PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE MATEMÁTICAS/DEPARTAMENTO DE ESTADÍSTICA

MODELO CON ERROR DE MEDICIÓN TWO-PIECE NORMAL

por

Karol I. Santoro Pizarro

Tesis presentada a la Facultad de Matemáticas de la
Pontificia Universidiad Católica de Chile,
como un requisito para optar al grado de
Doctor en Estadística.

Profesor Guía : **Reinaldo Boris Arellano Valle**

Septiembre 2019
Santiago, Chile.

AGRADECIMIENTOS

A Dios por ser mi guía y apoyo,
a mi familia por su comprensión
y constante aliento.

A mi profesor guía
Dr. Reinaldo B. Arellano-Valle,
por su paciencia y colaboración
en el desarrollo de esta tesis.
A los investigadores
Dr. Adelchi Azallini y Dr. Clecio S. Ferreira
por el apoyo que entregaron en este trabajo.

A la Pontifica Universidad Católica de Chile,
en especial al Departamento de Estadística
y a su Programa de Postgrado por darme
la oportunidad de estudiar este doctorado.

RESUMEN

En este trabajo se discute el modelo simple con errores de medición y una extensión multivariada, donde se considera principalmente el modelo estructual, suponiendo que el error de la regresión de respuesta sigue una distribución de dos piezas. Después de configurar una fórmula general para distribuciones de dos piezas, nos centramos en el caso donde la densidad base es una distribución normal. Lo interesante de usar como densidad base la distribución normal, es que el desarrollo entrega una mezcla de dos componentes de distribuciones skew-normal multivariada. Esta conexión facilita la construcción de un algoritmo tipo EM para realizar la estimación de máxima verosimilitud. Se obtiene la función de probabilidad de los datos observados, que se puede maximizar mediante el uso de software estadístico existente. Inferencia sobre los parámetros de interés puede ser abordado mediante el uso de la matriz de información observada, que también se puede calcular mediante el uso de software estadístico existente. Finalmente, se realiza algunas ilustraciones numéricas de la metodología, utilizando datos simulados y reales.

Índice general

Capítulo 1

Introducción

La idea de construir una línea recta para datos bivariados (x, y) es un procedimiento común en todas las disciplinas. La teoría estándar de una regresión lineal es dada cuando existe solo un error en una de sus variables, ya sea en x o en y. En particular, una regresión lineal simple a de tener la siguiente forma,

$$y_i = \alpha + \beta x_i + \varepsilon_i, \tag{1.1}$$

$i = 1, \ldots, n$, donde y_i es una variable de respuesta y x_i es una variable explicativa que puede observarse sin error, α y β, denominados como el intercepto y la pendiente de la regresión, son parámetros fijos y desconocidos, y ε_i es una variable aleatoria que representa el error de modelación.

El modelo lineal (1.1) puede ser extendido a modelos con errores en los predictores, denominados en la literatura estadística como modelos con errores en las variables (MEV). A continuación los presentamos en una forma muy general, ya que de esta manera las ideas resultan más claras.

Supongamos que hay n individuos en una muestra con valores verdaderos (ξ_i, η_i) y valores observados (x_i, y_i). Se supone que existe una relación lineal entre ξ_i y η_i,

$$\eta_i = \alpha + \beta x_i.$$

Sin embargo, en vez de observar ξ_i y η_i, observamos

$$x_i = \xi_i + \delta_i, \tag{1.2}$$

$$y_i = \eta_i + \varepsilon_i, \tag{1.3}$$

$i = 1 \ldots, n$, donde δ_i y ε_i son errores de medición asumidos independientes entre si y de la variable latente ξ_i, la cual representa la verdadera variable explicativa (covariable), mientras que x_i es la covariable observada.

Existen dos versiones del modelo con error de medición, dependiendo de los supuestos acerca de la variable regresora no observable. Si $\xi_1, \ldots, \xi_n$, son constantes desconocidas (parámetros incidentales), el modelo se denomina modelo con error de medición funcional, mientras que, si $\xi_1, \ldots, \xi_n$, son variables aletorias independientes e identicamente distribuidas e independientes de los errores de medición, el modelo es conocido como modelo de error de medición estructural. En el caso funcional es típico considerar que

$$\begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} \overset{iid}{\sim} N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\delta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix} \right),$$

$i = 1, \ldots, n$. En el caso estructural por lo general se supone que

$$\begin{aligned} \varepsilon_i &\overset{iid}{\sim} N(0, \sigma_\varepsilon^2), \\ \delta_i &\overset{iid}{\sim} N(0, \sigma_\delta^2), \\ \xi_i &\overset{iid}{\sim} N(\mu_\xi, \sigma_\xi^2), \end{aligned}$$

$i = 1, \ldots, n$, y que todos ellos son independientes entre si. Es bien sabido que el modelo normal estructural anterior no es identificable; ver Fuller (1987) y Kendall & Stuart (1979). En esta tesis, trabajaremos bajo el enfoque estructural. De esta forma, los supuestos de normalidad especificados arriba, nos conducen al modelo estructural normal definido por

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} \sim N\left(\begin{bmatrix} \alpha + \beta\mu_\xi \\ \mu_\xi \end{bmatrix}, \begin{bmatrix} \beta^2\sigma_\xi^2 + \sigma_\varepsilon^2 & \beta\sigma_\xi^2 \\ \beta\sigma_\xi^2 & \sigma_\xi + \sigma_\delta^2 \end{bmatrix} \right), \tag{1.4}$$

$i = 1, \ldots, n$. En general, se asume que $\mathbb{E}(\delta_i) = \mathbb{E}(\xi_i) = 0$ y que $\mathbb{V}ar(\delta_i) = \sigma_\delta^2$, $\mathbb{V}ar(\varepsilon_i) = \sigma_\varepsilon^2$, y que los errores δ_i y ε_i no están correlacionados:

$$\begin{aligned} Cov[\delta_i, \delta_j] &= Cov[\varepsilon_i, \varepsilon_j] = 0, \quad i \neq j, \\ Cov[\delta_i, \varepsilon_j] &= 0, \qquad \forall \ i, j. \end{aligned}$$

Es posible reescribir el modelo anterior como

$$y_i = \alpha + \beta x_i + (\varepsilon_i - \beta\delta_i),$$

$i = 1, \ldots, n$. Esto hace la diferencia entre este problema y una regresión lineal simple. El término del error es claramente dependiente de β. Además el término $(\varepsilon_i - \beta\delta_i)$ esta correlacionado con x_i. En efecto

$$Cov[x_i, \varepsilon_i - \beta\delta_i] = \mathbb{E}[x_i(\varepsilon_i - \beta\delta_i)] = \mathbb{E}[(\xi_i + \delta_i)(\varepsilon_i - \beta\delta_i)] = -\beta\sigma_\delta^2,$$

la cual es cero si $\beta = 0$ ó $\sigma_\delta^2 = 0$. Si $\sigma_\delta^2 = 0$, entonces $x = \xi$, es decir la covariable se mide sin error y por lo tanto el modelo es equivalente a la regresión estándar.

Como consecuencia del resultado anterior, en la presencia de errores de medida en la covariable, el ajuste del modelo ingenuo que no toma en cuenta los errores de medición puede conducir a estimaciones inconsistentes de los parámetros de regresión. Ver, Fuller (1987) para mayores detalles.

Los MEV han sido bastante estudiados, bajo el supuesto de normalidad. Mas recientemente, también han sido estudiados considerando otras distribuciones más flexibles que la normal. Vea por ejemplo, Arellano -Valle & Bolfarine (1996), Arellano-Valle et al. (1996) y Arellano-Valle et al. (2005b) donde los MEV son construidos usando distribuciones multivariadas simétricas y no simétricas. En este último trabajo se construye un MEV usando la distribución skew-normal multivariada propuesta por Azzalini & Dalla-Valle (1996). En esta tesis el principal objetivo es desarrollar un MEV, asumiendo que algunos de los errores o bien la covariable no-observada tiene distribución de dos piezas. A través de este supuesto se construye un MEV mucho mas flexible que el MEV normal usual. Como ya fue mencionado, nuestro estudio esta basado en el enfoque estructural.

Capítulo 2

Distribuciones Asimétricas

2.1. Distribución two-piece

El término "distribución de dos piezas"(o two-piece) se refiere a un mecanismo para construir una familia de densidades asimétricas a partir de una densidad simétrica, f, en la línea real. La densidad f se llamará densidad base TP. Asumiendo por conveniencia que f es simétrica alrededor de 0, el mecanismo TP funciona estirando f en una media línea y contrayéndolo en la otra media línea, con una escala adecuada para que la función resultante sea continua en 0 y aún su integral sea 1. El mecanismo de estiramiento y contracción generalmente está regulado por algún parámetro de asimetría, indicado como γ.

La referencia más antigua sobre una distribución de éste tipo es en el Capítulo XIX de Fechner (1897), quien propuso construir una distribución normal de dos piezas pegando dos normales truncadas con diferentes parámetros de escala, la densidad en este caso esta dada por

$$h(x;\mu,\sigma_1,\sigma_2) = \frac{2}{\sigma_1+\sigma_2}\left\{\phi\left(\frac{x-\mu}{\sigma_1}\right)I_{[\mu,\infty)}(x) + \phi\left(\frac{x-\mu}{\sigma_2}\right)I_{(-\infty,\mu)}(x)\right\},\ x\in\mathbb{R}, \tag{2.1}$$

donde $\phi(x)$ es la densidad de la distribución normal estandarizada, $\mu \in \mathbb{R}$ es un parámetro de localización, $\sigma_1, \sigma_2 > 0$ son parámetros de escala y $I_A(x)$ es la función indicadora. Este modelo fue propuesto posteriormente por Gibbons & Mylroie (1973), John (1982), Mudholkar & Hutson (2000), Fernandez & Steel (1998) entre otros, quienes han utilizado una parametrización alternativa pero equivalente. Wallis (2013) ha presentado una recopilación de redescubrimientos de la distribución normal de dos piezas (TPN). Otros autores han considerado extensiones de la construcción TP original utilizando densidades base distintas de la normal. Hansen (1994) ha considerado una distribución de TP con una densidad base t-Student como componentes paramétricos flexibles para modelos de densidad condicional autorregresivos. Las distribuciones de TP Laplace han aparecido en varios lugares de la literatura, a veces moviéndose desde una lógica completamente diferente, por lo que la conexión con el mecanismo de TP ni siquiera se menciona; un ejemplo de este tipo es el artículo de Koneker & Machado (1999). Una revisión de la distribución de TP Laplace se proporciona en el Capítulo 3 de Kotz & Kozubowski (2001). Fernandez & Steel (1998) presentan una construcción en términos más generales utilizando una densidad de referencia arbitraria, pero con especial énfasis en el caso especial de la densidad base t-Student.

La formula de TP también ha sido considerada por Arellano-Valle et al. (2005a) quienes introducen dos "funciones asimétricas" arbitrarias que regulan las acciones de estiramiento y contracción en las dos medias líneas. Además, considera un parámetro de localización μ y un parámetro de escala σ ($\sigma > 0$), de modo que la densidad base se escribe como

$$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}. \tag{2.2}$$

Jones (2006) ha señalado que la familia de distribuciones generada de esa forma no constituye un nivel adicional de generalidad con respecto a otras construcciones existentes, pero equivale a una reparametrización unificadas de otras familias una vez que se mantiene la proporción de las dos funciones de asimetría fijas, por lo tanto, hay realmente una sola función de asimetría. Estamos de acuerdo con su análisis, y luego explicaremos en detalle cómo las diferentes opciones de las funciones de asimetría pueden llevar a distribuciones equivalentes. Sin embargo, por varias razones, todavía preferimos trabajar con dos funciones de asimetría, denotadas (s, t). Un aspecto conveniente es la simplicidad de tratamiento en las secciones posteriores, y el hecho de que esta elección permite ver de inmediato como reconstruir a partir de versión unificada una serie de especificaciones existentes. Otra ventaja es que este esquema facilita el uso de las propiedades enumeradas más adelante, después de la Definición 2.1.1, bajo cualquier elección de las funciones (s, t) que uno prefiere adoptar.

Definición 2.1.1. *Dada una familia de localización y escala del tipo 2.1, donde la función de densidad f satisface las condiciones de simetría $f(z) = f(-z)$ para todo z, decimos que una v.a. continua Y sigue una distribución TP con densidad base f si su función de densidad es de la forma*

$$h(y) = \frac{2}{\sigma(s(\gamma)+t(\gamma))}\left\{f\left(\frac{y-\mu}{\sigma s(\gamma)}\right) I_{[\mu,\infty)}(y) + f\left(\frac{y-\mu}{\sigma t(\gamma)}\right) I_{(-\infty,\mu)}(y)\right\}, \quad y \in \mathbb{R}, \tag{2.3}$$

donde $I_A(x)$ representa la función indicadora del conjunto A, $s(\gamma)$ y $t(\gamma)$ son funciones de asimetría que satisfacen las siguientes propiedades:

(i) $s(\gamma)$ y $t(\gamma)$ son funciones que toman valores positivos para $\gamma \in (\gamma_L, \gamma_U)$, un posible intervalor infinito;

(ii) una de las funciones está (estrictamente) aumentando y la otra está (estrictamente) disminuyendo;

(iii) existe un valor γ_0 tal que $s(\gamma_0) = t(\gamma_0)$.

Usaremos la notación $Y \sim TPf(\mu, \sigma^2, \gamma)$ para referirnos a esta distribución. Cuando $\mu = 0$ y $\sigma = 1$, la distribución se denotara como $Y \sim TPf(\gamma)$. Cuando $\gamma = \gamma_0$, recuperamos la simetría, donde $h(y)$ se reduce a la densidad (2.2). A partir de la condición (ii), es inmediato que el valor γ_0 es único, y no hay pérdida de generalidad al suponer que $s(\gamma_0) = t(\gamma_0) = 1$.
Los casos específicos de (s, t) utilizados en la literatura son:

(a) $(s, t) = (1 + \gamma, 1 - \gamma)$, donde $\gamma \in (-1, 1)$ y $\gamma_0 = 0$, como en Mudholkar & Hutson (2000);

(b) $(s, t) = (\gamma, 1/\gamma)$, donde $\gamma > 0$ y $\gamma_0 = 1$, como en Fernandez & Steel (1998);

(c) $(s, t) = (\gamma, 1 - \gamma)$, donde $\gamma \in (0, 1)$ y $\gamma_0 = 1/2$, como en Yu & Zhang (2013).

La construcción de la distribución TP disfruta de varias propiedad formales, que se enumeran a continuación. Aunque algunos de estos hechos ya se han mencionado en la literatura citada, parece útil tenerlos reunidos en un solo lugar. Además, algunas de estas propiedades parecen no haber aparecido en otro lado. Uno de estos casos es la propiedad P1 que, aunque es elemental, es importante para nuestro desarrollo posterior.

2.2. Propiedades

P1. La densidad TP (2.3) puede escribirse como la mezcla de dos componenetes de densidades truncadas

$$f_s(y) = \frac{2}{\sigma\, s(\gamma)} f\left(\frac{y-\mu}{\sigma\, s(\gamma)}\right) I_{[\mu,\infty)}(y), \qquad f_t(y) = \frac{2}{\sigma\, t(\gamma)} f\left(\frac{y-\mu}{\sigma\, t(\gamma)}\right) I_{(-\infty,\mu)}(y), \tag{2.4}$$

en la configuración

$$h(y) = \pi_s f_s(y) + \pi_t f_t(y), \qquad y \in \mathbb{R}, \tag{2.5}$$

donde

$$\pi_s = \frac{s(\gamma)}{s(\gamma)+t(\gamma)}\,, \qquad \pi_t = \frac{t(\gamma)}{s(\gamma)+t(\gamma)}\,. \tag{2.6}$$

Una implicación de este hecho es una representación estocástica de la distribución TP. Otro uso es un enfoque simple para examinar la equivalencia formal, mencionada anteriormente, de diferentes opciones (s,t) que tienen igual proporción, $q = q(\gamma) = s(\gamma)/t(\gamma)$, por ejemplo. Para ello, se escribe las probabilidades (2.6) como $\pi_s = q/(q+1)$ y $\pi_t = 1/(q+1)$ y observe que el conjunto de componentes $(\mu, \sigma, s(\gamma), t(\gamma))$ en (2.4)-(2.5) corresponde a la misma distribución del conjunto $(\mu, \omega, q(\gamma), 1)$, donde $\omega = \sigma t(\gamma)$. Esta equivalencia significa que una especificación de tres componentes de la distribución es suficiente. Claramente, el caso simétrico se recupera cuando $q(\gamma) = 1$.

Debido a la condición de monotonía (ii), $q(\gamma)$ es una función monótona, de modo que $q = q(\gamma)$ puede usarse como parámetro en sí mismo. Esto justifica el uso de la notación más corta q que aveces adoptaremos.

P2. Otra representación estocástica de $Y \sim h$ es la siguiente, donde tomamos $\mu = 0$ y $\sigma = 1$ por simplicidad. Dada una variable aleatoria continua V con función

de densidad $2f(v)I_{[0,\infty)}(v)$ y una variable discreta independiente W_γ que tiene función de probabilidad

$$p(w;\gamma) = \begin{cases} \pi_s & \text{si } w = s(\gamma), \\ \pi_t & \text{si } w = -t(\gamma), \end{cases} \tag{2.7}$$

donde π_s y π_t se definen en (2.6), luego $Y \stackrel{d}{=} W_\gamma V$, donde la notación $\stackrel{d}{=}$ indica igualdad en distribución. De manera equivalente, podemos decir: si $Y = \mu + \sigma W_\gamma |X|$, donde $X \sim f$ y es independiente a W_γ, entonces $Y \sim h$.

Esta representación también aparece en Bauwens & Laurent (2005) y constituye el ingrediente clave para la versión multivariada de distribuciones de TP.

P3. Los momentos centrales de $Y \sim h$ se pueden expresar en términos de los momentos de $X \sim f$. Para este fin, el álgebra se simplifica sustancialmente si se trabaja con las cantidades $q = s/t$ y $\omega = \sigma t$. Entonces, se puede mostrar que, para un entero m no negativo,

$$\mathbb{E}\{Y^m\} = \frac{1}{q+1} \sum_{k=0}^{m} \binom{m}{k} \mu^k \omega^{m-k} d_{m-k} \left(q^{m-k+1} + (-1)^{m-k}\right), \tag{2.8}$$

$$\mathbb{E}\{(Y - \mathbb{E}\{Y\})^m\} = \frac{\omega^m}{q+1} \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} d_k \, d_{m-k} \left(q^{k+1} + (-1)^k\right) (q-1)^{m-k}, \tag{2.9}$$

previsto de $d_m < \infty$, donde

$$d_k = \mathbb{E}|X|^k = 2\int_0^\infty x^k f(x)dx, \qquad k = 0, 1, \ldots$$

Si $\mu = 0$, (2.8) se reduce a

$$\mathbb{E}\{Y^m\} = \frac{\omega^m}{q+1} d_m \left(q^{m+1} + (-1)^m\right).$$

P4. En el caso estandarizado con $\mu = 0$ y $\sigma = 1$, la función de distribución de h es

$$H(x;\gamma) = \begin{cases} 2\pi_t\, F\left(\dfrac{x}{t(\gamma)}\right), & si \quad x < 0 \\ \pi_t - \pi_s + 2\,\pi_s F\left(\dfrac{x}{s(\gamma)}\right), & si \quad x \geq 0 \end{cases}$$

donde F denota la función de distribución de f. En particular, obtenemos

$$H(0;\gamma) = \pi_t, \qquad \frac{\mathbb{P}\{X \geq 0|\gamma\}}{\mathbb{P}\{X < 0|\gamma\}} = \frac{s(\gamma)}{t(\gamma)} = q(\gamma) = \frac{\pi_s}{\pi_t}.$$

lo que muestra que $q(\gamma) > 1$ implica una inclinación hacia la derecha de la distribución y lo contrario sucede cuando $q(\gamma) < 1$. La mediana es

$$H^{-1}(1/2;\gamma) = \begin{cases} t(\gamma)\, F^{-1}\left(\dfrac{1}{4\pi_t}\right), & si \quad s(\gamma) < t(\gamma) \\ s(\gamma)\, F^{-1}\left(1 - \dfrac{1}{4\pi_s}\right), & c.o.c. \end{cases}$$

P5. Si f es unimodal, entonces también lo es h, con media μ. Rubio & Steel (2014) demostraron que en caso de una densidad base unimodal f, con algunas condiciones débiles adicionales, la matriz de información de Fisher para el modelo (2.3) es no-singular.

P6. Si cambiamos las funciones (s,t) a (t,s) en (2.3), esto equivale a considerar la distribución de $Y' = \mu - (Y - \mu)$, es decir, la variable Y se obtiene al reflejar μ en su lado opuesto en el eje de las abscisas. En cuanto a la parametrización (μ, ω, q), que se utilizará, por ejemplo, en (2.8) y (2.9), usando cálculos algebraicos simples, muestra que Y' tiene parámetros $(\mu, \omega, 1, 1/q)$. Una propiedad de Y' es

$$\text{median}(Y' - \mu) = -\text{median}(Y - \mu)\,.$$

P7. Si elegimos $s(\gamma) = 1 + \gamma$ y $t(\gamma) = 1 - \gamma$, se obtiene el caso particular

$$h(y) \quad = \quad f\left(\frac{y}{1+\gamma}\right) I_{[0,\infty)}(y) + f\left(\frac{y}{1-\gamma}\right) I_{(-\infty,0)}(y), \quad x \in \mathbb{R}. \qquad (2.10)$$

De (2.10), se tiene las siguientes propiedades:

a) $h(y;\gamma = 0) = f(y)$,

b) $\lim\limits_{\gamma\to\infty} h(y;\gamma) = 0$,

c) $\lim\limits_{\gamma\to 1} h(y;\gamma) = f\left(\frac{y}{2}\right) I_{[0,\infty)}(y)$ y $\lim\limits_{\gamma\to -1} h(y;\gamma) = f\left(\frac{y}{2}\right) I_{(-\infty,0)}(y)$.

P8. Para $f = \phi(y)$ y $\gamma = \varepsilon$, (2.10) corresponde a la densidad de la distribución ε-skew-normal de Mudholkar & Hutson (2000).

2.3. Algunos casos especiales

La tabla 2.1 entrega resultados específicos para casos particulares de densidades f:

(i) Exponencial potencia:$f(x) = c_1 \exp\{-\frac{1}{2}|x|^{2/(1+\beta)}\}$, con $\beta > -1$ y $c_1^{-1} = 2^{(3+\beta)/2}\Gamma((3+\beta)/2)$, $x \in \mathbb{R}$;

(ii) t-Student:$f(x) = c_2(1+x^2/\nu)^{-(1+\nu)/2}$, con $\gamma > 0$ y $c_2 = (\nu\pi)^{-1/2}\Gamma(\nu/2)^{-1}\Gamma((\nu+1)/2)$, $x \in \mathbb{R}$;

(iii) Logística: $f(x) = e^{-x}/(1+e^{-x})^2$, $x \in \mathbb{R}$;

(iv) Gama simetrizada:$f(x) = (2\Gamma(\nu))^{-1}\beta^{\nu}|x|^{\nu-1}e^{-\beta|x|}$, con $\beta > 0$ y $\nu > 0$, $x \in \mathbb{R}$.

Observación 2.3.1. *La distribución exponencial potencia (EP) contiene a las distribuciones normal y doble exponencial cuando $\beta = 0$ y $\beta = 1$ respectivamente. También*

se obtiene mayor flexibilidad en los coeficientes de asimetría y kurtosis. Para $\beta > 0$ se tienen colas mas pesadas que la distribución normal, y ocurre lo contrario cuando $\beta < 0$.

Distribución	d_k	$\mathbb{E}(Z)$	$\mathbb{V}ar(Z)$
$TPEP(\gamma,\beta)$	$\frac{2^{r(1+\beta)/2}\Gamma((1+\beta)(1+r)/2)}{\Gamma((1+\beta)/2)}$	$\mu - \frac{2^{(3+\beta)/2}\gamma\sigma\Gamma(1+\beta)}{\Gamma((1+\beta)/2)}$	$\frac{\sigma^2 2^{1+\beta}}{\Gamma((1+\beta)/2)}\left\{(1+3\gamma^2)\Gamma\left(\frac{3(1+\beta)}{2}\right) - \frac{4\gamma^2\Gamma(1+\beta)^2}{\Gamma((1+\beta)/2)}\right\}$
$TPt(\gamma,\nu)(*)$	$2c_2\nu^{(r+1)/2}\sum_{j=0}^{\frac{(r-1)}{2}}\frac{(-1)^{r-1-2j}}{(\nu-2j-1)}\begin{pmatrix}\frac{(r-1)}{2}\\ j\end{pmatrix}$, si r es impar $\frac{\nu^{r/2}\cdot 1\cdot 3\cdots(r-1)}{(\nu-r)(\nu-r-2)\cdot(\nu-2)}$, si r es par	$\mu - \frac{4c_2\nu\gamma\sigma}{\nu-1}$	$\sigma^2\left\{\frac{\nu(1+3\gamma^2)}{\nu-1} - \frac{16(c_2\gamma\nu)^2}{(\nu-1)^2}\right\}$
$TPL(\gamma)(**)$	$b_1 = 2\log(2)$ $b_r = \frac{2\Gamma(r+1)(2^r-2)\zeta(r)}{2^r}$, r=2,3,...	$\mu - 4\gamma\sigma\log(2)$	$\sigma^2\left\{\frac{(1+3\gamma^2)\pi^2}{3} - 16(\gamma\log(2))^2\right\}$
$TPGS(\beta,\nu)$	$\frac{\Gamma(r+\nu)}{\beta^r\Gamma(\nu)}$	$\mu - \frac{2\gamma\sigma\nu}{\beta}$	$\sigma^2\left\{\frac{(1+3\gamma^2)\nu(\nu+1)-4(\gamma\nu)^2}{\beta^2}\right\}$

Tabla 2.1: Momentos de $X \sim TPf$. $(*)$ En este caso $d_k < \infty$ si y sólo si $\nu > r$. $(**)$ $\zeta(r) = \sum_{j=1}^{\infty}\frac{1}{j^r}$ es la función Zeta de Riemann, $r = 2, 3, \ldots$

La Figura 2.1 muestra los gráficos para las densidades TPf para diferentes valores de $\gamma \in (-1, 1)$. El Gráfico (a) corresponde a la distribución TP$t_{(5)}(\gamma)$, la cual tiene colas

más pesadas que la distribución TPN(γ) estudiada por Mudholkar & Hutson (2000). El Gráfico (b) corresponde al caso cuando la densidad es TP$L(\gamma)$. Notemos que estos modelos no son simétricos con respecto al origen, y la asimetría depende del valor γ. Los demás gráficos representan varias formas que resultan cuando $f(x)$ es la distribución EP para diferentes valores de β.

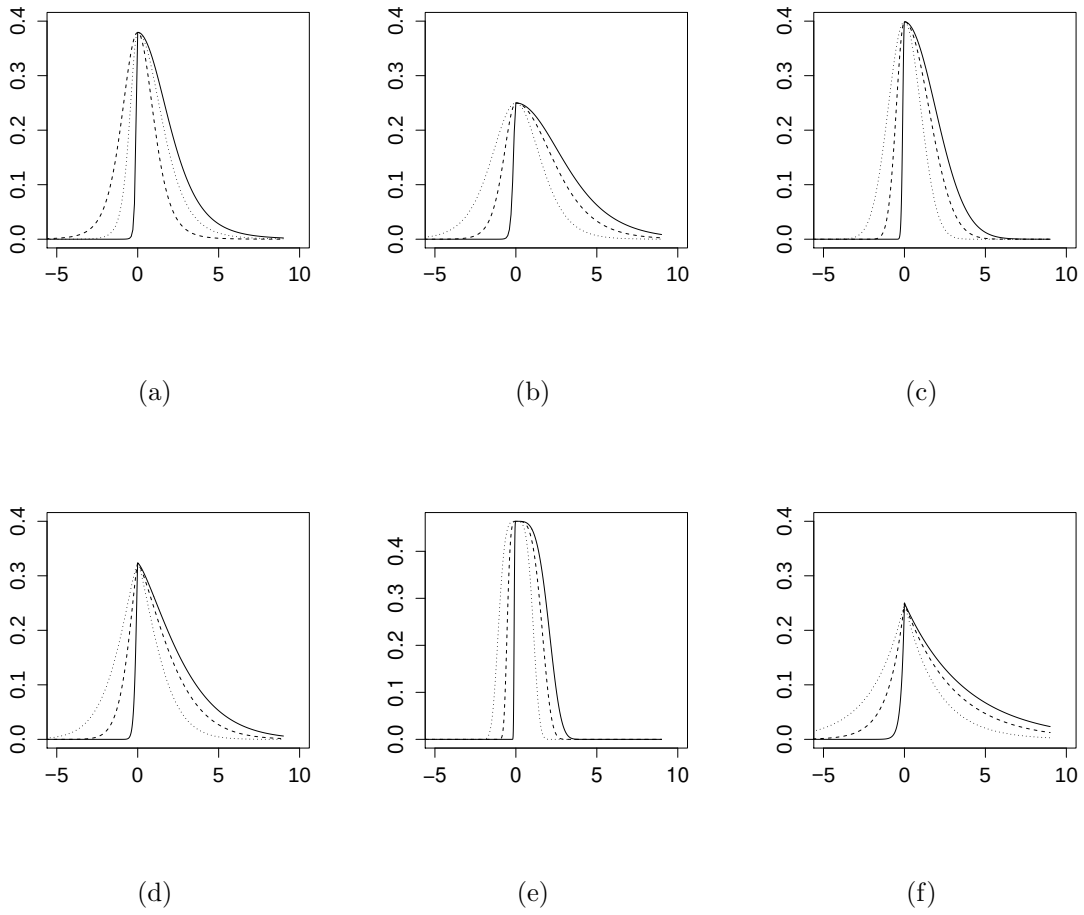

Figure 2.1: Ejemplos de la densidad TPf para $\gamma = 0$ (línea discontinua), γ=-0.5 (línea punteada) y γ=-0.9 (línea solida). Desde izquierda a derecha: (a) TPt_5; (b) TPL; (c) TP$EP(0, 1, \gamma, 0)$. En (d), (e) y (f) se muestran los casos TP$EP(0, 1, \gamma, \beta)$, con β=0.5, β=-0.5 y $\beta = 1$, respectivamente.

2.4. Estimación de máxima verosimilitud

La estimación de máxima verosimilitud (EMV) para la familia TP$f(\mu, \sigma, \gamma)$ es conceptualmente directa. En la práctica, sin embargo, calcular el EMV de $\theta = (\mu, \sigma, \gamma)$ representa alguna dificultad. Se ilustran estos cálculos, en el caso de la distribución TP$EP(\mu, \sigma, \gamma, \beta)$, asumiendo β conocido y tomando $s(\gamma) = 1 + \gamma$ y $t(\gamma) = 1 - \gamma$, $|\gamma| < 1$.

Sean $z_{(1)} \leq z_{(2)} \leq \ldots \leq z_{(n)}$ los estadísticos de orden asociados a una muestra aleatoria simple $Z_1, \ldots, Z_n$ desde la distribución TP$EP(\mu, \sigma, \gamma, \beta)$. Se denota $z_{(0)} = -\infty$ y $z_{(n+1)} = \infty$. Sea $k \equiv k(z_{(1)}, \ldots, z_{(n)}, \mu)$ un entero tal que $z_{(k)} < \mu < z_{(k+1)}$. Por la continuidad de la distribución TPEP, k es también definido con probabilidad 1, y el rango de variación es $\{0, 1, \ldots, n\}$. La función log-verosimilitud puede ser expresada como

$$\ell(\mu, \sigma^2, \gamma) \quad = \quad n \log(c_1) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^{2/(1+\beta)}} g_k(\mu, \gamma),$$

donde

$$g_k(\mu, \gamma) \quad = \quad \left\{ (1+\gamma)^{-2/(1+\beta)} \sum_{i=1}^{k} |z_{(i)} - \mu|^{2/(1+\beta)} + (1-\gamma)^{-2/(1+\beta)} \sum_{i=k+1}^{n} |z_{(i)} - \mu|^{2/(1+\beta)} \right\}. \quad (2.11)$$

El cálculo del EMV de $\boldsymbol{\theta}$ se necesita maximizar (2.11), en el cual se puede fijar primero k, y entonces se encuentra los correspondientes valores óptimos de $\widehat{\boldsymbol{\theta}}_k$. El EMV, es $\widehat{\boldsymbol{\theta}}_k^* = \arg\max \ell(\widehat{\boldsymbol{\theta}}_k)$ en el conjunto $k = 0, \ldots, n$. Para mayor detalle ver Arellano-Valle et al. (2005a) y Mudholkar & Hutson (2000).

2.5. Matriz de información de Fisher

En esta sección se aborda el cálculo de la matriz de información en el contexto de los modelos TP . Esta matriz fue calculada por Arellano-Valle et al. (2005a) para el modelo TPEP con $s(\gamma) = 1 + \gamma$ y $t(\gamma) = 1 - \gamma$. Posteriormente, ella es derivada por Rubio & Steel (2014) en el contexto del modelo (2.3) considerando condiciones adicionales sobre la densidad base y las funciones de asimetría.

Considerando la familia (2.3), Rubio & Steel (2014) asumen que las funciones de asimetría $s(\gamma) > 0$ y $t(\gamma) > 0$, $\gamma \in \Gamma$, son diferenciables y tales que

$$0 < |\lambda(\gamma)| \quad \text{con} \quad \lambda(\gamma) \equiv \log\left[\frac{s(\gamma)}{t(\gamma)}\right] \tag{2.12}$$

La condición (2.12) implica que la transformación de (σ_1, σ_2) para $(\sigma, s(\gamma), t(\gamma))$ es una aplicación no-singular, la cual es una condición necesaria para que la transformación sea uno a uno. Con esta reparametrización se obtiene (2.1) a partir de (2.3). Como se mencionó al inicio, un recuento histórico de las muchas formas de esta distribución se proporciona en Wallis (2013).

La matriz de información de Fisher para el modelo (2.3) es presentada en la siguiente proposición:

Proposición 2.5.1. *Sea f la densidad base TP. Suponga que f cumple las siguientes condiciones:*

1. $\int_0^\infty \left[\frac{f'(t)}{f(t)}\right]^2 f(t)dt < \infty$,
2. $\int_0^\infty t^2 \left[\frac{f'(t)}{f(t)}\right]^2 < \infty$,
3. $\lim\limits_{t\to\infty} tf(t) = 0$ *o* $\int_0^\infty tf'(t)dt = -\frac{1}{2}$.

Entonces la matriz de información de Fisher del modelo (2.3) es

$$I(\mu,\sigma,\gamma)=\begin{pmatrix} \frac{2\alpha_1}{s(\gamma)t(\gamma)\sigma^2} & 0 & \frac{2\alpha_3}{\sigma[s(\gamma)+t(\gamma)]}\left[\frac{s'(\gamma)}{t(\gamma)}-\frac{t'(\gamma)}{t(\gamma)}\right] \\ 0 & \frac{\alpha_2}{\sigma^2} & \frac{\alpha_2}{\sigma}\left[\frac{s'(\gamma)+t'(\gamma)}{s(\gamma)+t(\gamma)}\right] \\ \frac{2\alpha_3}{\sigma[s(\gamma)+t(\gamma)]}\left[\frac{s'(\gamma)}{s(\gamma)}-\frac{t'(\gamma)}{t(\gamma)}\right] & \frac{\alpha_2}{\sigma}\left[\frac{s'(\gamma)+t'(\gamma)}{s(\gamma)+t(\gamma)}\right] & \frac{\alpha_2+1}{s(\gamma)+t(\gamma)}\left[\frac{t'(\gamma)^2}{t(\gamma)}+\frac{s'(\gamma)^2}{s(\gamma)}\right]-\left[\frac{s'(\gamma)+t'(\gamma)}{s(\gamma)+t(\gamma)}\right]^2 \end{pmatrix},$$

donde

$$\begin{aligned} \alpha_1 &= \int_0^\infty \left[\frac{f'(t)}{f(t)}\right]^2 f(t)dt, \\ \alpha_2 &= 2\int_0^\infty \left[1+t\frac{f'(t)}{f(t)}\right]^2 f(t)dt = -1+2\int_0^\infty t^2\left[\frac{f'(t)}{f(t)}\right]^2 f(t)dt, \\ \alpha_3 &= \int_0^\infty t\left[\frac{f'(t)}{f(t)}\right]^2 f(t)dt. \end{aligned}$$

Note que las entradas I_{12} y I_{21} son ceros. Esto indica que se tiene ortogonalidad entre los paramétros μ y σ eligiendo cualquier parametrización para $\{s(\gamma),t(\gamma)\}$. Además, mediante la adecuada elección de $\{s(\gamma),t(\gamma)\}$, se puede generar mas ceros en las entradas de la matriz de información de Fisher. Una cosecuencia de esto se presenta en el siguiente corolario.

Corolario 2.5.1. *Si* $\frac{d}{d\gamma}\log[s(\gamma)+t(\gamma)]=0$, *entonces* $I_{23}=I_{32}=0$. *En particular si* $s(\gamma)+t(\gamma)$ *es constante, entonces* $I_{23}=I_{32}=0$.

Esto ocurre, por ejemplo, cuando $s(\gamma)=1+\gamma$ y $t(\gamma)=1-\gamma$, $|\gamma|<1$. En tal caso,

$$I(\mu,\sigma,\gamma)=\begin{pmatrix} \frac{2\alpha_1}{\sigma^2(1-\gamma^2)} & 0 & -\frac{2\alpha_3}{\sigma(1-\gamma^2)} \\ 0 & \frac{\alpha_2}{\sigma^2} & 0 \\ -\frac{2\alpha_3}{\sigma(1-\gamma^2)} & 0 & \frac{\alpha_2+1}{1-\gamma^2} \end{pmatrix}.$$

En particular, para $\gamma = 0$ (simetría) la matriz de información $I(\mu, \sigma, \gamma)$ es no sigular. Este resultado aporta una ventaja importante, por ejemplo, del modelo two-piece normal con respeccto al modelo skew-normal Azzalini (1985), ya que permite aplicar la teoría asintótica regular para testear la hipótesis de asimetría.

2.6. Otras distribuciones asimétricas

En esta sección entregaremos algunas propiedades de la distribución skew-normal(SN) multivariada y mezclas finitas de distribuciones SN, las que serán utilizadas en los capítulos posteriores.

2.6.1. Distribución skew normal multivariada

La distribución SN multivariada se ha parametrizado de varias maneras diferentes, vea las diversas alternativas equivalentes adoptadas por Azzalini & Dalla-Valle (1996), Azzalini & Capitano (1999), Arellano-Valle & Genton (2005), Arellano-Valle& Azzalini (2006). Por conveniencia matemática, en este trabajo adoptamos la parametrización, adoptada por Arellano-Valle & Genton (2005).

Para establecer la notación, denotamos por $\phi_d(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ la función de densidad de la distribución normal d-dimensional $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y por $\Phi_1(\cdot)$ a la función de distribución de la distribución normal estandar $N_1(0, 1)$, con derivada $\Phi' = \phi$. Así, dada una matriz simétrica $\boldsymbol{\Sigma}$ definida positiva $d \times d$, denotada por $\boldsymbol{\Sigma}^{1/2}$, su raíz cuadrada es simétrica definida positiva, de tal manera que $(\boldsymbol{\Sigma}^{1/2})^2 = \boldsymbol{\Sigma}$, donde $\boldsymbol{\Sigma}^{1/2}$ tiene una representación única. Luego decimos que un vector aleatorio d-dimensional $\boldsymbol{Y}$ tiene distribución SN, con parámetro de localización $\boldsymbol{\mu} \in \mathbb{R}^d$, parámetro de escala $\boldsymbol{\Sigma}$ y parámetro de forma $\boldsymbol{\lambda} \in \mathbb{R}^d$, su densidad viene dada por

$$f(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda}) \;=\; 2\phi_d(\boldsymbol{y};\boldsymbol{\mu},\boldsymbol{\Sigma})\Phi_1(\boldsymbol{\eta}^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})), \quad \boldsymbol{y}\in\mathbb{R}^d, \tag{2.13}$$

donde $\boldsymbol{\eta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$ se define de manera única gracias a la unicidad de $\boldsymbol{\Sigma}^{1/2}$. En este caso, usaremos la notación $\boldsymbol{Y} \sim SN_d(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda})$.

Un papel clave en nuestra formulación lo desempeña la denominada representación aditiva de las variables SN. Esta representación fue la ruta a través de la cual Azzalini & Dalla-Valle (1996) introdujeron la densidad SN multivariable (2.13), excepto por la diferencia ya mencionada en la parametrización. Esa construcción se puede redefinir en la siguiente forma equivalente, la que es más conveniente para el desarrollo posterior.

Teniendo en cuenta esta parametrización y la representación estocástica de la distribución multivariada SN, introducida por Arellano-Valle& Azzalini (2006), el vector aleatorio $\boldsymbol{Y} \sim SN_d(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda})$ puede ser especificada jerarquicamente en terminos de la normal multivariada y la distribuión half-normal, dada por

$$\boldsymbol{Y}|W=\omega \;\sim\; N_d(\boldsymbol{\mu}+\Delta\omega,\Psi), \tag{2.14}$$

$$\omega \;\sim\; HN_1(0,1). \tag{2.15}$$

donde la notación $HN_1(\mu,\sigma^2)$ denota la distribución obtenida por truncamiento de una variable $N(\mu,\sigma^2)$ bajo μ y

$$\Delta = \frac{1}{\sqrt{1+\boldsymbol{\eta}\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}}}\boldsymbol{\eta}, \qquad \boldsymbol{\Psi} = \boldsymbol{\Sigma} - \boldsymbol{\eta}\boldsymbol{\eta}^\top,$$

entonces la distribución marginal de $\boldsymbol{Y}$ tiene densidad (2.13).

La representación (2.14)-(2.15) nos dice que $\boldsymbol{Y}-\boldsymbol{\mu}$ tiene la misma ditribución que $\Delta\omega+\boldsymbol{\epsilon}$, donde $\boldsymbol{\epsilon}\sim N_d(\boldsymbol{0},\boldsymbol{\Psi})$ y es independiente de W, y es muy útil para estudiar

las propiedades probabilisticas de (2.13), como así para obtener una estimación de máxima verosimilitud a través del algoritmo EM. Además este hecho proporciona un camino para derivar varias propiedades formales de (2.13), por ejemplo, utilizando las propiedades de la esperanza condicional obtenemos fácilmente (2.14) y (2.15) donde el vector de medias y matriz de covarianzas de $\boldsymbol{Y}$ es dado por

$$\mathbb{E}\{\boldsymbol{Y}\} \;=\; \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\Delta \;\text{ y }\; V\text{ar}\{\boldsymbol{Y}\} = \boldsymbol{\Sigma} - \frac{2}{\pi}\Delta\Delta^{\top}. \tag{2.16}$$

Alternativamente, las mismas expresiones podrían obtenerse usando la función de generación de momento de $\boldsymbol{Y}$, que tiene una forma explícita simple.

Otro papel importante de la representación (2.14) - (2.15) es proporcionar un camino para abordar los EMV a través del algoritmo EM, o algunas de sus formas, ver por ejemplo Lacho et al. (2010) y Lin (2010) entre otros. También haremos uso de esta representación para el cálculo de EMV en nuestro desarrollo.

2.6.2. Función generadora de momentos

El siguiente resultado se ha presentado en varias ocasiones en la literatura para la distribución normal, con o sin demostraciones. Autores que han proporcionado una desmostración de este resultado son Ellison (1964) y Zacks (1981).

Lema 2.6.1. *Si* $\boldsymbol{U} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$*, entonces*

$$E\{\Phi(h^T\boldsymbol{U} + k)\} = \Phi\left(\frac{k}{\sqrt{1 + \boldsymbol{h}^T\boldsymbol{\Sigma}\boldsymbol{h}}}\right), \qquad \boldsymbol{h} \in \mathbb{R}^d, k \in \mathbb{R}. \tag{2.17}$$

Demostración: Ver Ellison (1964).

Este resultado por lo general se presenta para el caso univariado.

Lema 2.6.2. *Sea* $\boldsymbol{Y} \sim SN_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. *Entonces la función generatriz de momento es*

$$M(\boldsymbol{t}) = 2\exp\left\{\boldsymbol{t}^\top\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t}\right\}\Phi\left(\boldsymbol{\eta}^\top\boldsymbol{t}\right), \qquad \boldsymbol{t} \in \mathbb{R}^d. \tag{2.18}$$

donde

$$\boldsymbol{\eta} = (1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda})^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\lambda}.$$

Demostración: Por definición

$$M(\boldsymbol{t}) \;=\; 2\int_{\mathbb{R}^d} \exp(\boldsymbol{t}^\top\boldsymbol{y})\phi_d(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}))d\boldsymbol{y}.$$

Sea $\boldsymbol{k} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})$. Se tiene que $d\boldsymbol{y} = |\boldsymbol{\Sigma}|^{1/2}d\boldsymbol{k}$, de modo que

$$\begin{aligned}
M(\boldsymbol{t}) &= \exp\{\boldsymbol{t}^\top\boldsymbol{\mu}\}\int_{\mathbb{R}^d} \exp\{\boldsymbol{t}^\top\boldsymbol{\Sigma}^{1/2}\boldsymbol{k}\}\phi_2(\boldsymbol{k}|0, \boldsymbol{\Sigma})\Phi(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{k})d\boldsymbol{k}, \\
&= \frac{2\exp\{\boldsymbol{t}^\top\boldsymbol{\mu}\}}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}}\int_{\mathbb{R}^d} \exp\left\{-\frac{1}{2}(\boldsymbol{k}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{k} - 2\boldsymbol{t}^\top\boldsymbol{\Sigma}^{1/2}\boldsymbol{k})\right\}\Phi(\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}\boldsymbol{k})d\boldsymbol{k}, \\
&= 2\exp\left\{\frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}t + \boldsymbol{t}^\top\boldsymbol{\mu}\right\}E\left[\Phi(\boldsymbol{\lambda}^\top(U + \boldsymbol{\Sigma}\boldsymbol{t}))\right], \\
&= 2\exp\left\{\frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t} + \boldsymbol{t}^\top\boldsymbol{\mu}\right\}\Phi\left(\frac{\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{t}}{\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}\boldsymbol{\lambda}}}\right), \\
&= 2\exp\left\{\boldsymbol{t}^\top\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t}\right\}\Phi\left(\boldsymbol{\eta}^\top\ \boldsymbol{t}\right), \quad \boldsymbol{t} \in \mathbb{R}^d.
\end{aligned}$$

El resultado sigue del Lema (2.6.1).

■

Para mas propiedades de la distribución SN multivariada, vea Azzalini & Dalla-Valle (1996) y Azzalini & Capitano (1999).

2.6.3. Mezcla de distribuciones skew normal

Una distribución de mezclas finitas es obtenida como un promedio ponderado de densidades $f_1, \ldots, f_K$ con pesos no negativos $\pi_1, \ldots, \pi_K$, que representa la probabilidad de las subpoblaciones, de modo que $\sum_{k=1}^{K} \pi_k = 1$. Luego una densidad de mezclas finitas toma la siguiente forma:

$$f(\boldsymbol{y}; \boldsymbol{\theta}, \pi) \quad = \quad \sum_{k=1}^{K} \pi_k f_k(\boldsymbol{y}|\boldsymbol{\theta}_k), \quad \boldsymbol{y} \in \mathbb{R}^d,$$

donde $\boldsymbol{\theta}_k$ representa el vector de parámetros de la k-ésima subpoblación, $\pi = (\pi_1, \ldots, \pi_K)$ y $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$. Una revisión completa de los modelos de mezcla finita es el libro de Mclachlan & Peel (2000).

Si $f_k(\boldsymbol{y})$ es la densidad SN dada en (2.13), obtenemos una mezcla finita de distribuciones SN multivariadas, cuya densidad es dada por

$$f(\boldsymbol{y}; \boldsymbol{\theta}, \pi) \quad = \quad 2\sum_{k=1}^{K} \pi_k \phi_p(\boldsymbol{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\Phi(\boldsymbol{\lambda}_k \boldsymbol{\Sigma}_k^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu}_k)), \quad \boldsymbol{y} \in \mathbb{R}^d, \tag{2.19}$$

donde $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{\lambda}_k)$. Se denotara esto por $\boldsymbol{y} \sim MFSN_d(\boldsymbol{\theta})$.

Proposición 2.6.1. *Sea $\boldsymbol{Y} \sim MFSN_d(\boldsymbol{\theta})$, entonces la función generatriz de momento de $\boldsymbol{Y}$ en el caso K=2 es dada por*

$$M(\boldsymbol{t}) = 2\pi_1 \exp\left\{\boldsymbol{t}^\top \boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma}_1 \boldsymbol{t}\right\} \Phi\left(\boldsymbol{\delta}_1^\top \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{t}\right) + 2\pi_2 \exp\left\{\boldsymbol{t}^\top \boldsymbol{\mu}_2 + \frac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma}_2 \boldsymbol{t}\right\} \Phi(\boldsymbol{\delta}_1^\top \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{t}), \quad \boldsymbol{t} \in \mathbb{R}^d, \tag{2.20}$$

donde $\boldsymbol{\delta}_i = \frac{\boldsymbol{\lambda}_i}{\sqrt{1+\boldsymbol{\lambda}_i^\top \boldsymbol{\lambda}_i}}, i = 1, 2.$

Demostración: Por definición,

$$
\begin{aligned}
M(\boldsymbol{t}) &= 2\int\limits_{\mathbb{R}^p} \exp\{\boldsymbol{t}^\top \boldsymbol{y}\} \sum_{k=1}^{2} \pi_k \phi_p(\boldsymbol{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\Phi(\boldsymbol{\lambda}_k \boldsymbol{\Sigma}_k^{-1/2}(\boldsymbol{z}-\boldsymbol{\mu}_k))d\boldsymbol{y}, \\
&= \sum_{k=1}^{2} \pi_k \left[2\int_{\mathbb{R}^p} \exp(\boldsymbol{t}^\top \boldsymbol{y})\phi_p(\boldsymbol{y}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\Phi_1\left\{\boldsymbol{\lambda}_k \boldsymbol{\Sigma}_k^{-1/2}(\boldsymbol{y}-\boldsymbol{\mu}_k)\right\} d\boldsymbol{y}\right], \\
&= \sum_{k=1}^{2} \pi_k M_k(\boldsymbol{t}),
\end{aligned}
$$

donde $M_k(\boldsymbol{t})$ esta dado en (2.18) con $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\delta})$ reemplazado por $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{\delta}_k)$.

■

A partir de (2.20) podemos encontrar la esperanza y la varianza de $\boldsymbol{Y} \sim MFSN(\boldsymbol{\theta})$, para $K = 2$,

$$
\mathbb{E}\{\boldsymbol{Y}\} = \pi_1 \left\{\boldsymbol{\mu}_1 + \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\delta}_1\right\} + \pi_2 \left\{\boldsymbol{\mu}_2 + \sqrt{\frac{2}{\pi}}\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\delta}_2\right\},
$$

y

$$
\mathbb{V}ar\{\boldsymbol{Y}\} = \pi_1 \left\{\boldsymbol{\Sigma}_1 - \frac{2}{\pi}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\delta}_1\boldsymbol{\delta}_1^\top\boldsymbol{\Sigma}_1^{1/2}\right\} + \pi_2 \left\{\boldsymbol{\Sigma}_2 - \frac{2}{\pi}\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\delta}_2\boldsymbol{\delta}_2^T\boldsymbol{\Sigma}_2^{1/2}\right\}.
$$

Capítulo 3

Modelo con Error de Medición Two-Piece Normal

En este capítulo se estudiará el MEV definido en (1.2) y (1.3), en donde suponemos que la covariable ξ o el error de la ecuación ε sigue una distribución TP. A partir de estos supuestos entregamos luego una extension multivariada, hacemos distintos tipos de análisis, tanto de inferencia como estudios numéricos utilizando datos simulados y reales. Cabe destacar que al momento de incorporar algunos de los dos supuestos al MEV obtenemos una mezcla de distribuciones skew multivariadas. En Mclachlan & Peel (2000) encontramos un estudio detallado de este tipo de distribuciones y una descripción de algunas metodologias de estimación.

Los resultados que se presentan en la siguiente sección son basados en la distribución TP normal(TPN) que es obtenida desde (2.3) para $f(x) = \phi(x) = (2\pi)^{-1/2}e^{-\frac{1}{2}x^2}$, $x \in \mathbb{R}$, i.e, es generada por la densidad de la distribución normal. Nos referiremos a esta distribución como TPN(μ, σ, γ), diremos que la variable Y sigue una distribución TPN con parámetro de localización μ, parámetro de escala σ y parámetro de asimetría

γ, cuya densidad esta dada por

$$h(y) = \frac{2}{\sigma(s(\gamma)+t(\gamma))}\left\{\phi\left(\frac{y-\mu}{\sigma s(\gamma)}\right) I_{[\mu,\infty)}(y) + \phi\left(\frac{y-\mu}{\sigma t(\gamma)}\right) I_{(-\infty,\mu)}(y)\right\}, \tag{3.1}$$

$$= \frac{2\pi_s}{\sigma s(\gamma)}\phi\left(\frac{y-\mu}{\sigma s(\gamma)}\right) I_{[\mu,\infty)}(y) + \frac{2\pi_t}{\sigma t(\gamma)}\phi\left(\frac{y-\mu}{\sigma t(\gamma)}\right) I_{(-\infty,\mu)}(y), \quad y \in \mathbb{R}, \tag{3.2}$$

donde la última expresión usa la propiedad P1 de la Sección 2, y se debe pretender que las funciones s y t se hayan fijado de una vez por todas, obviamente en las condiciones establecidas en la Definición 2.1.1.

Para la distribución de TPN, las propiedades mencionadas anteriormente para la familia TP, pueden especificarse aún más. En particular, a partir de (2.8) y los resultados intermedios proporcionados en un apéndice, obtenemos las siguientes expresiones para el valor medio, la varianza y las medidas de asimetría y kurtosis de Pearson, donde $\omega = \sigma t(\gamma)$:

$$\begin{aligned}
\mathbb{E}\{Y\} &= \mu + \sqrt{\frac{2}{\pi}}\big(q-1\big)\,\omega, \\
\mathbb{V}ar\{Y\} &= \left\{\frac{q^3+1}{q+1} - \frac{2}{\pi}\big(q-1\big)^2\right\}\omega^2, \\
\sqrt{\beta_1} &= \frac{\sqrt{\frac{2}{\pi}}\big\{\frac{4}{\pi}\big(q+1\big)\big(q-1\big)^3 - 3\big(q^3+1\big)\big(q-1\big) + 2\big(q^4-1\big)\big\}\big(q+1\big)^{1/2}}{\Big\{q^3+1-\frac{2}{\pi}\big(q-1\big)^2\big(q+1\big)\Big\}^{3/2}}, \\
\beta_2 &= \frac{\big\{-\frac{12}{\pi^2}(q+1)(q-1)^4 + \frac{12}{\pi}(q^3+1)(q-1)^2 - \frac{8}{\pi}(q^4-1)(q-1) + 3(q^5+1)\big\}\big(q+1\big)}{\Big\{q^3+1-\frac{2}{\pi}\big(q-1\big)^2\big(q+1\big)\Big\}^2}
\end{aligned}$$

En este simple ajuste, uno puede verificar de forma algebraica el efecto que ocurre al cambiar las funciones (s,t) por (t,s). Como se indica en P6 de la Sección 2, esta acción implica la modificación de q en $1/q$, y de ω en ωq. La sustitución de estos términos

modificados en las cuatro expresiones basadas en los momentos del párrafo anterior confirma lo siguiente: la diferencia $\mathbb{E}\{Y\} - \mu$ y $\sqrt{\beta_1}$ el signo se invierte, $\mathbb{V}ar\{Y\}$ y β_2 se matienen. Todos estos cambios y los que no, son esperables de la relación de espejo de la densidad con respecto a μ.

Los casos de especial interés se dan a continuación. Los valores del límite cuando $q \to \infty$ se han calculado como se indica en el apéndice.

$q(\gamma)$	$\sqrt{\beta_1}$	β_2
$q = 1$	0	3
$q \to 0$	$\dfrac{\sqrt{2}\left(-\frac{4}{\pi}+1\right)}{\sqrt{\pi}\left(1-\frac{2}{\pi}\right)^{3/2}}$	$\dfrac{-\frac{12}{\pi^2}+\frac{4}{\pi}+3}{\left(1-\frac{2}{\pi}\right)^2}$
$q \to \infty$	$\dfrac{\sqrt{2}\left(\frac{4}{\pi}-1\right)}{\sqrt{\pi}\left(1-\frac{2}{\pi}\right)^{3/2}}$	$\dfrac{-\frac{12}{\pi^2}+\frac{4}{\pi}+3}{\left(1-\frac{2}{\pi}\right)^2}$

Una medida alternativa de la asimetría de una distribución unimodal con función de distribución H es el índice de Arnold-Groeneveld (AG) definido por AG$=1-2H(m_0) \in (-1,1)$ donde m_0 denota la moda (único). Como lo demuestra Fernandez & Steel (1998), para la familia TP en (2.1.1) esta medida se convierte en

$$AG(\gamma) = \frac{q(\gamma)-1}{q(\gamma)+1}.$$

3.1. Caso univariado

Los MEV son generalizaciones de los modelos de regresión clásicos, donde al menos uno de los predictores está contaminado con un error de medición.

Consideremos las ecuaciones (1.1) y (1.2), es decir

$$y = \alpha + \beta\xi + \varepsilon, \tag{3.3}$$

donde ξ no se observa directamente. En su lugar, observamos una variable x, de modo que

$$x = \xi + \delta. \tag{3.4}$$

El modelo de error de medición (3.4) establece que ξ no es observable, y la variable observada x puede ser vista como una estimación no sesgada de ξ. La estructura (3.4) es llamada aditivo ya que se define en términos de errores de medición aditivos, especificada por δ. De esta forma, el sistema conformado por las ecuaciones (3.3) y (3.4) se llama MEV, como detallamos en la capítulo de introducción.
En este trabajo nos centraremos en el modelo estructural, en primer lugar supondremos que

$$\xi \sim TPN(\mu_\xi, \sigma_\xi^2, \gamma). \tag{3.5}$$

Utilizando las regresiones definidas en (3.3) y (3.4), se obtiene el modelo funcional definido por

$$y|\xi \sim N(\alpha + \beta\xi, \sigma_\varepsilon^2), \tag{3.6}$$

$$x|\xi \sim N(\xi, \sigma_\delta^2), \tag{3.7}$$

se tiene que dado ξ, x e y son independientes. Para obtener el modelo estructural TPN propuesto en esta sección, consideremos en (3.6)-(3.7) la hipótesis (3.5) para ξ,

llamaremos a este MEV estructural como TPN-MEV, notar que para $\gamma = 0$, este se reduce al modelo estructuralal normal. Luego la densidad condicional de $(x, y)^\top$ dado ξ puede ser definida como,

$$f(x, y|\xi) = \frac{1}{\sigma_\varepsilon \sigma_\delta} \phi\left(\frac{y - \alpha - \beta\xi}{\sigma_\varepsilon}\right) \phi\left(\frac{x - \xi}{\sigma_\delta}\right).$$

Así, considerando (3.5), se tiene que la densidad marginal conjunta de $(x, y)^\top$ es dada por

$$\begin{aligned} f(x, y) &= \int_{-\infty}^{\infty} f(x, y|\xi) f(\xi) d\xi \\ &= \varphi_s \int_{-\infty}^{\infty} \phi_1\left(\frac{y - \alpha - \beta\xi}{\sigma_\varepsilon}\right) \phi_1\left(\frac{x - \xi}{\sigma_\delta}\right) \phi_1\left(\frac{\xi - \mu_\xi}{s(\gamma)\sigma_\xi}\right) I_{[\mu_\xi, \infty)}(\xi) d\xi + \\ & \quad \varphi_t \int_{-\infty}^{\infty} \phi_1\left(\frac{y - \alpha - \beta\xi}{\sigma_\varepsilon}\right) \phi_1\left(\frac{x - \xi}{\sigma_\delta}\right) \phi_1\left(\frac{\xi - \mu_\xi}{t(\gamma)\sigma_\xi}\right) I_{(-\infty, \mu_\xi)}(\xi) d\xi, \end{aligned}$$

donde $\varphi_r = \dfrac{2r(\gamma)}{\sigma_\delta \sigma_\varepsilon \sigma_\xi (s(\gamma) + t(\gamma))}$, para $r = s, t$.

Para resolver este tipo de integrales, es necesario mencionar el siguiente lema.

Lema 3.1.1. *Sean* $\boldsymbol{\tau} = (t_1, \ldots, t_k)^\top$ *y* $\boldsymbol{\omega} = (w_1, \ldots, w_m)^\top$*, vectores aleatorios tales que*

$$\boldsymbol{\tau}|\boldsymbol{\omega} \sim N_k(\boldsymbol{a} + \boldsymbol{A}\boldsymbol{\omega}, \boldsymbol{\Omega}) \quad y \quad \boldsymbol{\omega} \sim N_m(\boldsymbol{b}, \boldsymbol{B}),$$

y denote por $\phi_k(\boldsymbol{\tau}; \boldsymbol{a} + \boldsymbol{A\Omega})$ *y* $\phi_m(\boldsymbol{\omega}; \boldsymbol{b}, \boldsymbol{B})$ *las respectivas densidades. Entonces,*

$$\phi_k(\boldsymbol{\tau}; \boldsymbol{a} + \boldsymbol{A}\boldsymbol{\omega}, \boldsymbol{\Omega}) \phi_m(\boldsymbol{\omega}; \boldsymbol{b}, \boldsymbol{B}) = \phi_k(\boldsymbol{\tau}; \boldsymbol{a} + \boldsymbol{A}\boldsymbol{b}, \boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^\top) \phi_m(\boldsymbol{\omega}; \boldsymbol{\mu}_{\boldsymbol{\omega}|\boldsymbol{\tau}}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}|\boldsymbol{\tau}}), \quad (3.8)$$

donde

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\omega}|\boldsymbol{\tau}} &= \boldsymbol{b} + (\boldsymbol{A}\boldsymbol{B})^\top (\boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^\top)^{-1} (\boldsymbol{\tau} - \boldsymbol{a} - \boldsymbol{A}\boldsymbol{b}), \\ \boldsymbol{\Sigma}_{\boldsymbol{\omega}|\boldsymbol{\tau}} &= \boldsymbol{B} - (\boldsymbol{A}\boldsymbol{B})^\top (\boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^\top)^{-1} (\boldsymbol{A}\boldsymbol{B}). \end{aligned}$$

Demostración: Usando las propiedades de la esperanza condicional,

$$\mathbb{V}ar\{\boldsymbol{\tau}\} = \boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^\top, \quad \mathbb{E}\{\boldsymbol{\tau}\} = \boldsymbol{a} + \boldsymbol{A}\boldsymbol{b}, \quad \mathrm{Cov}(\boldsymbol{\tau}, \boldsymbol{\omega}) = \boldsymbol{A}\boldsymbol{B}.$$

De aquí se concluye que la distribución conjunta de $\boldsymbol{\tau}$ y $\boldsymbol{\omega}$ esta dada por

$$\begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{\omega} \end{pmatrix} \sim N_{k+m}\left(\begin{bmatrix} \boldsymbol{a} + \boldsymbol{A}\boldsymbol{b} \\ \boldsymbol{b} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^\top & \boldsymbol{A}\boldsymbol{B} \\ (\boldsymbol{A}\boldsymbol{B})^\top & \boldsymbol{B} \end{bmatrix} \right),$$

de donde se concluye que,

$$\boldsymbol{\tau} \sim N_k(\boldsymbol{a} + \boldsymbol{A}\boldsymbol{b}, \boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^\top) \;\; y \;\; \boldsymbol{\omega}|\boldsymbol{\tau} \sim N_m\left(\boldsymbol{\mu}_{\boldsymbol{\omega}|\boldsymbol{\tau}}, \boldsymbol{\Sigma}_{\boldsymbol{\omega}|\boldsymbol{\tau}}\right). \tag{3.9}$$

■

Introduciendo los vectores $\boldsymbol{z} = (x, y)^\top$, $\boldsymbol{\epsilon} = (\delta, \varepsilon)^\top$, $\boldsymbol{a} = (0, \alpha)^\top$, $\boldsymbol{b} = (1, \beta)^\top$, $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_\delta^2, \sigma_\varepsilon^2)$, esto se puede escribir de una forma mas resumida como

$$\boldsymbol{z} = \boldsymbol{a} + \boldsymbol{b}\xi + \boldsymbol{\epsilon}, \tag{3.10}$$

entonces,

$$\phi_2\left(\boldsymbol{z}; \boldsymbol{a} + \boldsymbol{b}\xi, \boldsymbol{\Sigma}\right) = \phi_1\left(y; \alpha + \beta\xi, \sigma_\varepsilon^2\right) \phi_1\left(x; \xi, \sigma_\delta^2\right). \tag{3.11}$$

Teorema 3.1.1. *Bajo el modelo de error de medición TPN-MEV definido en (3.6)-(3.7). Entonces, la variable observada* $\boldsymbol{z} = (x, y)^\top$ *tiene densidad marginal dada por*

$$f(\boldsymbol{z}; \theta) = 2\pi_s \phi_2\left(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}_s\right) \Phi_1\left(\boldsymbol{h}_s^\top(\boldsymbol{z} - \boldsymbol{\mu})\right) + 2\pi_t \phi_2\left(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}_t\right) \Phi_1\left(\boldsymbol{h}_t^\top(\boldsymbol{z} - \boldsymbol{\mu})\right), \tag{3.12}$$

donde $\boldsymbol{\theta} = (\alpha, \beta, \mu_\xi, \sigma_\xi^2, \sigma_\varepsilon^2, \sigma_\delta^2, \gamma)^\top$ *y cada uno de sus componentes son definidos como*

$$\boldsymbol{\mu} = \boldsymbol{a} + \boldsymbol{b}\mu_\xi, \qquad \boldsymbol{b}_s = s(\gamma)\boldsymbol{b}_\xi, \qquad \boldsymbol{b}_t = -t(\gamma)\boldsymbol{b}_\xi, \quad \boldsymbol{b}_\xi = \sigma_\xi \boldsymbol{b}, \tag{3.13}$$

$$\boldsymbol{\Sigma}_r = \boldsymbol{\Sigma} + \boldsymbol{b}_r\boldsymbol{b}_r^\top, \qquad \pi_r = \frac{r(\gamma)}{s(\gamma)+t(\gamma)}, \qquad \boldsymbol{h}_r = (1 - \boldsymbol{b}_r^\top \boldsymbol{\Sigma}_r^{-1}\boldsymbol{b}_r)^{-1/2}\boldsymbol{\Sigma}_r^{-1}\boldsymbol{b}_r \tag{3.14}$$

para $r = s, t$.

Demostración: Usando la representación jerárquica (3.6)-(3.7) y usando el hecho que ξ tiene densidad TPN como en (3.5), tenemos que

$$\begin{aligned}
f(\boldsymbol{z};\boldsymbol{\theta}) &= \int_{-\infty}^{\infty} \phi_2\left(\boldsymbol{z};\boldsymbol{a}+\boldsymbol{b}\xi,\boldsymbol{\Sigma}\right) h(\xi;\mu_\xi,\sigma_\xi,\gamma)d\xi, \\
&= \int_{-\infty}^{\infty} \phi_2\left(\boldsymbol{z};\boldsymbol{a}+\boldsymbol{b}\xi,\boldsymbol{\Sigma}\right) \frac{2}{\sigma_\xi(s(\gamma)+t(\gamma))}\left\{\phi\left(\frac{\xi-\mu_\xi}{\sigma_\xi s(\gamma)}\right) I_{[\mu_\xi,\infty)}(\xi) + \phi\left(\frac{\xi-\mu_\xi}{\sigma_\xi t(\gamma)}\right) I_{(-\infty,\mu_\xi)}(\xi)\right\} d\xi, \\
&= \frac{2}{\sigma_\xi(s(\gamma)+t(\gamma))}\left\{\int_{-\infty}^{\infty} \phi_2\left(\boldsymbol{z};\boldsymbol{a}+\boldsymbol{b}\xi,\boldsymbol{\Sigma}\right)\phi\left(\frac{\xi-\mu_\xi}{\sigma_\xi s(\gamma)}\right) I_{[\mu_\xi,\infty)}(\xi)d\xi + \right. \\
&\qquad \left. \int_{-\infty}^{\infty} \phi_2\left(\boldsymbol{z};\boldsymbol{a}+\boldsymbol{b}\xi,\boldsymbol{\Sigma}\right)\phi\left(\frac{\xi-\mu_\xi}{\sigma_\xi t(\gamma)}\right) I_{(-\infty,\mu_\xi)}(\xi)d\xi\right\}, \\
&= \frac{2}{\sigma_\xi(s(\gamma)+t(\gamma))}\left\{\int_{\mu_\xi}^{\infty} \phi_2\left(\boldsymbol{z};\boldsymbol{a}+\boldsymbol{b}\xi,\boldsymbol{\Sigma}\right)\phi\left(\frac{\xi-\mu_\xi}{\sigma_\xi s(\gamma)}\right) d\xi + \int_{-\infty}^{\mu_\xi} \phi_2\left(\boldsymbol{z};\boldsymbol{a}+\boldsymbol{b}\xi,\boldsymbol{\Sigma}\right)\phi\left(\frac{\xi-\mu_\xi}{\sigma_\xi t(\gamma)}\right) d\xi\right\}.
\end{aligned}$$

Haciendo cambio de variables $\xi_s = (\xi-\mu_\xi)/\sigma_\xi s(\gamma)$ y $\xi_t = (\xi-\mu_\xi)/\sigma_\xi t(\gamma)$ y notando que $\phi(\xi) = \phi(\xi;0,1)$ concluimos que

$$\begin{aligned}
f(\boldsymbol{z};\boldsymbol{\theta},\gamma) &= \frac{2}{s(\gamma)+t(\gamma)}\left\{s(\gamma)\int_0^{\infty} \phi_2\left(\boldsymbol{z};\boldsymbol{\mu}+\boldsymbol{b}_s\xi_s,\boldsymbol{\Sigma}\right)\phi(\xi_s;0,1)d\xi_s + \right. \\
&\qquad \left. t(\gamma)\int_0^{\infty} \phi_2\left(\boldsymbol{z};\boldsymbol{\mu}+\boldsymbol{b}_t\xi_t,\boldsymbol{\Sigma}\right)\phi(\xi_t;0,1)d\xi_t\right\}, \\
&= 2\pi_s\phi_2\left(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma}_s\right)\int_0^{\infty}\phi\left(\xi_s;\boldsymbol{b}_s^\top\Sigma_s^{-1}(\boldsymbol{z}-\boldsymbol{\mu}),1-\boldsymbol{b}_s^\top\boldsymbol{\Sigma}_s^{-1}\boldsymbol{b}_s\right)d\xi_s + \\
&\quad 2\pi_t\phi_2\left(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma}_t\right)\int_0^{\infty}\phi\left(\xi_t;\boldsymbol{b}_t^\top\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}-\boldsymbol{\mu}),1-\boldsymbol{b}_t^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{b}_t\right)d\xi_t, \\
&= 2\pi_s\phi_2\left(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma}_s\right)\Phi_1\left(\frac{\boldsymbol{b}_s^\top\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{z}-\boldsymbol{\mu})}{\sqrt{1-\boldsymbol{b}_s^\top\boldsymbol{\Sigma}_s^{-1}\boldsymbol{b}_s}}\right) + 2\pi_t\phi_2\left(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma}_t\right)\Phi_1\left(\frac{\boldsymbol{b}_t^\top\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{z}-\boldsymbol{\mu})}{\sqrt{1-\boldsymbol{b}_t^\top\boldsymbol{\Sigma}_t^{-1}\boldsymbol{b}_t}}\right),
\end{aligned}$$

donde usamos la representación-marginal de la densidad normal multivariada dada en (3.8). Vale la pena señalar que $1 - \boldsymbol{b}_r^\top \boldsymbol{\Sigma}_r^{-1} \boldsymbol{b}_r > 0$ para $r = s, t$. Para ver esto, se utilizó la conocida fórmula de Sherman-Morrison.

$$\boldsymbol{\Sigma}_r^{-1} = \left(\boldsymbol{\Sigma} + \boldsymbol{b}_r \boldsymbol{b}_r^\top\right) = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{b}_r \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1}}{1 + \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{b}_r}, \tag{3.15}$$

para obtener fácilmente $1 - \boldsymbol{b}_r^\top \boldsymbol{\Sigma}_r^{-1} \boldsymbol{b}_r = (1 + \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{b}_r)^{-1} > 0$, para $r = s, t$.

■

Una caraterística interesante del modelo estructural TPN-MEV con densidad en (3.12) para los datos observados $\boldsymbol{z}$, es que representa una mezcla de dos componentes de densidades SN dada en (2.13). Especificamente del Teorema 3.1.1 tenemos que $\boldsymbol{z}$ tiene una distritución dada por

$$\pi_s SN_2(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\lambda}_s) + \pi_t SN_2(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t, \boldsymbol{\lambda}_t), \tag{3.16}$$

donde $\boldsymbol{\mu}_r = \boldsymbol{\mu}$ y

$$\boldsymbol{\lambda}_r = \frac{\boldsymbol{\Sigma}_r^{-1/2} \boldsymbol{b}_r}{\sqrt{1 - \boldsymbol{b}_r^\top \boldsymbol{\Sigma}_r^{-1} \boldsymbol{b}_r}}, \quad r = s, t.$$

Por lo tanto, usando (3.15), obtenemos

$$\boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1} = \left(1 + \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{b}_r^\top\right)^{-1} \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1}, \qquad 1 - \boldsymbol{b}_r^\top \boldsymbol{\Sigma}_r^{-1} \boldsymbol{b}_r = \left(1 + \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{b}_r\right)^{-1}, \tag{3.17}$$

así el termino $\boldsymbol{h}_r$ en (3.14) se puede reescribir como

$$\boldsymbol{h}_r = \left(1 + \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{b}_r\right)^{-1/2} \boldsymbol{\Sigma}^{-1} \boldsymbol{b}_r$$

así (3.12), quedaría representada de la siguiente manera

$$f(\boldsymbol{z}; \boldsymbol{\theta}) = 2\pi_s \phi_2\left(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{b}_s \boldsymbol{b}_s^\top\right) \Phi\left(\frac{\boldsymbol{b}_s \boldsymbol{\Sigma}^{-1}(\boldsymbol{z} - \boldsymbol{\mu})}{\sqrt{1 + \boldsymbol{b}_t^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{b}_t}}\right) + 2\pi_t \phi_2\left(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{b}_t \boldsymbol{b}_t^\top\right) \Phi\left(\frac{\boldsymbol{b}_t \boldsymbol{\Sigma}^{-1}(\boldsymbol{z} - \boldsymbol{\mu})}{\sqrt{1 + \boldsymbol{b}_t^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{b}_t}}\right). \tag{3.18}$$

3.2. Extensión multivariada

En esta sección se estudiará una extensión del MEV, en donde disponemos de varias respuestas y un predictor medido con error.

Consideremos el modelo de regresión multivariado con error de medición aditivo definido por las siguientes ecuaciones

$$\boldsymbol{y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta}\xi_i + \boldsymbol{e}_i, \tag{3.19}$$

$$x_i = \xi_i + \delta_i, \tag{3.20}$$

$i = 1, \ldots, n$, donde $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{id})^\top$ es un vector de respuestas para la unidad experimental i, ξ_i es la covariable no observada (verdadera) en unidad i, $\boldsymbol{e}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{id})^\top$ vector aleatorio en la ecuación y el error de medición de la respuesta, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top$ y $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top$ son vectores de parámetros desconocidos. Aquí, para cada unidad de muestra $i = 1, \ldots, n$, la ecuación (3.20) especifica un modelo de regresión multivariado, en donde el vector de respuestas $\boldsymbol{y}_i$ esta relacionado a una covariable no observada ξ_i. Por ejemplo, las observaciones y_{ij}, $j = 1, \ldots, d$ provienen de d instrumentos que miden la caracaterística ξ_i asociada a la i-ésima unidad muestral. Como se indica en la ecuación (3.19), x_i es una medida sin sesgo obtenida con error de ξ_i, donde el error de medición es representado por δ_i. Dicha medición se puede interpretar como la realizada por un intrumento padrón o de referencia. Considerando los vectores $\boldsymbol{z}_i = (x_i, \boldsymbol{y}_i^\top)^\top$, $\boldsymbol{\epsilon}_i = (\delta_i, \boldsymbol{e}_i^\top)$, $\boldsymbol{a} = (0, \boldsymbol{\alpha}^\top)^\top$ y $\boldsymbol{b} = (1, \boldsymbol{\beta}^\top)^\top$, una representación mas compacta del modelo definido en (3.19)-(3.20) es

$$\boldsymbol{z}_i = \boldsymbol{a} + \boldsymbol{b}\xi_i + \boldsymbol{\epsilon}_i, \tag{3.21}$$

$i = 1, \ldots, n$. En este trabajo, se estudia el enfoque estructural, como fue mencionado anteriormente, donde las covariables latentes $\xi_1, \ldots, \xi_n$ son consideradas variables aleatorias.

Además, podemos suponer la independencia entre ξ_i y $\boldsymbol{\epsilon}_i$ para cada $i = 1, \ldots, n$. Por lo tanto, de acuerdo con (3.21) la distribución de $\boldsymbol{z}_i$ se define una vez que se especifica la distribución de ξ_i y $\boldsymbol{\epsilon}_i$. Especificamente, consideremos el MEV estructural dado por (3.21) con las siguientes suposiciones

$$\begin{aligned} \xi_i &\sim TPN(\mu_\xi, \sigma_\xi, \gamma), \\ \boldsymbol{\epsilon}_i &\sim N_{d+1}(0, \boldsymbol{\Sigma}), \end{aligned}$$

$i = 1, \ldots, n$, y donde

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\delta^2 & 0 \\ 0 & \boldsymbol{\Sigma}_e \end{pmatrix}. \tag{3.22}$$

Aquí $\boldsymbol{\Sigma}_e$ es una matriz de covarianza $d \times d$ arbitraria, y asumimos independencia entre todas las variables $(\xi_1, \boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n, \xi_n)$. Nos referiremos a este MEV multivariado como TPN-MEVM. Este modelo puede especificarse de forma equivalente usando la representación jerárquica siguiente,

$$\boldsymbol{z}_i \mid \xi_i \sim N_{d+1}(\boldsymbol{a} + \boldsymbol{b}\xi_i, \boldsymbol{\Sigma}), \tag{3.23}$$

$$\xi_i \sim TPN(\mu_\xi, \sigma_\xi, \gamma). \tag{3.24}$$

El siguiente teorema representa el principal resultado de esta sección. Donde $\boldsymbol{\theta}$ es el vector de parámetros formado por diferentes componentes $(\alpha, \beta, \mu_\xi, \sigma_\xi^2, \sigma_\delta^2, \Sigma_e, \gamma)$.

Teorema 3.2.1. *Sea el modelo TPN-MEVM estructural definido por (3.23)-(3.24) y el supuesto de independencia entre los vectores de n componentes, los vectores aleatorios observados $\boldsymbol{z}_i$ son independientes e idénticamente distribuidos (i.i.d.) con densidad dada por*

$$f(\boldsymbol{z}_i; \boldsymbol{\theta}) = 2\pi_s \phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}_s)\Phi_1\left(\boldsymbol{h}_s^\top(\boldsymbol{z}_i - \boldsymbol{\mu})\right) + 2\pi_t \phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}_t)\Phi_1\left(\boldsymbol{h}_t^\top(\boldsymbol{z}_i - \boldsymbol{\mu})\right), \tag{3.25}$$

$i = 1, \ldots, n$, donde sus componentes están dados por

$$\boldsymbol{\mu} = \boldsymbol{a} + \boldsymbol{b}\mu_\xi, \qquad \boldsymbol{b}_s = s(\gamma)\boldsymbol{b}_\xi, \qquad \boldsymbol{b}_t = -t(\gamma)\boldsymbol{b}_\xi, \quad \boldsymbol{b}_\xi = \sigma_\xi \boldsymbol{b}, \tag{3.26}$$

$$\boldsymbol{\Sigma}_r = \boldsymbol{\Sigma} + \boldsymbol{b}_r\boldsymbol{b}_r^\top, \qquad \pi_r = \frac{r(\gamma)}{s(\gamma) + t(\gamma)}, \qquad \boldsymbol{h}_r = (1 - \boldsymbol{b}_r^\top \boldsymbol{\Sigma}_r^{-1}\boldsymbol{b}_r)^{-1/2}\boldsymbol{\Sigma}_r^{-1}\boldsymbol{b}_r. \tag{3.27}$$

para $r = s, t$.

Demostración: Claramente los $\boldsymbol{z}_i$ son vectores aleatorios iid. Por lo tanto es suficiente mostrar solo para la unidad i. Luego, la demostración es análoga a la desarrollada en el Teorema 3.1.1.

■

En este caso que la densidad del TPN-MEVM estructural en (3.25) para los datos observados $\boldsymbol{z}_i$ pertenece a la clase de mezcla de densidades SN multivariada donde sus componentes estan definidos en (2.13). Especificamente, desde el Teorema 3.2.1 tenemos que $\boldsymbol{z}_i, \ldots, \boldsymbol{z}_n$ son vectores aleatorios iid con distribución dada por

$$\pi_s SN_{1+d}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\lambda}_s) + \pi_t SN_{1+d}(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t, \boldsymbol{\lambda}_t), \tag{3.28}$$

donde $\boldsymbol{\mu}_r = \boldsymbol{\mu}$ y

$$\boldsymbol{\lambda}_r = \frac{\boldsymbol{\Sigma}_r^{-1/2}\boldsymbol{b}_r}{\sqrt{1 - \boldsymbol{b}_r^\top \boldsymbol{\Sigma}_r^{-1}\boldsymbol{b}_r}}, \quad r = s, t.$$

Por lo tanto, usando (3.15), obtenemos el caso multivarido, es decir

$$\boldsymbol{b}_r^\top \boldsymbol{\Sigma}_r^{-1} = (1 + \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b}_r)^{-1}\boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1}, \qquad 1 - \boldsymbol{b}_r^\top \boldsymbol{\Sigma}_r^{-1}\boldsymbol{b}_r = (1 + \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b}_r)^{-1},$$

así, el término $\boldsymbol{h}_r$ en 3.27 puede escribirse como

$$\boldsymbol{h}_r = \left(1 + \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b}_r\right)^{-1/2} \boldsymbol{\Sigma}^{-1}\boldsymbol{b}_r\,.$$

Por otro lado, si $b_r = wb_\xi$ para $r = |w|$ y $w \in \{-t(\gamma), s(\gamma)\}$, que es una v.a. con función de probabilidad $p(w;\gamma) = \pi_s I_{\{w=s(\gamma)\}} + \pi_t I_{\{w=-t(\gamma)\}}$, entonces tenemos la siguiente formulación la que es equivalente a nuestro modelo:

Proposición 3.2.1. *El modelo estructural TPN-MEVM definido por (3.23)-(3.24) tiene una representación jerárquica de la siguiente manera:*

$$\boldsymbol{z}_i|v_i, w_i \overset{iid}{\sim} N_{1+d}(\boldsymbol{\mu} + \boldsymbol{b}v_i w_i, \boldsymbol{\Sigma}), \tag{3.29}$$

$$v_i|w_i \overset{iid}{\sim} HN_1(0, \sigma_\xi^2), \tag{3.30}$$

$$w_i \overset{iid}{\sim} p(w_i;\gamma) = \pi_s I_{\{w_i=s(\gamma)\}} + \pi_t I_{\{w_t=-t(\gamma)\}}, \tag{3.31}$$

donde v_i y w_i son variables independientes, para $i = 1, \ldots, n$.

Demostración: De acuerdo a la propiedad P2 de la distribución TP, combinado con el supuesto (3.24), podemos decir que existen variables aleatorias auxiliares $u_1, \ldots, u_n$ y $w_1, \ldots, w_n$ que son mutuamente independientes, de tal manera que $\xi_i = \mu_\xi + \sigma_\xi u_i w_i$ con $u_i \overset{iid}{\sim} HN_1(0,1)$ y $w_i \overset{iid}{\sim} p(w_i;\gamma) = \pi_s I_{\{w_i=s(\gamma)\}} + \pi_t I_{\{w_t=-t(\gamma)\}}$, para $i = 1, \ldots, n$. Por suposición $\xi_1, \ldots \xi_n$ son mutuamente independientes de $\epsilon_1, \ldots, \epsilon_n$, por lo tanto $(u_1, w_1), \ldots, (u_n, w_n)$ son también mutuamente independientes de $\epsilon_1, \ldots, \epsilon_n$. Usando esto en (3.21), podemos escribir,

$$\boldsymbol{z}_i = \boldsymbol{a} + \boldsymbol{b}(\mu_\xi + \sigma_\xi w_i u_i) + \epsilon_i = \boldsymbol{\mu} + \boldsymbol{b}w_i v_i + \epsilon_i,$$

con $\boldsymbol{\mu} = \boldsymbol{a} + \boldsymbol{b}\mu_\xi$ y $v_i = \sigma_\xi u_i$. Dado esto, la prueba sigue al condicionar el vector de observación $\boldsymbol{z}_i$ en las variables auxiliares (v_i, w_i).

■

Como primera consecuencia de la Proposicion 3.2.1, calculamos los dos primeros momentos de las variables observadas $\boldsymbol{z}_i$. Donde, usando que

$$\mathbb{E}\{v_i w_i\} = \sigma_\xi \mathbb{E}\{u_i\}\mathbb{E}\{w_i\} = \sqrt{\frac{2}{\pi}}\sigma_\xi\{s(\gamma) - t(\gamma)\}$$
$$\mathbb{E}\{v_i^2 w_i^2\} = \sigma_\xi^2 \mathbb{E}\{u_i^2\}\mathbb{E}\{w_i^2\} = \sigma_\xi^2 \frac{s(\gamma)^3 + t(\gamma)^3}{s(\gamma) + t(\gamma)},$$

obtenemos

$$\begin{aligned}
\mathbb{E}\{\boldsymbol{z}_i\} &= \mathbb{E}\{\mathbb{E}\{\boldsymbol{z}_i \mid v_i, w_i\}\} \\
&= \boldsymbol{a} + \boldsymbol{b}\boldsymbol{\mu}_\xi + \sqrt{\frac{2}{\pi}}\left(s(\gamma) - t(\gamma)\right)\sigma_\xi \boldsymbol{b}, \\
\mathbb{V}ar\{\boldsymbol{z}_i\} &= \mathbb{E}\{\mathbb{V}ar\{\boldsymbol{z}_i \mid v_i, w_i\}\} + \mathbb{V}ar\{\mathbb{E}\{\boldsymbol{z}_i \mid v_i, w_i\}\} \\
&= \boldsymbol{\Sigma} + \left\{\frac{s(\gamma)^3 + t(\gamma)^3}{s(\gamma) + t(\gamma)} - \frac{2}{\pi}\left(s(\gamma) - t(\gamma)\right)^2\right\}\sigma_\xi^2 \boldsymbol{b}\boldsymbol{b}^\top.
\end{aligned}$$

La representación jerárquica (3.29) a (3.31) será muy útil en el análisis estadístico del modelo TPN-MEV. Específicamente, veremos en la siguiente sección que facilita considerablemente la implementación de un algoritmo de tipo EM para la busqueda del EMV. Para este fín, necesitamos desarrollar los resultados que se presentan a continuación en la Proposición 3.2.2. Para simplificar la notación, definimos $\boldsymbol{\mu}_{w_i} = \sigma_\xi \boldsymbol{b}_{w_i}^\top \boldsymbol{\Sigma}_{w_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu})$ y $\sigma_{w_i}^2 = \sigma_\xi^2(1 - \boldsymbol{b}_{w_i}^\top \boldsymbol{\Sigma}_{w_i}^{-1}\boldsymbol{b}_{w_i})$, donde $\boldsymbol{b}_{w_i} = w_i \boldsymbol{b}_\xi$ y $\boldsymbol{\Sigma}_{w_i} = \boldsymbol{\Sigma} + \boldsymbol{b}_{w_i}\boldsymbol{b}_{w_i}^\top$. Notar que por (3.15)

$$\boldsymbol{\mu}_{w_i} = \frac{\sigma_\xi \boldsymbol{b}_{w_i}^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu})}{1 + \boldsymbol{b}_{w_i}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b}_{w_i}}, \ \sigma_{w_i}^2 = \frac{\sigma_\xi^2}{1 + \boldsymbol{b}_{w_i}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b}_{w_i}}, \tag{3.32}$$

para $i = 1, \ldots, n$, dando lugar a

$$\tau_{w_i} = \frac{\boldsymbol{\mu}_{w_i}}{\sigma_{w_i}} = \frac{\boldsymbol{b}_{w_i}^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu})}{\sqrt{1 + \boldsymbol{b}_{w_i}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b}_{w_i}}}. \tag{3.33}$$

Proposición 3.2.2. *Bajo el modelo jerárquico 3.29 a 3.31, obtenemos*

$$p(w_i \mid \boldsymbol{z}_i; \boldsymbol{\theta}) = \pi_{si} I_{\{w_i = s(\gamma)\}} + \pi_{ti} I_{\{w_i = -t(\gamma)\}}, \tag{3.34}$$

$$f(v_i \mid w_i, \boldsymbol{z}_i; \boldsymbol{\theta}) = \phi_1\left(v_i; \mu_{w_i}, \sigma_{w_i}^2\right) I_{\{v_i > 0\}} / \Phi_1\left(\tau_{w_i}\right), \tag{3.35}$$

$i = 1, \ldots, n$, donde μ_{w_i} y $\sigma^2_{w_i}$ estan definidas en (3.32), y

$$\begin{aligned}
\pi_{ri} = \pi_r(\boldsymbol{z}_i) &= \frac{\pi_r \phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{b}_r \boldsymbol{b}_r^\top) \Phi_1(\tau_{r_i})}{\pi_s \phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{b}_s \boldsymbol{b}_s^\top) \Phi(\tau_{s_i}) + \pi_t \phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{b}_t \boldsymbol{b}_t^\top) \Phi(\tau_{t_i})}, \\
\tau_{ri} &= \frac{\boldsymbol{\mu}_{ri}}{\sigma_r}, \quad \boldsymbol{\mu}_{ri} = \sigma_\xi \boldsymbol{h}_r^\top (\boldsymbol{z}_i - \boldsymbol{\mu}), \quad \sigma_r^2 = \frac{\sigma_\xi^2}{1 + \boldsymbol{b}_r^\top \boldsymbol{\Sigma}^{-1} b_r}.
\end{aligned}$$

para $r = s, t$ y $i = 1, \ldots, n$. Además

$$\mathbb{E}\{s_i \mid \boldsymbol{z}_i; \boldsymbol{\theta}\} = \pi_{si} - \pi_{ti} = 2\pi_{si} - 1, \tag{3.36}$$

$$\mathbb{E}\{v_i \mid \boldsymbol{z}_i; \boldsymbol{\theta}\} = \{\mu_{si} + \sigma_s \zeta_1(\tau_{si})\} \pi_{si} + \{\mu_{ti} + \sigma_t \zeta_1(\tau_{ti})\} \pi_{ti}, \tag{3.37}$$

$$\mathbb{E}\{v_i^2 \mid \boldsymbol{z}_i; \boldsymbol{\theta}\} = \left\{\mu_{si}^2 + \sigma_s^2 + \mu_{si}\sigma_s \zeta_1(\tau_{si})\right\} \pi_{si} + \left\{\mu_{ti}^2 + \sigma_t^2 + \mu_{ti}\sigma_t \zeta_1(\tau_{ti})\right\} \pi_{ti}, \tag{3.38}$$

$$\mathbb{E}\{v_i w_i \mid \boldsymbol{z}_i; \boldsymbol{\theta}\} = s(\gamma) \{\mu_{si} + \sigma_s \zeta_1(\tau_{si})\} \pi_{si} - t(\gamma) \{\mu_{ti} + \sigma_t \zeta_1(\tau_{ti})\} \pi_{ti}, \tag{3.39}$$

$$\mathbb{E}\{v_i^2 w_i^2 \mid \boldsymbol{z}_i; \boldsymbol{\theta}\} = s(\gamma)^2 \left\{\mu_{si}^2 + \sigma_s^2 + \mu_{si}\sigma_s \zeta_1(\tau_{si})\right\} \pi_{si} + t(\gamma)^2 \left\{\mu_{ti}^2 + \sigma_t^2 + \mu_{ti}\sigma_t \zeta_1(\tau_{ti})\right\} \pi_{ti}, \tag{3.40}$$

$i = 1, \ldots, n$, donde s_i=sign(w_i)$\in \{-1, 1\}$ y $\zeta_1(u) = \phi_1(u)/\Phi(u)$.

Demostración: Como para cada $i = 1, \ldots, n$, v_i y w_i son independientes, entonces por (3.29) y (3.30) y las propiedades de la distribución normal multivariada tenemos

$$\begin{aligned}
f(v_i \mid w_i, \boldsymbol{z}_i, \boldsymbol{\theta}) &= \frac{f(\boldsymbol{z}_i \mid v_i, w_i; \boldsymbol{\theta}) f(v_i; \boldsymbol{\theta})}{f(\boldsymbol{z}_i \mid w_i; \boldsymbol{\theta})} \\
&= \frac{\phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu} + v_i w_i \boldsymbol{b}, \boldsymbol{\Sigma}) 2\phi_1(v_i; 0, \sigma_\xi^2) I_{\mathbb{R}^+}(v_i)}{f(\boldsymbol{z}_i \mid w_i; \boldsymbol{\theta})} \\
&= \frac{2\phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma} + w_i \boldsymbol{b}_\xi \boldsymbol{b}_\xi^\top) \phi_1(v_i; \mu_{w_i}, \sigma_{w_i}^2) I_{\mathbb{R}^+}(v_i)}{f(\boldsymbol{z}_i \mid w_i; \boldsymbol{\theta})},
\end{aligned} \tag{3.41}$$

y, haciendo cambio de variable $u_i = (v_i - \boldsymbol{\mu}_{\omega_i})/\sigma_{w_i}$,

$$\begin{aligned}
f(\boldsymbol{z}_i \mid w_i; \boldsymbol{\theta}) &= \int_0^\infty 2\phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma} + w_i \boldsymbol{b}_\xi \boldsymbol{b}_\xi^\top) \phi_1(v_i; \mu_{w_i}, \sigma_{w_i}^2) dv_i \\
&= 2\phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma} + w_i^2 \boldsymbol{b}_\xi \boldsymbol{b}_\xi^\top) \int_{-\tau_{w_i}}^\infty \phi_1(u_i; 0, 1) du_i \\
&= 2\phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma} + w_i^2 \boldsymbol{b}_\xi \boldsymbol{b}_\xi^\top) \Phi(\tau_{w_i}) \\
&= 2\phi_{1+d}(\boldsymbol{z}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma} + w_i^2 \boldsymbol{b}_\xi \boldsymbol{b}_\xi^\top) \Phi\left(\frac{w_i \boldsymbol{b}_\xi^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_i - \boldsymbol{\mu})}{\sqrt{1 + w_i^2 \boldsymbol{b}_\xi^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{b}_\xi}}\right).
\end{aligned} \tag{3.42}$$

Notar que de (3.42), (3.25) se puede reescribir como

$$f(\boldsymbol{z}_i;\boldsymbol{\theta}) \quad = \quad \pi_s f(\boldsymbol{z}_i \mid w_i = s(\gamma)) + \pi_t f(\boldsymbol{z}_i \mid w_i = -t(\gamma)). \tag{3.43}$$

Por lo tanto, la función de probabilidad condicional en (3.34) se obtiene usando la fórmula de Bayes, (3.42) y (3.43), mientras que la densidad condicional en (3.35) se obtiene aplicando (3.42) en (3.41). Para (3.39) y (3.40), tenemos por (3.35) y los dos primeros momentos de la normal truncada (ver por ejemplo, Arellano-Valle et al. (2005b)) que

$$\begin{aligned}
\mathbb{E}\{s_i|\boldsymbol{z}_i;\boldsymbol{\theta}\} &= 2\pi_{si} - 1, \\
\mathbb{E}\{v_i \mid \boldsymbol{z}_i;\boldsymbol{\theta}\} &= \mathbb{E}\{\mathbb{E}\{v_i \mid w_i, \boldsymbol{z}_i;\boldsymbol{\theta}\}\} \\
&= \mathbb{E}\{[\mu_{w_i} + \sigma_{w_i}\zeta_1(\tau_{w_i})] \mid \boldsymbol{z}_i;\boldsymbol{\theta}\}, \\
\mathbb{E}\{v_i^2 \mid \boldsymbol{z}_i;\boldsymbol{\theta}\} &= \mathbb{E}\{\mathbb{E}\{v_i^2 \mid w_i, \boldsymbol{z}_i;\boldsymbol{\theta}\}\} \\
&= \mathbb{E}\{[\mu_{w_i}^2 + \sigma_{w_i}^2 + \mu_{w_i}\sigma_{w_i}\zeta_1(\tau_{w_i})] \mid \boldsymbol{z}_i;\boldsymbol{\theta}\}, \\
\mathbb{E}\{v_i w_i \mid \boldsymbol{z}_i;\boldsymbol{\theta}\} &= \mathbb{E}\{w_i\mathbb{E}\{v_i \mid w_i, \boldsymbol{z}_i;\boldsymbol{\theta}\}\} \\
&= \mathbb{E}\{w_i[\mu_{w_i} + \sigma_{w_i}\zeta_1(\tau_{w_i})] \mid \boldsymbol{z}_{w_i};\boldsymbol{\theta}\}, \\
\mathbb{E}\{v_i^2 w_i^2 \mid \boldsymbol{z}_i;\boldsymbol{\theta}\} &= \mathbb{E}\{w_i^2\mathbb{E}\{v_i^2 \mid w_i, z_i;\boldsymbol{\theta}\}\} \\
&= \mathbb{E}\{w_i^2[\mu_{w_i}^2 + \sigma_{w_i}^2 + \mu_{w_i}\sigma_{w_i}\zeta_1(\tau_{w_i})] \mid \boldsymbol{z}_i;\boldsymbol{\theta}\}.
\end{aligned}$$

Finalmente, (3.39) y (3.40) se obtienen usando (3.34).

■

En muchas situaciones también es interesante estimar la variable explicativa no observada ξ_i, para $i = 1, \ldots, n$. Para ello, es habitual considerar la media a posteriori

$$\mathbb{E}\{\xi_i \mid \boldsymbol{z}_i;\boldsymbol{\theta}\} = \mathbb{E}\{(\mu_\xi + v_i w_i) \mid \boldsymbol{z}_i;\boldsymbol{\theta}\} = \mu_\xi + \mathbb{E}\{v_i w_i \mid \boldsymbol{z}_i;\boldsymbol{\theta}\},$$

para $i = 1, \ldots, n$. Por lo tanto, si $\widehat{\boldsymbol{\theta}}$ denota el EMV de $\boldsymbol{\theta}$, entonces ξ_i, puede ser estimado por $\widehat{\xi}_i = \widehat{\mu}_\xi + \widehat{\sigma}_\xi(\widehat{vw})_i$, donde $(\widehat{vw})_i = \mathbb{E}\{v_i w_i \mid \boldsymbol{z}_i;\widehat{\boldsymbol{\theta}}\}$.

Capítulo 4

Máxima Verosimilitud

En este capítulo se desarrollan aspectos de inferencia para el modelo propuesto TPN-MEVM definido en (3.25) asumiendo $d > 1$. Dentro de la literatura se pueden encontrar varios trabajos en donde se aborda el problema de estimación de mezclas finitas, uno de ellos es el realizado por Tzy & Tsung (2010).

4.1. Estimación de máxima verosimilitud

Es bien sabido que la función log-verosimilitud asociada a una densidad de mezcla finita es generalmente bastante engorrosa para maximizar cuando se busca EMV. El enfoque empleado habitualmente en este contexto se basa en algunos de los algoritmos de la familia EM, ya que su esquema subyacente coincide naturalmente con la estructura de mezclas finitas y, además tienden a ser numéricamente estables. Específicamente, para estimar los parámetros del modelo (3.25), utilizaremos la técnica de ECM estudiada por Meng & Rubin (1993).
Para llevar a cabo este programa, comenzamos considerando la representación jerárquica (3.29)-(3.31) y el hecho de que

$$p(w_i;\gamma) = \pi_s I_{\{w_i = s(\gamma)\}} + \pi_t I_{\{w_i = -t(\gamma)\}} = \pi_s^{(1+s_i)/2}\pi_t^{(1-s_i)/2} I_{\{s_i=\pm 1\}},$$

donde $s_i = \text{sign}(w_i)$. Ya que $f(\boldsymbol{z}_i, v_i, w_i; \theta) = f(\boldsymbol{z}_i \mid v_i, w_i; \theta) f(v_i \mid w_i; \theta) p(w_i; \gamma)$, la log-verosimilitud de $\boldsymbol{\theta}$ basada en los datos completos $c_i = (\boldsymbol{z}_i, v_i, w_i)$, es $\ell_c(\boldsymbol{\theta}; c_1, \ldots, c_n) = \sum_{i=1}^n \ell_{ci}(\boldsymbol{\theta}; c_i)$, es entonces

$$\ell_c(\boldsymbol{\theta}; c_1, \ldots, c_n) = \sum_{i=1}^n \ell_{ci}(\boldsymbol{\theta}; c_i),$$

donde

$$\begin{aligned}
\ell_{c_i}(\boldsymbol{\theta}; c_i) &= \frac{1}{2}(1+s_i)\log\pi_s + \frac{1}{2}(1-s_i)\log\pi_t + \log\{2\phi_1(v_i; 0, \sigma_\xi^2)\} + \log\phi_{d+1}(\boldsymbol{z}_i; \boldsymbol{\mu} + v_i w_i \boldsymbol{b}, \boldsymbol{\Sigma}) \\
&= \frac{1}{2}(1+s_i)\log\pi_s + \frac{1}{2}(1-s_i)\log\pi_t + \log 2 - \frac{1}{2}\log(2\pi\sigma_\xi^2) - \frac{1}{2\sigma_\xi^2}v_i^2 - \frac{1}{2}\log|2\pi\boldsymbol{\Sigma}| \\
&\quad -\frac{1}{2}(\boldsymbol{z}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}) + v_i w_i (\boldsymbol{z}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b} - \frac{1}{2}(v_i w_i)^2 \boldsymbol{b}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b} \\
&= \frac{1}{2}(1+s_i)\log\pi_s + \frac{1}{2}(1-s_i)\log\pi_t + \log 2 - \frac{1}{2}\log(2\pi\sigma_\xi^2) - \frac{1}{2\sigma_\xi^2}v_i^2 - \frac{1}{2}\log(2\pi\sigma_\delta^2) \\
&\quad -\frac{1}{2\sigma_\delta^2}(x_i - \mu_x)^2 + \frac{1}{\sigma_\delta^2}v_i w_i (x_i - \mu_x) - \frac{1}{2\sigma_\delta^2}(v_i w_i)^2 - \frac{1}{2}\log \mid 2\pi\boldsymbol{\Sigma}_e \mid \\
&\quad -\frac{1}{2}(y_i - \mu_y)^\top \boldsymbol{\Sigma}_e^{-1}(y_i - \mu_y) + v_i w_i (y_i - \mu_y)^\top \boldsymbol{\Sigma}_e^{-1}\boldsymbol{\beta} - \frac{1}{2}(v_i w_i)^2 \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_e^{-1}\boldsymbol{\beta},
\end{aligned}$$

teniendo en cuenta las expresiones (2.6), (3.22), $\boldsymbol{\mu} = (\mu_x, \mu_y^\top)^\top = (\mu_\xi, \boldsymbol{\alpha}^\top + \boldsymbol{\beta}^\top \mu_\xi)^\top$, y $\boldsymbol{b} = (1, \boldsymbol{\beta}^\top)^\top$.

Por razones operativas que quedarán claras en un su momento, el componente final de $\boldsymbol{\theta}$ se reemplaza por π_s para el desarrollo del algoritmo ECM. Donde γ y π_s están relacionados, la modificación es solo superficial. Por este motivo y para evitar la introducción de un nuevo símbolo, conservamos el símbolo $\boldsymbol{\theta}$ después de esta modificación. El algoritmo ECM propuesto procede mediante la iteración de los siguientes pasos, donde denotamos el valor del vector de parámetros en la k-ésima iteración del algoritmo mediante $\boldsymbol{\theta}^{(k)}$, con componentes $\left(\alpha^{(k)}, \beta^{(k)}, \mu_\xi^{(k)}, \sigma_\xi^{2(k)}, \sigma_\delta^{2(k)}, \Sigma_e^{(k)}, \pi_s^{(k)}\right)$.

Paso-E De (3.36) a (3.40) con $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$, calculando la esperanza condicional

$$
\begin{aligned}
s_i^{(k)} &= \mathbb{E}\{s_i \mid z_;\boldsymbol{\theta}^{(k)}\}, \\
(vw)_i^{m(k)} &= \mathbb{E}\{(v_i w_i)^m \mid z_i; \boldsymbol{\theta}^{(k)}\} \quad \textit{param=1,2}, \\
v_i^{2(k)} &= \mathbb{E}\{v_i^2 \mid \boldsymbol{\theta}^{(k)}; \boldsymbol{z}_i\}.
\end{aligned}
$$

con estos ingredientes, computar la función objetivo

$$
Q(\theta \mid \theta^{(k)}) = \sum_{i=1}^{n} Q_i(\theta \mid \theta^{(k)}),
$$

donde

$$
\begin{aligned}
Q_i(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) &= \mathbb{E}\{\ell_{ci}(\boldsymbol{\theta}; c_i) \mid \boldsymbol{\theta}^{(k)}; \boldsymbol{z}_i\} \\
&= \frac{1}{2}(1 + s_i^{(k)}) \log \pi_s + \frac{1}{2}(1 - s_i^{(k)}) \log \pi_t + \log 2 - \frac{1}{2} \log(2\pi\sigma_\xi^2) - \frac{1}{2\sigma_\xi^2} v_i^{2(k)} \\
&\quad -\frac{1}{2} \log(2\pi\sigma_\delta^2) - \frac{1}{2\sigma_\delta^2}(x_i - \mu_x)^2 + \frac{1}{\sigma_\delta^2}(vw)^{(k)}(x_i - \mu_x) - \frac{1}{2\sigma_\delta^2}(vw)_i^{2(k)} \\
&\quad -\frac{1}{2} \log \mid 2\pi\boldsymbol{\Sigma}_e \mid -\frac{1}{2}(y_i - \mu_y)^\top \boldsymbol{\Sigma}_e^{-1}(y_i - \mu_y) + (vw)_i^{(k)}(y_i - \mu_y)^\top \boldsymbol{\Sigma}_e^{-1}\boldsymbol{\beta} \\
&\quad -\frac{1}{2}(vw)_i^{2(k)} \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_e^{-1}\boldsymbol{\beta}.
\end{aligned}
$$

Paso-CM.: Maximizando la función $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$ con respecto a $\boldsymbol{\theta}$ sobre la restricción $\pi_s + \pi_t = 1$. Esto se logra mediante los siguientes pasos:

Paso-CM.1: Actualizar $\pi_s^{(k)}$, $\pi_t^{(k)}$,$\sigma_\xi^{2(k)}$, $\mu_x^{(k)}$ y $\mu_y^{(k)}$ por

$$
\begin{aligned}
\pi_s^{(k+1)} &= \frac{1+\overline{s}^{(k)}}{2}, \\
\pi_t^{(k+1)} &= \frac{1-\overline{s}^{(k)}}{2}, \\
\sigma_\xi^{2(k+1)} &= \frac{1}{n}\sum_{i=1}^{n} v_i^{2(k)}, \\
\mu_x^{(k+1)} &= \bar{x} - (\overline{vw})^{(k)}, \\
\mu_y^{(k+1)} &= \bar{y} - (\overline{vw})^{(k)},
\end{aligned}
$$

donde $\overline{x} = \sum_{i=1}^{n} x_i/n$ y $\overline{y} = \sum_{i=1}^{n} y_i/n$, $\overline{s}^{(k)} = \sum_{i=1}^{n} s_i^{(k)}/n$, y $(\overline{vw})^{(k)} = \sum_{i=1}^{n}(vw)^{(k)}/n$.

Paso-CM.2: Fijando $\mu_x = \mu_x^{(k+1)}$ y $\mu_y = \mu_y^{(k+1)}$, actualizando $\boldsymbol{\beta}^{(k)}$, $\sigma_\delta^{2(k)}$ y $\boldsymbol{\Sigma}_e^{(k)}$ por

$$
\begin{aligned}
\boldsymbol{\beta}^{(k+1)} &= \frac{\sum_{i=1}^{n}(vw)_i^{(k)}(y_i - \mu_y^{(k+1)})}{\sum_{i=1}^{n}(vw)_i^{2(k)}}, \\
\sigma_\delta^{2(k+1)} &= \frac{1}{n}\left\{\sum_{i=1}^{n}(x_i - \mu_x^{(k+1)})^2 - 2\sum_{i=1}^{n}(vw)_i^{(k)}(x_i - \mu_x^{(k+1)}) + \sum_{i=1}^{n}(vw)_i^{2(k)}\right\}, \\
\Sigma_e^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \mu_y^{(k+1)}\right)\left(y_i - \mu_y^{(k+1)}\right)^\top + \frac{1}{n}\sum_{i=1}^{n}(vw)_i^{2(k)}\boldsymbol{\beta}^{(k+1)}\left(\boldsymbol{\beta}^{(k+1)}\right)^\top \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}(vw)_i^{(k)}\left(y_i - \mu_y^{(k+1)}\right)\left(\boldsymbol{\beta}^{(k+1)}\right)^\top - \boldsymbol{\beta}^{(k+1)}\frac{1}{n}\sum_{i=1}^{n}(vw)_i^{(k)}\left(y_i - \mu_y^{(k+1)}\right)^\top.
\end{aligned}
$$

Paso-CM.3: Actualizar $\mu_\xi^{(k)}$ y $\alpha^{(k)}$ por

$$
\begin{aligned}
\mu_\xi^{(k+1)} &= \mu_x^{(k+1)}, \\
\boldsymbol{\alpha}^{(k+1)} &= \mu_y^{(k+1)} - \boldsymbol{\beta}^{(k+1)}\mu_\xi^{(k+1)}.
\end{aligned}
$$

Observación 4.1.1. *Tras la convergencia iterativa de ECM, la estimación de γ puede ser obtenida mediante una adecuada representación de (π_s, π_t), vistas como funciones de γ. Como ya se señaló, la condición de monotonía en la Definición 2.1.1 implica que π_s y π_t son*

funciones invertibles de γ. La expresión explícita de la representación puede resolverse para elecciones específicas de $(s(\gamma), t(\gamma))$. Por ejemplo, si hemos establecido $(s(\gamma), t(\gamma)) = (1+\gamma, 1-\gamma)$ con la condición que $|\gamma| < 1$, entonces $\pi_s = (1+\gamma)/2$ y $\pi_t = (1-\gamma)/2$, y la correspondiente estimación de γ es

$$\gamma^{(k+1)} = 2\pi^{(k+1)} - 1 = \overline{s}^{(k)}$$

Observación 4.1.2. *El caso más común considerado tanto en la teoría como en las aplicaciones es cuando los errores de medición no están correlacionados, y por lo tanto la matriz $\mathbf{\Sigma}_e$ es diagonal, es decir $\mathbf{\Sigma}_e = diag(\sigma^2_{e_1}, \ldots, \sigma^2_{e_d})$. En este caso, las últimas ecuaciones del paso-CM.2 del algoritmo anterior se simplifica a:*

actualizar $\sigma^{(k)}_{ej}$ a

$$\sigma^{2(k+1)}_{e_j} = \frac{1}{n}\left\{\sum_{i=1}^{n}(y_{ij} - \mu^{(k+1)}_{yj})^2 - 2\boldsymbol{\beta}^{(k+1)}_j \sum_{i=1}^{n}(vw)^{(k)}(y_{ij} - \mu^{(k+1)}_{yj}) + (\boldsymbol{\beta}^{(k+1)}_j)^2 \sum_{i=1}^{n}(vw)^{2(k)}_i\right\},$$

para $j = 1, \ldots, d$.

Observación 4.1.3. *Las iteraciones del algoritmo ECM se repiten hasta que la diferencia entre dos valores sucesivos de la log-verosimilitud, $|\ell(\boldsymbol{\theta}^{(k+1)}; \boldsymbol{z}_1, \ldots, \boldsymbol{z}_n) - \ell(\boldsymbol{\theta}^{(k)}; \boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)|$, es suficientemente pequeña, podemos suponer que esta diferencia es de 10^{-6}. Como valor inicial del algoritmo, podemos usar el vector de medias de la muestra de $\overline{\boldsymbol{z}}$ para $\boldsymbol{\mu}$, la varianza muestral y la matriz de covarianza $S^2_x = \sum_{i=1}^n (x_i - \overline{x})^2/n$ y $S_{\boldsymbol{y}} = \sum_{i=1}^n (\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})^\top/n$ para $\sigma^2_x = \sigma^2_\xi + \sigma^2_\delta$ y $\mathbf{\Sigma}_e$, respectivamente, y la asimetría de las variables explicativas observadas $x_1, \ldots, x_n$, es es $g_1 = m_{3x}/m_{2x}$, donde $m_{kx} = \sum_{i=1}^n (x_i - \overline{x})^3/n$, para γ. Como una opción simple para los puntos de partida de σ^2_ξ y $\boldsymbol{\beta}$ uno puede tomar S^2_x y $\widehat{\beta} = S_{yx}/S^2_x$, respectivamente, como si no hubiera error de medición. Como para σ^2_δ, se puede establecer igual a una fracción de σ^2_ξ; en nuestro trabajo numérico del próximo capítulo, hemos iniciado el proceso iterativo con $\sigma^2_\delta = \sigma^2_\xi$.*

4.2. Matriz de información empírica

Para calcular la covarianza asintótica del EMV de los parámetros del TPN-MEVM propuesto, seguimos a Lin (2010). Según lo definido por Meilijson (1989), la matriz de información empírica puede ser aproximada por

$$I_e(\boldsymbol{\theta}; \boldsymbol{z}) = \sum_{i=1}^{n} U(\boldsymbol{z}_i; \boldsymbol{\theta}) U^\top(\boldsymbol{z}_i; \boldsymbol{\theta}) - n\overline{U}(\boldsymbol{z}; \boldsymbol{\theta}), \tag{4.1}$$

donde $\overline{U}(\boldsymbol{z}_i, \boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} U(\boldsymbol{z}_i; \boldsymbol{\theta})$ y $U(\boldsymbol{z}_i, \boldsymbol{\theta})$ es la función de score empírica para el índice i. Según Louis (1982), es posible relacionar la función de score de la log-verosimilitud incompleta con la experanza condicional de la función log-verosimilitud de datos completos. Por lo tanto, una función de score individual puede ser determinada como

$$U(\boldsymbol{z}_i, \boldsymbol{\theta}) = \frac{\partial \log f(\boldsymbol{z}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbb{E}\left\{ \frac{\partial \ell_{c_i}(\boldsymbol{\theta}; ci)}{\partial \theta} \middle| \boldsymbol{z}_i; \boldsymbol{\theta}. \right\},$$

donce c_i es el vector de datos completos y $\ell_{ci}(\boldsymbol{\theta}; ci)$, es la verosimilitud completa de los datos formados a partir de la i-ésima observación. Donde $\widehat{\boldsymbol{\theta}}$ es el estimador de MV de $\boldsymbol{\theta}$, dado que de (4.1) se tiene que $U(\boldsymbol{z}, \widehat{\boldsymbol{\theta}}) = 0$. Por lo tanto la matriz de covarianza asintótica de EMV puede ser aproximada por la matrix de información empírica I_e dada por

$$I_e(\widehat{\boldsymbol{\theta}}; \boldsymbol{z}) = \sum_{i=1}^{n} \widehat{U}_i \widehat{U}_i^\top, \tag{4.2}$$

donde $\widehat{U}_i = U(\boldsymbol{z}_i; \widehat{\boldsymbol{\theta}}) = \left(\widehat{U}_{i,\boldsymbol{\alpha}}, \widehat{U}_{i,\boldsymbol{\beta}}, \widehat{U}_{i,\mu_\xi}, \widehat{U}_{i,\sigma_\xi^2}, \widehat{U}_{i,\sigma_\delta^2}, \widehat{U}_{i,\sigma_{\boldsymbol{e}}^2}, \widehat{U}_{i,\gamma}\right)^\top$ con

$$
\begin{aligned}
\widehat{U}_{i,\boldsymbol{\alpha}} &= \widehat{\boldsymbol{\Sigma}}_e^{-1}\{\boldsymbol{y}_i - \widehat{\mu}_y - \widehat{vw}_i\widehat{\boldsymbol{\beta}}\}, && (4.3)\\
\widehat{U}_{i,\boldsymbol{\beta}} &= \widehat{\boldsymbol{\Sigma}}_e^{-1}\{\widehat{\mu}_\xi(\boldsymbol{y}_i - \widehat{\mu}_y) + \widehat{uw}_i(\boldsymbol{y}_i - \widehat{\mu}_y - \widehat{\mu}_\xi\widehat{\boldsymbol{\beta}}) - \widehat{v^2w^2}_i\widehat{\boldsymbol{\beta}}\}, && (4.4)\\
\widehat{U}_{i,\mu_\xi} &= \frac{1}{\widehat{\sigma}_\delta^2}(x_i - \widehat{\mu}_\xi - \widehat{vw}_i) - \widehat{\mu}_\xi\widehat{\boldsymbol{\beta}}^\top\widehat{\boldsymbol{\Sigma}}_e^{-1}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\Sigma}}_e^{-1}(\boldsymbol{y}_i - \widehat{\alpha} - \widehat{vw}_i\widehat{\boldsymbol{\beta}}), && (4.5)\\
\widehat{U}_{i,\sigma_\xi^2} &= -\frac{1}{2\widehat{\sigma}_\xi^2} + \frac{1}{2\widehat{\sigma}_\xi^4}\widehat{v^2}_i, && (4.6)\\
\widehat{U}_{i,\sigma_\delta^2} &= -\frac{1}{2\widehat{\sigma}_\delta^2} + \frac{1}{2\widehat{\sigma}_\delta^4}\{(x_i - \widehat{\mu}_\xi)^2 - 2\widehat{vw}_i(x_i - \widehat{\mu}_\xi) + \widehat{v^2w^2}_i\}, && (4.7)\\
\widehat{U}_{i,\gamma} &= \frac{1+\widehat{s}_i}{2\pi_s}\frac{\partial\pi_s}{\partial\gamma} + \frac{1-\widehat{s}_i}{2\pi_t}\frac{\partial\pi_t}{\partial\gamma}. && (4.8)
\end{aligned}
$$

Para el caso especial de la parametrización simétrica $\{s(\gamma), t(\gamma)\} = \{1+\gamma, 1-\gamma\}$, se tiene que

$$
\widehat{U}_{i,\gamma} = \frac{1}{2}\left(\frac{1+\widehat{s}_i}{1+\widehat{\gamma}} - \frac{1-\widehat{s}_i}{1-\widehat{\gamma}}\right). \tag{4.9}
$$

Además, si $\boldsymbol{\Sigma}_e$=diag$(\sigma_{e1}^2, \ldots, \sigma_{ed}^2)$, las estimaciones para $\widehat{U}_{i,\sigma_{ej}^2}$ se simplifican a

$$
\widehat{U}_{i,\sigma_{ej}^2} = -\frac{1}{2\widehat{\sigma}_{ej}^2} + \frac{1}{2\widehat{\sigma}_{ej}^4}\{(y_{ij} - \widehat{\mu}_{yj})^2 - 2\widehat{vw}_i(y_{ij} - \widehat{\mu}_{yj})\widehat{\beta}_j + \widehat{v^2w^2}_i\widehat{\beta}_j^2\}, j = 1, \ldots, d. \tag{4.10}
$$

En particular, si $\boldsymbol{\Sigma}_e = \sigma_e^2 I_d$ la ecuación (4.10) tendría una simplificación adicional, esto es

$$
\widehat{U}_{i,\sigma_{ej}^2} = -\frac{1}{2\widehat{\sigma}_{ej}^2} + \frac{1}{2\widehat{\sigma}_{ej}^4}\{(\boldsymbol{y}_i - \widehat{\mu}_y)^\top(\boldsymbol{y}_i - \widehat{\mu}_y) - 2\widehat{vw}_i\widehat{\boldsymbol{\beta}}^\top(\boldsymbol{y}_i - \widehat{\mu}_y) + \widehat{v^2w^2}_i\widehat{\beta}^\top\widehat{\beta}\}.
$$

Capítulo 5

Evaluaciones Numéricas

En este capítulo presentamos varios resultados de simulación con el objetivo de ilustrar el funcionamiento del algoritmo ECM propuesto, verificar si los errores estándar proporcionados por la aproximación de la matriz de información empírica son apropiados, y evaluar el desempeño del modelo propuesto frente perturbaciones de la distribución de la covariable latente.

5.1. Evaluación del modelo TPN-MEV cuando ξ distribuye TPN

Se realizará un pequeño estudio de simulación para investigar la funcionalidad del algoritmo ECM propuesto para el modelo TPN-MEV y verificar si estimación de los errores estándar proporcionados por la aproximación de la matriz de información empírica es apropiada.

Se simularon datos para (3.23)-(3.24) con $d = 2$, y los siguientes valores para los parámetros: $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top = (0, 0)^\top$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (1, 1)^\top$, $\mu_\xi = 1$, $\sigma^2_\xi = 1$, σ^2_δ=0.5, $\boldsymbol{\Sigma}_e = \text{diag}(\sigma^2_{e_1}, \sigma^2_{e_2})$=diag(0.25,0.25) y γ=0.25,0.5 y 0.75. Se generaron N=2000 replicas, para muestras de tamaño n=100, 500 y 1000 del modelo. En cada repetición, se obtuvieron las esti-

maciones de los parámetros y los errores estándar usando la matriz de información empírica. Finalmente, para cada componente θ_m de

$$\boldsymbol{\theta} = (\alpha_1, \alpha_2, \beta_1, \beta_2, \mu_\xi, \sigma_\xi^2, \sigma_\delta^2, \sigma_{e_1}^2, \sigma_{e_2}^2, \gamma)^\top$$

se calculó el "Promedio", la desviación estándar de las estimaciones, "SD" y los errores empíricos estándar como

$$\overline{\widehat{\theta}}_m = \frac{1}{N}\sum_{j=1}^{N}\widehat{\theta}_{m_j}, \qquad \mathrm{SD} = \left(\frac{1}{N-1}\sum_{j=1}^{N}(\widehat{\theta}_{m_j} - \overline{\widehat{\theta}}_m)^2\right)^{1/2}, \qquad \mathrm{SE_{emp}} = \frac{1}{N}\sum_{j=1}^{N}\left(I_j^{mm}\right)^{1/2},$$

respectivamente, donde I_j^{mm} denota el m-ésimo elemento diagonal de la inversa de la matriz de información empírica.

En la Tabla 5.1 se puede ver que, al aumentar el tamaño de la muestra, el sesgo de EMV $(\overline{\widehat{\theta}_k} - \theta_k)$ disminuye para cada m. Además, la desviación estándar de las estimaciones, así como los errores empíricos estándar, se acercan entre sí y disminuyen con el tamaño de la muestra, esto muestra que el cálculo de la matriz de información empírica sería correcto.

En las Tablas 5.2 y 5.3, se observa que para los valores de γ, γ=0.75 y γ=0.25, el resultado de las simulaciones es similar como cuando γ=0.5, lo obtenido es bastante signifcativo, ya que γ se encuentran cercano a su frontera ($\gamma \in (-1, 1)$), y esto puidiese ser un problema para la convergencia del algortimo EMC, lo que no ocurre.

En la Tabla 5.4, se puede ver que los valores simulados para el modelo propuesto son muy cercanos al modelo clásico TPN, a medida que el tamaño de muestra aumenta los valores se acercan y la desviación estandar disminuye, esto muestra que el comportamiento del modelo estudiado es bastante bueno.

Tabla 5.1: Estimación de la Media y desviación estandar (SD) para las estimaciones de máxima verosimilitud y estimación de los errores estandar empírico (SE_{emp}) basada en N=2000 replicas del modelo TPN-MEVM (γ=0.5).

Para-	Valor	n=100			n=500			n=1000		
metro	Verdadero	Media	SD	SE_{emp}	Media	SD	SE_{emp}	Media	SD	SE_{emp}
α_1	0	-0.013	0.187	0.204	-0.004	0.083	0.084	0.003	0.057	0.058
α_2	0	-0.010	0.183	0.204	-0.007	0.083	0.084	0.002	0.058	0.059
β_1	1	1.008	0.094	0.103	1.002	0.041	0.042	0.999	0.028	0.029
β_2	1	1.007	0.091	0.103	1.004	0.041	0.042	1.000	0.028	0.029
μ_ξ	1	0.976	0.356	0.415	1.003	0.143	0.123	1.002	0.099	0.090
σ^2_ξ	1	0.962	0.202	0.259	0.990	0.093	0.094	1.000	0.066	0.066
σ^2_δ	0.5	0.487	0.090	0.097	0.497	0.039	0.041	0.498	0.027	0.028
σ^2_{e1}	0.25	0.243	0.065	0.069	0.248	0.028	0.029	0.249	0.020	0.020
σ^2_{e2}	0.25	0.244	0.064	0.069	0.249	0.029	0.029	0.249	0.020	0.020
γ	0.5	0.532	0.225	0.424	0.503	0.087	0.129	0.496	0.059	0.091

Tabla 5.2: Estimación de la Media y desviación estandar (SD) para las estimaciones de máxima verosimilitud y estimación de los errores estandar empírico (SE_{emp}) basada en N=2000 replicas del modelo TPN-MEVM(γ=0.75).

Para-	Valor	n=100			n=500			n=1000		
metro	Verdadero	Media	SD	SE_{emp}	Media	SD	SE_{emp}	Media	SD	SE_{emp}
α_1	0	-0.033	0.227	0.250	-0.005	0.089	0.093	-0.002	0.067	0.065
α_2	0	-0.034	0.220	0.198	-0.001	0.089	0.093	-0.004	0.067	0.065
β_1	1	1.013	0.089	0.112	1.001	0.037	0.039	1.000	0.028	0.027
β_2	1	1.015	0.095	0.089	1.001	0.036	0.038	1.002	0.027	0.027
μ_ξ	1	1.000	0.296	1.484	0.987	0.171	0.246	1.004	0.129	0.143
σ^2_ξ	1	0.975	0.217	0.596	0.988	0.099	0.124	0.995	0.074	0.078
σ^2_δ	0.5	0.497	0.086	0.105	0.499	0.039	0.040	0.501	0.028	0.028
σ^2_{e1}	0.25	0.240	0.062	0.065	0.250	0.027	0.028	0.249	0.020	0.019
σ^2_{e2}	0.25	0.244	0.067	0.062	0.248	0.026	0.028	0.249	0.019	0.019
γ	0.75	0.785	0.066	1.382	0.766	0.119	0.236	0.751	0.088	0.139

Tabla 5.3: Estimación de la Media y desviación estandar (SD) para las estimaciones de máxima verosimilitud y estimación de los errores estandar empírico (SE_{emp}) basada en N=2000 replicas del modelo TPN-MEVM(γ=0.25).

Para-	Valor	n=100			n=500			n=1000		
metro	Verdadero	Media	SD	SE_{emp}	Media	SD	SE_{emp}	Media	SD	SE_{emp}
α_1	0	-0.033	0.151	0.250	-0.005	0.070	0.072	0.002	0.053	0.051
α_2	0	-0.022	0.161	0.198	0.003	0.069	0.072	0.002	0.052	0.050
β_1	1	1.018	1.098	0.112	1.003	0.043	0.043	0.999	0.031	0.030
β_2	1	1.010	0.108	0.088	0.999	0.042	0.043	0.999	0.030	0.030
μ_ξ	1	1.044	0.326	1.484	1.008	0.131	0.111	1.003	0.094	0.078
σ^2_ξ	1	0.955	0.207	0.259	1.004	0.092	0.095	1.004	0.065	0.066
σ^2_δ	0.5	0.481	0.802	0.097	0.500	0.041	0.041	0.498	0.027	0.029
σ^2_{e1}	0.25	0.239	0.069	0.069	0.248	0.028	0.029	0.249	0.021	0.021
σ^2_{e2}	0.25	0.249	0.072	0.069	0.249	0.028	0.029	0.249	0.021	0.021
γ	0.25	0.244	0.207	0.424	0.247	0.075	0.119	0.248	0.056	0.083

5.2. Evaluación del modelo TPN-MEV cuando ξ no distribuye TPN

Se han realizado simulaciones adicionales para examinar el comportamiento de las estimaciones del modelo propuesto cuando la distribución supuesta de la variable latente ξ no se especifica correctamente, es decir no es del tipo TPN. Específicamente, hemos considerado dos alternativas: en la primera, las variables ξ siguen una distribución t de Student con dos grados de libertad, denotada por t_2; en el segundo caso, se ha utilizado una distribución TP con una densidad base t_2, denominada TPt_2. Estas configuraciones son bastante diferentes de la distribución de TPN supuesta, ya que la densidad t_2 (que impulsa ambas alternativas) tiene colas mucho más pesadas que la TPN; recordar que la varianza no existe para una distribución t_2. Los otros parámetros se mantienen de la misma forma como se hicieron para las simulaciones de la Tabla 5.1.

Los resultados de estas simulaciones adicionales se presentan en la Tabla 5.5 para el caso t_2

Tabla 5.4: Estimación de los paramétros y desviación estandar (SD) usando el modelo N-MEV comparado con el modelo TPN-MEVM(γ=0.5).

Para-	Valor	n=100(N-MEV)		n=100(NTP-MEVM)		n=500(N-MEV)		n=500(NTP-MEVM)	
ámetro	Verdadero	Estimador	SD	Estimador	SD	Estimador	SD	Estimador	SD
α_1	0	0.010	0.138	-0.013	0.187	0.005	0.056	-0.004	0.083
α_2	0	-0.112	0.146	-0.010	0.183	-0.001	0.054	-0.007	0.083
β_1	1	0.964	0.106	1.008	0.094	0.977	0.042	1.002	0.041
β_2	1	1.127	0.122	1.007	0.091	0.997	0.042	1.004	0.041
μ_ξ	1	1.020	0.063	0.976	0.356	0.979	0.029	1.003	0.143
σ^2_ξ	1	0.825	0.074	0.962	0.202	1.097	0.029	0.990	0.093
σ^2_δ	0.5	0.491	0.117	0.487	0.090	0.487	0.028	0.497	0.039
σ^2_{e1}	0.25	0.272	0.196	0.243	0.065	0.254	0.057	0.248	0.028
σ^2_{e2}	0.25	0.162	0.102	0.244	0.064	0.231	0.040	0.249	0.029
γ	0.5	-	-	0.532	0.225	-	-	0.503	0.087

Para-	Valor	n=1000(N-MEV)		n=1000(NTP-MEVM)	
ámetro	Verdadero	Estimador	SD	Estimador	SD
α_1	0	-0.064	0.045	0.003	0.057
α_2	0	-0.062	0.044	0.002	0.058
β_1	1	1.073	0.034	0.999	0.028
β_2	1	1.061	0.032	1.000	0.028
μ_ξ	1	0.959	0.020	1.002	0.099
σ^2_ξ	1	0.924	0.022	1.000	0.066
σ^2_δ	0.5	0.557	0.038	0.498	0.027
σ^2_{e1}	0.25	0.213	0.065	0.249	0.020
σ^2_{e2}	0.25	0.271	0.030	0.249	0.020
γ	-	-	-	0.496	0.059

y en la Tabla 5.6 para el caso TPt_2. Los valores reportados son la estimación del sesgo y la estimación de probabilidad de cobertura (PC) de los intervalos de confianza de tipo Wald a un nivel del 95 %, es decir, intervalos del tipo $\widehat{\theta}\pm1.96\times$(SE). El indicador clave entregado por estas tablas es que los parámetros no relacionados con la distribución de ξ aún se estiman con muy bajo sesgo. Esto es especialmente importante para los α y β que típicamente representan los parámetros de primer interés en un estudio del MEV. Los valores de PC, aparte de los de la distribución ξ, generalmente están cerca del valor objetivo 0.95; a veces son algo superiores a 0.95, pero no mucho. Inevitablemente, la estimación de los parámetros de la distribución ξ está completamente sesgada. Este hecho se manifiesta especificamente para σ^2_ξ e indica que las estimaciones de los componentes de escala de la distribución TPN se han inflado para hacer frente a la variación infinita de la distribución subyacente. Gracias a este ajuste, las estimaciones de los parámetros de interés se han conservado.

Tabla 5.5: Probabilidades de cobertura (PC, en porcentaje) a un nivel nominal del 95 % y sesgo basado en muestras N=2000 del modelo MEV con $\xi_i \sim t_2$.

Para-	Valor	n=100		n=500		n=1000	
amétro	Verdadero	Sesgo	PC	Sesgo	CP	Sesgo	CP
α_1	0	-2.7$\times10^{-3}$	96.6	-8.5$\times10^{-4}$	95.3	5.4$\times10^{-5}$	94.7
α_2	0	-2.0$\times10^{-3}$	96.6	-7.8$\times10^{-4}$	96.1	1.6$\times10^{-4}$	95.1
β_1	1	1.9$\times10^{-3}$	98.4	-2.5$\times10^{-4}$	97.7	-1.7$\times10^{-4}$	97.3
β_2	1	2.0$\times10^{-3}$	98.4	2.6$\times10^{-4}$	98.1	-1.3$\times10^{-4}$	97.4
μ_ξ	1	6.4$\times10^{-3}$	75.8	7.8$\times10^{-3}$	51.9	-1.0$\times10^{-2}$	42.2
σ^2_ξ	1	6.6	0.3	8.8	0.0	8.5	0.0
σ^2_δ	0.5	9.5$\times10^{-3}$	95.0	-1.9$\times10^{-3}$	95.0	-5.7$\times10^{-4}$	94.6
σ^2_{e1}	0.25	4.3$\times10^{-3}$	95.5	-1.1$\times10^{-4}$	94.9	-8.4$\times10^{-4}$	95.1
σ^2_{e2}	0.25	5.7$\times10^{-3}$	95.8	-2.6$\times10^{-4}$	95.5	3.7$\times10^{-4}$	95.6
γ	0	6.3$\times10^{-4}$	-	5.6$\times10^{-4}$	-	1.9$\times10^{-3}$	-

Tabla 5.6: Probabilidades de cobertura (PC, en porcentaje) a un nivel nominal del 95 % y sesgo basado en muestras N=2000 del modelo MEV con $\xi_i \sim TP - t_2$.

Para-	Valor	n=100		n=500		n=1000	
amétro	Verdadero	Sesgo	PC	Sesgo	CP	Sesgo	CP
α_1	0	-4.2×10^{-3}	97.4	-7.2×10^{-4}	96.8	3.4×10^{-4}	96.6
α_2	0	-3.1×10^{-3}	97.1	-6.0×10^{-4}	96.7	3.9×10^{-4}	96.4
β_1	1	1.1×10^{-3}	98.3	2.3×10^{-4}	97.5	1.4×10^{-4}	97.3
β_2	1	3.7×10^{-4}	98.1	2.6×10^{-4}	97.5	2.1×10^{-4}	97.3
μ_ξ	1	-0.4×10^{-3}	84.5	-0.3	43.1	-2.2×10^{-1}	34.1
σ^2_ξ	1	6.8	2.1	10.9	0.0	10.9	0.0
σ^2_δ	0.5	-1.2×10^{-2}	94.8	-5.3×10^{-4}	95.0	-6.3×10^{-4}	95.1
σ^2_{e1}	0.25	-5.3×10^{-3}	94.9	-9.2×10^{-4}	94.9	-4.1×10^{-4}	95.0
σ^2_{e2}	0.25	-2.3×10^{-3}	95.1	-4.0×10^{-4}	95.5	2.8×10^{-5}	95.3
γ	0	0.1	90.3	5.2×10^{-2}	51.1	3.7×10^{-2}	41.8

Capítulo 6

Aplicaciones

En este capítulo se presentan dos ilustraciones basadas en datos reales. En este estudio se consideran los modelos N-MEV y el modelo TPN-MEVM propuesto. En tal caso es deseable disponer de un estadístico que permita seleccionar entre un modelo u otro. Los índices AIC y BIC (Criterio de información de Akaike y criterio de información bayesiano, respectivamente) son dos criterios de uso frecuente para la selección de modelos.

6.1. Aplicación 1

Primero, aplicamos el modelo TPN-MEV al conjunto de datos de Barnett (1969), donde el objetivo era evaluar las precisiones relativas y las calibraciones relativas de cuatro diferentes combinaciones operativas de instrumentos para medir varias características de la función pulmonar humana. Más recientemente, este conjunto de datos fue considerado por Galea et al. (2002) bajo análisis de influencia local de la estructura N-MEV y también por Lachos et al. (2009) bajo una perspectiva bayesiana. Estas mediciones fueron obtenidas por operadores calificados y no calificados en un grupo común de 72 pacientes.

Según lo considerado en Barnett (1969), las cuatro combinaciones de instrumentos y operadores se considerarán instrumentos 1 a 4 y el primer instrumento será el instrumento de

Tabla 6.1: Estadística descriptiva para los datos de Barnett.

	Media	Matriz de Covarianza				Asimetría	Kurtosis
x	22.461	59.252				0.493	2.637
y_1	21.757	58.145	62.755			0.655	2.657
y_2	21.486	64.022	68.223	79.903		0.511	2.505
y_3	21.022	60.530	64.618	73.683	73.166	0.478	2.273

referencia. Entonces x_i representa el valor observado del intrumento de referencia, mientras que y_{i1}, y_{i2} y y_{i3} representan los valores observados de los intrumento 2 al 4, respectivamente, $i = 1, \ldots, 72$ y $d = 3$. Se proporcionan cantidades resumidas de estos datos en la Tabla 6.1, se puede ver presencia de asimetría no despreciable en cada distribución marginal.

Los resultados del ajuste de los modelos considerados se presentan en la Tabla 6.2. La fila LRT se refiere al test de razón de verosimilitud, es decir, el doble de la diferencia de log-verosimilitud maximizada entre los dos modelos ajustados. Dado que el modelo N-MEV está anidado dentro del modelo TPN-MEV, el LRT permite probar la importancia del parámetro γ tomando la distribución χ^2_1 como la de referencia bajo la hipótesis nula $\gamma = 0$. El nivel de significancia observado 0.00235 da una indicación marcadamente significativa de la desviación de la simetría, por lo tanto, de la clara mejora lograda de la formulación TPN-MEV.

Se observa que la mayoría de las estimaciones son muy similares, y a veces casi idénticas, para el ajuste N-MEV y TPN-MEV. Las únicas diferencias sustanciales se producen con los parámetros μ_ξ y σ^2_ξ, es decir, aquellos que están relacionados con los supuestos de la distribución de la variable latente ξ_i. Estas diferencias indican que el modelo TPN-MEV propociona una descripción mas adecuada de la distribución de probabilidad subyacente. Al tener en cuenta la asimetría de los datos, la formulación del modelo TPN-MEV requiere un factor de escala más pequeño para ajustar el comportamiento de los datos, mientras que el supuesto de normalidad necesita un factor de escala mas grande, lo que implica una fracción

mayor de variabilidad explicada.

Tabla 6.2: Estimaciones de máxima verosimilitud y errores estándar (SE) de MV penalizadas (S.E.) bajo N-MEV y TPN-MEV ajustado a los datos de Barnett.

	N-MEV		TPN-MEV	
Efecto	Estimador	SE	Estimador	SE
α_1	-2.043	1.532	-2.076	1.590
α_2	-5.283	1.570	-5.271	1.619
α_3	-4.370	1.574	-4.373	1.619
β_1	1.060	0.066	1.061	0.068
β_2	1.192	0.066	1.191	0.068
β_3	1.131	0.070	1.131	0.071
μ_ξ	22.461	0.710	15.899	0.703
σ^2_ξ	53.412	0.973	46.754	1.016
σ^2_δ	5.025	0.894	5.047	0.909
σ^2_{e1}	1.915	1.051	1.795	1.639
σ^2_{e2}	2.294	13.392	3.053	11.191
σ^2_{e3}	3.884	1.012	3.926	1.013
γ			0.595	0.256
$\ell(\widehat{\theta})$	-738.373		-733.746	
LRT	9.254	(p-value= 0.00235)		

Dado el conjunto de datos, los modelos N-MEV y TPN-MEV pueden ser clasificados de acuerdo a su AIC y BIC, se observa que con el modelo que tiene el mínimo AIC es el mejor, en este caso sería el modelo propuesto.

Tabla 6.3: Estadística descriptiva de los datos de presión arterial sistólica.

	Media	Matriz de Covarianzas			Asimetría	Kurtosis
x	141.506	993.896			0.907	4.248
y_1	126.388	758.646	931.193		1.140	4.450
y_2	126.106	757.458	919.208	912.096	1.158	4.500

6.2. Aplicación 2

A continuación, se considera un conjunto de datos, extraídos de Bland & Altman (2005), que se refieren a la presión arterial sistólica de 85 pacientes, cada uno medido por tres métodos: un instrumento estándar que consiste en un monitor de presión arterial semiautomático (denotado por la variable x) y dos (por lo tanto, $d = 2$) mediciones manuales realizadas por dos observadores experimentados que usan un esfigmomanómetro (denotado como variables y_1 y y_2). La Tabla 6.3 proporciona un resumen estadístico de los datos. La suposición de que las tres mediciones están sujetas a errores aleatorios no correlacionados parece razonable. Los datos originales consisten de tres réplicas hechas por cada instrumento en cada paciente, pero aquí consideramos los datos solo con la tercera réplica para ilustrar nuestro análisis para datos sin réplicas.

Los resultados se resumen en la Tabla 6.3. En este caso, la indicación del LRT contra la simetría es particularmente fuerte, lo que indica una mejora importante con respecto al supuesto de normalidad. Una característica interesante es la disminución apreciable del valor de σ_ξ^2, aunque una evaluación formal de su significado estadístico es difícil, de todas formas podriamos pensar en una interpretación de σ_ξ^2 muy similar a la de los datos de Barnett.

Tabla 6.4: Estimaciones de máxima verosimilitud y errores estándar (SE) sobre N-MEV y TPN-MEV ajustadas al conjunto de datos de presión arterial sistólica.

	N-MEV		TPN-MEV	
Efecto	Estimador	SE	Estimador	SE
α_1	-17.661	22.283	-38.299	29.507
α_2	-16.823	22.265	-37.289	29.626
β_1	1.018	0.153	1.164	0.203
β_2	1.010	0.151	1.155	0.203
μ_ξ	141.465	6.558	126.708	5.572
σ^2_ξ	882.931	366.257	566.722	225.963
σ^2_δ	383.461	47.969	361.809	44.772
σ^2_{e1}	4.728	10.207	4.659	8.637
σ^2_{e2}	0.033	10.198	0.101	8.699
γ			0.439	0.159
$\ell(\widehat{\theta})$	-970.149		-962.404	
LRT	15.49	(p-value= 0.000083)		

Capítulo 7

Conclusiones

El desarrollo del presente trabajo consistió en el estudio de un modelo con error de medición, donde la distribución de la covariable latente ξ sigue una distribución two-piece con una distribución normal como base. En el proceso para obtener la distribución conjunta de nuestro modelo pudimos observar que esta distribución estaba completamente relacionada con una mezcla finita de componentes skew-normales.

Esta idea de suponer la distribución antes mencionada para la covarible latente fue estudiada para el caso univariado y luego se realizó una extensión multivariada. Para esta última se logró obtener distintas proposiciones, una de ellas fue la representación jerárquica de nuestro modelo TPN-MEV. Bajo esta representación se obtuvieron las esperanzas condicionales requerida por el algoritmo EM, las cuales fueron muy necesarias para la estimación de máxima verosimilitud. El enfoque empleado para este contexto fue el algoritmo EM ya que el esquema coincide con la estructura de mezclas finitas. Especificamente para estimar los parámetros del modelo, se utilizó la técnica de ECM.

Se realizaron algunos estudios de simulación, donde el resultado obtenido es bastante favo-

rable al modelo propuesto, ya que al aumentar el tamaño de la muestra, el sesgo y el EMV disminuyen. Además, la desviaciones estándar, así como los errores empíricos se acercan entre sí y disminuyen con el tamaño de la muestra. También se realizaron dos estudios de simulación adicionales, uno de ellos fue suponer que la covariable latente ξ sigue una distribución t-Student con dos grados de libertad (t_2) y la otra simulación se desarrolló suponiendo que la covariable latente ξ sigue una distribución TP pero con densidad base t_2. Los valores entregados fueron el sesgo y la estimación de la probabilidad de cobertura de los intervalos de confianza de tipo Wald a un nivel del 95 %. Pudimos observar de estas simulaciones adicionales es que los parámetros no relacionados directamente con la distribución ξ se estiman con muy bajo sesgo, lo que no sucede con algunos parámetros que están relacionados con dicha distribución, ya que estos están completamente sesgados. Se puede decir que las estimaciones de estos parámetros se han elevado para poder hacer frente a las variaciones de la distribución subyacente, gracias a este ajuste los parámetros relevantes siguen teniendo buen comportamiento.

Este trabajo fue realizado con la suposición de que la variable latente sigue una distribución TPN, sin embargo en el futuro se podría estudiar el supuesto donde la variable aleatoria en la ecuación ε (error de medición en la respuesta) sigue una distribución TPN. También se tiene en cuenta para un posible trabajo a futuro, el caso donde la densidad base se representa como mezcla de escala de densidad normal o bien tomar lo estudiado en esta tesis y llevarlo a un plano Bayesiano.

Bibliografía

Arellano-Valle, R. B. & Bolfarine, H. (1996). Elliptical structural models. *Communications in Statistics: Theory and Method*, *25*, 2319-2341.

Arellano-Valle, R. B., Bolfarine, H. & Lachos, V. H. (2005). Skew-normal linear mixed models. *Journal of Data Science*, *3*, 415-438.

Arellano-Valle,R.B., Bolfarine, H. & Vilca-Labra, F. (1996). Ultrastructural elliptical models. *The Canad. J. Statist. 24*, 207-216.

Arellano-Valle,R.B. & Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal Statistic,. 33*, 561-574.

Arellano-Valle,R.B., Gómez,H.W., & Quintana.F.A. (2005a). Statistical Inference for a General Class of Asymmetric Distributions. *Journal of Statistical Planning and Inference*, *128*, 427-443.

Arellano-Valle,R.B. & Genton, M.G. (2005). On fundamental skew distributions. *J. of Multivariate Analysis 96, 93-116.*

Arellano-Valle,R.B. S. Ozan, H. Bolfarine & V.H. Lachos (2005b). Skew normal measurement error models, *Jornual Multivariate Analysis*, *96*, 265-281.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*, 171-178.

Azzalini, A.& Capitano, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. Roy. Statist. Soc. Ser. B*, *61*, 579-602.

Azzalini, A. & Dalla-Valle, A.. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*, 715-726.

Barnett,V. D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics*, *25*, 129-142.

Bauwens, L. and Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic*

Statistic, 23, 346–354.

Bland, J.M. and Altman, D.G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*, 135-160.

Bolfarine, H & Arellano-Valle,R. (1998). Weak Nondifferential Error Models. *Statistics and Probability Letter, 40*, 279-287.

Bolfarine, H & Gale-Rojas, M. (1996). On structural comparative calibration under a t-model *Comput. Stat. 11*, 63-85.

Cheng,C.L. & Van. Ness, J.W. (1999). Statistical Regression with Measurement Error. *Arnold, London.*

Ellison,B.E. (1964). Two theorems for inferences about the normal distribution with aplications in acceptance sampling. *J.Amer: Statist. Assoc.*, *59*, 89-95.[26,233].

Dunn, G. (1992). Design and Analysis of Reliability: The statistical Evaluation of Measurement Errors. *Edward Arnold, NewYork.*

Fernández & Steel, M.F.J. (1998). On Bayesian Modeling of Fat Tails and Skewness. *J. of the American Statistical Association, 93* , 359-371.

Fechner,G.T. (1897). Kollectivmasslehre. Engleman, Leipzig.

Fuller, W.A. (1897). *Measurement error models.* Wiley, New York.

Gibbons, J. F. & Mylroie, S. (1973). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-Gaussian distributions. *Appl. Phys. Lett. 22*, 568-569.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *Internat. Econ. Rev 35*, 705-730.

John, S. (1982). The three-parameter two-piece normal family of distributions and its fitting. *Comm. Statist. A, Theory Methods*, *11*, 879-885.

Jones, M. C. (2006). *A note on rescalings, reparametrizations and classes of distributions.* J. Statistical Planning and Inference 136, 3730–3733.

Koenker, R. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *J. Amer. Stat. Assoc.*, 94, 1296-1310.

Kotz, S., Kozubowski, T. J. and Podgörski, K. (2001). *The Laplace Distribution and Generalizations.* Birkhäuser, Boston, USA.

Kendall, M.G & Stuart, A. (1979). *The Advanced Theory of Statistics,* 2, 4th ed. Hafner.New York.

Kenneth F. Wallis (2013). The Two-Piece Normal, Binormal, or Double Gaussian Distribution: Its Origin and Rediscoveries. *Statistical Science, 1*, 106-112.

Kimura, D.K.. (1992). *Finite Mixture Models* New York: J. Wiley & Sons *221.*

Lachos, V.H.Garibay, V. & Labra,F.V. (2009) A robust multivariate measurement error model with skew-normal/independent distributions and Bayesian MCMC implementation. *Statistical Methodology 6*, 527-541.

Lachos, V.H., Labra, F.V., Bolfarine, H. & Ghosh, P. (2010) Multivariate measurement error models based on scale mixtures of the skew - normal distribution. *Statistics Vol. 44 (6)*, 541-556.

Lin.T.I. (2010) Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing 20*, *25*, 343-356.

Louis, T.A., (2010). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 226-233.

Meilijson, I. (1993). A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society. Series B (Methodological),*, 127-138.

Meng,X.L. & Rubin,D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, *80*, 267-278.

Mudholkar, G. S. & Hutson, A. D. (2000). The epsilonskew-normal distribution for analyzing near-normal data. J. *J. Statist. Plann. Inference*, *83*, 291-309.

Rubio, F. J. & Steel, M. F. J. (2014). Inference in two-piece location-scale models with Jeffreys priors (with discussion) *Bayesian Analysis* , *9*, 9-22.

Tzy-Chy Lin,& Tsung-I, Lin. (2010) Supervised learning of multivariate skew normal mixture models with missing information. *Compu Stat.*, *25*, 183-201.

Wallis, K. (2013). The two-piece normal, binormal, or double Gaussian distribution: its origin and rediscoveries. *Statistical Science, forthcoming.*

Yu, K. and Zhang, J. (2013). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics-Theory and Methods* 34, 1867-1879.

Zacks,S. (1981). *Parametric Statistical Inference* Oxford:Pergammon Press, *26.*

Capítulo 8

Apéndice

Algunos resultados sobre la distribución TPN.

Los momentos centrales de orden inferior de una variable aleatoria Y con función de densidad de TPN (3.1) se pueden calcular utilizando (2.9), lo que lleva a

$$
\begin{aligned}
\mathbb{E}(Y-\mathbb{E}Y)^2 &= \frac{\omega^2}{q+1}\{\mathbb{E}|X|^2\,(q+1)(q-1)^2 - 2\,\mathbb{E}|X|^2(q^2-1)(q-1) + \mathbb{E}|X|^2(q^3+1)\} \\
&= \frac{\omega^2}{q+1}\{-\mathbb{E}|X|^2(q^2-1)(q-1) + \mathbb{E}|X|^2(q^3+1)\}, \\
\mathbb{E}(Y-\mathbb{E}Y)^3 &= \frac{\omega^3}{q+1}\{-\mathbb{E}|X|^3\,(q+1)(q-1)^3 + 3\,\mathbb{E}|X|^3\,(q^2-1)(q-1)^2 \\
&\quad -3\,\mathbb{E}|X|^2\,\mathbb{E}|X|(q^3+1)(q-1) + \mathbb{E}|X|^3\,(q^4-1)\}, \\
&= \frac{\omega^3}{q+1}\{2\,\mathbb{E}|X|^3\,(q+1)(q-1)^3 - 3\,\mathbb{E}|X|^2\,\mathbb{E}|X|\,(q^3+1)(q-1) + \mathbb{E}|X|^3\,(q^4-1)\},
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}(Y-\mathbb{E}Y)^4 &= \frac{\omega^4}{q+1}\{\mathbb{E}|X|^4(q+1)(q-1)^4 - 4\,\mathbb{E}|X|^4(q^2-1)(q-1)^3 \\
&\quad +6\,\mathbb{E}|X|^2\,\mathbb{E}|X|^2(q^3+1)(q-1)^2 - 4\,\mathbb{E}|X|^3\,\mathbb{E}|X|(q^4-1)(q-1) \\
&\quad +\mathbb{E}|X|^4\,(q^5+1)\} \\
&= \frac{\omega^4}{q+1}\{-3\,\mathbb{E}|X|^4(q+1)(q-1)^4 + 6\,\mathbb{E}|X|^2\,\mathbb{E}|X|^2(q^3+1)(q-1)^2 \\
&\quad -4\,\mathbb{E}|X|^3\,\mathbb{E}|X|(q^4-1)(q-1) + \mathbb{E}|X|^4\,(q^5+1)\}.
\end{aligned}
$$

El cálculo del límite del comportamiento de las medidas de asimetría de Pearson y kurtosis cuando $q \to \infty$, visto en el Capítulo 7, se deduce de:

$$
\begin{aligned}
\sqrt{\beta_1} &= \frac{\sqrt{\frac{2}{\pi}}\big\{\frac{4}{\pi}\big(1+1/q\big)\big(1-1/q\big)^3 - 3\big(1+1/q^3\big)\big(1-1/q\big) + 2\big(1-1/q^4\big)\big\}q^{4+1/2}\big(1+1/q\big)^{1/2}}{\Big\{1+1/q^3 - \frac{2}{\pi}\big(1-1/q^2\big)^2\big(1+1/q\big)\Big\}^{3/2}\,q^{9/2}} \\
&\to \frac{\sqrt{\frac{2}{\pi}}\big\{\frac{4}{\pi}-3+2\big\}}{\big\{1-\frac{2}{\pi}\big\}^{3/2}} \\
\beta_2 &= \frac{\big\{-\frac{12}{\pi^2}(1+1/q)(1-1/q)^4 + \frac{12}{\pi}(1+1/q^3)(1-1/q)^2 - \frac{8}{\pi}(1-1/q^4)(1-1/q) + 3(1+1/q^5)\big\}q^6\big(1+1/q\big)}{\Big\{1+1/q^3 - \frac{2}{\pi}\big(1-1/q^2\big)^2\big(1+1/q\big)\Big\}^2\,q^6} \\
&\to \frac{\big\{-\frac{12}{\pi^2}+\frac{12}{\pi}-\frac{8}{\pi}+3\big\}}{\big\{1-\frac{2}{\pi}\big\}^2}.
\end{aligned}
$$