Sigma}) = 0,$$

o equivalentemente,

$$\text{diag} (\widehat{\Psi}) = \text{diag} (\widetilde{\Sigma} - \widehat{\Lambda} \widehat{\Lambda}'),$$

con lo que obtenemos la ecuación (24).

Veamos que si $\widehat{\Lambda}$ y $\widehat{\Psi}$ satisfacen las ecuaciones (23) y (24), entonces

$$F_k (\widehat{\Lambda}, \widehat{\Psi}, \widetilde{\Sigma}) = \min_{(\Lambda, \Psi) \in \mathcal{R} \times \mathcal{D}_+(p)} F_k (\Lambda, \Psi, \widetilde{\Sigma}).$$

Por Lema 19, basta ver $\widetilde{\Sigma} = \widehat{\Sigma}$, o equivalentemente que $\widetilde{\Sigma} \widehat{\Sigma}^{-1} = I_p$.

Por la identidad (78), (95) y (97)

$$\widetilde{\Sigma} \widehat{\Sigma}^{-1} = \widetilde{\Sigma} \widehat{\Psi}^{-1} - \widetilde{\Sigma} \widehat{\Psi}^{-1} \widehat{\Lambda} \widehat{A}^{-1} \widehat{\Lambda}' \widehat{\Psi}^{-1},$$

y como se satisface (23)

$$\begin{aligned}
\widetilde{\Sigma} \widehat{\Sigma}^{-1} &= \widetilde{\Sigma} \widehat{\Psi}^{-1} - \widehat{\Lambda} \widehat{A} \widehat{A}^{-1} \widehat{\Lambda}' \widehat{\Psi}^{-1} \\
&= \widetilde{\Sigma} \widehat{\Psi}^{-1} - \widehat{\Lambda} \widehat{\Lambda}' \widehat{\Psi}^{-1} \\
&= (\widetilde{\Sigma} - \widehat{\Lambda} \widehat{\Lambda}) \widehat{\Psi}^{-1}.
\end{aligned}$$

Luego, como se satisface (24)

$$\begin{aligned} (\tilde{\Sigma}\hat{\Sigma}^{-1})_{ii} &= (\tilde{\Sigma} - \hat{\Lambda}\hat{\Lambda})_{ii} (\hat{\Psi}^{-1})_{ii} \\ &= \hat{\Psi}_{ii}\hat{\Psi}_{ii}^{-1} = 1; \end{aligned}$$

$$(\tilde{\Sigma}\hat{\Sigma}^{-1}\hat{\Psi})_{ij} = (\tilde{\Sigma} - \hat{\Lambda}\hat{\Lambda})_{ij} \hat{\Psi}_{ij} = 0,$$

es decir, $(\tilde{\Sigma}\hat{\Sigma}^{-1})_{ij} = \delta_{ij}$.

Demostración de la Proposición 2

Por la Proposición 1, basta ver que Λ^* y $\hat{\Psi}$ satisfacen las ecuaciones (23) y (24).

Como $(\hat{\Lambda}, \hat{\Psi}) = \arg \min_{(\Lambda, \Psi) \in \mathcal{R} \times \mathcal{D}_+(p)} F_k(\Lambda, \Psi, \tilde{\Sigma})$, entonces, por Proposición 1,

$$\tilde{\Sigma}\hat{\Psi}^{-1}\hat{\Lambda} = \hat{\Lambda} (I + \hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda}).$$

Postmultiplicando por M , y por ser M ortogonal

$$\begin{aligned} \tilde{\Sigma}\hat{\Psi}^{-1}\hat{\Lambda}M &= \hat{\Lambda} (MM' + \hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda}) M \\ &= \hat{\Lambda}MM' (I + MM'\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda}MM') M, \end{aligned}$$

que se simplifica a:

$$\begin{aligned} \tilde{\Sigma}\hat{\Psi}^{-1}\Lambda^* &= \Lambda^* (MM' + M\Lambda^{*'}\hat{\Psi}^{-1}\Lambda^*M') M \\ &= \Lambda^*MM' (I + \Lambda^{*'}\hat{\Psi}^{-1}\Lambda^*) M'M \\ &= \Lambda^* (I + \Lambda^{*'}\hat{\Psi}^{-1}\Lambda^*), \end{aligned}$$

y por lo tanto se satisface la ecuación (23).

Por ser M ortogonal

$$\Lambda^*\Lambda^{*'} = \hat{\Lambda}MM'\hat{\Lambda}' = \hat{\Lambda}\hat{\Lambda}'.$$

Luego, por la ecuación (24) de la Proposición 1,

$$\text{diag}(\hat{\Psi}) = \text{diag}(\tilde{\Sigma} - \hat{\Lambda}\hat{\Lambda}') = \text{diag}(\tilde{\Sigma} - \Lambda^*\Lambda^{*'}),$$

y por lo tanto se satisface (24).

Demostración del Corolario 1

Sean $\tilde{\Lambda}$ y $\tilde{\Psi}$ tales que satisfacen las ecuaciones (23) y (24), pero $\tilde{\Lambda}'\tilde{\Psi}^{-1}\tilde{\Lambda} \notin \mathcal{D}^*(k)$.

Como $\text{rg}(\tilde{\Lambda}) = k$ y $\tilde{\Psi} \in \mathcal{D}_+(p)$, $\tilde{\Lambda}'\tilde{\Psi}^{-1}\tilde{\Lambda}$ es una matriz simétrica definida positiva, y por Teorema 16, existen matrices $T \in \mathfrak{R}^{k \times k}$ ortogonal y $D \in \mathcal{D}^*(k)$ tales que $\text{rg}(T) = k$ y

$$T'\tilde{\Lambda}'\tilde{\Psi}^{-1}\tilde{\Lambda}T = D.$$

Sea $\Lambda^* = \tilde{\Lambda}T$. Entonces, por construcción, Λ^* satisface (25), y por Proposición 2, $F_k(\Lambda^*, \tilde{\Psi}, \tilde{\Sigma}) = \min_{(\Lambda, \Psi)} F_k(\Lambda, \Psi, \tilde{\Sigma})$ como queríamos probar.

Demostración de la Proposición 3

i) Como $\tilde{\Theta}$ es la matriz cuya diagonal contiene a los k autovalores más grandes de $\tilde{\Sigma}^* = \tilde{\Sigma}^*(\tilde{\Psi})$ y $\tilde{\Omega}$ es la matriz cuyas columnas son los correspondientes k autovectores normalizados, entonces

$$\tilde{\Sigma}^*\tilde{\Omega} = \tilde{\Omega}\tilde{\Theta}.$$

Luego,

$$\begin{aligned} \tilde{\Sigma}^*\tilde{\Omega}(\tilde{\Theta} - I)^{1/2} &= \tilde{\Omega}\tilde{\Theta}(\tilde{\Theta} - I)^{1/2} \\ &= \tilde{\Omega}(\tilde{\Theta} - I)^{1/2}\tilde{\Theta}, \end{aligned}$$

por ser $\tilde{\Theta}$ y $(\tilde{\Theta} - I)$ matrices diagonales, es decir

$$\left(\tilde{\Psi}^{-1/2}\tilde{\Sigma}\tilde{\Psi}^{-1/2}\right)\left[\tilde{\Omega}(\tilde{\Theta} - I)^{1/2}\right] = \left[\tilde{\Omega}(\tilde{\Theta} - I)^{1/2}\right]\tilde{\Theta}. \quad (99)$$

Por otra parte, como $\tilde{\Lambda} = \arg \min_{\Psi \in \mathcal{A}(\tilde{\Psi})} F_k(\Lambda, \Psi, \tilde{\Sigma})$, haciendo un desarrollo análogo al de las demostraciones de la Proposición 1 y el Corolario 1, se puede ver que $\tilde{\Lambda}$ debe satisfacer

$$\tilde{\Sigma}\tilde{\Psi}^{-1}\tilde{\Lambda} = \tilde{\Lambda}(I + \tilde{\Lambda}'\tilde{\Psi}^{-1}\tilde{\Lambda});$$

$$\tilde{\Lambda}'\tilde{\Psi}^{-1}\tilde{\Lambda} = \tilde{\Delta},$$

donde $\tilde{\Delta} \in \mathcal{D}^*(k)$.

Luego,

$$\tilde{\Sigma}\Psi^{-1}\tilde{\Lambda} = \tilde{\Lambda} (I + \tilde{\Delta}).$$

Premultiplicando por $\Psi^{-1/2}$

$$\left(\Psi^{-1/2}\tilde{\Sigma}\Psi^{-1/2}\right)\Psi^{-1/2}\tilde{\Lambda} = \Psi^{-1/2}\tilde{\Lambda} (I + \tilde{\Delta}). \quad (100)$$

Como $\tilde{\Delta}$ es una matriz diagonal, las columnas de $\Psi^{-1/2}\tilde{\Lambda}$ son los autovectores de la matriz $\tilde{\Sigma}^*$ y los elementos diagonales de $I + \tilde{\Delta}$ sus correspondientes autovalores. Luego, por (99) y (100)

$$\tilde{\Omega} (\tilde{\Theta} - I)^{1/2} = \Psi^{-1/2}\tilde{\Lambda}; \quad (101)$$

$$I + \tilde{\Delta} = \tilde{\Theta},$$

y por lo tanto, de la primera ecuación se obtiene que

$$\tilde{\Lambda} = \Psi^{1/2}\tilde{\Omega} (\tilde{\Theta} - I)^{1/2}$$

ii) Se obtiene como corolario de la Proposición 1 y de la parte i).

Además,

$$\begin{aligned} \hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda} &= (\tilde{\Theta} - I)^{1/2}\tilde{\Omega}'\hat{\Psi}^{1/2}\hat{\Psi}^{-1}\hat{\Psi}^{1/2}\tilde{\Omega} (\tilde{\Theta} - I)^{1/2} \\ &= \tilde{\Theta} (\hat{\Psi}) - I. \end{aligned}$$

Notación: Por simplicidad de notación, en la demostración del Corolario 2, notaremos con $\tilde{\Sigma}$ a $\tilde{\Sigma}_n$, con $\hat{\Lambda}$ a $\hat{\Lambda}_n$ y con $\hat{\Psi}$ a $\hat{\Psi}_n$.

Demostración del Corolario 2

Sean

$$\begin{aligned} \hat{\Sigma} &= \tilde{\Lambda}\tilde{\Lambda}' + \Psi; \\ \tilde{A} &= I + \tilde{\Lambda}'\Psi^{-1}\tilde{\Lambda}, \end{aligned}$$

donde $\tilde{\Lambda}$ está dada por la ecuación (28).

Para probar (29) primero calcularemos el determinante y la inversa de $\Psi^{-1/2}\hat{\Sigma}\Psi^{-1/2}$.

Por (101) y el Lema 7

$$\begin{aligned}
\left| \Psi^{-1/2} \widehat{\Sigma} \Psi^{-1/2} \right| &= \left| \Psi^{-1/2} (\widetilde{\Lambda} \widetilde{\Lambda}' + \Psi) \Psi^{-1/2} \right| = \left| \Psi^{-1/2} \widetilde{\Lambda} \widetilde{\Lambda}' \Psi^{-1/2} + I \right| \\
&= \left| I + \widetilde{\Omega} (\widetilde{\Theta} - I)^{1/2} (\widetilde{\Theta} - I)^{1/2} \widetilde{\mathcal{L}} \right| \\
&= \left| I + \widetilde{\Omega} (\widetilde{\Theta} - I) \widetilde{\mathcal{L}} \right| = \left| I + \widetilde{\Omega} \widetilde{\mathcal{L}} (\widetilde{\Theta} - I) \right| \\
&= \left| I + \widetilde{\Theta} - I \right| = \left| \widetilde{\Theta} \right| = \prod_{m=1}^k \widetilde{\theta}_m.
\end{aligned}$$

Por la Proposición 4 y (101)

$$\begin{aligned}
\Psi^{1/2} \widehat{\Sigma}^{-1} \Psi^{1/2} &= \Psi^{1/2} \left(\Psi^{-1} - \Psi^{-1} \widetilde{\Lambda} \widetilde{A}^{-1} \widetilde{\Lambda}' \Psi^{-1} \right) \Psi^{1/2} \\
&= \Psi^{1/2} \Psi^{-1} \Psi^{1/2} - \Psi^{1/2} \Psi^{-1} \widetilde{\Lambda} \widetilde{A}^{-1} \widetilde{\Lambda}' \Psi^{-1} \Psi^{1/2} \\
&= I - \Psi^{-1/2} \widetilde{\Lambda} \widetilde{A}^{-1} \widetilde{\Lambda}' \Psi^{-1/2} \\
&= I - \Psi^{-1/2} \widetilde{\Lambda} \widetilde{\Theta}^{-1} \widetilde{\Lambda}' \Psi^{-1/2}.
\end{aligned}$$

Además, tenemos que

$$\left| \widetilde{\Sigma} \right| = \left| \Psi^{1/2} \widetilde{\Sigma}^* \Psi^{1/2} \right| = \left| \widetilde{\Sigma}^* \right| |\Psi| = |\Psi| \left(\prod_{m=1}^p \widetilde{\theta}_m \right).$$

Por lo tanto:

$$\left| \widehat{\Sigma} \widetilde{\Sigma}^{-1} \right| = \frac{|\Psi^{-1/2}| \left| \widehat{\Sigma} \right| |\Psi^{-1/2}|}{|\Psi^{-1/2}| \left| \widetilde{\Sigma} \right| |\Psi^{-1/2}|} = \frac{|\Psi^{-1/2} \widehat{\Sigma} \Psi^{-1/2}|}{|\Psi^{-1/2} \widetilde{\Sigma} \Psi^{-1/2}|} = \frac{\prod_{m=1}^k \widetilde{\theta}_m}{\prod_{l=1}^p \widetilde{\theta}_l} = \frac{1}{\prod_{l=k+1}^p \widetilde{\theta}_l}.$$

Luego

$$\ln \left| \widehat{\Sigma} \right| - \ln \left| \widetilde{\Sigma} \right| = \ln \left(\left| \widehat{\Sigma} \widetilde{\Sigma}^{-1} \right| \right) = -\ln \left(\prod_{m=k+1}^p \widetilde{\theta}_m \right) = -\sum_{m=k+1}^p \ln \widetilde{\theta}_m. \quad (102)$$

Por (100)

$$\begin{aligned}
\operatorname{tr} \left(\tilde{\Sigma} \hat{\Sigma}^{-1} \right) &= \operatorname{tr} \left(\tilde{\Sigma} \Psi^{-1/2} \Psi^{1/2} \hat{\Sigma}^{-1} \Psi^{1/2} \Psi^{-1/2} \right) \\
&= \operatorname{tr} \left(\Psi^{-1/2} \tilde{\Sigma} \Psi^{-1/2} \Psi^{1/2} \hat{\Sigma}^{-1} \Psi^{1/2} \right) \\
&= \operatorname{tr} \left[\Psi^{-1/2} \tilde{\Sigma} \Psi^{-1/2} \left(I - \Psi^{-1/2} \tilde{\Lambda} \tilde{\Theta}^{-1} \tilde{\Lambda}' \Psi^{-1/2} \right) \right] \\
&= \operatorname{tr} \left(\Psi^{-1/2} \tilde{\Sigma} \Psi^{-1/2} \right) - \operatorname{tr} \left(\Psi^{-1/2} \tilde{\Lambda} \tilde{\Lambda}'^{-1} \tilde{\Theta}^{-1} \tilde{\Lambda}' \Psi^{-1/2} \right) \\
&= \operatorname{tr} \left(\Psi^{-1/2} \tilde{\Sigma} \Psi^{-1/2} \right) - \operatorname{tr} \left(\Psi^{-1/2} \tilde{\Lambda} \tilde{\Theta} \tilde{\Theta}^{-1} \tilde{\Lambda}' \Psi^{-1/2} \right) \\
&= \operatorname{tr} \left(\Psi^{-1/2} \tilde{\Sigma} \Psi^{-1/2} \right) - \operatorname{tr} \left(\tilde{\Lambda}' \Psi^{-1} \tilde{\Lambda} \right) \\
&= \sum_{m=1}^p \tilde{\theta}_m - \sum_{m=1}^k (\tilde{\theta}_m - 1) = \sum_{m=k+1}^p \tilde{\theta}_m + k.
\end{aligned} \tag{103}$$

Sustituyendo (102) y (103) en (21) obtenemos (29).

Veamos que $f_k(\Psi)$ puede expresarse en términos de los $p - k$ autovalores más pequeños $\tilde{\theta}_{k+1} \cdots \tilde{\theta}_p$ solamente.

Sea \mathcal{M} cualquier conjunto de k autovalores $\tilde{\theta}_1^* \cdots \tilde{\theta}_k^*$ de $\tilde{\Sigma}^* = \tilde{\Sigma}^*(\Psi)$ y $\overline{\mathcal{M}}$ el conjunto complementario consistente en los otros $p - k$ autovalores.

Entonces en forma análoga a lo hecho anteriormente, se obtiene que:

$$f_k(\Psi) = \sum \left(\tilde{\theta}_i^* - \ln \tilde{\theta}_i^* \right) - (p - k),$$

donde la suma se hace sobre todos los $\tilde{\theta}_i^*$ en $\overline{\mathcal{M}}$.

Ahora supongamos que intercambiamos $\tilde{\theta}_{k+b}^*$ con $\tilde{\theta}_a^*$ en (29), donde $a \leq k$ y $b \geq 1$. Al intercambiar los autovalores, la diferencia entre el $f_k(\Psi)$ viejo y el nuevo es:

$$d = \left(\tilde{\theta}_a^* - \ln \tilde{\theta}_a^* \right) - \left(\tilde{\theta}_{k+b}^* - \ln \tilde{\theta}_{k+b}^* \right).$$

Por Lema 5, si $\theta_1 \geq \cdots \geq \theta_p$ son los autovalores de $\Sigma_0^* = \Psi_0^{-1/2} \Sigma_0 \Psi_0^{-1/2}$, los mismos verifican que

$$\theta_1 > \cdots > \theta_k > 1 = \theta_{k+1} = \cdots = \theta_p.$$

Como $\Psi \in C$ y $\tilde{\Sigma} \xrightarrow{c.s.} \Sigma_0$, entonces por Teorema 17, por continuidad de los autovalores

$$\tilde{\theta}_i \geq 1 \quad \text{para } 1 \leq i \leq p.$$

Como $\tilde{\theta}_a^* > \tilde{\theta}_{k+b}^*$ y $g(x) = x - \ln x$ es una función monótona creciente de x para $x > 1$, ya que $g'(x) = 1 - 1/x > 0$ para $x > 1$, d será siempre positiva.

Por lo tanto, el uso de cualquier conjunto de $p - k$ autovalores en (29) distinto de los $p - k$ más pequeños producirá un mayor valor de $f_k(\Psi)$.

C.4 Demostración de la consistencia de los estimadores de los parámetros del modelo de análisis factorial

Demostración del Teorema 6

Vamos a probar que:

i) $F_k(\Lambda, \Psi, \Sigma_0)$ se minimiza con respecto a Λ y Ψ únicamente en $\Lambda = \Lambda_0$ y $\Psi = \Psi_0$.

ii) Si

$$\hat{\Sigma}_n = \hat{\Lambda}_n \hat{\Lambda}'_n + \hat{\Psi}_n, \quad (104)$$

entonces, $\hat{\Sigma}_n \xrightarrow{c.s.} \Sigma_0$.

iii) $\hat{\Psi}_n \xrightarrow{c.s.} \Psi_0$ y $\hat{\Lambda}_n \xrightarrow{c.s.} \Lambda_0$.

Veamos i)

Sea

$$F(\Sigma, \Sigma^*) = \text{tr}(\Sigma^* \Sigma^{-1}) - \ln |\Sigma^* \Sigma^{-1}|,$$

donde $\Sigma \in \text{PSD}(p)$ y $\Sigma^* \in \text{PSD}(p)$, $\Sigma = \Lambda \Lambda' + \Psi$.

Entonces

$$F_k(\Lambda, \Psi, \Sigma_0) = F(\Sigma, \Sigma_0). \quad (105)$$

Por Lema 19, $F(\Sigma, \Sigma_0)$ se minimiza con respecto a Σ únicamente en $\Sigma = \Sigma_0$. Luego, por identificabilidad del modelo y por (105), $F_k(\Lambda, \Psi, \Sigma_0)$ se minimiza con respecto a Λ y Ψ únicamente en $\Lambda = \Lambda_0$ y $\Psi = \Psi_0$.

Veamos ii)

Por construcción

$$F(\hat{\Sigma}_n, \tilde{\Sigma}_n) = \min_{\Sigma} F(\Sigma, \tilde{\Sigma}_n). \quad (106)$$

Sea

$$A = \left\{ \omega : \tilde{\Sigma}_n(\omega) \xrightarrow{n \rightarrow \infty} \Sigma_0 \right\},$$

y consideremos $\omega \in A$.

Entonces, por Lema 19

$$\lim_{n \rightarrow \infty} F(\Sigma_0, \tilde{\Sigma}_n(\omega)) = F(\Sigma_0, \Sigma_0) = p. \quad (107)$$

Sea

$$B = \left\{ \omega : \hat{\Sigma}_n(\omega) \xrightarrow{n \rightarrow \infty} \Sigma_0 \right\}.$$

Queremos probar que $A \cap B^C = \emptyset$.

Consideremos $\omega \in A \cap B^C$.

Si $(\hat{\Sigma}_n(\omega))_{n \geq 1}$ es una sucesión no acotada, por Lema 20

$$F(\hat{\Sigma}_n(\omega), \tilde{\Sigma}_n(\omega)) \xrightarrow{n \rightarrow \infty} \infty,$$

que junto con (107), contradice (106).

Si $(\hat{\Sigma}_n(\omega))_{n \geq 1}$ es una sucesión acotada, entonces existe una subsucesión

$(\hat{\Sigma}_{n_k}(\omega))_{k \geq 1}$ tal que $\hat{\Sigma}_{n_k}(\omega) \xrightarrow{k \rightarrow \infty} \Sigma_1$, con $\Sigma_1 \neq \Sigma_0$.

Luego, por Lema 19

$$F(\hat{\Sigma}_{n_k}(\omega), \tilde{\Sigma}_n(\omega)) \xrightarrow{k \rightarrow \infty} F(\Sigma_1, \Sigma_0) > p,$$

que junto con (107), contradice (106).

Por lo tanto, debe ser $A \cap B^C = \emptyset$.

Veamos iii)

Sean

$$C = \left\{ \omega \in B : \hat{\Lambda}_n(\omega) \xrightarrow{n \rightarrow \infty} \Lambda_0 \right\};$$

$$D = \left\{ \omega \in B : \hat{\Psi}_n(\omega) \xrightarrow{n \rightarrow \infty} \Psi_0 \right\}.$$

Veamos que $P(C^C \cup D^C) = 0$.

Supongamos que $P(C^C \cup D^C) > 0$ y sea $\omega \in C^C \cup D^C$.

Como $\hat{\Sigma}_n(\omega) \xrightarrow{n \rightarrow \infty} \Sigma_0$, entonces $\exists M < \infty$ tal que

$$\|\hat{\Sigma}_n(\omega)\| = \|\hat{\Lambda}_n(\omega)\hat{\Lambda}'_n(\omega) + \hat{\Psi}_n(\omega)\| \leq M \quad \forall n.$$

Por Lema 2, $\hat{\Lambda}_n(\omega)\hat{\Lambda}'_n(\omega)$ es una matriz semidefinida positiva. Luego, por Lema 10

$$\|\hat{\Psi}_n(\omega)\| \leq \|\hat{\Sigma}_n(\omega)\| \leq M \quad \forall n. \quad (108)$$

Por construcción, $\hat{\Psi}_n(\omega) \in \mathcal{D}_+(p)$, entonces

$$\hat{\Psi}_n(\omega) = \hat{\Sigma}_n(\omega) - \hat{\Lambda}_n(\omega)\hat{\Lambda}'_n(\omega),$$

es una matriz definida positiva.

Luego, por el Lema 10

$$\left\| \widehat{\Lambda}_n(\omega) \widehat{\Lambda}'_n(\omega) \right\| \leq \left\| \widehat{\Sigma}_n(\omega) \right\| \leq M \quad \forall n. \quad (109)$$

Sean $\widehat{\theta}_1^n \geq \dots \geq \widehat{\theta}_p^n$ los autovalores de $\widehat{\Sigma}_n^*(\omega) = \widehat{\Psi}_n^{-1/2}(\omega) \widehat{\Sigma}_n(\omega) \widehat{\Psi}_n^{-1/2}(\omega)$ y $\widehat{\mathbf{w}}_1^n, \dots, \widehat{\mathbf{w}}_p^n$ sus respectivos autovectores normalizados. Definamos

$$\widehat{T}_n(\omega) = \begin{bmatrix} \widehat{\theta}_1^n & 0 \\ & \vdots \\ 0 & \widehat{\theta}_p^n \end{bmatrix}; \quad \widehat{W}_n(\omega) = [\widehat{\mathbf{w}}_1^n \dots \widehat{\mathbf{w}}_p^n].$$

Entonces,

$$\begin{aligned} \widehat{W}_n \widehat{T}_n &= \widehat{\Sigma}_n^* \widehat{W}_n = \widehat{\Psi}_n^{-1/2} \widehat{\Sigma}_n \widehat{\Psi}_n^{-1/2} \widehat{W}_n \\ &= \left[\widehat{\Psi}_n^{-1/2} \left(\widehat{\Lambda}_n \widehat{\Lambda}'_n + \widehat{\Psi}_n \right) \widehat{\Psi}_n^{-1/2} \right] \widehat{W}_n \\ &= \left(\widehat{\Psi}_n^{-1/2} \widehat{\Lambda}_n \widehat{\Lambda}'_n \widehat{\Psi}_n^{-1/2} + I \right) \widehat{W}_n. \end{aligned}$$

Luego,

$$\begin{aligned} \widehat{W}_n \left(\widehat{T}_n - I \right) &= \widehat{\Psi}_n^{-1/2} \widehat{\Lambda}_n \widehat{\Lambda}'_n \widehat{\Psi}_n^{-1/2} \widehat{W}_n; \\ \widehat{T}_n - I &= \left(\widehat{W}_n' \widehat{\Psi}_n^{-1/2} \right) \widehat{\Lambda}_n \widehat{\Lambda}'_n \left(\widehat{\Psi}_n^{-1/2} \widehat{W}_n \right), \end{aligned}$$

y por lo tanto los autovalores de $\widehat{\Lambda}_n \widehat{\Lambda}'_n$ son $\widehat{\theta}_1^n - 1 \geq \dots \geq \widehat{\theta}_p^n - 1$.

Como $\widehat{\Lambda}_n \widehat{\Lambda}'_n$ es una matriz semidefinida positiva, sus autovalores deben ser no negativos. Es decir

$$\widehat{\theta}_1^n \geq \dots \geq \widehat{\theta}_p^n \geq 1. \quad (110)$$

Luego, por (110)

$$\begin{aligned} \left\| \widehat{\Lambda}_n(\omega) \right\| &= \left[\text{tr} \left(\widehat{\Lambda}_n(\omega) \widehat{\Lambda}'_n(\omega) \right) \right]^{1/2} = \left[\sum_{h=1}^p (\theta_h - 1) \right]^{1/2} \\ &\leq \left[\sum_{h=1}^p (\theta_h - 1)^2 \right]^{1/2} = \left[\text{tr} \left(\widehat{\Lambda}_n(\omega) \widehat{\Lambda}'_n(\omega) \right)^2 \right]^{1/2} \\ &= \left\| \widehat{\Lambda}_n(\omega) \widehat{\Lambda}'_n(\omega) \right\|. \end{aligned}$$

Luego, por (109)

$$\left\| \widehat{\Lambda}_n(\omega) \right\| \leq \left\| \widehat{\Lambda}_n(\omega) \widehat{\Lambda}'_n(\omega) \right\| \leq M \quad \forall n. \quad (111)$$

Por lo tanto, si $\omega \in C^C$, por (111), existe una subsucesión $(\widehat{\Lambda}_{n_k}(\omega))_{k \geq 1}$ tal que $\widehat{\Lambda}_{n_k}(\omega) \xrightarrow{k \rightarrow \infty} \Lambda_1$, con $\Lambda_1 \neq \Lambda_0$.

Además, por (104),

$$\widehat{\Psi}_{n_k}(\omega) = \widehat{\Sigma}_{n_k}(\omega) - \widehat{\Lambda}_{n_k}(\omega) \widehat{\Lambda}'_{n_k}(\omega) \xrightarrow{k \rightarrow \infty} \Sigma_0 - \Lambda_1 \Lambda'_1 = \Psi_1.$$

Como $\widehat{\Psi}_{n_k}(\omega) \in \mathcal{D}_+(p)$, entonces $\Psi_1 \in \mathcal{D}_+(p)$, y por lo tanto

$$\Sigma_0 = \Lambda_1 \Lambda'_1 + \Psi_1,$$

lo que contradice la identificabilidad del modelo, ya que, si $\Psi_1 = \Psi_0$, entonces

$$\Lambda_1 \Lambda'_1 + \Psi_0 \neq \Sigma_0.$$

Luego, debe ser $P(C^C \cup D^C) = 0$.

C.5 Demostración de la normalidad asintótica de los estimadores de los parámetros del modelo de análisis factorial

Antes de demostrar el Teorema 11, veamos algunos resultados sobre derivadas de los autovectores y autovalores de una matriz simétrica respecto a los elementos de la misma.

Lema 21 Sea $A \in \mathbb{R}^{p \times p}$ una matriz simétrica con autovalores $\lambda_1, \dots, \lambda_p$ y autovectores respectivos $\mathbf{x}_1, \dots, \mathbf{x}_p$ tales que $\mathbf{x}_i' \mathbf{x}_j = \delta_{ij}$, es decir los autovectores están normalizados.

Supongamos que λ_k es autovalor simple de A .

Sean

$$U_k = [\mathbf{x}_1, \dots, \mathbf{x}_{k-1}, \mathbf{x}_{k+1}, \dots, \mathbf{x}_p];$$

$$\Lambda_k = \begin{bmatrix} \lambda_1 - \lambda_k & 0 & 0 & 0 \\ \vdots & & & \vdots \\ 0 & \lambda_{k-1} - \lambda_k & 0 & 0 \\ 0 & 0 & \lambda_{k+1} - \lambda_k & 0 \\ \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \lambda_p - \lambda_k \end{bmatrix}.$$

Entonces

$$(A - \lambda_k I)^+ = U_k \Lambda_k^{-1} U_k'.$$

Dem. Sea $M = A - \lambda_k I$. Por la Definición 24, para probar que $M^+ = U_k \Lambda_k^{-1} U_k'$ basta ver que M^+ verifica:

1. $MM^+M = M$.
2. $M^+MM^+ = M^+$.
3. $(M^+M)' = M^+M$.
4. $(MM^+)' = MM^+$.

Sea

$$D_k = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \lambda_{k-1} & 0 & 0 \\ 0 & 0 & \lambda_{k+1} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \lambda_p \end{bmatrix} = \Lambda_k + \lambda_k I.$$

Entonces,

$$\begin{aligned} AU_k &= A[\mathbf{x}_1, \dots, \mathbf{x}_{k-1}, \mathbf{x}_{k+1}, \dots, \mathbf{x}_p] \\ &= [\lambda_1 \mathbf{x}_1, \dots, \lambda_{k-1} \mathbf{x}_{k-1}, \lambda_{k+1} \mathbf{x}_{k+1}, \dots, \lambda_p \mathbf{x}_p] \\ &= U_k D_k. \end{aligned} \quad (112)$$

Por ser A simétrica y D_k una matriz diagonal:

$$(AU_k)' = U_k' A = (U_k D_k)' = D_k U_k'. \quad (113)$$

Por ser D_k y Λ_k matrices diagonales:

$$D_k \Lambda_k^{-1} = \Lambda_k^{-1} D_k. \quad (114)$$

Luego, por (112) a (114)

$$U_k \Lambda_k^{-1} U_k' A = U_k \Lambda_k^{-1} D_k U_k' = U_k D_k \Lambda_k^{-1} U_k' = AU_k \Lambda_k^{-1} U_k'. \quad (115)$$

Además,

$$\begin{aligned}
& (A - \lambda_k I)U_k \Lambda_k^{-1} U_k' (A - \lambda_k I) = \\
& = AU_k \Lambda_k^{-1} U_k' A - \lambda_k AU_k \Lambda_k^{-1} U_k' - \lambda_k U_k \Lambda_k^{-1} U_k' A + \lambda_k^2 U_k \Lambda_k^{-1} U_k' \\
& = U_k \Lambda_k^{-1} U_k' A^2 - 2\lambda_k U_k \Lambda_k^{-1} U_k' A + \lambda_k^2 U_k \Lambda_k^{-1} U_k' \\
& = U_k \Lambda_k^{-1} U_k' (A - \lambda_k I)^2 \\
& = (U_k \Lambda_k^{-1} U_k' A - \lambda_k U_k \Lambda_k^{-1} U_k') (A - \lambda_k I) \\
& = (U_k \Lambda_k^{-1} D_k U_k' - \lambda_k U_k \Lambda_k^{-1} U_k') (A - \lambda_k I) \\
& = (U_k D_k \Lambda_k^{-1} U_k' - \lambda_k U_k \Lambda_k^{-1} U_k') (A - \lambda_k I) \\
& = U_k (D_k - \lambda_k I) \Lambda_k^{-1} U_k' (A - \lambda_k I) \\
& = U_k U_k' (A - \lambda_k I).
\end{aligned}$$

Sea $U = [\mathbf{x}_1, \dots, \mathbf{x}_p]$. Entonces $U \in \mathfrak{R}^{p \times p}$ es una matriz ortogonal, $U'U = UU' = I$ y $U' = U^{-1}$. Luego:

$$U_k U_k' + \mathbf{x}_k \mathbf{x}_k' = \sum_{i=1}^p \mathbf{x}_i \mathbf{x}_i' = UU' = I,$$

y por lo tanto,

$$\begin{aligned}
(A - \lambda_k I)U_k \Lambda_k^{-1} U_k' (A - \lambda_k I) &= (I - \mathbf{x}_k \mathbf{x}_k') (A - \lambda_k I) \\
&= A - \lambda_k I - \mathbf{x}_k (\mathbf{x}_k' A - \lambda_k \mathbf{x}_k') \\
&= A - \lambda_k I - \mathbf{x}_k (\lambda_k \mathbf{x}_k' - \lambda_k \mathbf{x}_k') \\
&= A - \lambda_k I,
\end{aligned}$$

es decir se verifica 1.

Como $U_k' U_k = I$, entonces

$$\begin{aligned}
& U_k \Lambda_k^{-1} U_k' (A - \lambda_k I) U_k \Lambda_k^{-1} U_k' = \\
& = U_k \Lambda_k^{-1} U_k' A U_k \Lambda_k^{-1} U_k' - \lambda_k U_k \Lambda_k^{-1} U_k' U_k \Lambda_k^{-1} U_k' \\
& = A U_k \Lambda_k^{-1} U_k' U_k \Lambda_k^{-1} U_k' - \lambda_k U_k \Lambda_k^{-1} U_k' U_k \Lambda_k^{-1} U_k' \\
& = A U_k \Lambda_k^{-2} U_k' - \lambda_k U_k \Lambda_k^{-2} U_k' \\
& = U_k D_k \Lambda_k^{-2} U_k' - \lambda_k U_k \Lambda_k^{-2} U_k' \\
& = U_k (D_k - \lambda_k I) \Lambda_k^{-2} U_k' \\
& = U_k \Lambda_k^{-1} U_k',
\end{aligned}$$

es decir, se verifica 2.

Por ser $A - \lambda_k I$ una matriz simétrica y por (115)

$$\begin{aligned}
 [U_k \Lambda_k^{-1} U_k' (A - \lambda_k I)]' &= (A - \lambda_k I) U_k \Lambda_k^{-1} U_k' \\
 &= A U_k \Lambda_k^{-1} U_k' - \lambda_k U_k \Lambda_k^{-1} U_k' \\
 &= U_k \Lambda_k^{-1} U_k' A - \lambda_k U_k \Lambda_k^{-1} U_k' \\
 &= U_k \Lambda_k^{-1} U_k' (A - \lambda_k I);
 \end{aligned}$$

$$\begin{aligned}
 [(A - \lambda_k I) U_k \Lambda_k^{-1} U_k']' &= U_k \Lambda_k^{-1} U_k' (A - \lambda_k I) \\
 &= U_k \Lambda_k^{-1} U_k' A - \lambda_k U_k \Lambda_k^{-1} U_k' \\
 &= A U_k \Lambda_k^{-1} U_k' - \lambda_k U_k \Lambda_k^{-1} U_k' \\
 &= (A - \lambda_k I) U_k \Lambda_k^{-1} U_k',
 \end{aligned}$$

es decir, se verifican 3. y 4.

Teorema 18 Sea $A \in \mathfrak{R}^{p \times p}$ es una matriz simétrica con autovalores simples $\lambda_1, \dots, \lambda_p$ y autovectores respectivos $\mathbf{x}_1, \dots, \mathbf{x}_p$ ortogonales.

Entonces, si $\mathbf{x}_k = (x_{1k}, \dots, x_{pk})'$ es el autovector correspondiente al autovalor λ_k se satisfacen:

$$\begin{aligned}
 \frac{\partial x_{lk}}{\partial a_{ij}} &= x_{jk} \sum_{s=1, s \neq k}^p \frac{x_{is} x_{ls}}{\lambda_k - \lambda_s}; \\
 \frac{\partial \lambda_k}{\partial a_{ij}} &= x_{ik} x_{jk},
 \end{aligned}$$

para $1 \leq i \leq j \leq l \leq k \leq p$.

Dem. Como

$$\frac{\partial \lambda_k}{\partial (\text{vec } A)'} = \left[\frac{\partial \lambda_k}{\partial a_{11}}, \dots, \frac{\partial \lambda_k}{\partial a_{p1}}, \dots, \frac{\partial \lambda_k}{\partial a_{1p}}, \dots, \frac{\partial \lambda_k}{\partial a_{pp}} \right];$$

$$\mathbf{x}_k' \otimes \mathbf{x}_k = [x_{1k} \mathbf{x}_k \dots x_{pk} \mathbf{x}_k],$$

entonces, por el Teorema 17

$$\left[\frac{\partial \lambda_k}{\partial a_{1j}}, \dots, \frac{\partial \lambda_k}{\partial a_{pj}} \right] = x_{jk} \mathbf{x}_k,$$

es decir,

$$\frac{\partial \lambda_k}{\partial a_{ij}} = x_{ik} x_{jk}.$$

Sea $H_k = \lambda_k I - A$. Entonces por el Lema 21, $H_k^+ = -U_k \Lambda_k^{-1} U_k'$ donde U_k y Λ_k son las dadas en dicho Lema.

Como

$$\frac{\partial \mathbf{x}_k}{\partial (\text{vec } A)'} = \begin{bmatrix} \frac{\partial x_{1k}}{\partial a_{11}} & \frac{\partial x_{1k}}{\partial a_{p1}} & \frac{\partial x_{1k}}{\partial a_{1p}} & \frac{\partial x_{1k}}{\partial a_{pp}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial x_{pk}}{\partial a_{11}} & \frac{\partial x_{pk}}{\partial a_{p1}} & \frac{\partial x_{pk}}{\partial a_{1p}} & \frac{\partial x_{pk}}{\partial a_{pp}} \end{bmatrix};$$

$$\mathbf{x}_k' \otimes H_k^+ = [x_{1k} H_k^+ \dots x_{pk} H_k^+],$$

entonces, por el Teorema 17

$$\begin{bmatrix} \frac{\partial x_{1k}}{\partial a_{1j}} & \frac{\partial x_{1k}}{\partial a_{pj}} \\ \vdots & \vdots \\ \frac{\partial x_{pk}}{\partial a_{1j}} & \frac{\partial x_{pk}}{\partial a_{pj}} \end{bmatrix} = x_{jk} H_k^+,$$

es decir,

$$\frac{\partial x_{lk}}{\partial a_{ij}} = x_{jk} (H_k^+)_{li}.$$

Calculemos los elementos de la matriz H_k^+ .

Como

$$(U_k)_{ij} = \begin{cases} x_{ij} & \text{si } j < k \\ x_{i,j+1} & \text{si } j \geq k \end{cases}$$

$$(\Lambda_k^{-1})_{ij} = \begin{cases} 0 & \text{si } i \neq j \\ \frac{1}{\lambda_j - \lambda_k} & \text{si } i = j \quad j < k \\ \frac{1}{\lambda_{j+1} - \lambda_k} & \text{si } i = j \quad j \geq k \end{cases}$$

entonces,

$$\begin{aligned} (U_k \Lambda_k^{-1})_{ij} &= \sum_{s=1}^{p-1} (U_k)_{is} (\Lambda_k^{-1})_{sj} = (U_k)_{ij} (\Lambda_k^{-1})_{jj} \\ &= \begin{cases} \frac{x_{ij}}{\lambda_j - \lambda_k} & \text{si } j < k \\ \frac{x_{i,j+1}}{\lambda_{j+1} - \lambda_k} & \text{si } j \geq k \end{cases} \end{aligned}$$

$$\begin{aligned}
(U_k \Lambda_k^{-1} U'_k)_{ij} &= \sum_{s=1}^{p-1} (U_k \Lambda_k^{-1})_{is} (U'_k)_{sj} \\
&= \sum_{s=1}^{k-1} \frac{x_{is} x_{js}}{\lambda_s - \lambda_k} + \sum_{s=k}^{p-1} \frac{x_{i,s+1} x_{j,s+1}}{\lambda_{s+1} - \lambda_k} \\
&= \sum_{s=1, s \neq k}^p \frac{x_{is} x_{js}}{\lambda_s - \lambda_k}.
\end{aligned}$$

Por lo tanto,

$$\frac{\partial x_{lk}}{\partial a_{ij}} = x_{jk} \sum_{s=1, s \neq k}^p \frac{x_{is} x_{ls}}{\lambda_k - \lambda_s}.$$

El Teorema 18 proporciona resultados sobre las derivadas de los autovalores y autovectores ortogonales con respecto a los elementos de una matriz simétrica definida positiva. Los resultados de este Teorema son necesarios para obtener los desarrollos de Taylor utilizados en la demostración del Teorema 11.

Antes de obtener la distribución asintótica de los parámetros del modelo de análisis factorial, necesitamos probar unos resultados previos.

Sea $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathfrak{R}^p$ una muestra aleatoria que satisface el modelo (1) bajo las hipótesis (10), (11) y (14), de modo que el modelo sea identificable. Entonces, la matriz de covarianza de las observaciones está dada por

$$\Sigma_0 = \Lambda_0 \Lambda_0' + \Psi_0,$$

siendo $\Lambda_0 \in \mathfrak{R}^{p \times k}$ y $\Psi_0 \in \mathcal{D}_+(p)$.

Notación: Sea

$$S_n^* = \widehat{\Psi}_n^{-1/2} \widetilde{\Sigma}_n \widehat{\Psi}_n^{-1/2},$$

donde $\widetilde{\Sigma}_n$ es un estimador consistente de Σ_0 y $\widehat{\Psi}_n$ es el estimador de Ψ_0 basado en $\widetilde{\Sigma}_n$. Notaremos con $\widehat{\theta}_1^n > \dots > \widehat{\theta}_p^n$ a los autovalores de S_n^* y con $\widehat{\mathbf{w}}_1^n, \dots, \widehat{\mathbf{w}}_p^n$ a sus correspondientes autovectores normalizados.

Sean

$$\widehat{\Theta}_n = \widetilde{\Theta}(\widehat{\Psi}_n) = \begin{bmatrix} \widehat{\theta}_1^n & 0 \\ \vdots & \vdots \\ 0 & \widehat{\theta}_k^n \end{bmatrix}; \quad \widehat{\Omega}_n = \widetilde{\Omega}(\widehat{\Psi}_n) = (\widehat{\mathbf{w}}_1^n, \dots, \widehat{\mathbf{w}}_k^n).$$

Por el resultado obtenido en la Proposición 3, si $\widehat{\Lambda}_n$ y $\widehat{\Psi}_n$ son los estimadores de los parámetros del modelo de análisis factorial basados en $\widetilde{\Sigma}_n$ de modo que

$$\left(\widehat{\Lambda}_n, \widehat{\Psi}_n\right) = \arg \min_{(\Lambda, \Psi) \in \mathcal{B}(\Lambda, \Psi)} \left\{ \text{tr} \left[\widetilde{\Sigma}_n \Sigma(\Lambda, \Psi)^{-1} \right] - \ln \left| \widetilde{\Sigma}_n \Sigma(\Lambda, \Psi)^{-1} \right| \right\},$$

donde $\mathcal{B}(\Lambda, \Psi) = \{(\Lambda, \Psi) \in \mathcal{R} \times \mathcal{D}_+(p) : \Lambda' \Psi^{-1} \Lambda \text{ es diagonal}\}$, $\widehat{\Lambda}_n$ y $\widehat{\Psi}_n$ deben satisfacer las ecuaciones

$$\widehat{\Lambda}_n = \widehat{\Psi}_n^{1/2} \widehat{\Omega}_n \left(\widehat{\Theta}_n - I \right)^{1/2}; \quad (116)$$

$$\text{diag} \left(\widehat{\Psi}_n \right) = \text{diag} \left(\widetilde{\Sigma}_n - \widehat{\Lambda}_n \widehat{\Lambda}_n' \right); \quad (117)$$

$$\widehat{\Lambda}_n' \widehat{\Psi}_n^{-1} \widehat{\Lambda}_n = \widehat{\Theta}_n - I. \quad (118)$$

Veamos cómo escribir a los elementos de la matriz Λ en función de los autovalores y autovectores de $\Sigma^* = \Psi^{-1/2} \Sigma \Psi^{-1/2}$ (ver Capítulo 6). De este modo, se puede considerar que $\lambda_{ij} = \lambda_{ij}(\Psi, \Sigma)$.

Lema 22 Para $1 \leq i \leq p$, $1 \leq j \leq k$

1.

$$\lambda_{ij} = w_{ij} \sqrt{\psi_{ii}(\theta_j - 1)}. \quad (119)$$

2.

$$(\Lambda \Lambda')_{ii} = \psi_{ii} \sum_{h=1}^k w_{ih}^2 (\theta_h - 1). \quad (120)$$

Dem.

1. Por Proposición 3

$$\Lambda = \Psi^{1/2} \Omega (\Theta - I)^{1/2},$$

con

$$\left(\Psi^{1/2}\right)_{ij} = \begin{cases} \sqrt{\psi_{ii}} & \text{si } i = j \\ 0 & \text{sino} \end{cases}$$

$$\left((\Theta - I)^{1/2}\right)_{ij} = \begin{cases} \sqrt{\theta_i - 1} & \text{si } i = j \\ 0 & \text{sino} \end{cases}$$

$$(\Omega)_{ij} = w_{ij};$$

$$(\Omega')_{ij} = w_{ji}.$$

Luego:

$$\left(\Omega(\Theta - I)^{1/2}\right)_{ij} = q_{ij} = \sum_{h=1}^k w_{ih} \left((\Theta - I)^{1/2}\right)_{hj} = w_{ij} \sqrt{\theta_j - 1};$$

$$\begin{aligned} (\Psi^{1/2} \Omega (\Theta - I)^{1/2})_{ij} &= \sum_{h=1}^p (\Psi^{1/2})_{ih} q_{hj} = \sqrt{\psi_{ii}} q_{ij} \\ &= \sqrt{\psi_{ii} (\theta_j - 1)} w_{ij}, \end{aligned}$$

y por lo tanto

$$\lambda_{ij} = w_{ij} \sqrt{\psi_{ii} (\theta_j - 1)}.$$

2. Por lo anterior

$$\Lambda \Lambda' = \Psi^{1/2} \Omega (\Theta - I) \Omega' \Psi^{1/2}.$$

Pero:

$$(\Omega(\Theta - I))_{ij} = t_{ij} = \sum_{h=1}^k w_{ih} (\Theta - I)_{hj} = w_{ij} (\theta_j - 1);$$

$$(\Omega(\Theta - I) \Omega')_{ij} = u_{ij} = \sum_{h=1}^k t_{ih} (\Omega')_{hj} = \sum_{h=1}^k w_{ih} (\theta_h - 1) w_{jh};$$

$$\begin{aligned} (\Psi^{1/2} \Omega (\Theta - I) \Omega')_{ij} &= v_{ij} = \sum_{h=1}^p (\Psi^{1/2})_{ih} u_{hj} = \sqrt{\psi_{ii}} u_{ij} \\ &= \sqrt{\psi_{ii}} \sum_{h=1}^k w_{ih} (\theta_h - 1) w_{jh}; \end{aligned}$$

$$\begin{aligned} (\Psi^{1/2} \Omega (\Theta - I) \Omega' \Psi^{1/2})_{ij} &= (\Lambda \Lambda')_{ij} = \sum_{h=1}^p v_{ih} (\Psi^{1/2})_{hj} = v_{ij} \sqrt{\psi_{jj}} \\ &= \sqrt{\psi_{ii} \psi_{jj}} \sum_{h=1}^k w_{ih} (\theta_h - 1) w_{jh}, \end{aligned}$$

y por lo tanto,

$$(\Lambda\Lambda')_{ii} = \psi_{ii} \sum_{h=1}^k w_{ih}^2 (\theta_h - 1).$$

En los siguientes Lemas obtendremos las expresiones de las derivadas de los elementos de la matriz Λ respecto de Σ y Ψ , las cuales utilizaremos en la demostración del Teorema 11.

Lema 23 Sea $g_i(\Psi, \Sigma) = (\Lambda\Lambda')_{ii}$ para $1 \leq i \leq p$. Entonces:

$$\begin{aligned} \frac{\partial g_i}{\partial \psi_{jj}} &= \frac{\psi_{ii}}{\psi_{jj}} \sum_{h=1}^k w_{ih} w_{jh} \left\{ -\theta_h w_{ih} w_{jh} + \right. \\ &\quad \left. + (\theta_h - 1) \left[\delta_{ij} + \sum_{r=1, r \neq h}^p \left(\frac{\theta_r + \theta_h}{\theta_r - \theta_h} \right) w_{ir} w_{jr} \right] \right\}; \\ \frac{\partial g_i}{\partial \sigma_{\alpha\beta}} &= \frac{(2 - \delta_{\alpha\beta}) \psi_{ii}}{\sqrt{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \sum_{h=1}^k w_{ih} \left\{ w_{ih} w_{\alpha h} w_{\beta h} + \right. \\ &\quad \left. + (\theta_h - 1) \left[\sum_{r=1, r \neq h}^p \frac{w_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r})}{\theta_h - \theta_r} \right] \right\}; \end{aligned}$$

Dem. Calculemos las derivadas $\partial g_i / \partial \psi_{jj}$.

Por (120) del Lema 22

$$\frac{\partial g_i}{\partial \psi_{ii}} = \sum_{h=1}^k \left\{ (\theta_h - 1) w_{ih}^2 + \psi_{ii} \left[\frac{\partial \theta_h}{\partial \psi_{ii}} w_{ih}^2 + 2(\theta_h - 1) w_{ih} \frac{\partial w_{ih}}{\partial \psi_{ii}} \right] \right\}, \quad (121)$$

y si $i \neq j$

$$\frac{\partial g_i}{\partial \psi_{jj}} = \psi_{ii} \sum_{h=1}^k \left[\frac{\partial \theta_h}{\partial \psi_{jj}} w_{ih}^2 + 2(\theta_h - 1) w_{ih} \frac{\partial w_{ih}}{\partial \psi_{jj}} \right]. \quad (122)$$

Entonces es necesario calcular las derivadas $\partial \theta_h / \partial \psi_{jj}$ y $\partial w_{ih} / \partial \psi_{jj}$.

Como $w_{ih} = w_{ih}(\Sigma^*(\Psi))$ y $\theta_h = \theta_h(\Sigma^*(\Psi))$, por el Lema 16

$$\frac{\partial w_{ih}}{\partial \psi_{jj}} = \sum_{\gamma} \sum_{\delta} \frac{\partial w_{ih}}{\partial (\Sigma^*)_{\gamma\delta}} \frac{\partial (\Sigma^*)_{\gamma\delta}}{\partial \psi_{jj}}; \quad (123)$$

$$\frac{\partial \theta_h}{\partial \psi_{jj}} = \sum_{\gamma} \sum_{\delta} \frac{\partial \theta_h}{\partial (\Sigma^*)_{\gamma\delta}} \frac{\partial (\Sigma^*)_{\gamma\delta}}{\partial \psi_{jj}}. \quad (124)$$

Para calcular $\partial w_{ih}/\partial (\Sigma^*)_{\gamma\delta}$ y $\partial \theta_h/\partial (\Sigma^*)_{\gamma\delta}$ usaremos el Teorema 18, debido al cual

$$\frac{\partial w_{ih}}{\partial (\Sigma^*)_{\gamma\delta}} = w_{\delta h} \sum_{r=1, r \neq h}^p \frac{w_{\gamma r} w_{ir}}{\theta_h - \theta_r}; \quad (125)$$

$$\frac{\partial \theta_h}{\partial (\Sigma^*)_{\gamma\delta}} = w_{\gamma h} w_{\delta h}. \quad (126)$$

Como $\Psi \in \mathcal{D}_+(p)$,

$$\left(\Psi^{-1/2} \Sigma \right)_{ij} = \sum_{h=1}^p \left(\Psi^{-1/2} \right)_{ih} \sigma_{hj} = \frac{\sigma_{ij}}{\sqrt{\psi_{ii}}};$$

$$\left(\Psi^{-1/2} \Sigma \Psi^{-1/2} \right)_{ij} = \sum_{h=1}^p \left(\Psi^{-1/2} \Sigma \right)_{ih} \left(\Psi^{-1/2} \right)_{hj} = \frac{\sigma_{ij}}{\sqrt{\psi_{ii} \psi_{jj}}} = \sigma_{ij}^* = \sigma_{ji}^*.$$

Luego, si $i \neq j$

$$\frac{\partial \sigma_{ij}^*}{\partial \psi_{ii}} = \frac{\partial \sigma_{ji}^*}{\partial \psi_{ii}} = -\frac{\sigma_{ij}}{2\psi_{ii} \sqrt{\psi_{ii} \psi_{jj}}},$$

y si $i = j$

$$\frac{\partial \sigma_{ii}^*}{\partial \psi_{ii}} = -\frac{\sigma_{ii}}{\psi_{ii}^2}.$$

Además,

$$\frac{\partial \sigma_{ij}^*}{\partial \psi_{ll}} = 0 \quad \text{si } l \neq i, l \neq j.$$

Reemplazando en (123)

$$\begin{aligned}
\frac{\partial w_{ih}}{\partial \psi_{jj}} &= \sum_{\gamma \neq j} \left[\left(w_{jh} \sum_{r=1, r \neq h}^p \frac{w_{\gamma r} w_{ir}}{\theta_h - \theta_r} \right) \left(-\frac{\sigma_{j\gamma}}{2\psi_{jj} \sqrt{\psi_{jj} \psi_{\gamma\gamma}}} \right) \right] + \\
&+ \sum_{\delta \neq j} \left[\left(w_{\delta h} \sum_{r=1, r \neq h}^p \frac{w_{jr} w_{ir}}{\theta_h - \theta_r} \right) \left(-\frac{\sigma_{j\delta}}{2\psi_{jj} \sqrt{\psi_{jj} \psi_{\delta\delta}}} \right) \right] + \\
&+ \left(w_{jh} \sum_{r=1, r \neq h}^p \frac{w_{jr} w_{ir}}{\theta_h - \theta_r} \right) \left(-\frac{\sigma_{jj}}{\psi_{jj}^2} \right) \tag{127} \\
&= \frac{1}{2\sqrt{\psi_{jj}^3}} \left[\sum_{\gamma} \frac{w_{jh} \sigma_{j\gamma}}{\sqrt{\psi_{\gamma\gamma}}} \left(\sum_{r=1, r \neq h}^p \frac{w_{\gamma r} w_{ir}}{\theta_r - \theta_h} \right) + \right. \\
&\left. + \sum_{\delta} \frac{w_{\delta h} \sigma_{j\delta}}{\sqrt{\psi_{\delta\delta}}} \left(\sum_{r=1, r \neq h}^p \frac{w_{jr} w_{ir}}{\theta_r - \theta_h} \right) \right].
\end{aligned}$$

Por otro lado,

$$\begin{aligned}
(\Sigma^*)_{ij} &= \sigma_{ij}^* = \frac{\sigma_{ij}}{\sqrt{\psi_{ii} \psi_{jj}}} \\
\Rightarrow \Sigma^* \mathbf{w}_h &= \theta_h \mathbf{w}_h \iff \sum_{\gamma=1}^p \sigma_{i\gamma}^* w_{\gamma h} = \theta_h w_{ih} \\
&\iff \sum_{\gamma=1}^p \frac{\sigma_{i\gamma}}{\sqrt{\psi_{\gamma\gamma}}} w_{\gamma h} = \theta_h w_{ih} \sqrt{\psi_{ii}}.
\end{aligned}$$

Luego,

$$\begin{aligned}
\sum_{\gamma} \frac{w_{jh} \sigma_{j\gamma}}{\sqrt{\psi_{\gamma\gamma}}} \left(\sum_{r=1, r \neq h}^p \frac{w_{\gamma r} w_{ir}}{\theta_r - \theta_h} \right) &= \sum_{r=1, r \neq h}^p \frac{w_{ir} w_{jh}}{\theta_r - \theta_h} \left(\sum_{\gamma} \frac{\sigma_{j\gamma} w_{\gamma r}}{\sqrt{\psi_{\gamma\gamma}}} \right) \\
&= \sum_{r=1, r \neq h}^p \left(\frac{w_{ir} w_{jh}}{\theta_r - \theta_h} \right) \theta_r w_{jr} \sqrt{\psi_{jj}}.
\end{aligned}$$

Reemplazando en (127)

$$\begin{aligned}
\frac{\partial w_{ih}}{\partial \psi_{jj}} &= \frac{\sqrt{\psi_{jj}}}{2\sqrt{\psi_{jj}^3}} \left[\left(\sum_{r=1, r \neq h}^p \left(\frac{\theta_r w_{ir} w_{jr}}{\theta_r - \theta_h} \right) w_{jh} \right) + \right. \\
&\quad \left. + \left(\sum_{r=1, r \neq h}^p \frac{w_{jr} w_{ir}}{\theta_r - \theta_h} \right) \theta_h w_{jh} \right] \\
&= \frac{w_{jh}}{2\psi_{jj}} \left[\sum_{r=1, r \neq h}^p \left(\frac{\theta_r + \theta_h}{\theta_r - \theta_h} \right) w_{ir} w_{jr} \right].
\end{aligned} \tag{128}$$

Reemplazando en (124)

$$\begin{aligned}
\frac{\partial \theta_h}{\partial \psi_{jj}} &= \sum_{\gamma \neq j} \left[(w_{\gamma h} w_{jh}) \left(-\frac{\sigma_{j\gamma}}{2\psi_{jj} \sqrt{\psi_{jj} \psi_{\gamma\gamma}}} \right) \right] + \\
&\quad + \sum_{\delta \neq j} \left[(w_{jh} w_{\delta h}) \left(-\frac{\sigma_{j\delta}}{2\psi_{jj} \sqrt{\psi_{jj} \psi_{\delta\delta}}} \right) \right] + \\
&\quad + w_{jh}^2 \left(-\frac{\sigma_{jj}}{\psi_{jj}^2} \right) \\
&= -\frac{w_{jh}}{\sqrt{\psi_{jj}^3}} \left(\sum_{\gamma} \frac{w_{\gamma h} \sigma_{j\gamma}}{\sqrt{\psi_{\gamma\gamma}}} \right) \\
&= -\frac{\theta_h w_{jh}^2}{\psi_{jj}}.
\end{aligned} \tag{129}$$

Entonces, reemplazando en (121)

$$\begin{aligned}
\frac{\partial g_i}{\partial \psi_{ii}} &= \sum_{h=1}^k w_{ih} \left\{ w_{ih} (\theta_h - 1) + \frac{1}{\sqrt{\psi_{ii}}} \left\{ -w_{ih}^2 \sum_{\gamma=1}^p \frac{w_{\gamma h} \sigma_{i\gamma}}{\sqrt{\psi_{\gamma\gamma}}} + \right. \right. \\
&\quad + (\theta_h - 1) \left[w_{ih} \sum_{\gamma=1}^p \frac{\sigma_{i\gamma}}{\sqrt{\psi_{\gamma\gamma}}} \left(\sum_{r=1, r \neq h}^p \frac{w_{\gamma r} w_{ir}}{\theta_r - \theta_h} \right) + \right. \\
&\quad \left. \left. + \sum_{\delta=1}^p \frac{w_{\delta h} \sigma_{i\delta}}{\sqrt{\psi_{\delta\delta}}} \left(\sum_{r=1, r \neq h}^p \frac{w_{ir}^2}{\theta_r - \theta_h} \right) \right] \right\} \\
&= \sum_{h=1}^k w_{ih}^2 \left\{ (\theta_h - 1) \left[1 + \sum_{r=1, r \neq h}^p \left(\frac{\theta_r + \theta_h}{\theta_r - \theta_h} \right) w_{ir}^2 \right] - \theta_h w_{ih}^2 \right\}.
\end{aligned}$$

Si $i \neq j$, reemplazando en (122)

$$\begin{aligned}
\frac{\partial g_i}{\partial \psi_{jj}} &= \frac{\psi_{ii}}{\sqrt{\psi_{jj}^3}} \sum_{h=1}^k w_{ih} \left\{ -w_{ih} w_{jh} \sum_{\gamma=1}^p \frac{w_{\gamma h} \sigma_{j\gamma}}{\sqrt{\psi_{\gamma\gamma}}} + \right. \\
&\quad + (\theta_h - 1) \left[w_{jh} \sum_{\gamma=1}^p \frac{\sigma_{j\gamma}}{\sqrt{\psi_{\gamma\gamma}}} \left(\sum_{r=1, r \neq h}^p \frac{w_{\gamma r} w_{ir}}{\theta_r - \theta_h} \right) + \right. \\
&\quad \left. \left. + \sum_{\delta=1}^p \frac{w_{\delta h} \sigma_{j\delta}}{\sqrt{\psi_{\delta\delta}}} \left(\sum_{r=1, r \neq h}^p \frac{w_{jr} w_{ir}}{\theta_r - \theta_h} \right) \right] \right\} \\
&= \frac{\psi_{ii}}{\psi_{jj}} \sum_{h=1}^k w_{ih} w_{jh} \left[-\theta_h w_{ih} w_{jh} + \right. \\
&\quad \left. + (\theta_h - 1) \sum_{r=1, r \neq h}^p \left(\frac{\theta_r + \theta_h}{\theta_r - \theta_h} \right) w_{ir} w_{jr} \right].
\end{aligned}$$

Por (120)

$$\frac{\partial g_i}{\partial \sigma_{\alpha\beta}} = \psi_{ii} \sum_{h=1}^k \left[w_{ih}^2 \frac{\partial \theta_h}{\partial \sigma_{\alpha\beta}} + 2w_{ih} (\theta_h - 1) \frac{\partial w_{ih}}{\partial \sigma_{\alpha\beta}} \right]. \quad (130)$$

Entonces es necesario calcular las derivadas $\partial \theta_h / \partial \sigma_{\alpha\beta}$ y $\partial w_{ih} / \partial \sigma_{\alpha\beta}$.
Como $w_{ih} = w_{ih}(\Sigma^*(\Psi))$ y $\theta_h = \theta_h(\Sigma^*(\Psi))$, por el Lema 16

$$\frac{\partial w_{ih}}{\partial \sigma_{\alpha\beta}} = \sum_{\gamma} \sum_{\delta} \frac{\partial w_{ih}}{\partial (\Sigma^*)_{\gamma\delta}} \frac{\partial (\Sigma^*)_{\gamma\delta}}{\partial \sigma_{\alpha\beta}}; \quad (131)$$

$$\frac{\partial \theta_h}{\partial \sigma_{\alpha\beta}} = \sum_{\gamma} \sum_{\delta} \frac{\partial \theta_h}{\partial (\Sigma^*)_{\gamma\delta}} \frac{\partial (\Sigma^*)_{\gamma\delta}}{\partial \sigma_{\alpha\beta}}. \quad (132)$$

Debido a que

$$\sigma_{ij}^* = \frac{\sigma_{ij}}{\sqrt{\psi_{ii} \psi_{jj}}},$$

entonces, por la simetría de la matriz Σ

$$\frac{\partial \sigma_{ij}^*}{\partial \sigma_{\alpha\beta}} = \begin{cases} \frac{1}{\sqrt{\psi_{\alpha\alpha} \psi_{\beta\beta}}} & \text{si } i = \alpha, j = \beta \text{ ó } i = \beta, j = \alpha \\ 0 & \text{sino} \end{cases}$$

es decir:

$$\frac{\partial \sigma_{ij}^*}{\partial \sigma_{ij}} = \frac{\partial \sigma_{ij}^*}{\partial \sigma_{ji}} = \frac{\partial \sigma_{ji}^*}{\partial \sigma_{ij}} = \frac{\partial \sigma_{ji}^*}{\partial \sigma_{ji}} = \frac{1}{\sqrt{\psi_{ii}\psi_{jj}}}.$$

Sea $\alpha \neq \beta$. Reemplazando en (131) y usando (125)

$$\begin{aligned} \frac{\partial w_{ih}}{\partial \sigma_{\alpha\beta}} &= \frac{\partial w_{ih}}{\partial \sigma_{\alpha\beta}^*} \frac{\partial \sigma_{\alpha\beta}^*}{\partial \sigma_{\alpha\beta}} + \frac{\partial w_{ih}}{\partial \sigma_{\beta\alpha}^*} \frac{\partial \sigma_{\beta\alpha}^*}{\partial \sigma_{\alpha\beta}} \\ &= \frac{1}{\sqrt{\psi_{\alpha\alpha}\psi_{\beta\beta}}} \left[w_{\beta h} \left(\sum_{r=1, r \neq h}^p \frac{w_{\alpha r} w_{ir}}{\theta_h - \theta_r} \right) + \right. \\ &\quad \left. + w_{\alpha h} \left(\sum_{r=1, r \neq h}^p \frac{w_{\beta r} w_{ir}}{\theta_h - \theta_r} \right) \right]. \end{aligned} \quad (133)$$

Sea $\alpha = \beta$. Reemplazando en (131) y usando (125)

$$\begin{aligned} \frac{\partial w_{ih}}{\partial \sigma_{\alpha\alpha}} &= \frac{\partial w_{ih}}{\partial \sigma_{\alpha\alpha}^*} \frac{\partial \sigma_{\alpha\alpha}^*}{\partial \sigma_{\alpha\alpha}} \\ &= \frac{w_{\alpha h}}{\psi_{\alpha\alpha}} \left(\sum_{r=1, r \neq h}^p \frac{w_{\alpha r} w_{ir}}{\theta_h - \theta_r} \right). \end{aligned} \quad (134)$$

Sea $\alpha \neq \beta$. Reemplazando en (132) y usando (126)

$$\begin{aligned} \frac{\partial \theta_h}{\partial \sigma_{\alpha\beta}} &= \frac{\partial \theta_h}{\partial \sigma_{\alpha\beta}^*} \frac{\partial \sigma_{\alpha\beta}^*}{\partial \sigma_{\alpha\beta}} + \frac{\partial \theta_h}{\partial \sigma_{\beta\alpha}^*} \frac{\partial \sigma_{\beta\alpha}^*}{\partial \sigma_{\alpha\beta}} \\ &= 2 \frac{w_{\alpha h} w_{\beta h}}{\sqrt{\psi_{\alpha\alpha}\psi_{\beta\beta}}}. \end{aligned} \quad (135)$$

Sea $\alpha = \beta$. Reemplazando en (132) y usando (126)

$$\begin{aligned} \frac{\partial \theta_h}{\partial \sigma_{\alpha\alpha}} &= \frac{\partial \theta_h}{\partial \sigma_{\alpha\alpha}^*} \frac{\partial \sigma_{\alpha\alpha}^*}{\partial \sigma_{\alpha\alpha}} \\ &= \frac{w_{\alpha h}^2}{\psi_{\alpha\alpha}}. \end{aligned} \quad (136)$$

Entonces, si $\alpha \neq \beta$, reemplazando en (130)

$$\begin{aligned} \frac{\partial g_i}{\partial \sigma_{\alpha\beta}} &= \frac{2\psi_{ii}}{\sqrt{\psi_{\alpha\alpha}\psi_{\beta\beta}}} \sum_{h=1}^k w_{ih} \left\{ w_{ih} w_{\alpha h} w_{\beta h} + \right. \\ &\quad \left. + (\theta_h - 1) \left[\sum_{r=1, r \neq h}^p \left(\frac{w_{ir}}{\theta_h - \theta_r} \right) (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) \right] \right\}, \end{aligned}$$

y si $\alpha = \beta$

$$\frac{\partial g_i}{\partial \sigma_{\alpha\alpha}} = \frac{\psi_{ii}}{\psi_{\alpha\alpha}} \sum_{h=1}^k w_{ih} w_{\alpha h} \left[w_{ih} w_{\alpha h} + 2(\theta_h - 1) \left(\sum_{r=1, r \neq h}^p \frac{w_{\alpha r} w_{ir}}{\theta_h - \theta_r} \right) \right].$$

Lema 24

$$\begin{aligned} \frac{\partial \lambda_{ij}}{\partial \psi_{ii}} &= \frac{w_{ij} \sqrt{\psi_{ii}(\theta_j - 1)}}{2\psi_{ii}} \left[\delta_{il} + \sum_{r=1, r \neq j}^p \left(\frac{\theta_r + \theta_j}{\theta_r - \theta_j} \right) w_{ir} w_{lr} - \frac{\theta_j w_{ij} w_{lj}}{\theta_j - 1} \right]; \\ \frac{\partial \lambda_{ij}}{\partial \sigma_{\alpha\beta}} &= \sqrt{\frac{\psi_{ii}(\theta_j - 1)}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \left(1 - \frac{\delta_{\alpha\beta}}{2} \right) \left[\sum_{r=1, r \neq j}^p \frac{w_{ir}}{\theta_j - \theta_r} (w_{\alpha r} w_{\beta j} + \right. \\ &\quad \left. + w_{\alpha j} w_{\beta r}) + \frac{w_{\alpha j} w_{ij} w_{\beta j}}{\theta_j - 1} \right]. \end{aligned}$$

Dem. Calculemos las derivadas $\partial \lambda_{ij} / \partial \psi_{ii}$.

Sea $l = i$. Por (119) del Lema 22

$$\begin{aligned} \frac{\partial \lambda_{ij}}{\partial \psi_{ii}} &= \frac{\partial w_{ij}}{\partial \psi_{ii}} \sqrt{\psi_{ii}(\theta_j - 1)} + \frac{1}{2} w_{ij} \frac{\sqrt{\theta_j - 1}}{\sqrt{\psi_{ii}}} + \\ &\quad + \frac{w_{ij} \sqrt{\psi_{ii}}}{2\sqrt{\theta_j - 1}} \frac{\partial \theta_j}{\partial \psi_{ii}}. \end{aligned}$$

Por (128) y (129)

$$\begin{aligned} \frac{\partial \lambda_{ij}}{\partial \psi_{ii}} &= \frac{w_{ij}}{2\sqrt{\psi_{ii}}} \sqrt{\psi_{ii}(\theta_j - 1)} \sum_{r=1, r \neq j}^p \left(\frac{\theta_r + \theta_j}{\theta_r - \theta_j} \right) w_{ir}^2 + \frac{w_{ij} \sqrt{\theta_j - 1}}{2\sqrt{\psi_{ii}}} + \\ &\quad + \frac{w_{ij} \sqrt{\psi_{ii}}}{2\sqrt{\theta_j - 1}} \left(-\frac{\theta_j w_{ij}^2}{\psi_{ii}} \right) \\ &= \frac{w_{ij}}{2} \sqrt{\frac{\theta_j - 1}{\psi_{ii}}} \left[\sum_{r=1, r \neq j}^p \left(\frac{\theta_r + \theta_j}{\theta_r - \theta_j} \right) w_{ir}^2 + 1 - \frac{\theta_j w_{ij}^2}{\theta_j - 1} \right]. \end{aligned}$$

Sea $l \neq i$. Por (119)

$$\frac{\partial \lambda_{ij}}{\partial \psi_{ii}} = \frac{\partial w_{ij}}{\partial \psi_{ii}} \sqrt{\psi_{ii}(\theta_j - 1)} + \frac{w_{ij} \sqrt{\psi_{ii}}}{2\sqrt{\theta_j - 1}} \frac{\partial \theta_j}{\partial \psi_{ii}}.$$

Por (128) y (129)

$$\begin{aligned}
\frac{\partial \lambda_{ij}}{\partial \psi_{ii}} &= \frac{w_{ij}}{2\psi_{ii}} \sqrt{\psi_{ii}(\theta_j - 1)} \sum_{r=1, r \neq j}^p \left(\frac{\theta_r + \theta_j}{\theta_r - \theta_j} \right) w_{ir} w_{lr} + \\
&+ \frac{w_{ij} \sqrt{\psi_{ii}}}{2\sqrt{\theta_j - 1}} \left(-\frac{\theta_j w_{ij}^2}{\psi_{ii}} \right) \\
&= \frac{w_{ij} \sqrt{\psi_{ii}(\theta_j - 1)}}{2\psi_{ii}} \left[\sum_{r=1, r \neq j}^p \left(\frac{\theta_r + \theta_j}{\theta_r - \theta_j} \right) w_{ir} w_{lr} - \frac{\theta_j w_{ij} w_{ij}}{\theta_j - 1} \right].
\end{aligned}$$

Calculemos las derivadas $\partial \lambda_{ij} / \partial \sigma_{\alpha\beta}$. Por (119)

$$\frac{\partial \lambda_{ij}}{\partial \sigma_{\alpha\beta}} = \sqrt{\psi_{ii}} \left(\sqrt{\theta_j - 1} \frac{\partial w_{ij}}{\partial \sigma_{\alpha\beta}} + \frac{w_{ij}}{2\sqrt{\theta_j - 1}} \frac{\partial \theta_j}{\partial \sigma_{\alpha\beta}} \right).$$

Sea $\alpha \neq \beta$. Luego, por (133) y (135)

$$\begin{aligned}
\frac{\partial \lambda_{ij}}{\partial \sigma_{\alpha\beta}} &= \sqrt{\psi_{ii}} \left\{ \sqrt{\frac{\theta_j - 1}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \left[w_{\beta j} \left(\sum_{r=1, r \neq j}^p \frac{w_{\alpha r} w_{ir}}{\theta_j - \theta_r} \right) + \right. \right. \\
&+ \left. \left. w_{\alpha j} \left(\sum_{r=1, r \neq j}^p \frac{w_{\beta r} w_{ir}}{\theta_j - \theta_r} \right) \right] + \frac{w_{ij}}{2\sqrt{\theta_j - 1}} \frac{w_{\alpha j} w_{\beta j}}{\sqrt{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \right\} \\
&= \sqrt{\frac{\psi_{ii}(\theta_j - 1)}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \left[w_{\beta j} \sum_{r=1, r \neq j}^p \left(\frac{w_{\alpha r} w_{ir}}{\theta_j - \theta_r} \right) + \right. \\
&+ \left. w_{\alpha j} \sum_{r=1, r \neq j}^p \left(\frac{w_{\beta r} w_{ir}}{\theta_j - \theta_r} \right) + \frac{w_{\alpha j} w_{ij} w_{\beta j}}{\theta_j - 1} \right].
\end{aligned}$$

Sea $\alpha = \beta$. Luego, por (134) y (136)

$$\begin{aligned}
\frac{\partial \lambda_{ij}}{\partial \sigma_{\alpha\alpha}} &= \sqrt{\psi_{ii}} \left[\sqrt{\theta_j - 1} \frac{w_{\alpha j}}{\psi_{\alpha\alpha}} \left(\sum_{r=1, r \neq j}^p \frac{w_{\alpha r} w_{ir}}{\theta_j - \theta_r} \right) + \frac{w_{ij}}{2\sqrt{\theta_j - 1}} \frac{w_{\alpha j}^2}{\psi_{\alpha\alpha}} \right] \\
&= \frac{w_{\alpha j} \sqrt{\psi_{ii}(\theta_j - 1)}}{\psi_{\alpha\alpha}} \left[\sum_{r=1, r \neq j}^p \frac{w_{\alpha r} w_{ir}}{\theta_j - \theta_r} + \frac{w_{\alpha j} w_{ij}}{2(\theta_j - 1)} \right].
\end{aligned}$$

Lema 25 Sea $g_i(\Psi, \Sigma) = (\Lambda\Lambda')_{ii}$ para $1 \leq i \leq p$. Entonces:

$$\frac{\partial^2 g_i}{\partial \psi_{jj} \partial \psi_{ss}} = \frac{1}{\psi_{jj}} \left\{ \psi_{ii} \frac{\partial B_{ij}}{\partial \psi_{ii}} + B_{ij} \left[\delta_{si} - \frac{\psi_{ii}}{\psi_{jj}} \delta_{sj} + \left(\frac{\psi_{ii}}{\psi_{jj}} - 1 \right) \delta_{si} \delta_{sj} \right] \right\}; \quad (137)$$

$$\begin{aligned} \frac{\partial^2 g_i}{\partial \psi_{jj} \partial \sigma_{\alpha\beta}} &= \frac{\psi_{ii}}{\sqrt{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \left\{ (2 - \delta_{\alpha\beta}) \frac{\partial F_{i\alpha\beta}}{\partial \psi_{ss}} + \right. \\ &\quad \left. + \left\{ \frac{\delta_{si}}{\psi_{ii}} \left[2 + \left(\frac{\psi_{ii}}{\psi_{\alpha\alpha}} - 1 \right) \delta_{s\alpha} + \left(\frac{\psi_{ii}}{\psi_{\beta\beta}} - 1 \right) \delta_{s\beta} \right] - \right. \right. \\ &\quad \left. \left. - \frac{\delta_{s\alpha}}{\psi_{\alpha\alpha}} - \frac{\delta_{s\beta}}{\psi_{\beta\beta}} + \frac{\delta_{s\alpha} \delta_{s\beta}}{\psi_{\beta\beta}} (1 - \delta_{si}) \right\} F_{i\alpha\beta} \right\}; \end{aligned} \quad (138)$$

$$\frac{\partial^2 g_i}{\partial \sigma_{\alpha\beta} \partial \sigma_{uv}} = \frac{(2 - \delta_{\alpha\beta}) \psi_{ii}}{\sqrt{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \frac{\partial F_{i\alpha\beta}}{\partial \sigma_{uv}}, \quad (139)$$

donde

$$A_{ij}^h = \sum_{r=1, r \neq h}^p \left(\frac{\theta_r + \theta_h}{\theta_r - \theta_h} \right) w_{ir} w_{jr};$$

$$B_{ij} = \sum_{h=1}^k w_{ih} w_{jh} \left[-\theta_h w_{ih} w_{jh} + (\theta_h - 1) (\delta_{ij} + A_{ij}^h) \right];$$

$$C_{ih} = \frac{w_{sh}}{2\psi_{ss}} A_{is}^h;$$

$$D_h = -\frac{\theta_h w_{sh}^2}{\psi_{ss}};$$

$$E_{i\alpha\beta}^h = \sum_{r=1, r \neq h}^p \frac{w_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r})}{\theta_h - \theta_r};$$

$$F_{i\alpha\beta} = \sum_{h=1}^k w_{ih} \left[w_{ih} w_{\alpha h} w_{\beta h} + (\theta_h - 1) E_{i\alpha\beta}^h \right];$$

$$G_{ih} = \frac{1}{(1 + \delta_{uv}) \sqrt{\psi_{uu} \psi_{vv}}} \left(w_{uh} \sum_{r=1, r \neq h}^p \frac{w_{ir} w_{vr}}{\theta_h - \theta_r} + w_{vh} \sum_{r=1, r \neq h}^p \frac{w_{ir} w_{ur}}{\theta_h - \theta_r} \right);$$

$$H_h = (2 - \delta_{uv}) \frac{w_{uh} w_{vh}}{\sqrt{\psi_{uu} \psi_{vv}}};$$

$$\frac{\partial A_{ij}^h}{\partial \psi_{ss}} = \sum_{r=1, r \neq h}^p \left(\frac{1}{\theta_r - \theta_h} \right)^2 [(\theta_r^2 - \theta_h^2) (C_{ir} w_{jr} + C_{jr} w_{ir}) + 2w_{ir} w_{jr} (\theta_r D_h - \theta_h D_r)];$$

$$\frac{\partial B_{ij}}{\partial \psi_{ss}} = \sum_{h=1}^k \left\{ (C_{ih} w_{jh} + C_{jh} w_{ih}) [-\theta_h w_{ih} w_{jh} + (\theta_h - 1) (\delta_{ij} + A_{ij}^h)] + w_{ih} w_{jh} [(-w_{ih} w_{jh} + \delta_{ij} + A_{ij}^h) D_h - \theta_h (C_{ih} w_{jh} + C_{jh} w_{ih}) + (\theta_h - 1) \frac{\partial A_{ij}^h}{\partial \psi_{ss}}] \right\};$$

$$\frac{\partial E_{i\alpha\beta}^h}{\partial \psi_{ss}} = \sum_{r=1, r \neq h}^p \left(\frac{1}{\theta_h - \theta_r} \right)^2 \{ (\theta_h - \theta_r) [C_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) + w_{ir} (C_{\alpha r} w_{\beta h} + C_{\beta h} w_{\alpha r} + C_{\beta r} w_{\alpha h} + C_{\alpha h} w_{\beta r})] + w_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) (D_r - D_h) \};$$

$$\frac{\partial E_{i\alpha\beta}^h}{\partial \sigma_{uv}} = \sum_{r=1, r \neq h}^p \left(\frac{1}{\theta_h - \theta_r} \right)^2 \{ (\theta_h - \theta_r) [G_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) + w_{ir} (G_{\alpha r} w_{\beta h} + G_{\beta h} w_{\alpha r} + G_{\beta r} w_{\alpha h} + G_{\alpha h} w_{\beta r})] + w_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) (H_r - H_h) \};$$

$$\frac{\partial F_{i\alpha\beta}}{\partial \psi_{ss}} = \sum_{h=1}^k \left\{ C_{ih} [w_{ih} w_{\alpha h} w_{\beta h} + (\theta_h - 1) E_{i\alpha\beta}^h] + w_{ih} [C_{ih} w_{\alpha h} w_{\beta h} + C_{\alpha h} w_{ih} w_{\beta h} + C_{\beta h} w_{ih} w_{\alpha h} + D_h E_{i\alpha\beta}^h + (\theta_h - 1) \frac{\partial E_{i\alpha\beta}^h}{\partial \psi_{ss}}] \right\};$$

$$\frac{\partial F_{i\alpha\beta}}{\partial \sigma_{uv}} = \sum_{h=1}^k \left\{ G_{ih} [w_{ih} w_{\alpha h} w_{\beta h} + (\theta_h - 1) E_{i\alpha\beta}^h] + w_{ih} [G_{ih} w_{\alpha h} w_{\beta h} + G_{\alpha h} w_{ih} w_{\beta h} + G_{\beta h} w_{ih} w_{\alpha h} + H_h E_{i\alpha\beta}^h + (\theta_h - 1) \frac{\partial E_{i\alpha\beta}^h}{\partial \sigma_{uv}}] \right\}.$$

Dem. Calculemos las derivadas $\partial^2 g_i / \partial \psi_{jj} \partial \psi_{ss}$.

Por Lema 23

$$\frac{\partial g_i}{\partial \psi_{jj}} = \frac{\psi_{ii}}{\psi_{jj}} B_{ij}.$$

Si $s \neq i$ y $s \neq j$

$$\frac{\partial^2 g_i}{\partial \psi_{jj} \partial \psi_{ss}} = \frac{\psi_{ii}}{\psi_{jj}} \frac{\partial B_{ij}}{\partial \psi_{ss}}.$$

Si $s = i$ y $s \neq j$

$$\frac{\partial^2 g_i}{\partial \psi_{jj} \partial \psi_{ii}} = \frac{1}{\psi_{jj}} \left(B_{ij} + \psi_{ii} \frac{\partial B_{ij}}{\partial \psi_{ii}} \right).$$

Si $s \neq i$ y $s = j$

$$\frac{\partial^2 g_i}{\partial^2 \psi_{jj}} = \frac{\psi_{ii}}{\psi_{jj}} \left(-\frac{B_{ij}}{\psi_{jj}} + \frac{\partial B_{ij}}{\partial \psi_{jj}} \right).$$

Si $s = i = j$

$$\frac{\partial^2 g_i}{\partial^2 \psi_{ii}} = \frac{\partial B_{ii}}{\partial \psi_{ii}}.$$

Por lo tanto,

$$\begin{aligned} \frac{\partial^2 g_i}{\partial \psi_{jj} \partial \psi_{ss}} &= \frac{1}{\psi_{jj}} \left\{ \psi_{ii} \frac{\partial B_{ij}}{\partial \psi_{ii}} + B_{ij} \left[\delta_{si} - \frac{\psi_{ii}}{\psi_{jj}} \delta_{sj} + \right. \right. \\ &\quad \left. \left. + \left(\frac{\psi_{ii}}{\psi_{jj}} - 1 \right) \delta_{si} \delta_{sj} \right] \right\}. \end{aligned}$$

Calculemos $\partial B_{ij} / \partial \psi_{ss}$.

$$\begin{aligned} \frac{\partial B_{ij}}{\partial \psi_{ss}} &= \sum_{h=1}^k \left\{ \left(\frac{\partial w_{ih}}{\partial \psi_{ss}} w_{jh} + \frac{\partial w_{jh}}{\partial \psi_{ss}} w_{ih} \right) [-\theta_h w_{ih} w_{jh} + \right. \\ &\quad \left. + (\theta_h - 1) (\delta_{ij} + A_{ij}^h)] + w_{ih} w_{jh} [(-w_{ih} w_{jh} + \right. \\ &\quad \left. + \delta_{ij} + A_{ij}^h) \frac{\partial \theta_h}{\partial \psi_{ss}} - \theta_h \left(\frac{\partial w_{ih}}{\partial \psi_{ss}} w_{jh} + \frac{\partial w_{jh}}{\partial \psi_{ss}} w_{ih} \right) + \right. \\ &\quad \left. + (\theta_h - 1) \frac{\partial A_{ij}^h}{\partial \psi_{ss}} \right] \left. \right\}. \end{aligned} \quad (140)$$

Sean

$$C_{ih} = \frac{\partial w_{ih}}{\partial \psi_{ss}};$$

$$D_h = \frac{\partial \theta_h}{\partial \psi_{ss}}.$$

Entonces, por (128) y (129) del Lema 23

$$C_{ih} = \frac{w_{sh}}{2\psi_{ss}} A_{is}^h; \quad (141)$$

$$D_h = -\frac{\theta_h w_{sh}^2}{\psi_{ss}}. \quad (142)$$

Luego, reemplazando en (140)

$$\begin{aligned} \frac{\partial B_{ij}}{\partial \psi_{ss}} = & \sum_{h=1}^k \left\{ (C_{ih} w_{jh} + C_{jh} w_{ih}) [-\theta_h w_{ih} w_{jh} + \right. \\ & + (\theta_h - 1) (\delta_{ij} + A_{ij}^h)] + w_{ih} w_{jh} [(-w_{ih} w_{jh} + \\ & + \delta_{ij} + A_{ij}^h) D_h - \theta_h (C_{ih} w_{jh} + C_{jh} w_{ih}) + \\ & \left. + (\theta_h - 1) \frac{\partial A_{ij}^h}{\partial \psi_{ss}} \right\}. \end{aligned}$$

Además,

$$\begin{aligned} \frac{\partial A_{ij}^h}{\partial \psi_{ss}} = & \sum_{r=1, r \neq h}^p \left\{ \left(\frac{\theta_r + \theta_h}{\theta_r - \theta_h} \right) \left(\frac{\partial w_{ir}}{\partial \psi_{ss}} w_{jr} + \frac{\partial w_{jr}}{\partial \psi_{ss}} w_{ir} \right) + \right. \\ & + \frac{w_{ir} w_{jr}}{(\theta_r - \theta_h)^2} \left[(\theta_r - \theta_h) \left(\frac{\partial \theta_r}{\partial \psi_{ss}} + \frac{\partial \theta_h}{\partial \psi_{ss}} \right) - \right. \\ & \left. \left. - (\theta_r + \theta_h) \left(\frac{\partial \theta_r}{\partial \psi_{ss}} - \frac{\partial \theta_h}{\partial \psi_{ss}} \right) \right] \right\}. \end{aligned}$$

Por (141) y (142)

$$\begin{aligned} \frac{\partial A_{ij}^h}{\partial \psi_{ss}} = & \sum_{r=1, r \neq h}^p \left(\frac{1}{\theta_r - \theta_h} \right)^2 [(\theta_r^2 - \theta_h^2) (C_{ir} w_{jr} + C_{jr} w_{ir}) + \\ & + 2w_{ir} w_{jr} (\theta_r D_h - \theta_h D_r)]. \end{aligned}$$

Calculemos las derivadas $\partial^2 g_i / \partial \psi_{jj} \partial \sigma_{\alpha\beta}$.

Por Lema 23

$$\frac{\partial g_i}{\partial \sigma_{\alpha\beta}} = \frac{(2 - \delta_{\alpha\beta}) \psi_{ii}}{\sqrt{\psi_{\alpha\alpha} \psi_{\beta\beta}}} F_{i\alpha\beta}.$$

Si $s \neq i, s \neq \alpha$ y $s \neq \beta$

$$\frac{\partial^2 g_i}{\partial \psi_{ss} \partial \sigma_{\alpha\beta}} = \frac{2\psi_{ii}}{\sqrt{\psi_{\alpha\alpha}\psi_{\beta\beta}}} \frac{\partial F_{i\alpha\beta}}{\partial \psi_{ss}}.$$

Si $s \neq i, s = \alpha$ y $s \neq \beta$

$$\frac{\partial^2 g_i}{\partial \psi_{\alpha\alpha} \partial \sigma_{\alpha\beta}} = \frac{\psi_{ii}}{\sqrt{\psi_{\alpha\alpha}\psi_{\beta\beta}}} \left(-\frac{F_{i\alpha\beta}}{\psi_{\alpha\alpha}} + 2 \frac{\partial F_{i\alpha\beta}}{\partial \psi_{\alpha\alpha}} \right).$$

Si $s \neq i, s \neq \alpha$ y $s = \beta$

$$\frac{\partial^2 g_i}{\partial \psi_{\beta\beta} \partial \sigma_{\alpha\beta}} = \frac{\psi_{ii}}{\sqrt{\psi_{\alpha\alpha}\psi_{\beta\beta}}} \left(-\frac{F_{i\alpha\beta}}{\psi_{\beta\beta}} + 2 \frac{\partial F_{i\alpha\beta}}{\partial \psi_{\beta\beta}} \right).$$

Si $s \neq i, s = \alpha = \beta$

$$\frac{\partial^2 g_i}{\partial \psi_{\alpha\alpha} \partial \sigma_{\alpha\alpha}} = \frac{\psi_{ii}}{\psi_{\alpha\alpha}} \left(-\frac{F_{i\alpha\alpha}}{\psi_{\alpha\alpha}} + \frac{\partial F_{i\alpha\alpha}}{\partial \psi_{\alpha\alpha}} \right).$$

Si $s = i, s \neq \alpha$ y $s \neq \beta$

$$\frac{\partial^2 g_i}{\partial \psi_{ii} \partial \sigma_{\alpha\beta}} = \frac{2}{\sqrt{\psi_{\alpha\alpha}\psi_{\beta\beta}}} \left(F_{i\alpha\beta} + \psi_{ii} \frac{\partial F_{i\alpha\beta}}{\partial \psi_{ii}} \right).$$

Si $s = i = \alpha = \beta$

$$\frac{\partial^2 g_i}{\partial \psi_{ii} \partial \sigma_{ii}} = \frac{\partial F_{iii}}{\partial \psi_{ii}}.$$

Si $s = i = \alpha$ y $s \neq \beta$

$$\frac{\partial^2 g_i}{\partial \psi_{ii} \partial \sigma_{i\beta}} = \frac{1}{\sqrt{\psi_{ii}\psi_{\beta\beta}}} \left(F_{ii\beta} + 2\psi_{ii} \frac{\partial F_{ii\beta}}{\partial \psi_{ii}} \right).$$

Si $s = i = \beta$ y $s \neq \alpha$

$$\frac{\partial^2 g_i}{\partial \psi_{ii} \partial \sigma_{\alpha i}} = \frac{1}{\sqrt{\psi_{ii}\psi_{\alpha\alpha}}} \left(F_{i\alpha i} + 2\psi_{ii} \frac{\partial F_{i\alpha i}}{\partial \psi_{ii}} \right).$$

Por lo tanto,

$$\begin{aligned} \frac{\partial^2 g_i}{\partial \psi_{jj} \partial \sigma_{\alpha\beta}} &= \frac{\psi_{ii}}{\sqrt{\psi_{\alpha\alpha}\psi_{\beta\beta}}} \left\{ (2 - \delta_{\alpha\beta}) \frac{\partial F_{i\alpha\beta}}{\partial \psi_{ss}} + \right. \\ &+ \left\{ \frac{\delta_{si}}{\psi_{ii}} \left[2 + \left(\frac{\psi_{ii}}{\psi_{\alpha\alpha}} - 1 \right) \delta_{s\alpha} + \left(\frac{\psi_{ii}}{\psi_{\beta\beta}} - 1 \right) \delta_{s\beta} \right] - \right. \\ &\left. \left. - \frac{\delta_{s\alpha}}{\psi_{\alpha\alpha}} - \frac{\delta_{s\beta}}{\psi_{\beta\beta}} + \frac{\delta_{s\alpha}\delta_{s\beta}}{\psi_{\beta\beta}} (1 - \delta_{si}) \right\} F_{i\alpha\beta} \right\}. \end{aligned}$$

Calculemos $\partial F_{i\alpha\beta} / \partial \psi_{ss}$

$$\begin{aligned} \frac{\partial F_{i\alpha\beta}}{\partial \psi_{ss}} &= \sum_{h=1}^k \left\{ \frac{\partial w_{ih}}{\partial \psi_{ss}} \left[w_{ih} w_{\alpha h} w_{\beta h} + (\theta_h - 1) E_{i\alpha\beta}^h \right] + \right. \\ &\quad + w_{ih} \left[\frac{\partial w_{ih}}{\partial \psi_{ss}} w_{\alpha h} w_{\beta h} + \frac{\partial w_{\alpha h}}{\partial \psi_{ss}} w_{ih} w_{\beta h} + \frac{\partial w_{\beta h}}{\partial \psi_{ss}} w_{ih} w_{\alpha h} + \right. \\ &\quad \left. \left. + \frac{\partial \theta_h}{\partial \psi_{ss}} E_{i\alpha\beta}^h + (\theta_h - 1) \frac{\partial E_{i\alpha\beta}^h}{\partial \psi_{ss}} \right] \right\}. \end{aligned}$$

Por (141) y (142)

$$\begin{aligned} \frac{\partial F_{i\alpha\beta}}{\partial \psi_{ss}} &= \sum_{h=1}^k \left\{ C_{ih} \left[w_{ih} w_{\alpha h} w_{\beta h} + (\theta_h - 1) E_{i\alpha\beta}^h \right] + \right. \\ &\quad + w_{ih} \left[C_{ih} w_{\alpha h} w_{\beta h} + C_{\alpha h} w_{ih} w_{\beta h} + C_{\beta h} w_{ih} w_{\alpha h} + \right. \\ &\quad \left. \left. + D_h E_{i\alpha\beta}^h + (\theta_h - 1) \frac{\partial E_{i\alpha\beta}^h}{\partial \psi_{ss}} \right] \right\}. \end{aligned}$$

Además,

$$\begin{aligned} \frac{\partial E_{i\alpha\beta}^h}{\partial \psi_{ss}} &= \sum_{r=1, r \neq h}^p \left\{ \left(\frac{1}{\theta_h - \theta_r} \right) \left[\frac{\partial w_{ir}}{\partial \psi_{ss}} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) + \right. \right. \\ &\quad \left. \left. + w_{ir} \left(\frac{\partial w_{\alpha r}}{\partial \psi_{ss}} w_{\beta h} + \frac{\partial w_{\beta h}}{\partial \psi_{ss}} w_{\alpha r} + \frac{\partial w_{\alpha h}}{\partial \psi_{ss}} w_{\beta r} + \frac{\partial w_{\beta r}}{\partial \psi_{ss}} w_{\alpha h} \right) \right] + \right. \\ &\quad \left. + \frac{w_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r})}{(\theta_h - \theta_r)^2} \left(\frac{\partial \theta_r}{\partial \psi_{ss}} - \frac{\partial \theta_h}{\partial \psi_{ss}} \right) \right\}. \end{aligned}$$

Luego, por (141) y (142)

$$\begin{aligned} \frac{\partial E_{i\alpha\beta}^h}{\partial \psi_{ss}} &= \sum_{r=1, r \neq h}^p \left(\frac{1}{\theta_h - \theta_r} \right)^2 \{ (\theta_h - \theta_r) [C_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) + \\ &\quad + w_{ir} (C_{\alpha r} w_{\beta h} + C_{\beta h} w_{\alpha r} + C_{\beta r} w_{\alpha h} + C_{\alpha h} w_{\beta r})] + \\ &\quad + w_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) (D_r - D_h) \}. \end{aligned}$$

Calculemos las derivadas $\partial^2 g_i / \partial \sigma_{\alpha\beta} \partial \sigma_{uv}$

Sean

$$G_{ih} = \frac{\partial w_{ih}}{\partial \sigma_{uv}};$$

$$H_h = \frac{\partial \theta_h}{\partial \sigma_{uv}}.$$

Entonces, por (133) y (134) del Lema 23

$$G_{ih} = \frac{1}{(1 + \delta_{uv}) \sqrt{\psi_{uu} \psi_{vv}}} \left(w_{uh} \sum_{r=1, r \neq h}^p \frac{w_{ir} w_{vr}}{\theta_h - \theta_r} + w_{vh} \sum_{r=1, r \neq h}^p \frac{w_{ir} w_{ur}}{\theta_h - \theta_r} \right), \quad (143)$$

y por (135) y (136) del Lema 23

$$H_h = (2 - \delta_{uv}) \frac{w_{uh} w_{vh}}{\sqrt{\psi_{uu} \psi_{vv}}}. \quad (144)$$

Por Lema 23

$$\frac{\partial^2 g_i}{\partial \sigma_{\alpha\beta} \partial \sigma_{uv}} = \frac{(2 - \delta_{\alpha\beta}) \psi_{ii}}{\sqrt{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \frac{\partial F_{i\alpha\beta}}{\partial \sigma_{uv}}.$$

Calculemos $\partial F_{i\alpha\beta} / \partial \sigma_{uv}$

$$\begin{aligned} \frac{\partial F_{i\alpha\beta}}{\partial \sigma_{uv}} &= \sum_{h=1}^k \left\{ \frac{\partial w_{ih}}{\partial \sigma_{uv}} \left[w_{ih} w_{\alpha h} w_{\beta h} + (\theta_h - 1) E_{i\alpha\beta}^h \right] + \right. \\ &\quad + w_{ih} \left[\frac{\partial w_{ih}}{\partial \sigma_{uv}} w_{\alpha h} w_{\beta h} + \frac{\partial w_{\alpha h}}{\partial \sigma_{uv}} w_{ih} w_{\beta h} + \frac{\partial w_{\beta h}}{\partial \sigma_{uv}} w_{ih} w_{\alpha h} + \right. \\ &\quad \left. \left. + \frac{\partial \theta_h}{\partial \sigma_{uv}} E_{i\alpha\beta}^h + (\theta_h - 1) \frac{\partial E_{i\alpha\beta}^h}{\partial \sigma_{uv}} \right] \right\}. \end{aligned}$$

Por (143) y (144)

$$\begin{aligned} \frac{\partial F_{i\alpha\beta}}{\partial \sigma_{uv}} &= \sum_{h=1}^k \left\{ G_{ih} \left[w_{ih} w_{\alpha h} w_{\beta h} + (\theta_h - 1) E_{i\alpha\beta}^h \right] + \right. \\ &\quad + w_{ih} \left[G_{ih} w_{\alpha h} w_{\beta h} + G_{\alpha h} w_{ih} w_{\beta h} + G_{\beta h} w_{ih} w_{\alpha h} + \right. \\ &\quad \left. \left. + H_h E_{i\alpha\beta}^h + (\theta_h - 1) \frac{\partial E_{i\alpha\beta}^h}{\partial \sigma_{uv}} \right] \right\}. \end{aligned}$$

Además,

$$\begin{aligned} \frac{\partial E_{i\alpha\beta}^h}{\partial \sigma_{uv}} &= \sum_{r=1, r \neq h}^p \left\{ \left(\frac{1}{\theta_h - \theta_r} \right) \left[\frac{\partial w_{ir}}{\partial \sigma_{uv}} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) + \right. \right. \\ &\quad \left. \left. + w_{ir} \left(\frac{\partial w_{\alpha r}}{\partial \sigma_{uv}} w_{\beta h} + \frac{\partial w_{\beta h}}{\partial \sigma_{uv}} w_{\alpha r} + \frac{\partial w_{\alpha h}}{\partial \sigma_{uv}} w_{\beta r} + \frac{\partial w_{\beta r}}{\partial \sigma_{uv}} w_{\alpha h} \right) \right] + \right. \\ &\quad \left. + \frac{w_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r})}{(\theta_h - \theta_r)^2} \left(\frac{\partial \theta_r}{\partial \sigma_{uv}} - \frac{\partial \theta_h}{\partial \sigma_{uv}} \right) \right\}. \end{aligned}$$

Luego, por (143) y (144)

$$\begin{aligned} \frac{\partial E_{\alpha\beta}^h}{\partial\sigma_{uv}} &= \sum_{r=1, r \neq h}^p \left(\frac{1}{\theta_h - \theta_r} \right)^2 \{ (\theta_h - \theta_r) [G_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) + \\ &+ w_{ir} (G_{\alpha r} w_{\beta h} + G_{\beta h} w_{\alpha r} + G_{\beta r} w_{\alpha h} + G_{\alpha h} w_{\beta r})] + \\ &+ w_{ir} (w_{\alpha r} w_{\beta h} + w_{\alpha h} w_{\beta r}) (H_r - H_h) \}. \end{aligned}$$

Lema 26

$$\begin{aligned} \frac{\partial^2 \lambda_{ij}}{\partial\psi_{ii} \partial\psi_{ss}} &= \frac{\sqrt{\psi_{ii}}}{2\psi_{ii}} \left\{ \frac{\partial I_{ijl}}{\partial\psi_{ss}} + \left[\frac{\delta_{si}}{2\psi_{ii}} - \frac{\delta_{sl}}{\psi_{ii}} + \right. \right. \\ &\left. \left. + \left(\frac{1}{\psi_{ii}} - \frac{3}{2\psi_{ii}} \right) \delta_{il} \right] I_{ijl} \right\}; \end{aligned} \quad (145)$$

$$\begin{aligned} \frac{\partial^2 \lambda_{ij}}{\partial\psi_{ss} \partial\sigma_{\alpha\beta}} &= \sqrt{\frac{\psi_{ii}}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \left\{ \frac{\partial J_{ij\alpha\beta}}{\partial\psi_{ss}} + \frac{J_{ij\alpha\beta}}{2} \left\{ \delta_{si} \left[\frac{1}{\psi_{ii}} + \right. \right. \right. \\ &+ \left. \left. \left(\frac{1}{\psi_{\alpha\alpha}} - \frac{1}{\psi_{\beta\beta}} - \frac{1}{\psi_{ii}} \right) \delta_{s\alpha} + \left(-\frac{1}{\psi_{\alpha\alpha}} + \frac{1}{\psi_{\beta\beta}} - \right. \right. \right. \\ &\left. \left. \left. - \frac{1}{\psi_{ii}} \right) \delta_{s\beta} + \frac{2\delta_{s\alpha} \delta_{s\beta}}{\psi_{\alpha\alpha}} \right] - \frac{\delta_{s\alpha}}{\psi_{\alpha\alpha}} - \frac{\delta_{s\beta}}{\psi_{\beta\beta}} \right\} \right\}; \end{aligned} \quad (146)$$

$$\frac{\partial^2 \lambda_{ij}}{\partial\sigma_{\alpha\beta} \partial\sigma_{uv}} = \sqrt{\frac{\psi_{ii}}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \frac{\partial J_{ij\alpha\beta}}{\partial\sigma_{uv}}, \quad (147)$$

donde

$$H_h = (2 - \delta_{uv}) \frac{w_{uh} w_{vh}}{\sqrt{\psi_{uu} \psi_{vv}}};$$

$$I_{ijl} = w_{lj} \sqrt{\theta_j - 1} \left(\delta_{il} + A_{il}^j - \frac{\theta_j w_{ij} w_{lj}}{\theta_j - 1} \right);$$

$$J_{ij\alpha\beta} = \sqrt{\theta_j - 1} \left[(2 - \delta_{\alpha\beta}) E_{i\alpha\beta}^j + \left(1 - \frac{\delta_{\alpha\beta}}{2} \right) \frac{w_{\alpha j} w_{ij} w_{\beta j}}{\theta_j - 1} \right];$$

$$\begin{aligned} \frac{\partial I_{ijl}}{\partial\psi_{ss}} &= \left(C_{lj} \sqrt{\theta_j - 1} + \frac{w_{lj}}{2\sqrt{\theta_j - 1}} D_j \right) \left(\delta_{il} + A_{il}^j - \frac{\theta_j w_{ij} w_{lj}}{\theta_j - 1} \right) + \\ &+ w_{lj} \sqrt{\theta_j - 1} \left\{ \frac{\partial A_{il}^j}{\partial\psi_{ss}} - \frac{1}{(\theta_j - 1)^2} [(w_{ij} w_{lj} D_j + \right. \\ &\left. + \theta_j w_{lj} C_{ij} + \theta_j w_{ij} C_{lj}) (\theta_j - 1) - \theta_j w_{ij} w_{lj} D_j] \right\}; \end{aligned}$$

$$\begin{aligned} \frac{\partial J_{ij\alpha\beta}}{\partial \psi_{ss}} &= \frac{1}{2\sqrt{\theta_j-1}} D_j \left[(2 - \delta_{\alpha\beta}) E_{i\alpha\beta}^j + \left(1 - \frac{\delta_{\alpha\beta}}{2}\right) \frac{w_{\alpha j} w_{i j} w_{\beta j}}{\theta_j - 1} \right] + \\ &+ \sqrt{\theta_j - 1} \left\{ (2 - \delta_{\alpha\beta}) \frac{\partial E_{i\alpha\beta}^j}{\partial \psi_{ss}} + \frac{(1 - \delta_{\alpha\beta}/2)}{(\theta_j - 1)^2} [(\theta_j - \right. \\ &- 1) (w_{i j} w_{\beta j} C_{\alpha j} + w_{\alpha j} w_{\beta j} C_{i j} + w_{i j} w_{\alpha j} C_{\beta j}) - \\ &\left. - w_{\alpha j} w_{i j} w_{\beta j} D_j] \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial J_{ij\alpha\beta}}{\partial \sigma_{uv}} &= \frac{1}{2\sqrt{\theta_j-1}} H_j \left[(2 - \delta_{\alpha\beta}) E_{i\alpha\beta}^j + \left(1 - \frac{\delta_{\alpha\beta}}{2}\right) \frac{w_{\alpha j} w_{i j} w_{\beta j}}{\theta_j - 1} \right] + \\ &+ \sqrt{\theta_j - 1} \left\{ (2 - \delta_{\alpha\beta}) \frac{\partial E_{i\alpha\beta}^j}{\partial \sigma_{uv}} + \frac{(1 - \delta_{\alpha\beta}/2)}{(\theta_j - 1)^2} [(\theta_j - \right. \\ &- 1) (w_{i j} w_{\beta j} G_{\alpha j} + w_{\alpha j} w_{\beta j} G_{i j} + w_{i j} w_{\alpha j} G_{\beta j}) - \\ &\left. - w_{\alpha j} w_{i j} w_{\beta j} H_j] \right\}; \end{aligned}$$

y $A_{ij}^h, C_{ih}, D_h, E_{i\alpha\beta}^h, G_{ih}, \frac{\partial A_{ij}^h}{\partial \psi_{ss}}, \frac{\partial E_{i\alpha\beta}^h}{\partial \psi_{ss}}, \frac{\partial E_{i\alpha\beta}^h}{\partial \sigma_{uv}}$ son los mismos que en el Lema 25.

Dem. Calculemos las derivadas $\partial^2 \lambda_{ij} / \partial \psi_{ii} \partial \psi_{ss}$.

Por Lema 24

$$\frac{\partial \lambda_{ij}}{\partial \psi_{ii}} = \frac{\sqrt{\psi_{ii}}}{2\psi_{ii}} I_{ijl}.$$

Si $s \neq i$ y $s \neq l$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{ii} \partial \psi_{ss}} = \frac{\sqrt{\psi_{ii}}}{2\psi_{ii}} \frac{\partial I_{ijl}}{\partial \psi_{ss}}.$$

Si $s = i$ y $s \neq l$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{ii} \partial \psi_{ii}} = \frac{\sqrt{\psi_{ii}}}{2\psi_{ii}} \left(\frac{I_{ijl}}{2\psi_{ii}} + \frac{\partial I_{ijl}}{\partial \psi_{ii}} \right).$$

Si $s \neq i$ y $s = l$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{ii} \partial \psi_{ll}} = \frac{\sqrt{\psi_{ii}}}{2\psi_{ii}} \left(-\frac{I_{ijl}}{\psi_{ll}} + \frac{\partial I_{ijl}}{\partial \psi_{ll}} \right).$$

Si $s = i = l$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{ii} \partial \psi_{ii}} = \frac{1}{2\sqrt{\psi_{ii}}} \left(-\frac{I_{iji}}{\psi_{ii}} + \frac{\partial I_{iji}}{\partial \psi_{ii}} \right).$$

Por lo tanto,

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{ii} \partial \psi_{ss}} = \frac{\sqrt{\psi_{ii}}}{2\psi_{ii}} \left\{ \frac{\partial I_{ijl}}{\partial \psi_{ss}} + \left[\frac{\delta_{si}}{2\psi_{ii}} - \frac{\delta_{sl}}{\psi_{ii}} + \left(\frac{1}{\psi_{ii}} - \frac{3}{2\psi_{ii}} \right) \delta_{il} \right] I_{ijl} \right\}.$$

Calculemos $\partial I_{ijl} / \partial \psi_{ss}$.

$$\begin{aligned} \frac{\partial I_{ijl}}{\partial \psi_{ss}} &= \left(\frac{\partial w_{lj}}{\partial \psi_{ss}} \sqrt{\theta_j - 1} + \frac{w_{lj}}{2\sqrt{\theta_j - 1}} \frac{\partial \theta_j}{\partial \psi_{ss}} \right) \left(\delta_{il} + A_{il}^j - \frac{\theta_j w_{ij} w_{lj}}{\theta_j - 1} \right) + \\ &+ w_{lj} \sqrt{\theta_j - 1} \left\{ \frac{\partial A_{il}^j}{\partial \psi_{ss}} - \frac{1}{(\theta_j - 1)^2} \left[\left(w_{ij} w_{lj} \frac{\partial \theta_j}{\partial \psi_{ss}} + \right. \right. \right. \\ &\left. \left. \left. + \theta_j w_{lj} \frac{\partial w_{ij}}{\partial \psi_{ss}} + \theta_j w_{ij} \frac{\partial w_{lj}}{\partial \psi_{ss}} \right) (\theta_j - 1) - \theta_j w_{ij} w_{lj} \frac{\partial \theta_j}{\partial \psi_{ss}} \right] \right\}. \end{aligned}$$

Entonces, por (141) y (139) del Lema 25

$$\begin{aligned} \frac{\partial I_{ijl}}{\partial \psi_{ss}} &= \left(C_{lj} \sqrt{\theta_j - 1} + \frac{w_{lj}}{2\sqrt{\theta_j - 1}} D_j \right) \left(\delta_{il} + A_{il}^j - \frac{\theta_j w_{ij} w_{lj}}{\theta_j - 1} \right) + \\ &+ w_{lj} \sqrt{\theta_j - 1} \left\{ \frac{\partial A_{il}^j}{\partial \psi_{ss}} - \frac{1}{(\theta_j - 1)^2} \left[(w_{ij} w_{lj} D_j + \right. \right. \\ &\left. \left. + \theta_j w_{lj} C_{ij} + \theta_j w_{ij} C_{lj}) (\theta_j - 1) - \theta_j w_{ij} w_{lj} D_j \right] \right\}. \end{aligned}$$

Calculemos las derivadas $\partial^2 \lambda_{ij} / \partial \psi_{ii} \partial \sigma_{\alpha\beta}$.

Por Lema 24

$$\frac{\partial \lambda_{ij}}{\partial \sigma_{\alpha\beta}} = \sqrt{\frac{\psi_{ii}}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} J_{ij\alpha\beta}.$$

Si $s \neq i$, $s \neq \alpha$ y $s \neq \beta$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{ss} \partial \sigma_{\alpha\beta}} = \sqrt{\frac{\psi_{ii}}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \frac{\partial J_{ij\alpha\beta}}{\partial \psi_{ss}}.$$

Si $s \neq i$, $s = \alpha$ y $s \neq \beta$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{\alpha\alpha} \partial \sigma_{\alpha\beta}} = \sqrt{\frac{\psi_{ii}}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \left(-\frac{J_{ij\alpha\beta}}{2\psi_{\alpha\alpha}} + \frac{\partial J_{ij\alpha\beta}}{\partial \psi_{\alpha\alpha}} \right).$$

Si $s \neq i, s \neq \alpha$ y $s = \beta$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{\beta\beta} \partial \sigma_{\alpha\beta}} = \sqrt{\frac{\psi_{ii}}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \left(-\frac{J_{ij\alpha\beta}}{2\psi_{\beta\beta}} + \frac{\partial J_{ij\alpha\beta}}{\partial \psi_{\beta\beta}} \right).$$

Si $s \neq i, s = \alpha = \beta$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{\alpha\alpha} \partial \sigma_{\alpha\alpha}} = \frac{\sqrt{\psi_{ii}}}{\psi_{\alpha\alpha}} \left(-\frac{J_{ij\alpha\alpha}}{\psi_{\alpha\alpha}} + \frac{\partial J_{ij\alpha\alpha}}{\partial \psi_{\alpha\alpha}} \right).$$

Si $s = i, s \neq \alpha$ y $s \neq \beta$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{ii} \partial \sigma_{\alpha\beta}} = \sqrt{\frac{\psi_{ii}}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \left(\frac{J_{ij\alpha\beta}}{2\psi_{ii}} + \frac{\partial J_{ij\alpha\beta}}{\partial \psi_{ii}} \right).$$

Si $s = i = \alpha = \beta$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{ii} \partial \sigma_{ii}} = \frac{1}{\sqrt{\psi_{ii}}} \left(-\frac{J_{ijii}}{2\psi_{ii}} + \frac{\partial J_{ijii}}{\partial \psi_{ii}} \right).$$

Si $s = i = \alpha$ y $s \neq \beta$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{ii} \partial \sigma_{i\beta}} = \frac{1}{\sqrt{\psi_{\beta\beta}}} \left(-\frac{J_{ij\beta}}{2\psi_{\beta\beta}} + \frac{\partial J_{ij\beta}}{\partial \psi_{ii}} \right).$$

Si $s = i = \beta$ y $s \neq \alpha$

$$\frac{\partial^2 \lambda_{ij}}{\partial \psi_{ii} \partial \sigma_{\alpha i}} = \frac{1}{\sqrt{\psi_{\alpha\alpha}}} \left(-\frac{J_{ij\alpha i}}{2\psi_{\alpha\alpha}} + \frac{\partial J_{ij\alpha i}}{\partial \psi_{ii}} \right).$$

Por lo tanto,

$$\begin{aligned} \frac{\partial^2 \lambda_{ij}}{\partial \psi_{ss} \partial \sigma_{\alpha\beta}} &= \sqrt{\frac{\psi_{ii}}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \left\{ \frac{\partial J_{ij\alpha\beta}}{\partial \psi_{ss}} + \frac{J_{ij\alpha\beta}}{2} \left\{ \delta_{si} \left[\frac{1}{\psi_{ii}} + \right. \right. \right. \\ &+ \left. \left. \left(\frac{1}{\psi_{\alpha\alpha}} - \frac{1}{\psi_{\beta\beta}} - \frac{1}{\psi_{ii}} \right) \delta_{s\alpha} + \left(-\frac{1}{\psi_{\alpha\alpha}} + \frac{1}{\psi_{\beta\beta}} - \right. \right. \right. \\ &\left. \left. \left. - \frac{1}{\psi_{ii}} \right) \delta_{s\beta} + \frac{2\delta_{s\alpha}\delta_{s\beta}}{\psi_{\alpha\alpha}} \right] - \frac{\delta_{s\alpha}}{\psi_{\alpha\alpha}} - \frac{\delta_{s\beta}}{\psi_{\beta\beta}} \right\} \right\}. \end{aligned}$$

Calculemos $\partial J_{ij\alpha\beta} / \partial \psi_{ss}$.

$$\begin{aligned} \frac{\partial J_{ij\alpha\beta}}{\partial \psi_{ss}} &= \frac{1}{2\sqrt{\theta_j-1}} \frac{\partial \theta_j}{\partial \psi_{ss}} \left[(2 - \delta_{\alpha\beta}) E_{i\alpha\beta}^j + \left(1 - \frac{\delta_{\alpha\beta}}{2}\right) \frac{w_{\alpha j} w_{ij} w_{\beta j}}{\theta_j - 1} \right] + \\ &+ \sqrt{\theta_j - 1} \left\{ (2 - \delta_{\alpha\beta}) \frac{\partial E_{i\alpha\beta}^j}{\partial \psi_{ss}} + \frac{(1 - \delta_{\alpha\beta}/2)}{(\theta_j - 1)^2} \left[(\theta_j - \right. \right. \\ &- 1) \left(w_{ij} w_{\beta j} \frac{\partial w_{\alpha j}}{\partial \psi_{ss}} + w_{\alpha j} w_{\beta j} \frac{\partial w_{ij}}{\partial \psi_{ss}} + w_{ij} w_{\alpha j} \frac{\partial w_{\beta j}}{\partial \psi_{ss}} \right) - \\ &\left. \left. - w_{\alpha j} w_{ij} w_{\beta j} \frac{\partial \theta_j}{\partial \psi_{ss}} \right] \right\}. \end{aligned}$$

Luego, por (141) y (142) del Lema 25

$$\begin{aligned} \frac{\partial J_{ij\alpha\beta}}{\partial \psi_{ss}} &= \frac{1}{2\sqrt{\theta_j-1}} D_j \left[(2 - \delta_{\alpha\beta}) E_{i\alpha\beta}^j + \left(1 - \frac{\delta_{\alpha\beta}}{2}\right) \frac{w_{\alpha j} w_{ij} w_{\beta j}}{\theta_j - 1} \right] + \\ &+ \sqrt{\theta_j - 1} \left\{ (2 - \delta_{\alpha\beta}) \frac{\partial E_{i\alpha\beta}^j}{\partial \psi_{ss}} + \frac{(1 - \delta_{\alpha\beta}/2)}{(\theta_j - 1)^2} \left[(\theta_j - \right. \right. \\ &- 1) (w_{ij} w_{\beta j} C_{\alpha j} + w_{\alpha j} w_{\beta j} C_{ij} + w_{ij} w_{\alpha j} C_{\beta j}) - \\ &\left. \left. - w_{\alpha j} w_{ij} w_{\beta j} D_j \right] \right\}. \end{aligned}$$

Calculemos las derivadas $\partial^2 \lambda_{ij} / \partial \sigma_{\alpha\beta} \partial \sigma_{uv}$.

Por Lema 24

$$\frac{\partial^2 \lambda_{ij}}{\partial \sigma_{\alpha\beta} \partial \sigma_{uv}} = \sqrt{\frac{\psi_{ii}}{\psi_{\alpha\alpha} \psi_{\beta\beta}}} \frac{\partial J_{ij\alpha\beta}}{\partial \sigma_{uv}}.$$

Calculemos $\partial J_{ij\alpha\beta} / \partial \sigma_{uv}$.

$$\begin{aligned} \frac{\partial J_{ij\alpha\beta}}{\partial \sigma_{uv}} &= \frac{1}{2\sqrt{\theta_j-1}} \frac{\partial \theta_j}{\partial \sigma_{uv}} \left[(2 - \delta_{\alpha\beta}) E_{i\alpha\beta}^j + \left(1 - \frac{\delta_{\alpha\beta}}{2}\right) \frac{w_{\alpha j} w_{ij} w_{\beta j}}{\theta_j - 1} \right] + \\ &+ \sqrt{\theta_j - 1} \left\{ (2 - \delta_{\alpha\beta}) \frac{\partial E_{i\alpha\beta}^j}{\partial \sigma_{uv}} + \frac{(1 - \delta_{\alpha\beta}/2)}{(\theta_j - 1)^2} \left[(\theta_j - \right. \right. \\ &- 1) \left(w_{ij} w_{\beta j} \frac{\partial w_{\alpha j}}{\partial \sigma_{uv}} + w_{\alpha j} w_{\beta j} \frac{\partial w_{ij}}{\partial \sigma_{uv}} + w_{ij} w_{\alpha j} \frac{\partial w_{\beta j}}{\partial \sigma_{uv}} \right) - \\ &\left. \left. - w_{\alpha j} w_{ij} w_{\beta j} \frac{\partial \theta_j}{\partial \sigma_{uv}} \right] \right\}. \end{aligned}$$

Por (143) y (144) del Lema 25

$$\begin{aligned} \frac{\partial J_{i\alpha\beta}}{\partial \sigma_{uv}} &= \frac{1}{2\sqrt{\theta_j - 1}} H_j \left[(2 - \delta_{\alpha\beta}) E_{i\alpha\beta}^j + \left(1 - \frac{\delta_{\alpha\beta}}{2}\right) \frac{w_{\alpha j} w_{i j} w_{\beta j}}{\theta_j - 1} \right] + \\ &+ \sqrt{\theta_j - 1} \left\{ (2 - \delta_{\alpha\beta}) \frac{\partial E_{i\alpha\beta}^j}{\partial \sigma_{uv}} + \frac{(1 - \delta_{\alpha\beta}/2)}{(\theta_j - 1)^2} [(\theta_j - \right. \\ &- 1) (w_{i j} w_{\beta j} G_{\alpha j} + w_{\alpha j} w_{\beta j} G_{i j} + w_{i j} w_{\alpha j} G_{\beta j}) - \\ &\left. - w_{\alpha j} w_{i j} w_{\beta j} H_j] \right\}. \end{aligned}$$

En la demostración del Teorema 11 utilizaremos el siguiente Lema, que se obtiene como una aplicación del Lema 4.1 de Lehmann [19] (1983). El mismo nos permitirá obtener la distribución asintótica de los parámetros del modelo de análisis factorial a partir de la distribución asintótica de un estimador consistente de la matriz de covarianza.

Lema 27 Sea $\mathbf{T}_n \in \mathbb{R}^p$ una sucesión de vectores aleatorios tales que $\mathbf{T}_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Delta)$, y sea $Q_n \in \mathbb{R}^{p \times p}$ una sucesión de matrices aleatorias tales que $Q_n \xrightarrow{P} Q_0$, de modo que Q_0 es una matriz constante no singular. Si $\mathbf{Y}_n \in \mathbb{R}^{p \times n}$ satisfice

$$Q_n \mathbf{Y}_n = \mathbf{T}_n,$$

entonces

$$\mathbf{Y}_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, Q_0^{-1} \Delta Q_0^{-1}).$$

Notación: Por simplicidad de notación, en la demostración del Teorema 11, notaremos con $\tilde{\Sigma}$ a $\tilde{\Sigma}_n$, con $\hat{\Lambda}$ a $\hat{\Lambda}_n$, con $\hat{\Psi}$ a $\hat{\Psi}_n$, con Σ a Σ_0 , con Λ a Λ_0 y con Ψ a Ψ_0 .

Demostración del Teorema 11

Para encontrar la distribución asintótica de los parámetros del modelo de análisis factorial, veremos que $\hat{\Lambda}$ y $\hat{\Psi}$ se pueden expresar como función de $\tilde{\Sigma}$, de quien conocemos su distribución asintótica por hipótesis.

Luego, aplicando el Lema 27, obtendremos la distribución asintótica de dichos estimadores.

Como $\Sigma = \Lambda \Lambda' + \Psi$ con $\Lambda \in \mathcal{R}$, $\Psi \in \mathcal{D}_+(p)$, entonces:

$$\text{diag}(\Psi) = \text{diag}(\Sigma - \Lambda \Lambda').$$

Además, como $\hat{\Lambda}$ y $\hat{\Psi}$ satisfacen las ecuaciones de máxima verosimilitud, por (117):

$$\text{diag}(\widehat{\Psi}) = \text{diag}(\widetilde{\Sigma} - \widehat{\Lambda}\widehat{\Lambda}').$$

Es decir, para $1 \leq i \leq p$

$$\Psi_{ii} = \sigma_{ii} - (\Lambda\Lambda')_{ii}; \quad (148)$$

$$\widehat{\Psi}_{ii} = \widetilde{\sigma}_{ii} - (\widehat{\Lambda}\widehat{\Lambda}')_{ii}. \quad (149)$$

Sea $S^* = \widehat{\Psi}^{-1/2}\widetilde{\Sigma}\widehat{\Psi}^{-1/2}$ y sean $\widehat{\theta}_1 > \dots > \widehat{\theta}_p$ sus autovalores y $\widehat{\mathbf{w}}_1, \dots, \widehat{\mathbf{w}}_p$ sus correspondientes autovectores normalizados.

Como $\widetilde{\Sigma}$ es un estimador consistente de Σ , y por el Teorema 6, $\widehat{\Psi}$ es un estimador consistente de Ψ , entonces $\widetilde{\Sigma} \xrightarrow{p} \Sigma$ y $\widehat{\Psi} \xrightarrow{p} \Psi$, y por lo tanto

$$S^* = \widehat{\Psi}^{-1/2}\widetilde{\Sigma}\widehat{\Psi}^{-1/2} \xrightarrow{p} \Sigma^* = \Psi^{-1/2}\Sigma\Psi^{-1/2}.$$

Por el Teorema 18, si $\widehat{\theta}_i$ es autovalor simple de S^* con autovector normalizado $\widehat{\mathbf{w}}_i$ para $1 \leq i \leq p$, entonces $\widehat{\theta}_i$ y $\widehat{\mathbf{w}}_i$ son funciones continuas y derivables en un entorno de S^* y por lo tanto

$$\begin{aligned} \widehat{\theta}_i &\xrightarrow{p} \theta_i; \\ \widehat{\mathbf{w}}_i &\xrightarrow{p} \mathbf{w}_i. \end{aligned} \quad (150)$$

para $1 \leq i \leq p$, siendo θ_i autovalor de Σ^* con autovector \mathbf{w}_i .

Sea

$$\Upsilon = \begin{bmatrix} \text{diag}(\Psi) \\ \text{vec}_r(\Sigma) \end{bmatrix},$$

donde, siguiendo la notación usada anteriormente,

$$\text{vec}_r(\Sigma)' = (\sigma_{11}, \dots, \sigma_{p1}, \sigma_{22}, \dots, \sigma_{p2}, \dots, \sigma_{p-1,p-1}, \sigma_{p-1,p}, \sigma_{pp}).$$

Por la Proposición 3 y el Lema 22

$$(\widehat{\Lambda}\widehat{\Lambda}')_{ii} = g_i(\widehat{\Upsilon}) = \widehat{\Psi}_{ii} \sum_{h=1}^k \widehat{w}_{ih}^2 (\widehat{\theta}_h - 1),$$

luego, haciendo el desarrollo de Taylor

$$\begin{aligned} g_i(\widehat{\Upsilon}) &= g_i(\Upsilon) + \left. \frac{\partial g_i}{\partial \widehat{\Upsilon}} \right|_{\widehat{\Upsilon}=\Upsilon} (\widehat{\Upsilon} - \Upsilon) + \\ &+ \frac{1}{2} (\widehat{\Upsilon} - \Upsilon)' \left. \frac{\partial^2 g_i}{\partial \widehat{\Upsilon} \partial \widehat{\Upsilon}'} \right|_{\widehat{\Upsilon}=\Upsilon} (\widehat{\Upsilon} - \Upsilon), \end{aligned} \quad (151)$$

donde $\tilde{\Upsilon}$ es un punto en el segmento que une Υ con $\hat{\Upsilon}$.

Por (148) y (149)

$$\begin{aligned}
\hat{\Psi}_{ii} - \Psi_{ii} &= \\
&= \tilde{\sigma}_{ii} - \left(\hat{\Lambda} \hat{\Lambda}' \right)_{ii} - \sigma_{ii} + (\Lambda \Lambda')_{ii} \\
&= \tilde{\sigma}_{ii} - \sigma_{ii} - g_i \left(\text{diag} \left(\hat{\Psi} \right), \text{vec}_r \left(\tilde{\Sigma} \right) \right) + g_i \left(\text{diag} \left(\Psi \right), \text{vec}_r \left(\Sigma \right) \right),
\end{aligned} \tag{152}$$

y por (151)

$$\begin{aligned}
\hat{\Psi}_{ii} - \Psi_{ii} &= \\
&= \tilde{\sigma}_{ii} - \sigma_{ii} - \left. \frac{\partial g_i}{\partial \left[\text{diag} \left(\hat{\Psi} \right) \right]} \right|_{\hat{\Upsilon}=\Upsilon} \text{diag} \left(\hat{\Psi} - \Psi \right) - \\
&\quad - \left. \frac{\partial g_i}{\partial \left[\text{vec}_r \left(\tilde{\Sigma} \right) \right]} \right|_{\hat{\Upsilon}=\Upsilon} \text{vec}_r \left(\tilde{\Sigma} - \Sigma \right) - \\
&\quad - \frac{1}{2} \text{diag} \left(\hat{\Psi} - \Psi \right)' \left. \frac{\partial^2 g_i}{\partial \left[\text{diag} \left(\hat{\Psi} \right) \right] \partial \left[\text{diag} \left(\hat{\Psi} \right) \right]'} \right|_{\hat{\Upsilon}=\tilde{\Upsilon}} \text{diag} \left(\hat{\Psi} - \Psi \right) - \\
&\quad - \text{vec}_r \left(\tilde{\Sigma} - \Sigma \right)' \left. \frac{\partial^2 g_i}{\partial \left[\text{vec}_r \left(\tilde{\Sigma} \right) \right] \partial \left[\text{diag} \left(\hat{\Psi} \right) \right]'} \right|_{\hat{\Upsilon}=\tilde{\Upsilon}} \text{diag} \left(\hat{\Psi} - \Psi \right) - \\
&\quad - \frac{1}{2} \text{vec}_r \left(\tilde{\Sigma} - \Sigma \right)' \left. \frac{\partial^2 g_i}{\partial \left[\text{vec}_r \left(\tilde{\Sigma} \right) \right] \partial \left[\text{vec}_r \left(\tilde{\Sigma} \right) \right]'} \right|_{\hat{\Upsilon}=\tilde{\Upsilon}} \text{vec}_r \left(\tilde{\Sigma} - \Sigma \right).
\end{aligned}$$

$$\text{Sean } D^i = \frac{\partial g_i'}{\partial \left[\text{vec}_r \left(\tilde{\Sigma} \right) \right]}; \quad E^i = \frac{\partial g_i'}{\partial \left[\text{diag} \left(\hat{\Psi} \right) \right]};$$

$$F^i = \frac{\partial^2 g_i}{\partial \left[\text{diag} \left(\hat{\Psi} \right) \right] \partial \left[\text{diag} \left(\hat{\Psi} \right) \right]'}; \quad G^i = \frac{\partial^2 g_i}{\partial \left[\text{vec}_r \left(\tilde{\Sigma} \right) \right] \partial \left[\text{diag} \left(\hat{\Psi} \right) \right]'};$$

$$J^i = \frac{\partial^2 g_i}{\partial \left[\text{vec}_r \left(\tilde{\Sigma} \right) \right] \partial \left[\text{vec}_r \left(\tilde{\Sigma} \right) \right]'},$$

donde los gradientes están evaluados en $\hat{\Upsilon} = \Upsilon$.

Entonces:

$$\begin{aligned}
& \sqrt{n} \left[\tilde{\sigma}_{ii} - \sigma_{ii} - D^i \text{vec}_r (\tilde{\Sigma} - \Sigma) - \frac{1}{2} \text{vec}_r (\tilde{\Sigma} - \Sigma)' J^i \text{vec}_r (\tilde{\Sigma} - \Sigma) \right] = \\
& \sqrt{n} \left[\widehat{\Psi}_{ii} - \Psi_{ii} + E^i \text{diag} (\widehat{\Psi} - \Psi) + \frac{1}{2} \text{diag} (\widehat{\Psi} - \Psi)' F^i \text{diag} (\widehat{\Psi} - \Psi) + \right. \\
& \left. + \text{vec}_r (\tilde{\Sigma} - \Sigma)' G^i \text{diag} (\widehat{\Psi} - \Psi) \right].
\end{aligned} \tag{153}$$

Por ecuación (119) del Lema 22

$$\widehat{\lambda}_{ij} = \widehat{\lambda}_{ij} (\widehat{\Upsilon}) = \widehat{w}_{ij} \sqrt{\widehat{\Psi}_{ii} (\widehat{\theta}_j - 1)},$$

luego, haciendo el desarrollo de Taylor

$$\begin{aligned}
\widehat{\lambda}_{ij} = \lambda_{ij} (\widehat{\Upsilon}) &= \lambda_{ij} (\Upsilon) + \frac{\partial \lambda_{ij}}{\partial \widehat{\Upsilon}} \Big|_{\widehat{\Upsilon}=\Upsilon} (\widehat{\Upsilon} - \Upsilon) + \\
&+ \frac{1}{2} (\widehat{\Upsilon} - \Upsilon)' \frac{\partial^2 \lambda_{ij}}{\partial \widehat{\Upsilon} \partial \widehat{\Upsilon}'} \Big|_{\widehat{\Upsilon}=\tilde{\Upsilon}^*} (\widehat{\Upsilon} - \Upsilon),
\end{aligned}$$

donde $\tilde{\Upsilon}^*$ es un punto en el segmento que une Υ con $\widehat{\Upsilon}$.

Luego:

$$\begin{aligned}
& \widehat{\lambda}_{ij} - \lambda_{ij} = \\
& = \frac{\partial \lambda'_{ij}}{\partial [\text{diag} (\widehat{\Psi})]} \Big|_{\widehat{\Upsilon}=\Upsilon} \text{diag} (\widehat{\Psi} - \Psi) + \frac{\partial \lambda'_{ij}}{\partial [\text{vec}_r (\tilde{\Sigma})]} \Big|_{\widehat{\Upsilon}=\Upsilon} \text{vec}_r (\tilde{\Sigma} - \Sigma) + \\
& + \frac{1}{2} \text{diag} (\widehat{\Psi} - \Psi)' \frac{\partial^2 \lambda_{ij}}{\partial [\text{diag} (\widehat{\Psi})] \partial [\text{diag} (\widehat{\Psi})]'} \Big|_{\widehat{\Upsilon}=\tilde{\Upsilon}^*} \text{diag} (\widehat{\Psi} - \Psi) + \\
& + \text{vec}_r (\tilde{\Sigma} - \Sigma)' \frac{\partial^2 \lambda_{ij}}{\partial [\text{vec}_r (\tilde{\Sigma})] \partial [\text{diag} (\widehat{\Psi})]'} \Big|_{\widehat{\Upsilon}=\tilde{\Upsilon}^*} \text{diag} (\widehat{\Psi} - \Psi) + \\
& + \frac{1}{2} \text{vec}_r (\tilde{\Sigma} - \Sigma)' \frac{\partial^2 \lambda_{ij}}{\partial [\text{vec}_r (\tilde{\Sigma})] \partial [\text{vec}_r (\tilde{\Sigma})]'} \Big|_{\widehat{\Upsilon}=\tilde{\Upsilon}^*} \text{vec}_r (\tilde{\Sigma} - \Sigma).
\end{aligned}$$

$$\begin{aligned} \text{Sean } L^{ij} &= \frac{\partial \lambda'_{ij}}{\partial [\text{vec}_r(\tilde{\Sigma})]}; \quad H^{ij} = \frac{\partial \lambda'_{ij}}{\partial [\text{diag}(\hat{\Psi})]}; \\ N^{ij} &= \frac{\partial^2 \lambda_{ij}}{\partial [\text{diag}(\hat{\Psi})] \partial [\text{diag}(\hat{\Psi})]'}; \quad O^{ij} = \frac{\partial^2 \lambda_{ij}}{\partial [\text{vec}_r(\tilde{\Sigma})] \partial [\text{diag}(\hat{\Psi})]'}; \\ P^{ij} &= \frac{\partial^2 \lambda_{ij}}{\partial [\text{vec}_r(\tilde{\Sigma})] \partial [\text{vec}_r(\tilde{\Sigma})]'}, \end{aligned}$$

donde los gradientes están evaluados en $\hat{\Upsilon} = \Upsilon$.

Entonces:

$$\begin{aligned} &\sqrt{n} \left[L^{ij} \text{vec}_r(\tilde{\Sigma} - \Sigma) + \frac{1}{2} \text{vec}_r(\tilde{\Sigma} - \Sigma)' P^{ij} \text{vec}_r(\tilde{\Sigma} - \Sigma) \right] = \\ &= \sqrt{n} \left[\hat{\lambda}_{ij} - \lambda_{ij} - H^{ij} \text{diag}(\hat{\Psi} - \Psi) - \right. \\ &\quad \left. - \frac{1}{2} \text{diag}(\hat{\Psi} - \Psi)' N^{ij} \text{diag}(\hat{\Psi} - \Psi) - \right. \\ &\quad \left. - \text{vec}_r(\tilde{\Sigma} - \Sigma)' O^{ij} \text{diag}(\hat{\Psi} - \Psi) \right]. \end{aligned} \quad (154)$$

Sea $\tilde{E}^i \in \mathfrak{R}^p$ el vector tal que

$$(\tilde{E}^i)_j = \begin{cases} E_j^i & \text{si } i \neq j \\ 1 + E_i^i & \text{si } i = j \end{cases}$$

Sean: $m = p + pk$, $\beta = [(i-p)/(p+1)] + 1$, $\alpha = (i-p) - (\beta-1)p$ para $p+1 \leq i \leq m$, $\mathbf{T}'_n = (T'_{1n}, \dots, T'_{mn})$, donde

$$\frac{T'_{in}}{\sqrt{n}} = \begin{cases} \bar{\sigma}_{ii} - \sigma_{ii} - D^i \text{vec}_r(\tilde{\Sigma} - \Sigma) - \\ \quad - \frac{1}{2} \text{vec}_r(\tilde{\Sigma} - \Sigma)' J^i \text{vec}_r(\tilde{\Sigma} - \Sigma) & \text{para } 1 \leq i \leq p \\ L^{\alpha\beta} \text{vec}_r(\tilde{\Sigma} - \Sigma) + \\ \quad + \frac{1}{2} \text{vec}_r(\tilde{\Sigma} - \Sigma)' P^{ij} \text{vec}_r(\tilde{\Sigma} - \Sigma) & \text{para } p+1 \leq i \leq m \end{cases}$$

Definamos

$$\begin{aligned} \mathbf{Y}_n &= \sqrt{n} \left(\text{diag}(\hat{\Psi} - \Psi), \text{vec}(\hat{\Lambda} - \Lambda) \right)'; \\ Q_n &= \begin{bmatrix} Q_n^1 & \mathbf{0} \\ Q_n^3 & I \end{bmatrix}, \end{aligned} \quad (155)$$

donde

$$(Q_n^1)_{ij} = \widetilde{E}_j^i + \frac{1}{2} \sum_{s=1}^p F_{sj}^i (\widehat{\Psi}_{ss} - \Psi_{ss}) + \sum_{u=1}^p \sum_{v=u}^p C_{(2p-u)(u-1)/2+v,j}^i (\widetilde{\sigma}_{uv} - \sigma_{uv})$$

para $1 \leq i \leq p$, $1 \leq j \leq p$;

$$\begin{aligned} (Q_n^3)_{i+(h-1)p,s} &= \\ &= -H_s^{ih} - \frac{1}{2} \sum_{r=1}^p N_{rs}^{ih} (\widehat{\Psi}_{rr} - \Psi_{rr}) - \sum_{u=1}^p \sum_{v=u}^p O_{(2p-u)(u-1)/2+v,s}^{ih} (\widetilde{\sigma}_{uv} - \sigma_{uv}) \end{aligned}$$

para $1 \leq i \leq p$, $1 \leq h \leq k$, $1 \leq s \leq p$.

Luego:

$$\mathbf{T}_n = Q_n \mathbf{Y}_n.$$

Vamos a probar que:

i) $\mathbf{T}_n \xrightarrow{D} \mathcal{N}(\mathbf{0}, RNR')$ donde N es la matriz de covarianza asintótica de $\widetilde{\Sigma}$ y $R \in \mathfrak{R}^{m \times p(p+1)/2}$ es la matriz dada por:

$$R = \begin{bmatrix} D \\ L \end{bmatrix},$$

siendo

$$D = \begin{bmatrix} \left(C^1 - \frac{\partial g_1}{\partial [\text{vec}_r(\widetilde{\Sigma})]} \right)' \\ \vdots \\ \left(C^p - \frac{\partial g_p}{\partial [\text{vec}_r(\widetilde{\Sigma})]} \right)' \end{bmatrix}; \quad L = \begin{bmatrix} \frac{\partial \lambda'_{11}}{\partial [\text{vec}_r(\widetilde{\Sigma})]} \\ \vdots \\ \frac{\partial \lambda'_{pk}}{\partial [\text{vec}_r(\widetilde{\Sigma})]} \end{bmatrix};$$

y $C^i \in \mathfrak{R}^{1 \times p(p+1)/2}$ tal que

$$(C^i)_{1h} = \begin{cases} 1 & \text{si } h = \frac{2 + (i-1)(p-i+2)}{2} \\ 0 & \text{sino} \end{cases}$$

ii) $(Q_n^1)_{ij} \xrightarrow{P} \widetilde{E}_j^i$.

iii) $(Q_n^3)_{i+(h-1)p,s} \xrightarrow{P} -H_s^{ih}$.

Veamos i)

Como por hipótesis $\sqrt{n} \text{vec}_r(\tilde{\Sigma} - \Sigma) \xrightarrow{D} \mathcal{N}(0, N)$, entonces, para $1 \leq u \leq v \leq p$, $\sqrt{n}(\tilde{\sigma}_{uv} - \sigma_{uv})$ está acotado en probabilidad.

Además, de la ecuación (139) del Lema 25 se deduce que $\partial^2 g_i / \partial \tilde{\sigma}_{uv} \partial \tilde{\sigma}_{lt}$ es una función continua de los elementos de $\text{diag}(\hat{\Psi})$ y de los autovalores y autovectores de S^* , luego por (150)

$$J_{(2p-u)(u-1)/2+v,(2p-l)(l-1)/2+t}^i = \frac{\partial^2 g_i}{\partial \tilde{\sigma}_{uv} \partial \tilde{\sigma}_{lt}} \Big|_{\hat{\Upsilon}=\tilde{\Upsilon}} \xrightarrow{P} \frac{\partial^2 g_i}{\partial \tilde{\sigma}_{uv} \partial \tilde{\sigma}_{lt}} \Big|_{\hat{\Upsilon}=\Upsilon}$$

Por lo tanto

$$\sqrt{n} \left[\text{vec}_r(\tilde{\Sigma} - \Sigma)' J^i \text{vec}_r(\tilde{\Sigma} - \Sigma) \right] \xrightarrow{P} 0. \quad (156)$$

Siguiendo un razonamiento análogo, por la ecuación (147) del Lema 26 y por (150)

$$P_{(2p-u)(u-1)/2+v,(2p-l)(l-1)/2+t}^{ij} = \frac{\partial^2 \lambda_{ij}}{\partial \tilde{\sigma}_{uv} \partial \tilde{\sigma}_{lt}} \Big|_{\hat{\Upsilon}=\tilde{\Upsilon}} \xrightarrow{P} \frac{\partial^2 \lambda_{ij}}{\partial \tilde{\sigma}_{uv} \partial \tilde{\sigma}_{lt}} \Big|_{\hat{\Upsilon}=\Upsilon},$$

y por lo tanto

$$\sqrt{n} \left[\text{vec}_r(\tilde{\Sigma} - \Sigma)' P^{ij} \text{vec}_r(\tilde{\Sigma} - \Sigma) \right] \xrightarrow{P} 0. \quad (157)$$

Sea

$$\frac{T_{in}^*}{\sqrt{n}} = \begin{cases} \tilde{\sigma}_{ii} - \sigma_{ii} - D^i \text{vec}_r(\tilde{\Sigma} - \Sigma) & \text{para } 1 \leq i \leq p \\ L^{(i-p) - \lfloor \frac{i-p}{p+1} \rfloor p, \lfloor \frac{i-p}{p+1} \rfloor + 1} \text{vec}_r(\tilde{\Sigma} - \Sigma) & \text{para } p+1 \leq i \leq m \end{cases}$$

entonces, por (156) y (157), la distribución asintótica de T_{in} es la misma que la de T_{in}^* .

Como $\mathbf{T}_n^* = \sqrt{n} \left[R \text{vec}_r(\tilde{\Sigma} - \Sigma) \right]$, $\mathbf{T}_n^* \xrightarrow{D} \mathcal{N}(0, RNR')$ y por lo tanto $\mathbf{T}_n \xrightarrow{D} \mathcal{N}(0, RNR')$ como queríamos probar.

Veamos ii)

Como de la ecuación (137) del Lema 25 se deduce que $\partial^2 g_i / \partial \widehat{\Psi}_{jj} \partial \widehat{\Psi}_{ss}$ es una función continua de los elementos de $\text{diag}(\widehat{\Psi})$ y de los autovalores y autovectores de S^* , luego por (150)

$$F_{js}^i = \frac{\partial^2 g_i}{\partial \widehat{\Psi}_{jj} \partial \widehat{\Psi}_{ss}} \Big|_{\widehat{\Gamma}=\widetilde{\Gamma}} \xrightarrow{p} \frac{\partial^2 g_i}{\partial \widehat{\Psi}_{jj} \partial \widehat{\Psi}_{ss}} \Big|_{\widehat{\Gamma}=\Gamma}$$

Luego

$$F_{js}^i (\widehat{\Psi}_{ss} - \Psi_{ss}) \xrightarrow{p} 0.$$

Siguiendo un razonamiento análogo, por la ecuación (138) del Lema 25 y por (150)

$$G_{(2p-u)(u-1)/2+v,t}^i = \frac{\partial^2 g_i}{\partial \widetilde{\sigma}_{uv} \partial \widehat{\Psi}_{tt}} \Big|_{\widehat{\Gamma}=\widetilde{\Gamma}} \xrightarrow{p} \frac{\partial^2 g_i}{\partial \widetilde{\sigma}_{uv} \partial \widehat{\Psi}_{tt}} \Big|_{\widehat{\Gamma}=\Gamma}$$

Luego

$$G_{(2p-u)(u-1)/2+v,t}^i (\widetilde{\sigma}_{uv} - \sigma_{uv}) \xrightarrow{p} 0.$$

De esta manera $(Q_n^1)_{ij} \xrightarrow{p} E_j^i$.

Veamos *iii*)

Siguiendo un razonamiento análogo a *ii*), como $\widehat{\Psi}_{ss} \xrightarrow{p} \Psi_{ss}$ para $1 \leq s \leq p$, $\widetilde{\sigma}_{uv} \xrightarrow{p} \sigma_{uv}$ para $1 \leq u \leq v \leq p$ y de las ecuaciones (145) y (146) del Lema 26 se deduce que $\partial^2 \lambda_{ih} / \partial \widehat{\Psi}_{ss} \partial \widehat{\Psi}_{rr}$ y $\partial^2 \lambda_{ih} / \partial \widetilde{\sigma}_{uv} \partial \widehat{\Psi}_{rr}$ son funciones continuas de los elementos de $\text{diag}(\widehat{\Psi})$ y de los autovalores y autovectores de S^* por lo cual

$$N_{sr}^{ih} = \frac{\partial^2 \lambda_{ih}}{\partial \widehat{\Psi}_{ss} \partial \widehat{\Psi}_{rr}} \Big|_{\widehat{\Gamma}=\widetilde{\Gamma}} \xrightarrow{p} \frac{\partial^2 \lambda_{ih}}{\partial \widehat{\Psi}_{ss} \partial \widehat{\Psi}_{rr}} \Big|_{\widehat{\Gamma}=\Gamma};$$

$$O_{(2p-u)(u-1)/2+v,s}^{ih} = \frac{\partial^2 \lambda_{ih}}{\partial \widetilde{\sigma}_{uv} \partial \widehat{\Psi}_{ss}} \Big|_{\widehat{\Gamma}=\widetilde{\Gamma}} \xrightarrow{p} \frac{\partial^2 \lambda_{ih}}{\partial \widetilde{\sigma}_{uv} \partial \widehat{\Psi}_{ss}} \Big|_{\widehat{\Gamma}=\Gamma},$$

entonces

$$N_{sr}^{ih} (\widehat{\Psi}_{rr} - \Psi_{rr}) \xrightarrow{p} 0;$$

$$O_{(2p-u)(u-1)/2+v,s}^{ih} (\widetilde{\sigma}_{uv} - \sigma_{uv}) \xrightarrow{p} 0.$$

De esta manera

$$(Q_n^3)_{i+(h-1)p,s} \xrightarrow{p} -H_s^{ih}.$$

Sean

$$E = \begin{bmatrix} \frac{\partial g'_1}{\partial [\text{diag}(\widehat{\Psi})]} \\ \vdots \\ \frac{\partial g'_p}{\partial [\text{diag}(\widehat{\Psi})]} \end{bmatrix}; \quad H = \begin{bmatrix} \frac{\partial \lambda'_{11}}{\partial [\text{diag}(\widehat{\Psi})]} \\ \vdots \\ \frac{\partial \lambda'_{pk}}{\partial [\text{diag}(\widehat{\Psi})]} \end{bmatrix},$$

donde los gradientes están evaluados en $\widehat{\Upsilon} = \Upsilon$, entonces

$$Q_n \xrightarrow{p} \begin{bmatrix} I + E & \mathbf{0} \\ -H & I \end{bmatrix} = Q_0.$$

Luego, por Lema 27, si Q_0 es inversible

$$\mathbf{Y} = Q_0^{-1} \mathbf{T} \xrightarrow{D} \mathcal{N}(\mathbf{0}, Q_0^{-1} R N R' Q_0^{-1}).$$

Pero,

$$Q_0^{-1} = \begin{bmatrix} (I + E)^{-1} & \mathbf{0} \\ H(I + E)^{-1} & I \end{bmatrix},$$

luego,

$$\begin{aligned} C &= Q_0^{-1} R N R' Q_0^{-1} \\ &= \begin{bmatrix} (I + E)^{-1} D N D' (I + E)^{-1} & (I + E)^{-1} D N [H(I + E)^{-1} + L]' \\ [H(I + E)^{-1} + L] N D' (I + E)^{-1} & [H(I + E)^{-1} + L] N [H(I + E)^{-1} + L]' \end{bmatrix}. \end{aligned}$$

Es decir,

$$\mathbf{Y}_n \xrightarrow{D} \mathcal{N}(\mathbf{0}, C),$$

de lo cual se deduce junto con la ecuación (155) que

$$\begin{aligned} \sqrt{n} (\widehat{\Psi} - \Psi) &\xrightarrow{D} \mathcal{N}(\mathbf{0}, (I + E)^{-1} D N D' (I + E)^{-1}); \\ \sqrt{n} (\widehat{\Lambda} - \Lambda) &\xrightarrow{D} \mathcal{N}(\mathbf{0}, [H(I + E)^{-1} + L] N [H(I + E)^{-1} + L]'). \end{aligned}$$

D Métodos numéricos para calcular los estimadores

D.1 El procedimiento iterativo para encontrar los estimadores de los parámetros del modelo

Basado en los resultados obtenidos en la Proposición 3, se pueden obtener estimadores robustos de los parámetros del modelo de análisis factorial, utilizando el procedimiento de dos pasos propuesto por Jöreskog [17] (1967). Para esto, se reemplaza en dicho procedimiento la matriz de covarianza muestral S_n por un estimador robusto y consistente $\tilde{\Sigma}_n$ de la matriz de covarianza de las observaciones.

Este procedimiento consiste en lo siguiente.

Primero, se minimiza $F_k(\Lambda, \Psi, \tilde{\Sigma}_n)$ (dada por (21)) sobre $\Lambda \in \mathcal{R}$ para $\Psi \in \mathcal{D}_+(p)$ dado, es decir se encuentra

$$f_k(\Psi) = \min_{\Lambda \in \mathcal{A}(\Psi)} F_k(\Lambda, \Psi, \tilde{\Sigma}_n)$$

donde $\mathcal{A}(\Psi) = \{\Lambda : \Lambda' \Psi^{-1} \Lambda \in \mathcal{D}^*(k)\}$.

Segundo, se minimiza $f_k(\Psi)$ sobre $\Psi \in \mathcal{D}_+(p)$.

De esta manera,

$$\min_{\Psi \in \mathcal{D}_+(p)} f_k(\Psi) = \min_{(\Lambda, \Psi) \in \mathcal{B}(\Lambda, \Psi)} F_k(\Lambda, \Psi, \tilde{\Sigma}_n),$$

donde $\mathcal{B}(\Lambda, \Psi) = \{(\Lambda, \Psi) \in \mathcal{R} \times \mathcal{D}_+(p) : \Lambda' \Psi^{-1} \Lambda \in \mathcal{D}^*(k)\}$.

Como hemos visto en la Proposición 3, si $\tilde{\Lambda}_n = \arg \min_{\Lambda \in \mathcal{A}(\Psi)} F_k(\Lambda, \Psi, \tilde{\Sigma}_n)$, entonces,

$$\tilde{\Lambda}_n = \Psi^{1/2} \tilde{\Omega}_n (\tilde{\Theta}_n - I)^{1/2}, \quad (158)$$

donde

$$\tilde{\Theta}_n = \tilde{\Theta}_n(\Psi) = \begin{bmatrix} \tilde{\theta}_1^n & 0 \\ \vdots & \vdots \\ 0 & \tilde{\theta}_k^n \end{bmatrix}; \quad \tilde{\Omega}_n = \tilde{\Omega}_n(\Psi) = [\tilde{\mathbf{w}}_1^n \cdots \tilde{\mathbf{w}}_k^n],$$

siendo $\Sigma_n^*(\Psi) = \Psi^{-1/2} \tilde{\Sigma}_n \Psi^{-1/2}$, $\tilde{\theta}_1^n > \cdots > \tilde{\theta}_k^n$ los k autovalores más grandes de $\Sigma_n^*(\Psi)$ y $\tilde{\mathbf{w}}_1^n, \dots, \tilde{\mathbf{w}}_k^n$ sus correspondientes autovectores normalizados.

Es decir,

$$f_k(\Psi) = F_k(\tilde{\Lambda}_n, \Psi, \tilde{\Sigma}_n).$$

Para encontrar el mínimo de $f_k(\Psi)$ se utilizará el algoritmo desarrollado por Clarke [6] (1970), basado en el método de Newton-Raphson que consiste en lo siguiente.

Sean $g(\Psi)$ y $J(\Psi)$ las matrices correspondientes a las derivadas primeras y segundas de $f_k(\Psi)$ respecto de Ψ . Haciendo el desarrollo de Taylor de g

$$g(\Psi + \delta\Psi) = g(\Psi) + J(\Psi)\delta\Psi + \text{términos de orden superior.}$$

Si $\Psi^{(s)}$ es la s -ésima aproximación al punto mínimo entonces

$$\Psi^{(s+1)} = \Psi^{(s)} + \delta\Psi^{(s)}, \quad (159)$$

donde $\delta\Psi^{(s)}$ es la solución de la ecuación

$$J(\Psi^{(s)})\delta\Psi^{(s)} = -g(\Psi^{(s)}).$$

Luego, partiendo de un punto inicial adecuado $\Psi^{(0)}$, el algoritmo para encontrar $\hat{\Lambda}_n$ y $\hat{\Psi}_n$ consiste en resolver iterativamente las ecuaciones (158) y (159) hasta obtener la convergencia.

Se sugieren tomar como Ψ iniciales $\Psi^{(0)} = \text{diag}(\tilde{\Sigma}_n)$ o en el caso de que las variables estén estandarizadas $\Psi^{(0)} = I$. Jöreskog [16] (1963) demostró que una mejor elección para $\Psi^{(0)}$ que requiere menos iteraciones para lograr la convergencia está dada por

$$\psi_{ii}^{(0)} = \left(1 - \frac{k}{2p}\right) \frac{1}{\tilde{\sigma}^{ii}} \quad i = 1, \dots, p, \quad (160)$$

donde $\tilde{\sigma}^{ij}$ denota al elemento ij de la matriz $\tilde{\Sigma}_n^{-1}$. Utilizaremos este valor inicial en nuestras simulaciones.

En la Sección siguiente daremos los algoritmos que permiten encontrar estimadores robustos de la matriz de covarianza.

D.2 Algoritmos para el cálculo de los estimadores robustos de la matriz de covarianza muestral

D.2.1 Estimador MVE

Un procedimiento para la computación de un valor aproximado del MVE (ver Rousseeuw y Leroy [32](1987) o Woodruff y Rocke [43] (1994)) consiste

en lo siguiente:

PASO 1

Para todos los subconjuntos posibles de submuestras de tamaño $(p + 1)$ de \mathcal{X}_n , indexadas por $J = \{i_1, \dots, i_{p+1}\}$

- (a) Calcular el vector de medias y la matriz de covarianza muestrales para dicha submuestra

$$\bar{\mathbf{x}}_J = \frac{1}{p+1} \sum_{i \in J} \mathbf{x}_i;$$

$$C_J = \frac{1}{p} \sum_{i \in J} (\mathbf{x}_i - \bar{\mathbf{x}}_J) (\mathbf{x}_i - \bar{\mathbf{x}}_J)',$$

donde C_J es no singular siempre que $x_{i_1}, \dots, x_{i_{p-1}}$ estén en posición general.

- (b) Encontrar el factor de “inflación” de modo tal que la mitad de los puntos de la muestra \mathcal{X}_n estén incluidos en el elipsoide cuya “forma” está dada por C_J y su centro es $\bar{\mathbf{x}}_J$

$$m_J^2 = \text{med}_{1 \leq i \leq n} (\mathbf{x}_i - \bar{\mathbf{x}}_J)' C_J^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_J).$$

- (c) Calcular el determinante de la matriz “inflacionada”, que es el valor de la función objetivo

$$|m_J^2 C_J|^{1/2} = |C_J|^{1/2} (m_J)^p. \quad (161)$$

PASO 2

- (a) Determina la submuestra con la que se obtuvo el menor valor de la función (161).
- (b) Para dicha submuestra se calculan los estimadores de la media y la covarianza:

$$\mathbf{t}(\mathcal{X}_n) = \bar{\mathbf{x}}_J;$$

$$C(\mathcal{X}_n) = \frac{m_J^2 C_J}{\chi_{p,0.5}^2},$$

donde $\chi_{p,0.5}^2$ es la mediana de la distribución χ^2 , y este factor se considera para obtener consistencia bajo normalidad.

El total de submuestras de tamaño $(p + 1)$ es $\binom{n}{p+1}$. Como este valor es grande, aún para p pequeño, para calcular el MVE se considerará un número N fijo de submuestras elegidas al azar.

D.2.2 Estimador de Donoho-Stahel

Para calcular los $u_i = r(\mathbf{x}_i, \mathcal{X}_n)$ (ver (32)), se debería en principio buscar en todas las direcciones \mathbf{a} posibles, lo cual no es numéricamente factible para valores de $p > 2$.

Por lo tanto emplearemos el método sugerido por Stahel [39] (1981) para obtener una aproximación de los u_i , donde en vez de tomar en (32) el supremo sobre todas las direcciones \mathbf{a} tales que $\|\mathbf{a}\| = 1$, hallaremos el supremo sobre un conjunto \mathcal{A} finito definido así: para cada submuestra $\widetilde{\mathcal{X}}_p$ de tamaño p de \mathcal{X}_n , sea \mathbf{a} la dirección ortogonal al hiperplano que contiene a $\widetilde{\mathcal{X}}_p$; sea \mathcal{A} el conjunto de todas las direcciones \mathbf{a} así obtenidas.

Como la cantidad de elementos de \mathcal{A} puede ser muy grande aún para n y p pequeños, se reemplaza al conjunto \mathcal{A} por el que se obtiene considerando sólo N submuestras $\widetilde{\mathcal{X}}_p$, y se calculan los u_i tomando en (32) el supremo en \mathcal{A} sobre dicho conjunto.

De las simulaciones realizadas por Maronna y Yohai [26] (1995), se deduce que para $4 \leq p \leq 6$ es suficiente tomar $N = 1000$. Comparando los resultados que obtienen para algunas funciones de peso, la función w que produce menores sesgos entre las estudiadas es la función de Huber dada por

$$w_H(u) = I_{(u \leq c)} + \left(\frac{c}{u}\right)^2 I_{(u > c)},$$

donde c es el punto donde los pesos comienzan a decrecer. Ellos obtienen las constantes c de modo que se minimice el sesgo para distintos valores de n y p y distintos porcentajes de contaminación. Sin embargo, elegiremos c de modo de lograr un compromiso entre minimizar el sesgo y la eficiencia esférica. Mediante simulaciones comparamos los resultados que se obtienen al tomar distintos valores de c .

D.2.3 S-estimador

Una forma de calcular en forma aproximada los S-estimador sería mediante el algoritmo de submuestreo utilizado para calcular en forma aproximada los estimadores MVE. En este caso, el elipsoide que se construye con cada subconjunto de la muestra elegido aleatoriamente será "expandido" de modo que se satisfaga la restricción (35) en vez de (161).

Este estimador inicial se puede mejorar considerando la relación que existe entre los S- y los M-estimadores. Lopuhaä [20] (1989) y [21] (1991) mostró, aplicando multiplicadores de Lagrange a (34) que el S-estimador (\mathbf{t}_n, C_n) es tanto una solución de una ecuación que define un M-estimador

como una media muestral pesada y una matriz de covarianza pesada. Esto se deduce de lo siguiente.

El problema de minimización \mathcal{P}_n definido en la Sección 4.3 es equivalente a minimizar $\ln|C|$ sujeto a (35). La función lagrangiana para este problema es:

$$L_n(\mathbf{t}, C, \lambda) = \ln|C| - \lambda \left\{ \frac{1}{n} \sum_{i=1}^n \rho[d(\mathbf{x}_i, \mathbf{t}, C)] - b \right\},$$

donde λ es el multiplicador de Lagrange.

Entonces cualquier solución (\mathbf{t}_n, C_n) del problema \mathcal{P}_n será un cero de las derivadas parciales $\partial L_n / \partial \mathbf{t}$ y $\partial L_n / \partial C$, es decir si $\psi = \rho'$ y $d_i = d(\mathbf{x}_i, \mathbf{t}, C)$

$$\sum_{i=1}^n \frac{\psi(d_i)}{d_i} (\mathbf{x}_i - \mathbf{t}) = \mathbf{0}; \quad (162)$$

$$\sum_{i=1}^n \left[p \frac{\psi(d_i)}{d_i} (\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})' - \psi(d_i) d_i C \right] = \mathbf{0}. \quad (163)$$

(La deducción de estas ecuaciones que permiten ver la relación existente entre los S- y los M-estimadores puede verse en Lopuhaä [20] (1989)).

Las ecuaciones (162) y (163) son linealmente dependientes, lo cual se puede ver multiplicando la segunda ecuación por C^{-1} y tomando traza. Esta dependencia puede eliminarse introduciendo en la segunda ecuación el término (35), ya que (\mathbf{t}_n, C_n) serán también solución de las ecuaciones

$$\begin{aligned} \sum_{i=1}^n \frac{\psi(d_i)}{d_i} (\mathbf{x}_i - \mathbf{t}) &= \mathbf{0}; \\ \sum_{i=1}^n \left\{ p \frac{\psi(d_i)}{d_i} (\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})' - \{ \psi(d_i) d_i - [\rho(d_i) - b] \} C \right\} &= \mathbf{0}. \end{aligned}$$

Luego si $u(x) = \psi(x)/x$, (definiendo $u(0)$ por continuidad), a partir de las ecuaciones anteriores, se obtiene que:

$$\mathbf{t} = \sum_{i=1}^n u(d_i) \mathbf{x}_i / \sum_{i=1}^n u(d_i), \quad (164)$$

$$C = \sum_{i=1}^n u(d_i) (\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})' / \left[\sum_{i=1}^n u(d_i) d_i^2 / p \right]. \quad (165)$$

Además de estas ecuaciones, los S-estimadores deben satisfacer la restricción (35). A partir de (164) y (165) se puede construir un algoritmo de punto fijo que permita obtener la solución de las ecuaciones (162) y (163).

D.2.4 τ -estimador

De la misma forma que para los S-estimadores, se puede obtener un estimador inicial mediante el algoritmo de submuestreo utilizado para calcular en forma aproximada los estimadores MVE, donde en este caso, el elipsoide que se construye con cada subconjunto de la muestra elegido aleatoriamente será "expandido" de modo que se satisfaga la restricción (39) en vez de (161). Este estimador también se puede mejorar considerando la relación entre los τ - y los M-estimadores.

En este caso, el lagrangiano correspondiente al problema de minimización \mathcal{P}_n^τ definido en la Sección 4.4 está dado por:

$$L_n(\mathbf{t}, C, \lambda) = \ln |C| + p \ln \left[\frac{1}{n} \sum_{i=1}^n \rho_2(d_i) \right] - \lambda \left[\frac{1}{n} \sum_{i=1}^n \rho_1(d_i) - b_1 \right].$$

Entonces todas las soluciones (\mathbf{t}_n, C_n) del problema \mathcal{P}_n^τ deben ser un cero de las derivadas parciales de L_n , y por lo tanto, además de la restricción (39) se deben satisfacer

$$\frac{1}{n} \sum_{i=1}^n \left\{ \left[\frac{1}{n} \sum_{j=1}^n \rho_2(d_j) \right]^{-1} \frac{p\psi_2(d_i)}{d_i} - \frac{\lambda\psi_1(d_i)}{d_i} \right\} (\mathbf{x}_i - \mathbf{t}) = \mathbf{0}; \quad (166)$$

$$\frac{1}{2n} \sum_{i=1}^n \left\{ \left[\frac{1}{n} \sum_{j=1}^n \rho_2(d_j) \right]^{-1} \frac{p\psi_2(d_i)}{d_i} - \frac{\lambda\psi_1(d_i)}{d_i} \right\} (\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})' = C. \quad (167)$$

Podemos ver que estas ecuaciones son similares a las que obteníamos en el caso del S-estimador si consideramos ahora a la función ψ como un promedio pesado adaptativo de las funciones ψ_1 y ψ_2 , es decir

$$\psi_n(\cdot, \mathbf{t}, C) = A_n(\mathbf{t}, C)\psi_1(\cdot) + B_n(\mathbf{t}, C)\psi_2(\cdot),$$

donde los pesos, que dependen de $\mathbf{x}_1, \dots, \mathbf{x}_n$ están dados por:

$$A_n(\mathbf{t}, C) = \frac{1}{n} \sum_{i=1}^n [2\rho_2(d_i) - \psi_2(d_i) d_i];$$

$$B_n(\mathbf{t}, C) = \frac{1}{n} \sum_{i=1}^n \psi_1(d_i) d_i,$$

y por lo tanto, las ecuaciones anteriores se pueden escribir como:

$$\sum_{i=1}^n \frac{\psi_n(d_i)}{d_i} (\mathbf{x}_i - \mathbf{t}) = \mathbf{0};$$

$$\sum_{i=1}^n \left\{ p \frac{\psi_n(d_i)}{d_i} (\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})' - \psi_n(d_i) d_i C \right\} = \mathbf{0}.$$

Estas ecuaciones son linealmente dependientes. Sin embargo, agregando un múltiplo conveniente de la restricción (39) a la segunda ecuación se puede evitar la dependencia lineal. De esta manera, las soluciones (\mathbf{t}_n, C_n) del problema \mathcal{P}_n^τ deben satisfacer:

$$\sum_{i=1}^n \frac{\psi_n(d_i)}{d_i} (\mathbf{x}_i - \mathbf{t}) = \mathbf{0};$$

$$\sum_{i=1}^n \left\{ p \frac{\psi_n(d_i)}{d_i} (\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})' - \{ \psi_n(d_i) d_i - 2b_2 [\rho(d_i) - b_1] \} C \right\} = \mathbf{0},$$

lo que permite, como en el caso del S-estimador, construir un algoritmo de punto fijo que permita encontrar la solución de las ecuaciones (166) y (167).

Referencias

- [1] AMEMIYA, Y., FULLER, W. A. & PANTULA, S. G. (1987) The asymptotic distributions of some estimators for a factor analysis model *Journal of multivariate analysis* **22**, 51-64
- [2] ANDERSON, T. W. (1958) *An introduction to multivariate statistical analysis* John Wiley & Sons: New York
- [3] ANDERSON, T. W. (1984) Estimating linear statistical relationships *The Annals of Statistics* **12**, 1-45
- [4] ANDERSON, T. W. & AMEMIYA, Y. (1988) The asymptotic normal distribution of estimators in factor analysis under general conditions *The Annals of Statistics* **16**, 759-771
- [5] ANDERSON, T. W. & RUBIN, H. (1956) Statistical inference in factor analysis *Proceedings of the Third Berkeley Symposium of Mathematical Statistics* (J. Neyman Ed.) Univ. of California Press, **5**, 111-156
- [6] CLARKE, M. R. B. (1970) A rapidly convergent method for maximum likelihood factor analysis *Brit. J. Math. Statistic. Psych.* **23**, 43-52
- [7] DAVIES, L. (1992) The asymptotics of Rousseeuw's minimum volume ellipsoid estimator *The Annals of Statistics* **20**, 1828-1843
- [8] DAVIES, P. L. (1987) Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices *The Annals of Statistics* **15**, 1269-1292
- [9] DONOHO, D. L. (1982) *Breakdown properties of multivariate location estimators*, qualifying paper, Harvard University, Boston, M.A.
- [10] DWYER, P. S. (1967) Some applications of matrix derivatives in multivariate analysis. *Journal of the American Statistical Association* **62**, 607-625
- [11] EVERITT, B. S. (1984) *An introduction to latent variable models* Chapman & Hall: New York
- [12] GIRI, N. C. (1977) *Multivariate statistical inference* Academic Press: New York

- [13] HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. & STAHEL, W. A. (1986) *Robust Statistics: the approach based on influence functions* John Wiley & Sons: New York
- [14] HENNIG, C. (1995) Efficient high-breakdown point estimators in robust regression: which function to choose? *Statistics and Decisions* **13**, 221-241
- [15] JENNRICH, R. I. & THAYER, D. (1973) A note on Lawley's formulas for standard errors in maximum likelihood factor analysis *Psychometrika* **38**, 571-581
- [16] JÖRESKOG, K. G. (1963) *Statistical Estimation in Factor Analysis* Almqvist & Wiksell: Stockholm
- [17] JÖRESKOG, K. G. (1967) Some contributions to maximum likelihood factor analysis *Psychometrika* **32**, 443-482
- [18] LAWLEY, D. N. & MAXWELL, A. E. (1971) *Factor analysis as a Statistical Method* 2nd. ed. Butterworths: London
- [19] LEHMANN, E. L. (1983) *Theory of point estimation* John Wiley & Sons: New York
- [20] LOPUHAA, H. P. (1989) On the relation between S-estimators and M-estimators of multivariate location and covariance *The Annals of Statistics* **17**, 1662-1683
- [21] LOPUHAA, H. P. (1991) Breakdown point and asymptotic properties of multivariate S-estimators and τ -estimators: a summary *Directions in Robust Statistic and Diagnostics Part I*, 167-182
- [22] LOPUHAA, H. P. (1991) Multivariate τ -estimators for location and scatter *The Canadian Journal of Statistics* **19**, 307-321
- [23] LOPUHAA, H. P. & ROUSSEEUW, P. J. (1991) Breakdown points of affine equivariant estimators of multivariate location and covariance matrices *The Annals of Statistics* **19**, 229-248
- [24] MAGNUS, J. R. (1985) On differentiating eigenvalues and eigenvectors *Econometric Theory* **1**, 179-191
- [25] MAGNUS, J. R. & NEUDECKER, H. (1985) Matrix differential calculus with applications to simple, Hadamard and Kronecker products *Journal of Mathematical Psychology* **29**, 474-492

- [26] MARONNA, R. A. & YOHAI, V. J. (1995) The behavior of the Stahel-Donoho robust multivariate estimator *Journal of the American Statistical Association* **90**, 330-341
- [27] MUIRHEAD, R. J. (1982) *Aspects of multivariate statistical theory* John Wiley & Sons: New York
- [28] NEUDECKER, H. (1967) On matrix procedures for optimizing differentiable scalar functions of matrices *Statistica Neerlandica* **21**, 101-107
- [29] RAO, C. R. (1967) Calculus of generalized inverses of matrices - Part I: general theory *Sankhyā, Ser A* **29**, 317-342
- [30] RAO, C. R. (1973) *Linear statistical inference and its applications* John Wiley & Sons: New York
- [31] ROCKE, D. M. & WOODRUFF, D. L. (1993) Computation of robust estimates of multivariate location and shape *Statistica Neerlandica* **47**, 27-42
- [32] ROUSSEEUW, P. J. & LEROY, A. M. (1987) *Robust regression and outliers detection* John Wiley & Sons: New York
- [33] ROUSSEEUW, P. J. & VAN ZOMEREN, B. C. (1990) Unmasking multivariate outliers and leverage points *Journal of the American Statistical Association* **85**, 633-639
- [34] ROUSSEEUW, P. J. & YOHAI, V. J. (1984) Robust regression by means of S-estimators - *Robust and nonlinear time series analysis (Lecture notes in Statistics)* Springer-Verlag, 256-272
- [35] RUPPERT, D. (1992) Computing S estimators for regression and multivariate location/dispersion *Journal of computational and graphical statistical* **1**, 253-270
- [36] SEBER, G. A. F. (1984) *Multivariate observations* John Wiley & sons. U.S.A.
- [37] SHAPIRO, A. (1982) Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis *Psychometrika* **47**, 187-199
- [38] SHAPIRO, A. (1985) Identifiability of factor analysis: some results and open problems *Linear algebra and its applications* **70**, 1-7

- [39] STAHEL, W.A. (1981). *Breakdown of covariance estimators*. Research report 31, Fachgruppe für Statistik, E.T.H. Ztrich.
- [40] STYAN, G. P. H. (1973) Hadamard products and multivariate statistical analysis *Linear algebra and its applications* **6**, 217-240
- [41] SWAIN, A. J. (1975) A class of factor analysis estimation procedures with common asymptotic sampling properties *Psychometrika* **40**, 315-335
- [42] TYLER, D. E. (1994) Finite-sample breakdown points of projection-based multivariate location and scatter statistics *The Annals of Statistics* **22**, 1024-1044
- [43] WOODRUFF, D. L. & ROCKE, D. M. (1994) Computable robust estimation of multivariate location and shape in high dimension using compound estimators *Journal of the American Statistical Association* **89**, 888-896
- [44] YOHAI, V. J. & ZAMAR, R. H. (1988) High-breakdown point estimators of regression by means of the minimization of an efficient scale *Journal of the American Statistical Association* **83**, 406-413
- [45] YOHAI, V. J. & ZAMAR, R. H. (1991) Discussion of "Least median squares estimation in power systems" *IEEE Transaction on Power Systems* **6**, 520