

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA

INFERENCIA Y REDUCCIÓN DE SESGO EN
MODELOS ESTRUCTURALES ELÍPTICOS

POR

JUAN FRANCISCO OLIVARES PACHECO

TESIS PRESENTADA A LA FACULTAD DE MATEMÁTICA DE LA
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE PARA OPTAR
AL GRADO ACADÉMICO DE DOCTOR EN ESTADÍSTICA
SANTIAGO, CHILE
DICIEMBRE 2017

Orientador: Reinaldo Boris Arellano Valle
Co-orientador: Manuel Jesús Galea Rojas
©2017, Juan Francisco Olivares Pacheco

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA

Autor: Juan F. Olivares

Título: Inferencia y reducción de sesgo en modelos estructurales elípticos

Fecha: Diciembre 2017

Orientador: Reinaldo Arellano Valle

Co-orientador: Manuel Galea Rojas

Comisión examinadora: _____

Reinaldo Arellano Valle
Pontificia Universidad Católica de Chile

Manuel Galea Rojas
Pontificia Universidad Católica de Chile

Luis Castro Cepero
Pontificia Universidad Católica de Chile

Orietta Nicolis
Universidad de Valparaíso

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA

ABSTRACT

Inference and bias-reduction for structural elliptical models

Juan F. Olivares

In this thesis we study elliptical structural models following the approach that errors are independent and have elliptical marginal distribution. We will adopt an approach studied by Gleser (1992), for which we obtain an alternative and convenient representation to the usual form of log-likelihood function. This representation involves a new parameterization which produces a more parsimonious model, transforming the classic measurement error linear regression model into a regression model with a usual randomly designed but heteroskedastic matrix. The identification problems present in normal structural models are also present in elliptical structural models. Therefore, additional assumptions are necessary to make the estimation problem feasible. These assumptions can be considered standard in the literature of models with measurement errors. On the other hand, we determine the score vector and the expected information matrix for the reparameterized model. Simple expressions to calculate the elements of the information matrix are given, in which only a few univariate moments must be calculated numerically. Two illustrations of the elliptical distributions are considered. Firstly, we consider the normal case, for which in the reparameterized model the maximum likelihood estimators correspond to the estimates of least squares and the information matrix is diagonal by blocks. The second illustration corresponds to the structural model t , which shows that the inference can be made iteratively. Finally, given that in the context of models with measurement errors, the maximum likelihood estimators are biased, we implemented two algorithms to correct the bias of the estimates. In particular, we consider the bias reduction method proposed by Firth (1993), which has the advantage that it does not depend directly on the existence of maximum likelihood estimators.

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA

RESUMEN

Inferencia y reducción de sesgo en modelos estructurales elípticos

Juan F. Olivares

En esta tesis estudiamos los modelos estructurales elípticos siguiendo el enfoque de que los errores son independientes y tienen distribución marginal elíptica. Adoptaremos un enfoque estudiado por Gleser (1992), para el cual obtenemos una representación alternativa y conveniente a la forma usual de la función de log-verosimilitud. Esta representación involucra una nueva parametrización la cual produce un modelo más parsimonioso, transformando el modelo de regresión lineal con error de medida clásico en un modelo de regresión con matriz de diseño aleatoria usual, pero heteroscedástico. Los problemas de identificación presentes en los modelos estructurales normales, se tienen también en los modelos estructurales elípticos. Por tanto, supuestos adicionales son necesarios para hacer el problema de estimación factible. Estos supuestos pueden ser considerados como estándar en la literatura de modelos con error de medida. Por otro lado, determinamos el vector de score y la matriz de información esperada para el modelo reparametrizado. Expresiones simples para calcular los elementos de la matriz de información son dados, en el cual sólo algunos momentos univariados deben ser calculados numéricamente. Dos ilustraciones de las distribuciones elípticas son considerados. En primer lugar, consideramos el caso normal, para el cual en el modelo reparametrizado los estimadores de máxima verosimilitud corresponden a los estimados de mínimos cuadrados y la matriz de información es diagonal por bloques. La segunda ilustración corresponde al modelo estructural t , en el cual se muestra que la inferencia puede ser realizada iterativamente. Finalmente, dado que en el contexto de modelos con error de medida, los estimadores de máxima verosimilitud son sesgados, implementamos dos algoritmos para corregir el sesgo de las estimaciones. En particular, consideramos el método de reducción de sesgo propuesto por Firth (1993), el cual tiene la ventaja de que este no depende directamente de la existencia de los estimadores de máxima verosimilitud.

Índice de contenidos

1	Introducción	3
1.1	Motivación	3
1.2	Definición de objetivos	7
1.3	Resumen	8
2	Distribuciones elípticas	11
2.1	Introducción	11
2.2	Definiciones	12
2.3	Representación estocástica	16
2.4	Momentos	18
2.5	Transformaciones lineales	19
2.6	Distribuciones marginales	19
2.7	Distribución condicional	20
2.8	Discusión	21
3	Modelos estructurales elípticos	23
3.1	Introducción	23
3.2	Parametrización directa	24
3.3	Parametrización tipo Gleser	26
3.4	Identificación del modelo estructural elíptico	29
3.5	Estimación de máxima verosimilitud	30
3.6	Representación jerárquica	35
3.7	Matriz de información	38
3.7.1	Matriz de información observada	39
3.7.2	Matriz de información esperada	42
3.7.3	Parametrización directa	44
3.8	El modelo estructural normal	45
3.9	El modelo estructural t	48

3.10	Modelo estructural de mezcla de escala de normales	56
3.11	Discusión	58
4	Reducción de sesgo en modelos estructurales elípticos	61
4.1	Introducción	61
4.2	Método de reducción de sesgo	65
4.3	Algoritmos de corrección y reducción de sesgo	66
4.4	Reducción de sesgo en modelos estructurales	68
4.4.1	Modelo estructural normal simple	71
4.4.2	Modelo estructural t	78
4.4.3	Aplicación	90
4.5	Discusión	98
5	Conclusiones y trabajos futuros	99
5.1	Comentarios finales	99
5.2	Trabajos futuros	100
5.2.1	Modelos con medida de error generalizados	100
5.2.2	Distribuciones skew-elípticas	101
5.2.3	Distribuciones slash-elípticas	104
	Referencias	106

Capítulo 1

Introducción

1.1. Motivación

Es bien conocido que los modelos de regresión paramétricos, especialmente los modelos de regresión lineal son herramientas útiles para el análisis de conjuntos de datos y proveen una descripción parsimoniosa de la relación entre la variable de respuesta y sus covariables. Sin embargo, en muchas situaciones prácticas, pueden existir covariables medidas con error. Ejemplos bien documentados en la literatura son los trabajos de Carroll et al. (1995), Wang et al. (1996) y Tosteson et al. (1989).

Los *modelos con error de medida*, donde es asumido que las covariables están sujetas a un error de medición, se han transformado en conceptos útiles en muchas disciplinas, por ejemplo, tenemos la comparación de instrumentos (Barnett (1969), Bolfarine & Galea-Rojas (1996) y Chipkevitch et al. (1996)), en el área médica han sido reportados resultados por Chipkevitch et al. (1996) y Kelly (1984), ejemplos en agricultura son considerados en Cheng & Van Ness (1999), y casos en psicología y educación han sido estudiados por Dunn (1992).

El modelo de regresión lineal simple con error de medida clásico es definido por las ecuaciones

$$\begin{cases} Y_i = \beta_0 + \beta_1 \xi_i + \epsilon_i, \\ X_i = \xi_i + \delta_i, \end{cases} \quad (1.1)$$

para $i = 1, 2, \dots, n$, donde

$$\begin{aligned} E[\epsilon_i] = E[\delta_i] = 0, \quad V(\epsilon_i) = \sigma_\epsilon^2, \quad V(\delta_i) = \sigma_\delta^2, \quad \forall i, \\ \text{Cov}(\delta_i, \delta_j) = \text{Cov}(\epsilon_i, \epsilon_j) = 0, \quad \forall i \neq j, \quad \text{y} \\ \text{Cov}(\epsilon_i, \delta_j) = 0, \quad \forall i, j. \end{aligned}$$

El supuesto de considerar que la covariable es medida con error puede causar que la estimación de los parámetros asociados al modelo difieran dramáticamente de los resultados que se pueden obtener a través de los análisis de los modelos de regresión estándar. Los primeros resultados en esta área han sido resumidos por Fuller (1987). Actualizaciones de estos resultados en modelos lineales y no-lineales con error de medida han sido recopilados por Cheng & Van Ness (1999) y Carroll et al. (1995), respectivamente.

La principal idea detrás de las ecuaciones en (1.1) es que la variable verdadera, ξ_i , no es directamente observable y sólo es posible observar una estimación de ésta, dada por X_i , para $i = 1, 2, \dots, n$. En (1.1), la ecuación $X_i = \xi_i + \delta_i$ es conocida como error aditivo. Usualmente, es típicamente asumido que ϵ_i y δ_i sean variables aleatorias independientes normalmente distribuidas con media cero y varianzas constantes

Además, los modelos con error de medida se pueden clasificar en dos clases: si para una muestra (X_i, Y_i) , con $i = 1, 2, \dots, n$, la variable verdadera ξ_i es no-aleatoria (parámetros incidentales), el modelo con error de medida es denominado *modelo funcional*. Por otro lado, si ξ_i son variables aleatorias independientes e idénticamente distribuidas se denomina *modelo estructural*.

Existe un tercer modelo, el *modelo ultra-estructural* (Dolby (1976)), el cual asume que ξ_i son variables aleatorias independientes como en el caso estructural pero no idénticamente distribuidas, a su vez se permite la posibilidad de tener diferentes medias pero con varianza común. Claramente, el modelo ultra-estructural es una generalización de los modelos funcional y estructural; se tienen muchas otras modificaciones y extensiones del modelo con error de medida anteriormente descrito, estas versiones se pueden encontrar en los libros de Fuller (1987) y Cheng & Van Ness (1999).

Al considerar el modelo con error de medida como en (1.1), se tiene que la estimación de β_1 presenta dificultades, tanto en el caso funcional como en el caso estructural. En el caso funcional, la función de verosimilitud puede ser no-acotada, mientras que en el caso estructural, el modelo puede ser no-identificable.

Para hacer el problema de estimación factible se deben considerar supuestos adicionales. Algunos de los supuestos considerados en la literatura son: (i) $\lambda_\epsilon = \sigma_\epsilon^2/\sigma_\delta^2$ conocido; (ii) σ_ϵ^2 o σ_δ^2 (o ambos) conocidos o (iii) $\lambda = \sigma_\xi^2/(\sigma_\xi^2 + \sigma_\epsilon^2)$ conocido (ver Cheng & Van Ness (1999) para revisar otros supuestos comúnmente utilizados).

Otro problema que se puede presentar al considerar el error de medición aditivo como en (1.1) es que la estimación de β_1 puede ser sesgada negativamente, con sesgo aproximadamente igual a la proporción de la varianza de los errores. Para notar esto, se tiene que los modelos con error de medida lucen como un modelo de regresión

estándar. Pero los modelos de regresión, sin embargo, son un caso especial de los modelos con error de medida cuando el error, δ_i , es idénticamente cero. Si tratamos de escribir (1.1) como un modelo de regresión estándar obtenemos

$$Y_i = \beta_0 + \beta_1 X_i + (\epsilon_i - \beta_1 \delta_i) = \beta_0 + \beta_1 X_i + \zeta_i.$$

Este no es un modelo de regresión usual, ya que por un lado, X_i es aleatorio (esto ocurre en cualquier modelo con medida de error) y además está correlacionado con el término de error ζ_i , es decir, $\text{Cov}(X, \zeta) = -\beta_1 \sigma_\delta^2$. Notemos que, esta covarianza sólo es cero cuando $\sigma_\delta^2 = 0$, lo cual corresponde al modelo de regresión estándar, o cuando $\beta_1 = 0$, lo cual es el caso trivial.

Si utilizamos la estimación de mínimos cuadrados en modelos con error de medida, obtenemos estimadores inconsistentes. Por ejemplo, consideremos el modelo estructural, la estimación de β_1 es $\hat{\beta}_1 = s_{yx}/S_{xx}$, donde s_{yx} y S_{xx} son las covarianzas muestrales, con esperanza

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 \frac{\sigma_\xi^2}{\sigma_\xi^2 + \sigma_\delta^2} = \beta_1 \lambda < \beta_1.$$

El error de medición atenúa la estimación de β_1 , además si σ_δ^2 aumenta, el sesgo aumenta. Existen varias formas solucionar éste problema, es decir, se pueden construir estimadores consistentes cuando las variables son sujetas a error de medición, por ejemplo, si asumimos que λ es conocido, se puede proponer que λ corrija el estimador, es decir, obtener un estimador corregido de β_1 dado por $\hat{\beta}_1/\lambda$.

Debido a que los estimadores de mínimos cuadrados no son consistentes, entonces no pueden ser considerados como estimadores razonables para los parámetros de los modelos con medida de error (Fuller (1980)). Entonces, es claro que el sesgo en la estimación es una preocupación común entre los profesionales e investigadores en el tema. Su magnitud juega un importante rol en la estimación y éste puede dar lugar a inferencias que pueden inducir a error.

En los modelos de regresión lineal con error de medida, la estimación por mínimos cuadrados, así como otras, también presenta problemas. Si la ecuación incluye dos covariables, y una de ellas es medida con error y la otra está libre de éste, se puede derivar analíticamente, con los supuestos habituales sobre los errores, el sesgo de los coeficientes asociados a cada covariable. Éste sesgo depende de la proporción de la varianza de los errores y también de la correlación entre las covariables (Griliches & Ringstad (1971)). Cuando, sin embargo, el error está presente en ambas covariables, el sesgo que se produce en el estimador de mínimos cuadrados no puede ser evaluado

fácilmente (Maddala (1988)).

En general, para modelos con dos o más covariables, se tiene que los errores de medición producen sesgo en las estimaciones de los coeficientes aún cuando alguna covariable este libre de error (Johnston (1997)).

En la literatura estadística existen muchos trabajos sobre las formas que se podría reducir el sesgo en las estimaciones en modelos paramétricos, y dependiendo del proceso de reducción, se pueden diferenciar dos clases, las cuales se pueden clasificar como: estimadores con sesgo corregido y estimadores con sesgo reducido.

En el caso de corrección de sesgo, este método está basado en los siguientes dos pasos: (i) obtener el sesgo de primer orden de los estimadores, y (ii) restar este sesgo a dichos estimadores. Este método ha sido ampliamente utilizado, pero un problema que se puede distinguir en los métodos de corrección de sesgo, es que por definición, los estimadores obtenidos en el proceso de corrección serán indefinidos si los estimadores de máxima verosimilitud divergen. Ejemplos característicos de la aplicación de estos métodos son Cox & Snell (1968), Cook et al. (1986), Cordeiro & McCullagh (1991), Breslow & Lin (1995), Lin & Breslow (1996), Cordeiro & Vasconcellos (1997) y Cordeiro & Toyama Udo (2008).

La principal diferencia entre los métodos de corrección y reducción de sesgo es que estos últimos no dependen directamente de los estimadores de máxima verosimilitud. Quenouille (1956) ha sido uno de los primeros en desarrollar un método de reducción de sesgo que es aplicable a una familia general de distribuciones.

Por otro lado, Firth (1993) desarrolla un método general para remover el término de primer orden en la expansión asintótica del sesgo de los estimadores de máxima verosimilitud. En este método, el vector de score es apropiadamente modificado de tal forma que las raíces resultantes de este nuevo vector de score resulten en un estimador modificado.

A diferencia de los métodos de corrección, nos podemos dar cuenta que la naturaleza de este método es preventiva en vez de correctiva. Ejemplos de tales estudios son Firth (1995), Pettitt et al. (1998), Heinze & Schemper (2001), Bull et al. (2002), Sartori (2006), Kosmidis (2007), Kosmidis & Firth (2009) y Kosmidis & Firth (2010). Considerando estos métodos de reducción, estudiaremos el comportamiento del sesgo en las estimaciones de los parámetros involucrados en el modelo con error de medida. En otro orden de ideas, como se mencionó anteriormente, típicamente es asumido que los errores son distribuidos normalmente, pero la normalidad es un supuesto fuerte que no siempre puede ser sustentable en la práctica. Entonces, una generalización natural del modelo normal multivariado es dado por la clase de las distribuciones

elípticas.

Estas distribuciones han sido introducidas por Kelker (1970) y bien investigadas por Cambanis et al. (1981), Fang et al. (1990) y Arellano-Valle (1994). En adición a la distribución normal muchas otras distribuciones multivariadas bien conocidas son elípticas, por ejemplo, la distribución t (Fang et al. (1990) y Arellano-Valle & Bolfarine (1995)), y la distribución simétrica hiperbólica generalizada (Barndorff-Nielsen et al. (1982)), por nombrar algunas.

Una de las características de las distribuciones elípticas es que comparten algunas de las propiedades de la distribución normal. Es decir, cualquier transformación lineal de un vector aleatorio elíptico es también elíptico. Y, cualquier distribución marginal de un vector aleatorio elíptico es elíptico. Esto ocurre también para las distribuciones condicionales (Kelker (1970)).

En el contexto de los modelos lineales, varios autores han usado las distribuciones elípticas para explorar las consecuencias de tener errores no-normales o distribuciones con colas más (o menos) pesadas que la distribución normal (Zellner (1976*a*), Zellner (1976*b*), Ghosh & Sinha (1980), Ullah & Zinde-Walsh (1984), Chib et al. (1988), Anderson et al. (1986), Lange et al. (1989), Taylor (1992), Kano et al. (1993) y Arellano-Valle (1994)).

En la literatura existen dos formas de especificar un modelo de regresión elíptico, tales modelos pueden ser referidos como: (a) *modelo elíptico dependiente*, el cual asume que los errores (o las observaciones) tienen distribución conjunta elíptica multivariada, es decir, pueden ser no-correlacionados sin implicar independencia, y (b) *modelo elíptico independiente*, estos modelos asumen que los errores (o las observaciones) son independientes y tienen distribución marginal elíptica. En este caso, la distribución conjunta es el producto de las distribuciones marginales elípticas, pero debemos notar que ésta puede no ser elíptica. Si se asume que la distribución conjunta es elíptica junto con independencia es equivalente a asumir normalidad. Por lo tanto, en una primera etapa, el interés radicará en estudiar una versión multivariada del modelo (1.1), bajo el enfoque de que las distribuciones marginales de los errores y las covariables son elípticas.

1.2. Definición de objetivos

Aunque el caso funcional de los modelos con error de medida pueden ser más naturales en un contexto de regresión, donde las covariables usualmente son consideradas no-aleatorias (parámetros incidentales), existen importantes ventajas conceptuales y

técnicas en el uso del correspondiente caso estructural para derivar los estimadores (Gleser (1992)).

En lo sucesivo estudiamos las propiedades del modelo estructural multivariado, es decir, la variable verdadera subyacente ξ_i es considerada como un vector aleatorio, con localización $\eta_\xi \in \mathbb{R}^p$ y matriz de dispersión $\Sigma_{\xi\xi}$ (matriz definida positiva de dimensión $p \times p$). Por otro lado, para evitar problemas de identificabilidad, el supuesto considerado en este trabajo es que $\Sigma_{\delta\delta}$ es conocido, desde el punto de vista aplicado, este parámetro puede ser aproximado a través de replicaciones (Buonaccorsi (2010)). Entonces, tomando en cuenta el supuesto de que los errores (conjuntamente con ξ) tienen distribución elíptica multivariada, son investigados e implementados algoritmo correctivos del sesgo para las estimaciones de máxima verosimilitud. Por tanto, los objetivos específicos de este estudio pueden ser resumidos como sigue:

1. Estudiar el modelo estructural elíptico independiente con error de medida.
2. Investigar e implementar algoritmo basados en la modificación del vector de score para la reducción del sesgo en la estimación de los coeficientes de regresión del modelo estructural elíptico.
3. Implementar esquemas de simulación para estudiar el comportamiento de las estimaciones bajo el concepto de reducción de sesgo.

1.3. Resumen

En cuanto al cumplimiento de los objetivos descritos anteriormente, este trabajo está organizado en cuatro capítulos principales. Por lo tanto, estos capítulos pueden resumirse como sigue.

Capítulo 2: En este capítulo resumimos las propiedades básicas de las distribuciones elípticas y además establecemos la notación usada en este trabajo. Definimos las distribuciones esféricas y elípticas, bajo el supuesto de la existencia de la densidad presentamos la forma de la función de densidad de probabilidad para los modelos elípticos y, la representación estocástica, momentos y mostramos que cualquier transformación de un vector aleatorio elíptico es también elíptico, además cualquier distribución marginal de un vector aleatorio elíptico es elíptico. Esto también ocurre para las distribuciones condicionales. Las referencias básicas usadas en éste capítulo son Kelker (1970), Cambanis et al. (1981),

Anderson et al. (1986), Muirhead (1982), Fang et al. (1990) y Arellano-Valle (1994).

Capítulo 3: En este capítulo son dada algunas ideas generales sobre los modelos estructurales elípticos siguiendo el enfoque de que los errores son independientes y tienen distribución marginal elíptica. Presentamos un breve desarrollo de los modelos estructurales con error de medida bajo la parametrización directa. Para simplificar nuestro análisis, adoptamos un enfoque alternativo para obtener una expresión diferente para la forma usual de la función de log-verosimilitud, ver Gleser (1992) y Cheng & Tsai (2004). Estudiamos la identificabilidad del modelo estructural elíptico y la estimación de máxima verosimilitud. Mostramos que el modelo estructural elíptico en términos de la parametrización de Gleser permite una simple representación jerárquica. También se dan expresiones simples para obtener los elementos de la matriz de información esperada, en el cual sólo algunos momentos univariados deben ser calculados numéricamente.

Capítulo 4: En este capítulo se estudia el método de reducción de sesgo. Mostramos la conexión entre los métodos de reducción y corrección de sesgo. Se presenta el algoritmo de reducción como una procedimiento de corrección iterado. Aplicamos estos métodos a un conjunto de datos e implementamos un esquema de simulación para estudiar el comportamiento de la estimación de los coeficientes de regresión.

Capítulo 5: En este capítulo se resume los resultados obtenidos en los capítulos previos y algunas conclusiones son dadas. Se concluye este trabajo con algunas recomendaciones para trabajos futuros.

Capítulo 2

Distribuciones elípticas

2.1. Introducción

Una generalización natural de la distribución normal multivariada es dada por la clase de distribuciones elípticas, donde el primer desarrollo teórico de esta clase es dado por Kelker (1970) e investigado detalladamente por Cambanis et al. (1981). Desde entonces, estas distribuciones han sido estudiadas tanto en términos de sus propiedades teóricas como sus aplicaciones. Un desarrollo más actualizado puede ser encontrado en Fang et al. (1990) y Arellano-Valle (1994).

Una de las características de las distribuciones elípticas es que pueden ser generadas por transformaciones lineales de localización y escala de las distribuciones esféricas. A su vez, las distribuciones esféricas son caracterizadas por su invarianza con respecto a transformaciones ortogonales y forman una clase general de distribuciones con la misma simetría esférica que la distribución normal estándar. Por lo tanto, cada distribución esférica genera una familia paramétrica de distribuciones elípticas a través de transformaciones lineales. Recíprocamente, cada distribución esférica es un elemento particular de la familia de distribuciones elípticas. Consecuentemente, las distribuciones elípticas preservan la estructura simétrica (simetría elíptica) de las distribuciones normales.

En adición a la distribución normal, otros típicos elementos en la clase de distribuciones elípticas son, por ejemplo, la familia de normales compuestas y la familia de distribuciones t . Sin embargo, existe un amplio espectro de distribuciones paramétricas alternativas al modelo normal que pueden ser obtenidas desde la clase de distribuciones elípticas (Tabla 2.1).

Una de las principales características de las distribuciones elípticas, es que heredan una gran cantidad de buenas propiedades de la distribución normal. Así, cualquier

transformación (lineal) de un vector aleatorio elíptico también es elíptico. Además, cualquier función de distribución marginal de un vector aleatorio elíptico también tiene distribución elíptica. Esto ocurre también para las funciones condicionales.

En este capítulo resumimos las propiedades básicas de las distribuciones elípticas y se establece la notación usada en este trabajo. En la sección 2.2 definimos las distribuciones esférica y elípticas, además bajo el supuesto de la existencia de la densidad presentamos la forma de la función de densidad de probabilidad para las distribuciones elípticas. La sección 2.3 muestra la representación estocástica. Los momentos son estudiados en la sección 2.4. La sección 2.5 muestra que cualquier transformación lineal de un vector aleatorio elíptico es también elíptico, además cualquier distribución marginal de un vector aleatorio elíptico tiene distribución elíptica, sección 2.6. Esto ocurre también para las distribuciones condicionales, sección 2.7. Una discusión del capítulo es presentada en la sección 2.8.

2.2. Definiciones

Definición 2.1 (Distribución esférica). Un vector aleatorio \mathbf{X} de dimensión $n \times 1$, tiene distribución esférica si para cada matriz $\mathbf{\Gamma} \in \mathcal{O}_n$, se tiene que

$$\mathbf{X} \stackrel{d}{=} \mathbf{\Gamma} \mathbf{X}, \quad (2.1)$$

donde \mathcal{O}_n denota el conjunto de matrices ortogonales de dimensión $n \times n$. Aquí (2.1) denota que el vector aleatorio \mathbf{X} tiene la misma distribución que $\mathbf{\Gamma} \mathbf{X}$.

Las distribuciones esféricas y sus correspondientes vectores aleatorios son algunas veces también llamadas distribuciones *radiales* (Kelker (1970)) o *isotrópicas* (Bingham & Kiesel (2002)). El siguiente teorema muestra una representación de las distribuciones esféricas a través de su función característica.

Teorema 2.1. *Un vector aleatorio \mathbf{X} de dimensión $n \times 1$, tiene distribución esférica si y sólo si, su función característica $\phi_{\mathbf{X}}(\mathbf{t})$ satisface alguna de las siguientes dos condiciones equivalentes:*

(i) $\phi_{\mathbf{X}}(\mathbf{\Gamma}^\top \mathbf{t}) = \phi_{\mathbf{X}}(\mathbf{t})$, para cada matriz $\mathbf{\Gamma} \in \mathcal{O}_n$.

(ii) Existe una función $\phi(\cdot)$ de variable escalar tal que $\phi_{\mathbf{X}}(\mathbf{t}) = \phi(\mathbf{t}^\top \mathbf{t})$.

Demostración. Para una matriz cuadrada \mathbf{A} la función característica de $\mathbf{A}\mathbf{X}$ es igual a $\phi_{\mathbf{X}}(\mathbf{A}^\top \mathbf{t})$. Por lo tanto (i) es equivalente a la Definición 2.1. La condición (ii)

implica (i), ya que $\phi_{\mathbf{X}}(\mathbf{\Gamma}^\top \mathbf{t}) = \phi((\mathbf{\Gamma}^\top \mathbf{t})^\top (\mathbf{\Gamma}^\top \mathbf{t})) = \phi(\mathbf{t}^\top \mathbf{\Gamma} \mathbf{\Gamma}^\top \mathbf{t}) = \phi(\mathbf{t}^\top \mathbf{t}) = \phi_{\mathbf{X}}(\mathbf{t})$. Recíprocamente, (i) implica que $\phi_{\mathbf{X}}(\mathbf{t})$ es invariante con respecto a transformaciones ortogonales, pero desde las propiedades de invarianza de \mathcal{O}_n (Fang et al. (1990)), se tiene que $\phi_{\mathbf{X}}(\mathbf{t})$ debe ser una función de $\mathbf{t}^\top \mathbf{t}$. \square

Una de las principales características de las distribuciones elípticas y por tanto las distribuciones esféricas, es que heredan muchas de las buenas propiedades de la distribución normal. Esto es debido a que la función característica de la distribución normal estándar multivariada, es decir, $\mathbf{t} \mapsto \exp(-\mathbf{t}^\top \mathbf{t}/2)$, es simplemente relajada a $\mathbf{t} \mapsto \phi(\mathbf{t}^\top \mathbf{t})$, por la condición (ii) del teorema anterior, donde $\phi(\cdot)$, usualmente llamada *generadora de funciones características*, es una función arbitraria que sólo garantiza que $\phi(\mathbf{t}^\top \mathbf{t})$ sea una función característica.

Por otro lado, el vector aleatorio \mathbf{X} no necesariamente posee densidad. Sin embargo, si existe debe ser de la forma $g(\mathbf{x}^\top \mathbf{x})$, donde g es una función no-negativa que satisface

$$\int g(\mathbf{x}^\top \mathbf{x}) d\mathbf{x} = s_n \int_0^\infty y^{n/2-1} g(y) dy = 1,$$

con $s_n = \pi^{n/2}/\Gamma(n/2)$ y puede estar indexada por un vector de parámetros $\boldsymbol{\nu}$. Por lo tanto, g puede ser usada para definir una densidad $cg(\mathbf{x}^\top \mathbf{x})$ para alguna distribución esférica si y sólo si

$$\int_0^\infty y^{n/2-1} g(y) dy < \infty. \quad (2.2)$$

La función g es llamada *función generadora de densidades* de \mathbf{X} (Fang et al. (1990)). Algunos ejemplos de las distribuciones esféricas son dados en Muirhead (1982) y Fang et al. (1990).

Ejemplo 2.1 (Distribución normal estándar multivariada). Sea $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I}_n)$, la distribución normal estándar n -variada. Entonces, \mathbf{X} tiene función característica $\phi_{\mathbf{X}}(\mathbf{t}) = \phi(\mathbf{t}^\top \mathbf{t}) = e^{-\mathbf{t}^\top \mathbf{t}/2}$, $\mathbf{t} \in \mathbb{R}^n$ y función de densidad $f_{\mathbf{X}}(\mathbf{x}) = cg(\mathbf{x}^\top \mathbf{x}) = (2\pi)^{-n/2} e^{-\mathbf{x}^\top \mathbf{x}/2}$, $\mathbf{x} \in \mathbb{R}^n$.

Ejemplo 2.2 (Distribución t estándar multivariada). Sea $\mathbf{X} \sim t_n(\mathbf{0}, \mathbf{I}_n, \nu)$, la distribución t estándar n -variada con ν grados de libertad ($\nu > 0$). Entonces, \mathbf{X} tiene función de densidad

$$f_{\mathbf{X}}(\mathbf{x}; \nu) = cg(\mathbf{x}^\top \mathbf{x}) = \frac{\Gamma((\nu + n)/2)}{(\nu\pi)^{n/2} \Gamma(\nu/2)} \left(1 + \frac{1}{\nu} \mathbf{x}^\top \mathbf{x}\right)^{-(\nu+n)/2}, \quad \mathbf{x} \in \mathbb{R}^n,$$

la función característica es derivada en Sutradhar (1986) y Sutradhar (1988).

Proposición 2.1. Sea \mathbf{Y} un vector aleatorio m -variado con distribución esférica y función característica $\phi_{\mathbf{Y}}$. Además, sea $\mathbf{A} \in \mathbb{R}^{n \times m}$ una matriz arbitraria y $\boldsymbol{\mu} \in \mathbb{R}^n$. Entonces, la función característica de $\mathbf{X} := \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$ viene dada por $\phi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \phi(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t})$, $\mathbf{t} \in \mathbb{R}^n$ donde $\boldsymbol{\Sigma} := \mathbf{A}\mathbf{A}^\top$.

Demostración. La función característica de \mathbf{X} corresponde a

$$\phi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \phi_{\mathbf{Y}}(\mathbf{A}^\top \mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \phi((\mathbf{A}^\top \mathbf{t})^\top (\mathbf{A}^\top \mathbf{t})) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \phi(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}),$$

dados los resultados presentados en el Teorema 2.1. □

Podemos decir, que la Proposición 2.1 es la base para la definición clásica de las distribuciones elípticas dada en Cambanis et al. (1981). Debemos notar que la clase de distribuciones elípticas ha sido definida en un número equivalentes de formas (Kelker (1970), Eaton (1981), Muirhead (1982), Fang & Zhang (1990) y Arellano-Valle (1994)).

Definición 2.2 (Distribución elíptica). Un vector aleatorio \mathbf{X} de dimensión $n \times 1$, tiene distribución elíptica con $\boldsymbol{\mu} \in \mathbb{R}^n$ y $\boldsymbol{\Sigma} \geq 0$ de dimensión $n \times n$, si la función característica de \mathbf{X} es de la forma

$$\phi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \phi(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}), \mathbf{t} \in \mathbb{R}^n. \quad (2.3)$$

Si un vector aleatorio \mathbf{X} tiene función característica como en (2.3), diremos que \mathbf{X} tiene distribución elíptica con parámetros $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ (y ϕ), lo cual denotaremos como $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ o $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$ cuando sea necesario indicar la dependencia de ϕ en la representación paramétrica.

De la Definición 2.2, la matriz $\boldsymbol{\Sigma}$ corresponde a la *matriz de dispersión* de \mathbf{X} y $\boldsymbol{\mu}$ determina su localización. Así, cada distribución elíptica pertenece a una familia de localización-escala (Kelker (1970)) definida por una distribución esférica subyacente, es decir, $\boldsymbol{\mu} = \mathbf{0}$ y $\boldsymbol{\Sigma} = \mathbf{I}_n$, donde \mathbf{I}_n es la matriz identidad de dimensión $n \times n$. Para $n = 1$, la clase de distribuciones elípticas coincide con la clase de distribuciones simétricas univariadas (Cambanis et al. (1981)).

Al igual como en el caso esférico, un vector aleatorio $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$, no posee necesariamente una densidad. Una condición necesaria para que \mathbf{X} posea una densidad es que $r(\boldsymbol{\Sigma}) = n$, donde $r(\boldsymbol{\Sigma})$ es el rango de la matriz $\boldsymbol{\Sigma}$ tal que $\boldsymbol{\Sigma}$ sea una matriz de dimensión $n \times n$ definida positiva ($\boldsymbol{\Sigma} > 0$). En este caso, de la Proposición 2.1, la representación estocástica es $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$. Como la distribución de \mathbf{Y} es esférica,

su densidad es de la forma $g(\mathbf{y}^\top \mathbf{y})$, donde g es la función generadora de densidades. Entonces, se tiene que la densidad de \mathbf{X} es proporcional a

$$|\Sigma|^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})), \mathbf{x} \in \mathbb{R}^n. \quad (2.4)$$

En este caso algunas veces usaremos la notación $\text{El}_n(\boldsymbol{\mu}, \Sigma; g)$ en vez de $\text{El}_n(\boldsymbol{\mu}, \Sigma; \phi)$. Cualquier función g que satisface (2.2) define una densidad (2.4) de una distribución elíptica con constante de normalización $C_n(\nu)$ donde

$$C_n(\nu) = \frac{\Gamma(n/2)}{\pi^{n/2} \int_0^\infty r^{n/2-1} g(r) dr}. \quad (2.5)$$

Por lo tanto, si $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \Sigma; g)$, el vector aleatorio \mathbf{X} tiene una función de densidad de la forma

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = C_n(\nu) |\Sigma|^{-1/2} g((\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})), \mathbf{x} \in \mathbb{R}^n. \quad (2.6)$$

Ejemplo 2.3 (Distribución normal multivariada). Sea $\mathbf{X} \sim \text{N}_n(\boldsymbol{\mu}, \Sigma)$, la distribución normal n -variada con función característica $\phi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu} - \mathbf{t}^\top \Sigma \mathbf{t}/2}$, $\mathbf{t} \in \mathbb{R}^n$. Si $\Sigma > 0$, entonces \mathbf{X} tiene función de densidad

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-n/2} |\Sigma|^{-1/2} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}, \mathbf{x} \in \mathbb{R}^n.$$

Ejemplo 2.4 (Distribución t multivariada). Sea $\mathbf{X} \sim t_n(\boldsymbol{\mu}, \Sigma, \nu)$, la distribución t n -variada con ν grados de libertad y parámetro de localización $\boldsymbol{\mu}$ y matriz de dispersión Σ . Si $\Sigma > 0$, entonces \mathbf{X} tiene función de densidad

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \Sigma, \nu) = \frac{\Gamma((\nu + n)/2)}{(\nu\pi)^{n/2} \Gamma(\nu/2)} |\Sigma|^{-1/2} \left(1 + \frac{1}{\nu} (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)^{-\frac{\nu+n}{2}}, \mathbf{x} \in \mathbb{R}^n.$$

La simetría de la densidad elíptica implica que los contornos de las superficies de iguales densidades tienen la misma forma en una clase dada. Debido a esto surge el nombre de *distribuciones de contornos elípticos* el cual es una extensión de las distribuciones esféricas introduciendo parámetros adicionales. La Tabla 2.1, la cual es esencialmente debida a Jensen (1985), muestra algunos ejemplos del generador de densidades para diferentes modelos.

2.3. Representación estocástica

Sea $\mathbf{U}^{(n)}$ un vector aleatorio n -variado uniformemente distribuido en

$$\mathcal{S}^n = \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\| = 1\},$$

donde \mathcal{S}^n es la superficie de la esfera unitaria en \mathbb{R}^n , y

$$\|\mathbf{u}\| = (\mathbf{u}^\top \mathbf{u})^{1/2}, \quad \mathbf{u} \in \mathbb{R}^n.$$

Cabe señalar que la distribución uniforme es la única distribución en \mathcal{S}^n la cual es invariante respecto a transformaciones ortogonales. Entonces, cada vector aleatorio n -dimensional $\mathbf{X} \sim \text{El}_n(\mathbf{0}, \mathbf{I}_n; \phi)$, puede ser representado como

$$\mathbf{X} \stackrel{d}{=} R\mathbf{U}^{(n)},$$

donde R es una variable aleatoria no-negativa independiente de $\mathbf{U}^{(n)}$. La variable aleatoria R es denominada como la *variable radial* o *variable generadora* de \mathbf{X} .

El siguiente teorema de la representación estocástica para las distribuciones elípticas es debido a Cambanis et al. (1981).

Teorema 2.2 (Representación estocástica). *Sea $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$ con $r(\boldsymbol{\Sigma}) = k$, si y sólo si*

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + R\mathbf{A}\mathbf{U}^{(k)}, \tag{2.7}$$

donde $\mathbf{U}^{(k)}$ es un vector aleatorio k -dimensional uniformemente distribuido en \mathcal{S}^k , R es una variable aleatoria no-negativa independiente de $\mathbf{U}^{(k)}$, $\boldsymbol{\mu} \in \mathbb{R}^n$ y \mathbf{A} es una matriz de dimensión $n \times k$ con $r(\mathbf{A}) = k$ tal que $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^\top$.

Notemos que la matriz \mathbf{A} corresponde a una matriz de transformación, el vector aleatorio esférico \mathbf{U} , produce densidades elípticas, mientras que la variable radial R determina la forma de la distribución, en particular el peso de las colas de la distribución. Además, si $\mathbf{X} \stackrel{d}{=} R\mathbf{U}^{(n)} \sim \text{El}_n(\mathbf{0}, \mathbf{I}_n; \phi)$ con $\mathbb{P}(\mathbf{X} = \mathbf{0}) = 0$, entonces

$$\mathbf{U}^{(n)} \stackrel{d}{=} \frac{\mathbf{X}}{\|\mathbf{X}\|} \quad \text{y} \quad R \stackrel{d}{=} \|\mathbf{X}\|,$$

son independientes.

La representación estocástica de un vector aleatorio distribuido elípticamente es usualmente más conveniente para propósitos prácticos que su representación a través de

la función característica. Especialmente, de la representación estocástica, es sencillo presentar esquemas de simulación para la generación de vectores aleatorios.

Los siguientes ejemplos, muestran que a partir de la representación estocástica, la distribución normal n -variada es caracterizada por la distribución chi-cuadrada con n grados de libertad. Similarmente, la distribución t n -variada con ν grados de libertad es caracterizada por la distribución F con n y ν grados de libertad.

Ejemplo 2.5 (Distribución normal multivariada). Sea $\boldsymbol{\mu} \in \mathbb{R}^n$ y $\mathbf{A} \in \mathbb{R}^{n \times k}$ tal que $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^\top \in \mathbb{R}^{n \times n}$ es definida positiva. Sea $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces \mathbf{X} posee distribución elíptica y por tanto puede ser representado como

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{\chi_k^2} \mathbf{A}\mathbf{U}^{(k)}, \quad (2.8)$$

donde $\mathbf{U}^{(k)}$ es un vector aleatorio k -dimensional uniformemente distribuido. La distribución esférica subyacente es el normal estándar multivariada.

Ejemplo 2.6 (Distribución t multivariada). Consideremos el vector aleatorio

$$\mathbf{Y} \stackrel{d}{=} \frac{\mathbf{X}}{\sqrt{\chi_\nu^2/\nu}}, \quad \nu > 0, \quad (2.9)$$

donde $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I}_n)$, con χ_ν^2 y \mathbf{X} siendo independientes. Entonces, \mathbf{Y} se dice que es distribuido t multivariado con ν grados de libertad (Fang et al. (1990)). \mathbf{X} puede ser representado por $\sqrt{\chi_n^2} \mathbf{U}^{(n)}$, donde $\mathbf{U}^{(n)}$, χ_n^2 y χ_ν^2 son mutuamente independientes. Así \mathbf{Y} puede ser representado por

$$\mathbf{Y} \stackrel{d}{=} \frac{\sqrt{\chi_n^2}}{\sqrt{\chi_\nu^2/\nu}} \mathbf{U}^{(n)} = \sqrt{n \frac{\chi_n^2/n}{\chi_\nu^2/\nu}} \mathbf{U}^{(n)} \stackrel{d}{=} \sqrt{n F_{n,\nu}} \mathbf{U}^{(n)},$$

donde $F_{n,\nu}$ es una variable aleatoria con distribución F con n y ν grados de libertad y es independiente de $\mathbf{U}^{(n)}$. Además

$$\frac{\sqrt{\chi_n^2}}{\sqrt{\chi_\nu^2/\nu}} \xrightarrow{d} \sqrt{\chi_n^2}, \quad \text{cuando } \nu \rightarrow \infty,$$

como una consecuencia de que $\chi_\nu^2/\nu \xrightarrow{a.s.} 1$ debido a la ley fuerte de los grandes números. Por lo tanto, $\mathbf{Y} \xrightarrow{d} N_n(\mathbf{0}, \mathbf{I}_n)$, cuando $\nu \rightarrow \infty$. Notemos que, el vector aleatorio $\boldsymbol{\mu} + \sqrt{n F_{n,\nu}} \mathbf{A}\mathbf{U}^{(n)}$ tiene una distribución t multivariada con vector de localización $\boldsymbol{\mu}$ y matriz de dispersión $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^\top$ siempre que \mathbf{A} tenga rango completo.

Como es notado en Kelker (1970) y en Cambanis et al. (1981), si $\mathbf{X} \stackrel{d}{=} R\mathbf{U}^{(n)} \sim$

$\text{El}_n(\mathbf{0}, \mathbf{I}_n; \phi)$, entonces $\mathbf{X} \sim \text{El}_n(\mathbf{0}, \mathbf{I}_n; g)$, es decir, el vector aleatorio \mathbf{X} tiene densidad g , si y sólo si, R tiene densidad f_R y las dos densidades están relacionadas como sigue

$$f_R(r) = 2s_n r^{n-1} g(r^2), \quad r \geq 0. \quad (2.10)$$

Además, si $\text{P}(\mathbf{X} = \mathbf{0}) = 0$, entonces todas las distribuciones marginales de \mathbf{X} tienen densidad. La demostración de estos resultados pueden ser encontrados en Kelker (1970) o Fang et al. (1990). Con estos resultados y usando el hecho de que si $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$ y $r(\boldsymbol{\Sigma}) = n$, entonces

$$\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}) \sim \text{El}_n(\mathbf{0}, \mathbf{I}_n; g).$$

Cambanis et al. (1981), nota que la densidad de \mathbf{X} existe y toma la forma (2.4), si y sólo si la variable aleatoria R en la representación estocástica (2.7), con $k = n$ y $\mathbf{A} = \boldsymbol{\Sigma}^{1/2}$ tiene densidad dada por (2.10). Algunos detalles de estos y otros resultados pueden ser encontrados en Fang et al. (1990).

2.4. Momentos

Sea $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$, de (2.7), \mathbf{X} tiene primer momento finito si y sólo si $\text{E}[R] < \infty$ y

$$\text{E}[\mathbf{X}] = \text{E}[\boldsymbol{\mu} + R\mathbf{A}\mathbf{U}^{(k)}] = \boldsymbol{\mu} + \mathbf{A}\text{E}[R]\text{E}[\mathbf{U}^{(k)}],$$

dado que R y $\mathbf{U}^{(k)}$ son independientes y como $\text{E}[\mathbf{U}^{(k)}] = \mathbf{0}$ obtenemos que $\text{E}[\mathbf{X}] = \boldsymbol{\mu}$. Similarmente, el segundo momento de los componentes de \mathbf{X} es finito si y sólo si $\text{E}[R^2] < \infty$ y

$$\text{Cov}(\mathbf{X}) = \text{E}[(R\mathbf{A}\mathbf{U}^{(k)})(R\mathbf{A}\mathbf{U}^{(k)})^\top] = \text{E}[R^2]\mathbf{A}\text{E}[\mathbf{U}^{(k)}\mathbf{U}^{(k)\top}]\mathbf{A}^\top,$$

dado que $\sqrt{\chi_k^2}\mathbf{U}^{(k)} \sim \text{N}_k(\mathbf{0}, \mathbf{I}_k)$, por tanto

$$\mathbf{I}_k = \text{E}\left[(\sqrt{\chi_k^2}\mathbf{U}^{(k)})(\sqrt{\chi_k^2}\mathbf{U}^{(k)})^\top\right] = \text{E}[\chi_k^2]\text{E}[\mathbf{U}^{(k)}\mathbf{U}^{(k)\top}] = k\text{E}[\mathbf{U}^{(k)}\mathbf{U}^{(k)\top}],$$

obtenemos $\text{E}[\mathbf{U}^{(k)}\mathbf{U}^{(k)\top}] = \text{Cov}(\mathbf{U}^{(k)}) = \mathbf{I}_k/k$, entonces

$$\text{Cov}(\mathbf{X}) = \frac{\text{E}[R^2]}{k}\boldsymbol{\Sigma}.$$

Notemos que, k , no es necesariamente el rango de Σ o la dimensión de \mathbf{X} , éste corresponde al número de componentes de $\mathbf{U}^{(k)}$. Además, nos podemos dar cuenta que la matriz de dispersión generalmente no coincide con la matriz de covarianza. La distribución normal es el caso excepcional, debido a que $E[R^2] = k$ y por tanto $\text{Cov}(\mathbf{X}) = \Sigma$.

2.5. Transformaciones lineales

Sea $\mathbf{a} \in \mathbb{R}^m$ y $\mathbf{B} \in \mathbb{R}^{m \times n}$, consideremos la siguiente transformación lineal $\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X}$, donde $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + R\mathbf{A}\mathbf{U}^{(m)} \sim \text{El}_n(\boldsymbol{\mu}, \Sigma; \phi)$ con $\mathbf{A} \in \mathbb{R}^{n \times m}$ y $\Sigma = \mathbf{A}\mathbf{A}^\top$. Por tanto

$$\mathbf{Y} \stackrel{d}{=} \mathbf{a} + \mathbf{B}(\boldsymbol{\mu} + R\mathbf{A}\mathbf{U}^{(m)}) = (\mathbf{a} + \mathbf{B}\boldsymbol{\mu}) + R(\mathbf{B}\mathbf{A})\mathbf{U}^{(m)}.$$

Por lo tanto, cada transformación lineal y particularmente cada combinación lineal de vectores aleatorios elípticos posee distribución elíptica.

Un interesante hecho es que la variable radial de un vector aleatorio transformado permanece constante. Por tanto, los vectores aleatorios transformados no sólo son elípticos sino también cerrados bajo la correspondiente familia de localización-escala. Entonces

$$\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X} \sim \text{El}_m(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\Sigma\mathbf{B}^\top; \phi). \quad (2.11)$$

2.6. Distribuciones marginales

De (2.11), si $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \Sigma; \phi)$, entonces la distribución marginal también es elíptica. Para ver esto, consideremos la partición de \mathbf{X} , $\boldsymbol{\mu}$ y Σ dada por

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \text{y} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad (2.12)$$

donde \mathbf{X}_1 y $\boldsymbol{\mu}_1$ son vectores de dimensión $m \times 1$ y Σ_{11} es una matriz $m \times m$, para $1 \leq m < n$. Entonces

$$\mathbf{X}_1 \sim \text{El}_m(\boldsymbol{\mu}_1, \Sigma_{11}; \phi) \quad \text{y} \quad \mathbf{X}_2 \sim \text{El}_{n-m}(\boldsymbol{\mu}_2, \Sigma_{22}; \phi),$$

con

$$\phi_{\mathbf{X}_1}(\mathbf{t}_1) = e^{it_1^\top \boldsymbol{\mu}_1} \phi(\mathbf{t}_1^\top \Sigma_{11} \mathbf{t}_1) \quad \text{y} \quad \phi_{\mathbf{X}_2}(\mathbf{t}_2) = e^{it_2^\top \boldsymbol{\mu}_2} \phi(\mathbf{t}_2^\top \Sigma_{22} \mathbf{t}_2),$$

donde $\mathbf{t}_1 \in \mathbb{R}^m$ y $\mathbf{t}_2 \in \mathbb{R}^{n-m}$. En efecto, a través (2.11) y definiendo $\mathbf{B} = (\mathbf{I}_m, \mathbf{0})$ y $\mathbf{a} = \mathbf{0}$, se tiene la demostración para \mathbf{X}_1 y en forma similar para \mathbf{X}_2 . Una forma más directa para derivar $\phi_{\mathbf{X}_1}(\mathbf{t}_1)$ es considerar que $\mathbf{t} = (\mathbf{t}_1^\top, \mathbf{0}^\top)^\top$ en (2.3). La distribución de \mathbf{X}_2 es obtenido por simetría.

En el caso donde $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$ y se asume la existencia de la densidad, es decir, $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$, entonces todas las densidades marginales de dimensión $m < n$ no difiere en su forma funcional.

La densidad marginal de \mathbf{X}_1 , es dada por

$$f_{\mathbf{X}_1}(\mathbf{x}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) = C_{n,n-m}(\boldsymbol{\nu}) |\boldsymbol{\Sigma}_{11}|^{-1/2} \int_0^\infty r^{n-m-1} g(u + r^2) dr, \quad (2.13)$$

con $\mathbf{x}_1 \in \mathbb{R}^m$ donde $u = (\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1)$ y

$$C_{p,q}(\boldsymbol{\nu}) = \frac{\pi^{(q-p)/2} \Gamma(p/2)}{\Gamma(q/2) \int_0^\infty r^{p-1} g(r^2) dr}. \quad (2.14)$$

La densidad marginal de \mathbf{X}_2 , es obtenida por simetría.

2.7. Distribución condicional

Sea $\mathbf{X} \sim \text{El}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$, es decir, \mathbf{X} tiene densidad como en (2.4), y considerando la partición dada en (2.12), se tiene que la distribución condicional de \mathbf{X}_1 dado $\mathbf{X}_2 = \mathbf{x}_2$, es elíptica

$$(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) \sim \text{El}_m(\boldsymbol{\mu}_1(\mathbf{x}_2), \boldsymbol{\Sigma}_{11.2}; g_{q(\mathbf{x}_2)}). \quad (2.15)$$

Además, asumiendo que existen

$$\begin{aligned} E[\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2] &= \boldsymbol{\mu}_1(\mathbf{x}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2), \\ \text{Cov}(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) &= \alpha_{q(\mathbf{x}_2)} \boldsymbol{\Sigma}_{11.2} = \alpha_{q(\mathbf{x}_2)} (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}), \end{aligned}$$

donde $g_{q(\mathbf{x}_2)}$ y $\alpha_{q(\mathbf{x}_2)}$, son funciones de \mathbf{x}_2 a través la forma cuadrática $q(\mathbf{x}_2) = (\mathbf{x}_2 - \boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2)$. Notemos que, de (2.4)

$$f_{\mathbf{X}_1, \mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{C_n(\boldsymbol{\nu}) g((\mathbf{x}_1 - \boldsymbol{\mu}_1(\mathbf{x}_2))^\top \boldsymbol{\Sigma}_{11.2}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1(\mathbf{x}_2)) + q(\mathbf{x}_2))}{|\boldsymbol{\Sigma}_{22}|^{-1/2} |\boldsymbol{\Sigma}_{11.2}|^{-1/2}},$$

entonces, la densidad condicional de \mathbf{X}_1 dado $\mathbf{X}_2 = \mathbf{x}_2$, es dada por

$$f_{\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2}(\mathbf{x}_1; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}_{11.2}|^{-1/2} g_{q(\mathbf{x}_2)}((\mathbf{x}_1 - \boldsymbol{\mu}_1(\mathbf{x}_2))^\top \boldsymbol{\Sigma}_{11.2}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1(\mathbf{x}_2))), \quad (2.16)$$

para $\mathbf{x}_1 \in \mathbb{R}^m$ y donde para cada $a \geq 0$ se tiene que

$$g_a(u) = \frac{\Gamma(m/2)g^{(n)}(u+a)}{2\pi^{m/2} \int_0^\infty r^{m-1}g(r^2+a) dr}, \quad u \geq 0,$$

es una función no-negativa. Además, si $\phi_{\mathbf{X}_1}(\mathbf{t}_1|\mathbf{x}_2)$, $\mathbf{t}_1 \in \mathbb{R}^m$, denota la función generadora de densidades condicionales de $(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2)$ por tanto

$$\phi_{\mathbf{X}_1}(\mathbf{t}_1|\mathbf{x}_2) = \mathbb{E}[e^{i\mathbf{t}_1^\top \mathbf{X}_1} | \mathbf{X}_2 = \mathbf{x}_2] = e^{i\mathbf{t}_1^\top \boldsymbol{\mu}_1(\mathbf{x}_2)} \mathbb{E}[e^{i\mathbf{t}_1^\top (\mathbf{X}_1 - \boldsymbol{\mu}_1(\mathbf{x}_2))} | \mathbf{X}_2 = \mathbf{x}_2],$$

Usando (2.16), $\mathbb{E}[e^{i\mathbf{t}_1^\top (\mathbf{X}_1 - \boldsymbol{\mu}_1(\mathbf{x}_2))} | \mathbf{X}_2 = \mathbf{x}_2] = \phi_{q(\mathbf{x}_2)}(\mathbf{t}_1^\top \boldsymbol{\Sigma}_{11.2} \mathbf{t}_1)$ y consecuentemente

$$\phi_{\mathbf{X}_1}(\mathbf{t}_1|\mathbf{x}_2) = e^{i\mathbf{t}_1^\top \boldsymbol{\mu}_1(\mathbf{x}_2)} \phi_{q(\mathbf{x}_2)}(\mathbf{t}_1^\top \boldsymbol{\Sigma}_{11.2} \mathbf{t}_1), \quad \mathbf{t}_1 \in \mathbb{R}^m. \quad (2.17)$$

Además, para cada $a \geq 0$ se tiene que $\phi_a(\mathbf{t}^\top \mathbf{t})$, $\mathbf{t} \in \mathbb{R}^m$, es la función característica de una distribución esférica m -variada.

2.8. Discusión

En este capítulo han sido definidos las distribuciones esféricas y elípticas multivariadas y se provee un desarrollo teórico de las propiedades básicas de estas familias de distribuciones. Estos conceptos discutidos son necesarios para la definición de los modelos estructurales elípticos tratados en el capítulo 3. Las referencias básicas usadas son Kelker (1970), Cambanis et al. (1981), Anderson et al. (1986), Muirhead (1982), Fang et al. (1990) y Arellano-Valle (1994).

Como hemos notado, las distribuciones elípticas son radialmente simétricas, así una interesante extensión que se puede estudiar sobre esta familia es modelar la asimetría radial sin perder muchas de las propiedades básicas de las distribuciones elípticas. Es decir, por un lado, se debería considerar la parsimonia con respecto a la parametrización de un modelo multivariado.

Y por otro lado, todos los componentes de las distribuciones elípticas, es decir, la variable generadora R , el vector de localización $\boldsymbol{\mu}$ y la matriz de dispersión $\boldsymbol{\Sigma}$ (la cual contiene la dependencia lineal de los componentes) deben mantenerse en la nueva clase de las distribuciones asimétricas.

Tabla 2.1: Algunos ejemplos del generador de densidades.

Tipo	$g(u)$ ($u = \mathbf{x}^\top \mathbf{x}$, $\mathbf{x} \in \mathbb{R}^n$) o $\phi(v)$ ($v = \mathbf{t}^\top \mathbf{t}$, $\mathbf{t} \in \mathbb{R}^n$)
Kotz	$g(u) = u^{m-1} \exp(-ru^s)$, $r, s > 0$, $2m + n > 2$
Normal	$g(u) = \exp(-u/2)$
Pearson tipo VII	$g(u) = (1 + u/s)^{-m}$, $m > n/2$, $s > 0$
t multivariada	$g(u) = (1 + u/s)^{-(s+n)/2}$, $s > 0$
Cauchy multivariada	$g(u) = (1 + u)^{-(n+1)/2}$, $s > 0$
Pearson Tipo II	$g(u) = (1 - u)^m$, $m > 0$, $0 \leq u < 1$
Logística	$g(u) = \exp(-u)[1 + \exp(-u)]^2$
Mezcla de escala	$g(u) = \int_0^\infty t^{-n/2} \exp(-u/2t) dF(t)$
Ley estable	$\phi(v) = \exp(rv^{\alpha/2})$, $0 < \alpha \leq 2$, $r < 0$
Multi-uniforme	$\phi(v) = {}_0F_1(n/2; -v/4)$

Capítulo 3

Modelos estructurales elípticos

3.1. Introducción

Los modelos de regresión lineal son comúnmente utilizados en muchas disciplinas, dado que son herramientas útiles para el análisis de conjuntos de datos y proveen una descripción parsimoniosa de la relación entre la variable de respuesta y sus covariables. El uso de estos modelos comúnmente radica en los llamados supuestos clásicos. Estos incluyen que las covariables son medidas sin error y son no-correlacionadas. También, los errores son independientes e idénticamente distribuidos, posiblemente normales, y son no-correlacionados con las covariables.

Desde luego, cualquiera de estos supuestos pueden ser relajados de muchas maneras. En este capítulo, nos enfocaremos en dos supuestos asociados a los modelos de regresión lineal mencionados anteriormente, por un lado consideraremos que las covariables están sujetas a cierto error de medición (ver Fuller (1987), Cheng & Van Ness (1999) y Carroll et al. (1995)).

Y por otro lado, asumir la normalidad en los errores es un supuesto fuerte que no siempre puede ser satisfecho en la práctica. Entonces, una generalización natural de la distribución normal multivariada es dada por la clase de las distribuciones elípticas, las cuales serán el eje central de este capítulo.

En el contexto de los modelos de regresión lineal, varios autores han usado las distribuciones elípticas para explorar las consecuencias de tener errores no-normales o distribuciones con colas más (o menos) pesadas que la distribución normal (Zellner (1976a), Zellner (1976b), Ghosh & Sinha (1980), Ullah & Zinde-Walsh (1984), Chib et al. (1988), Anderson et al. (1986), Lange et al. (1989), Taylor (1992), Kano et al. (1993) y Arellano-Valle (1994)), en este capítulo nos enfocaremos en estudiar los modelos de regresión lineal con error de medición, asumiendo que *los errores son in-*

dependientes y tienen distribuciones marginales elípticas independientes. En este caso, la distribución conjunta es el producto de las distribuciones marginales elípticas, pero debemos notar que esta puede no ser elíptica. Si se asume que la distribución conjunta es elíptica junto con independencia es equivalente a asumir normalidad.

Para simplificar el análisis, realizaremos una reparametrización del modelo de regresión lineal con error de medida, con el fin de obtener una versión más simple de la función de verosimilitud. Esta reparametrización tiene la ventaja de establecer una útil conexión entre los modelos de regresión lineal y los modelos de regresión lineal con error de medida.

Este capítulo está estructurado de la siguiente forma, la sección 3.2 presenta un breve desarrollo del modelo estructural elíptico bajo la parametrización directa. Para simplificar nuestro análisis, adoptamos un enfoque alternativo, en la sección 3.3, se obtiene una expresión diferente para la forma usual de la función de log-verosimilitud (ver Gleser (1992) y Cheng & Tsai (2004)). En la sección 3.4, se estudia la identificabilidad del modelo estructural elíptico. La sección 3.5, estudiamos la estimación de máxima verosimilitud, siguiendo el enfoque de cálculo diferencial, como en Magnus & Neudecker (1999). La sección 3.6, muestra que el modelo estructural elíptico, en término de la parametrización de Gleser, permite una simple representación jerárquica. La sección 3.7, es esencialmente de naturaleza técnica, y concierne al estudio de la matriz de información para el modelo estructural elíptico, damos expresiones simples para calcular los elementos de la matriz de información esperada, en el cual sólo algunos momentos univariados deben ser calculados numéricamente.

Estos resultados son ilustrados con dos ejemplos para las distribuciones normal y t en las secciones 3.8 y 3.9, respectivamente. En la sección 3.10, estudiamos el modelo de mezcla de escala. Finalmente, una discusión del capítulo es presentada en la sección 3.11.

3.2. Parametrización directa

Consideremos el modelo de regresión lineal con error de medida clásico definido por

$$\begin{cases} Y_i = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{\xi}_i + \epsilon_i, \\ \mathbf{X}_i = \boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \end{cases} \quad (3.1)$$

para $i = 1, 2, \dots, n$, donde $\boldsymbol{\beta}$ y $\boldsymbol{\xi}_i$ son vectores de dimensión $p \times 1$ y ϵ_i es el error aleatorio de media 0 y varianza $\sigma_\epsilon^2 > 0$.

La covariable observada \mathbf{X}_i , es una medida del valor verdadero no observado $\boldsymbol{\xi}_i$ más un error de medición $\boldsymbol{\delta}_i$, con vector de localización $\mathbf{0}$ y matriz de dispersión $\boldsymbol{\Sigma}_{\delta\delta} > 0$.

Asumiendo el modelo estructural, las covariables subyacentes $\boldsymbol{\xi}_i$ son consideradas vectores aleatorios, con vector de localización $\boldsymbol{\eta}_\xi \in \mathbb{R}^p$ y matriz de dispersión $\boldsymbol{\Sigma}_{\xi\xi} > 0$.

Definamos el vector aleatorio $\mathbf{r}_i = (\epsilon_i, \boldsymbol{\delta}_i^\top, \boldsymbol{\xi}_i^\top)^\top$, donde asumimos que $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ son independientes y $\mathbf{r}_i \sim \text{El}_{2p+1}(\boldsymbol{\eta}, \boldsymbol{\Omega}; \phi)$, para $i = 1, 2, \dots, n$, con

$$\boldsymbol{\eta} = \begin{bmatrix} 0 \\ \mathbf{0} \\ \boldsymbol{\eta}_\xi \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Omega} = \begin{bmatrix} \sigma_\epsilon^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\delta\delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{\xi\xi} \end{bmatrix}. \quad (3.2)$$

Aquí, $\mathbf{0}$ representa al vector cero o a la matriz cero siempre y cuando la dimensión sea clara. Definiendo el vector observado como $\mathbf{Z}_i = (\mathbf{X}_i^\top, Y_i)^\top$, para $i = 1, 2, \dots, n$, y

$$\mathbf{a} = \begin{bmatrix} \mathbf{0} \\ \beta_0 \end{bmatrix} \quad \text{y} \quad \mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_p & \mathbf{I}_p \\ 1 & \mathbf{0} & \boldsymbol{\beta}^\top \end{bmatrix},$$

entonces, la relación (3.1) puede ser re-escrita como $\mathbf{Z}_i = \mathbf{a} + \mathbf{B}\mathbf{r}_i$, para $i = 1, 2, \dots, n$. A través de esta relación lineal no es difícil obtener la distribución de \mathbf{Z}_i , una vez postulada la distribución de \mathbf{r}_i .

En efecto, como $\mathbf{r}_i \stackrel{\text{ind.}}{\sim} \text{El}_{2p+1}(\boldsymbol{\eta}, \boldsymbol{\Omega}; \phi)$, para $i = 1, 2, \dots, n$, entonces $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$ serán independientes y $\mathbf{Z}_i \sim \text{El}_{p+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \phi)$, con vector de localización $\boldsymbol{\mu} = \mathbf{a} + \mathbf{B}\boldsymbol{\eta}$ y matriz de dispersión $\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^\top$, es decir,

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\eta}_\xi \\ \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{\eta}_\xi \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{\xi\xi} + \boldsymbol{\Sigma}_{\delta\delta} & \boldsymbol{\Sigma}_{\xi\xi}\boldsymbol{\beta} \\ \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\xi\xi} & \sigma_\epsilon^2 + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\xi\xi}\boldsymbol{\beta} \end{bmatrix},$$

donde $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta}^\top, \sigma_\epsilon^2, \boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\delta\delta})^\top)^\top$, aquí $\mathbf{v}(\mathbf{A})$ es el vector que contiene el triángulo inferior (incluida la diagonal) de elementos de una matriz simétrica \mathbf{A} .

Por otra parte, bajo el supuesto de la existencia de la densidad, es decir, $\mathbf{r}_i \stackrel{\text{ind.}}{\sim} \text{El}_{2p+1}(\boldsymbol{\eta}, \boldsymbol{\Omega}; g)$ con $\boldsymbol{\Omega} > 0$, de (2.6), se tiene que el vector aleatorio observado \mathbf{Z}_i tiene distribución elíptica con densidad dada por

$$f_{\mathbf{Z}_i}(\mathbf{z}_i; \boldsymbol{\theta}, \boldsymbol{\nu}) = C_{p+1}(\boldsymbol{\nu}) |\boldsymbol{\Sigma}|^{-1/2} g((\mathbf{z}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{z}_i - \boldsymbol{\mu})), \quad i = 1, 2, \dots, n,$$

y $\boldsymbol{\nu}$ es el vector de parámetros que indexa a la función generadora de densidades g .

Por lo tanto, la función de verosimilitud para $\boldsymbol{\theta}$ y $\boldsymbol{\nu}$ es

$$L(\boldsymbol{\theta}, \boldsymbol{\nu}) = \prod_{i=1}^n f_{\mathbf{Z}_i}(\mathbf{z}_i; \boldsymbol{\theta}, \boldsymbol{\nu}) = C_{p+1}^n(\boldsymbol{\nu}) |\boldsymbol{\Sigma}|^{-n/2} \prod_{i=1}^n g((\mathbf{z}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{z}_i - \boldsymbol{\mu})), \quad (3.3)$$

donde $C_{p+1}(\boldsymbol{\nu})$ es la constante de normalización definida en (2.5).

3.3. Parametrización tipo Gleser

Bajo la parametrización directa definida en la sección anterior, la función de verosimilitud (3.3) asociada al modelo estructural elíptico definido en (3.1), es complicada de estudiar. Para simplificar nuestro análisis, realizamos una extensión elíptica al enfoque de Gleser (ver Gleser (1992) y Cheng & Tsai (2004)) para obtener a través de una reparametrización una versión de la función de verosimilitud más simple. Esta extensión tiene la ventaja de establecer una útil conexión entre los modelos de regresión lineal y los modelos de regresión lineal con error de medida.

Dado que $\mathbf{r}_i \stackrel{ind.}{\sim} \text{El}_{2p+1}(\boldsymbol{\eta}, \boldsymbol{\Omega}; g)$, para $i = 1, 2, \dots, n$, donde $\boldsymbol{\eta}$ y $\boldsymbol{\Omega}$ son definidos en (3.2), es posible obtener la distribución del vector aleatorio $(\boldsymbol{\xi}_i^\top, \mathbf{X}_i^\top)^\top$, la cual es dada por

$$\begin{bmatrix} \boldsymbol{\xi}_i \\ \mathbf{X}_i \end{bmatrix} \sim \text{El}_{2p} \left(\begin{bmatrix} \boldsymbol{\eta}_\xi \\ \boldsymbol{\eta}_x \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\xi\xi} & \boldsymbol{\Sigma}_{\xi x} \\ \boldsymbol{\Sigma}_{x\xi} & \boldsymbol{\Sigma}_{xx} \end{bmatrix}; g \right), \quad i = 1, 2, \dots, n.$$

De la sección 2.7, se tiene que la distribución condicional de la variable aleatoria no observada $\boldsymbol{\xi}_i$ dada la covariable observada \mathbf{X}_i es elíptica donde

$$(\boldsymbol{\xi}_i | \mathbf{X}_i = \mathbf{x}_i) \sim \text{El}_p(\boldsymbol{\mu}_\xi(\mathbf{x}_i), \boldsymbol{\Sigma}_{\xi\xi \cdot \mathbf{x}}; g_{q(\mathbf{x}_i)}), \quad (3.4)$$

donde

$$\begin{aligned} \boldsymbol{\mu}_\xi(\mathbf{x}_i) &= (\mathbf{I}_p - \boldsymbol{\Lambda}^\top) \boldsymbol{\eta}_\xi + \boldsymbol{\Lambda}^\top \mathbf{x}_i, \\ \boldsymbol{\Sigma}_{\xi\xi \cdot \mathbf{x}} &= \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda}, \end{aligned}$$

y $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{x\xi} = (\boldsymbol{\Sigma}_{\delta\delta} + \boldsymbol{\Sigma}_{\xi\xi})^{-1} \boldsymbol{\Sigma}_{\xi\xi}$. La matriz $\boldsymbol{\Lambda}$ es conocida como la *matriz de confiabilidad* y está relacionada al sesgo de las estimaciones de $\boldsymbol{\beta}$. Si $\boldsymbol{\Lambda}$ es singular, entonces $\boldsymbol{\beta}$ no es identificable y no puede ser estimado correctamente (Gleser (1992)).

Así, la densidad condicional de $\boldsymbol{\xi}_i | \mathbf{X}_i = \mathbf{x}_i$ es

$$f_{\boldsymbol{\xi}_i | \mathbf{X}_i}(\boldsymbol{\xi}_i; \boldsymbol{\mu}_\xi(\mathbf{x}_i), \boldsymbol{\Sigma}_{\xi\xi \cdot \mathbf{x}}) = |\boldsymbol{\Sigma}_{\xi\xi \cdot \mathbf{x}}|^{-1/2} g_{q(\mathbf{x}_i)}(q(\mathbf{x}_i) + u), \quad \boldsymbol{\xi}_i \in \mathbb{R}^p,$$

con $u = (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\xi(\mathbf{x}_i))^\top \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}|\mathbf{x}}^{-1} (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\xi(\mathbf{x}_i))$, $q(\mathbf{x}_i) = (\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi)$ y para cada $a \geq 0$,

$$g_a(v) = \frac{\Gamma(p/2)g(v+a)}{2\pi^{p/2} \int_0^\infty r^{p-1}g(r^2+a) dr}, u \geq 0,$$

es la función generadora de densidades condicionales (ver sección 2.7).

Considerando el vector de localización de la distribución condicional en (3.4), podemos re-escribir (3.1) como

$$Y_i = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{\xi}_i + \epsilon_i + \boldsymbol{\beta}^\top \boldsymbol{\mu}_\xi(\mathbf{x}_i) - \boldsymbol{\beta}^\top \boldsymbol{\mu}_\xi(\mathbf{x}_i), i = 1, 2, \dots, n.$$

Entonces, el modelo dado por (3.1) puede ser re-definido como

$$Y_i = \beta_0^* + \boldsymbol{\beta}^{*\top} \mathbf{X}_i + \epsilon_i^*, i = 1, 2, \dots, n, \quad (3.5)$$

donde,

$$\beta_0^* = \beta_0 + \boldsymbol{\beta}^\top (\mathbf{I}_p - \boldsymbol{\Lambda}^\top) \boldsymbol{\eta}_\xi, \quad (3.6)$$

$$\boldsymbol{\beta}^* = \boldsymbol{\Lambda} \boldsymbol{\beta}, \quad (3.7)$$

y

$$\epsilon_i^* = \boldsymbol{\beta}^\top [\boldsymbol{\xi}_i - (\mathbf{I}_p - \boldsymbol{\Lambda}^\top) \boldsymbol{\eta}_\xi - \boldsymbol{\Lambda}^\top \mathbf{X}_i] + \epsilon_i$$

Además, de (3.5), $\mathbf{X}_i = (\mathbf{0}, \mathbf{I}_p, \mathbf{I}_p) \mathbf{r}_i$ y

$$\epsilon_i^* = -\boldsymbol{\beta}^\top (\mathbf{I}_p - \boldsymbol{\Lambda}^\top) \boldsymbol{\eta}_\xi + (1, -\boldsymbol{\beta}^\top \boldsymbol{\Lambda}^\top, \boldsymbol{\beta}^\top (\mathbf{I}_p - \boldsymbol{\Lambda})^\top) \mathbf{r}_i. \quad (3.8)$$

Por lo tanto, la distribución conjunta de $(\mathbf{X}_i^\top, \epsilon_i^*)^\top$ es elíptica $(p+1)$ -variada, es decir

$$\begin{bmatrix} \mathbf{X}_i \\ \epsilon_i^* \end{bmatrix} \stackrel{ind.}{\sim} \text{El}_{p+1} \left(\begin{bmatrix} \boldsymbol{\eta}_\xi \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} & \mathbf{0} \\ \mathbf{0} & \sigma_{\epsilon^*}^2 \end{bmatrix}; g \right), i = 1, 2, \dots, n, \quad (3.9)$$

y como se ha definido anteriormente $\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} = \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}} + \boldsymbol{\Sigma}_{\boldsymbol{\delta}\boldsymbol{\delta}}$ y

$$\sigma_{\epsilon^*}^2 = \sigma_\epsilon^2 + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\boldsymbol{\delta}\boldsymbol{\delta}} \boldsymbol{\Lambda} \boldsymbol{\beta}. \quad (3.10)$$

De (3.9), ϵ_i^* y \mathbf{X}_i son no-correlacionados con distribución marginal

$$\epsilon_i^* \stackrel{iid}{\sim} \text{El}_1(0, \sigma_{\epsilon^*}^2; g) \quad \text{y} \quad \mathbf{X}_i \stackrel{iid}{\sim} \text{El}_p(\boldsymbol{\eta}_\xi, \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}; g).$$

Por lo tanto, es clara la ventaja de esta reparametrización, ya que el modelo definido en (3.5), es un modelo de regresión clásico con matriz de diseño aleatoria y el modelo está definido sólo a través de la información observada, además se tiene que ϵ_i^* y \mathbf{X}_i son no-correlacionados.

Entonces, en lo que sigue, se estudiarán las características relacionadas con el modelo reparametrizado y su conexión con el modelo de regresión lineal con error de medida clásico definido en (3.1). Por otra parte, definiendo

$$\mathbf{a}^* = \begin{bmatrix} \mathbf{0} \\ \beta_0^* \end{bmatrix} \quad \text{y} \quad \mathbf{B}^* = \begin{bmatrix} \mathbf{I}_p & \mathbf{0} \\ \boldsymbol{\beta}^{*\top} & 1 \end{bmatrix},$$

de (3.9) se tiene que

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{X}_i \\ Y_i \end{bmatrix} = \mathbf{a}^* + \mathbf{B}^* \begin{bmatrix} \mathbf{X}_i \\ \epsilon_i^* \end{bmatrix}.$$

De (2.11), obtenemos

$$\mathbf{Z}_i \stackrel{ind.}{\sim} \text{El}_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; g), \quad i = 1, 2, \dots, n,$$

donde,

$$\boldsymbol{\mu}^* = \begin{bmatrix} \boldsymbol{\eta}_\xi \\ \beta_0^* + \boldsymbol{\beta}^{*\top} \boldsymbol{\eta}_\xi \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Sigma}^* = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \boldsymbol{\beta}^* \\ \boldsymbol{\beta}^{*\top} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} & \sigma_{\epsilon^*}^2 + \boldsymbol{\beta}^{*\top} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \boldsymbol{\beta}^* \end{bmatrix}, \quad (3.11)$$

con $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) = (\boldsymbol{\mu}(\boldsymbol{\theta}^*), \boldsymbol{\Sigma}(\boldsymbol{\theta}^*))$ y $\boldsymbol{\theta}^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2, \boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\delta\delta})^\top)^\top$.

Entonces, la densidad de \mathbf{Z}_i es

$$f_{\mathbf{Z}_i}(\mathbf{z}_i; \boldsymbol{\theta}^*, \boldsymbol{\nu}) = C_{p+1}(\boldsymbol{\nu}) |\boldsymbol{\Sigma}^*|^{-1/2} g((\mathbf{z}_i - \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*-1} (\mathbf{z}_i - \boldsymbol{\mu}^*)), \quad \mathbf{z}_i \in \mathbb{R}^{p+1},$$

para $i = 1, 2, \dots, n$, con constante de normalización

$$C_{p+1}(\boldsymbol{\nu}) = \frac{\Gamma((p+1)/2)}{2\pi^{(p+1)/2} \int_0^\infty r^p g(r^2) dr}.$$

Además, en este caso se tiene que $|\boldsymbol{\Sigma}^*| = \sigma_{\epsilon^*}^2 |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}|$ y

$$\boldsymbol{\Sigma}^{*-1} = \frac{1}{\sigma_{\epsilon^*}^2} \begin{bmatrix} \sigma_{\epsilon^*}^2 \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} + \boldsymbol{\beta}^* \boldsymbol{\beta}^{*\top} & -\boldsymbol{\beta}^* \\ -\boldsymbol{\beta}^{*\top} & 1 \end{bmatrix}.$$

Entonces,

$$(\mathbf{z}_i - \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{*-1} (\mathbf{z}_i - \boldsymbol{\mu}^*) = (\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi) + \frac{\epsilon_i^{*2}}{\sigma_{\epsilon^*}^2}.$$

Por lo tanto, la densidad de \mathbf{Z}_i puede ser expresada en la forma

$$f_{\mathbf{Z}_i}(\mathbf{z}_i; \boldsymbol{\theta}^*, \boldsymbol{\nu}) = \frac{C_{p+1}(\boldsymbol{\nu})}{(\sigma_{\epsilon^*}^2)^{1/2} |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}|^{1/2}} g \left((\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi) + \frac{\epsilon_i^{*2}}{\sigma_{\epsilon^*}^2} \right), \quad (3.12)$$

luego, la función de verosimilitud de $\boldsymbol{\theta}^*$ y $\boldsymbol{\nu}$ es dada por

$$L(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = \frac{C_{p+1}^n(\boldsymbol{\nu})}{(\sigma_{\epsilon^*}^2)^{n/2} |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}|^{n/2}} \prod_{i=1}^n g \left(q(\mathbf{x}_i) + \frac{\epsilon_i^{*2}}{\sigma_{\epsilon^*}^2} \right). \quad (3.13)$$

donde $q(\mathbf{x}_i) = (\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi)$.

3.4. Identificación del modelo estructural elíptico

Notemos que, de (3.13), se tiene que la función de verosimilitud es igual a

$$L(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = f_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, \boldsymbol{\nu}),$$

es decir, corresponde a la distribución conjunta de $\mathbf{Z} = (\mathbf{Z}_1^\top, \mathbf{Z}_2^\top, \dots, \mathbf{Z}_n^\top)^\top$, que depende de $\boldsymbol{\theta}^*$, a través de $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ dados en (3.11).

Definiendo la función generadora de densidades g en (3.13), la distribución conjunta de $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n$ esta completamente caracterizada por los parámetros de localización y dispersión definidos por

$$\boldsymbol{\mu}^* = \begin{bmatrix} \boldsymbol{\mu}_X \\ \mu_Y \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Sigma}^* = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} & \boldsymbol{\sigma}_{\mathbf{X}Y} \\ \boldsymbol{\sigma}_{\mathbf{X}Y}^\top & \sigma_{Y Y}^2 \end{bmatrix}.$$

Comparando el número de elementos libres en $\boldsymbol{\mu}^*$ y $\boldsymbol{\Sigma}^*$, el cual corresponde a un total de $(p+1)(p+4)/2$ parámetros a estimar, con el número de elementos en $\boldsymbol{\theta}^*$ donde se tienen $(p+1)(p+2)$ parámetros, se tiene que el modelo dado por (3.13) está sobre-parametrizado y por tanto es no-identificado (ver Gleser (1992) y Arellano-Valle (1994)). Esto muestra que cada modelo estructural elíptico definido por (3.13), incluyendo el modelo estructural normal, es no-identificable (ver Vidal & Castro (2010)). Entonces, para hacer el problema de estimación factible, se requiere de un mínimo de $p(p+1)/2$ restricciones necesarias para identificar el modelo.

Por lo tanto, en este trabajo consideramos el supuesto que $\Sigma_{\delta\delta}$ es conocido, entonces el modelo es identificado y la relación $\boldsymbol{\theta}^* \mapsto (\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) = (\boldsymbol{\mu}(\boldsymbol{\theta}^*), \boldsymbol{\Sigma}(\boldsymbol{\theta}^*))$, es uno a uno. En este caso, $\boldsymbol{\theta}^*$ puede ser estimado consistentemente usando procedimiento de máxima verosimilitud.

Notemos que, si $\Sigma_{\delta\delta}$ es conocido, y definiendo la estimación de $\boldsymbol{\theta}^*$ por $\hat{\boldsymbol{\theta}}^*$, podemos estimar en una primera etapa los parámetros de localización y dispersión $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ y entonces resolver para $\boldsymbol{\theta}^*$ desde el sistema

$$\boldsymbol{\mu}(\hat{\boldsymbol{\theta}}^*) = \hat{\boldsymbol{\mu}}^* \quad \text{y} \quad \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}^*) = \hat{\boldsymbol{\Sigma}}^*.$$

A través de la reparametrización definida en (3.6), (3.7) y (3.10), es posible estimar la matriz de confiabilidad como

$$\hat{\boldsymbol{\Lambda}} = (\boldsymbol{\Sigma}_{\delta\delta} + \hat{\boldsymbol{\Sigma}}_{\xi\xi})^{-1} \hat{\boldsymbol{\Sigma}}_{\xi\xi}.$$

Considerando la propiedad de invarianza de los estimadores de máxima verosimilitud, se tiene de (3.6), (3.7) y (3.10) que los estimadores de máxima verosimilitud de β_0 , $\boldsymbol{\beta}$ y σ_ϵ^2 son

$$\begin{cases} \hat{\beta}_0 = \hat{\beta}_0^* - \hat{\boldsymbol{\beta}}^{*\top} (\hat{\boldsymbol{\Lambda}}^{-\top} - \mathbf{I}_p) \hat{\boldsymbol{\eta}}_\xi, \\ \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Lambda}}^{-1} \hat{\boldsymbol{\beta}}^*, \\ \hat{\sigma}_\epsilon^2 = \hat{\sigma}_{\epsilon^*}^2 - \hat{\boldsymbol{\beta}}^{*\top} \hat{\boldsymbol{\Lambda}}^{-\top} \boldsymbol{\Sigma}_{\delta\delta} \hat{\boldsymbol{\beta}}^*, \end{cases} \quad (3.14)$$

donde se define $\hat{\boldsymbol{\Lambda}}^{-\top} := (\hat{\boldsymbol{\Lambda}}^{-1})^\top = (\hat{\boldsymbol{\Lambda}}^\top)^{-1}$.

Otro supuesto usual considerado en la literatura es que la matriz de confiabilidad $\boldsymbol{\Lambda}$ es conocida. En este caso, en los estimadores previos sólo se debe reemplazar $\hat{\boldsymbol{\Lambda}}$ por $\boldsymbol{\Lambda}$, y la estimación de la matriz de dispersión de los errores de medición es dada por $\hat{\boldsymbol{\Sigma}}_{\delta\delta} = \hat{\boldsymbol{\Sigma}}_{\xi\xi}(\boldsymbol{\Lambda}^{-1} - \mathbf{I}_p)$. Es más, de (3.14) es posible recuperar la distribución asintótica de $\hat{\boldsymbol{\theta}}$ de la distribución asintótica de $\hat{\boldsymbol{\theta}}^*$ considerando el método del Jacobiano (ver sección 3.7). Para una revisión más detallada de la identificación en los modelos con errores de medición, particularmente considerando la normalidad como supuesto distribucional, se puede ver en Wansbeek & Meijer (2000).

3.5. Estimación de máxima verosimilitud

Para determinar los estimadores de máxima verosimilitud del modelo estructural elíptico definido en (3.13), debemos también considerar el vector de parámetros $\boldsymbol{\nu}$ que indexa la función generadora de densidades g , que puede ser de interés para

propósitos inferenciales.

De (3.13)

$$L(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = C_{p+1}^n(\boldsymbol{\nu})(\sigma_{\epsilon^*}^2)^{-n/2} |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}|^{-n/2} \prod_{i=1}^n g(v_i),$$

donde

$$v_i = (\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi) + \frac{\epsilon_i^{*2}}{\sigma_{\epsilon^*}^2},$$

con

$$\epsilon_i^* = y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i.$$

Considerando que $\boldsymbol{\Sigma}_{\delta\delta}$ es conocido, definamos $\boldsymbol{\theta}^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2, \boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top)^\top$.

Notemos que v_i sólo depende de $\boldsymbol{\theta}^*$ y la constante de normalización C_{p+1} depende de $\boldsymbol{\nu}$ debido a que su derivación involucra la función g , indexada por $\boldsymbol{\nu}$. Por lo tanto, la función de log-verosimilitud para $\boldsymbol{\theta}^*$ y $\boldsymbol{\nu}$, definida por $\ell(\boldsymbol{\theta}^*, \boldsymbol{\nu}) := \log L(\boldsymbol{\theta}^*, \boldsymbol{\nu})$, es

$$\ell(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = n \log C_{p+1}(\boldsymbol{\nu}) - \frac{n}{2} \log \sigma_{\epsilon^*}^2 - \frac{n}{2} \log |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}| + \sum_{i=1}^n \log g(v_i). \quad (3.15)$$

Para derivar el primer diferencial de (3.15), seguimos la metodología usada en Arellano-Valle (2010) para obtener el vector de score. Adoptamos el enfoque de cálculo matricial como es desarrollado en Magnus & Neudecker (1999).

Sea la partición $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*\top}, \boldsymbol{\theta}_2^{*\top})^\top$, con $\boldsymbol{\theta}_1^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2)^\top$ y $\boldsymbol{\theta}_2^* = (\boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top)^\top$. Esta partición de $\boldsymbol{\theta}^*$ no es arbitraria, $\boldsymbol{\theta}_1^*$ corresponde a los parámetros relacionados al modelo de regresión dado en (3.5), y usualmente son los parámetros de interés en el análisis de regresión y $\boldsymbol{\theta}_2^*$ son los parámetros relacionados a las covariables medidas con error.

Entonces, el primer diferencial de $\ell^* := \ell(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ es dado por $d\ell^* = d_1\ell^* + d_2\ell^* + d_\nu\ell^*$, donde $d_\alpha\ell^*$ corresponde al primer diferencial de ℓ^* con respecto a α , por tanto

$$\begin{aligned} d_1\ell^* &= S_{\beta_0^*} d\beta_0^* + \mathbf{S}_{\boldsymbol{\beta}^*}^\top d\boldsymbol{\beta}^* + S_{\sigma_{\epsilon^*}^2} d\sigma_{\epsilon^*}^2, \\ d_2\ell^* &= \mathbf{S}_{\boldsymbol{\eta}_\xi}^\top d\boldsymbol{\eta}_\xi + \mathbf{S}_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})}^\top d\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}), \\ d_\nu\ell^* &= \mathbf{S}_\nu^\top d\boldsymbol{\nu}. \end{aligned}$$

A continuación, veremos algunos detalles en el cálculo de éstos diferenciales.

Primer diferencial parcial con respecto a θ_1^* :

Usando las reglas $d_1 \log \sigma_{\epsilon^*}^2 = \sigma_{\epsilon^*}^{-2} d\sigma_{\epsilon^*}^2$, $d_1 \epsilon_i^* = -d\beta_0^* - \mathbf{x}_i^\top d\boldsymbol{\beta}^*$, donde d_1 corresponde al primer diferencial parcial con respecto a θ_1^* , y

$$d_1 v_i = -\frac{2}{\sigma_{\epsilon^*}^2} \epsilon_i^* d\beta_0^* - \frac{2}{\sigma_{\epsilon^*}^2} \epsilon_i^* \mathbf{x}_i^\top d\boldsymbol{\beta}^* - \frac{1}{(\sigma_{\epsilon^*}^2)^2} \epsilon_i^{*2} d\sigma_{\epsilon^*}^2,$$

después de una extensa álgebra, el primer diferencial parcial de ℓ^* con respecto a θ_1^* es dado por

$$\begin{aligned} d_1 \ell^* &= \left(-\frac{2}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) \epsilon_i^* \right) d\beta_0^* + \left(-\frac{2}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) \epsilon_i^* \mathbf{x}_i^\top \right) d\boldsymbol{\beta}^* \\ &+ \left(-\frac{n}{2\sigma_{\epsilon^*}^2} - \frac{1}{(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) \epsilon_i^{*2} \right) d\sigma_{\epsilon^*}^2, \end{aligned}$$

donde

$$W_g^{(v_i)}(v_i) = \frac{\partial \log g(v_i)}{\partial v_i}.$$

Por lo tanto, el primer diferencial parcial, $d_1 \ell^*$, es dado por

$$d_1 \ell^* = S_{\beta_0^*} d\beta_0^* + \mathbf{S}_{\boldsymbol{\beta}^*}^\top d\boldsymbol{\beta}^* + S_{\sigma_{\epsilon^*}^2} d\sigma_{\epsilon^*}^2,$$

con

$$\begin{cases} S_{\beta_0^*} = -\frac{2}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) \epsilon_i^*, \\ \mathbf{S}_{\boldsymbol{\beta}^*} = -\frac{2}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) \epsilon_i^* \mathbf{x}_i, \\ S_{\sigma_{\epsilon^*}^2} = -\frac{n}{2\sigma_{\epsilon^*}^2} - \frac{1}{(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) \epsilon_i^{*2}. \end{cases} \quad (3.16)$$

Primer diferencial parcial con respecto a θ_2^* :

Considerando que

$$\begin{aligned} d_2 \log |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}| &= \text{tr}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} d\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}}), \\ d_2 \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} &= -\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (d\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}}) \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}, \\ d_2 \mathbf{x}_{i,0} &= -d\boldsymbol{\eta}_{\boldsymbol{\xi}}. \end{aligned}$$

donde $\mathbf{x}_{i,0} = \mathbf{x}_i - \boldsymbol{\eta}_\xi$ y

$$d_2 v_i = -\text{tr}(\mathbf{x}_{i,0}^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (d\boldsymbol{\Sigma}_{\xi\xi}) \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{x}_{i,0}) - 2\mathbf{x}_{i,0}^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} d\boldsymbol{\eta}_\xi.$$

Tenemos que el primer diferencial parcial de ℓ^* con respecto a $\boldsymbol{\theta}_2^*$ es dado por

$$\begin{aligned} d_2 \ell^* &= -\frac{n}{2} \text{tr}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} d\boldsymbol{\Sigma}_{\xi\xi}) - 2 \sum_{i=1}^n W_g^{(v_i)}(v_i) \mathbf{x}_{i,0}^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} d\boldsymbol{\eta}_\xi \\ &\quad - \sum_{i=1}^n W_g^{(v_i)}(v_i) \text{tr}(\mathbf{x}_{i,0}^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (d\boldsymbol{\Sigma}_{\xi\xi}) \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{x}_{i,0}). \end{aligned}$$

Sea \mathbf{D} , la matrix de duplicación $p^2 \times p(p+1)/2$ -dimensional tal que

$$\mathbf{D}\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}) = \text{vec}(\boldsymbol{\Sigma}_{\xi\xi})$$

o en forma equivalente, $\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}) = \mathbf{D}_+ \text{vec}(\boldsymbol{\Sigma}_{\xi\xi})$, donde $\mathbf{D}_+ = (\mathbf{D}^\top \mathbf{D})^{-1} \mathbf{D}^\top$ (ver Magnus & Neudecker (1999)).

Entonces, se tiene que de la expresión anterior para $d_2 \ell^*$, que

$$\text{tr}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} d\boldsymbol{\Sigma}_{\xi\xi}) = \text{vec}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1})^\top \mathbf{D} d\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}),$$

y

$$\text{tr}(\mathbf{x}_{i,0}^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (d\boldsymbol{\Sigma}_{\xi\xi}) \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \mathbf{x}_{i,0}) = (\mathbf{x}_{i,0}^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \otimes \mathbf{x}_{i,0}^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D} d\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}),$$

entonces el primer diferencial parcial ℓ^* con respecto a $\boldsymbol{\theta}_2^*$ es dado por

$$d_2 \ell^* = \mathbf{S}_{\boldsymbol{\eta}_\xi}^\top d\boldsymbol{\eta}_\xi + \mathbf{S}_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})}^\top d\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})$$

con

$$\begin{cases} \mathbf{S}_{\boldsymbol{\eta}_\xi} = -2\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \sum_{i=1}^n W_g^{(v_i)}(v_i) \mathbf{x}_{i,0}, \\ \mathbf{S}_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})} = -\frac{n}{2} \mathbf{D}^\top \text{vec}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) - \mathbf{D}^\top (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \sum_{i=1}^n W_g^{(v_i)}(v_i) (\mathbf{x}_{i,0} \otimes \mathbf{x}_{i,0}). \end{cases} \quad (3.17)$$

Primer diferencial parcial con respecto a $\boldsymbol{\nu}$:

Respecto a $\boldsymbol{\nu}$ tenemos que $d_{\boldsymbol{\nu}}\ell^* = \mathbf{S}_{\boldsymbol{\nu}}^{\top} d\boldsymbol{\nu}$, con

$$\mathbf{S}_{\boldsymbol{\nu}} = n\mathbf{C}'_{p+1}(\boldsymbol{\nu}) + \sum_{i=1}^n \mathbf{W}_g^{(\boldsymbol{\nu})}(v_i), \quad (3.18)$$

donde

$$\mathbf{W}_g^{(\boldsymbol{\nu})}(v_i) = \left[\frac{\partial \log g(v_i)}{\partial \nu_1}, \dots, \frac{\partial \log g(v_i)}{\partial \nu_q} \right]^{\top},$$

y $\mathbf{C}'_{p+1}(\boldsymbol{\nu})$ denota al vector de las primeras derivadas parciales de la constante de normalización definido por

$$\mathbf{C}'_{p+1}(\boldsymbol{\nu}) = \left[\frac{\partial \log C_{p+1}(\boldsymbol{\nu})}{\partial \nu_1}, \dots, \frac{\partial \log C_{p+1}(\boldsymbol{\nu})}{\partial \nu_q} \right]^{\top}.$$

Por lo tanto, obtenemos el siguiente resultado para el primer diferencial de ℓ^* .

Proposición 3.1 (Vector de score del modelo estructural elíptico). *El primer diferencial de la función de log-verosimilitud (3.15) es dado por*

$$d\ell^* = S_{\beta_0^*} d\beta_0^* + \mathbf{S}_{\boldsymbol{\beta}^*}^{\top} d\boldsymbol{\beta}^* + S_{\sigma_{\epsilon^*}^2} d\sigma_{\epsilon^*}^2 + \mathbf{S}_{\boldsymbol{\eta}_{\xi}}^{\top} d\boldsymbol{\eta}_{\xi} + \mathbf{S}_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})}^{\top} d\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}) + \mathbf{S}_{\boldsymbol{\nu}}^{\top} d\boldsymbol{\nu}.$$

Así, el vector de score es definido por $\mathbf{S}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = (S_{\beta_0^*}, \mathbf{S}_{\boldsymbol{\beta}^*}^{\top}, S_{\sigma_{\epsilon^*}^2}, \mathbf{S}_{\boldsymbol{\eta}_{\xi}}^{\top}, \mathbf{S}_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})}^{\top}, \mathbf{S}_{\boldsymbol{\nu}}^{\top})^{\top}$, con componentes definidos en (3.16)-(3.18).

Si ℓ^* es la función de log-verosimilitud del vector de parámetros $(\boldsymbol{\theta}^{*\top}, \boldsymbol{\nu}^{\top})^{\top}$, el estimador de máxima verosimilitud $(\hat{\boldsymbol{\theta}}^{*\top}, \hat{\boldsymbol{\nu}}^{\top})^{\top}$ es la solución a las ecuaciones de verosimilitud, $\mathbf{S}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = \mathbf{0}$, ya que la matriz de información observada es definida positiva cuando es evaluada en $(\hat{\boldsymbol{\theta}}^{*\top}, \hat{\boldsymbol{\nu}}^{\top})^{\top}$. Para ciertas elecciones de la función generadora de densidades, el resultado presentado en la Proposición 3.1, facilita la implementación del modelo con las rutinas estándar de maximización.

Por otro lado, la matriz de covarianza asintótica de los estimadores de máxima verosimilitud puede ser estimada usando la matriz Hessiana, la cual puede ser determinada numéricamente usando, por ejemplo, la rutina `optim` en la plataforma R (R Core Team (2018)).

3.6. Representación jerárquica

La representación de los modelos estructurales elípticos en términos de la parametrización de Gleser, permite la siguiente representación jerárquica

$$\begin{aligned} Y_i | \mathbf{X}_i = \mathbf{x}_i &\stackrel{ind.}{\sim} \text{El}_1(\beta_0^* + \boldsymbol{\beta}^{*\top} \mathbf{x}_i, \sigma_{\epsilon^*}^2; g_{q(\mathbf{x}_i)}), \\ \mathbf{X}_i &\stackrel{iid}{\sim} \text{El}_p(\boldsymbol{\eta}_\xi, \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}; g), \end{aligned}$$

para $i = 1, 2, \dots, n$, donde

$$q(\mathbf{x}_i) = (\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi) \quad \text{y} \quad \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} = \boldsymbol{\Sigma}_{\xi\xi} + \boldsymbol{\Sigma}_{\delta\delta}.$$

Considerando el vector de parámetros $\boldsymbol{\theta}^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2, \boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top)^\top$ y $\boldsymbol{\nu}$ el vector de parámetros que indexa a la función generadora de densidades g .

Sea la partición $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*\top}, \boldsymbol{\theta}_2^{*\top})^\top$ con $\boldsymbol{\theta}_1^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2)^\top$ y $\boldsymbol{\theta}_2^* = (\boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top)^\top$, entonces la densidad conjunta de $\mathbf{Z}_i = (\mathbf{X}_i^\top, Y_i)^\top$ dada la representación jerárquica es

$$f_{\mathbf{Z}_i}(\mathbf{z}_i; \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\nu}) = f_{Y_i | \mathbf{X}_i}(y_i; \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\nu}) f_{\mathbf{X}_i}(\mathbf{x}_i; \boldsymbol{\theta}_2^*, \boldsymbol{\nu}). \quad (3.19)$$

Notemos que la dependencia de $f_{Y_i | \mathbf{X}_i}$ respecto a $\boldsymbol{\theta}_2^*$ viene dada por $q(\mathbf{x}_i)$ a través de la función generadora de densidades condicionales $g_{q(\mathbf{x}_i)}$.

Luego, la densidad condicional de $Y_i | \mathbf{X}_i = \mathbf{x}_i$ es

$$f_{Y_i | \mathbf{X}_i}(y_i; \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\nu}) = \frac{1}{(\sigma_{\epsilon^*}^2)^{1/2}} g_{q(\mathbf{x}_i)} \left(\frac{\epsilon_i^{*2}}{\sigma_{\epsilon^*}^2} \right), \quad y_i \in \mathbb{R},$$

donde $\epsilon_i^* = y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i$ y

$$g_{q(\mathbf{x}_i)} \left(\frac{\epsilon_i^{*2}}{\sigma_{\epsilon^*}^2} \right) = \frac{g(q(\mathbf{x}_i) + \epsilon_i^{*2}/\sigma_{\epsilon^*}^2)}{2 \int_0^\infty g(q(\mathbf{x}_i) + r^2) dr}.$$

Por otro lado, la densidad de \mathbf{X}_i es dada por

$$f_{\mathbf{X}_i}(\mathbf{x}_i; \boldsymbol{\theta}_2^*, \boldsymbol{\nu}) = \frac{2C_{p+1}(\boldsymbol{\nu})}{|\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}|^{1/2}} \int_0^\infty g(q(\mathbf{x}_i) + r^2) dr, \quad \mathbf{x}_i \in \mathbb{R}^p.$$

La función de verosimilitud para $\boldsymbol{\theta}_1^*$, $\boldsymbol{\theta}_2^*$ y $\boldsymbol{\nu}$ es

$$L(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\nu}) = \prod_{i=1}^n f_{\mathbf{Z}_i}(\mathbf{z}_i; \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\nu}).$$

De (3.19) y dada la partición de $\boldsymbol{\theta}^*$, la función de verosimilitud puede ser re-escrita como sigue

$$L(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\nu}) = \prod_{i=1}^n f_{Y_i|\mathbf{X}_i}(y_i; \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\nu}) f_{\mathbf{X}_i}(\mathbf{x}_i; \boldsymbol{\theta}_2^*, \boldsymbol{\nu}). \quad (3.20)$$

Si consideremos el caso normal, es decir, $\mathbf{Z}_i \sim N_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, donde $\boldsymbol{\mu}^*$ y $\boldsymbol{\Sigma}^*$ son definidas en (3.11), la representación jerárquica del modelo estructural normal viene dada por

$$Y_i|\mathbf{X}_i = \mathbf{x}_i \stackrel{ind.}{\sim} N_1(\beta_0^* + \boldsymbol{\beta}^{*\top} \mathbf{x}_i, \sigma_{\epsilon^*}^2) \quad \text{y} \quad \mathbf{X}_i \stackrel{iid}{\sim} N_p(\boldsymbol{\eta}_\xi, \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}),$$

entonces es fácil notar que $\boldsymbol{\theta}_1^*$ y $\boldsymbol{\theta}_2^*$ están en variación libre, es decir, $\boldsymbol{\theta}_i^*$ no variará con $\boldsymbol{\theta}_j^*$ para $i \neq j$ (ver Florens & Mouchart (1980)), y por tanto \mathbf{X}_i es débilmente exógeno con respecto a $\boldsymbol{\theta}_1^*$ (ver Engle et al. (1983)).

Esto nos permite ignorar $f_{\mathbf{X}_i}(\mathbf{x}_i; \boldsymbol{\theta}_2^*)$ en el proceso de estimación de $\boldsymbol{\theta}^*$, considerando que usualmente el interés radica en los parámetros de $\boldsymbol{\theta}_1^*$. Es decir, bajo el supuesto de normalidad, $f_{\mathbf{Z}_i}(\mathbf{z}_i; \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*)$, se puede reducir a $\prod_{i=1}^n f_{Y_i|\mathbf{X}_i}(y_i; \boldsymbol{\theta}_1^*)$ (ver Spanos (1994) y Castro (2008)).

Además, el modelo de regresión normal puede ser especificado en términos de los dos primeros momentos de $f_{Y_i|\mathbf{X}_i}(y_i; \boldsymbol{\theta}_1^*)$ como sigue

$$E[Y_i|\mathbf{X}_i = \mathbf{x}_i] = \beta_0^* + \boldsymbol{\beta}^{*\top} \mathbf{x}_i \quad \text{y} \quad V(Y_i|\mathbf{X}_i = \mathbf{x}_i) = \sigma_{\epsilon^*}^2,$$

notando que la varianza condicional es homoscedástica. El supuesto de reducción del modelo normal es útil para los propósitos de especificación debido que provee un enlace entre la información proporcionada por los datos (que se relacionan con $f_{\mathbf{Z}_i}(\mathbf{z}_i; \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*)$) y el modelo ($f_{Y_i|\mathbf{X}_i}(y_i; \boldsymbol{\theta}_1^*)$). Esto nos permite evaluar (al menos indirectamente) lo apropiado del modelo mediante la evaluación de la idoneidad de la información proporcionada por los datos.

Por otro lado, considerando el caso elíptico, es decir, $\mathbf{Z}_i \sim \text{El}_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; g)$, la media y varianza condicional del modelo son

$$E[Y_i|\mathbf{X}_i = \mathbf{x}_i] = \beta_0^* + \boldsymbol{\beta}^{*\top} \mathbf{x}_i \quad \text{y} \quad V(Y_i|\mathbf{X}_i = \mathbf{x}_i) = \alpha_{q(\mathbf{x}_i)} \sigma_{\epsilon^*}^2,$$

donde $\alpha_{q(\mathbf{x}_i)} = -2\phi'_{q(\mathbf{x}_i)}(0)$ (ver Arellano-Valle (1994)). Entonces, en el caso elíptico las varianzas condicionales de la regresión $Y_i|\mathbf{X}_i = \mathbf{x}_i$ son heteroscedástica.

Podemos notar que los modelos de regresión elípticos surgen naturalmente como una extensión en el contexto de la reducción (3.20) reemplazando la distribución normal

con otro miembro de la familia de distribución elíptica (ver Tabla 2.1). Además, del capítulo 2, las distribuciones elípticas comparten varias propiedades de la distribución normal (Muirhead (1982), Fang et al. (1990), Arellano-Valle (1994)), pero ellas difieren del caso normal en dos importantes aspectos:

1. las varianzas condicionales son heteroscedásticas (Kelker (1970)), y
2. la no-correlación no implica independencia (Zellner (1976a)).

Desde el punto de vista de la especificación del modelo, la elección de la familia elíptica es motivada por el hecho que permite modelar algunas importantes fuentes de heteroscedasticidad: heterogeneidad individual, colas más (o menos) pesadas, y dependencia no-lineal. En el caso de la heterogeneidad individual, es conocido que un subconjunto de distribuciones en esta familia puede ser generada por mezcla de normales. Es también conocido que estas familias incluyen las distribuciones leptocúrticas (por ejemplo, la distribución t) y platicúrticas (por ejemplo, la distribución Pearson Tipo II) en ambos casos se tienen varianzas heteroscedásticas.

Por lo tanto, con lo mencionado anteriormente, la distribución normal es la única distribución en la familia elíptica para la cual \mathbf{X}_i es débilmente exógeno (ver Engle et al. (1983) y Spanos (1994)). Entonces, esto sugiere que en la familia elíptica los parámetros θ_1^* y θ_2^* en $f_{Y_i|\mathbf{X}_i}(y_i; \theta_1^*, \theta_2^*, \nu)$ y $f_{\mathbf{X}_i}(\mathbf{x}_i; \theta_2^*, \nu)$, respectivamente, no están en variación libre y por tanto no son exógenos, debido a esto no podemos ignorar la distribución marginal de \mathbf{X}_i en la estimación de θ_1^* .

En el modelo estructural elíptico en términos de la parametrización de Gleser, el interés radica en la estimación del vector de parámetros θ_1^* , por tanto θ_2^* y ν pueden ser tratados como parámetros molestos y debido a la no exogeneidad no es factible eliminar θ_2^* y ν , por ejemplo, a través de condicionamientos o factorización, es usual considerar el enfoque de verosimilitud para la estimación.

Varios estudios en la literatura muestran las ventajas del enfoque de verosimilitud en los modelos con error de medida sobre alternativas basadas en supuestos más débiles. En particular, las ventajas son sustanciales en términos de grandes muestras (Schafer & Purdy (1996) y Kuchenhoff & Carroll (1997)). A pesar de esto, el enfoque de verosimilitud no ha tenido una cantidad considerable de aplicaciones comparadas con la cantidad de alternativas propuestas (Carroll et al. (2006)). La principal razón es la dificultad computacional relacionada a la evaluación y maximización de la función de verosimilitud.

Por lo tanto, dada la representación jerárquica, una alternativa para obtener los estimadores de máxima verosimilitud, es considerar el enfoque de pseudo-verosimilitud

(ver Gong & Samaniego (1981) y Guolo (2011)). Donde el objetivo es expresar la verosimilitud (3.20) como una función de los parámetros de interés $\boldsymbol{\theta}_1^*$. Por lo tanto, la maximización se debe realizar en dos etapas:

1. Los parámetros molestos $\boldsymbol{\theta}_2^*$ y $\boldsymbol{\nu}$ son estimados convenientemente.
2. La estimación de $\boldsymbol{\theta}_1^*$ es obtenida maximizando la pseudo-verosimilitud, donde se consideran las estimaciones de $\boldsymbol{\theta}_2^*$ y $\boldsymbol{\nu}$.

Entonces, la estimación de $\boldsymbol{\theta}_1^*$ maximiza la pseudo-verosimilitud definida por

$$L^{(p)}(\boldsymbol{\theta}_1^*, \hat{\boldsymbol{\theta}}_2^*, \hat{\boldsymbol{\nu}}) = \prod_{i=1}^n f_{Y_i|X_i}(y_i; \boldsymbol{\theta}_1^*, \hat{\boldsymbol{\theta}}_2^*, \hat{\boldsymbol{\nu}}) f_{X_i}(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_2^*, \hat{\boldsymbol{\nu}}). \quad (3.21)$$

La estimación de los parámetros molestos $\boldsymbol{\theta}_2$ y $\boldsymbol{\nu}$ puede ser obtenida maximizando la función de verosimilitud reducida

$$L^{(r)}(\boldsymbol{\theta}_2^*, \boldsymbol{\nu}) = \prod_{i=1}^n f_{X_i}(\mathbf{x}_i; \boldsymbol{\theta}_2^*, \boldsymbol{\nu}). \quad (3.22)$$

Definiendo la función de log-pseudo-verosimilitud de (3.21), como

$$\begin{aligned} \ell^{(p)}(\boldsymbol{\theta}_1^*, \hat{\boldsymbol{\theta}}_2^*, \hat{\boldsymbol{\nu}}) &:= \log L^{(p)}(\boldsymbol{\theta}_1^*, \hat{\boldsymbol{\theta}}_2^*, \hat{\boldsymbol{\nu}}) \\ &= \sum_{i=1}^n \log f_{Y_i|X_i}(y_i; \boldsymbol{\theta}_1^*, \hat{\boldsymbol{\theta}}_2^*, \hat{\boldsymbol{\nu}}) + \sum_{i=1}^n \log f_{X_i}(\mathbf{x}_i; \hat{\boldsymbol{\theta}}_2^*, \hat{\boldsymbol{\nu}}). \end{aligned} \quad (3.23)$$

La maximización de (3.23) puede ser simplificada removiendo el último término, ya que no depende de los parámetros de interés $\boldsymbol{\theta}_1^*$. Guolo (2011), propone un conveniente enfoque computacional para maximizar la función de pseudo-verosimilitud (3.21) a través de un algoritmo tipo Monte Carlo-EM.

3.7. Matriz de información

Esta sección concierne el estudio de la matriz de información del modelo estructural elíptico bajo la reparametrización de Gleser dado en (3.5). En este sentido, se dan expresiones simples para calcular los componentes de la matriz de información esperada, en la cual sólo algunos momentos univariados deben ser determinados numéricamente.

3.7.1. Matriz de información observada

La matriz de información observada corresponde a menos la matriz Hessiana definida como $\mathbf{H}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$, la cual es determinada a través del segundo diferencial de ℓ^* definido por

$$d(d\ell^*) = d^2\ell^* = (d(\boldsymbol{\theta}^*, \boldsymbol{\nu}))^\top \mathbf{H}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) d(\boldsymbol{\theta}^*, \boldsymbol{\nu}),$$

donde

$$d(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = (d\beta_0^*, (d\boldsymbol{\beta}^*)^\top, d\sigma_{\epsilon^*}^2, (d\boldsymbol{\eta}_\xi)^\top, (d\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}))^\top, (d\boldsymbol{\nu})^\top)^\top$$

En este caso

$$d^2\ell^* = S'_{\beta_0^*} d\beta_0^* + (\mathbf{S}'_{\boldsymbol{\beta}^*})^\top d\boldsymbol{\beta}^* + (S'_{\sigma_{\epsilon^*}^2}) d\sigma_{\epsilon^*}^2 + (\mathbf{S}'_{\boldsymbol{\eta}_\xi})^\top d\boldsymbol{\eta}_\xi + (\mathbf{S}'_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})})^\top d\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}) + (\mathbf{S}'_{\boldsymbol{\nu}})^\top d\boldsymbol{\nu},$$

y, $S'_{\beta_0^*}$, $\mathbf{S}'_{\boldsymbol{\beta}^*}$, $S'_{\sigma_{\epsilon^*}^2}$, $\mathbf{S}'_{\boldsymbol{\eta}_\xi}$, $\mathbf{S}'_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})}$ y $\mathbf{S}'_{\boldsymbol{\nu}}$, corresponden al primer diferencial de (3.16)-(3.18).

Sea la partición $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*\top}, \boldsymbol{\theta}_2^{*\top})^\top$, con $\boldsymbol{\theta}_1^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2)^\top$ y $\boldsymbol{\theta}_2^* = (\boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top)^\top$, entonces se tiene que la matriz de información observada $\mathbf{I}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ es una matriz por bloques dada por

$$\mathbf{I}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = -\mathbf{H}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = \begin{bmatrix} \mathbf{I}_{\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_1^*} & \mathbf{I}_{\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*} & \mathbf{I}_{\boldsymbol{\theta}_1^*, \boldsymbol{\nu}} \\ & \mathbf{I}_{\boldsymbol{\theta}_2^*, \boldsymbol{\theta}_2^*} & \mathbf{I}_{\boldsymbol{\theta}_2^*, \boldsymbol{\nu}} \\ & & \mathbf{I}_{\boldsymbol{\nu}, \boldsymbol{\nu}} \end{bmatrix}, \quad (3.24)$$

donde los elementos de $\mathbf{I}_{\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_1^*}$ son

$$\mathbf{I}_{\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_1^*} = \begin{bmatrix} I_{\beta_0^*, \beta_0^*} & \mathbf{I}_{\beta_0^*, \boldsymbol{\beta}^*} & I_{\beta_0^*, \sigma_{\epsilon^*}^2} \\ & \mathbf{I}_{\boldsymbol{\beta}^*, \boldsymbol{\beta}^*} & \mathbf{I}_{\boldsymbol{\beta}^*, \sigma_{\epsilon^*}^2} \\ & & I_{\sigma_{\epsilon^*}^2, \sigma_{\epsilon^*}^2} \end{bmatrix}, \quad (3.25)$$

con

$$\begin{aligned} I_{\beta_0^*, \beta_0^*} &= -\frac{2}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) - \frac{4}{(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^{*2}, \\ \mathbf{I}_{\beta_0^*, \boldsymbol{\beta}^*} &= -\frac{2}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) \mathbf{x}_i^\top - \frac{4}{(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^{*2} \mathbf{x}_i^\top, \\ I_{\beta_0^*, \sigma_{\epsilon^*}^2} &= -\frac{2}{(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) \epsilon_i^* - \frac{2}{(\sigma_{\epsilon^*}^2)^3} \sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^{*3}, \\ \mathbf{I}_{\boldsymbol{\beta}^*, \boldsymbol{\beta}^*} &= -\frac{2}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) \mathbf{x}_i \mathbf{x}_i^\top - \frac{4}{(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^{*2} \mathbf{x}_i \mathbf{x}_i^\top, \end{aligned}$$

$$\begin{aligned}\mathbf{I}_{\beta^*, \sigma_{\epsilon^*}^2} &= -\frac{2}{(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n W_g^{(v_i)}(v_i) \epsilon_i^* \mathbf{x}_i - \frac{2}{(\sigma_{\epsilon^*}^2)^3} \sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^{*3} \mathbf{x}_i, \\ \mathbf{I}_{\sigma_{\epsilon^*}^2, \sigma_{\epsilon^*}^2} &= -\frac{n}{2(\sigma_{\epsilon^*}^2)^2} - \frac{2}{(\sigma_{\epsilon^*}^2)^3} \sum_{i=1}^n W_g^{(v_i)}(v_i) \epsilon_i^{*2} - \frac{1}{(\sigma_{\epsilon^*}^2)^4} \sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^{*4}.\end{aligned}$$

y

$$W_g^{(v_i, v_i)}(v_i) = \frac{\partial W_g^{(v_i)}}{\partial v_i}.$$

Similarmente, tenemos que para $\mathbf{I}_{\theta_2^*, \theta_2^*}$

$$\mathbf{I}_{\theta_2^*, \theta_2^*} = \begin{bmatrix} \mathbf{I}_{\eta_\xi, \eta_\xi} & \mathbf{I}_{\eta_\xi, v(\Sigma_{\xi\xi})} \\ \mathbf{I}_{v(\Sigma_{\xi\xi}), v(\Sigma_{\xi\xi})} & \end{bmatrix}, \quad (3.26)$$

con elementos dados por

$$\begin{aligned}\mathbf{I}_{\eta_\xi, \eta_\xi} &= -2\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \sum_{i=1}^n W_g^{(v_i)}(v_i) - 4\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \left(\sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \mathbf{x}_{i,0} \mathbf{x}_{i,0}^\top \right) \Sigma_{\mathbf{X}\mathbf{X}}^{-1}, \\ \mathbf{I}_{\eta_\xi, v(\Sigma_{\xi\xi})} &= -2 \left(\sum_{i=1}^n W_g^{(v_i)}(v_i) (\mathbf{x}_{i,0}^\top \otimes \mathbf{I}_p) \right) (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D} \\ &\quad - 2\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \left(\sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \mathbf{x}_{i,0} (\mathbf{x}_{i,0}^\top \otimes \mathbf{x}_{i,0}^\top) \right) (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D}, \\ \mathbf{I}_{v(\Sigma_{\xi\xi}), v(\Sigma_{\xi\xi})} &= -\frac{n}{2} \mathbf{D}^\top (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D} \\ &\quad - \mathbf{D}^\top (\Sigma_{\mathbf{X}\mathbf{X}}^{-1/2} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \left(\sum_{i=1}^n W_g^{(v_i)}(v_i) (\mathbf{I}_p \otimes \mathbf{x}_{i,0} \mathbf{x}_{i,0}^\top) \right) (\Sigma_{\mathbf{X}\mathbf{X}}^{-1/2} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D} \\ &\quad - \mathbf{D}^\top (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1/2}) \left(\sum_{i=1}^n W_g^{(v_i)}(v_i) (\mathbf{x}_{i,0} \mathbf{x}_{i,0}^\top \otimes \mathbf{I}_p) \right) (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1/2}) \mathbf{D} \\ &\quad - \mathbf{D}^\top (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \left(\sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) (\mathbf{x}_{i,0} \mathbf{x}_{i,0}^\top \otimes \mathbf{x}_{i,0} \mathbf{x}_{i,0}^\top) \right) (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D}.\end{aligned}$$

Además, $\mathbf{I}_{\nu, \nu}$ es dado por

$$\mathbf{I}_{\nu, \nu} = n\mathbf{C}_{p+1}''(\boldsymbol{\nu}) + \sum_{i=1}^n \mathbf{W}_g^{(\nu, \nu)}(v_i),$$

donde $\mathbf{C}_{p+1}''(\boldsymbol{\nu})$ es la matriz $q \times q$ -dimensional de las segundas derivadas de $\log C_{p+1}(\boldsymbol{\nu})$

con elementos

$$\mathbf{C}_{p+1}''(\boldsymbol{\nu}) = \left[\frac{\partial^2 \log C_{p+1}(\boldsymbol{\nu})}{\partial \nu_p \partial \nu_r} \right]_{r,p},$$

con $1 \leq r, p \leq q$.

De igual forma, tenemos que la matriz $q \times q$ -dimensional, $\mathbf{W}_g^{(\boldsymbol{\nu}, \boldsymbol{\nu})}(v_i)$ tiene elementos

$$\mathbf{W}_g^{(\boldsymbol{\nu}, \boldsymbol{\nu})}(v_i) = \left[\frac{\partial^2 \log g(v_i)}{\partial \nu_p \partial \nu_r} \right]_{r,p},$$

donde $1 \leq r, p \leq q$.

Por otro lado,

$$\mathbf{I}_{\theta_1^*, \theta_2^*} = \begin{bmatrix} \mathbf{I}_{\beta_0^*, \eta_\xi} & \mathbf{I}_{\beta_0^*, \mathbf{v}(\Sigma_{\xi\xi})} \\ \mathbf{I}_{\beta^*, \eta_\xi} & \mathbf{I}_{\beta^*, \mathbf{v}(\Sigma_{\xi\xi})} \\ \mathbf{I}_{\sigma_{\epsilon^*}^2, \eta_\xi} & \mathbf{I}_{\sigma_{\epsilon^*}^2, \mathbf{v}(\Sigma_{\xi\xi})} \end{bmatrix},$$

donde

$$\begin{aligned} \mathbf{I}_{\beta_0^*, \eta_\xi} &= -\frac{4}{\sigma_{\epsilon^*}^2} \left(\sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^* \mathbf{x}_{i,0}^\top \right) \Sigma_{\mathbf{X}\mathbf{X}}^{-1}, \\ \mathbf{I}_{\beta_0^*, \mathbf{v}(\Sigma_{\xi\xi})} &= -\frac{2}{\sigma_{\epsilon^*}^2} \left(\sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^* (\mathbf{x}_{i,0}^\top \otimes \mathbf{x}_{i,0}^\top) \right) (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D}, \\ \mathbf{I}_{\beta^*, \eta_\xi} &= -\frac{4}{\sigma_{\epsilon^*}^2} \left(\sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^* \mathbf{x}_i \mathbf{x}_{i,0}^\top \right) \Sigma_{\mathbf{X}\mathbf{X}}^{-1}, \\ \mathbf{I}_{\beta^*, \mathbf{v}(\Sigma_{\xi\xi})} &= -\frac{2}{\sigma_{\epsilon^*}^2} \left(\sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^* \mathbf{x}_i (\mathbf{x}_{i,0}^\top \otimes \mathbf{x}_{i,0}^\top) \right) (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D}, \\ \mathbf{I}_{\sigma_{\epsilon^*}^2, \eta_\xi} &= -\frac{2}{(\sigma_{\epsilon^*}^2)^2} \left(\sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^{*2} \mathbf{x}_{i,0}^\top \right) \Sigma_{\mathbf{X}\mathbf{X}}^{-1}, \\ \mathbf{I}_{\sigma_{\epsilon^*}^2, \mathbf{v}(\Sigma_{\xi\xi})} &= -\frac{1}{(\sigma_{\epsilon^*}^2)^2} \left(\sum_{i=1}^n W_g^{(v_i, v_i)}(v_i) \epsilon_i^{*2} (\mathbf{x}_{i,0}^\top \otimes \mathbf{x}_{i,0}^\top) \right) (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D}, \end{aligned}$$

Además,

$$\mathbf{I}_{\theta_1^*, \boldsymbol{\nu}} = \begin{bmatrix} \mathbf{I}_{\beta_0^*, \boldsymbol{\nu}} \\ \mathbf{I}_{\beta^*, \boldsymbol{\nu}} \\ \mathbf{I}_{\sigma_{\epsilon^*}^2, \boldsymbol{\nu}} \end{bmatrix} = \begin{bmatrix} (2/\sigma_{\epsilon^*}^2) \sum_{i=1}^n (\mathbf{W}_g^{(\boldsymbol{\nu}, v_i)}(v_i))^\top \epsilon_i^* \\ (2/\sigma_{\epsilon^*}^2) \sum_{i=1}^n (\mathbf{W}_g^{(\boldsymbol{\nu}, v_i)}(v_i))^\top \epsilon_i^* \mathbf{x}_i \\ (\sigma_{\epsilon^*}^2)^{-2} \sum_{i=1}^n (\mathbf{W}_g^{(\boldsymbol{\nu}, v_i)}(v_i))^\top \epsilon_i^{*2} \end{bmatrix}.$$

Finalmente,

$$\mathbf{I}_{\theta_2^*, \nu} = \begin{bmatrix} \mathbf{I}_{\eta_\xi, \nu} \\ \mathbf{I}_{v(\Sigma_{\xi\xi}), \nu} \end{bmatrix} = \begin{bmatrix} 2\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \sum_{i=1}^n \mathbf{x}_{i,0} (\mathbf{W}_g^{(\nu, v_i)}(v_i))^\top \\ \mathbf{D}^\top (\Sigma_{\mathbf{X}\mathbf{X}}^{-1} \otimes \Sigma_{\mathbf{X}\mathbf{X}}^{-1}) \sum_{i=1}^n (\mathbf{x}_{i,0} \otimes \mathbf{x}_{i,0}) (\mathbf{W}_g^{(\nu, v_i)}(v_i))^\top \end{bmatrix},$$

donde

$$\mathbf{W}_g^{(\nu, v_i)}(v_i) = \frac{\partial \mathbf{W}_g^{(\nu)}(v_i)}{\partial v_i}.$$

3.7.2. Matriz de información esperada

En esta sección, se dan expresiones simples para calcular los elementos de la matriz de información esperada, la cual depende de algunos momentos univariados que pueden ser determinados numéricamente.

De (3.24) es fácil obtener la matriz de información esperada (Fisher) para el modelo estructural elíptico. En efecto, de (3.24) la matriz de información esperada de $(\boldsymbol{\theta}^*, \nu)$ es dada por

$$\mathbf{F}(\boldsymbol{\theta}^*, \nu) = \mathbb{E}[\mathbf{I}(\boldsymbol{\theta}^*, \nu)] = \begin{bmatrix} \mathbf{F}_{\theta_1^*, \theta_1^*} & \mathbf{F}_{\theta_1^*, \theta_2^*} & \mathbf{F}_{\theta_1^*, \nu} \\ & \mathbf{F}_{\theta_2^*, \theta_2^*} & \mathbf{F}_{\theta_2^*, \nu} \\ & & \mathbf{F}_{\nu, \nu} \end{bmatrix}. \quad (3.27)$$

Definamos el vector aleatorio

$$\mathbf{T}_i = \left[\frac{\epsilon_i^*}{\sigma_{\epsilon^*}}, (\mathbf{X}_i - \boldsymbol{\eta}_\xi)^\top \Sigma_{\mathbf{X}\mathbf{X}}^{-1/2} \right]^\top,$$

de dimensión $(p+1) \times 1$, para el cual de (3.9), se tiene que $\mathbf{T}_i \stackrel{ind.}{\sim} \text{El}_{p+1}(\mathbf{0}, \mathbf{I}_{p+1}; g)$. Considerando la representación estocástica de las distribuciones elípticas (sección 2.3), tenemos que

$$\mathbf{U}_i^{(p+1)} = \frac{\mathbf{T}_i}{\|\mathbf{T}_i\|}, \quad (3.28)$$

como se define en el capítulo 2, $\mathbf{U}_i^{(p+1)}$ es un vector aleatorio $(p+1) \times 1$ distribuido uniformemente en \mathcal{S}^{p+1} y es independiente de $R_i = \|\mathbf{T}_i\|$, una variable aleatoria no-negativa. Los cálculos de los momentos requeridos pueden ser simplificados considerando (3.28) y la independencia de \mathbf{U}_i y R_i . Por lo tanto, de (3.25) y los resultados en Arellano-Valle (1994) y Arellano-Valle (2010), obtenemos el siguiente resultado para la matriz de información $\mathbf{F}(\boldsymbol{\theta}^*, \nu)$.

Proposición 3.2 (Matriz de información del modelo estructural elíptico). *Consideremos el modelo estructural elíptico dado en (3.5), entonces la matriz de información esperada para $\boldsymbol{\theta}^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2, \boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top)^\top$ y $\boldsymbol{\nu}$ es dada en (3.27), donde*

$$\mathbf{F}_{\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_1^*} = \mathbb{E}[\mathbf{I}_{\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_1^*}] = \begin{bmatrix} F_{\beta_0^*, \beta_0^*} & \mathbf{F}_{\beta_0^*, \boldsymbol{\beta}^*} & 0 \\ & \mathbf{F}_{\boldsymbol{\beta}^*, \boldsymbol{\beta}^*} & \mathbf{0} \\ & & F_{\sigma_{\epsilon^*}^2, \sigma_{\epsilon^*}^2} \end{bmatrix},$$

con

$$\begin{aligned} F_{\beta_0^*, \beta_0^*} &= -\frac{2}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n \left(\mathbb{E}[W_g^{(v_i)}(v_i)] + \frac{2}{p+1} \mathbb{E}[W_g^{(v_i, v_i)}(v_i)v_i] \right), \\ \mathbf{F}_{\beta_0^*, \boldsymbol{\beta}^*} &= -\frac{2}{\sigma_{\epsilon^*}^2} \boldsymbol{\eta}_\xi^\top \sum_{i=1}^n \left(\mathbb{E}[W_g^{(v_i)}(v_i)] + \frac{2}{p+1} \mathbb{E}[W_g^{(v_i, v_i)}(v_i)v_i] \right), \\ \mathbf{F}_{\boldsymbol{\beta}^*, \boldsymbol{\beta}^*} &= -\frac{2}{\sigma_{\epsilon^*}^2} \boldsymbol{\eta}_\xi \boldsymbol{\eta}_\xi^\top \sum_{i=1}^n \left(\mathbb{E}[W_g^{(v_i)}(v_i)] + \frac{2}{p+1} \mathbb{E}[W_g^{(v_i, v_i)}(v_i)v_i] \right) \\ &\quad - \frac{2}{(p+1)\sigma_{\epsilon^*}^2} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \sum_{i=1}^n \left(\mathbb{E}[W_g^{(v_i)}(v_i)v_i] + \frac{2}{p+3} \mathbb{E}[W_g^{(v_i, v_i)}(v_i)v_i^2] \right), \\ F_{\sigma_{\epsilon^*}^2, \sigma_{\epsilon^*}^2} &= -\frac{n}{2(\sigma_{\epsilon^*}^2)^2} - \frac{1}{(p+1)(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n \left(2\mathbb{E}[W_g^{(v_i)}(v_i)v_i] - \frac{1}{p+3} \mathbb{E}[W_g^{(v_i, v_i)}(v_i)v_i^2] \right), \end{aligned}$$

donde

$$W_g^{(v_i)}(v_i) = \frac{\partial \log g(v_i)}{\partial v_i} \quad y \quad W_g^{(v_i, v_i)}(v_i) = \frac{\partial W_g^{(v_i)}(v_i)}{\partial v_i}.$$

Además

$$\mathbf{F}_{\boldsymbol{\theta}_2^*, \boldsymbol{\theta}_2^*} = \mathbb{E}[\mathbf{I}_{\boldsymbol{\theta}_2^*, \boldsymbol{\theta}_2^*}] = \begin{bmatrix} \mathbf{F}_{\boldsymbol{\eta}_\xi, \boldsymbol{\eta}_\xi} & \mathbf{0} \\ & \mathbf{F}_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}), \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})} \end{bmatrix},$$

con

$$\begin{aligned} \mathbf{F}_{\boldsymbol{\eta}_\xi, \boldsymbol{\eta}_\xi} &= -2\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \sum_{i=1}^n \left(\mathbb{E}[W_g^{(v_i)}(v_i)] + \frac{2}{p+1} \mathbb{E}[W_g^{(v_i, v_i)}(v_i)v_i] \right), \\ \mathbf{F}_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}), \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})} &= -\frac{n}{2} \mathbf{D}^\top (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D} - \frac{2}{p+1} \mathbf{D}^\top (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D} \sum_{i=1}^n \mathbb{E}[W_g^{(v_i)}(v_i)v_i] \\ &\quad - \frac{2}{(p+1)(p+3)} \mathbf{D}^\top (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D} \sum_{i=1}^n \mathbb{E}[W_g^{(v_i, v_i)}(v_i)v_i^2] \\ &\quad - \frac{1}{(p+1)(p+3)} \mathbf{D}^\top \text{vec}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \text{vec}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1})^\top \mathbf{D} \sum_{i=1}^n \mathbb{E}[W_g^{(v_i, v_i)}(v_i)v_i^2]. \end{aligned}$$

$$Y, \mathbf{F}_{\nu, \nu} = n\mathbf{C}_{p+1}''(\nu) + \sum_{i=1}^n \mathbb{E}[\mathbf{W}_g^{(\nu, \nu)}(v_i)], \mathbf{F}_{\theta_1^*, \nu} = \mathbf{0} \text{ y } \mathbf{F}_{\theta_2^*, \nu} = \mathbf{0}.$$

Notemos que, de la Proposición 3.2, la matriz de información esperada depende de momentos univariados los cuales dependiendo de la elección de la función generadora de densidades pueden ser determinados explícitamente o pueden ser calculados numéricamente.

3.7.3. Parametrización directa

Considerando la reparametrización dada en (3.6), (3.7) y (3.10), notemos que la matriz de información esperada en (3.27) corresponde a la matriz de información esperada del modelo definido en (3.5).

Por otro lado, el vector de parámetros para el modelo con error de medida definido en (3.1) es dado por $(\boldsymbol{\theta}^\top, \boldsymbol{\nu}^\top)^\top$. Así, el vector de score en la parametrización directa es dado por

$$\mathbf{S}(\boldsymbol{\theta}, \boldsymbol{\nu}) = \mathbf{J}(\boldsymbol{\theta}, \boldsymbol{\nu})\mathbf{S}(\boldsymbol{\theta}^*(\boldsymbol{\theta}), \boldsymbol{\nu}), \quad (3.29)$$

donde $\mathbf{J}(\boldsymbol{\theta}, \boldsymbol{\nu})$ es la matriz Jacobiana y $\mathbf{S}(\boldsymbol{\theta}^*(\boldsymbol{\theta}), \boldsymbol{\nu})$ es el vector de score dado en la Proposición 3.1 y la matriz de información esperada es (Lehmann & Casella (1998))

$$\mathbf{F}(\boldsymbol{\theta}, \boldsymbol{\nu}) = \mathbf{J}(\boldsymbol{\theta}, \boldsymbol{\nu})\mathbf{F}(\boldsymbol{\theta}^*(\boldsymbol{\theta}), \boldsymbol{\nu})\mathbf{J}^\top(\boldsymbol{\theta}, \boldsymbol{\nu}), \quad (3.30)$$

donde $\mathbf{F}(\boldsymbol{\theta}^*(\boldsymbol{\theta}), \boldsymbol{\nu})$ es la matriz de información esperada dada en (3.27). Entonces,

$$d(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = \mathbf{J}(\boldsymbol{\theta}, \boldsymbol{\nu})d(\boldsymbol{\theta}, \boldsymbol{\nu}),$$

con $d(\boldsymbol{\theta}, \boldsymbol{\nu}) = (d\beta_0, (d\boldsymbol{\beta})^\top, d\sigma_\epsilon^2, (d\boldsymbol{\eta}_\xi)^\top, (d\boldsymbol{\nu}(\boldsymbol{\Sigma}_{\xi\xi}))^\top, (d\boldsymbol{\nu})^\top)^\top$. En este caso, la matriz Jacobiana es

$$\mathbf{J}(\boldsymbol{\theta}, \boldsymbol{\nu}) = \begin{bmatrix} \mathbf{J}_{11}^{\boldsymbol{\theta}, \boldsymbol{\nu}} & \mathbf{J}_{12}^{\boldsymbol{\theta}, \boldsymbol{\nu}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p(p+3)/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{bmatrix}, \quad (3.31)$$

donde

$$\mathbf{J}_{11}^{\boldsymbol{\theta}, \boldsymbol{\nu}} = \begin{bmatrix} 1 & \boldsymbol{\eta}_\xi^\top & 0 \\ \mathbf{0} & \boldsymbol{\Lambda} & \mathbf{0} \\ 0 & 2\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda} & 1 \end{bmatrix}, \quad (3.32)$$

y

$$\mathbf{J}_{12}^{\boldsymbol{\theta}, \boldsymbol{\nu}} = \begin{bmatrix} \boldsymbol{\beta}^\top (\mathbf{I}_p - \boldsymbol{\Lambda})^\top & (\boldsymbol{\eta}_\xi^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \otimes \boldsymbol{\beta}^\top (\mathbf{I}_p - \boldsymbol{\Lambda})^\top) \mathbf{D} \\ \mathbf{0} & (\boldsymbol{\beta}^\top (\mathbf{I}_p - \boldsymbol{\Lambda})^\top \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D} \\ \mathbf{0} & (\boldsymbol{\beta}^\top (\mathbf{I}_p - \boldsymbol{\Lambda})^\top \otimes \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D} \end{bmatrix}. \quad (3.33)$$

3.8. El modelo estructural normal

El modelo estructural normal, es dado por las ecuaciones (3.1), con

$$\mathbf{r}_i = \begin{bmatrix} \epsilon_i \\ \delta_i \\ \xi_i \end{bmatrix} \stackrel{iid}{\sim} N_{2p+1} \left(\begin{bmatrix} 0 \\ \mathbf{0} \\ \boldsymbol{\eta}_\xi \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\delta\delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{\xi\xi} \end{bmatrix} \right),$$

para $i = 1, 2, \dots, n$. La función generadora de densidades es

$$g(v_i) = \exp\left(-\frac{v_i}{2}\right).$$

Parametrización tipo Gleser:

El modelo reparametrizado es dado en (3.5), es decir, $Y_i = \beta_0^* + \boldsymbol{\beta}^{*\top} \mathbf{X}_i + \epsilon_i^*$, donde

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{X}_i \\ Y_i \end{bmatrix} \stackrel{iid}{\sim} N_{p+1} \left(\begin{bmatrix} \boldsymbol{\eta}_\xi \\ \beta_0^* + \boldsymbol{\beta}^{*\top} \boldsymbol{\eta}_\xi \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \boldsymbol{\beta}^* \\ \boldsymbol{\beta}^{*\top} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} & \sigma_{\epsilon^*}^2 + \boldsymbol{\beta}^{*\top} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \boldsymbol{\beta}^* \end{bmatrix} \right),$$

para $i = 1, 2, \dots, n$. Bajo las condiciones del modelo reparametrizado, tenemos que $v_i = (\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi) + \epsilon_i^{*2} / \sigma_{\epsilon^*}^2$ y $\epsilon_i^* = y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i$. El error aleatorio ϵ_i^* se distribuye normal con media 0 y varianza $\sigma_{\epsilon^*}^2$, con densidad

$$f_{\epsilon^*}(\epsilon_i^*; \beta_0^*, \boldsymbol{\beta}^*, \sigma_{\epsilon^*}^2) = \frac{1}{(2\pi\sigma_{\epsilon^*}^2)^{1/2}} e^{-\epsilon_i^{*2}/2\sigma_{\epsilon^*}^2}, \quad \epsilon_i^* \in \mathbb{R},$$

el cual es independiente de $\mathbf{X}_i \sim N_p(\boldsymbol{\eta}_\xi, \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}})$, y

$$f_{\mathbf{X}}(\mathbf{x}_i; \boldsymbol{\eta}_\xi, \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi)}, \quad \mathbf{x}_i \in \mathbb{R}^p,$$

y en este caso $v_i \sim \chi_{p+1}^2$. Además, de (3.12) la densidad de \mathbf{Z}_i , $i = 1, 2, \dots, n$, puede ser representada como

$$f_{\mathbf{Z}_i}(\mathbf{z}_i; \boldsymbol{\theta}^*) = f_{\mathbf{X}}(\mathbf{x}_i; \boldsymbol{\eta}_\xi, \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}) f_{\epsilon^*}(\epsilon_i^*; \beta_0^*, \boldsymbol{\beta}^*, \sigma_{\epsilon^*}^2).$$

Esto simplifica la forma de la función de verosimilitud y por tanto la obtención de las estimaciones, como se discute más adelante. Notemos que, \mathbf{Z}_i , $i = 1, 2, \dots, n$, es normalmente distribuido, entonces la distribución conjunta de $\mathbf{Z} = (\mathbf{Z}_1^\top, \mathbf{Z}_2^\top, \dots, \mathbf{Z}_n^\top)^\top$, también es normalmente distribuida (ver Cheng & Tsai (2004)).

Representación jerárquica:

De la sección 3.6, los modelos estructurales elípticos en términos de la parametrización de Gleser permite una simple representación jerárquica dada por

$$Y_i | \mathbf{X}_i = \mathbf{x}_i \stackrel{ind.}{\sim} N_1(\beta_0^* + \boldsymbol{\beta}^{*\top} \mathbf{x}_i, \sigma_{\epsilon^*}^2),$$

$$\mathbf{X}_i \stackrel{iid.}{\sim} N_p(\boldsymbol{\eta}_\xi, \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}).$$

Definiendo $\boldsymbol{\theta}_1^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2)^\top$ y $\boldsymbol{\theta}_2^* = (\boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top)^\top$, se tiene que $\boldsymbol{\theta}_1^*$ y $\boldsymbol{\theta}_2^*$ están en variación libre, y por tanto \mathbf{X}_i es débilmente exógeno respecto a $\boldsymbol{\theta}_1^*$ (ver Engle et al. (1983)); esto nos permite trabajar con el modelo reducido (ver sección 3.6).

Además, tenemos que el modelo puede ser especificado en términos de los primeros dos momentos de $f_{Y_i | \mathbf{X}_i = \mathbf{x}_i}(y_i; \boldsymbol{\theta}_1^*)$ como sigue

$$E[Y_i | \mathbf{X}_i = \mathbf{x}_i] = \beta_0^* + \boldsymbol{\beta}^{*\top} \mathbf{x}_i \quad y \quad V(Y_i | \mathbf{X}_i = \mathbf{x}_i) = \sigma_{\epsilon^*}^2,$$

donde las varianzas condicionales son homoscedásticas.

Estimación de máxima verosimilitud:

Asumiendo que $\boldsymbol{\Sigma}_{\delta\delta}$ es conocido, de la Proposición 3.1, tenemos que

$$S_{\beta_0^*} = \frac{1}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n (y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i),$$

$$\mathbf{S}_{\boldsymbol{\beta}^*} = \frac{1}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n (y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i) \mathbf{x}_i,$$

$$S_{\sigma_{\epsilon^*}^2} = -\frac{n}{2\sigma_{\epsilon^*}^2} + \frac{1}{2(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n (y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i)^2,$$

$$\mathbf{S}_{\boldsymbol{\eta}_\xi} = \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\eta}_\xi),$$

$$\mathbf{S}_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})} = -\frac{n}{2} \mathbf{D}^\top \text{vec}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) + \frac{1}{2} \mathbf{D}^\top (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \sum_{i=1}^n ((\mathbf{x}_i - \boldsymbol{\eta}_\xi) \otimes (\mathbf{x}_i - \boldsymbol{\eta}_\xi)).$$

En este caso, los estimadores de máxima verosimilitud de β_0^* y $\boldsymbol{\beta}^*$, corresponden a los

estimadores mínimos cuadrados ordinarios

$$\hat{\beta}_0^* = \bar{y} - \mathbf{s}_{yx}^\top \mathbf{S}_{xx}^{-1} \bar{\mathbf{x}}, \quad (3.34)$$

$$\hat{\boldsymbol{\beta}}^* = \mathbf{S}_{xx}^{-1} \mathbf{s}_{yx}, \quad (3.35)$$

donde $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_p)^\top$, con $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$, para $j = 1, 2, \dots, p$. Además, el estimador de máxima verosimilitud de $\sigma_{\epsilon^*}^2$, es dado por

$$\hat{\sigma}_{\epsilon^*}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_i^{*2} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_0^* - \hat{\boldsymbol{\beta}}^{*\top} \mathbf{x}_i)^2, \quad (3.36)$$

También se tiene que $\hat{\boldsymbol{\eta}}_\xi = \bar{\mathbf{x}}$ y $\mathbf{v}(\hat{\boldsymbol{\Sigma}}_{\xi\xi}) = \mathbf{v}(\mathbf{S}_{xx}) - \mathbf{v}(\boldsymbol{\Sigma}_{\delta\delta})$, donde \mathbf{s}_{yx} y \mathbf{S}_{xx} son las covarianzas muestrales y $\mathbf{s}_{yx} = (s_{y1}, \dots, s_{yp})^\top$ con $s_{yj} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(x_{ij} - \bar{x}_j)$, para $j = 1, 2, \dots, p$, y $\mathbf{S}_{xx} = [s_{kj}]_{kj}$, con $s_{kj} = \frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)(x_{ij} - \bar{x}_j)$, para $j = 1, 2, \dots, p$ y $k = 1, 2, \dots, p$.

Por la propiedad de invarianza se tiene que los estimadores de máxima verosimilitud de β_0 y $\boldsymbol{\beta}$ son

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\boldsymbol{\beta}}^\top \bar{\mathbf{x}} = \bar{y} - \mathbf{s}_{xy}^\top (\mathbf{S}_{xx} - \boldsymbol{\Sigma}_{\delta\delta})^{-1} \bar{\mathbf{x}}, \\ \hat{\boldsymbol{\beta}} &= \hat{\boldsymbol{\Lambda}}^{-1} \hat{\boldsymbol{\beta}}^* = (\mathbf{S}_{xx} - \boldsymbol{\Sigma}_{\delta\delta})^{-1} \mathbf{s}_{xy}, \end{aligned}$$

donde $\hat{\boldsymbol{\Lambda}} = \mathbf{S}_{xx}^{-1} (\mathbf{S}_{xx} - \boldsymbol{\Sigma}_{\delta\delta})$. Además

$$\hat{\sigma}_\epsilon^2 = s_{yy} - \mathbf{s}_{yx}^\top (\mathbf{S}_{xx} - \boldsymbol{\Sigma}_{\delta\delta})^{-1} \mathbf{s}_{yx},$$

donde $s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 / n$.

Por otro lado, si consideramos $\boldsymbol{\Lambda}$ conocido, los estimadores de máxima verosimilitud β_0^* , $\boldsymbol{\beta}^*$, $\sigma_{\epsilon^*}^2$, y $\boldsymbol{\eta}_\xi$ son dadas por las ecuaciones (3.34)-(3.36) y $\hat{\boldsymbol{\eta}}_\xi = \bar{\mathbf{x}}$, respectivamente. Pero, se debe tener en cuenta que $\hat{\boldsymbol{\Sigma}}_{\xi\xi} = \mathbf{S}_{xx} \boldsymbol{\Lambda}$ y $\hat{\boldsymbol{\Sigma}}_{\delta\delta} = \mathbf{S}_{xx} (\mathbf{I}_p - \boldsymbol{\Lambda})$.

Matriz de información:

Sea $\boldsymbol{\theta}_1^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2)^\top$ y $\boldsymbol{\theta}_2^* = (\boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi}))^\top$, entonces de (3.24) y (3.27), podemos obtener la matriz de información observada y esperada.

Notemos que, los parámetros $\boldsymbol{\theta}_1^*$ y $\boldsymbol{\theta}_2^*$ son ortogonales, donde

$$\mathbf{F}(\boldsymbol{\theta}^*) = \begin{bmatrix} \frac{n}{\sigma_{\epsilon^*}^2} & \frac{n}{\sigma_{\epsilon^*}^2} \boldsymbol{\eta}_{\xi}^{\top} & 0 & \mathbf{0} & 0 \\ \frac{n}{\sigma_{\epsilon^*}^2} (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} + \boldsymbol{\eta}_{\xi} \boldsymbol{\eta}_{\xi}^{\top}) & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \\ & & \frac{n}{2\sigma_{\epsilon^*}^4} & \mathbf{0} & 0 \\ & & & n\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} & 0 \\ & & & & \frac{n}{2} \mathbf{D}^{\top} (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \mathbf{D} \end{bmatrix}.$$

De (3.31), la matriz Jacobiana es definida por

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{J}_{11}^{\boldsymbol{\theta}} & \mathbf{J}_{12}^{\boldsymbol{\theta}} \\ \mathbf{0} & \mathbf{I}_{p(p+3)/2} \end{bmatrix}$$

donde $\mathbf{J}_{11}^{\boldsymbol{\theta}}$ y $\mathbf{J}_{12}^{\boldsymbol{\theta}}$ son definidos en (3.32) y (3.33), respectivamente. Así, la matriz de información esperada para la parametrización directa, es $\mathbf{F}(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta}) \mathbf{F}(\boldsymbol{\theta}^*(\boldsymbol{\theta})) \mathbf{J}^{\top}(\boldsymbol{\theta})$, donde $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta}^{\top}, \sigma_{\epsilon}^2, \boldsymbol{\eta}_{\xi}^{\top}, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^{\top})^{\top}$.

3.9. El modelo estructural t

La distribución t es otro miembro de la familia de distribuciones elípticas. Consecuentemente, sus propiedades pueden ser derivadas de la teoría presentada en el capítulo 2. Esta distribución a sido aceptada como una alternativa robusta al modelo normal y esto motiva una importante teoría alrededor de la distribución t . En este sentido, algunos resultados son dados en Fang et al. (1990), Arellano-Valle (1994), Arellano-Valle & Bolfarine (1995) y Vidal & Castro (2010).

En esta sección, consideramos el modelo (3.1), donde

$$\mathbf{r}_i = \begin{bmatrix} \epsilon_i \\ \boldsymbol{\delta}_i \\ \boldsymbol{\xi}_i \end{bmatrix} \stackrel{ind.}{\sim} t_{2p+1} \left(\begin{bmatrix} 0 \\ \mathbf{0} \\ \boldsymbol{\eta}_{\xi} \end{bmatrix}, \begin{bmatrix} \sigma_{\epsilon}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\delta\delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{\xi\xi} \end{bmatrix}, \nu \right),$$

para $i = 1, 2, \dots, n$ con $\nu > 0$ grados de libertad.

Por lo tanto la función generadora de densidades es dada por

$$g(v_i) = \left(1 + \frac{v_i}{\nu}\right)^{-\frac{\nu+p+1}{2}},$$

donde $v_i = (\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi) + \epsilon_i^{*2} / \sigma_{\epsilon^*}^2$.

Parametrización tipo Gleser:

Del modelo reparametrizado dado en (3.5), tenemos que

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{X}_i \\ Y_i \end{bmatrix} \stackrel{ind.}{\sim} t_{p+1} \left(\begin{bmatrix} \boldsymbol{\eta}_\xi \\ \beta_0^* + \boldsymbol{\beta}^{*\top} \boldsymbol{\eta}_\xi \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \boldsymbol{\beta}^* \\ \boldsymbol{\beta}^{*\top} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} & \sigma_{\epsilon^*}^2 + \boldsymbol{\beta}^{*\top} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \boldsymbol{\beta}^* \end{bmatrix}, \nu \right),$$

para $i = 1, \dots, n$ y de (3.9)

$$\epsilon_i^* \sim t(0, \sigma_{\epsilon^*}^2, \nu) \quad \text{y} \quad \mathbf{X}_i \sim t_p(\boldsymbol{\eta}_\xi, \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}, \nu),$$

La densidad de \mathbf{Z}_i , para $i = 1, 2, \dots, n$, es

$$f_{\mathbf{Z}_i}(\mathbf{z}_i; \boldsymbol{\theta}^*, \nu) = \frac{C_{p+1}(\nu)}{(\sigma_{\epsilon^*}^2)^{1/2} |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}|^{1/2}} \left(1 + \frac{1}{\nu} v_i \right)^{-\frac{\nu+p+1}{2}},$$

y constante de normalización

$$C_{p+1}(\nu) = \frac{\Gamma((\nu + p + 1)/2)}{(\pi\nu)^{(p+1)/2} \Gamma(\nu/2)}.$$

La función de verosimilitud para $\boldsymbol{\theta}^*$ y ν , es dada por

$$L(\boldsymbol{\theta}^*, \nu) = \frac{C_{p+1}^n(\nu)}{(\sigma_{\epsilon^*}^2)^{n/2} |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}|^{n/2}} \prod_{i=1}^n \left(1 + \frac{1}{\nu} v_i \right)^{-\frac{\nu+p+1}{2}}, \quad (3.37)$$

En este caso, la distribución conjunta de $\mathbf{Z} = (\mathbf{Z}_1^\top, \mathbf{Z}_2^\top, \dots, \mathbf{Z}_n^\top)$ no se distribuye t .

Representación jerárquica:

La media y varianza condicional bajo el modelo reparametrizado son

$$\mathbb{E}[Y_i | \mathbf{X}_i = \mathbf{x}_i] = \beta_0^* + \boldsymbol{\beta}^* \mathbf{x}_i \quad \text{y} \quad \mathbb{V}(Y_i | \mathbf{X}_i = \mathbf{x}_i) = \frac{\nu + q(\mathbf{x}_i)}{\nu + p - 2} \sigma_{\epsilon^*}^2,$$

donde $q(\mathbf{x}_i) = (\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi)$, entonces el modelo definido por $Y_i | \mathbf{X}_i = \mathbf{x}_i$ es heteroscedástico.

Estimación de máxima verosimilitud:

A diferencia del modelo normal tratado en la sección anterior, la distribución conjunta de n marginales independientes t , no se distribuye t . En este caso, tenemos que

$$W_g^{(v_i)}(v_i) = -\frac{\omega_i}{2},$$

donde

$$\omega_i = \frac{\nu + p + 1}{\nu + v_i}.$$

De la Proposición 3.1, los elementos del vector de score $\mathbf{S}(\boldsymbol{\theta}^*, \nu)$ son dados por

$$\begin{aligned} S_{\beta_0^*} &= \frac{1}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n \omega_i (y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i), \\ \mathbf{S}_{\boldsymbol{\beta}^*} &= \frac{1}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n \omega_i (y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i) \mathbf{x}_i, \\ S_{\sigma_{\epsilon^*}^2} &= -\frac{n}{2\sigma_{\epsilon^*}^2} + \frac{1}{2(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n \omega_i (y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i)^2, \\ \mathbf{S}_{\boldsymbol{\eta}_\xi} &= \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \sum_{i=1}^n \omega_i (\mathbf{x}_i - \boldsymbol{\eta}_\xi), \\ \mathbf{S}_{\nu(\boldsymbol{\Sigma}_{\xi\xi})} &= -\frac{n}{2} \mathbf{D}^\top \text{vec}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) + \frac{1}{2} \mathbf{D}^\top (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \sum_{i=1}^n \omega_i ((\mathbf{x}_i - \boldsymbol{\eta}_\xi) \otimes (\mathbf{x}_i - \boldsymbol{\eta}_\xi)), \end{aligned}$$

y de (3.18)

$$S_\nu = n \left(\psi \left(\frac{\nu + p + 1}{2} \right) - \psi \left(\frac{\nu}{2} \right) - \frac{p + 1}{2\nu} \right) + \frac{1}{2\nu} \sum_{i=1}^n v_i \omega_i - \frac{1}{2} \sum_{i=1}^n \log \left(1 + \frac{v_i}{\nu} \right),$$

donde $\psi(a) = \Gamma'(a)/\Gamma(a)$, es la función digamma y

$$\Gamma'(a) = \frac{\partial}{\partial a} \Gamma(a) = \int_0^\infty t^{a-1} e^{-t} \log t \, dt.$$

En este caso las ecuaciones de log-verosimilitud pueden ser resueltas iterativamente usando, por ejemplo, el algoritmo EM (Dempster et al. (1977)).

Modelo de mezcla de escala normal:

La implementación del algoritmo EM para obtener los estimadores de máxima verosimilitud de los parámetros del modelo definido en (3.37) es facilitado por el hecho

de que este modelo puede ser especificado como una mezcla de escala del modelo estructural normal.

Definamos las variables aleatorias $T_i = \nu/\chi_\nu^2$, para $i = 1, 2, \dots, n$, independientes e idénticamente distribuidas con distribución Gamma inversa con parámetros $(\nu/2, \nu/2)$, denotando esto como $T_i \stackrel{iid}{\sim} \text{IG}(\nu/2, \nu/2)$, con densidad

$$f_{T_i}(t_i; \nu) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} t_i^{-(\nu+2)/2} e^{-\nu/2t_i}, \quad t_i > 0,$$

y consideremos el vector aleatorio $\mathbf{W}_i \sim N_{p+1}(\mathbf{0}, \mathbf{I}_{p+1})$ independientes, para $i = 1, 2, \dots, n$. Entonces, de (2.9) es fácil ver que $\mathbf{Z}_i = (\mathbf{X}_i^\top, Y_i)^\top$ es representado estocásticamente como

$$\mathbf{Z}_i = \boldsymbol{\mu}^* + T_i^{1/2} \boldsymbol{\Sigma}^{*1/2} \mathbf{W}_i. \quad (3.38)$$

Considerando la familia de normales compuestas (Andrews & Mallows (1974))

$$(\mathbf{Z}_i | T_i = t_i) \sim N_{p+1}(\boldsymbol{\mu}^*, t_i \boldsymbol{\Sigma}^*), \quad (3.39)$$

y $T_i \stackrel{iid}{\sim} \text{IG}(\nu/2, \nu/2)$, para $i = 1, 2, \dots, n$.

Andrews & Mallows (1974) muestra que $\mathbf{Z}_i \sim t_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, \nu)$ si y sólo si \mathbf{Z}_i tiene representación estocástica (3.38), donde $T_i \sim \text{IG}(\nu/2, \nu/2)$ y $\mathbf{W}_i \sim N_{p+1}(\mathbf{0}, \mathbf{I}_{p+1})$ independientes. Arellano-Valle & Bolfarine (1995) también presentan caracterizaciones de la versión generalizada de la distribución t , es decir, considerando que $T_i \sim \text{IG}(\nu/2, \lambda/2)$ (en nuestro caso, consideramos $\nu = \lambda$) en términos de las distribuciones marginales, distribuciones condicionales, formas cuadráticas y en la clase de normales compuestas.

Algoritmo EM:

El algoritmo EM aumenta los datos $\mathbf{Z} = (\mathbf{Z}_1^\top, \mathbf{Z}_2^\top, \dots, \mathbf{Z}_n^\top)^\top$ con datos adicionales, en este caso consideramos $\mathbf{T} = (T_1, T_2, \dots, T_n)^\top$, tal que los estimadores de verosimilitud de $(\boldsymbol{\theta}^{*\top}, \nu)^\top$, sean basados en la función de verosimilitud $\ell_C(\boldsymbol{\theta}^*, \nu)$ de los datos aumentados $(\mathbf{Z}^\top, \mathbf{T}^\top)^\top$.

Si $(\boldsymbol{\theta}^{*(l)\top}, \nu^{(l)})^\top$ es el valor de $(\boldsymbol{\theta}^{*\top}, \nu)^\top$ obtenido en la l -ésima iteración, entonces la $(l+1)$ -ésima iteración del algoritmo EM consiste de dos pasos: E y M. El paso E, determina el valor esperado de $\ell_C(\boldsymbol{\theta}^*, \nu)$ con respecto a la distribución condicional de \mathbf{T} dado $\mathbf{Z} = \mathbf{z}$ y $(\boldsymbol{\theta}^{*(l)\top}, \nu^{(l)})^\top$. El paso M, maximiza la función resultante respecto a $(\boldsymbol{\theta}^{*\top}, \nu)^\top$ y da una estimación $(\boldsymbol{\theta}^{*(l+1)\top}, \nu^{(l+1)})^\top$ de $(\boldsymbol{\theta}^{*\top}, \nu)^\top$.

Notemos que, la implementación del algoritmo EM para obtener los estimadores de

máxima verosimilitud es facilitada por el hecho de que el modelo puede ser especificado de la forma (3.39). Por lo tanto, de (3.39), la densidad conjunta de $(\mathbf{Z}_i^\top, T_i)^\top$ es dada por

$$f_{\mathbf{Z}_i, T_i}(\mathbf{z}_i, t_i; \boldsymbol{\theta}^*, \nu) = f_{\mathbf{Z}_i | T_i}(\mathbf{z}_i; \boldsymbol{\theta}^*) f_{T_i}(t_i; \nu). \quad (3.40)$$

Entonces, la función de log-verosimilitud de $(\mathbf{Z}^\top, \mathbf{T}^\top)^\top$ es

$$\begin{aligned} \ell_C(\boldsymbol{\theta}^*, \nu) = n & \left\{ \frac{\nu}{2} \log \left(\frac{\nu}{2} \right) - \frac{p+1}{2} \log(2\pi) - \frac{1}{2} \log \sigma_{\epsilon^*}^2 - \frac{1}{2} \log |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}| - \log \Gamma \left(\frac{\nu}{2} \right) \right\} \\ & - \frac{1}{2} (\nu + 3) \sum_{i=1}^n \log t_i - \frac{1}{2} \sum_{i=1}^n \frac{1}{t_i} (v_i + \nu). \end{aligned}$$

En el paso E calculamos

$$\ell_E(\boldsymbol{\theta}^*, \nu) = \mathbb{E}[\ell_C(\boldsymbol{\theta}^*, \nu) | \mathbf{Z} = \mathbf{z}, \boldsymbol{\theta}^{*(l)\top}, \nu^{(l)}],$$

sobre la distribución de $(\mathbf{T} | \mathbf{Z} = \mathbf{z})$. Notemos que,

$$\sum_{i=1}^n \frac{v_i + \nu}{T_i} = \text{tr}(\mathbf{D}^{-1}(\mathbf{T})(\mathbf{D}(\mathbf{v}) + \nu)),$$

donde $\mathbf{D}(\mathbf{T}) = \text{diag}\{T_1, T_2, \dots, T_n\}$ y $\mathbf{D}(\mathbf{v}) = \text{diag}\{v_1, v_2, \dots, v_n\}$. Por lo tanto

$$\begin{aligned} \mathbb{E}[\mathbf{D}^{-1}(\mathbf{T}) | \mathbf{Z} = \mathbf{z}] &= \text{diag}\{\mathbb{E}[T_1^{-1} | \mathbf{Z} = \mathbf{z}], \mathbb{E}[T_2^{-1} | \mathbf{Z} = \mathbf{z}], \dots, \mathbb{E}[T_n^{-1} | \mathbf{Z} = \mathbf{z}]\}, \\ &= \text{diag}\{\omega_1, \omega_2, \dots, \omega_n\}, \\ &= \mathbf{D}(\boldsymbol{\omega}), \end{aligned}$$

entonces

$$\mathbb{E} \left[\sum_{i=1}^n \frac{1}{T_i} (v_i + \nu) \middle| \mathbf{Z} = \mathbf{z}, \boldsymbol{\theta}^{*(l)\top}, \nu^{(l)} \right] = \sum_{i=1}^n \omega_i^{(l)} (v_i + \nu).$$

Por otro lado, $\sum_{i=1}^n \log T_i = \log |\mathbf{D}(\mathbf{T})|$ entonces

$$\begin{aligned} \mathbb{E} [\log |\mathbf{D}(\mathbf{T})| | \mathbf{Z} = \mathbf{z}] &= \sum_{i=1}^n \mathbb{E} [\log T_i | \mathbf{Z}_i = \mathbf{z}_i], \\ &= - \sum_{i=1}^n \mathbb{E} [\log T_i^{-1} | \mathbf{Z}_i = \mathbf{z}_i], \\ &= - \sum_{i=1}^n \log \omega_i + n \log \left(\frac{\nu + p + 1}{2} \right) - n \psi \left(\frac{\nu + p + 1}{2} \right). \end{aligned}$$

Por lo tanto

$$\begin{aligned} \mathbb{E} \left[\sum_{i=1}^n \log T_i \middle| \mathbf{Z} = \mathbf{z}, \boldsymbol{\theta}^{*(l)\top}, \nu^{(l)} \right] &= - \sum_{i=1}^n \log \omega_i^{(l)} \\ &\quad + n \log \left(\frac{\nu^{(l)} + p + 1}{2} \right) - n\psi \left(\frac{\nu^{(l)} + p + 1}{2} \right), \end{aligned}$$

Entonces, en el paso E tenemos que

$$\begin{aligned} \ell_E(\boldsymbol{\theta}^*, \nu) &\propto n \left\{ \frac{\nu}{2} \log \left(\frac{\nu}{2} \right) - \frac{1}{2} \log \sigma_{\epsilon^*}^2 - \frac{1}{2} \log |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}| - \log \Gamma \left(\frac{\nu}{2} \right) \right\} \\ &\quad - \frac{1}{2}(\nu + 3) \left\{ - \sum_{i=1}^n \log \omega_i^{(l)} + n \log \left(\frac{\nu^{(l)} + p + 1}{2} \right) - n\psi \left(\frac{\nu^{(l)} + p + 1}{2} \right) \right\} \\ &\quad - \frac{1}{2} \sum_{i=1}^n \omega_i^{(l)} (v_i + \nu), \end{aligned}$$

donde

$$\omega_i^{(l)} = \frac{\nu^{(l)} + p + 1}{\nu^{(l)} + v_i^{(l)}}. \quad (3.41)$$

Ahora el algoritmo EM procede al paso M, donde $\ell_E(\boldsymbol{\theta}^*, \nu)$ es maximizado con respecto a $(\boldsymbol{\theta}^{*\top}, \nu)^\top$. Entonces, definiendo $\boldsymbol{\beta}_C^* = (\beta_0^*, \boldsymbol{\beta}^{*\top})^\top$, $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$, $\mathbf{X}_{i,C} = (1, \mathbf{X}_i^\top)^\top$, para $i = 1, 2, \dots, n$, y la matriz de diseño $\mathbf{X} = (\mathbf{X}_{1,C}, \mathbf{X}_{2,C}, \dots, \mathbf{X}_{n,C})^\top$, es fácil ver que

$$\begin{cases} \boldsymbol{\beta}_C^{*(l+1)} = (\mathbf{X}^\top \mathbf{D}(\boldsymbol{\omega}^{(l)}) \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{D}(\boldsymbol{\omega}^{(l)}) \mathbf{Y}, \\ \sigma_{\epsilon^*}^{2(l+1)} = \frac{1}{n} \boldsymbol{\epsilon}^{*\top(l+1)} \mathbf{D}(\boldsymbol{\omega}^{(l)}) \boldsymbol{\epsilon}^{*(l+1)}, \end{cases} \quad (3.42)$$

donde $\boldsymbol{\epsilon}^{*(l+1)} = \mathbf{Y} - \mathbf{X} \boldsymbol{\beta}_C^{*(l+1)}$ es el vector de residuos de la $(l + 1)$ -ésima iteración.

Además tenemos

$$\boldsymbol{\eta}_\xi^{(l+1)} = \frac{\sum_{i=1}^n \omega_i^{(l)} \mathbf{x}_i}{\sum_{i=1}^n \omega_i^{(l)}}, \quad (3.43)$$

$$\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^{(l+1)} = \frac{1}{n} \sum_{i=1}^n \omega_i^{(l)} \mathbf{v}(\mathbf{x}_{i,0} \mathbf{x}_{i,0}^\top) - \mathbf{v}(\boldsymbol{\Sigma}_{\delta\delta}). \quad (3.44)$$

Para los grados de libertad, se tiene que $\nu^{(l+1)}$ es la solución de la ecuación dada por

$$\log\left(\frac{\nu^{(l+1)}}{2}\right) - \psi\left(\frac{\nu^{(l+1)}}{2}\right) + 1 = \frac{1}{n} \sum_{i=1}^n (\omega_i^{(l)} - \log \omega_i^{(l)}) + \log\left(\frac{\nu^{(l)} + p + 1}{2}\right) - \psi\left(\frac{\nu^{(l)} + p + 1}{2}\right). \quad (3.45)$$

Notemos que la conveniencia del algoritmo EM es particularmente clara cuando ν es conocido, ya que en este caso el algoritmo se reduce a evaluar iterativamente los estimadores mínimos cuadrados ponderados dados por (3.42).

El valor inicial $(\boldsymbol{\theta}^{*(0)}, \nu^{(0)})$ puede ser elegido tomando $\mathbf{D}(\boldsymbol{\omega}^{(0)}) = \mathbf{I}_n$, el cual es equivalente a tomar como valores iniciales los estimadores de mínimos cuadrados ordinarios para $\boldsymbol{\beta}_C^{*(0)}$ y $\sigma_{\epsilon^*}^{2(0)}$. Para ν desconocido, es vez de resolver (3.45), un método alternativo, es usar algoritmo auxiliares tal como el algoritmo de scoring.

Matriz de información:

Respecto a la matriz de información del modelo estructural t , bajo la parametrización de Gleser, se tiene que de (3.27), esta matriz depende de las cantidades, $\mathbf{E}[W_g^{(v_i)}(v_i)]$, $\mathbf{E}[W_g^{(v_i)}(v_i)v_i]$, $\mathbf{E}[W_g^{(v_i, v_i)}(v_i)v_i]$, $\mathbf{E}[W_g^{(v_i, v_i)}(v_i)v_i^2]$ y $\mathbf{E}[W_g^{(\nu, \nu)}(v_i)]$, donde

$$W_g^{(v_i)}(v_i) = -\frac{1}{2} \frac{\nu + p + 1}{\nu + v_i}, \quad W_g^{(v_i, v_i)}(v_i) = \frac{1}{2} \frac{\nu + p + 1}{(\nu + v_i)^2},$$

y

$$W_g^{(\nu, \nu)}(v_i) = \frac{\nu - (p + 1)}{2\nu} \frac{v_i^2}{(\nu + v_i)^2} - (p + 1) \frac{v_i}{(\nu + v_i)^2}.$$

Entonces, es necesario determinar la esperanza $\mathbf{E}[v_i^k (\nu + v_i)^{-m}]$. Notemos que, $\mathbf{Z}_i \sim t_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, \nu)$, $\boldsymbol{\Sigma}^* > 0$, donde $v_i = \|\boldsymbol{\Sigma}^{*-1/2}(\mathbf{Z}_i - \boldsymbol{\mu}^*)\|^2$. También, $v_i/(p + 1) = T_i \|\mathbf{W}_i\|^2/(p + 1) \sim F_{p+1, \nu}$ donde $T_i \sim \nu/\chi_\nu^2$ y $\|\mathbf{W}_i\|^2/(p + 1) \sim \chi_{p+1}^2/(p + 1)$ independientes. Consideremos el siguiente resultado.

Corolario 3.1 (Arellano-Valle (1994)). *Sea $S_1 \sim \chi_r^2/r$ y $S_2 \sim s/\chi_s^2$ independientes. Entonces, $F = S_1 S_2 \sim F_{r, s}$ con función de densidad dada por*

$$f_F(t) = \frac{\Gamma\left(\frac{r+s}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\Gamma\left(\frac{s}{2}\right)} \left(\frac{r}{s}\right)^{r/2} t^{(r-2)/2} \left(1 + \frac{r}{s}t\right)^{-\frac{1}{2}(r+s)}, \quad t > 0,$$

y

$$\mathbb{E}[F^{k/2}] = \left(\frac{s}{r}\right)^{k/2} \frac{\Gamma\left(\frac{r+k}{2}\right) \Gamma\left(\frac{s-k}{2}\right)}{\Gamma\left(\frac{r}{2}\right) \Gamma\left(\frac{s}{2}\right)}, \quad k < s.$$

Definiendo $F = v_i/(p+1) \sim F_{p+1,\nu}$, entonces

$$\frac{v_i^k}{(\nu + v_i)^m} = \nu^{k-m} \left(\frac{p+1}{\nu} F\right)^k \left(1 + \frac{p+1}{\nu} F\right)^{-m}.$$

Así, del Corolario 3.1

$$\begin{aligned} \mathbb{E}\left[\frac{v_i^k}{(\nu + v_i)^m}\right] &= \nu^{k-m} \mathbb{E}\left[\left(\frac{p+1}{\nu} F\right)^k \left(1 + \frac{p+1}{\nu} F\right)^{-m}\right], \\ &= \frac{B\left(\frac{p+1+2k}{2}, \frac{\nu+2m-2k}{2}\right)}{\nu^{m-k} B\left(\frac{p+1}{2}, \frac{\nu}{2}\right)}, \quad \text{si } \nu + 2m - 2k > 0, \end{aligned}$$

donde $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ es la función beta. Por lo tanto

$$\begin{aligned} \mathbb{E}[W_g^{(v_i)}(v_i)] &= -\frac{\nu + p + 1}{2\nu} \frac{B\left(\frac{p+1}{2}, \frac{\nu+2}{2}\right)}{B\left(\frac{p+1}{2}, \frac{\nu}{2}\right)}, \\ \mathbb{E}[W_g^{(v_i)}(v_i)v_i] &= -\frac{\nu + p + 1}{2} \frac{B\left(\frac{p+3}{2}, \frac{\nu}{2}\right)}{B\left(\frac{p+1}{2}, \frac{\nu}{2}\right)}, \\ \mathbb{E}[W_g^{(v_i, v_i)}(v_i)v_i] &= \frac{\nu + p + 1}{2\nu} \frac{B\left(\frac{p+3}{2}, \frac{\nu+2}{2}\right)}{B\left(\frac{p+1}{2}, \frac{\nu}{2}\right)}, \\ \mathbb{E}[W_g^{(v_i, v_i)}(v_i)v_i^2] &= \frac{\nu + p + 1}{2} \frac{B\left(\frac{p+5}{2}, \frac{\nu}{2}\right)}{B\left(\frac{p+1}{2}, \frac{\nu}{2}\right)}, \\ \mathbb{E}[W_g^{(\nu, \nu)}(v_i)] &= \frac{\nu - (p+1)}{2\nu} \frac{B\left(\frac{p+5}{2}, \frac{\nu}{2}\right)}{B\left(\frac{p+1}{2}, \frac{\nu}{2}\right)} - \frac{p+1}{\nu} \frac{B\left(\frac{p+3}{2}, \frac{\nu+2}{2}\right)}{B\left(\frac{p+1}{2}, \frac{\nu}{2}\right)}. \end{aligned}$$

Con estos resultados y (3.27) es fácil determinar la matriz de información esperada del modelo estructural t .

3.10. Modelo estructural de mezcla de escala de normales

Como notamos en la sección 3.9, la representación de mezcla de escala de normales, es útil para implementar el algoritmo de estimación EM. Entonces, se puede explorar la posibilidad de reemplazar la distribución normal por una mezcla de escala normal, la cual incluye la distribución t como un caso especial.

La mezcla de escala del modelo estructural normal es especificada considerando las variables latentes $\mathbf{T} = (T_1, \dots, T_n)^\top$, donde $T_i > 0$, para $i = 1, 2, \dots, n$, y asumiendo que,

$$\begin{aligned} Y_i | \mathbf{X}_i = \mathbf{x}_i, T_i = t_i &\sim N(\beta_0^* + \boldsymbol{\beta}^{*\top} \mathbf{x}_i, t_i \sigma_{\epsilon^*}^2), \\ \mathbf{X}_i | T_i = t_i &\sim N_p(\boldsymbol{\eta}_\xi, t_i \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}), \\ T_i &\sim F_{T_i}(t_i, \boldsymbol{\nu}). \end{aligned}$$

Las variables aleatorias T_i son asumidas independientes de $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*\top}, \boldsymbol{\theta}_2^{*\top})^\top$, donde $\boldsymbol{\theta}_1^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2)^\top$ y $\boldsymbol{\theta}_2^* = (\boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top)^\top$, para $i = 1, 2, \dots, n$.

El modelo anterior incluye varios casos especiales. En el caso usual, una distribución construida de esta forma es llamada mezcla de escala de distribución normal (ver, por ejemplo, Andrews & Mallows (1974)). La clase de distribuciones normales obtenidas modificando F_{T_i} es llamada distribuciones normales compuestas. Si $T_i = 1$, para $i = 1, 2, \dots, n$, entonces se obtiene el modelo estructural normal, como en la sección 3.6. Por otro lado, si se define $T_i \sim \text{IG}(\nu/2, \nu/2)$ se obtiene el modelo estructural t , como una alternativa más simple a la discutida en la sección 3.9.

Implementar el algoritmo EM en esta representación jerárquica es sencillo, considerando que las variables observadas $(Y_i, \mathbf{X}_i^\top)^\top$ son aumentadas con $\mathbf{T} = (T_1, \dots, T_n)^\top$, tal que los estimadores de verosimilitud estén basados en la función de log-verosimilitud $\ell_C(\boldsymbol{\theta}^*, \boldsymbol{\psi})$ de los datos aumentados $(Y_i, \mathbf{X}_i^\top, T_i)^\top$, para $i = 1, 2, \dots, n$. Por tanto, la densidad conjunta para $(Y_i, \mathbf{X}_i^\top, T_i)^\top$ es dada por

$$f_{Y_i, \mathbf{X}_i, T_i}(y_i, \mathbf{x}_i, t_i; \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \boldsymbol{\nu}) = f_{Y_i | \mathbf{X}_i, T_i}(y_i; \boldsymbol{\theta}_1^*) f_{\mathbf{X}_i | T_i}(\mathbf{x}_i; \boldsymbol{\theta}_2^*) f_{T_i}(t_i; \boldsymbol{\nu}).$$

Entonces, la función de log-verosimilitud de los datos aumentados es

$$\ell_C(\boldsymbol{\theta}^*, \boldsymbol{\nu}) \propto -\frac{n}{2} \{ \log \sigma_{\epsilon^*}^2 + \log |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}| \} - \sum_{i=1}^n \log t_i - \frac{1}{2} \sum_{i=1}^n \frac{v_i}{t_i} + \log f_{T_i}^{(I)}(t_i; \boldsymbol{\nu}),$$

donde $v_i = q(\mathbf{x}_i) + \epsilon_i^{*2} / \sigma_{\epsilon^*}^2$, con $\epsilon_i^* = y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i$, $q(\mathbf{x}_i) = (\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi)$ y

$$f_{T_i}^{(I)}(t_i; \boldsymbol{\nu}) = \prod_{i=1}^n f_{T_i}(t_i; \boldsymbol{\nu})$$

Si $(\boldsymbol{\theta}^{*(l)\top}, \boldsymbol{\nu}^{(l)\top})^\top$ es el valor de $(\boldsymbol{\theta}^{*\top}, \boldsymbol{\nu}^\top)^\top$ obtenido en la l -ésima iteración, la $(l+1)$ -ésima iteración del algoritmo EM consiste en maximizar $\ell_E(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ respecto a $(\boldsymbol{\theta}^{*(l)\top}, \boldsymbol{\nu}^{(l)\top})^\top$, con

$$\begin{aligned} \ell_E(\boldsymbol{\theta}^*, \boldsymbol{\nu}) &= \mathbb{E} [\ell_C(\boldsymbol{\theta}^*, \boldsymbol{\nu}) | \mathbf{Y} = \mathbf{y}, \mathbf{X} = \mathbf{x}, \boldsymbol{\theta}^{*(l)}, \boldsymbol{\nu}^{(l)}], \\ &= -\frac{n}{2} \{ \log \sigma_{\epsilon^*}^2 + \log |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}| \} - \sum_{i=1}^n m_{1,i}^{(l)} - \frac{1}{2} \sum_{i=1}^n v_i m_{2,i}^{(l)} + m_{3,i}^{(l)}, \end{aligned}$$

donde

$$\begin{aligned} m_{1,i}^{(l)} &= \mathbb{E} [\log T_i | \boldsymbol{\theta}^{*(l)}, \boldsymbol{\nu}^{(l)}], \\ m_{2,i}^{(l)} &= \mathbb{E} \left[\frac{1}{T_i} | \boldsymbol{\theta}^{*(l)}, \boldsymbol{\nu}^{(l)} \right], \\ m_{3,i}^{(l)} &= \mathbb{E} \left[\log f_{T_i}^{(I)}(t_i; \boldsymbol{\nu}) | \boldsymbol{\theta}^{*(l)}, \boldsymbol{\nu}^{(l)} \right], \end{aligned}$$

para $i = 1, 2, \dots, n$. Como las variables aleatorias T_i , son independientes de $\boldsymbol{\theta}^*$, las esperanzas $m_{k,i}^{(l)}$, para $k = 1, 2, 3$, dependerán tanto de $\boldsymbol{\nu}$ como de $\boldsymbol{\nu}^{(l)}$. Notemos que, una vez definida la distribución de las variables aleatorias T_i , quedara completamente especificada la función $\ell_E(\boldsymbol{\theta}^*, \boldsymbol{\nu})$.

Ahora el paso M comprende de la maximización de $\ell_E(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ respecto de $\boldsymbol{\theta}^*$ y $\boldsymbol{\nu}$. Por lo tanto, el vector de score $\mathbf{S}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ tiene componentes dadas por

$$\begin{aligned} S_{\beta_0^*} &= \frac{1}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n m_{2,i}^{(l)} \epsilon_i^*, \\ \mathbf{S}_{\boldsymbol{\beta}^*} &= \frac{1}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n m_{2,i}^{(l)} \epsilon_i^* \mathbf{x}_i, \\ S_{\sigma_{\epsilon^*}^2} &= -\frac{n}{2\sigma_{\epsilon^*}^2} + \frac{1}{2(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n m_{2,i}^{(l)} \epsilon_i^{*2}, \end{aligned}$$

$$\mathbf{S}_{\boldsymbol{\eta}_\xi} = \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \sum_{i=1}^n m_{2,i}^{(l)}(\mathbf{x}_i - \boldsymbol{\eta}_\xi),$$

$$\mathbf{S}_{\mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})} = -\frac{n}{2} \mathbf{D}^\top \text{vec}(\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) + \mathbf{D}^\top (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1}) \sum_{i=1}^n m_{2,i}^{(l)}((\mathbf{x}_i - \boldsymbol{\eta}_\xi) \otimes (\mathbf{x}_i - \boldsymbol{\eta}_\xi)),$$

por otro lado, se tiene que $\mathbf{S}_\nu = \partial \ell_E(\boldsymbol{\theta}^*, \boldsymbol{\nu}) / \partial \boldsymbol{\nu}$. Entonces, una vez definida la distribución de las variables aleatorias T_i , la solución al sistema $\mathbf{S}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = \mathbf{0}$, corresponderá a la estimación $(l+1)$ -ésima iteración, es decir, $\boldsymbol{\theta}^{*(l+1)}$ y $\boldsymbol{\nu}^{(l+1)}$.

3.11. Discusión

En este capítulo se han estudiado los modelos estructurales elípticos siguiendo el enfoque de que los errores son independientes y tienen distribución marginal elíptica. Adoptamos un enfoque alternativo estudiado por Gleser (1992) y Cheng & Tsai (2004), para obtener una expresión alternativa a la forma usual de la función de verosimilitud. Esta nueva parametrización produce un modelo más parsimonioso transformando el modelo con error de medida usual en un modelo de regresión heteroscedástico con matriz de diseño aleatoria.

El problema de identificación presente en el modelo estructural normal, se traslada a los modelos estructurales elípticos. Por lo tanto, supuestos adicionales son necesarios para hacer el problema de estimación factible, los supuestos que pueden ser considerados son estándar en la literatura relacionada a los modelos con error de medida (ver Fuller (1987) y Cheng & Van Ness (1999)).

Notemos que la representación de los modelos estructurales elípticos en términos de la parametrización de Gleser, nos permite una simple representación jerárquica del modelo (3.5), donde sólo bajo el supuesto de normalidad se tiene que $\boldsymbol{\theta}_1^* = (\beta_0^*, \boldsymbol{\beta}^{*\top}, \sigma_{\epsilon^*}^2)^\top$ y $\boldsymbol{\theta}_2^* = (\boldsymbol{\eta}_\xi^\top, \mathbf{v}(\boldsymbol{\Sigma}_{\xi\xi})^\top)^\top$ están en variación libre y en este caso \mathbf{X}_i es débilmente exógena con respecto a $\boldsymbol{\theta}_1^*$, es decir, podemos sólo considerar el modelo reducido en el proceso de estimación. Pero considerando las distribuciones elípticas, la representación no es exógena y no podemos ignorar la distribución marginal de \mathbf{X}_i en la estimación. Pero la ventaja de considerar las distribuciones elípticas es que permite modelar diferentes fuentes de heteroscedasticidad. Además, considerando la verosimilitud reducida (3.22) es posible definir un método alternativo para la estimación de modelo con error de medida por los métodos clásicos de regresión lineal, a través de un procedimiento de dos pasos, como en Gleser (1992): (i) A partir de la verosimilitud reducida dada en (3.22) estimar $\boldsymbol{\theta}_2^*$ y obtener un estimador $\hat{\boldsymbol{\Lambda}}$ de $\boldsymbol{\Lambda}$, es decir $\hat{\boldsymbol{\Lambda}} = (\hat{\boldsymbol{\Sigma}}_{\xi\xi} + \boldsymbol{\Sigma}_{\delta\delta})^{-1} \hat{\boldsymbol{\Sigma}}_{\xi\xi}$,

y (ii) obtener un estimador de $\boldsymbol{\beta}$ y otros parámetros a través del ajuste del modelo de regresión clásico

$$Y_i = \zeta + \mathbf{X}_i^\top \hat{\boldsymbol{\Lambda}} \boldsymbol{\beta} + \text{error}_i, \quad i = 1, 2, \dots, n,$$

usando un apropiado método de ajuste. Sea $\hat{\boldsymbol{\beta}}_0 = \hat{\zeta} - \hat{\boldsymbol{\beta}}^\top (\mathbf{I}_p - \hat{\boldsymbol{\Lambda}}^\top) \hat{\boldsymbol{\eta}}_\xi$, con $\hat{\boldsymbol{\eta}}_\xi = \bar{\mathbf{x}}$. Notemos que si en (ii), usamos el método de mínimos cuadrados para obtener las estimaciones de ζ y $\boldsymbol{\beta}$, los estimadores resultantes son generalizaciones de los estimadores de corrección por atenuación usados en psicometría.

Por otro lado, considerando el enfoque de cálculo matricial como es desarrollado por Magnus & Neudecker (1999), es fácil (después de un extenso álgebra), determinar el vector de score para el modelo reparametrizado, ver Proposición 3.1. Para ciertas elecciones de la función generadora de densidades, el resultado presentado en la Proposición 3.1 facilita la implementación de rutinas estándar de optimización. Se dan expresiones simples para calcular los elementos de la matriz de información esperada, la cual sólo algunos momentos univariados deben ser calculados numéricamente.

Dos ejemplos de distribuciones elípticas fueron considerados. Por un lado, en el caso normal, se muestra que los estimadores de máxima verosimilitud del modelo reparametrizado corresponden a los estimadores mínimos cuadrados y la matriz de información es diagonal por bloques, donde se tiene que los parámetros $\boldsymbol{\theta}_1^*$ y $\boldsymbol{\theta}_2^*$ son ortogonales. Por otro lado, en el caso de la distribución t , la inferencia debe ser realizada iterativamente, notando que este modelo puede ser representado como una mezcla de normales, es fácil el uso del algoritmo EM para la estimación. Otra forma de estudiar el modelo reparametrizado es considerar los modelos de mezcla de escala de normales, esto simplifica la implementación de métodos iterativos en el proceso de estimación.

Capítulo 4

Reducción de sesgo en modelos estructurales elípticos

4.1. Introducción

El método de estimación más común en la escuela frecuentista es la máxima verosimilitud (MV). Las razones para esto son principalmente las atractivas propiedades de los estimadores de MV, es decir, el principio de invarianza, normalidad asintótica, suficiencia asintótica, insesgamiento y eficiencia, y la fácil implementación de procedimientos de ajuste, si las condiciones de regularidad usuales son satisfechas (Cox & Hinkley (1974)), lo que hace que este método sea una elección por defecto en muchas aplicaciones. Sin embargo, en pequeñas muestras el desempeño de los estimadores de MV es un importante tema de la teoría asintótica, debido a que los estimadores de MV son típicamente estimadores sesgados del verdadero valor del parámetro, y su magnitud juega un importante rol en la estimación y éste puede dar lugar a inferencias que pueden inducir al error.

En el contexto de los modelos con error de medida, varios estudios en la literatura muestran las ventajas del enfoque de verosimilitud en los modelos con error de medida sobre alternativas basadas en supuestos más débiles. En particular, las ventajas son sustanciales en términos de grandes muestras (Schafer & Purdy (1996) y Küchenhoff & Carroll (1997)). A pesar de esto, el enfoque de verosimilitud no ha tenido una cantidad considerable de aplicaciones comparadas con la cantidad de alternativas propuestas (Carroll et al. (2006)). Es bien conocido que cuando el error de medición es ignorado en el proceso de estimación, los estimadores de MV son sesgados e inconsistentes y el principal interés es corregir estos problemas.

Específicamente, como se expuso en el Capítulo 1, la estimación de la pendiente de un modelo lineal simple con error de medida es atenuada por la razón de confiabilidad. Entonces, cuando las variables están sujetas a errores de medición, un tratamiento inferencial especial debe llevarse a cabo para la estimación con el fin de evitar los estimadores sesgados e inconsistentes. Una cuidadosa y detallada exposición en el proceso inferencial en modelos con error de medida puede ser visto en Fuller (1987). Si el modelo con error de medida incluye dos covariables, y una de ellas es medida con error y la otra está libre de éste, se puede derivar analíticamente, con los supuestos habituales sobre los errores, el sesgo de los coeficientes asociados a cada covariable. Este sesgo depende de la proporción de la varianza de los errores y también de la correlación entre las covariables (Griliches & Ringstad (1971)). Cuando, sin embargo, el error está presente en ambas covariables, el sesgo que se produce no puede ser evaluado fácilmente (Maddala (1988)).

En general, para modelos con dos o más covariables, se tiene que los errores de medición producen sesgo en las estimaciones de los coeficientes aún cuando alguna covariable esté libre de error (Johnston (1997)).

Aunque el sesgo no es un serio problema para tamaños de muestra relativamente grandes, en muestras de tamaño pequeño o moderado el sesgo puede constituir un problema. En la literatura estadística existen muchos trabajos sobre las formas que se podría corregir el sesgo de los estimadores de MV en modelos paramétricos, y dependiendo del proceso de corrección, se pueden diferenciar dos clases, las cuales se pueden clasificar como: estimadores con sesgo corregido (ESC) y estimadores con sesgo reducido (ESR).

Estimadores con sesgo corregido

Para determinar los ESC el proceso de estimación debe realizarse en dos pasos: (i) obtener el sesgo de primer orden de los estimadores de MV, y (ii) restar este sesgo a dichos estimadores. Para ilustrar esto consideremos el caso uni-paramétrico, donde el valor esperado del estimador de MV, digamos $\hat{\theta}$, del parámetro desconocido θ de un modelo paramétrico puede ser generalmente expresado como

$$E[\hat{\theta}] = \theta_0 + \frac{b_1(\theta_0)}{n} + \frac{b_2(\theta_0)}{n^2} + \frac{b_3(\theta_0)}{n^3} + \dots ,$$

donde n es usualmente interpretado como el número de observaciones, pero puede ser alguna otra medida sobre la tasa de acumulación de la información observada. Sea θ_0 el valor verdadero del parámetro desconocido y $b_r(\cdot)$, para $r = 1, 2, \dots$, funciones

$O(1)$ de θ , las cuales pueden ser obtenidas explícitamente una vez que el modelo ha sido especificado. Por lo tanto, a través de un simple arreglo

$$E[\hat{\theta}] - \frac{b_1(\theta_0)}{n} = \theta_0 + \frac{b_2(\theta_0)}{n^2} + \frac{b_3(\theta_0)}{n^3} + \dots,$$

es posible corregir el sesgo de $\hat{\theta}$ de orden $O(n^{-1})$. Ésta es la línea argumental de los métodos correctivos. Cox & Snell (1968) derivan las expresiones para $b_1(\theta_0)/n$ para una amplia familia de modelos. Donde, un enfoque estándar es simplemente sustituir $\hat{\theta}$ por θ_0 en $b_1(\theta_0)/n$, entonces los ESC son calculados como

$$\hat{\theta}_{SC} = \hat{\theta} - \frac{b_1(\hat{\theta})}{n}.$$

La principal ventaja de este método es la simplicidad de su aplicación. Una vez que la estimación del sesgo está disponible, la corrección del sesgo es simplemente una cuestión de un procedimiento de una etapa donde el sesgo estimado es abstraído desde el estimador. Sin embargo, debido a su dependencia explícita sobre $\hat{\theta}$, los ESC heredan directamente cualquier inestabilidad del estimador inicial. En particular, los ESC dependen de la existencia de los estimadores de MV. Entonces, por definición los ESC son indefinidos cuando los estimadores de MV divergen. Además, los ESC tiene la desventaja que sólo son aplicables cuando $b_1(\theta)/n$ puede ser obtenido explícitamente, lo cual puede ser tedioso o una tarea impracticable para muchos modelos. En esta línea, existe una extensa literatura en los métodos correctivos donde se han obtenido incluso expresiones analíticas para el sesgo de primer orden, ejemplos característicos de dichos estudios son Cox & Snell (1968), Cook et al. (1986), Cordeiro & McCullagh (1991), Breslow & Lin (1995), Lin & Breslow (1996), Cordeiro & Vasconcellos (1997), Cordeiro & Vasconcellos (1999), Cordeiro et al. (2000), Vasconcellos & Cordeiro (2000) y Cordeiro & Toyama Udo (2008). En el contexto de los modelos con medidas de error, Patriota & Lemonte (2009) y Patriota et al. (2011) estudian la corrección de sesgo en los modelos con medida de error heteroscedásticos asumiendo normalidad.

Estimadores con sesgo reducido

La principal diferencia entre los ESC y ESR, es que estos últimos no dependen directamente de los estimadores de MV. Estos estimadores son obtenidos a través de la modificación del vector de score, y en este sentido se puede considerar que la naturaleza de estos métodos es ser preventiva en vez de correctiva.

Quenouille (1956) ha sido el primero en desarrollar un método de reducción de sesgo que es aplicable a una familia general de distribuciones. Este es el procedimiento jackknife, donde el objetivo es remover el sesgo de un orden específico. Sin embargo, si los estimadores de MV no tienen forma cerrada, el método jackknife puede ser costoso computacionalmente, debido a que los estimadores de MV deben ser obtenidos iterativamente para cada uno de todos los posibles subconjuntos de la muestra de acuerdo al esquema de particionamiento considerado. Además, varias consideraciones deben ser realizadas en aquellos casos donde los estimadores de MV para un subconjunto de la muestra sea infinito.

Firth (1993) desarrolla un método general para remover el término de primer orden en la expansión asintótica del sesgo de los estimadores de MV. El vector de score es apropiadamente modificado tal que las raíces resultantes del vector de score modificado produzcan un estimador insesgado. Estos estimadores difieren de los estimadores obtenidos a través de los métodos correctivos, ya que no son obtenidos directamente de los estimadores de MV y por tanto, no dependen de la finitud de éstos. Este último hecho ha motivado el estudio y uso de los ESR, por ejemplo, Pettitt et al. (1998), Heinze & Schemper (2001), Bull et al. (2002), Sartori (2006) y Kosmidis & Firth (2009). Para una discusión sobre el sesgo en la estimación paramétrica, ver Kosmidis (2014).

Pero, una desventaja del método de reducción de sesgo desarrollado por Firth, es que su aplicación requiere de la solución de un conjunto de ecuaciones las cuales en muchos de los casos útiles requiere de optimización numérica. Esta tarea es a menudo más demandante computacionalmente en modelos generales debido a la necesidad de aproximar la función de sesgo en un espacio p dimensional. Por lo tanto, el objetivo de este capítulo es implementar los métodos corrección y reducción de sesgo en los modelos estructurales considerando la parametrización de Gleser.

Por tanto, este capítulo está estructurado de la siguiente forma, la sección 4.2 ilustra para el caso uni-paramétrico, el método de reducción de sesgo desarrollado por Firth (1993). Basado en el método de reducción de sesgo, en la sección 4.3, se desarrolla un algoritmo tipo Fisher-Scoring *modificado* para obtener versiones reducidas y corregidas de los estimadores de MV. La sección 4.4, ilustra el sesgo en los modelos estructurales elípticos y muestra algunas consideraciones para su implementación. Las secciones 4.4.1 y 4.4.2, se aplican los métodos de corrección y reducción de sesgo en los modelos estructurales, considerando los supuestos distribucionales normal y t . Se realizan estudios de simulación e ilustraciones con conjuntos de datos reales. Finalmente, una discusión del capítulo es presentada en la sección 4.5.

4.2. Método de reducción de sesgo

Asumamos que, a la luz de n realizaciones de las variables aleatorias independientes, formulamos la log-verosimilitud $\ell(\theta)$. Bajo las condiciones de regularidad usuales, Firth (1993) desarrolla una familia general de modificaciones del vector de score, donde se muestra que las raíces resultantes obtenidas a través de este vector de score modificado, resultan en un estimador con sesgo de segundo orden. Notemos que, estas condiciones dadas en Cox & Hinkley (1974), son utilizadas para justificar la expansión de Taylor, y el desarrollo de un algoritmo iterativo de reducción de sesgo.

Denotemos por $S(\theta)$ y $I(\theta)$, la función de score y la información observada, respectivamente. Además, $F(\theta) = \mathbf{E}[I(\theta)]$ la información esperada, y $P(\theta) = \mathbf{E}[S(\theta)^3]$ y $Q(\theta) = -\mathbf{E}[S(\theta)I(\theta)]$. Típicamente, se tiene que $I(\theta)$ es de orden $O_p(n)$ y $F(\theta)$, $P(\theta)$ y $Q(\theta)$ son cantidades de orden $O(n)$, además el estimador de máxima verosimilitud tiene, en general, un sesgo del orden $O(n^{-1})$. En particular, a través de la expansión de Taylor de la ecuación de verosimilitud, $S(\hat{\theta}) = 0$, se puede ver que $\mathbf{E}[\hat{\theta} - \theta] = b(\theta) + O(n^{-2})$ donde

$$b(\theta) = -\frac{1}{2}F(\theta)^{-2}[P(\theta) + Q(\theta)] = O(n^{-1}).$$

Por lo tanto, el estimador con sesgo corregido

$$\hat{\theta}_{SC} = \hat{\theta} - b(\hat{\theta}),$$

es tal que $\mathbf{E}[\hat{\theta}_{SC} - \theta] = O(n^{-2})$. Claramente, el inconveniente de los métodos de corrección de sesgo, es que se requiere que $\hat{\theta}$ sea finito.

La idea de Firth es prevenir el sesgo, $b(\theta)$, antes de calcular $\hat{\theta}$. Notemos que el sesgo en $\hat{\theta}$ surge de la combinación del insesgamiento de la función de score, $\mathbf{E}[S(\theta)] = 0$, en el verdadero valor de θ , y la curvatura de la función de score, $S''(\theta) \neq 0$. Si $S(\theta)$ fuera lineal en θ , entonces $\mathbf{E}[\hat{\theta}] = 0$, por ejemplo, una curvatura positiva combinada con el insesgamiento de la función de score inducirá un sesgo en $\hat{\theta}$, en este caso en la dirección positiva. Así, el sesgo en $\hat{\theta}$ puede ser reducido introduciendo una perturbación en la función de score. Si $\hat{\theta}$ está sujeto a un sesgo positivo $b(\theta)$, la función de score es desplazada hacia abajo en cada punto de θ por una cantidad $M(\theta)$, de orden $O(1)$. Esto produce la ecuación de verosimilitud modificada

$$S_M(\theta) = S(\theta) + M(\theta) = 0. \tag{4.1}$$

La solución de (4.1) produce el estimador con sesgo reducido, $\hat{\theta}_{SR}$, donde la modificación $M(\theta)$ es elegida tal que

$$\mathbb{E}[\hat{\theta}_{SR} - \theta] = O(n^{-2}). \quad (4.2)$$

Usando la expansión de Taylor para $S_M(\hat{\theta}_{SR}) = 0$ alrededor de θ , podemos encontrar una expansión para $\hat{\theta}_{SR} - \theta$. Entonces, imponiendo la condición (4.2), se tiene que

$$M(\theta) = -F(\theta)b(\theta). \quad (4.3)$$

El estimador $\hat{\theta}_{SR}$ es equivalente en segundo orden a $\hat{\theta}_{SC}$, con la ventaja de que no depende en la finitud de $\hat{\theta}$, y además no heredara cualquier inestabilidad que pueda tener el estimador inicial.

4.3. Algoritmos de corrección y reducción de sesgo

Consideramos un modelo con vector de parámetros dado por $\theta = (\theta_1, \theta_2, \dots, \theta_p)$. Si $\ell(\theta)$ es la función de log-verosimilitud para θ , el estimador de MV, $\hat{\theta}$, es la solución de las ecuaciones de score, $\mathbf{S}(\theta) = \mathbf{0}$, debido que la matriz de información observada $\mathbf{I}(\theta)$ es definida positiva cuando es evaluada en $\hat{\theta}$. Si definimos

$$M_r(\theta) = \frac{1}{2} \text{tr} [\mathbf{F}^{-1}(\theta)(\mathbf{P}_r(\theta) + \mathbf{Q}_r(\theta))], \quad r = 1, 2, \dots, p, \quad (4.4)$$

la solución de las ecuaciones del vector de score modificado

$$\mathbf{S}_M(\theta) = \mathbf{S}(\theta) + \mathbf{M}(\theta) = \mathbf{0}, \quad (4.5)$$

donde $\mathbf{M}(\theta) = (M_1(\theta), M_2(\theta), \dots, M_p(\theta))^\top$, resulta en un estimador, $\hat{\theta}_{SR}$, es cual está libre del término de primer orden en la expansión asintótica del sesgo, donde $\mathbf{M}(\theta)$ es de orden $O(1)$. Aquí, $\mathbf{F}(\theta) = \mathbb{E}[\mathbf{I}(\theta)]$ corresponde a la matriz de información esperada y $\mathbf{P}_r(\theta) = \mathbb{E}[S_r(\theta)\mathbf{S}(\theta)\mathbf{S}^\top(\theta)]$, es el r -ésimo bloque de la matriz, de dimensión $p^2 \times p$, de los cumulantes de tercer orden del vector de score y $\mathbf{Q}_r(\theta) = -\mathbb{E}[S_r(\theta)\mathbf{I}(\theta)]$, es el r -ésimo bloque de la matriz, de dimensión $p^2 \times p$, de la covarianza entre las primeras y segundas derivadas de la función de log-verosimilitud. El vector del sesgo de primer orden, $\mathbf{b}(\cdot)$, puede ser expresado en términos de $\mathbf{M}(\theta)$ como

$$\mathbf{b}(\theta) = -\mathbf{F}^{-1}(\theta)\mathbf{M}(\theta), \quad (4.6)$$

y por tanto, el estimador con sesgo corregido es

$$\hat{\boldsymbol{\theta}}_{SC} = \hat{\boldsymbol{\theta}} - \mathbf{b}(\hat{\boldsymbol{\theta}}).$$

El vector de modificaciones, $\mathbf{M}(\boldsymbol{\theta})$, también puede estar basado en la matriz de información observada, el cual lo podemos expresar como

$$M_r^{(O)}(\boldsymbol{\theta}) = \mathbf{I}_r(\boldsymbol{\theta})\mathbf{F}^{-1}(\boldsymbol{\theta})\mathbf{M}(\boldsymbol{\theta}), \quad r = 1, 2, \dots, p,$$

donde el vector $\mathbf{I}_r(\boldsymbol{\theta})$, de dimensión $1 \times p$, denota a la r -ésima fila de $\mathbf{I}(\boldsymbol{\theta})$, es decir, la matriz de información observada. Las dos formas de modificación, $\mathbf{M}(\boldsymbol{\theta})$ o $\mathbf{M}^{(O)}(\boldsymbol{\theta})$, son iguales en los modelos exponenciales. En modelos más generales, no existe ganancia en eficiencia significativa en el uso de uno o de otro (Firth (1995)).

Alternativamente, considerando (4.6) y la notación utilizada por Cordeiro & Klein (1994), el vector de modificaciones puede ser expresado como

$$\mathbf{M}(\boldsymbol{\theta}) = -\mathbf{W}\text{vec}(\mathbf{F}^{-1}(\boldsymbol{\theta})), \quad (4.7)$$

donde $\mathbf{W} = (\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(p)})$ es una matriz particionada de dimensión $p \times p^2$, las matrices $\mathbf{W}^{(l)}$ tienen el (i, j) -ésimo elemento definido por $w_{ij}^{(l)} = \kappa_{ij}^{(l)} - \frac{1}{2}\kappa_{ijl}$, para $l = 1, 2, \dots, p$ y los cumulantes nulos de las derivadas de la función de log-verosimilitud se definen como

$$\kappa_{ij} = \mathbf{E} \left[\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j} \right], \quad \kappa_{ijl} = \mathbf{E} \left[\frac{\partial^3 \ell}{\partial \theta_i \partial \theta_j \partial \theta_l} \right] \quad \text{y} \quad \kappa_{ij}^{(l)} = \frac{\partial \kappa_{ij}}{\partial \theta_l},$$

para $i, j, l = 1, 2, \dots, p$.

Una vez definido el vector de score modificado (4.5), su aplicación requiere de la solución de un conjunto de ecuaciones las cuales en muchos de los casos es necesaria la implementación de algún método de optimización numérica.

Una forma de resolver las ecuaciones del vector de score modificado puede ser a través de la implementación de un algoritmo tipo Fisher-Scoring *modificado*, es decir, el vector de score de la función de log-verosimilitud es reemplazado por el vector de score modificado (ver Kosmidis & Firth (2009) y Kosmidis & Firth (2010)). Por esta razón, el siguiente algoritmo de Fisher-Scoring modificado puede ser desarrollado

$$\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} + \mathbf{F}^{-1}(\boldsymbol{\theta}^{(j)})\mathbf{S}_M(\boldsymbol{\theta}^{(j)}). \quad (4.8)$$

Sustituyendo (4.5) y (4.6) en (4.8), se tiene que

$$\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} + \mathbf{F}^{-1}(\boldsymbol{\theta}^{(j)})\mathbf{S}(\boldsymbol{\theta}^{(j)}) + \mathbf{F}^{-1}(\boldsymbol{\theta}^{(j)})\mathbf{M}(\boldsymbol{\theta}^{(j)}), \quad (4.9)$$

los dos primeros términos en el lado derecho de la expresión anterior corresponde al algoritmo Fisher-Scoring estándar para maximizar la función de log-verosimilitud y el último término es el vector de sesgo de primer orden definido en (4.6), por tanto (4.9) puede ser re-expresado como

$$\boldsymbol{\theta}^{(j+1)} = \hat{\boldsymbol{\theta}}^{(j+1)} - \mathbf{b}(\boldsymbol{\theta}^{(j)}), \quad (4.10)$$

donde $\hat{\boldsymbol{\theta}}^{(j+1)}$ se define como el valor candidato para los estimadores de MV que son obtenidos en un paso del algoritmo Fisher-Scoring de $\boldsymbol{\theta}^{(j)}$. Si los estimadores de MV son finitos, estos pueden proveen una buena elección de los valores iniciales.

Notemos que, de (4.10) se tiene que en cada paso, el siguiente candidato del estimador de MV es corregido restando el sesgo, evaluado en el valor actual del estimador de sesgo reducido. Por lo tanto, si $\boldsymbol{\theta}^{(0)} = \hat{\boldsymbol{\theta}}$, el primer paso del esquema propuesto entrega el estimador de MV con sesgo corregido, e iterando hasta la convergencia donde se obtiene los estimadores de MV con sesgo reducido, basados en el ajuste $\mathbf{M}(\boldsymbol{\theta})$.

Si consideremos la reparametrización general de $\boldsymbol{\theta}$ a $\boldsymbol{\theta}^* = h(\boldsymbol{\theta})$, para alguna función $h : \mathbb{R}^p \rightarrow \mathbb{R}^q$. Debido a la propiedad de invarianza, el estimador de MV es equivariante bajo reparametrizaciones. Entonces, el ESR de $\boldsymbol{\theta}^*$ pueden ser obtenido usando la iteración (4.10) considerando que la matriz de información esperada, se puede determinar como $\mathbf{F}(\boldsymbol{\theta}^*) = \mathbf{J}(\boldsymbol{\theta}^*)\mathbf{F}(\boldsymbol{\theta}(\boldsymbol{\theta}^*))\mathbf{J}(\boldsymbol{\theta}^*)^\top$ (ver Lehmann & Casella (1998)) y $\mathbf{S}(\boldsymbol{\theta}^*) = \mathbf{J}(\boldsymbol{\theta}^*)\mathbf{S}(\boldsymbol{\theta}(\boldsymbol{\theta}^*))$, donde $\mathbf{J}(\boldsymbol{\theta}^*)$ es la matriz Jacobiana de la reparametrización.

4.4. Reducción de sesgo en modelos estructurales

En el análisis de regresión estándar, típicamente, el interés radica en la estimación de los coeficientes de regresión. El método de mínimos cuadrados, es el método usualmente seleccionado para realizar la estimación, ya que provee estimaciones consistentes de los coeficientes de regresión del modelo lineal cuando las covariables son medidas sin error. Pero, si las covariables son medidas con error, este hecho induce un sesgo en los estimadores de mínimos cuadrados. Si consideramos el modelo estructural elíptico definido en (3.1), bajo los supuestos sobre los errores definidos en el capítulo 3. Los estimadores de mínimos cuadrados de β_0 y $\boldsymbol{\beta}$ vienen dados por $\hat{\beta}_0 = \bar{y} - \mathbf{s}_{yx}^\top \mathbf{S}_{xx}^{-1} \bar{\mathbf{x}}$ y $\hat{\boldsymbol{\beta}} = \mathbf{S}_{xx}^{-1} \mathbf{s}_{yx}$, donde \mathbf{s}_{yx} y \mathbf{S}_{xx} son las covarianzas muestrales.

Notemos que, a partir de estos resultados, se tiene que

$$E[\hat{\beta}_0] = \beta_0 + \boldsymbol{\beta}^\top (\mathbf{I}_p - \boldsymbol{\Lambda}^\top) \boldsymbol{\eta}_\xi \quad \text{y} \quad E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\Lambda} \boldsymbol{\beta},$$

donde $\boldsymbol{\Lambda} = (\boldsymbol{\Sigma}_{\delta\delta} + \boldsymbol{\Sigma}_{\xi\xi})^{-1} \boldsymbol{\Sigma}_{\xi\xi}$. Entonces, tenemos que el sesgo resultante puede ser complejo. En el caso de una covariable, el sesgo se manifiesta como una atenuación a cero de la pendiente de la recta de regresión. En el caso multivariado, no existe una clara relación entre el signo y magnitud del coeficiente de regresión y su estimación. Incluso, si sólo una variable es medida con error, esto induce sesgo en las estimaciones de todos los coeficientes, incluyendo aquellos que son medidos sin error. Para una ilustración del sesgo en los modelos lineales, bajo diferentes escenarios, ver Buonaccorsi (2010).

Así, cuando las covariables están sujetas a un error de medición, un tratamiento inferencial especial debe llevarse a cabo para el proceso de estimación, con el fin de evitar los estimadores sesgados e inconsistentes. Por lo tanto, en lo que resta de este capítulo, implementaremos los algoritmos de corrección y reducción de sesgo descritos anteriormente, en algunos casos particulares. Para esto, realizaremos algunas consideraciones sobre los modelos estructurales elípticos que serán analizados.

Consideremos la situación en donde se tienen n vectores aleatorios independientes $\mathbf{Z}_i = (\mathbf{X}_i^\top, Y_i)^\top$, para $i = 1, 2, \dots, n$, observados. Si el modelo estructural elíptico definido en (3.1) es reparametrizado como en la sección 3.3, entonces el modelo lineal estará definido sólo a través de la información observada, es decir

$$Y_i = \beta_0^* + \boldsymbol{\beta}^{*\top} \mathbf{X}_i + \epsilon_i^*, \quad (4.11)$$

donde $\mathbf{Z}_i \sim \text{El}_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; g)$, con $\boldsymbol{\mu}^*$ y $\boldsymbol{\Sigma}^*$ definidos en (3.11). La función de log-verosimilitud asociada a (4.11) es

$$\ell^* := \ell(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = n \log C_{p+1}(\boldsymbol{\nu}) - \frac{n}{2} \log \sigma_{\epsilon^*}^2 - \frac{n}{2} \log |\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}| + \sum_{i=1}^n \log g(v_i),$$

con $v_i = (\mathbf{x}_i - \boldsymbol{\eta}_\xi)^\top \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}}^{-1} (\mathbf{x}_i - \boldsymbol{\eta}_\xi) + \epsilon_i^{*2} / \sigma_{\epsilon^*}^2$, $\epsilon_i^* = y_i - \beta_0^* - \boldsymbol{\beta}^{*\top} \mathbf{x}_i$ y constante de normalización $C_{p+1}(\boldsymbol{\nu})$ dada en (2.5). Para obtener los estimadores con sesgo reducido, se debe definir el vector de score modificado (4.5). Entonces, el vector de modificaciones $\mathbf{M}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$, se puede determinar ya sea a través de (4.4) o de (4.7). Dependiendo del modelo seleccionado, estas expresiones requieren de algún álgebra, y dependiendo de la complejidad de las expresiones resultantes, será necesaria la implementación de

algún método de optimización numérica. Si consideramos (4.4) para definir el vector de modificaciones, deben ser especificadas las matrices $\mathbf{P}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ y $\mathbf{Q}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$, donde la matriz de los cumulantes de tercer orden del vector de score, es dada por

$$\mathbf{P}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = \mathbb{E}[\mathbf{S}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) \otimes \mathbf{S}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) \mathbf{S}^\top(\boldsymbol{\theta}^*, \boldsymbol{\nu})],$$

donde $\mathbf{S}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ es el vector de score definido en la Proposición 3.1. Además, la matriz de las covarianzas entre las primeras y segundas derivadas de la función de log-verosimilitud, viene dada por

$$\mathbf{Q}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = -\mathbb{E}[\mathbf{S}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) \otimes \mathbf{I}(\boldsymbol{\theta}^*, \boldsymbol{\nu})],$$

donde $\mathbf{I}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ es la matriz de información observada definida en (3.24). Notemos que, si consideramos que el vector de parámetros $(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ es de dimensión $m \times 1$, donde $m = (p+1)(p+4)/2 + q$ y q representa el número de parámetros del vector $\boldsymbol{\nu}$. Entonces los elementos del vector de modificaciones $\mathbf{M}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$, vienen dados por $M_r(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ definido en (4.4), con $r = 1, \dots, m$. Además, se define a las matrices $\mathbf{P}_r(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ y $\mathbf{Q}_r(\boldsymbol{\theta}^*, \boldsymbol{\nu})$, como el r -ésimo bloque de la respectivas matrices, ambas de dimensión $m^2 \times m$.

Por otro lado, si consideramos (4.7) se tiene que el vector de modificaciones puede ser derivado utilizando

$$\mathbf{M}(\boldsymbol{\theta}^*, \boldsymbol{\nu}) = -\mathbf{W} \text{vec}(\mathbf{F}^{-1}(\boldsymbol{\theta}^*, \boldsymbol{\nu})),$$

donde $\mathbf{F}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ corresponde a la matriz de información esperada para el modelo dado en (4.11), dada por la Proposición 3.2, y los elementos de \mathbf{W} son descritos en (4.7). La elección de uno u otro método para determinar el vector de modificaciones, dependerá de la elección de la función generadora de densidades, definida por g , dado que facilitará la implementación del algoritmo de reducción de sesgo. Finalmente, con el vector $\mathbf{M}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$ y (4.6), es fácil determinar el vector de sesgo de primer orden $\mathbf{b}(\boldsymbol{\theta}^*, \boldsymbol{\nu})$.

Definiendo $(\hat{\boldsymbol{\theta}}_{SR}^*, \hat{\boldsymbol{\nu}}_{SR})$ y $(\hat{\boldsymbol{\theta}}_{SC}^*, \hat{\boldsymbol{\nu}}_{SC})$ como los estimadores con sesgo reducido y corregido, respectivamente, bajo la parametrización de Gleser. Debido a que los estimadores de MV son equivariantes bajo reparametrizaciones, los componentes del vector de sesgo y por lo tanto el vector de sesgo de primer orden, serán los mismos en las parametrizaciones de Gleser y la parametrización directa (ver Capítulo 3). Entonces, de (3.6), (3.7) y (3.10), obtenemos la correspondiente estimación para $(\boldsymbol{\theta}, \boldsymbol{\nu})$. A conti-

nuación, consideraremos algunos casos particulares para ilustrar la implementación del algoritmo de reducción de sesgo y el impacto sobre las estimaciones.

4.4.1. Modelo estructural normal simple

Consideremos la situación en donde n vectores aleatorios independientes, $\mathbf{Z}_i = [X_i, Y_i]^\top$, para $i = 1, 2, \dots, n$, son observados. El modelo de regresión lineal con error de medida simple es definido en (1.1), con $\mathbf{r}_i = [\epsilon_i, \delta_i, \xi_i]^\top$ normalmente distribuido, es decir

$$\mathbf{r}_i = \begin{bmatrix} \epsilon_i \\ \delta_i \\ \xi_i \end{bmatrix} \stackrel{iid}{\sim} N_3 \left(\begin{bmatrix} 0 \\ 0 \\ \eta_\xi \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2 & 0 & 0 \\ 0 & \sigma_\delta^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{bmatrix} \right), i = 1, 2, \dots, n.$$

Entonces, de (3.5), tenemos el modelo reparametrizado con $\epsilon_i^* \sim N(0, \sigma_{\epsilon^*}^2)$ y $X_i \sim N(\eta_\xi, \sigma_X^2)$ independientes, y parámetros definidos por

$$\begin{aligned} \beta_0^* &= \beta_0 + \beta_1(1 - \lambda)\eta_\xi, \\ \beta_1^* &= \lambda\beta_1, \\ \sigma_{\epsilon^*}^2 &= \sigma_\epsilon^2 + \beta_1^2\sigma_\delta^2\lambda, \end{aligned}$$

y razón de confiabilidad

$$\lambda = \sigma_\xi^2 / (\sigma_\delta^2 + \sigma_\xi^2) = \sigma_\xi^2 / \sigma_X^2.$$

El vector de parámetros que se deben estimar es $\boldsymbol{\theta}^* = [\beta_0^*, \beta_1^*, \sigma_{\epsilon^*}^2, \eta_\xi, \sigma_\xi^2]^\top$. De la Proposición 3.1, se define el vector de score $\mathbf{S}(\boldsymbol{\theta}^*)$. De la sección 3.8, y considerando la partición del vector $\boldsymbol{\theta}^*$, dada por $\boldsymbol{\theta}_1^* = [\beta_0^*, \beta_1^*, \sigma_{\epsilon^*}^2]^\top$ y $\boldsymbol{\theta}_2^* = [\eta_\xi, \sigma_\xi^2]^\top$, se puede determinar la matriz de información observada y esperada, las cuales son definidas por $\mathbf{I}(\boldsymbol{\theta}^*)$ y $\mathbf{F}(\boldsymbol{\theta}^*)$, respectivamente. Bajo el supuesto de normalidad, es fácil determinar las matrices $\mathbf{P}_r(\boldsymbol{\theta}^*)$ y $\mathbf{Q}_r(\boldsymbol{\theta}^*)$, ambas de dimensión 5×5 . Se debe notar que para cada parámetro involucrado en el modelo se deben calcular éstas matrices. Entonces, de (4.4), tenemos que el vector de modificaciones viene dado por

$$\mathbf{M}(\boldsymbol{\theta}^*) = \left[0, 0, \frac{1}{\sigma_{\epsilon^*}^2}, 0, \frac{1}{2\sigma_X^2} \right]^\top.$$

De (4.5), el vector de score modificado es $\mathbf{S}_M(\boldsymbol{\theta}^*) = \mathbf{S}(\boldsymbol{\theta}^*) + \mathbf{M}(\boldsymbol{\theta}^*)$. Resolviendo el sistema $\mathbf{S}_M(\boldsymbol{\theta}^*) = \mathbf{0}$, se tiene que el estimador con sesgo reducido, $\tilde{\boldsymbol{\theta}}_{SR}^*$, tiene

componentes dados por

$$\begin{cases} \tilde{\beta}_0^* = \bar{y} - \frac{s_{yx}\bar{x}}{S_{XX}}, \\ \tilde{\beta}_1^* = \frac{s_{yx}}{S_{XX}}, \\ \tilde{\sigma}_{\epsilon^*}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \tilde{\beta}_0^* - \tilde{\beta}_1^* x_i)^2, \end{cases}$$

y para los parámetros asociados al error de medición, se tiene que

$$\tilde{\eta}_\xi = \bar{x} \quad \text{y} \quad \tilde{\sigma}_\xi^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \tilde{\eta}_\xi)^2 - \sigma_\delta^2.$$

Ahora, a través de la parametrización dada por (3.6), (3.7) y (3.10), los ESR para el modelo estructural con parametrización directa son dados por

$$\begin{cases} \tilde{\beta}_0 = \tilde{\beta}_0^* - \tilde{\beta}_1^*(\tilde{\lambda}^{-1} - 1)\tilde{\eta}_\xi, \\ \tilde{\beta}_1 = \tilde{\lambda}^{-1}\tilde{\beta}_1^*, \\ \tilde{\sigma}_\epsilon^2 = \tilde{\sigma}_{\epsilon^*}^2 - \tilde{\beta}_1^{*2}\tilde{\lambda}^{-1}\sigma_\delta^2. \end{cases}$$

La razón de confiabilidad puede ser estimada como $\tilde{\lambda} = \tilde{\sigma}_\xi^2/(\tilde{\sigma}_\xi^2 + \sigma_\delta^2)$.

Por otra parte, de (4.6), y considerando la inversa de la matriz de información esperada y el vector de modificaciones, definidos por $\mathbf{F}^{-1}(\boldsymbol{\theta}^*)$ y $\mathbf{M}(\boldsymbol{\theta}^*)$, respectivamente, se tiene que el vector de sesgo de primer orden, $\mathbf{b}(\boldsymbol{\theta}^*)$, es dado por

$$\mathbf{b}(\boldsymbol{\theta}^*) = -\mathbf{F}^{-1}(\boldsymbol{\theta}^*)\mathbf{M}(\boldsymbol{\theta}^*) = \left[0, 0, -\frac{2\sigma_{\epsilon^*}^2}{n}, 0, -\frac{\sigma_X^2}{n} \right]^\top.$$

Los resultados de la simulación de Monte Carlo son presentados y discutidos a continuación. Los resultados numéricos muestran que la reducción de sesgo derivada es efectiva en pequeñas muestras.

Estudio de simulación

En materia de analizar los resultados de la estimación puntual, calculamos para cada tamaño muestral y para cada estimador: el sesgo (el sesgo del estimador $\hat{\theta}$ es definido como $\mathbf{E}[\hat{\theta}] - \theta$, su estimación es obtenida estimando $\mathbf{E}[\hat{\theta}]$, a través de la simulación de Monte Carlo), el sesgo relativo definido como $|\mathbf{E}[\hat{\theta}] - \theta|/\theta$ y la varianza relativa del estimador, obtenida como $|\mathbf{V}(\hat{\theta}^*) - \mathbf{V}(\hat{\theta})|/\mathbf{V}(\hat{\theta})$, para comparar la varianza de los

estimadores con sesgo corregido y reducido con respecto a la varianza del estimador de máxima verosimilitud.

Se consideran dos escenarios diferentes para la elección de los valores verdaderos de los parámetros de la regresión, $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma_\epsilon^2, \eta_\xi, \sigma_\xi^2, \sigma_\delta^2)^\top$, dados por Fuller (1987) y Patriota & Lemonte (2009)

$$\begin{aligned}\boldsymbol{\theta}_{\text{Fuller}} &= (67; 0,42; 43; 70; 247; 57)^\top, \\ \boldsymbol{\theta}_{\text{Patriota}} &= (76; 0,61; 93; 25; 398; 64)^\top,\end{aligned}$$

Las Figura 4.1 y Figura 4.2, muestran el comportamiento del sesgo promedio (en valor absoluto) para $\boldsymbol{\theta}_{\text{Fuller}}$ y $\boldsymbol{\theta}_{\text{Patriota}}$, respectivamente. Donde para cada muestra, los métodos de reducción y corrección de sesgo son bastantes estables y presentan un sesgo menor en comparación con los estimadores de MV. Notemos que, el método de reducción de sesgo tiene un rendimiento similar al método de corrección. Además, en particular la corrección de $\sigma_{\epsilon^*}^2$ y σ_ξ^2 , es muy evidente con ambos métodos. Los tamaños muestrales considerados van de $n = 30$ a $n = 150$, y para cada tamaño muestral se realizaron 10 000 replicaciones. Todas las simulaciones son realizadas usando R (R Core Team (2018)).

Las Tabla 4.1 y 4.2, muestra el sesgo relativo y la varianza relativa, para los diferentes métodos de estimación. Podemos notar la ganancia respecto al sesgo relativo de los métodos de corrección y reducción. Como es sabido, al disminuir el sesgo, aumenta la varianza, según los resultados reportados, el aumento no es significativo.

También se realizó un estudio simulación Monte Carlo para verificar la cobertura empírica y la longitud promedio del intervalo de confianza asintótico del 95 % para $\beta_1 = 0,42$ y $\beta_1 = 0,61$, es decir, para cada configuración de los valores verdaderos de los parámetros. Los intervalos son definidos como $\hat{\beta}_1 \pm 1,96 \times \text{es}(\hat{\beta}_1)$ (para el estimador de MV), y $\hat{\beta}_{1SC} \pm 1,96 \times \text{es}(\hat{\beta}_{1SC})$ y $\hat{\beta}_{1SR} \pm 1,96 \times \text{es}(\hat{\beta}_{1SR})$, para los respectivos estimadores con sesgo corregido y sesgo reducido, donde $\text{es}(\cdot)$ denota a la estimación del error estándar por el uso del respectivo estimador. Los resultados se pueden ver en la Tabla 4.3, donde podemos notar que los ESC y ESR son ligeramente mejores que los correspondientes EVM en ambos casos, es decir, en la cobertura empírica como en la longitud promedio. Además, bajo estos resultados los ESC y ESR son bastante similares para los diferentes tamaños muestrales, donde podemos notar que los ESR posee un rendimiento similar a los ESC, pero con la ventaja de que no depende de la finitud de los EMV.

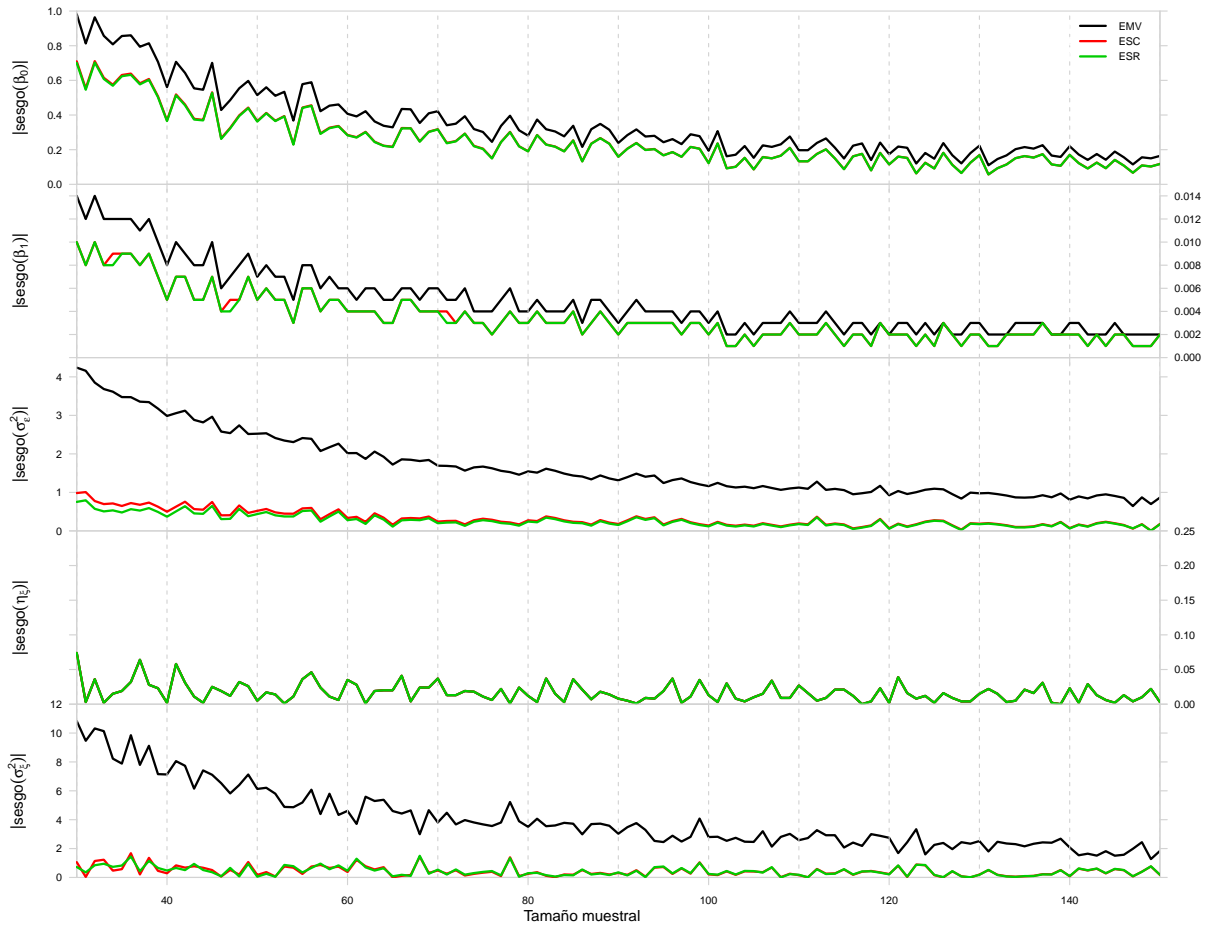


Figura 4.1: Sesgo promedio (en valor absoluto) para diferentes tamaños muestrales. Con valores verdaderos de los parámetros $\theta_{\text{Fuller}} = (67; 0,42; 43; 70; 247; 57)^\top$.

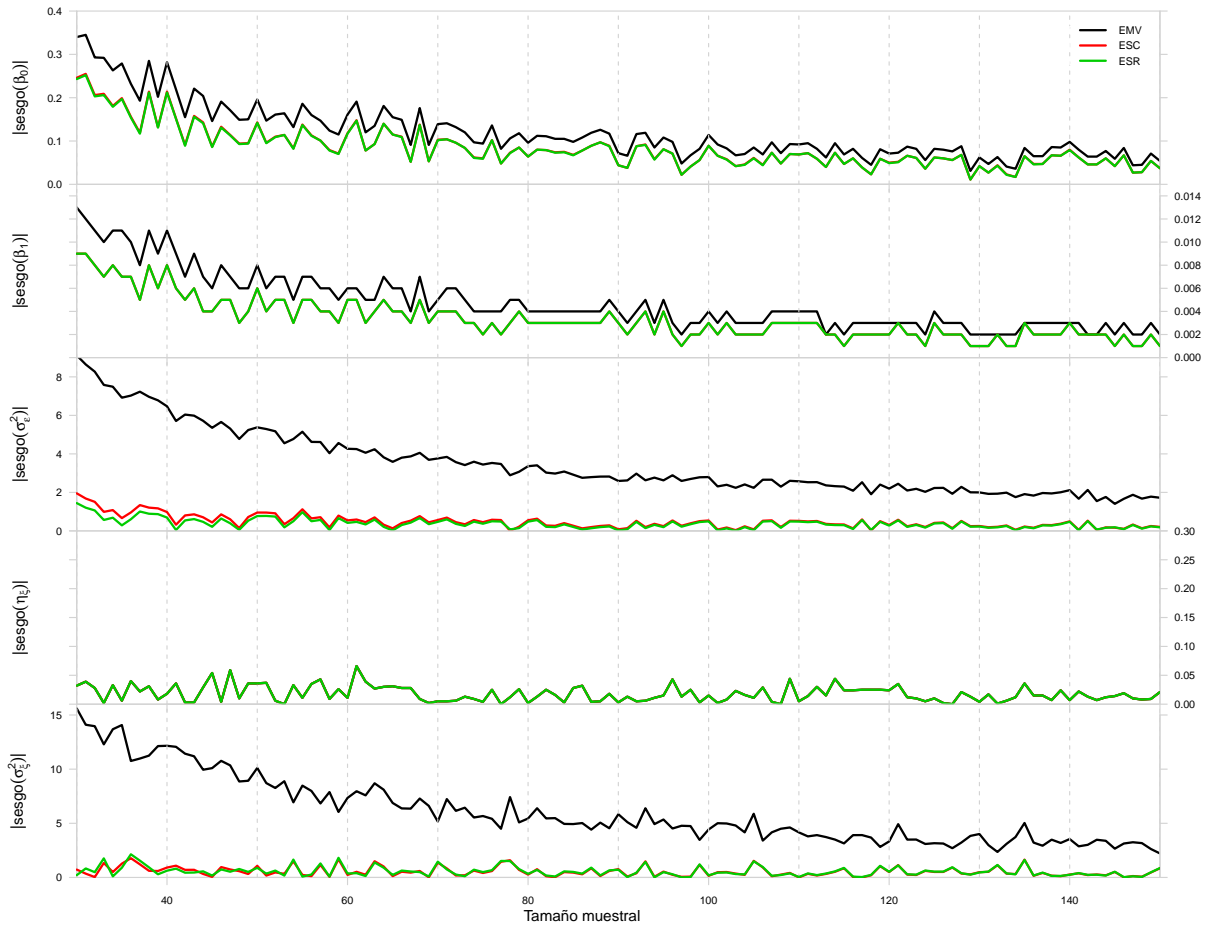


Figura 4.2: Sesgo promedio (en valor absoluto) para diferentes tamaños muestrales. Con valores verdaderos de los parámetros $\theta_{\text{Patriota}} = (76; 0,61; 93; 25; 398; 64)^\top$.

Tabla 4.1: Sesgo relativo y varianza relativa de los ESC y ESR. Con valores verdaderos de los parámetros $\theta_{\text{Fuller}} = (67; 0,42; 43; 70; 247; 57)^\top$.

n	θ	EMV	ESC		ESR	
		Sesgo rel.	Sesgo rel.	Varianza rel.	Sesgo rel.	Varianza rel.
25	β_0	0,0213	0,0161	0,0556	0,0159	0,0576
	β_1	0,0480	0,0362	0,0580	0,0357	0,0601
	σ_ϵ^2	0,1139	0,0235	0,1433	0,0157	0,1565
	η_ξ	0,0003	0,0003	0,0000	0,0003	0,0000
	σ_ξ^2	0,0506	0,0034	0,0816	0,0015	0,0851
35	β_0	0,0117	0,0083	0,0240	0,0083	0,0247
	β_1	0,0274	0,0198	0,0246	0,0196	0,0253
	σ_ϵ^2	0,0835	0,0180	0,1057	0,0140	0,1124
	η_ξ	0,0001	0,0001	0,0000	0,0001	0,0000
	σ_ξ^2	0,0348	0,0007	0,0580	0,0003	0,0597
50	β_0	0,0075	0,0052	0,0151	0,0052	0,0154
	β_1	0,0168	0,0117	0,0155	0,0116	0,0158
	σ_ϵ^2	0,0581	0,0115	0,0734	0,0096	0,0766
	η_ξ	0,0001	0,0001	0,0000	0,0001	0,0000
	σ_ξ^2	0,0280	0,0040	0,0404	0,0035	0,0412
100	β_0	0,0037	0,0026	0,0068	0,0026	0,0068
	β_1	0,0084	0,0059	0,0070	0,0059	0,0071
	σ_ϵ^2	0,0290	0,0052	0,0369	0,0047	0,0376
	η_ξ	0,0002	0,0002	0,0000	0,0002	0,0000
	σ_ξ^2	0,0131	0,0009	0,0201	0,0008	0,0203
150	β_0	0,0027	0,0020	0,0043	0,0020	0,0043
	β_1	0,0059	0,0043	0,0044	0,0043	0,0045
	σ_ϵ^2	0,0196	0,0037	0,0245	0,0034	0,0248
	η_ξ	0,0001	0,0001	0,0000	0,0001	0,0000
	σ_ξ^2	0,0071	0,0010	0,0134	0,0011	0,0135
200	β_0	0,0024	0,0019	0,0032	0,0019	0,0032
	β_1	0,0056	0,0044	0,0033	0,0044	0,0033
	σ_ϵ^2	0,0168	0,0048	0,0184	0,0047	0,0186
	η_ξ	0,0001	0,0001	0,0000	0,0001	0,0000
	σ_ξ^2	0,0054	0,0007	0,0100	0,0008	0,0101
250	β_0	0,0019	0,0015	0,0025	0,0015	0,0025
	β_1	0,0044	0,0034	0,0026	0,0034	0,0026
	σ_ϵ^2	0,0105	0,0009	0,0147	0,0008	0,0148
	η_ξ	0,0002	0,0002	0,0000	0,0002	0,0000
	σ_ξ^2	0,0041	0,0008	0,0080	0,0008	0,0080

Tabla 4.2: Sesgo relativo y varianza relativa de los ESC y ESR. Con valores verdaderos de los parámetros $\theta_{\text{Patriota}} = (76; 0,61; 93; 25; 398; 64)^\top$.

n	θ	EMV	ESC		ESR	
		Sesgo rel.	Sesgo rel.	Varianza rel.	Sesgo rel.	Varianza rel.
25	β_0	0,0056	0,0040	0,0210	0,0040	0,0218
	β_1	0,0301	0,0223	0,0275	0,0220	0,0286
	σ_ϵ^2	0,1106	0,0186	0,1518	0,0107	0,1656
	η_ξ	0,0016	0,0016	0,0000	0,0016	0,0000
	σ_ξ^2	0,0537	0,0094	0,0816	0,0076	0,0851
35	β_0	0,0034	0,0024	0,0124	0,0023	0,0128
	β_1	0,0175	0,0123	0,0169	0,0122	0,0174
	σ_ϵ^2	0,0775	0,0105	0,1088	0,0064	0,1156
	η_ξ	0,0026	0,0026	0,0000	0,0026	0,0000
	σ_ξ^2	0,0356	0,0035	0,0580	0,0025	0,0597
50	β_0	0,0026	0,0019	0,0079	0,0019	0,0080
	β_1	0,0126	0,0091	0,0109	0,0090	0,0111
	σ_ϵ^2	0,0560	0,0083	0,0759	0,0063	0,0791
	η_ξ	0,0006	0,0006	0,0000	0,0006	0,0000
	σ_ξ^2	0,0219	0,0009	0,0404	0,0013	0,0412
100	β_0	0,0012	0,0008	0,0035	0,0008	0,0035
	β_1	0,0055	0,0038	0,0048	0,0038	0,0049
	σ_ϵ^2	0,0251	0,0008	0,0379	0,0003	0,0387
	η_ξ	0,0006	0,0006	0,0000	0,0006	0,0000
	σ_ξ^2	0,0080	0,0036	0,0201	0,0037	0,0203
150	β_0	0,0007	0,0005	0,0022	0,0005	0,0022
	β_1	0,0035	0,0024	0,0031	0,0024	0,0032
	σ_ϵ^2	0,0203	0,0040	0,0252	0,0038	0,0256
	η_ξ	0,0003	0,0003	0,0000	0,0003	0,0000
	σ_ξ^2	0,0066	0,0011	0,0134	0,0011	0,0135
200	β_0	0,0007	0,0005	0,0017	0,0005	0,0017
	β_1	0,0035	0,0027	0,0024	0,0027	0,0024
	σ_ϵ^2	0,0147	0,0025	0,0189	0,0023	0,0191
	η_ξ	0,0010	0,0010	0,0000	0,0010	0,0000
	σ_ξ^2	0,0086	0,0028	0,0100	0,0028	0,0101
250	β_0	0,0004	0,0002	0,0013	0,0002	0,0013
	β_1	0,0018	0,0012	0,0019	0,0012	0,0019
	σ_ϵ^2	0,0104	0,0006	0,0151	0,0005	0,0152
	η_ξ	0,0006	0,0006	0,0000	0,0006	0,0000
	σ_ξ^2	0,0053	0,0007	0,0080	0,0007	0,0080

Tabla 4.3: Cobertura empírica y longitud promedio del intervalo de confianza del 95 %.

	n	Cobertura empírica			Longitud promedio		
		EMV	ESC	ESR	EMV	ESC	ESR
$\beta_1 = 0,42$	25	0,9462	0,9474	0,9474	0,4903	0,4798	0,4794
	35	0,9482	0,9482	0,9481	0,3892	0,3843	0,3841
	50	0,9471	0,9474	0,9475	0,3173	0,3148	0,3147
	100	0,9503	0,9501	0,9501	0,2146	0,2139	0,2138
	150	0,9462	0,9469	0,9469	0,1697	0,1693	0,1693
	200	0,9471	0,9469	0,9469	0,1465	0,1462	0,1462
	250	0,9509	0,9509	0,9509	0,1313	0,1311	0,1311
$\beta_1 = 0,61$	25	0,9482	0,9496	0,9495	0,5272	0,5199	0,5196
	35	0,9470	0,9475	0,9475	0,4256	0,4221	0,4220
	50	0,9462	0,9462	0,9462	0,3415	0,3398	0,3397
	100	0,9468	0,9470	0,9470	0,2371	0,2365	0,2365
	150	0,9493	0,9497	0,9497	0,1924	0,1921	0,1921
	200	0,9503	0,9504	0,9504	0,1659	0,1657	0,1657
	250	0,9481	0,9477	0,9477	0,1467	0,1465	0,1465

4.4.2. Modelo estructural t

Como se ha mencionado en el capítulo 3, la distribución t es un miembro particular de la familia de distribuciones elípticas, donde se ha mostrado que esta distribución es una alternativa a la distribución normal, especialmente cuando se desea implementar un procedimiento robusto de estimación.

En esta sección, a modo de ilustración, se implementarán los algoritmos de reducción y corrección de sesgo, al modelo de regresión lineal con error de medida simple, donde asumimos que σ_δ^2 y ν son conocidos. Debemos notar, que conocer el valor de ν no es un supuesto muy fuerte para el propósito de este trabajo.

Lange et al. (1989) sugieren que ν puede ser considerado conocido para conjuntos de datos pequeños y debe ser estimado para conjuntos de datos relativamente grandes. En otras palabras, las muestras pequeñas no proveen (típicamente) suficiente información sobre ν para que este parámetro pueda ser estimado con suficiente precisión.

Además, como nuestra principal preocupación está en el análisis de conjuntos de datos pequeños, donde la corrección y reducción de sesgo es particularmente importante, es razonable fijar ν en un valor pequeño. El supuesto de conocer ν que se adopta en esta sección también es considerado en Ferrari & Arellano-Valle (1996), Cordeiro et al. (1998) y Vasconcellos & Cordeiro (2000).

Consideremos la situación en donde n variables aleatorias independientes $\mathbf{Z}_i = [X_i, Y_i]^\top$, para $i = 1, 2, \dots, n$, son observadas. El modelo con medida de error simple es definido en (1.1), con

$$\mathbf{r}_i = \begin{bmatrix} \epsilon_i \\ \delta_i \\ \xi_i \end{bmatrix} \stackrel{\text{ind.}}{\sim} t_3 \left(\begin{bmatrix} 0 \\ 0 \\ \eta_\xi \end{bmatrix}, \begin{bmatrix} \sigma_\epsilon^2 & 0 & 0 \\ 0 & \sigma_\delta^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{bmatrix}, \nu \right), i = 1, 2, \dots, n.$$

De igual forma que en el caso normal analizado previamente, de (3.5), el modelo reparametrizado es $Y_i = \beta_0^* + \beta_1^* X_i + \epsilon_i^*$, donde

$$\epsilon_i^* \sim t(0, \sigma_{\epsilon^*}^2, \nu) \quad \text{y} \quad X_i \sim t(\eta_\xi, \sigma_X^2, \nu),$$

no-correlacionados. Del modelo reparametrizado, $\mathbf{Z}_i = [X_i, Y_i]^\top \sim t_2(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, \nu)$ con

$$\boldsymbol{\mu}^* = \begin{bmatrix} \eta_\xi \\ \beta_0^* + \beta_1^* \eta_\xi \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Sigma}^* = \begin{bmatrix} \sigma_X^2 & \beta_1^* \sigma_X^2 \\ \beta_1^* \sigma_X^2 & \sigma_{\epsilon^*}^2 + \beta_1^{*2} \sigma_X^2 \end{bmatrix},$$

la función de log-verosimilitud asociada a $\mathbf{Z} = [\mathbf{Z}_1^\top, \dots, \mathbf{Z}_n^\top]^\top$ es dada por

$$\ell^* \propto -\frac{n}{2} \log \sigma_{\epsilon^*}^2 - \frac{n}{2} \log \sigma_X^2 - \frac{\nu + 2}{2} \sum_{i=1}^n \log \left(1 + \frac{v_i}{\nu} \right),$$

con $v_i = (x_i - \eta_\xi)^2 / \sigma_X^2 + \epsilon_i^{*2} / \sigma_{\epsilon^*}^2$, y $\epsilon_i^* = y_i - \beta_0^* - \beta_1^* x_i$.

Sea la partición del vector de parámetros $\boldsymbol{\theta}^* = [\boldsymbol{\theta}_1^{*\top}, \boldsymbol{\theta}_2^{*\top}]^\top$ donde $\boldsymbol{\theta}_1^* = [\beta_0^*, \beta_1^*, \sigma_{\epsilon^*}^2]^\top$ y $\boldsymbol{\theta}_2^* = [\eta_\xi, \sigma_\xi^2]^\top$, se define el vector de score como

$$\mathbf{S}(\boldsymbol{\theta}^*) = [\mathbf{S}_{\boldsymbol{\theta}_1^*}^\top, \mathbf{S}_{\boldsymbol{\theta}_2^*}^\top]^\top,$$

donde $\mathbf{S}_{\boldsymbol{\theta}_1^*} = [S_{\beta_0^*}, S_{\beta_1^*}, S_{\sigma_{\epsilon^*}^2}]^\top$ tiene elementos dados por

$$S_{\beta_0^*} = \frac{1}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n \omega_i \epsilon_i^*, \quad S_{\beta_1^*} = \frac{1}{\sigma_{\epsilon^*}^2} \sum_{i=1}^n \omega_i x_i \epsilon_i^* \quad \text{y} \quad S_{\sigma_{\epsilon^*}^2} = -\frac{n}{2\sigma_{\epsilon^*}^2} + \frac{1}{2(\sigma_{\epsilon^*}^2)^2} \sum_{i=1}^n \omega_i \epsilon_i^{*2},$$

y $\mathbf{S}_{\boldsymbol{\theta}_2^*} = [S_{\eta_\xi}, S_{\sigma_\xi^2}]^\top$ con elementos

$$S_{\eta_\xi} = \frac{1}{\sigma_X^2} \sum_{i=1}^n \omega_i (x_i - \eta_\xi) \quad \text{y} \quad S_{\sigma_\xi^2} = -\frac{n}{2\sigma_X^2} + \frac{1}{2(\sigma_X^2)^2} \sum_{i=1}^n \omega_i (x_i - \eta_\xi)^2,$$

donde

$$\omega_i = \frac{\nu + 2}{\nu + v_i}.$$

Entonces, considerando que

$$W_g^{(v_i)}(v_i) = -\frac{\omega_i}{2} \quad \text{y} \quad W_g^{(v_i, v_i)}(v_i) = \frac{\omega_i}{2(\nu + v_i)},$$

y de la Proposición 3.2 y de la sección 3.9, se puede definir los elementos de la matriz de información esperada, definida por $\mathbf{F}(\boldsymbol{\theta}^*)$.

Dado que en este caso no se tienen soluciones explícitas para determinar los EMV, ESC y ESR, se implementan en cada caso, diferentes versiones del algoritmo Fisher-Scoring. En el caso de los EMV, se implementa el algoritmo Fisher-Scoring usual, es decir

$$\boldsymbol{\theta}^{*(j+1)} = \boldsymbol{\theta}^{*(j)} + \mathbf{F}^{-1}(\boldsymbol{\theta}^{*(j)})\mathbf{S}(\boldsymbol{\theta}^{*(j)}),$$

donde $\mathbf{F}^{-1}(\cdot)$ y $\mathbf{S}(\cdot)$, corresponden a la inversa de la matriz de información esperada y el vector de score, respectivamente.

Por otro lado, para determinar los ESC y ESR, es necesario determinar el vector de modificaciones $\mathbf{M}(\boldsymbol{\theta}^*)$. Después de un extenso algebra, de (4.7), el vector de modificaciones viene dado por

$$\mathbf{M}(\boldsymbol{\theta}^*) = \left[0, 0, \frac{m_1(\nu)}{2\sigma_{\epsilon^*}^2}, 0, \frac{m_2(\nu)}{2\sigma_X^2} \right]^\top,$$

donde

$$m_1(\nu) = \frac{2\nu^4 + 5\nu^3 - 7\nu^2 - 32\nu - 36}{\nu^2(\nu + 2)(\nu + 6)},$$

$$m_2(\nu) = \frac{\nu^4 - 3\nu^3 - 19\nu^2 - 32\nu - 36}{\nu^2(\nu + 2)(\nu + 6)}.$$

Definiendo los EVM de $\boldsymbol{\theta}^*$ como $\hat{\boldsymbol{\theta}}^*$, los ESC viene dados por

$$\hat{\boldsymbol{\theta}}_{SC}^* = \hat{\boldsymbol{\theta}}^* - \mathbf{b}(\hat{\boldsymbol{\theta}}^*),$$

donde el vector de sesgo viene dado por

$$\mathbf{b}(\boldsymbol{\theta}^*) = \left[0, 0, -\frac{\sigma_{\epsilon^*}^2 b_1(\nu)}{n}, 0, -\frac{\sigma_X^2 b_2(\nu)}{n} \right]^\top,$$

con

$$b_1(\nu) = \frac{(\nu + 4)(2\nu^4 + 4\nu^3 - 13\nu^2 - 32\nu - 36)}{\nu^3(\nu + 2)(\nu + 6)},$$

$$b_2(\nu) = \frac{(\nu + 4)(\nu^4 - 2\nu^3 - 13\nu^2 - 32\nu - 36)}{\nu^3(\nu + 2)(\nu + 6)}.$$

Para determinar los ESR, se aplica el algoritmo Fisher-Scoring modificado definido en la sección 4.3, el cual viene dado por

$$\boldsymbol{\theta}^{*(j+1)} = \hat{\boldsymbol{\theta}}^{*(j+1)} - \mathbf{b}(\boldsymbol{\theta}^{*(j)}),$$

Donde $\hat{\boldsymbol{\theta}}^{*(j+1)}$ se define como el valor candidato para los EMV que son obtenidos en un paso del algoritmo Fisher-Scoring de $\boldsymbol{\theta}^{*(j)}$. Por lo tanto, a través de la parametrización dada por (3.6), (3.7) y (3.10), es fácil determinar las estimaciones bajo la parametrización directa.

Los resultados de la simulación Monte Carlo son presentados y discutidos a continuación, donde se observa que la corrección de sesgo derivada es efectiva en pequeñas muestras, tanto con el método de reducción como el de corrección de sesgo.

Estudio de simulación

Al igual que en el caso normal, es este estudio numérico usamos simulación Monte Carlo para evaluar el rendimiento de los estimadores de MV y sus respectivas versiones reducidas y corregidas en muestras finitas.

Para analizar los resultados de la estimación puntual, calculamos para cada tamaño muestral y para cada estimador: el sesgo, el sesgo relativo y la varianza relativa del estimador.

Los valores verdaderos de los parámetros de la regresión (Cheng & Tsai (2004)) son

$$\boldsymbol{\theta} = (1,41; 1,08; 4,11; 20,86; 20,13; 4,6)^\top.$$

Se consideraron tres valores diferentes para los grados de libertad, $\nu = 5, 10$ y 15 . Las Figuras 4.3-4.5, muestra el comportamiento del sesgo promedio (en valor absoluto). Los tamaños muestrales considerados son de $n = 30$ a $n = 150$, y para cada tamaño muestral se realizaron 10 000 replicaciones. A partir de estos resultados, se puede observar que los métodos de reducción y corrección de sesgo son bastante estables y presentan un sesgo menor o similar en comparación con los estimadores de

MV. Además, en particular la corrección de los parámetros asociados al modelo de regresión es, es muy evidente con ambos métodos.

Por otro lado, de las Tablas 4.4-4.6, podemos ver algunos resultados adicionales obtenidos a través de la simulación donde podemos destacar, que al igual que en el caso normal, la ganancia respecto al sesgo relativo de los métodos de corrección de sesgo y reducción de sesgo es significativo. Así, la varianza relativa muestra que la aplicación de estos métodos aumenta la variabilidad de las estimaciones, pero disminuye a medida que incrementamos el tamaño muestral.

Una consideración importante que debemos hacer notar, respecto a las simulación realizada es que puede existir problemas de convergencia cuando se maximiza la función log-verosimilitud por medio del algoritmo Fisher-Scoring o su versión modificada, debido a la inestabilidad numérica de esta función.

Para la optimización, el límite obtenido en cada maximización debe corresponder a un máximo local, ya que no podemos garantizar la maximización global, como es notado en Fernandez & Steel (1999). Como se esperaba, se puede notar empíricamente que la intensidad de este problema decrece con n , es decir, cuando el tamaño muestral aumenta.

Para muestras de tamaño $n < 100$, se tuvieron problemas de convergencia aproximadamente el 2,8 % de las veces. Para $n \geq 100$, el problema fue menos frecuente, y para $n \geq 200$, la inestabilidad numérica prácticamente desaparece.

La Tabla 4.7, muestra los resultados de un estudio de simulación para verificar la cobertura empírica y la longitud promedio del intervalo de confianza asintótico del 95 % para $\beta_1 = 1,08$, considerando $\nu = 5, 10$ y 15 .

Los intervalos son definidos como $\hat{\beta}_1 \pm 1,96 \times \text{es}(\hat{\beta}_1)$ (para el estimador de MV), y $\hat{\beta}_{1SC} \pm 1,96 \times \text{es}(\hat{\beta}_{1SC})$ y $\hat{\beta}_{1SR} \pm 1,96 \times \text{es}(\hat{\beta}_{1SR})$, para los respectivos estimadores con sesgo corregido y sesgo reducido, donde $\text{es}(\cdot)$ denota a la estimación del error estándar por el uso del respectivo estimador.

De los resultados obtenidos podemos notar que los ESC y ESR no presentan ganancia significativa en comparación con los correspondientes EMV, tanto en la cobertura empírica como en la longitud promedio. Además, para diferentes tamaños muestrales, podemos notar que los ESR posee un rendimiento similar a los ESC.

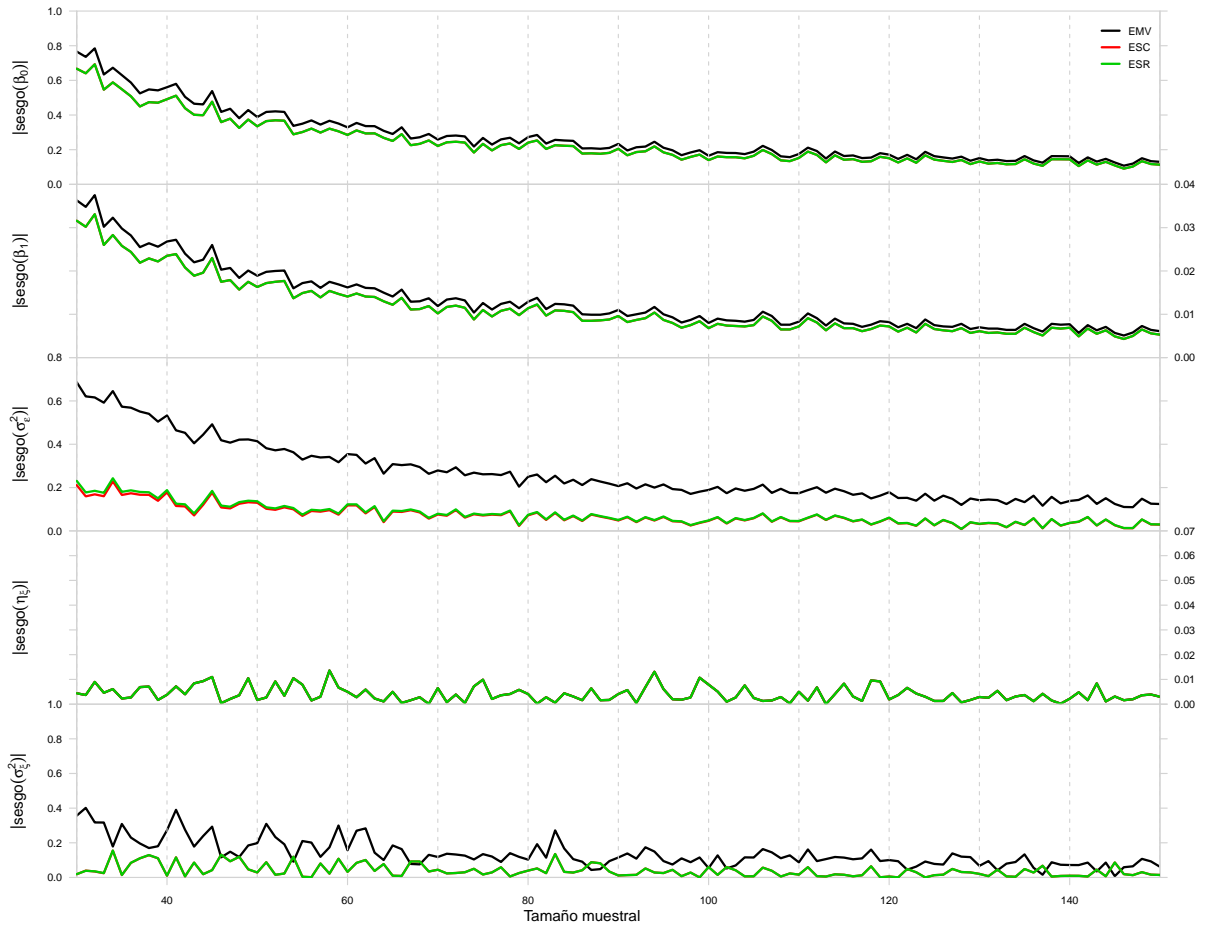


Figura 4.3: Sesgo promedio (en valor absoluto) para diferentes tamaños muestrales. Con valores verdaderos de los parámetros θ y ($\nu = 5$).

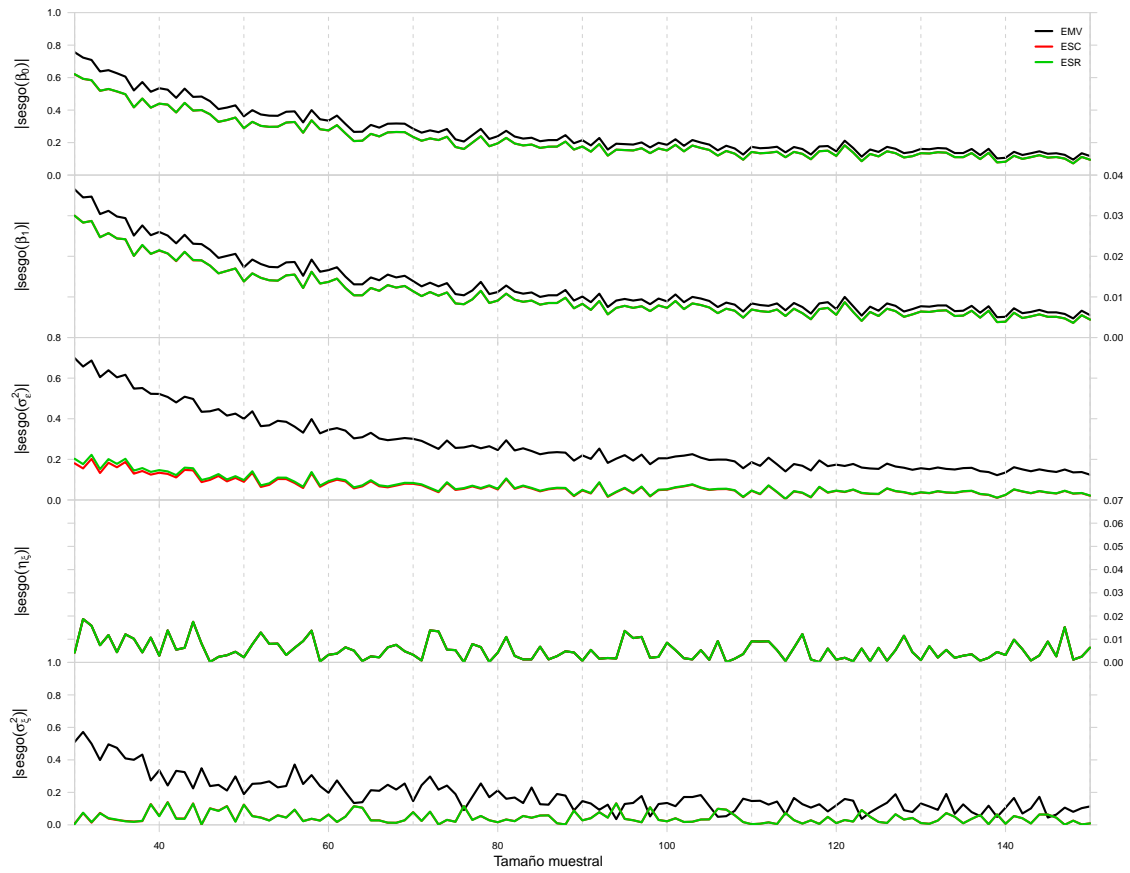


Figura 4.4: Sesgo promedio (en valor absoluto) para diferentes tamaños muestrales. Con valores verdaderos de los parámetros θ y ($\nu = 10$).

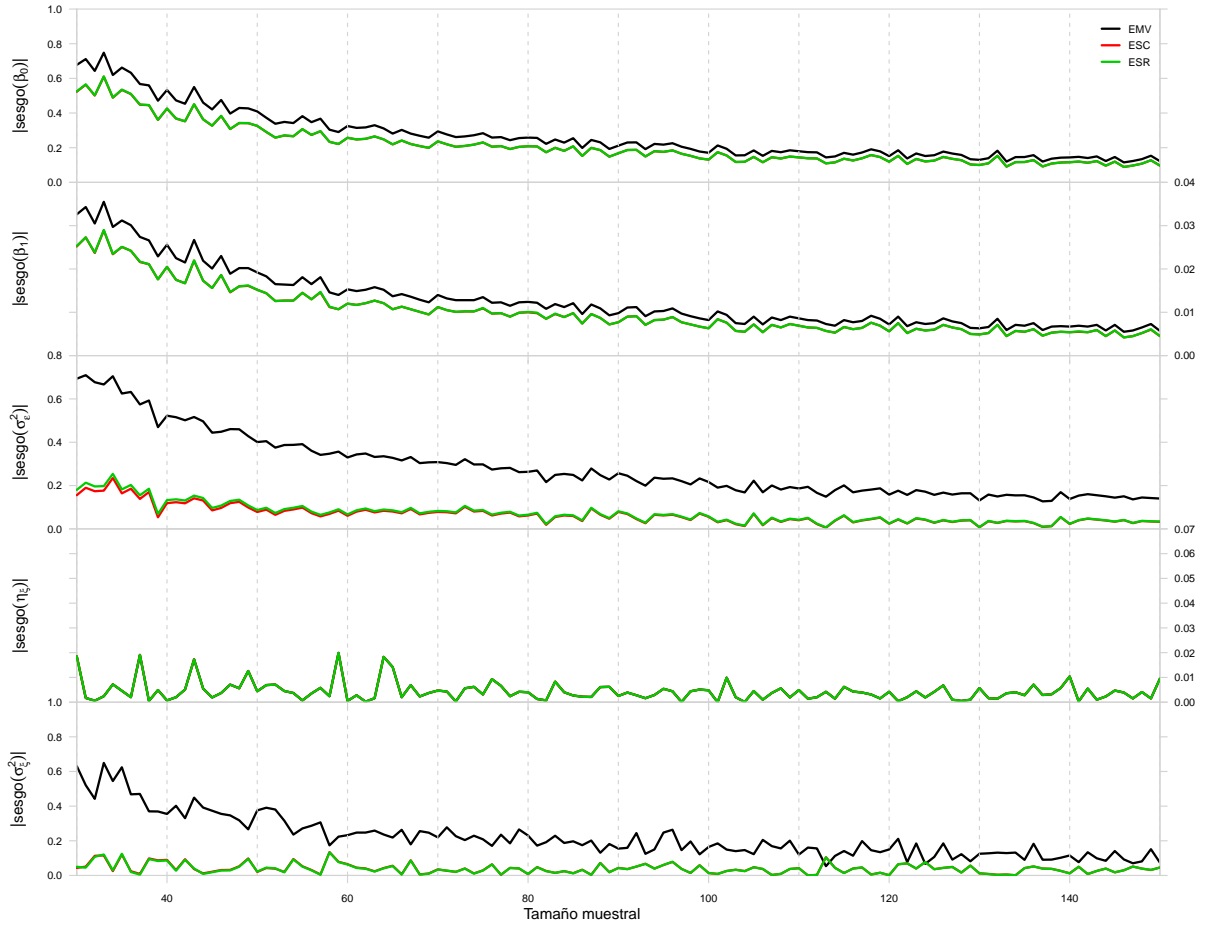


Figura 4.5: Sesgo promedio (en valor absoluto) para diferentes tamaños muestrales. Con valores verdaderos de los parámetros θ y ($\nu = 15$).

Tabla 4.4: Sesgo relativo y varianza relativa de los ESC y ESR. Con valores verdaderos de los parámetros $\theta = (1,41; 1,08; 4,11; 20,86; 20,13; 4,6)^\top$ y $\nu = 5$.

n	θ	EMV	ESC		ESR	
		Sesgo rel.	Sesgo rel.	Varianza rel.	Sesgo rel.	Varianza rel.
25	β_0	0,6257	0,5127	0,9206	0,5085	1,1420
	β_1	0,0391	0,0320	0,9302	0,0318	1,1540
	σ_ϵ^2	0,1951	0,0545	0,2773	0,0609	0,3176
	η_ξ	0,0005	0,0005	0,0001	0,0005	0,0000
	σ_ξ^2	0,0168	0,0056	0,0372	0,0056	0,0373
35	β_0	0,4294	0,3719	0,0175	0,3719	0,0173
	β_1	0,0266	0,0230	0,0178	0,0230	0,0177
	σ_ϵ^2	0,1379	0,0389	0,0719	0,0423	0,0689
	η_ξ	0,0001	0,0001	0,0000	0,0001	0,0000
	σ_ξ^2	0,0139	0,0021	0,0267	0,0021	0,0266
50	β_0	0,2900	0,2521	0,0112	0,2520	0,0113
	β_1	0,0188	0,0164	0,0114	0,0164	0,0115
	σ_ϵ^2	0,0951	0,0257	0,0511	0,0274	0,0497
	η_ξ	0,0002	0,0002	0,0000	0,0002	0,0000
	σ_ξ^2	0,0062	0,0051	0,0186	0,0051	0,0186
100	β_0	0,1386	0,1208	0,0048	0,1208	0,0048
	β_1	0,0086	0,0075	0,0049	0,0075	0,0049
	σ_ϵ^2	0,0460	0,0113	0,0261	0,0117	0,0258
	η_ξ	0,0006	0,0006	0,0000	0,0006	0,0000
	σ_ξ^2	0,0038	0,0019	0,0093	0,0019	0,0093
150	β_0	0,0856	0,0740	0,0032	0,0740	0,0032
	β_1	0,0054	0,0046	0,0033	0,0046	0,0032
	σ_ϵ^2	0,0319	0,0088	0,0172	0,0090	0,0171
	η_ξ	0,0001	0,0001	0,0000	0,0001	0,0000
	σ_ξ^2	0,0021	0,0017	0,0062	0,0017	0,0062
200	β_0	0,0752	0,0666	0,0023	0,0665	0,0023
	β_1	0,0048	0,0042	0,0024	0,0042	0,0024
	σ_ϵ^2	0,0322	0,0149	0,0128	0,0150	0,0127
	η_ξ	0,0001	0,0001	0,0000	0,0001	0,0000
	σ_ξ^2	0,0045	0,0016	0,0047	0,0016	0,0046
250	β_0	0,0531	0,0462	0,0018	0,0462	0,0018
	β_1	0,0034	0,0030	0,0018	0,0030	0,0018
	σ_ϵ^2	0,0190	0,0051	0,0104	0,0052	0,0103
	η_ξ	0,0001	0,0001	0,0000	0,0001	0,0000
	σ_ξ^2	0,0039	0,0016	0,0037	0,0016	0,0037

Tabla 4.5: Sesgo relativo y varianza relativa de los ESC y ESR. Con valores verdaderos de los parámetros $\theta = (1,41; 1,08; 4,11; 20,86; 20,13; 4,6)^\top$ y $\nu = 10$.

n	θ	EMV	ESC		ESR	
		Sesgo rel.	Sesgo rel.	Varianza rel.	Sesgo rel.	Varianza rel.
25	β_0	0,6574	0,5351	0,0462	0,5358	0,0459
	β_1	0,0417	0,0340	0,0473	0,0341	0,0469
	σ_ϵ^2	0,2042	0,0527	0,1057	0,0606	0,0992
	η_ξ	0,0003	0,0003	0,0001	0,0003	0,0000
	σ_ξ^2	0,0306	0,0001	0,0522	0,0001	0,0517
35	β_0	0,4492	0,3700	0,0249	0,3704	0,0249
	β_1	0,0284	0,0235	0,0254	0,0235	0,0255
	σ_ϵ^2	0,1611	0,0539	0,0771	0,0580	0,0735
	η_ξ	0,0003	0,0003	0,0000	0,0003	0,0000
	σ_ξ^2	0,0169	0,0053	0,0371	0,0052	0,0368
50	β_0	0,2658	0,2135	0,0145	0,2137	0,0144
	β_1	0,0167	0,0134	0,0148	0,0135	0,0147
	σ_ϵ^2	0,0943	0,0187	0,0542	0,0207	0,0525
	η_ξ	0,0005	0,0005	0,0000	0,0005	0,0000
	σ_ξ^2	0,0166	0,0010	0,0258	0,0011	0,0257
100	β_0	0,1638	0,1391	0,0070	0,1391	0,0069
	β_1	0,0103	0,0088	0,0071	0,0088	0,0070
	σ_ϵ^2	0,0544	0,0167	0,0274	0,0172	0,0269
	η_ξ	0,0003	0,0003	0,0000	0,0003	0,0000
	σ_ξ^2	0,0077	0,0001	0,0128	0,0001	0,0128
150	β_0	0,0874	0,0713	0,0042	0,0713	0,0042
	β_1	0,0055	0,0044	0,0043	0,0044	0,0043
	σ_ϵ^2	0,0319	0,0067	0,0184	0,0069	0,0182
	η_ξ	0,0005	0,0005	0,0000	0,0005	0,0000
	σ_ξ^2	0,0053	0,0001	0,0085	0,0001	0,0085
200	β_0	0,0691	0,0571	0,0031	0,0571	0,0031
	β_1	0,0043	0,0035	0,0032	0,0035	0,0032
	σ_ϵ^2	0,0267	0,0078	0,0137	0,0080	0,0136
	η_ξ	0,0002	0,0002	0,0000	0,0002	0,0000
	σ_ξ^2	0,0065	0,0026	0,0064	0,0026	0,0064
250	β_0	0,0570	0,0475	0,0024	0,0475	0,0024
	β_1	0,0035	0,0029	0,0025	0,0029	0,0025
	σ_ϵ^2	0,0201	0,0050	0,0110	0,0051	0,0109
	η_ξ	0,0000	0,0000	0,0000	0,0000	0,0000
	σ_ξ^2	0,0022	0,0009	0,0051	0,0009	0,0051

Tabla 4.6: Sesgo relativo y varianza relativa de los ESC y ESR. Con valores verdaderos de los parámetros $\theta = (1,41; 1,08; 4,11; 20,86; 20,13; 4,6)^\top$ y $\nu = 15$.

n	θ	EMV	ESC		ESR	
		Sesgo rel.	Sesgo rel.	Varianza rel.	Sesgo rel.	Varianza rel.
25	β_0	0,6627	0,5226	0,0544	0,5245	0,0534
	β_1	0,0416	0,0328	0,0552	0,0330	0,0543
	σ_ϵ^2	0,2195	0,0627	0,1062	0,0714	0,0994
	η_ξ	0,0005	0,0005	0,0000	0,0005	0,0000
	σ_ξ^2	0,0312	0,0040	0,0595	0,0035	0,0588
35	β_0	0,4457	0,3554	0,0280	0,3563	0,0276
	β_1	0,0277	0,0220	0,0286	0,0221	0,0282
	σ_ϵ^2	0,1465	0,0342	0,0788	0,0387	0,0753
	η_ξ	0,0000	0,0000	0,0000	0,0000	0,0000
	σ_ξ^2	0,0252	0,0001	0,0423	0,0003	0,0418
50	β_0	0,2945	0,2349	0,0168	0,2353	0,0165
	β_1	0,0184	0,0147	0,0171	0,0147	0,0168
	σ_ϵ^2	0,1079	0,0296	0,0564	0,0318	0,0546
	η_ξ	0,0004	0,0004	0,0000	0,0004	0,0000
	σ_ξ^2	0,0194	0,0018	0,0294	0,0019	0,0292
100	β_0	0,1335	0,1056	0,0074	0,1057	0,0073
	β_1	0,0081	0,0064	0,0076	0,0064	0,0075
	σ_ϵ^2	0,0493	0,0101	0,0284	0,0107	0,0280
	η_ξ	0,0000	0,0000	0,0000	0,0000	0,0000
	σ_ξ^2	0,0083	0,0006	0,0146	0,0005	0,0145
150	β_0	0,0818	0,0635	0,0047	0,0636	0,0047
	β_1	0,0051	0,0040	0,0048	0,0040	0,0048
	σ_ϵ^2	0,0332	0,0070	0,0190	0,0073	0,0188
	η_ξ	0,0001	0,0001	0,0000	0,0001	0,0000
	σ_ξ^2	0,0035	0,0024	0,0097	0,0024	0,0097
200	β_0	0,0570	0,0435	0,0035	0,0435	0,0035
	β_1	0,0036	0,0027	0,0036	0,0027	0,0036
	σ_ϵ^2	0,0269	0,0073	0,0141	0,0075	0,0140
	η_ξ	0,0001	0,0001	0,0000	0,0001	0,0000
	σ_ξ^2	0,0030	0,0015	0,0073	0,0015	0,0073
250	β_0	0,0612	0,0504	0,0028	0,0504	0,0028
	β_1	0,0037	0,0030	0,0029	0,0030	0,0029
	σ_ϵ^2	0,0203	0,0047	0,0114	0,0047	0,0113
	η_ξ	0,0000	0,0000	0,0000	0,0000	0,0000
	σ_ξ^2	0,0043	0,0007	0,0058	0,0007	0,0058

Tabla 4.7: Cobertura empírica y longitud promedio del intervalo de confianza del 95 %.

		Cobertura empírica			Longitud promedio		
		EMV	ESC	ESR	EMV	ESC	ESR
$\nu = 5$	25	0,9560	0,9560	0,9560	0,9717	0,9502	0,9501
	35	0,9410	0,9400	0,9400	0,6610	0,6549	0,6549
	50	0,9490	0,9490	0,9480	0,5094	0,5063	0,5062
	100	0,9450	0,9450	0,9450	0,3521	0,3513	0,3513
	150	0,9550	0,9560	0,9560	0,2849	0,2844	0,2844
	200	0,9430	0,9430	0,9430	0,2582	0,2579	0,2579
	250	0,9570	0,9570	0,9570	0,2128	0,2126	0,2126
$\nu = 10$	25	0,9480	0,9490	0,9490	0,8452	0,8276	0,8277
	35	0,9540	0,9520	0,9520	0,6770	0,6676	0,6677
	50	0,9510	0,9530	0,9530	0,5380	0,5334	0,5335
	100	0,9460	0,9490	0,9490	0,3547	0,3535	0,3535
	150	0,9510	0,9510	0,9520	0,2828	0,2822	0,2822
	200	0,9550	0,9530	0,9530	0,2420	0,2416	0,2416
	250	0,9420	0,9430	0,9430	0,2181	0,2178	0,2178
$\nu = 15$	25	0,9370	0,9370	0,9370	0,8156	0,7959	0,7959
	35	0,9430	0,9460	0,9460	0,6319	0,6232	0,6232
	50	0,9440	0,9470	0,9470	0,5055	0,5009	0,5011
	100	0,9500	0,9530	0,9520	0,3468	0,3455	0,3455
	150	0,9460	0,9460	0,9460	0,2824	0,2817	0,2817
	200	0,9510	0,9490	0,9490	0,2416	0,2412	0,2412
	250	0,9540	0,9550	0,9550	0,2164	0,2161	0,2161

4.4.3. Aplicación

Caso simple

Como un chequeo del efecto de los algoritmo de reducción y corrección, son ajustados tres conjuntos de datos con el modelo de regresión lineal simple con error de medida. El primero de ellos ha sido estudiado en Cheng & Tsai (2004) y previamente analizado por Kelly (1984) y Miller (1986). El conjunto de datos comprende medidas de los niveles de suero de kanamicina en muestras de sangre de 20 individuos.

La covariable, X , es obtenida por un método heelstick (procedimiento no invasivo para la extracción de muestras de sangre en recién nacidos) y la variable de respuesta, Y , es obtenida usando un catéter umbilical. Considerando el modelo de regresión estándar, es decir

$$Y_i = \beta_0^{(s)} + \beta_1^{(s)} X_i + \epsilon_i^{(s)},$$

se obtienen las siguientes estimaciones:

$$\hat{\beta}_0^{(s)} = 2,7864, \quad \hat{\beta}_1^{(s)} = 0,8805 \quad \text{y} \quad \hat{\sigma}_{\epsilon^{(s)}}^2 = 8,4936.$$

Pero, en el contexto de modelos con error de medida la covariable X es una observación medida con error. Entonces se define el modelo de la sección 4.4.1 bajo el supuesto de normalidad. Kelly (1984) considera el supuesto de que $\sigma_\delta^2 = \sigma_\epsilon^2$, para hacer el modelo identificable. En contraste al supuesto de Kelly, consideramos que el valor σ_δ^2 es conocido, al igual que en Kelly (1984), asumimos $\sigma_\delta^2 = 4,6$. Entonces, el modelo reparametrizado viene dado por

$$Y_i = \beta_0^* + \beta_1^* X_i + \epsilon_i^*,$$

donde se obtienen las siguientes estimaciones:

$$\hat{\beta}_0^* = 2,7864, \quad \hat{\beta}_1^* = 0,8805 \quad \text{y} \quad \hat{\sigma}_{\epsilon^*}^2 = 8,4936.$$

Como notamos anteriormente, los estimadores de máxima verosimilitud de β_0^* y β_1^* corresponden a los estimadores de mínimos cuadrados. De (3.6), (3.7) y (3.10), obtenemos las estimaciones:

$$\hat{\beta}_0 = -1,4109, \quad \hat{\beta}_1 = 1,0818, \quad \hat{\sigma}_\epsilon^2 = 4,1118, \quad \hat{\eta}_\xi = 20,8550 \quad \text{y} \quad \hat{\sigma}_\xi^2 = 20,1255.$$

A partir de la regresión de Y sobre X , el estimador de mínimos cuadrados de β_1 no

es un estimador consistente, donde a su vez tenemos $\beta_1^{(s)} = \lambda\beta_1$, con

$$\lambda = \frac{\sigma_\xi^2}{\sigma_\xi^2 + \sigma_\delta^2} < 1,$$

en este caso, $\hat{\lambda} = 0,8140$.

Por lo tanto, el estimador de mínimos cuadrados de la regresión Y sobre X produce un estimador que está atenuado a cero. En la Figura 4.6, se observa el efecto del coeficiente de atenuación, λ es particularmente pronunciada.

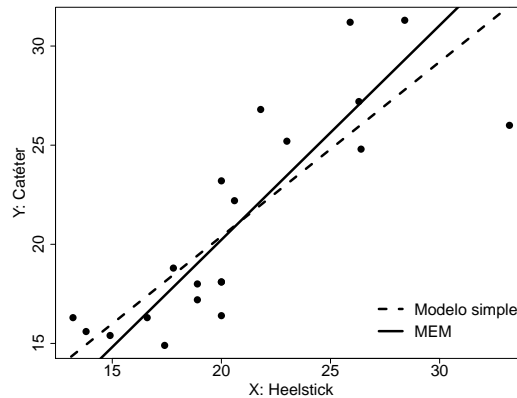


Figura 4.6: Muestras de sangre de 20 individuos analizados por Kelly (1984).

Debido a que X es propenso a error, uno debe esperar que exista una débil relación de la variable de respuesta y ξ , como se observa en la Figura 4.6. Esto puede ser visto tanto en la atenuación como también en el hecho de que la varianza residual de la regresión de Y sobre X es

$$\sigma_{\epsilon^{(s)}}^2 = \sigma_\epsilon^2 + \lambda\beta_1^2\sigma_\delta^2.$$

Esta faceta del problema es a menudo ignorada, pero es importante. El error de medición causa un doble problema, por un lado tenemos que la pendiente es atenuada, y por otro lado, los datos son más ruidosos con un incremento del error sobre la línea de regresión.

No es sorpresa que el error de medición, dado que es otra fuente de error, incremente la variabilidad sobre la recta de regresión. En efecto, podemos sustituir $\xi_i = X_i - \delta_i$ en el modelo de regresión para obtener el modelo $Y_i = \beta_0 + \beta_1 X_i + (\epsilon_i - \beta_1 \delta_i)$, con error $(\epsilon_i - \beta_1 \delta_i)$ que tiene varianza $\sigma_\epsilon^2 + \beta_1^2 \sigma_\delta^2 > \sigma_\epsilon^2$ y además es correlacionado con X . Lo que puede ser sorpresa es que este error adicional cause sesgo. Sin embargo, se tiene que el error y la covariable tienen un componente en común δ_i , el cual es la causa que sean correlacionados. Así, la correlación entre el error y la covariable es la

fuente del sesgo.

Los otros conjuntos de datos analizados, corresponden a los datos del proyecto WHO MONICA (World Health Organization Multinational Monitoring of Trends and Determinants in Cardiovascular Disease). Estos han sido previamente estudiado por Kuulasmaa et al. (2000), Kulathinal et al. (2002), de Castro et al. (2008) y Patriota et al. (2011). Uno de los principales objetivos de este proyecto consiste en relacionar los factores de riesgo conocidos de la mortalidad cardiovascular y/o las enfermedades cardiovasculares. Aquí, Y , es el índice medio anual en % de la ocurrencia de un evento, es decir, mortalidad cardiovascular o la presencia de enfermedad coronaria y la covariable observada, X , es el índice medio anual en % del riesgo asociado al evento. El riesgo es definido como una combinación lineal de la condición de fumador, presión sistólica, índice de masa corporal y nivel de colesterol. Notemos que, estas variables son índices no observables en alguna forma. El conjunto de datos esta dividido en dos grupos, hombres ($n = 38$) y mujeres ($n = 36$). El efecto de la medida de error puede ser visto en la Figura 4.7. La Tabla 4.8 muestra las estimaciones para los datos del proyecto WHO MONICA, considerando el modelo con y sin error.

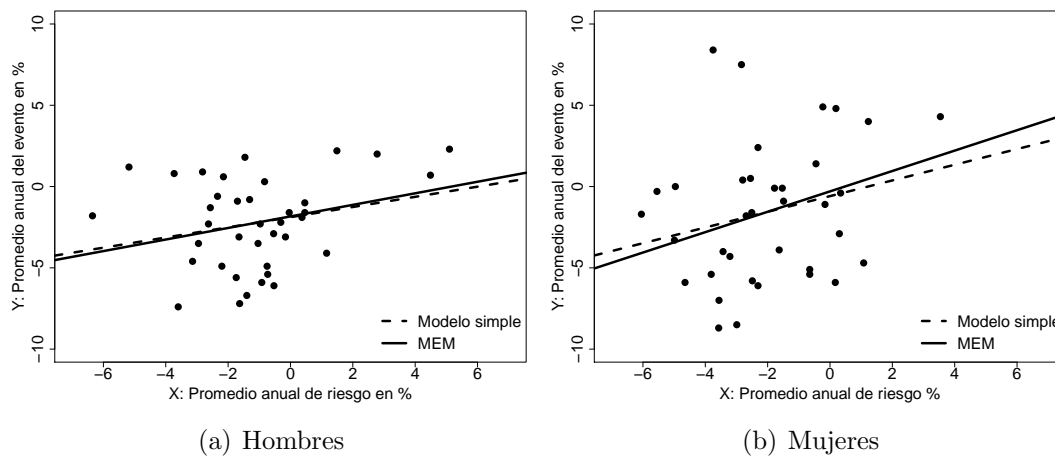


Figura 4.7: Proyecto WHO MONICA.

A partir de la aplicación de los resultados obtenidos en los conjunto de datos anteriores, podemos notar, por un lado el impacto de ignorar la medida de error en la covariable sobre la estimación de los parámetros del modelo, en particular en el sesgo en las estimaciones, y por otro lado, al considerar el supuesto de normalidad, tenemos que a través de la parametrización de Gleser, la estimación se ve simplificada, al ver que los estimadores de máxima verosimilitud, corresponden a los estimadores de mínimos cuadrados.

Tabla 4.8: Estimaciones y desviaciones estándar proyecto WHO MONICA.

Estimaciones	Hombres ($n = 38$)		Mujeres ($n = 36$)	
	Modelo simple	MEM	Modelo simple	MEM
$\hat{\beta}_0$	-1,8832(0,4944)	-1,7959(0,4853)	-0,5886(0,9879)	-0,2969(0,9685)
$\hat{\beta}_1$	0,3127(0,2012)	0,3934(1,5296)	0,4827(0,3403)	0,6271(5,8880)
$\hat{\sigma}_\epsilon^2$	2,7370(0,6451)	7,0970(1,6564)	4,2560(1,0031)	16,8031(4,1032)
$\hat{\eta}_\xi$	—	-1,0805(0,3602)	—	-2,0206(0,3744)
$\hat{\sigma}_\xi^2$	—	3,8714(1,1176)	—	3,3436(1,0278)

Ahora, comparamos los EMV con los estimadores con sesgo corregido (ESC) y los estimadores con sesgo reducido (ESR). Los resultados se muestran en la Tabla 4.9. La desviación estándar (SD) es basada en la matriz de información esperada.

A partir de la Tabla 4.9, podemos ver las diferencias en las estimaciones obtenidas a partir de los diferentes métodos. Notemos que, la diferencia en las estimaciones de las varianzas de los errores, es decir, la varianza del error en la ecuación σ_ϵ^2 y la varianza del error de medición σ_ξ^2 es significativa en comparación con las estimaciones de los parámetros β_0 y β_1 . Además, como se pudo ver en las simulaciones previas, la corrección en la estimación de η_ξ , es prácticamente despreciable.

Tabla 4.9: Estimaciones y desviación estándar.

θ	Miller (1986)					
	EMV	SD	ESC	SD	ESR	SD
β_0	-1,4110	2,8098	-1,1680	2,8798	-1,1560	2,8909
β_1	1,0818	76,9417	1,0702	78,7206	1,0696	79,0486
σ_ϵ^2	4,1118	2,6859	5,0083	2,9545	5,1051	2,9844
η_ξ	20,8550	1,2474	20,8550	1,2634	20,8350	1,2649
σ_ξ^2	20,1255	7,8204	21,3617	8,2111	21,4268	8,2316
WHO MONICA Hombres						
β_0	-1,7960	0,4812	-1,7987	0,4724	-1,7988	0,4931
β_1	0,3934	1,5072	0,3908	1,5627	0,3908	1,5661
σ_ϵ^2	6,9740	1,6282	7,3484	1,7139	7,3691	1,7186
η_ξ	-1,0805	0,3601	-1,0805	0,3648	-1,0805	0,3649
σ_ξ^2	3,8714	1,1176	3,9996	1,1470	4,0031	1,1478
WHO MONICA Mujeres						
β_0	-0,2969	0,9601	-0,3070	0,9799	-0,3073	0,9813
β_1	0,6271	5,7975	0,6220	6,0210	0,6219	6,0359
σ_ϵ^2	16,8031	4,0319	17,7559	4,2559	17,8118	4,2690
η_ξ	-2,0206	0,3740	-2,0206	0,3777	-2,0206	0,3778
σ_ξ^2	3,3436	1,0277	3,4643	1,0559	3,4677	1,0567

Los resultados que se muestran en la Tabla 4.10, están asociados a las estimaciones considerando la distribución t con $\nu = 6$. La desviación estándar (SD) es basada en la matriz de información esperada. A partir de estos resultados, podemos ver las diferencias en las estimaciones obtenidas a partir de los diferentes métodos. Además, como se pudo ver en las simulaciones previas, la corrección en la estimación de η_ξ , es prácticamente despreciable.

Tabla 4.10: Estimaciones y desviación estándar ($\nu = 6$).

Miller (1986)						
θ	EMV	SD	ESC	SD	ESR	SD
β_0	-6,6255	0,7746	-6,5868	0,8000	-6,5616	0,7992
β_1	1,3459	22,4673	1,3440	23,1626	1,3426	23,1370
σ_ϵ^2	0,2511	0,6060	0,6862	0,6241	0,6813	0,6235
η_ξ	20,1720	1,0834	20,1720	1,0873	20,1838	1,0878
σ_ξ^2	12,9891	6,6494	13,0604	6,6764	13,0871	6,6865
WHO MONICA Hombres						
β_0	-1,9899	0,5370	-1,9903	0,5463	-1,9920	0,5456
β_1	0,3851	1,2225	0,3847	1,2422	0,3841	1,2414
σ_ϵ^2	5,9416	0,4282	6,1533	0,4349	6,1387	0,4344
η_ξ	-1,1465	0,3240	-1,1465	0,3245	-1,1478	0,3245
σ_ξ^2	2,0519	0,8373	2,0584	0,8391	2,0574	0,8388
WHO MONICA Mujeres						
β_0	-0,3900	0,8361	-0,3914	0,8514	-0,3903	0,8512
β_1	0,7416	2,7791	0,7410	2,8272	0,7396	2,8265
σ_ϵ^2	13,1987	0,6430	13,7016	0,6543	13,6949	0,6541
η_ξ	-2,0589	0,3909	-2,0589	0,3925	-2,0590	0,3923
σ_ξ^2	2,4940	0,9863	2,5018	0,9886	2,4994	0,9879

Caso multivariado

Como ejemplo de la aplicación de los métodos de reducción y corrección de sesgo, consideramos los datos estudiados en Maddala (1988), los cuales corresponde a las importaciones (Y), producción (X_1) y el consumo (X_2), ya estudiados en el contexto de los modelos de regresión lineal con medida de error. La muestra consiste en $n = 18$ observaciones.

Así, el modelo es definido por las ecuaciones

$$\begin{cases} Y_i = \beta_0 + \beta_1 \xi_{1,i} + \beta_2 \xi_{2,i} + \epsilon_i, \\ X_{1,i} = \xi_{1,i} + \delta_{1,i}, \\ X_{2,i} = \xi_{2,i} + \delta_{2,i}, \end{cases}$$

Para hacer el modelo identificable consideramos los supuestos considerados en Maddala (1988), es decir, $\sigma_{\delta_1}^2 = 11,9378$ y $\sigma_{\delta_2}^2 = 3,2286$, con vector de parámetros a estimar dados por

$$\boldsymbol{\theta} = [\beta_0, \beta_1, \beta_2, \sigma_\epsilon^2, \eta_{\xi_1}, \eta_{\xi_2}, \sigma_{\xi_1}^2, \sigma_{\xi_2}^2]^\top,$$

lo cual implica que la implementación de los métodos de reducción y corrección sean costosos algebraicamente y computacionalmente, principalmente al implementar la estimación bajo es supuesto de la distribución t .

Considerando el modelo reparametrizado en (3.5), y la metodología desarrollada en este capítulo, se obtienen los resultados que se observan en la Tabla 4.11, la desviación estándar (SD) es basada en la matriz de información esperada, tanto en el caso normal, como en el caso t .

A partir de ésta tabla, podemos ver las diferencias en las estimaciones obtenidas a partir de los diferentes métodos. Notemos que, en los casos de distribucionales seleccionados, los métodos de reducción y corrección dan estimaciones similares, además podemos notar que el aumento en la desviación estándar en comparación con la desviación estándar de los estimadores de máxima verosimilitud, no es significativa para los propósitos del estudio.

Este último punto, es importante de destacar, así como se mencionó anteriormente, el costo de la implementación de algún método de reducción de sesgo implica que la variabilidad de los mismos aumentará, pero como es notado, tanto en este ejemplo y los anteriores, incluso las simulaciones realizadas, la perdida de la aplicación de los métodos para mejorar las estimaciones no es significativa.

Tabla 4.11: Estimaciones y desviación estándar.

Caso normal						
θ	EMV	SD	ESC	SD	ESR	SD
β_0	-17,6927	4,1237	-15,9094	4,3560	-15,9913	4,3502
β_1	1,6229	0,1907	1,1940	0,5763	1,1940	0,5018
β_2	-2,3786	0,2913	-2,2976	0,7113	-2,2976	0,7234
σ_ϵ^2	4,2753	1,9103	4,6903	2,3134	4,7544	2,3616
η_{ξ_1}	237,5167	63,5167	237,5167	64,7130	237,5167	64,7059
η_{ξ_2}	167,3778	41,5811	167,3778	45,6901	167,3778	45,6516
$\sigma_{\xi_1}^2$	3798,3150	240,0764	4010,1450	247,4193	4022,5968	248,2415
$\sigma_{\xi_2}^2$	1629,7010	154,9193	1621,9295	183,1110	1717,2048	182,8806
Caso t ($\nu = 4$)						
β_0	-18,2748	2,6397	-18,2881	2,6563	-18,2954	2,6511
β_1	1,2862	0,0249	1,2863	0,1842	1,2863	0,1876
β_2	-1,9842	0,1873	-1,9755	0,2539	-1,9757	0,2975
σ_ϵ^2	2,0573	0,3180	2,1922	0,3293	2,1903	0,3268
η_{ξ_1}	241,0335	54,3205	241,0335	54,3242	241,0644	54,3498
η_{ξ_2}	161,0564	38,9534	161,0563	38,9837	161,0564	38,9752
$\sigma_{\xi_1}^2$	2514,6574	233,0969	2482,8680	235,5398	2491,3450	235,5060
$\sigma_{\xi_2}^2$	1122,6574	172,3024	1090,5715	174,0740	1099,3450	174,2366

4.5. Discusión

En este capítulo se estudiaron los métodos de corrección y reducción de sesgo, en los modelos de regresión lineal con error de medida. A partir de los resultados obtenidos, se puede observar que los ESC y ESR tienen similar comportamiento en conjuntos de datos pequeños, donde la ventaja de los ESR es que no depende de la finitud de los estimadores de MV. A pesar de su buen comportamiento, la principal desventaja de los métodos de reducción de sesgo es que, en algunos casos, su aplicación requiere de la solución de un conjunto de ecuaciones en el cual en muchos de los casos útiles requiere de optimización numérica. Esta tarea es a menudo más demandante computacionalmente.

Además notamos que los algoritmos tipo Fisher-Scoring desarrollados, son una buena alternativa para la obtención de los estimadores, aunque se debe notar que ciertas consideraciones se deben tomar en algunos casos, para asegurar la convergencia.

En los ejemplos desarrollados, notamos que, tanto los ESC y los ESR corrigen las estimaciones asociadas a los parámetros del modelo con medida de error. Particularmente, notamos que la corrección realizada sobre la varianza de los errores es más evidente. Tanto en el caso normal como con la distribución t .

Una consideración final, que debemos realizar, respecto a los procesos de reducción o corrección de sesgo es que, por su definición, el sesgo necesariamente depende de la parametrización del modelo; si el sesgo de cualquier estimador $\hat{\theta}$ es reducido resultando en un estimador con un sesgo menor, $\tilde{\theta}$. Esto no necesariamente ocurrirá lo mismo para el estimador $h(\tilde{\theta})$, para alguna función uno-a-uno $h : \mathbb{R}^p \rightarrow \mathbb{R}^p$. En efecto, el sesgo de $h(\tilde{\theta})$ como un estimador de $h(\theta)$ puede aumentar considerablemente. Por lo tanto, la corrección o reducción de sesgo aplicada a los estimadores de MV puede tener el costo de anular la propiedad de invarianza bajo la reparametrización. Entonces, los métodos de reducción y corrección de sesgo desarrollados en este capítulo, deben ser vistos con cierto escepticismo si la propiedad de invarianza es un requerimiento necesario para el análisis. Por otro lado, si la parametrización es fijada por el problema, podemos mejorar mucho en términos de estimación e inferencia a través de la corrección o reducción de sesgo.

Capítulo 5

Conclusiones y trabajos futuros

5.1. Comentarios finales

En esta tesis se discutieron los siguientes puntos relacionados con los modelos con error de medida, bajo el supuesto de que los errores (y las observaciones) tienen distribución elíptica.

1. Se estudiaron los modelos de regresión lineal multivariados con error de medida, considerando una extensión elíptica al enfoque discutido en Gleser (1992) (Capítulo 3).
2. La nueva parametrización tiene la ventaja de establecer una útil conexión entre los modelos de regresión estándar y los modelos de regresión lineal multivariados con error de medida.
3. El enfoque de Gleser (1992) permite una simple representación jerárquica, donde notamos que sólo bajo el supuesto de normalidad el vector de parámetros está en variación libre, y podemos estudiar una versión reducida del modelo. Considerando las distribuciones elípticas, ésta representación no es exógena y no podemos ignorar la distribución de la variable medida con error en el proceso de estimación. Además, notamos que los modelos con error de medida son heteroscedásticos bajo las distribuciones elípticas no-normales.
4. El modelo reparametrizado permite dar expresiones simples para calcular los elementos de la matriz de información esperada, en donde sólo algunos momentos univariados deben ser calculados numéricamente.
5. Se investigaron e implementaron algoritmos que son basados en la modificación del vector de score para la reducción del sesgo en la estimación de los coeficientes

de regresión en modelos con error de medida (Capítulo 4). La principal ventaja de los métodos de reducción de sesgo es que estos no dependen directamente de los estimadores de máxima verosimilitud.

6. La simulación muestra que los métodos de reducción de sesgo tienen un comportamiento similar a los métodos de corrección de sesgo, por tanto constituye una alternativa cuando los estimadores de máxima verosimilitud divergen.

5.2. Trabajos futuros

Debido a la flexibilidad que podemos tener al considerar las distribuciones elípticas en el contexto de los modelos con error de medida, también es posible obtener diferentes extensiones.

Por un lado, se puede considerar una generalización de la estructura de los modelos con error de medida, y por otro lado se pueden considerar diferentes supuestos distribucionales.

En este último contexto podemos considerar una extensión asimétrica de las distribuciones elípticas, es decir, distribuciones skew-elípticas. Además, tenemos las distribuciones slash-elípticas, las cuales corresponden a distribuciones de mezcla de escala de normales, las distribuciones slash presentan colas más pesadas que la distribución normal y como una consecuencia de esto la distribución posee mayor kurtosis. Estas extensiones son discutidas a continuación.

5.2.1. Modelos con medida de error generalizados

Un interesante área de estudio es considerar el enfoque de Gleser (1992) en una generalización de los modelos con medida de error dados por

$$\begin{cases} \mathbf{Y}_i = \boldsymbol{\alpha}^\top \mathbf{Z}_i + \boldsymbol{\beta}^\top \boldsymbol{\xi}_i + \varepsilon_i, \\ \mathbf{X}_i = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 \odot \boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \end{cases}$$

para $i = 1, 2, \dots, n$, donde \mathbf{Z}_i son las covariables medidas sin error y $\boldsymbol{\xi}_i$ son las covariables medidas con error y pueden ser observadas a través de \mathbf{X}_i . Aquí, \mathbf{Z}_i , \mathbf{X}_i y $\boldsymbol{\xi}_i$ son vectores p -dimensionales, \odot es el producto Hadamard, $\boldsymbol{\tau}_0$ es el error aditivo y $\boldsymbol{\tau}_1$ es el error multiplicativo. Para $\boldsymbol{\tau}_0 = \mathbf{0}$ y $\boldsymbol{\tau}_1 = \mathbf{1}$ se recupera el modelo con medida de error clásico.

5.2.2. Distribuciones skew-elípticas

Las distribuciones elípticas son radialmente simétricas, así que la idea es modelar la asimetría radial sin perder muchas de las propiedades básicas de las distribuciones elípticas.

Por un lado, se debería considerar la parsimonia con respecto a la parametrización de un modelo multivariado. Y otro lado, todos los componentes de las distribuciones elípticas, es decir, la variable generadora R , el vector de localización $\boldsymbol{\mu}$ y la matriz de dispersión $\boldsymbol{\Sigma}$ (la cual contiene la dependencia lineal de los componentes) deben mantenerse en la nueva clase de las distribuciones asimétricas. Entonces, las siguientes definiciones describen dos métodos para introducir un parámetro que controle la asimetría en el modelo elíptico.

Definición 5.1 (Mezcla de media-varianza elíptica). Un vector aleatorio n -dimensional, \mathbf{X} , es llamado mezcla de media-varianza elíptica si este puede ser representado por

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + R\boldsymbol{\lambda} + \sqrt{R}\mathbf{Y},$$

donde $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\lambda} \in \mathbb{R}^n$, $\mathbf{Y} \sim \text{El}_n(\mathbf{0}, \boldsymbol{\Sigma}; \phi)$, $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ es definido positivo, y R es una variable aleatoria no-negativa independiente de \mathbf{Y} . Si $\boldsymbol{\lambda} = \mathbf{0}$, entonces \mathbf{X} es una mezcla de varianza elíptica.

Ya que $(\mathbf{X}|R = r) \stackrel{d}{=} \boldsymbol{\mu} + r\boldsymbol{\lambda} + \sqrt{r}\mathbf{Y}$, la función de distribución de \mathbf{X} es dada por una mezcla de $(\mathbf{X}|R = r)$ con distribución de mezcla $F_R(r)$. Esto es denotado por

$$F_{\mathbf{X}}(\mathbf{x}) = \int_0^{\infty} \text{El}_n(\boldsymbol{\mu} + r\boldsymbol{\lambda}, r\boldsymbol{\Sigma}; \phi) dF_R(r).$$

El vector $\boldsymbol{\lambda}$ no es un vector de localización pero determina la asimetría de la mezcla de varianza elíptica.

Definición 5.2 (Mezcla de localización-escala elíptica). Un vector aleatorio n -dimensional, \mathbf{X} , es llamado mezcla de localización-escala si este puede ser representado por

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + R\mathbf{Y},$$

donde $\boldsymbol{\mu} \in \mathbb{R}^n$, $\mathbf{Y} \sim \text{El}_n(\boldsymbol{\lambda}, \boldsymbol{\Sigma}; \phi)$, $\boldsymbol{\lambda} \in \mathbb{R}^n$, $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ es definido positivo, y R es una variable aleatoria no-negativa independiente de \mathbf{Y} . Si $\boldsymbol{\lambda} = \mathbf{0}$, entonces \mathbf{X} es una mezcla de escala elíptica.

Ahora, la función de distribución de \mathbf{X} puede ser representada por

$$F_{\mathbf{X}}(\mathbf{x}) = \int_0^\infty \text{El}_n(\boldsymbol{\mu} + r\boldsymbol{\lambda}, r^2\boldsymbol{\Sigma}) dF_R(r).$$

Si $\mathbf{Y} \sim N_n(\boldsymbol{\lambda}, \boldsymbol{\Sigma})$, entonces \mathbf{X} tiene distribución de mezcla de localización-escala normal. En ningún caso la distribución de mezcla de media-varianza normal y la distribución de mezcla de localización-escala normal son distribuciones elípticas si $\boldsymbol{\lambda} \neq 0$. Sin embargo, ambas clases son caracterizadas por los componentes de un vector elíptico. Únicamente el vector de parámetros adicional $\boldsymbol{\lambda}$ determina la asimetría, es decir, la asimetría radial.

Otra forma de incorporar asimetría en el contexto elíptico es dada por la técnica de la truncación oculta (hidden truncation, ver Arnold & Beaver (2002)).

Definición 5.3 (Distribución skew-elíptica). Sea $\mathbf{Y}^* = (Y_0, \mathbf{Y}^\top)^\top \sim \text{El}_{n+1}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*; \phi)$ donde $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\mu}^* = (0, \boldsymbol{\mu}^\top)^\top$, $\boldsymbol{\delta} \in \mathbb{R}^n$, $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ y

$$\boldsymbol{\Sigma}^* = \begin{pmatrix} 1 & \boldsymbol{\delta}^\top \\ \boldsymbol{\delta} & \boldsymbol{\Sigma} \end{pmatrix}.$$

Entonces el vector aleatorio n -dimensional, $\mathbf{X} = (\mathbf{Y} | Y_0 > 0)$, se dice que es distribuido skew-elíptico (Branco & Dey (2001)) lo cual es denotado por $\mathbf{X} \sim \text{SEl}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \phi)$, con

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\delta}) = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}}{(1 - \boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\delta})^{-1/2}}. \quad (5.1)$$

Nuevamente $\boldsymbol{\lambda}$ (o $\boldsymbol{\delta}$) sirve como parámetro de asimetría. Si ϕ corresponde a la función característica de la distribución normal entonces \mathbf{X} se distribuye skew-normal multivariada (Azzalini & Dalla Valle (1996)). Una revisión de la literatura concerniente a las distribuciones skew-elípticas puede ser encontrada en Azzalini (2008). De la Definición 5.3, si la densidad del vector aleatorio \mathbf{Y}^* existe y $\mathbf{P}(\mathbf{Y}^* = \mathbf{0}) = 0$, entonces la función de densidad de probabilidad de \mathbf{X} es de la forma

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = 2f_g(\mathbf{x})F_{g_{q(\mathbf{x})}}(\boldsymbol{\lambda}^\top(\mathbf{x} - \boldsymbol{\mu})),$$

donde $f_g(\cdot)$ es la función de densidad de probabilidad de $\text{El}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ y $F_{g_{q(\mathbf{x})}}$ es la función de distribución de $\text{El}(0, 1, g_{q(\mathbf{x})})$ donde $g_{q(\mathbf{x})}$ es definida en (2.16). Algunas propiedades son dadas por

1. Si $\mathbf{X} \sim \text{SEl}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; g)$, entonces $(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ tiene la misma distribución de R^2 , donde R es la variable radial en la representación estocástica de

\mathbf{Y} en la Definición 5.3.

2. Sea $\mathbf{a} \in \mathbb{R}^m$ y $\mathbf{B} \in \mathbb{R}^{m \times n}$, y consideremos el vector aleatorio

$$\mathbf{Y} = \mathbf{a} + \mathbf{B}\mathbf{X},$$

donde $\mathbf{X} \sim \text{SEl}_n(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X, \boldsymbol{\lambda}_X; g)$, entonces $\mathbf{Y} \sim \text{SEl}_m(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y, \boldsymbol{\lambda}_Y; g)$, donde $\boldsymbol{\mu}_Y = \mathbf{a} + \mathbf{B}\boldsymbol{\mu}_X$, $\boldsymbol{\Sigma}_Y = \mathbf{B}\boldsymbol{\Sigma}_X\mathbf{B}^\top$ y $\boldsymbol{\lambda}_Y = \boldsymbol{\lambda}(\mathbf{B}\boldsymbol{\delta})$ en (5.1).

3. Sea $\mathbf{X} \sim \text{SEl}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; g)$, consideremos la partición definida en (2.12) y $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\top, \boldsymbol{\lambda}_2^\top)^\top$, entonces el vector aleatorio m -dimensional \mathbf{X}_1 , tiene distribución dada por $\mathbf{X}_1 \sim \text{SEl}_n(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\lambda}_1; g)$.

Consideremos el modelo de regresión lineal multivariado con error de medida clásico descrito por

$$\begin{cases} Y_i = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{\xi}_i + \varepsilon_i, \\ \mathbf{X}_i = \boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \end{cases}$$

Como en (3.1). Sea $\mathbf{r}_i = (\varepsilon_i, \boldsymbol{\delta}_i^\top, \boldsymbol{\xi}_i^\top)^\top$, ahora asumamos que $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ son no-correlacionados $\mathbf{r}_i \sim \text{SEl}_{2p+1}(\boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\lambda}; \phi)$ donde $\boldsymbol{\eta}$ y $\boldsymbol{\Omega}$ son definidos en (3.2) y $\boldsymbol{\lambda} = (\lambda_\varepsilon, \boldsymbol{\lambda}_\delta^\top, \boldsymbol{\lambda}_\xi^\top)^\top$. Entonces, $(\mathbf{X}_i^\top, Y_i)^\top \sim \text{SEl}_{p+1}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}), \boldsymbol{\lambda}(\boldsymbol{\theta}); \phi)$ donde

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\eta}_\xi \\ \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{\eta}_\xi \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{\delta\delta} + \boldsymbol{\Sigma}_{\xi\xi} & \boldsymbol{\Sigma}_{\xi\xi}\boldsymbol{\beta} \\ \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\xi\xi} & \sigma_\varepsilon^2 + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\xi\xi}\boldsymbol{\beta} \end{bmatrix},$$

y

$$\boldsymbol{\lambda}(\boldsymbol{\theta}) = b(\boldsymbol{\theta}) \begin{bmatrix} -\lambda_\varepsilon \sigma_\varepsilon^2 \boldsymbol{\Lambda} \boldsymbol{\beta} + b \boldsymbol{\Lambda} \boldsymbol{\Sigma}_{\xi\xi}^{-1} \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\lambda}_\delta + b \boldsymbol{\Lambda} \boldsymbol{\lambda}_\xi + \boldsymbol{\Lambda} \boldsymbol{\beta} \boldsymbol{\beta}^\top \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}_{\delta\delta} (\boldsymbol{\lambda}_\delta - \boldsymbol{\lambda}_\xi) \\ \lambda_\varepsilon \sigma_\varepsilon^2 - \boldsymbol{\beta}^\top \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}_{\delta\delta} (\boldsymbol{\lambda}_\delta - \boldsymbol{\lambda}_\xi) \end{bmatrix},$$

donde

$$\begin{aligned} b^{-2}(\boldsymbol{\theta}) = & b^3 + b^2 \left[\lambda_\varepsilon^2 (b\sigma_\varepsilon^2 - \sigma_\varepsilon^4) - (\boldsymbol{\lambda}_\delta^\top - \boldsymbol{\lambda}_\xi^\top) \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda} \boldsymbol{\beta} \boldsymbol{\beta}^\top \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}_{\delta\delta} (\boldsymbol{\lambda}_\delta - \boldsymbol{\lambda}_\xi) \right. \\ & + \lambda_\varepsilon \sigma_\varepsilon^2 (\boldsymbol{\lambda}_\delta^\top - \boldsymbol{\lambda}_\xi^\top) \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda} \boldsymbol{\beta} + \lambda_\varepsilon \sigma_\varepsilon^2 \boldsymbol{\beta}^\top \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}_{\delta\delta} (\boldsymbol{\lambda}_\delta - \boldsymbol{\lambda}_\xi) \\ & - b (\boldsymbol{\lambda}_\delta^\top - \boldsymbol{\lambda}_\xi^\top) \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda} \boldsymbol{\lambda}_\varepsilon + b (\boldsymbol{\lambda}_\delta^\top - \boldsymbol{\lambda}_\xi^\top \boldsymbol{\Lambda}) \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\lambda}_\delta \\ & \left. - b \boldsymbol{\lambda}_\delta^\top \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda} \boldsymbol{\Sigma}_{\xi\xi}^{-1} \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\lambda}_\delta \right], \end{aligned}$$

y, $\boldsymbol{\Lambda}$ es la matriz de confiabilidad, $b = \sigma_\varepsilon^2 + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda} \boldsymbol{\beta}$ y $\boldsymbol{\theta} = [\beta_0, \boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\eta}_\xi, \boldsymbol{\Sigma}_{\xi\xi}, \boldsymbol{\Sigma}_{\delta\delta}]^\top$.

5.2.3. Distribuciones slash-elípticas

Otra extensión posible de la familia de distribuciones elípticas es a través de las distribuciones slash (mezcla de escala de normales). La distribución slash estándar es la razón de dos variables aleatorias independientes con representación estocástica dada por

$$S = \frac{Z}{U^{1/q}}, \quad (5.2)$$

donde $Z \sim N(0, 1)$, $U \sim U(0, 1)$ y $q > 0$.

Esta distribución presenta colas más pesadas que la distribución normal, y como una consecuencia, ésta distribución tiene mayor kurtosis. Algunas propiedades de esta familia han sido discutidas en Rogers & Tukey (1972) y Mosteller & Tukey (1977), los estimadores de máxima verosimilitud para los parámetros de localización y escala son estudiados en Kafadar (1982). Wang & Genton (2006) describen una extensión multivariada simétrica y asimétrica de la distribución slash, mientras Gómez et al. (2007) y Gómez & Venegas (2008) extiende las distribuciones slash introduciendo la familia de distribuciones slash-elípticas; versiones asimétricas de esta familia son discutidas en el trabajo de Arslan (2008). Arslan & Genç (2009) discute una extensión simétrica de las distribuciones slash multivariadas y Genç (2007) discute una generalización simétrica de la distribución slash. Recientemente Gómez et al. (2009) utiliza la familia slash-elíptica para extender la distribución Birnbaum-Saunders.

Motivados por la representación univariada (5.2), diremos que un vector aleatorio $\mathbf{Y} \in \mathbb{R}^k$ tiene distribución slash-elíptica multivariada con vector de localización $\boldsymbol{\mu}$, matriz de escala $\boldsymbol{\Sigma}$ definida positiva y parámetro de kurtosis $q > 0$, si esta puede ser representada como

$$\mathbf{Y} = \boldsymbol{\Sigma}^{1/2} \frac{\mathbf{X}}{U^{1/q}} + \boldsymbol{\mu}, \quad (5.3)$$

donde $\mathbf{X} \sim \text{El}_k(\mathbf{0}, \mathbf{I}_k; g)$ y es independiente de $U \sim U(0, 1)$. Denotamos esto como $\mathbf{Y} \sim \text{ShEl}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q; g)$. La función de densidad de \mathbf{Y} es dada por

$$f_{\mathbf{Y}} = \begin{cases} \frac{q|\boldsymbol{\Sigma}|^{-1/2}}{2\gamma^{(q+k)/2}} \int_0^{(q+k-2)/2} g(t) dt, & \mathbf{y} \neq \boldsymbol{\mu}, \\ \frac{q}{k+q} |\boldsymbol{\Sigma}|^{-1/2} g(0), & \mathbf{y} = \boldsymbol{\mu}, \end{cases}$$

donde $\gamma = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$, ver Gómez et al. (2007). La media y la matriz de covarianza de la familia slash-elíptica, es dada por

$$E[\mathbf{Y}] = \boldsymbol{\mu}, \text{ y } V(\mathbf{Y}) = \left(\frac{q\alpha}{q-2} \right) \boldsymbol{\Sigma}, \quad q > 2.$$

donde α es una constante positiva que depende de g (Fang et al. (1990)). Notemos, que las distribuciones skew-elípticas y slash-elípticas son dos extensiones de las distribuciones elípticas, donde se puede explorar la asimetría y la kurtosis de los datos. Además, se pueden considerar como una extensión sobre los supuestos distribucionales de los modelos con error de medición. Aspectos interesantes de estudio son el comportamiento del sesgo presente en la estimación de los nuevos parámetros.

Referencias

- Anderson, T. W., Fang, K. T. & Hsu, H. (1986), ‘Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions’, *The Canadian Journal of Statistics. La Revue Canadienne de Statistique* **14**(1), 55–59.
- Andrews, D. F. & Mallows, C. L. (1974), ‘Scale mixtures of normal distributions’, *Journal of the Royal Statistical Society. Series B. Methodological* **36**, 99–102.
- Arellano-Valle, R. B. (1994), Distribuições elípticas: propriedades, inferência e aplicações a modelos de regressão, Phd thesis, Universidad de São Paulo, Brasil.
- Arellano-Valle, R. B. (2010), ‘On the information matrix of the multivariate skew- t model’, *Metron* **68**(3), 371–386.
- Arellano-Valle, R. B. & Bolfarine, H. (1995), ‘On some characterizations of the t -distribution’, *Statistics & Probability Letters* **25**(1), 79–85.
- Arnold, B. C. & Beaver, R. J. (2002), ‘Skewed multivariate models related to hidden truncation and/or selective reporting’, *Test* **11**(1), 7–54. With discussion and a rejoinder by the authors.
- Arslan, O. (2008), ‘An alternative multivariate skew-slash distribution’, *Statistics & Probability Letters* **78**(16), 2756–2761.
- Arslan, O. & Genç, A. İ. (2009), ‘A generalization of the multivariate slash distribution’, *Journal of Statistical Planning and Inference* **139**(3), 1164–1170.
- Azzalini, A. (2008), ‘References on the skew-normal distribution and related ones’. Retrieved 2008-06-22 from <http://azzalini.stat.unipd.it/SN/list-publ.pdf>.
- Azzalini, A. & Dalla Valle, A. (1996), ‘The multivariate skew-normal distribution’, *Biometrika* **83**(4), 715–726.

- Barndorff-Nielsen, O., Kent, J. & Sørensen, M. (1982), ‘Normal variance-mean mixtures and z distributions’, *International Statistical Review. Revue Internationale de Statistique* **50**(2), 145–159.
- Barnett, V. D. (1969), ‘Simultaneous pairwise linear structural relationship’, *Biometrics* **25**(1), 129–142.
- Bingham, N. H. & Kiesel, R. (2002), ‘Semi-parametric modelling in finance: theoretical foundations’, *Quantitative Finance* **2**(4), 241–250.
- Bolfarine, H. & Galea-Rojas, M. (1996), ‘On structural comparative calibration under a t -model’, *Computational Statistics* **11**(1), 63–85.
- Branco, M. D. & Dey, D. K. (2001), ‘A general class of multivariate skew-elliptical distributions’, *Journal of Multivariate Analysis* **79**(1), 99–113.
- Breslow, N. E. & Lin, X. (1995), ‘Bias correction in generalised linear mixed models with a single component of dispersion’, *Biometrika* **82**(1), 81–91.
- Bull, S. B., Mak, C. & Greenwood, C. M. T. (2002), ‘A modified score function estimator for multinomial logistic regression in small samples’, *Computational Statistics & Data Analysis* **39**(1), 57–74.
- Buonaccorsi, J. P. (2010), *Measurement error*, CRC Press, Boca Raton, FL.
- Cambanis, S., Huang, S. & Simons, G. (1981), ‘On the theory of elliptically contoured distributions’, *Journal of Multivariate Analysis* **11**(3), 368–385.
- Carroll, R. J., Ruppert, D. & Stefanski, L. A. (1995), *Measurement error in nonlinear models*, Chapman & Hall, London.
- Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006), *Measurement error in nonlinear models*, second edn, Chapman & Hall/CRC, Boca Raton, FL.
- Castro, L. M. (2008), Distribución skew-normal: identificabilidad y reducción, y enfoque bayesiano de mezclas, Phd thesis, Pontificia Universidad Católica de Chile, Chile.
- Cheng, C.-L. & Tsai, C.-L. (2004), ‘The invariance of some score tests in the linear model with classical measurement error’, *Journal of the American Statistical Association* **99**(467), 805–809.

- Cheng, C. L. & Van Ness, J. W. (1999), *Statistical regression with measurement error*, Arnold, London.
- Chib, S., Tiwari, R. C. & Jammalamadaka, S. R. (1988), ‘Bayes prediction in regressions with elliptical errors’, *Journal of Econometrics* **38**(3), 349–360.
- Chipkevitch, E., Nishimura, R., Tu, D. & Galea-Rojas, M. (1996), ‘Clinical measurement of testicular volume in adolescents: comparison of the reliability of 5 methods’, *The Journal of Urology* **156**(6), 2050–2053.
- Cook, R. D., Tsai, C.-L. & Wei, B. C. (1986), ‘Bias in nonlinear regression’, *Biometrika* **73**(3), 615–623.
- Cordeiro, G. M., Ferrari, S. L. P. & Uribe-Opazo, M. A. (2000), ‘Improving score tests in the two-parameter exponential family’, *Revista de Matemática e Estatística. Brazilian Journal of Mathematics and Statistics* **18**, 247–267.
- Cordeiro, G. M. & Klein, R. (1994), ‘Bias correction in ARMA models’, *Statistics & Probability Letters* **19**(3), 169–176.
- Cordeiro, G. M. & McCullagh, P. (1991), ‘Bias correction in generalized linear models’, *Journal of the Royal Statistical Society. Series B. Methodological* **53**(3), 629–643.
- Cordeiro, G. M. & Toyama Udo, M. C. (2008), ‘Bias correction in generalized nonlinear models with dispersion covariates’, *Communications in Statistics - Theory and Methods* **37**(13-15), 2219–2225.
- Cordeiro, G. M. & Vasconcellos, K. L. P. (1997), ‘Bias correction for a class of multivariate nonlinear regression models’, *Statistics & Probability Letters* **35**(2), 155–164.
- Cordeiro, G. M. & Vasconcellos, K. L. P. (1999), ‘Second-order biases of the maximum likelihood estimates in von Mises regression models’, *Australian & New Zealand Journal of Statistics* **41**(2), 189–198.
- Cordeiro, G. M., Vasconcellos, K. L. P. & Santos, M. L. F. (1998), ‘On the second-order bias of parameter estimates in nonlinear regression models with student t errors’, *Journal of Statistical Computation and Simulation* **60**(1), 363–378.
- Cox, D. R. & Hinkley, D. V. (1974), *Theoretical statistics*, Chapman and Hall, London.

- Cox, D. R. & Snell, E. J. (1968), ‘A general definition of residuals’, *Journal of the Royal Statistical Society. Series B. Methodological* **30**, 248–275.
- de Castro, M., Galea, M. & Bolfarine, H. (2008), ‘Hypothesis testing in an errors-in-variables model with heteroscedastic measurement errors’, *Statistics in Medicine* **27**(25), 5217–5234.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), ‘Maximum likelihood from incomplete data via the EM algorithm’, *Journal of the Royal Statistical Society. Series B. Methodological* **39**(1), 1–38. With discussion.
- Dolby, G. R. (1976), ‘The ultrastructural relation: a synthesis of the functional and structural relations’, *Biometrika* **63**(1), 39–50.
- Dunn, G. (1992), *Design and analysis of reliability: The statistical evaluation of measurement errors*, Edward Arnold, New York.
- Eaton, M. L. (1981), ‘On the projections of isotropic distributions’, *The Annals of Statistics* **9**(2), 391–400.
- Engle, R. F., Hendry, D. F. & Richard, J.-F. (1983), ‘Exogeneity’, *Econometrica. Journal of the Econometric Society* **51**(2), 277–304.
- Fang, K. T., Kotz, S. & Ng, K. W. (1990), *Symmetric multivariate and related distributions*, Chapman and Hall Ltd., London.
- Fang, K. T. & Zhang, Y. T. (1990), *Generalized multivariate analysis*, Springer-Verlag, Berlin.
- Fernandez, C. & Steel, M. F. J. (1999), ‘Multivariate student t regression models: pitfall and inference’, *Biometrika* **86**(1), 153–167.
- Ferrari, S. L. P. & Arellano-Valle, R. B. (1996), ‘Modified likelihood ratio and score tests in linear regression models using the t distribution’, *Brazilian Journal of Probability and Statistics* **10**(1), 15–33.
- Firth, D. (1993), ‘Bias reduction of maximum likelihood estimates’, *Biometrika* **80**(1), 27–38.
- Firth, D. (1995), ‘Correction to: “Bias reduction of maximum likelihood estimates” [Biometrika **80** (1993), no. 1, 27–38; MR1225212 (94b:62017)]’, *Biometrika* **82**(3), 667.

- Florens, J. P. & Mouchart, M. (1980), ‘Initial and sequential reduction of bayesian experiment’, *CORE Discussion Papers 8015* .
- Fuller, W. A. (1980), ‘Properties of some estimators for the errors-in-variables model’, *The Annals of Statistics* **8**(2), 407–422.
- Fuller, W. A. (1987), *Measurement error models*, John Wiley & Sons Inc., New York.
- Genç, A. İ. (2007), ‘A generalization of the univariate slash by a scale-mixed exponential power distribution’, *Communications in Statistics. Simulation and Computation* **36**(4-6), 937–947.
- Ghosh, M. & Sinha, B. K. (1980), ‘On the robustness of least squares procedures in regression models’, *Journal of Multivariate Analysis* **10**(3), 332–342.
- Gleser, L. J. (1992), ‘The importance of assessing measurement reliability in multivariate regression’, *Journal of the American Statistical Association* **87**(419), 696–707.
- Gómez, H. W., Olivares-Pacheco, J. F. & Bolfarine, H. (2009), ‘An extension of the generalized Birnbaum-Saunders distribution’, *Statistics & Probability Letters* **79**(3), 331–338.
- Gómez, H. W., Quintana, F. A. & Torres, F. J. (2007), ‘A new family of slash-distributions with elliptical contours’, *Statistics & Probability Letters* **77**(7), 717–725.
- Gómez, H. W. & Venegas, O. (2008), ‘Erratum to: “A new family of slash-distributions with elliptical contours” [mr2356511]’, *Statistics & Probability Letters* **78**(14), 2273–2274.
- Gong, G. & Samaniego, F. J. (1981), ‘Pseudomaximum likelihood estimation: theory and applications’, *The Annals of Statistics* **9**(4), 861–869.
- Griliches, Z. & Ringstad, V. (1971), *Economies of Scale and the Form of the Production Function: An Econometric Study of Norwegian Manufacturing Establishment Data*, North-Holland Publishing Company.
- Guolo, A. (2011), ‘Pseudo-likelihood inference for regression models with misclassified and mismeasured variables’, *Statistica Sinica* **21**(4), 1639–1663.

- Heinze, G. & Schemper, M. (2001), ‘A solution to the problem of monotone likelihood in Cox regression’, *Biometrics. Journal of the International Biometric Society* **57**(1), 114–119.
- Jensen, D. R. (1985), *Multivariate distribution*, Wiley.
- Johnston, J. (1997), *Econometric Methods*, McGraw-Hill.
- Kafadar, K. (1982), ‘A biweight approach to the one-sample problem’, *Journal of the American Statistical Association* **77**(378), 416–424.
- Kano, Y., Berkane, M. & Bentler, P. M. (1993), ‘Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations’, *Journal of the American Statistical Association* **88**(421), 135–143.
- Kelker, D. (1970), ‘Distribution theory of spherical distributions and a location-scale parameter generalization’, *Sankhyā (Statistics). The Indian Journal of Statistics. Series A* **32**, 419–438.
- Kelly, G. (1984), ‘The influence function in the errors in variables problem’, *The Annals of Statistics* **12**(1), 87–100.
- Kosmidis, I. (2007), Bias reduction in exponential family nonlinear models, Phd thesis, University of Warwick.
- Kosmidis, I. (2014), ‘Bias in parametric estimation: reduction and useful side-effects’, *Wiley Interdisciplinary Reviews: Computational Statistics* **6**(3), 185–196.
- Kosmidis, I. & Firth, D. (2009), ‘Bias reduction in exponential family nonlinear models’, *Biometrika* **96**(4), 793–804.
- Kosmidis, I. & Firth, D. (2010), ‘A generic algorithm for reducing bias in parametric estimation’, *Electronic Journal of Statistics* **4**, 1097–1112.
- Kuchenhoff, H. & Carroll, R. J. (1997), ‘Segmented regression with errors in predictors: semi-parametric and parametric methods’, *Statistics in medicine* **16**(2), 169–188.
- Kulathinal, S. B., Kuulasmaa, K. & Gasbarra, D. (2002), ‘Estimation of an errors-in-variables regression model when the variances of the measurement error vary between the observations’, *Statistics in medicine* **21**(8), 1089–1101.

- Kuulasmaa, K., Tunstall-Pedoe, H., A., D., S., F., Tolonen, H., Evans, A., Ferrario, M. & J., T. (2000), 'Estimation of contribution of changes in classic risk factor to trends in coronary-event rates across the who monica project populations', *The Lancet* **355**(9205), 675–687.
- Lange, K. L., Little, R. J. A. & Taylor, J. M. G. (1989), 'Robust statistical modeling using the t distribution', *Journal of the American Statistical Association* **84**(408), 881–896.
- Lehmann, E. L. & Casella, G. (1998), *Theory of point estimation*, second edn, Springer-Verlag, New York.
- Lin, X. & Breslow, N. E. (1996), 'Bias correction in generalized linear mixed models with multiple components of dispersion', *Journal of the American Statistical Association* **91**(435), 1007–1016.
- Maddala, G. (1988), *Introduction to econometrics*, Macmillan.
- Magnus, J. R. & Neudecker, H. (1999), *Matrix differential calculus with applications in statistics and econometrics*, John Wiley & Sons Ltd., Chichester.
- Miller, R. G. (1986), *Beyond ANOVA, basics of applied statistics*, John Wiley & Sons Inc., New York.
- Mosteller, F. & Tukey, J. (1977), *Data analysis and regression: a second course in statistics*, Addison-Wesley Pub. Co.
- Muirhead, R. J. (1982), *Aspects of multivariate statistical theory*, John Wiley & Sons Inc., New York.
- Patriota, A. G. & Lemonte, A. J. (2009), 'Bias correction in a multivariate normal regression model with general parameterization', *Statistics & Probability Letters* **79**(15), 1655–1662.
- Patriota, A. G., Lemonte, A. J. & Bolfarine, H. (2011), 'Improved maximum likelihood estimators in a heteroskedastic errors-in-variables model', *Statistical Papers* **52**(2), 455–467.
- Pettitt, A. N., Kelly, J. M. & Gao, J. T. (1998), 'Bias correction for censored data with exponential lifetimes', *Statistica Sinica* **8**(3), 941–963.
- Quenouille, M. H. (1956), 'Notes on bias in estimation', *Biometrika* **43**, 353–360.

- R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: <https://www.R-project.org/>
- Rogers, W. H. & Tukey, J. W. (1972), ‘Understanding some long-tailed symmetrical distributions’, *Statistica Neerlandica. Journal of the Netherlands Society for Statistics and Operations Research* **26**(3), 211–226. Collection of articles in honour of Professor H. C. Hamaker.
- Sartori, N. (2006), ‘Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions’, *Journal of Statistical Planning and Inference* **136**(12), 4259–4275.
- Schafer, D. W. & Purdy, K. G. (1996), ‘Likelihood analysis for errors-in-variables regression with replicate measurements’, *Biometrika* **83**(4), 813–824.
- Spanos, A. (1994), ‘On modeling heteroskedasticity: the Student’s t and elliptical linear regression models’, *Econometric Theory* **10**(2), 286–315.
- Sutradhar, B. C. (1986), ‘On the characteristic function of multivariate Student t -distribution’, *The Canadian Journal of Statistics. La Revue Canadienne de Statistique* **14**(4), 329–337.
- Sutradhar, B. C. (1988), ‘Author’s revision: “On the characteristic function of multivariate Student t -distribution” [Canad. J. Statist. **14** (1986), no. 4, 329–337; MR0876759 (88g:62108)]’, *The Canadian Journal of Statistics. La Revue Canadienne de Statistique* **16**(3), 323.
- Taylor, J. M. G. (1992), ‘Properties of modelling the error distribution with an extra shape parameter’, *Computational Statistics & Data Analysis* **13**(1), 33–46.
- Tosteson, T. D., Stefanski, L. A. & Schafer, D. W. (1989), ‘A measurement error model for binary and ordinal regression’, *Statistics in Medicine* **8**, 1139–1147.
- Ullah, A. & Zinde-Walsh, V. (1984), ‘On the robustness of LM, LR, and W tests in regression models’, *Econometrica. Journal of the Econometric Society* **52**(4), 1055–1066.
- Vasconcellos, K. L. P. & Cordeiro, G. M. (2000), ‘Bias corrected estimates in multivariate student t regression models’, *Communications in Statistics - Theory and Methods* **29**(4), 797–822.

- Vidal, I. & Castro, L. M. (2010), ‘Influential observations in the independent Student- t measurement error model with weak nondifferential error’, *Chilean Journal of Statistics* **1**(2), 17–24.
- Wang, J. & Genton, M. G. (2006), ‘The multivariate skew-slash distribution’, *Journal of Statistical Planning and Inference* **136**(1), 209–220.
- Wang, N., Carroll, R. J. & Liang, K. Y. (1996), ‘Quasilikelihood and variance functions in measurement error models with replicates’, *Biometrics* **52**, 423–432.
- Wansbeek, T. & Meijer, E. (2000), *Measurement error and latent variables in econometrics*, North-Holland.
- Zellner, A. (1976*a*), ‘Bayesian and non-Bayesian analysis of the regression model with multivariate Student- t error terms’, *Journal of the American Statistical Association* **71**(354), 400–405.
- Zellner, A. (1976*b*), ‘Corrigendum: “Bayesian and non-Bayesian analysis of the regression model with multivariate Student- t error terms” (J. Amer. Statist. Assoc. **71** (1976), no. 354, 400–405)’, *Journal of the American Statistical Association* **71**(356), 1010.