

**MODELOS CON ERRORES EN LAS VARIABLES: APLICACIÓN A
COMPARACIÓN DE INSTRUMENTOS DE MEDICIÓN**

Manuel Galea Rojas

Departamento de Estadística, Universidad de Valparaíso, Chile

Minicurso XXXV Coloquio de la Sociedad Argentina de Estadística
Mar del Plata, Octubre de 2007

– Valparaíso, octubre de 2007 –

Resumen

Los modelos con errores en las variables, también conocidos como modelos con errores de medición corresponden a modelos con al menos un predictor medido con error. El problema de errores de medición aparece con frecuencia en diversas áreas del conocimiento. Campos, como agricultura, medicina, ingeniería, psicología, educación, y finanzas son algunas disciplinas que presentan problemas donde los predictores están contaminados por errores de medición. En este minicurso se discute el modelo simple con errores de medición y una extensión multivariada, suponiendo que los errores aleatorios siguen una distribución normal. Se presentan métodos de estimación, test de hipótesis y algunas técnicas de diagnóstico, como influencia local. Se considera principalmente el modelo estructural y se discuten extensiones dentro de la clase de distribuciones elípticas multivariadas. Finalmente, se ilustra el uso de estos modelos en la comparación de instrumentos o métodos de medición.

Contenido

Capítulo 1: Modelo Estructural Simple	1
1.1 Introducción	1
1.2 Modelo con error de medición simple	1
1.3 Estimación Máximo Verosímil	3
1.4 Test de hipótesis	4
1.5 Bondad de ajuste	5
1.6 Comparación de dos instrumentos de medición	6
Capítulo 2: Modelo Estructural Multivariado	12
2.1 Introducción	12
2.2 Modelo homocedástico	12
2.3 Modelo heterocedástico	14
2.4 Bondad de ajuste	18
2.5 Comparación de instrumentos de medición	18
Capítulo 3: Modelo Estructural t	25
3.1 Introducción	25
3.2 La distribución t multivariada	26
3.3 Modelo estructural t	29
3.4 Test de hipótesis	31
3.5 Bondad de ajuste	32
Capítulo 4: Diagnósticos de Influencia	37
4.1 Introducción	37
4.2 Influencia local	37
4.3 Derivación de la curvatura para el MEM multivariado	39
4.3.1 Matriz de información observada	40
4.3.2 Perturbación de casos	40
4.3.3 Perturbación del predictor	41
4.3.4 Perturbación de una respuesta	42
4.3.5 Perturbación de varianzas de los errores de medición	43
4.4 Derivación de la curvatura para el modelo estructural t	44
4.5 Aplicaciones	45

Apéndice A: Matriz de información observada para el modelo homocedástico	62
Apéndice B: Matriz de información observada para el modelo estructural t	63
Apéndice C: Matriz de información esperada para el modelo heterocedástico	66
Referencias	67

Capítulo 1

Modelo Estructural Simple

1.1. Introducción

Los modelos con errores de medición, *MEM* (también conocidos como modelos con errores en las variables) son generalizaciones de los modelos de regresión clásicos, donde al menos uno de los predictores está contaminado con un error de medición. En este capítulo discutiremos el *MEM* simple, que corresponde a una extensión del modelo de regresión simple.

1.2. Modelo con error de medición simple

El clásico *MEM* simple está dado por: (ver Fuller (1987))

$$\begin{aligned} X_i &= x_i + \epsilon_{i0} \\ Y_i &= y_i + \epsilon_{i1} \\ y_i &= \alpha + \beta x_i, \quad i = 1, 2, \dots, n; \end{aligned} \tag{1.1}$$

donde Y_i y X_i corresponden a la i -ésima observación, cuyos verdaderos valores son y_i y x_i , respectivamente, y ϵ_{i0} y ϵ_{i1} son errores de medición. En notación matricial podemos escribir:

$$\mathbf{Z}_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} 0 \\ \alpha \end{pmatrix} + \begin{pmatrix} 1 \\ \beta \end{pmatrix} x_i + \epsilon_i \quad i = 1, \dots, n \tag{1.2}$$

donde $\epsilon_i = \begin{pmatrix} \epsilon_{i0} \\ \epsilon_{i1} \end{pmatrix}$ son vectores aleatorios *iid* $N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \phi_0 & 0 \\ 0 & \phi_1 \end{pmatrix}\right)$.

Existen dos versiones del *MEM*, dependiendo de los supuestos acerca de la variable regresora no observable. Si x_1, x_2, \dots, x_n , son constantes desconocidas (parámetros incidentales), el modelo se denomina *MEM* Funcional; mientras que, si x_1, x_2, \dots, x_n , son

variables aleatorias *iid* $N(\mu_x, \phi_x)$ e independientes de los errores de medición, el modelo es conocido como *MEM Estructural (ME)*. En este curso nos centraremos en el modelo estructural. Luego, tenemos que:

$$\mathbf{Z}_i \text{ iid } N_2(\boldsymbol{\mu}, \mathbf{V}), \quad i = 1, \dots, n, \quad (1.3)$$

donde

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \alpha + \beta\mu_x \end{pmatrix} \quad \text{y} \quad \mathbf{V} = \begin{pmatrix} \phi_x + \phi_0 & \beta\phi_x \\ \beta\phi_x & \beta^2\phi_x + \phi_1 \end{pmatrix}. \quad (1.4)$$

Un hecho conocido es que este modelo no es identificable, ver Fuller (1987) y Kendall y Stuart (1979).

Definición 1.1 Sea Z un vector aleatorio observable con distribución $F(z, \theta)$; $\theta \in \Theta$ y $z \in \mathbf{Z}$.

(i) Se dice que el punto $\theta_* \in \Theta$ es identificable por $F(z, \theta_*)$ si no existe $\theta \in \Theta$ tal que

$$F(z, \theta_*) = F(z, \theta) \text{ para algún } z \in \mathbf{Z}. \quad (1.5)$$

(ii) El espacio paramétrico, Θ o el modelo F , se dice identificable si cada $\theta \in \Theta$ es identificable. En este caso diremos que θ es identificable. En caso contrario, diremos que θ no es identificable.

Note que el parámetro θ no será identificable si existen $\theta_* \neq \theta_{**} \in \Theta$ tales que

$$F(z; \theta_*) = F(z; \theta_{**}), \quad \forall z \in \mathbf{Z}. \quad (1.6)$$

Lema 1.1 En el modelo estructural (1.3)-(1.4), $\theta = (\alpha, \beta, \phi_0, \phi_1, \mu_x, \phi_x)^T$ no es identificable.

Ejemplo 1.1 En el modelo estructural (1.3)-(1.4) considere los siguientes valores de θ

$$\theta_* = (2, 1, 3, 7, 1, 1)^T \quad \text{y} \quad \theta_{**} = (1, 2, 7/2, 6, 1, 1/2)^T.$$

De (1.4) es fácil ver que ambos valores de θ generan la misma media y la misma matriz de covarianzas de \mathbf{Z}_i :

$$E(\mathbf{Z}_i) = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad \text{y} \quad \text{Var}(\mathbf{Z}_i) = \begin{pmatrix} 4 & 1 \\ 1 & 8 \end{pmatrix}, \quad i = 1, 2, \dots, n,$$

de donde sigue que θ no es identificable.

Una de las consecuencias de que θ no sea identificable es dada en el siguiente lema.

Lema 1.2 En el *ME*, si θ no es identificable, entonces no existe un estimador consistente para θ .

Del lema tenemos que falta de identificabilidad implica falta de estimadores consistentes, lo que tiene serias implicaciones en la teoría asintótica de estimación y prueba de hipótesis. Con todo, si θ es identificable no implica que exista un estimador consistente para θ . Para un ejemplo interesante ver Grabielsen (1978).

Una de las formas de resolver el problema de identificabilidad es colocar restricciones sobre θ . En este caso supondremos que la matriz de covarianzas de los errores de medición, ϵ_i , es $\phi\Sigma_0$, con Σ_0 conocida, y sin pérdida de generalidad, ver Gleser (1981), podemos suponer que $\Sigma_0 = \mathbf{I}_2$, la matriz identidad de orden 2. Este caso incluye la usual restricción, ver Fuller (1987), de razón de varianzas de los errores de medición $\lambda = \phi_1/\phi_0$ conocida, donde $\Sigma_0 = \text{Diag}(1, \lambda)$, y sin pérdida de generalidad podemos tomar $\lambda = 1$. Para otras restricciones de identificabilidad ver Fuller (1987), Shyr y Gleser (1986), Jaech (1985) y Cheng y van Ness (1999).

1.3. Estimación Máximo Verosímil

Desde (1.3)-(1.4) sigue que la función de log-verosimilitud está dada por:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n l_i(\boldsymbol{\theta}), \quad (1.7)$$

donde, $\boldsymbol{\theta} = (\alpha, \beta, \phi, \mu_x, \phi_x)^T$, $l_i(\boldsymbol{\theta}) = -\log(2\pi) - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\delta_i(\boldsymbol{\theta})$, con $\delta_i(\boldsymbol{\theta}) = (Z_i - \boldsymbol{\mu})^T \mathbf{V}^{-1}(Z_i - \boldsymbol{\mu})$. Luego los estimadores máximo verosímiles son, (ver Fuller, 1987):

$$\hat{\mu}_x = \bar{X}, \quad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}, \quad \hat{\beta} = \frac{-(S_{XX} - S_{YY}) + \sqrt{(S_{XX} - S_{YY})^2 + 4S_{XY}^2}}{2S_{XY}} \quad (1.8)$$

y

$$\hat{\phi} = \frac{1}{c}(S_{YY} - 2\hat{\beta}S_{XY} + \hat{\beta}^2S_{XX}), \quad \hat{\phi}_x = \frac{S_{XY}}{n\hat{\beta}},$$

donde,

$$c = (1 + \hat{\beta}^2), \quad S_{XX} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S_{XY} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad \text{y} \quad S_{YY} = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Para estimar la varianza de los EMV podemos usar la inversa de la matriz de información esperada. Este resultado es presentado en el siguiente teorema.

Teorema 1.1 Bajo el ME : $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_5(0; \boldsymbol{\Omega}_\theta)$, donde

$$\boldsymbol{\Omega}_\theta = \begin{pmatrix} (\mu_x^2 \phi^2 k / \phi_x^2) + c\phi & -\mu_x \phi^2 k & 0 & \beta \phi & (\mu_x \phi^2 \beta k) / \phi_x^{3/2} c \\ & c/k_1 \phi_x & 0 & 0 & \beta/k_1 \sqrt{\phi_x} \\ & & 2\phi^2 & 0 & -\phi \sqrt{\phi_x} / (k-1) \\ & & & \phi_x + \phi & 0 \\ & & & & (1/k_1) + (\phi_x/2) \end{pmatrix}, \quad (1.9)$$

con $k = 1 + \phi_x c / \phi$, $k_1 = (1 - k^{-1}) / \phi = \phi_x c / \phi^2 k$. Aquí, $\xrightarrow{\mathcal{D}}$ indica convergencia en distribución cuando $n \rightarrow \infty$. Es decir, para n grande, $\hat{\boldsymbol{\theta}} \sim AN_5(\boldsymbol{\theta}, \frac{1}{n} \boldsymbol{\Omega}_\theta)$

Demostración: Primero calculamos la matriz de información esperada usando la formula dada por Dolby (1976),

$$-E \left(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right) = n \left\{ \frac{1}{2} \text{tr} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_i} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \theta_j} \right) + \frac{\partial \boldsymbol{\mu}'}{\partial \theta_i} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_j} \right\},$$

y el resultado sigue, después de alguna algebra. Aquí, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)^T = (\alpha, \beta, \phi, \mu_x, \phi_x)^T$.

Notemos que $\sqrt{n}(\hat{\alpha} - \alpha, \hat{\beta} - \beta, \hat{\phi} - \phi) \xrightarrow{\mathcal{D}} N_3(0, \Gamma)$, donde $\Gamma = \Gamma_{\theta\theta}$ dada no Teorema 1.3.1 de Fuller (1987).

1.4. Test de hipótesis

Al igual que en regresión lineal simple, una hipótesis de interés en este caso es $H_{01} : \beta = 0$. Note que $Cov(Y_i, X_i) = \beta \phi_x$, de modo que $\beta = 0$ es equivalente con $\rho = \rho(Y_i, X_i) = 0$. Luego el test de la razón de verosimilitudes (TRV) conduce al siguiente estadístico,

$$T = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}, \quad (1.10)$$

el cual, bajo H_{01} , tiene una distribución t con $n-2$ grados de libertad, y $r = S_{XY} / \sqrt{S_{XX} S_{YY}}$ es el coeficiente de correlación muestral.

Otra hipótesis de interés es $H_{02} : \beta = \beta_0$, con β_0 conocido. El TRV esta dado por,

$$G = 2\{l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_0)\} = n \log \left\{ \frac{(\beta_0^2 S_{YY} + 2\beta_0 S_{XY} + S_{XX})(S_{YY} - 2\beta_0 S_{XY} + \beta_0^2 S_{XX})}{(1 + \beta_0^2)(S_{YY} - S_{XY}^2 / S_{XX})} \right\}, \quad (1.11)$$

el cual bajo H_{02} tiene, para n grande, aproximadamente una distribución chi cuadrado con 1 grado de libertad. Como ha sido discutido en la literatura, Wong (1989) y Arellano-Valle

y Bolfarine (1996a) la aproximación de la distribución del estadístico G a la distribución chi-cuadrado puede mejorarse usando la corrección de Bartlett. En efecto, tanto Wong (1989) como Arellano-Valle y Bolfarine (1996a) muestran, usando distintas técnicas, que el TRV corregido está dado por,

$$G^* = \frac{G}{(1 + 5/2n)}. \quad (1.12)$$

Arellano-Valle y Bolfarine (1998) proponen el test score para probar la hipótesis $H_{02} : \beta = \beta_0$. En efecto ellos muestran que el estadístico score esta dado por,

$$S_R = n(d_1^T S d_2)^2 / (d_1^T S d_1)(d_2^T S d_2), \quad (1.13)$$

donde $d_1 = \frac{1}{\sqrt{c}}(1, \beta_0)^T$, $d_2 = \frac{1}{\sqrt{c}}(-\beta_0, 1)^T$ y $\mathbf{S} = [(S_{ij})]$, con $S_{11} = S_{XX}$, $S_{12} = S_{XY}$ y $S_{22} = S_{YY}$.

Arellano-Valle y Bolfarine (1998) muestran que, bajo H_{02} ,

$$\frac{S_R}{n} \sim \text{Beta} \left(\frac{1}{2}, \frac{n-2}{2} \right),$$

una distribución beta con parámetros $\frac{1}{2}$ y $\frac{n-2}{2}$. De aquí sigue que, bajo H_{02} , la estadística

$$F_S = (n-2) \left(\frac{n}{S_R} - 1 \right)^{-1} = (n-2) \left(\frac{S_R/n}{1 - \frac{S_R}{n}} \right) \sim F(1, n-2), \quad (1.14)$$

una distribución F con 1 y $n-2$ grados de libertad. Es decir el estadístico score tiene una distribución exacta.

1.5. Bondad de ajuste

Para evaluar el ajuste del modelo estructural normal, podemos construir el gráfico de las distancias transformadas, ver Little (1988) y Lange et al. (1989).

Recordemos que

$$\delta_i(\boldsymbol{\theta}) = (\mathbf{Z}_i - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}), \quad i = 1, 2, \dots, n. \quad (1.15)$$

es la distancia de Mahalanobis. Bajo normalidad, sabemos que $\delta_i(\boldsymbol{\theta})$ tiene una distribución chi-cuadrado con 2 grados de libertad, $i = 1, 2, \dots, n$. Además, $\delta_i(\hat{\boldsymbol{\theta}})$ tiene asintóticamente la misma distribución χ^2 . Usando la aproximación de Wilson-Hilferty (Johnson y Kotz, 1970) tenemos que

$$z_i = 3(\delta_i(\hat{\boldsymbol{\theta}})/2)^{1/3} - \frac{8}{3}, \quad (1.16)$$

$i = 1, 2, \dots, n$, se distribuyen aproximadamente normal estándar. Un $Q - Q$ plot de las distancias transformadas z_i , $i = 1, 2, \dots, n$, dadas por (1.16), puede ser utilizado para evaluar el ajuste del modelo estructural.

1.6. Comparación de dos instrumentos de medición

La necesidad de comparar instrumentos de medición que difieren en costo, velocidad o otros factores, ha aparecido con frecuencia en diversas áreas del conocimiento. Grubbs (1948, 1973), por ejemplo, compara tres cronómetros, Barnett (1969) da un ejemplo donde cuatro combinaciones instrumento-operador concebidos para medir la capacidad vital en un grupo de pacientes son evaluados. Leurgans (1980) compara dos métodos para medir la concentración de glucosa en la sangre. Varios otros ejemplos, en el área médica, son mencionados por Kelly (1984, 1985) y Kaaks et al. (1994). Jaech (1985) da varios ejemplos en ingeniería, Fuller (1987) presenta casos en agricultura. Dunn (1992, 2004) presenta aplicaciones en Psicología y Educación.

El termino método o instrumento (de medición), se refiere a diferentes formas de medición de una característica dada. El termino también puede ser aplicado cuando la misma forma de medición es hecha por diferentes aparatos o diferentes operadores. Por ejemplo, cuando un análisis químico es realizado en diferentes laboratorios, cada laboratorio es identificado como un método o instrumento de medición. En algunos casos puede ocurrir, también, que los diferentes métodos sean un mismo instrumento avaluado en diferentes tiempos. En esta sección nos centraremos en el caso de dos instrumentos. En efecto, supongamos que una característica x es medida por dos métodos en un grupo común de n unidades experimentales. Como en Barnett (1969) suponemos que uno de los instrumentos es estándar (control), que mide x con un error de medición y otro nuevo, que se desea comparar con el estándar, y que mide x también con un error de medición y posiblemente con sesgo aditivo y multiplicativo. Es decir un modelo para esta situación es

$$X_i = x_i + \epsilon_{i0}, \quad (1.17)$$

$$Y_i = \alpha + \beta x_i + \epsilon_{i1}$$

donde los errores de medición $(\epsilon_{i0}, \epsilon_{i1})^T$ son no correlacionados.

Un concepto importante en la comparación de instrumentos de medición es la confiabilidad. En lo que sigue, $\beta_0 = 1$ y $\beta_1 = \beta$.

Definición 1.2 La confiabilidad del j -ésimo instrumento, denotada por κ_j se define

como,

$$\kappa_j = \beta_j^2 \phi_x / (\beta_j^2 \phi_x + \phi), \quad j = 0, 1. \quad (1.18)$$

Ahora, como $Cov(Y_i, x_i) = \beta \phi_x$, $Var(x_i) = \phi_x$ y $Var(Y_i) = \beta^2 \phi_x + \phi$, tenemos que κ_1 corresponde al cuadrado del coeficiente de correlación entre Y_i y x_i . O sea, es una medida tipo coeficiente de determinación usada en regresión lineal y puede ser interpretada como la proporción de la variación observada “explicada” por la recta $Y_i = \alpha + \beta x_i + \epsilon_{i1}$ y luego $0 \leq \kappa_1 \leq 1$. Similar interpretación tiene κ_0 .

Otra medida o índice usado para evaluar la calidad de un instrumento o método de medición es la precisión. Se conocemos los parámetros α y β , entonces las observaciones Y_i , $i = 1, \dots, n$, en el nuevo instrumento pueden ser calibradas como

$$Y_i^* = \frac{1}{\beta}(Y_i - \alpha) = x_i + \frac{1}{\beta}\epsilon_{i1} = x_i + \epsilon_{i1}^*,$$

las cuales son “no sesgadas” para x_i , $i = 1, 2, \dots, n$. Los errores de medición ϵ_{i1}^* de Y_i^* tienen varianza ϕ/β^2 . Ya que después de la calibración los dos instrumentos miden x_i sin sesgo, una definición natural de la precisión sería, ver Shyr y Gleser (1986) y Cochran (1968):

Definición 1.3 *La precisión del j -ésimo instrumento, denotada por π_j se define, como:*

$$\pi_j = \beta_j^2 / \phi, \quad j = 0, 1. \quad (1.19)$$

Theobald y Mallison (1978) usan la “precisión relativa”, denotada por τ_j , definida por:

$$\tau_j = \pi_j \phi_x = \beta_j^2 \phi_x / \phi, \quad j = 0, 1. \quad (1.20)$$

Sin embargo, Draper y Guttman (1975, 1976), definen la precisión relativa como

$$\delta_j \equiv \frac{1}{\tau_j} = \phi / \beta_j^2 \phi_x, \quad j = 0, 1. \quad (1.21)$$

Note que

$$\kappa_j = \phi_x \pi_j / (\phi_x \pi_j + 1) = \tau_j / (\tau_j + 1) = 1 / (1 + \delta_j), \quad j = 0, 1 \quad (1.22)$$

y que

$$\kappa_0 = \kappa_1 \Leftrightarrow \pi_0 = \pi_1 \Leftrightarrow \tau_0 = \tau_1 \Leftrightarrow \delta_0 = \delta_1. \quad (1.23)$$

De las medidas anteriores, las más utilizadas son la confiabilidad y la precisión. Note que la confiabilidad no depende de las unidades de medición de los x_i y es fácil de interpretar. Sin embargo, de (1.23) vemos que comparar confiabilidades es equivalente a

comparar precisiones o precisiones relativas.

En general el problema de comparar dos instrumentos se reduce a hacer inferencias sobre los sesgos (α, β) , confiabilidades (κ_j) y/o precisiones (π_j) . Es necesario considerar el sesgo y la confiabilidad (precisión) simultáneamente, pues un valor alto de κ_j (próximo de 1) implica sólo que Y_i (medida hecha por el nuevo instrumento) es una medida confiable de $\alpha + \beta x$. El nuevo instrumento proporcionaría medidas “exactas” de la característica x sólo si κ_1 es próximo de uno y si fuese insesgado ($\alpha = 0, \beta = 1$).

Si estamos interesados en comparar las precisiones de los instrumentos, π_0 y π_1 , las hipótesis relevantes pueden formularse como,

$$H_0 : \pi_1 \leq \pi_0, \quad H_1 : \pi_1 > \pi_0. \quad (1.24)$$

Es decir, H_1 es la hipótesis que el instrumento 1 es mejor, más preciso, que el instrumento estándar. Note que H_0 puede formularse equivalentemente en términos de los coeficientes de confiabilidad: $H_0 : \kappa_1 \leq \kappa_0$. Para probar H_0 podemos usar el test propuesto por Pitman (1939). En efecto, sea

$$\mathbf{W}_i = \begin{pmatrix} W_{i1} \\ W_{i2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \mathbf{T} \mathbf{Z}_i, \quad i = 1, 2, \dots, n. \quad (1.25)$$

con $\mathbf{T} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Luego, tenemos que

$$\mathbf{W}_i \sim N_2 \left(\begin{pmatrix} \alpha + \mu_x(\beta + 1) \\ \alpha + \mu_x(\beta - 1) \end{pmatrix}, \begin{pmatrix} 2\phi + \phi_x(1 + \beta)^2 & \phi_x(\beta^2 - 1) \\ \phi_x(\beta^2 - 1) & 2\phi + \phi_x(1 - \beta)^2 \end{pmatrix} \right). \quad (1.26)$$

Luego el coeficiente de correlación, ρ_{12} , entre W_{i1} y W_{i2} es,

$$\rho_{12} = \frac{\phi_x(\beta^2 - 1)}{\sqrt{(2\phi + \phi_x(1 + \beta)^2)(2\phi + \phi_x(1 - \beta)^2)}}. \quad (1.27)$$

Las hipótesis (1.24) son equivalentes a

$$H_0 : \rho_{12} \leq 0, \quad H_1 : \rho_{12} > 0. \quad (1.28)$$

Es bien sabido que el TRV para probar las hipótesis (1.28) es

$$t_n = \frac{r_{12}\sqrt{n-2}}{\sqrt{1-r_{12}^2}} \quad (1.29)$$

donde r_{12} es el coeficiente de correlación muestral entre W_1 y W_2 . En términos de los datos originales,

$$t_n = \frac{\sqrt{n-2}(S_{YY} - S_{XX})}{2\sqrt{|\mathbf{S}|}}. \quad (1.30)$$

Finalmente rechazamos H_0 si $t_n > t_\eta$, donde t_η es el percentil $100(1 - \eta)\%$ de la distribución \mathbf{t} con $n - 2$ grados de libertad. Shyr (1984) muestra que este test es insesgado y uniformemente más potente.

Otra hipótesis de interés podría ser: $H_{01} : \alpha = 0, \beta = 1, \kappa_0 = \kappa_1$. O sea, estamos interesados en probar si los instrumentos miden la característica x sin sesgo y con la misma confiabilidad. Esta hipótesis puede probarse, por ejemplo, usando el test de Wald. Naturalmente, dependiendo del problema, podría ser de interés probar otro tipo de hipótesis.

Ilustración: Datos conducto radicular. Se desea comparar dos métodos de medición del conducto radicular, uno fue el convencional usando radiografía (X), y otro fue un aparato electrónico llamado LAE, (Y). En la Tabla 1.1 se presentan los resultados para una muestra de $n = 110$ conductos. En la figura (1.1) se muestra el gráfico de dispersión de los datos, observándose una relación lineal entre las dos mediciones.

Al aplicar el modelo estructural se obtienen los siguientes resultados:

Tabla 1.2: EMV y sus errores estándar asintóticos

Parámetro	Estima	Error estándar
α	-0.4675	0.5405
β	1.0166	0.0264
ϕ	0.2133	0.0288
μ_x	20.3690	0.2335
σ_x	2.4046	0.1681

La figura 1.2 muestra el gráfico de las distancias transformadas para el ME . Se puede observar un ajuste razonable a pesar de que algunas observaciones podrían ser outliers.

Tabla 1.1: Medidas del conducto radicular, hechas por 2 instrumentos

Obs.	X	Y	Obs.	X	Y	Obs.	X	Y
1	16.5	15.5	38	22.0	22.5	75	21.0	19.5
2	19.0	17.5	39	18.0	17.5	76	21.0	21.5
3	17.5	17.5	40	25.0	24.0	77	20.5	20.0
4	18.0	17.5	41	23.2	23.5	78	20.5	20
5	20.8	20.5	42	22.4	22.0	79	21.5	21.5
6	17.9	17.5	43	19.0	20.0	80	22.0	21.0
7	19.5	19.5	44	25.0	23.5	81	22.0	22.0
8	18.0	17.0	45	24.0	24.5	82	22.5	22.0
9	16.0	15.5	46	20.0	20.5	83	21.0	20.5
10	14.0	13.5	47	20.0	21.0	84	20.0	20.5
11	20.5	20.5	48	18.1	18.5	85	19.0	19.5
12	14.5	15.5	49	22.0	22.5	86	25.0	25.5
13	14.5	14.5	50	15.0	15.5	87	21.0	20.5
14	23.5	23.5	51	15.0	14.5	88	20.5	20.0
15	18.0	18.5	52	22.0	23.0	89	25.5	24.5
16	19.5	19.0	53	22.0	22.5	90	24.0	24.0
17	19.5	18.5	54	18.5	18.5	91	17.5	17.5
18	23.0	22.5	55	18.5	18.5	92	20.0	18.5
19	19.0	19.5	56	19.0	20.5	93	19.0	19.5
20	19.0	19.0	57	27.6	27.5	94	21.6	21.5
21	22.5	22.0	58	20.5	21.0	95	21.5	22.5
22	22.0	22.5	59	23.0	22.5	96	24.5	24.5
23	20.0	20.0	60	22.0	22.0	97	22.5	22.5
24	20.0	20.5	61	22.5	22.5	98	20.0	20.5
25	23.0	23.0	62	20.5	18.5	99	19.0	19.5
26	21.0	22.0	63	23.0	23.0	100	18.0	17.5
27	21.0	19.0	64	22.0	22.0	101	19.0	19.5
28	19.5	19.0	65	21.5	21.5	102	19.5	19.5
29	19.5	19.0	66	19.0	18.5	103	19.5	19.5
30	19.5	19.0	67	18.0	18.0	104	21.5	21.5
31	17.5	17.0	68	23.0	21.5	105	21.5	22.0
32	19.0	18.5	69	22.0	21.0	106	20.5	21.0
33	20.5	19.5	70	25.0	25.0	107	21.0	20.5
34	19.0	18.5	71	21.0	20.5	108	22.0	22.5
35	20.0	19.5	72	19.0	19.0	109	16.5	16.5
36	22.0	22.0	73	18.0	18.0	110	20.0	19.5
37	20.0	20.5	74	18.5	19.0			

Fuente: Trabajo de titulación UV

Figura 1.1: Gráfico de dispersión

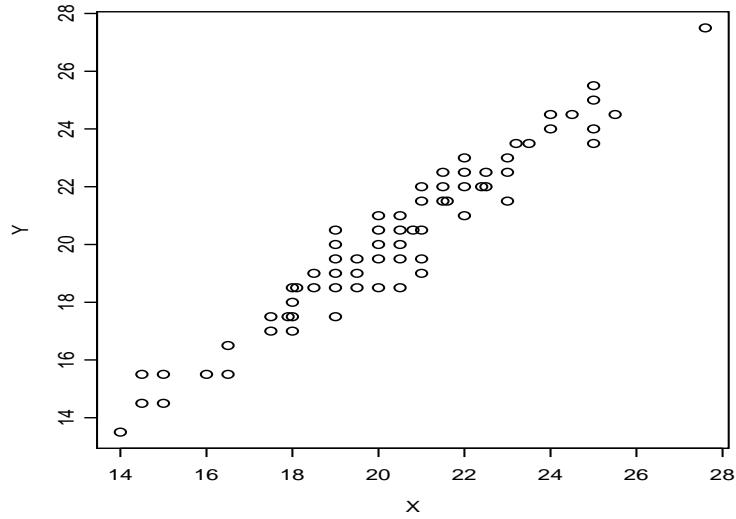
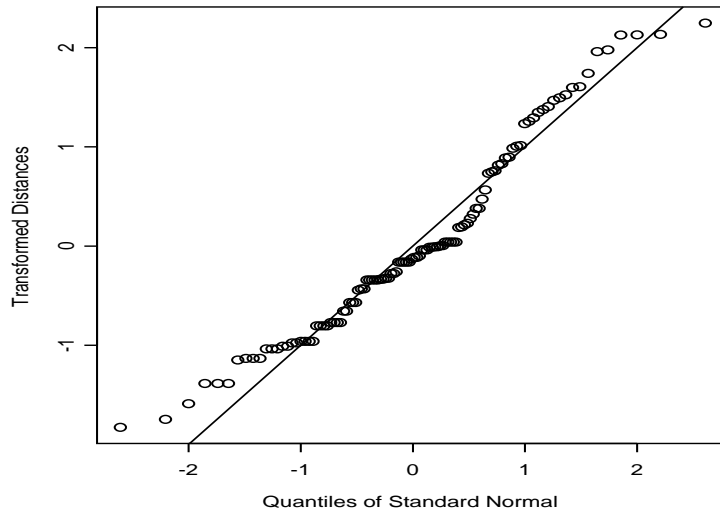


Figura 1.2: Gráfico de la distancia transformada para el modelo normal



Las estimas de la precisión de cada instrumento son $\hat{\pi}_0 = 4,6882$ y $\hat{\pi}_1 = 4,8452$. El valor del estadístico $t_n = \sqrt{n-2}(S_{YY} - S_{XX})/2\sqrt{|\mathbf{S}|} = \sqrt{108}(6,1896 - 5,9954)/2\sqrt{2,5546} = 0,3949$. Luego no podemos rechazar la hipótesis, $H_0 : \pi_1 \leq \pi_0$.

Capítulo 2

Modelo Estructural Multivariado

2.1. Introducción

En este capítulo consideraremos una extensión del *MEM* simple, al caso donde disponemos de varias respuestas y un predictor medido con error.

2.2. Modelo homocedástico

Sean x_i y X_i los valores verdadero y observado del predictor x en la unidad experimental i , y Y_{ij} las respuestas observadas en la unidad i , $i = 1, \dots, n$ y $j = 1, \dots, p$. Como en Barnett (1969), Shyr (1984), Kelly (1984) y Kimura (1992) el *MEM* multivariado puede escribirse como,

$$Y_{ij} = \alpha_j + \beta_j x_i + \epsilon_{ij} , \quad (2.1)$$

$i = 1, \dots, n$ y $j = 1, \dots, p$, and

$$X_i = x_i + \epsilon_{i0} , \quad (2.2)$$

$i = 1, \dots, n$. Usaremos la siguiente notación:

$$\mathbf{Z}_i = (X_i, Y_{i1}, \dots, Y_{ip})^T , \quad \boldsymbol{\epsilon}_i = (\epsilon_{i0}, \epsilon_{i1}, \dots, \epsilon_{ip})^T ,$$

los cuales son vectores aleatorios de dimension $r = p + 1$,

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^T \quad \text{y} \quad \boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T ,$$

que corresponden a vectores de parámetros de dimension p . Luego, en notación matricial podemos escribir,

$$\mathbf{Z}_i = \begin{pmatrix} 0 \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix} x_i + \boldsymbol{\epsilon}_i , \quad (2.3)$$

$i = 1, \dots, n$. También suponemos que

$$\boldsymbol{\epsilon}_i \sim \mathbf{N}_r(\mathbf{0}, \phi \boldsymbol{\Sigma}_0),$$

$i = 1, \dots, n$, con $\boldsymbol{\Sigma}_0$ conocido, la cual podemos tomar, sin pérdida de generalidad, igual a la matriz identidad de dimension r , que denotaremos por \mathbf{I}_r . Note que si $p = 1$ tenemos el *MEM* simple discutido en el capítulo 1.

Al igual que en el *MEM* simple podemos definir dos versiones del *MEM* multivariado, la funcional y la estructural. Aquí nos centraremos en el caso estructural donde suponemos que x_i es independiente de $\boldsymbol{\epsilon}_i$, $i = 1, \dots, n$. Luego, tenemos que

$$\mathbf{Z}_i \sim \mathbf{N}_r\left(\begin{pmatrix} 0 \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix} \mu_x, \mathbf{V}\right), \quad (2.4)$$

son vectores aleatorios iid, $i = 1, \dots, n$, con

$$\mathbf{V} = \phi_x \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix}^T + \phi \mathbf{I}_r.$$

Sean $d_1 \geq d_2 \geq \dots \geq d_r$ los valores propios de la matriz de covarianza muestral, la cual denotamos por

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T,$$

con $\bar{\mathbf{Z}} = \sum_{i=1}^n \mathbf{Z}_i/n$. Sabemos que \mathbf{S} puede ser escrita como

$$\mathbf{S} = \mathbf{G} \mathbf{D}(\mathbf{d}) \mathbf{G}^T,$$

donde \mathbf{G} es una matriz ortogonal $r \times r$ con la i -ésima columna correspondiente al i -ésimo vector propio de \mathbf{S} y $\mathbf{D}(\mathbf{d}) = \text{diag}(d_1, \dots, d_r)$, una matriz diagonal. Finalmente, particionamos \mathbf{G} como sigue:

$$\mathbf{G} = \begin{pmatrix} g_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{pmatrix},$$

donde g_{11} es un escalar y \mathbf{G}_{22} es $p \times p$. Así, los EMV están dados por, Gleser (1981) y Shyr (1984),

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \bar{\mathbf{Y}} - \bar{X} \hat{\boldsymbol{\beta}}, \\ \hat{\boldsymbol{\beta}} &= \frac{\mathbf{G}_{21}}{g_{11}} \quad \text{y} \quad \hat{\phi} = \frac{1}{p} \sum_{i=2}^r d_i \end{aligned}$$

y

$$\hat{\mu}_x = \bar{X} = \sum_{i=1}^n X_i/n \quad \text{y} \quad \hat{\phi}_x = \frac{pd_1 - \sum_{i=2}^r d_i}{p\hat{c}},$$

con $\bar{\mathbf{Y}} = \sum_{i=1}^n \mathbf{Y}_i/n$, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip})^T$, $i = 1, \dots, n$ y $\hat{c} = 1 + \hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\beta}}$. Alternativamente, los EMV pueden obtenerse usando el algoritmo EM, ver Kimura (1992) y Bolarine y Galea (1995).

Para estimar la varianza de los estimadores podemos usar la matriz de información observada o la esperada. Nosotros usamos la matriz de información observada, la cual se encuentra en el apéndice A. Dado que el vector $\boldsymbol{\beta}$ es de interés primario en este modelo, a continuación presentamos su distribución asintótica. Ver Shyr (1984).

Teorema 2.1 Bajo el modelo (2.4): $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N_p(0; \Omega_\beta)$, donde

$$\Omega_\beta = \left\{ \frac{1}{1 + \boldsymbol{\beta}^T \boldsymbol{\beta}} \frac{1}{\tau_0^2} + \frac{1}{\tau_0} \right\} (I_p + \boldsymbol{\beta} \boldsymbol{\beta}^T), \quad (2.5)$$

con $\tau_0 = \phi_x / \phi$.

Usando este resultado podemos construir, por ejemplo, el test de Wald para probar hipótesis lineales de la forma $H : \mathbf{A}\boldsymbol{\beta} = \mathbf{a}$.

2.3. Modelo heterocedástico

En el modelo anterior hemos supuesto que la matriz de covarianza de los errores es de la forma $\phi \boldsymbol{\Sigma}_0$, con $\boldsymbol{\Sigma}_0$ conocida, lo que conduce a un modelo homocedástico, donde las varianzas de los errores de medición son iguales a ϕ . En esta sección consideraremos el modelo (2.3), pero las varianzas de los errores de medición pueden ser distintas, dando origen a un modelo heterocedástico. En efecto, desde (2.3) tenemos que

$$\mathbf{Z}_i = \begin{pmatrix} 0 \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix} x_i + \boldsymbol{\epsilon}_i, \quad (2.6)$$

$i = 1, \dots, n$, donde ahora suponemos que

$$\boldsymbol{\epsilon}_i \sim N_r(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon),$$

$i = 1, \dots, n$, con $\boldsymbol{\Sigma}_\epsilon = \text{Diag}(\sigma_0^2, \sigma_1^2, \dots, \sigma_p^2)$, una matriz diagonal $r \times r$. Para $p \geq 2$ este modelo es identificable y no necesitamos imponer restricciones sobre los parámetros. Así entonces tenemos que

$$\mathbf{Z}_i \text{ iid } N_r(\boldsymbol{\mu}, \mathbf{V}), \quad i = 1, \dots, n, \quad (2.7)$$

donde

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \boldsymbol{\alpha} + \boldsymbol{\beta} \mu_x \end{pmatrix} \quad \text{y} \quad \mathbf{V} = \sigma_x^2 \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix} \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix}^T + \boldsymbol{\Sigma}_\epsilon. \quad (2.8)$$

Se sabe que (ver Barnett, 1969) cuando $p = 2$, los EMV tienen forma explícita y están dados por:

$$\begin{aligned}\hat{\mu}_x &= \bar{X}, & \hat{\sigma}_x^2 &= S_{12}^{-1}S_{01}S_{02}, & \hat{\alpha}_1 &= \bar{Y}_1 - \hat{\beta}_1\bar{X}, & \hat{\alpha}_2 &= \bar{Y}_2 - \hat{\beta}_2\bar{X}, \\ & & & & \hat{\beta}_1 &= S_{02}^{-1}S_{12}, & \hat{\beta}_2 &= S_{01}^{-1}S_{12}\end{aligned}\quad (2.9)$$

y,

$$\hat{\sigma}_0^2 = S_{00} - \hat{\sigma}_x^2, \quad \hat{\sigma}_1^2 = S_{11} - \hat{\beta}_1^2\hat{\sigma}_x^2, \quad \hat{\sigma}_2^2 = S_{22} - \hat{\beta}_2^2\hat{\sigma}_x^2,$$

donde,

$$\begin{aligned}S_{00} &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, & S_{01} &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_{i1} - \bar{Y}_1), & S_{02} &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_{i2} - \bar{Y}_2), \\ S_{11} &= \frac{1}{n} \sum_{i=1}^n (Y_{i1} - \bar{Y}_1)^2, & S_{12} &= \frac{1}{n} \sum_{i=1}^n (Y_{i1} - \bar{Y}_1)(Y_{i2} - \bar{Y}_2) & y & S_{22} = \frac{1}{n} \sum_{i=1}^n (Y_{i2} - \bar{Y}_2)^2.\end{aligned}$$

y

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y}_j = \frac{1}{n} \sum_{i=1}^n Y_{ij}, \quad j = 1, 2.$$

Para $p > 2$, los EMV de μ_x y α_j son, ver Shyr (1984),

$$\hat{\mu}_x = \bar{X}, \quad \hat{\alpha}_j = \bar{Y}_j - \hat{\beta}_j\bar{X}, \quad j = 1, \dots, p.$$

Los EMV de los restantes parámetros satisfacen las siguientes ecuaciones:

$$\hat{\mathbf{V}}(\mathbf{S} - \hat{\mathbf{V}})\hat{\mathbf{V}}^{-1} \begin{pmatrix} 1 \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \mathbf{0}, \quad (2.10)$$

$$\hat{\sigma}_0^2 = S_{00} - \hat{\sigma}_x^2, \quad \hat{\sigma}_j^2 = S_{jj} - \hat{\beta}_j^2\hat{\sigma}_x^2, \quad j = 1, \dots, p,$$

donde

$$\hat{\mathbf{V}} = \begin{pmatrix} 1 \\ \hat{\boldsymbol{\beta}} \end{pmatrix} \begin{pmatrix} 1 \\ \hat{\boldsymbol{\beta}} \end{pmatrix}^T \hat{\sigma}_x^2 + \text{Diag}(\hat{\sigma}_0^2, \dots, \hat{\sigma}_p^2).$$

Alternativamente podemos usar el algoritmo EM para encontrar los EMV de $\boldsymbol{\theta}$. Como es conocido, el algoritmo EM (Dempster et al., 1977; Little y Rubin, 1987) aumenta los datos observados $Z = (Z_1, \dots, Z_n)$ con datos hipotéticos o no observados $x = (x_1, x_2, \dots, x_n)$ de tal forma que el EMV de $\boldsymbol{\theta}$ basado en $Z^c = (Z, x)$ sea fácil de calcular. Dada una

estimativa $\boldsymbol{\theta}^{(m)}$, en la iteración m , la iteración $m + 1$ del EM consiste de dos etapas; una etapa E de esperanza y una etapa M de maximización. En la etapa E calculamos el valor esperado de la función de log-verosimilitud completa $\ell(\boldsymbol{\theta}|Z^c)$ con respecto a la distribución condicional de x dado Z y $\boldsymbol{\theta}^{(m)}$. En la etapa M maximizamos la función resultante con respecto a $\boldsymbol{\theta}$, y obtenemos así una nueva estimativa $\boldsymbol{\theta}^{(m+1)}$. Cada iteración del algoritmo EM incrementa la función de log-verosimilitud observada $\ell(\boldsymbol{\theta}|Z)$; $\ell(\boldsymbol{\theta}^{(m)}|Z) \leq \ell(\boldsymbol{\theta}^{(m+1)}|Z)$.

En efecto, sea $Z_i^c = (x_i, \mathbf{Z}_i)^T$, con $i = 1, \dots, n$ el conjunto de datos completos, los no observados (x_i) y los observados (\mathbf{Z}_i). Luego tenemos que

$$Z_i^c \sim N_{r+1}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z), \quad i = 1, \dots, n, \quad (2.11)$$

independientes, donde

$$\boldsymbol{\mu}_z = \begin{pmatrix} \mu_x \\ \mu_x \\ \boldsymbol{\alpha} + \boldsymbol{\beta}\mu_x \end{pmatrix} \quad \text{y} \quad \boldsymbol{\Sigma}_z = \sigma_x^2 \begin{pmatrix} 1 & \mathbf{b}^T \\ \mathbf{b} & \mathbf{b}\mathbf{b}^T + \mathbf{B} \end{pmatrix}$$

con

$$\mathbf{b} = \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix} \quad \text{y} \quad \mathbf{B} = \text{Diag}\left(\frac{\sigma_0^2}{\sigma_x^2}, \dots, \frac{\sigma_p^2}{\sigma_x^2}\right).$$

De (2.11) sigue que la función de log-verosimilitud completa está dada por,

$$l_c(\boldsymbol{\theta}) = \sum_{i=1}^n l_{ci}(\boldsymbol{\theta}), \quad (2.12)$$

donde, $l_{ci}(\boldsymbol{\theta}) = c - \frac{1}{2} \log |\boldsymbol{\Sigma}_z| - \frac{1}{2} \delta_{ci}(\boldsymbol{\theta})$, con $\delta_{ci}(\boldsymbol{\theta}) = (Z_i^c - \boldsymbol{\mu}_z)^T \boldsymbol{\Sigma}_z^{-1} (Z_i^c - \boldsymbol{\mu}_z)$.

De (2.11) tenemos que, ver Bolfarine y Galea (1995),

$$\hat{x}_i = E(x_i | \mathbf{Z}_i, \boldsymbol{\theta}) = \mu_x + \frac{1}{c} \left(\frac{\sigma_x^2}{\sigma_0^2} X_i - \mu_i(c-1) + \boldsymbol{\beta}^T \mathbf{B}_1^{-1} (\mathbf{Y}_i - \boldsymbol{\alpha}) \right), \quad (2.13)$$

y

$$\hat{x}_i^2 = E(x_i^2 | \mathbf{Z}_i, \boldsymbol{\theta}) = \hat{x}_i^2 + \frac{\sigma_x^2}{c}, \quad i = 1, \dots, n,$$

donde, $\mathbf{B}_1 = \text{Diag}\left(\frac{\sigma_1^2}{\sigma_x^2}, \dots, \frac{\sigma_p^2}{\sigma_x^2}\right)$, una matriz $p \times p$.

Maximizando (2.12) nos queda

$$\hat{\mu}_x = \bar{x}, \quad \hat{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (2.14)$$

$$\hat{\alpha}_j = \bar{Y}_j - \hat{\beta}_j \bar{x}, \quad \hat{\beta}_j = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_{ij} - \bar{Y}_j)}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad j = 1, \dots, p$$

y,

$$\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \hat{\sigma}_j^2 = \frac{1}{n} \sum_{i=1}^n (Y_{ij} - \bar{Y}_j - \hat{\beta}_j(x_i - \bar{x}))^2, \quad j = 1, \dots, p.$$

Note que no necesitamos de un procedimiento iterativo para obtener el máximo de la función de log-verosimilitud completa. La implementación del EM es simple a pesar de que para alcanzar la convergencia se requiere de varias iteraciones.

En efecto las etapas del algoritmo, ver Bolfarine y Galea (1995), son:

Etapla E: Dado $\boldsymbol{\theta}^{(m)}$, en la iteración m , calcule \hat{x}_i y \hat{x}_i^2 usando (2.13), $i = 1, \dots, n$.

Etapla M: Usando (2.14) obtenga un nuevo estimador $\boldsymbol{\theta}^{(m+1)}$. Repita el proceso para $m = 0, 1, \dots$

Existen varios criterios de convergencia utilizados en la literatura. Aquí citamos tres de ellos. Sea $\boldsymbol{\theta}^{(m)}$ la estimativa de $\boldsymbol{\theta}$ en la m -ésima iteración, $m = 0, 1, 2, \dots$

Como criterio de convergencia del algoritmo EM podemos usar:

$$\|\boldsymbol{\theta}^{(m)} - \boldsymbol{\theta}^{(m-1)}\| < \varepsilon, \quad (2.15)$$

donde $\|a\|$ denota la norma del vector a y $\varepsilon > 0$, o alternativamente

$$\frac{\|\boldsymbol{\theta}^{(m)} - \boldsymbol{\theta}^{(m-1)}\|}{\|\boldsymbol{\theta}^{(m)}\|} < \varepsilon. \quad (2.16)$$

Finalmente, si $\ell(\boldsymbol{\theta}|Z)$ es la función de log-verosimilitud observada, podemos decir que el algoritmo EM converge en la m -ésima iteración si

$$\ell(\boldsymbol{\theta}^{(m)}|Z) - \ell(\boldsymbol{\theta}^{(m-1)}|Z) < \varepsilon. \quad (2.17)$$

Para iniciar el algoritmo EM, o sea, para asignar un valor a $\boldsymbol{\theta}^{(0)}$, podemos usar, por ejemplo, el método de estimación propuesto por Barnett (1969), basado en que para $p = 2$ se obtienen soluciones explícitas, ver ecuación (2.9).

La matriz de información esperada toma la forma:

$$\mathbf{I}_F(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_{\mu_x \mu_x} & \mathbf{I}_{\mu_x \alpha} & \mathbf{I}_{\mu_x \beta} & \mathbf{0} \\ \mathbf{I}_{\alpha \mu_x} & \mathbf{I}_{\alpha \alpha} & \mathbf{I}_{\alpha \beta} & \mathbf{0} \\ \mathbf{I}_{\beta \mu_x} & \mathbf{I}_{\beta \alpha} & \mathbf{I}_{\beta \beta} & \mathbf{I}_{\beta \phi} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{\phi \beta} & \mathbf{I}_{\phi \phi} \end{pmatrix},$$

donde $\mathbf{I}_{\mu_x \mu_x} = [(I_{\mu_x \alpha_k})]$, una matriz con elementos $I_{\mu_x \alpha_k}$, $\mathbf{0}$, una matriz nula, y así sucesivamente. Los elementos de estas matrices son presentados en el apéndice *C*. La inversa de la matriz de información, $\mathbf{I}_F^{-1}(\boldsymbol{\theta})$, la denotamos como:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_{\mu_x \mu_x} & \boldsymbol{\Sigma}_{\mu_x \alpha} & \boldsymbol{\Sigma}_{\mu_x \beta} & \boldsymbol{\Sigma}_{\mu_x \phi} \\ \boldsymbol{\Sigma}_{\alpha \mu_x} & \boldsymbol{\Sigma}_{\alpha \alpha} & \boldsymbol{\Sigma}_{\alpha \beta} & \boldsymbol{\Sigma}_{\alpha \phi} \\ \boldsymbol{\Sigma}_{\beta \mu_x} & \boldsymbol{\Sigma}_{\beta \alpha} & \boldsymbol{\Sigma}_{\beta \beta} & \boldsymbol{\Sigma}_{\beta \phi} \\ \boldsymbol{\Sigma}_{\phi \mu_x} & \boldsymbol{\Sigma}_{\phi \alpha} & \boldsymbol{\Sigma}_{\phi \beta} & \boldsymbol{\Sigma}_{\phi \phi} \end{pmatrix}.$$

Finalmente, sabemos que $\hat{\boldsymbol{\theta}}$ se distribuye asintóticamente normal con vector de media $\boldsymbol{\theta}$ y matriz de covarianza $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, la cual estimamos por $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$. Este resultado puede ser usado para construir intervalos de confianza asintóticos para los parámetros de interés y/o para funciones lineales de la forma $\mathbf{a}^T \boldsymbol{\theta}$. También, usando el test de Wald, podemos probar hipótesis lineales del tipo $H_0 : \mathbf{A}\boldsymbol{\theta} = \mathbf{a}$.

2.4. Bondad de ajuste

Al igual que el capítulo 1, para evaluar el ajuste del modelo estructural normal multivariado, podemos construir el gráfico de las distancias transformadas, ver Little (1988) y Lange et al. (1989). Sea

$$\delta_i(\boldsymbol{\theta}) = (\mathbf{Z}_i - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{Z}_i - \boldsymbol{\mu}), \quad i = 1, 2, \dots, n. \quad (2.18)$$

la distancia de Mahalanobis. Bajo el supuesto de normalidad, sabemos que $\delta_i(\boldsymbol{\theta})$ tiene una distribución χ^2 con r grados de libertad, $i = 1, 2, \dots, n$. Además, $\delta_i(\hat{\boldsymbol{\theta}})$ tiene, asintóticamente, la misma distribución χ^2 , $i = 1, 2, \dots, n$. Usando la aproximación de Wilson-Hilferty, ver Johnson y Kotz (1970), tenemos que

$$z_i = \frac{\left\{ (\delta_i(\hat{\boldsymbol{\theta}})/r)^{1/3} - \left(1 - \frac{2}{9r}\right) \right\}}{\{2/9r\}^{1/2}}, \quad (2.19)$$

$i = 1, 2, \dots, n$, se distribuye aproximadamente $N(0, 1)$. Un $Q - Q$ plot de las distancias transformadas z_i , $i = 1, 2, \dots, n$, puede ser utilizado para evaluar el ajuste del modelo estructural normal multivariado.

2.5. Comparación de instrumentos de medición

Al igual que el simple con error de medición, el MEM multivariado ha sido utilizado, ver Barnett (1969), Theobald y Mallison (1978), Shyr y Gleser (1986) y Fuller (1987),

para comparar instrumentos de medición. En efecto, supongamos que una característica x es medida por varios métodos en un grupo común de n unidades experimentales. Como en Barnett (1969) suponemos que uno de los instrumentos es estándar, que mide x con un error de medición. Es decir, para modelar esta situación experimental usamos el modelo heterocedástico dado por

$$\mathbf{Z}_i = \begin{pmatrix} 0 \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix} x_i + \boldsymbol{\epsilon}_i, \quad (2.20)$$

con

$$x_i \sim N(\mu_x, \sigma_x^2) \quad \text{independientes de} \quad \boldsymbol{\epsilon}_i \sim \mathbf{N}_r(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon),$$

$i = 1, \dots, n$, donde $\boldsymbol{\Sigma}_\epsilon = \text{Diag}(\sigma_0^2, \sigma_1^2, \dots, \sigma_p^2)$. En este contexto $\boldsymbol{\alpha}$ es el vector de sesgos aditivos y $\boldsymbol{\beta}$ el vector de sesgo multiplicativo. La confiabilidad del j -ésimo instrumento está dada por,

$$\kappa_j = \beta_j^2 \sigma_x^2 / (\beta_j^2 \sigma_x^2 + \sigma_j^2), \quad j = 0, 1, \dots, p, \quad (2.21)$$

donde $\beta_0 = 1$.

Como antes sea $\boldsymbol{\theta} = (\mu_x, \boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\phi}^T)^T$, con $\boldsymbol{\phi} = (\sigma_x^2, \sigma_0^2, \dots, \sigma_p^2)^T$ y considere la transformación

$$\boldsymbol{\theta}^* = g(\boldsymbol{\theta}) = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \boldsymbol{\kappa}^T)^T = (g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}), \dots, g_{3r-2}(\boldsymbol{\theta}))^T$$

Entonces tenemos que

$$g(\hat{\boldsymbol{\theta}}) \quad \text{es} \quad AN(g(\boldsymbol{\theta}), \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^T)$$

donde

$$\mathbf{D} = \left[\left(\frac{\partial g_i(\boldsymbol{\theta})}{\partial \theta_j} \right) \right] = \begin{pmatrix} \mathbf{0}_{px1} & \mathbf{I}_p & \mathbf{0}_{pxp} & \mathbf{0}_{px(r+1)} \\ \mathbf{0}_{px1} & \mathbf{0}_{pxp} & \mathbf{I}_p & \mathbf{0}_{px(r+1)} \\ \mathbf{0}_{rx1} & \mathbf{0}_{rxp} & \mathbf{D}_1 & \mathbf{D}_2 \end{pmatrix},$$

con las matrices \mathbf{D}_1 , $r \times p$, y \mathbf{D}_2 , $r \times (r+1)$ dadas por

$$\mathbf{D}_1 = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\Psi}_1 \end{pmatrix} \quad \text{y} \quad \mathbf{D}_2 = (\mathbf{d} \quad \boldsymbol{\Psi}_2),$$

donde $\boldsymbol{\Psi}_1 = 2\text{Diag}(\beta_1^{-1}\kappa_1(1-\kappa_1), \dots, \beta_p^{-1}\kappa_p(1-\kappa_p))$, $\boldsymbol{\Psi}_2 = -\sigma_x^{-2}\text{Diag}(\kappa_0^2, \frac{\kappa_1^2}{\beta_1^2}, \dots, \frac{\kappa_p^2}{\beta_p^2})$ y $\mathbf{d} = -\sigma_x^{-2}(\kappa_0(1-\kappa_0), \dots, \kappa_p(1-\kappa_p))^T$, un vector $r \times 1$.

Luego, podemos escribir

$$\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^T = \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{\alpha\alpha} & \boldsymbol{\Omega}_{\alpha\beta} & \boldsymbol{\Omega}_{\alpha\kappa} \\ \boldsymbol{\Omega}_{\beta\alpha} & \boldsymbol{\Omega}_{\beta\beta} & \boldsymbol{\Omega}_{\beta\kappa} \\ \boldsymbol{\Omega}_{\kappa\alpha} & \boldsymbol{\Omega}_{\kappa\beta} & \boldsymbol{\Omega}_{\kappa\kappa} \end{pmatrix},$$

donde $\mathbf{\Omega}_{\alpha\alpha} = \mathbf{\Sigma}_{\alpha\alpha}$, $\mathbf{\Omega}_{\alpha\beta} = \mathbf{\Sigma}_{\alpha\beta}$, $\mathbf{\Omega}_{\beta\beta} = \mathbf{\Sigma}_{\beta\beta}$, $\mathbf{\Omega}_{\alpha\kappa} = \mathbf{\Sigma}_{\alpha\beta}\mathbf{D}_1^T + \mathbf{\Sigma}_{\alpha\phi}\mathbf{D}_2^T$, $\mathbf{\Omega}_{\beta\kappa} = \mathbf{\Sigma}_{\beta\beta}\mathbf{D}_1^T + \mathbf{\Sigma}_{\beta\phi}\mathbf{D}_2^T$, y

$$\mathbf{\Omega}_{\kappa\kappa} = \mathbf{D}_1\mathbf{\Sigma}_{\beta\beta}\mathbf{D}_1^T + \mathbf{D}_2\mathbf{\Sigma}_{\phi\phi}\mathbf{D}_2^T + \mathbf{D}_1\mathbf{\Sigma}_{\beta\phi}\mathbf{D}_2^T + \mathbf{D}_2\mathbf{\Sigma}_{\phi\beta}\mathbf{D}_1^T,$$

con $\boldsymbol{\kappa} = (\kappa_0, \dots, \kappa_p)^T$. En muchos problemas puede ser de interés probar la hipótesis nula $H_0 : \boldsymbol{\alpha} = \mathbf{0}, \boldsymbol{\beta} = \mathbf{1}, \kappa_0 = \kappa_1 = \dots = \kappa_p$, la cual podemos escribir como $H_0 : \mathbf{A}\boldsymbol{\theta}^* = \mathbf{a}$, donde

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_1 \end{pmatrix},$$

una matriz $3p \times (3r - 2)$ de rango $3p$, con

$$\mathbf{A}_1 = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 0 & 1 & -1 \end{pmatrix},$$

una matriz $p \times r$ y $\mathbf{a} = (0, \dots, 0, 1, \dots, 1, 0, \dots, 0)^T$ de dimensión $3p \times 1$. El estadístico de Wald para probar H_0 está dado por

$$W_a = (\mathbf{A}\hat{\boldsymbol{\theta}}^* - \mathbf{a})^T (\mathbf{A}\hat{\mathbf{\Omega}}\mathbf{A}^T)^{-1} (\mathbf{A}\hat{\boldsymbol{\theta}}^* - \mathbf{a}), \quad (2.22)$$

donde $\hat{\mathbf{\Omega}}$ denota el EMV de $\mathbf{\Omega}$. Así, H_0 es rechazada al nivel α si $W_a > \chi_{1-\alpha}^2$, donde $\chi_{1-\alpha}^2$ denota el percentil $100(1-\alpha)\%$ de la distribución chi cuadrado con $3p$ grados de libertad. Otra hipótesis de interés puede ser $H_0 : \boldsymbol{\alpha} = \mathbf{0}, \boldsymbol{\beta} = \mathbf{1}$, o equivalentemente $H_0 : \mathbf{A}\boldsymbol{\theta}^* = \mathbf{a}$, con \mathbf{a} de dimensión $2p \times 1$, donde

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p & \mathbf{0} \end{pmatrix},$$

una matriz de dimensión $2p \times (3r - 2)$, de rango $2p$. En este caso el estadístico de Wald está también dado por (2.22) el cual, en muestras grandes, se distribuye aproximadamente χ_{2p}^2 . Para comparar la confiabilidad de los instrumentos la hipótesis relevante puede ser $H_0 : \kappa_0 = \kappa_1 = \dots = \kappa_p$, o equivalentemente $H_0 : \mathbf{A}\boldsymbol{\theta}^* = \mathbf{0}$, con $\mathbf{0}$ de dimensión $p \times r$, con $\mathbf{A} = (\mathbf{0}, \mathbf{A}_1)$, una matriz $p \times (3r - 2)$ de rango p donde \mathbf{A}_1 como fue definida anteriormente. En este caso,

$$W_a = \hat{\boldsymbol{\theta}}^{*T} \mathbf{A}^T (\mathbf{A}\hat{\mathbf{\Omega}}\mathbf{A}^T)^{-1} \mathbf{A}\hat{\boldsymbol{\theta}}^*, \quad (2.23)$$

el cual, en muestras grandes, se distribuye aproximadamente χ_p^2 . Para más detalles sobre esta sección ver Bolfarine y Galea (1995).

Ilustración: Datos de Barnett Considere los siguientes datos, tomados de Barnett (1969), relativos a un estudio médico para evaluar la calidad de dos instrumentos usados para medir la capacidad vital pulmonar (x) en un grupo de pacientes. Un instrumento es de tipo estándar y otro nuevo, portátil y más fácil de operar. Además, el instrumento estándar requiere de un operador especializado. Los dos instrumentos fueron comparados en un grupo común de $n = 72$ pacientes, operados por un operador especializado y por uno no especializado. Las cuatro combinaciones instrumento-operador serán denominadas como (Theobald y Mallison, 1978):

Instrumento 0: Instrumento estándar y operador especializado;

Instrumento 1: Instrumento estándar y operador no especializado;

Instrumento 2: Instrumento nuevo y operador especializado;

Instrumento 3: Instrumento nuevo y operador no especializado.

Sea X_i la medida de la capacidad vital para el paciente i proporcionada por el instrumento 0, considerado estándar y sean Y_{ij} la medida de la capacidad vital obtenida por el instrumento j para el paciente i , con $j = 1, 2, 3$ y $i = 1, 2, \dots, 72$. El conjunto de observaciones \mathbf{Z}_i es presentado en la tabla (2.1)

Siguiendo a Barnett (1969) y Theobald y Mallison (1978), vamos a suponer un modelo estructural normal, con $p = 3$ y $n = 72$.

La Tabla 2.2 muestra varias iteraciones del algoritmo EM, para algunos estimadores, mostrando que la convergencia es alcanzada aproximadamente en la iteración 200.

Tabla 2.2: Convergencia del algoritmo EM

m	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}_x^2$	$\hat{\sigma}_0^2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_3^2$
50	1.075	1.212	1.149	517731.2	54174.1	20448.3	27561.4	37534.2
100	1.059	1.192	1.307	534003.7	50249.3	19151.6	29234.6	38842.2
200	1.059	1.192	1.131	534042.4	50248.1	19150.7	29235.3	38843.2
500	1.059	1.192	1.131	534042.4	50248.1	19150.7	29235.3	38843.2
2500	1.059	1.192	1.131	534043.4	50248.1	19150.7	29235.7	38843.2

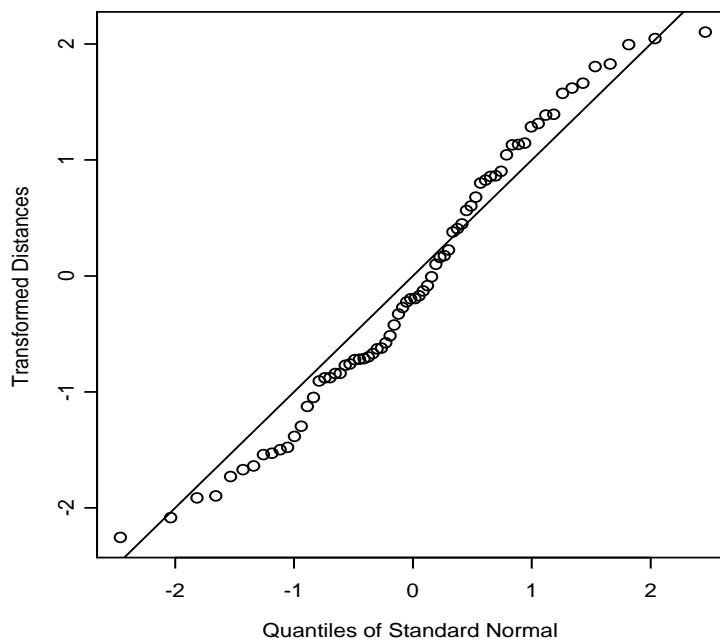
La Tabla 2.3 muestra las estimas de máxima verosimilitud y sus errores estándar asintóticos.

Tabla 2.1: Medidas de la capacidad vital para 72 pacientes, hechas por 4 instrumentos

Obs.	X	Y_1	Y_2	Y_3	Obs.	X	Y_1	Y_2	Y_3
1	3450	3530	4030	3720	37	1060	1000	850	600
2	1310	1320	1610	1600	38	2000	1800	1270	1700
3	3820	3720	4150	3700	39	2280	2280	2380	2350
4	2110	2880	2740	2520	40	1940	1800	1670	1580
5	1860	1420	1540	1690	41	2580	2700	2850	2110
6	1940	1780	2020	1800	42	1400	1440	1680	1480
7	2360	2260	2430	2350	43	1260	1100	1000	1030
8	2880	2920	2650	2860	44	2320	2420	2360	2360
9	1980	1720	1800	1660	45	2000	1940	1980	1980
10	3120	3180	3250	3040	46	2400	1900	1470	1740
11	1760	1630	1390	1200	47	2880	2980	3240	3140
12	1480	1760	1700	1640	48	3420	3150	3200	3200
13	1840	1660	1400	1650	49	1000	1130	650	840
14	3580	3480	3680	3960	50	1400	1400	1350	1380
15	1880	2000	2090	2070	51	1880	1710	1600	1350
16	2400	2320	2550	2480	52	1280	1260	1160	1330
17	2220	2120	2290	2270	53	3120	3000	3110	3250
18	2540	2500	2620	1960	54	3770	3340	3900	3700
19	920	1200	640	1030	55	3420	3220	3120	3290
20	2240	2160	2300	2300	56	2740	2880	2850	2880
21	2240	2130	2030	2140	57	2840	2920	2710	2750
22	2260	2510	2400	2450	58	3800	3740	3440	3400
23	3860	4180	3980	3680	59	2100	1680	1650	1930
24	2780	2100	1890	2000	60	1820	1400	1060	1050
25	2220	1400	1840	1360	61	1400	1320	1350	1100
26	1880	1820	1900	1840	62	2200	1680	1640	1110
27	940	960	1060	1000	63	1940	1900	1820	1270
28	2480	2220	2150	2150	64	3260	3200	3250	3270
29	1660	1780	1760	1800	65	1960	1940	1890	1920
30	4040	4180	4000	3770	66	1320	1260	1140	1000
31	2540	2560	2080	2250	67	2840	3060	3650	3510
32	1780	1700	1390	1200	68	2060	1840	1720	1780
33	1280	1300	800	1130	69	2200	1970	1900	2270
34	1940	2060	2030	1880	70	1260	1150	860	1150
35	1760	2000	1860	1860	71	3040	2840	2850	2670
36	2040	1660	1470	1160	72	2140	2180	2560	2720

Fuente: Barnett (1969).

Figura 2.1: Gráfico de la distancia transformada para el modelo normal



Capítulo 3

Modelo Estructural t

3.1. Introducción

En los capítulos anteriores hemos considerado modelos estructurales normales, es decir, donde tanto el predictor como la respuesta siguen una distribución normal multivariada. Como método de estimación usamos máxima verosimilitud.

Ciertamente la distribución normal es útil en muchos casos y gran parte de la inferencia estadística para variables continuas ha sido desarrollada bajo el supuesto de normalidad. Esto es particularmente el caso de análisis multivariado y regresión lineal.

Sin embargo, es conocido que la inferencia basada en la distribución normal es vulnerable a la presencia de “outliers” en los datos, que provienen de distribuciones con colas más pesadas que la normal. O sea, es un hecho reconocido que los outliers tienen una gran influencia sobre los estimadores y test de hipótesis basados en máxima verosimilitud.

Esto ha provocado innumerables estudios sobre procedimientos o métodos para el análisis de datos que contienen outliers o observaciones extremas. Es así como en la actualidad existen en diversas áreas de la estadística una variedad de métodos para analizar datos en presencia de outliers.

Un enfoque es a través de la identificación y eliminación de outliers seguido de inferencia estadística (máxima verosimilitud) basada en la distribución normal. De hecho, una amplia área de investigación es la detección de datos atípicos, especialmente en modelos lineales y análisis multivariado. Algunas referencias básicas son Barnett y Lewis (1994), Belsley et al. (1980), Chatterjee y Hadi (1988), Cook y Weisberg (1982), Hawkins (1980) y Peña y Yohai (1995). En el área de modelos de regresión con errores en las variables dos trabajos pioneros son Kelly (1984), Wellman y Gunst (1991) y Abdullah (1995). Una referencia básica para influencia local es Cook (1986).

Otro enfoque es la acomodación de observaciones aberrantes usando procedimientos robustos clásicos. En esta dirección, han sido desarrollados métodos robustos, de modo

que los “outliers” tengan menos influencia sobre las estimativas finales; ver Huber (1981), Hampel et al. (1986), Rousseeau y Leroy (1987), Staudte y Sheather (1990) y Maronna et al. (2006). En el área de modelos de regresión con errores en las variables algunas referencias son Kelly (1984), Zamar (1989) y Cheng y Van Ness (1990).

Una tercera posibilidad es usar distribuciones simétricas, con colas más pesadas que la normal, las cuales permiten reducir la influencia de los outliers sobre los EMV. Una clase de distribuciones que ofrece varias alternativas en este sentido es la clase de distribuciones elípticas. Esta familia de distribuciones, que contiene las distribuciones normal, t y normal contaminada multivariadas, entre otras, ha tenido un interés creciente en los últimos años; ver por ejemplo, Fang y Zhang (1990), Fang et al. (1990), Fang y Anderson (1990), Kano et al. (1993), Berkane et al. (1994) y Arellano-Valle (1994). Dentro de esta clase una de las distribuciones más utilizadas como alternativa a la normal es la distribución t .

En los años recientes, varios autores han sugerido la distribución t multivariada como alternativa a la normal multivariada por tener las colas más pesadas y, por lo tanto, “acomoda” posibles outliers presentes en los datos. Así, el uso de la distribución t proporcionaría un procedimiento robusto de análisis de datos. Por ejemplo, Rubin (1983) obtiene los EMV para los parámetros de la distribución t multivariada usando el algoritmo EM; Little (1988) extiende los resultados de Rubin (1983) a datos incompletos, Sutradhar y Ali (1986) usan máxima verosimilitud en el modelo de regresión multivariado con errores distribuidos t . Lange et al. (1989) ilustran el uso de la distribución t en regresión y en algunas áreas de análisis multivariado, ver también Taylor (1992). Sutradhar (1993) propone un test score para probar que la matriz de covarianza es igual a una matriz dada, usando también la distribución t ; Bolfarine y Arellano (1994) usan la distribución t en el modelo lineal simple con errores en las variables y Bolfarine y Galea (1996) usan esta distribución en modelos de calibración comparativa.

En este capítulo, seguimos el enfoque propuesto por Lange et al. (1989) y extendemos el modelo estructural normal simple del capítulo 1, suponiendo que las observaciones siguen una distribución t bivariada.

3.2. La distribución t multivariada

En esta sección presentamos algunas propiedades de la distribución t multivariada.

Definición 3.1 Diremos que un vector aleatorio $\mathbf{Z} = (Z_1, \dots, Z_p)^T$ p -dimensional tiene una distribución t multivariada, con vector de posición $\boldsymbol{\mu} \in \mathbb{R}^p$, matriz de escala $\boldsymbol{\Lambda}$, $p \times p$, definida positiva y ν grados de libertad, si su densidad está dada:

$$f_{\mathbf{Z}}(\mathbf{z}) = K(\nu, p) (1 + (\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\mathbf{z} - \boldsymbol{\mu}) / \nu)^{-\frac{1}{2}(\nu+p)}, \quad \mathbf{z} \in \mathbb{R}^p; \quad (3.1)$$

donde

$$K(\nu, p) = |\Lambda|^{-\frac{1}{2}} \Gamma\left(\frac{\nu+p}{2}\right) / \Gamma\left(\frac{1}{2}\right)^p \Gamma\left(\frac{\nu}{2}\right) \nu^{\frac{p}{2}}, \quad \nu > 0.$$

Si un vector aleatorio \mathbf{Z} tiene una densidad dada por (3.1) lo denotaremos por $\mathbf{Z} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$.

Recientemente, esta distribución t , introducida por Dunnett y Sobel (1954) y Bennett (1961), ha sido bastante utilizada, para ν pequeño, en análisis multivariada debido a sus colas pesadas y, consecuentemente, los EMV de $\boldsymbol{\mu}$ y $\boldsymbol{\Lambda}$ son resistentes a outliers. Ver Maronna (1976), Rubin (1983), Little (1988), Lange et al. (1989), Berkane y Bentler (1994), Kent et al. (1994).

Propiedad 3.1 Sea $\mathbf{Z} = \boldsymbol{\mu} + \mathbf{W}/\sqrt{U}$, donde $\mathbf{W} \sim N_p(\mathbf{0}, \boldsymbol{\Lambda})$ independiente de $U \sim \chi^2(\nu)/\nu$. Entonces $\mathbf{Z} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$.

Note que dado $U = u$, tenemos que

$$(\mathbf{Z}|U = u) \sim N_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Lambda}). \quad (3.2)$$

De aquí es fácil ver que la densidad de \mathbf{Z} es (3.1). Es decir,

$$f_{\mathbf{Z}}(\mathbf{z}) = \int_0^\infty |\boldsymbol{\Lambda}|^{-1/2} (2\pi u^{-1})^{-p/2} e^{-\frac{u}{2}(\mathbf{z}-\boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1}(\mathbf{z}-\boldsymbol{\mu})} f_U(u) du,$$

donde $f_U(u)$ es densidad de U . O sea, la distribución t multivariada es una mezcla de una distribución normal multivariada y una distribución χ^2 .

Propiedad 3.2 Sea $(\mathbf{Z}|U = u) \sim N_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Lambda})$, donde $U \sim \chi^2(\nu)/\nu$, $\nu > 0$. Entonces tenemos que

- (i) $\mathbf{Z} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$;
- (ii) $(U|\mathbf{Z} = \mathbf{z}) \sim \chi^2(\nu + p)/(\nu + \delta)$, donde $\delta = (\mathbf{Z} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1}(\mathbf{Z} - \boldsymbol{\mu})$;
- (iii) $\delta/p \sim F(p, \nu)$.

Note que esta propiedad sigue directamente de la propiedad (3.1).

Propiedad 3.3 Sea $\mathbf{Z} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$. Entonces,

$$\begin{aligned} E(\mathbf{Z}) &= \boldsymbol{\mu} && \text{si } \nu > 1 \quad \text{y} \\ \text{Var}(\mathbf{Z}) &= (\nu/(\nu - 2))\boldsymbol{\Lambda} && \text{si } \nu > 2. \end{aligned}$$

Usando la propiedad (3.1), tenemos que

$$E(\mathbf{Z}) = \boldsymbol{\mu} + E(1/\sqrt{U})E(\mathbf{W}) \quad \text{y} \quad \text{Var}(\mathbf{Z}) = E(U^{-1})E(\mathbf{W}\mathbf{W}^T)$$

y el resultado sigue, ya que $E(1/\sqrt{U}) = (\frac{\nu}{2})^{1/2}\Gamma(\frac{\nu-1}{2})/\Gamma(\frac{\nu}{2})$ si $\nu > 1$ y $E(U^{-1}) = (\nu/(\nu-2))$ si $\nu > 2$.

También, usando la propiedad (3.1), tenemos:

Propiedad 3.4 Sea $\mathbf{Z} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$ y $\mathbf{a} \in \mathbb{R}^q$, $q \leq p$, \mathbf{A} una matriz $q \times p$ de rango q y $\mathbf{X} = \mathbf{a} + \mathbf{AZ}$. Entonces

$$\mathbf{X} \sim t_q(\mathbf{a} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Lambda}\mathbf{A}^T, \nu) .$$

Particionemos \mathbf{Z} , $\boldsymbol{\mu}$ y $\boldsymbol{\Lambda}$ como

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{y} \quad \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix},$$

donde \mathbf{Z}_1 y $\boldsymbol{\mu}_1$ son $q \times 1$, $q \leq p$ y $\boldsymbol{\Lambda}_{11}$ es $q \times q$.

Propiedad 3.5 Sea $\mathbf{Z} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$. Entonces, tenemos que

$$\mathbf{Z}_1 \sim t_q(\boldsymbol{\mu}_1, \boldsymbol{\Lambda}_{11}, \nu).$$

Note que esta propiedad sigue de la propiedad (3.4).

Propiedad 3.6 Sea $\mathbf{Z} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$. Entonces

$$(\mathbf{Z}_1 | \mathbf{Z}_2 = \mathbf{z}_2) \sim t_q(\boldsymbol{\mu}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}(\mathbf{z}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Lambda}_{1,2}, \nu + p - q) ,$$

donde $\boldsymbol{\Lambda}_{1,2} = \boldsymbol{\Lambda}_{11} - \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{\Lambda}_{21}$. Ver Arellano-Valle (1994).

De esta propiedad sigue que

$$E(\mathbf{Z}_1 | \mathbf{Z}_2 = \mathbf{z}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}(\mathbf{z}_2 - \boldsymbol{\mu}_2) \quad \text{si } \nu + p - q > 1 \quad (3.3)$$

y

$$\text{Var}(\mathbf{Z}_1 | \mathbf{Z}_2 = \mathbf{z}_2) = \left[\frac{\nu + (\mathbf{z}_2 - \boldsymbol{\mu}_2)^T \boldsymbol{\Lambda}_{22}^{-1} (\mathbf{z}_2 - \boldsymbol{\mu}_2)}{\nu + p - q - 2} \right] \boldsymbol{\Lambda}_{1,2}, \quad (3.4)$$

si $\nu + p - q > 2$.

Mardia (1970) define el coeficiente de curtosis multivariado, $\beta_{2,p}$, por

$$\beta_{2,p} = E\{(\mathbf{Z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\mu})\}^2, \quad (3.5)$$

donde $\boldsymbol{\Sigma}$ es la matriz de covarianza de \mathbf{Z} . En el caso elíptico tenemos, ver Berkane y Bentler (1987), que

$$\kappa + 1 = \beta_{2,p}/p(p+2) \quad (3.6)$$

donde κ es denominado parámetro de curtosis.

Propiedad 3.7 Sea $\mathbf{Z} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$. Entonces el coeficiente de curtosis multivariado está dado

$$\beta_{2,p} = p(p+2)(\nu-2)/(\nu-4) \text{ para } \nu > 4.$$

Usando (3.5) y la propiedad (3.1), tenemos que

$$\beta_{2,p} = ((\nu-2)/\nu)^2 E(U^{-2}) E[(\mathbf{W}^T \boldsymbol{\Lambda}^{-1} \mathbf{W})^2]$$

y el resultado sigue, ya que $E(U^{-2}) = \nu^2/(\nu-2)(\nu-4)$, $\nu > 4$ y como $\mathbf{W} \sim N_p(\mathbf{0}, \boldsymbol{\Lambda})$, $E[(\mathbf{W}^T \boldsymbol{\Lambda}^{-1} \mathbf{W})^2] = p(p+2)$, ver Graybill (1983).

Ahora, como $\kappa + 1 = \beta_{2,p}/p(p+2)$, ver (3.6), sigue que

$$\kappa = 2/(\nu-4) \text{ si } \nu > 4. \quad (3.7)$$

Propiedad 3.8 Sea $\mathbf{Z} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$ con densidad $f_Z(\mathbf{z}, \nu)$ dada por (3.1) y sea $f_N(\mathbf{z})$ la función de densidad de un vector aleatorio con $\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$. Entonces,

$$\lim_{\nu \rightarrow \infty} f_Z(\mathbf{z}, \nu) = f_N(\mathbf{z}), \quad \forall \mathbf{z} \in \mathbb{R}^p.$$

Esta propiedad sigue directamente del hecho que (Tong, 1990):

$$\lim_{\nu \rightarrow \infty} \left(1 + \frac{1}{\nu} (\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\mathbf{z} - \boldsymbol{\mu})\right)^{-\frac{1}{2}(\nu+p)} = \exp \left\{ -\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right\}$$

y que

$$\lim_{\nu \rightarrow \infty} K(\nu, p) = 1/(2\pi)^{p/2} |\boldsymbol{\Lambda}|^{1/2}.$$

Este resultado muestra que la distribución normal multivariada puede ser obtenida a partir de la distribución t multivariada, haciendo $\nu \mapsto \infty$.

3.3. Modelo estructural t

Recordemos que el ME está dado por,

$$\begin{aligned} X_i &= x_i + \epsilon_{i0} \\ Y_i &= y_i + \epsilon_{i1} \\ y_i &= \alpha + \beta x_i, \quad i = 1, 2, \dots, n; \end{aligned} \quad (3.8)$$

donde Y_i y X_i corresponden a la i -ésima observación, cuyos verdaderos valores son y_i y x_i , respectivamente, y ϵ_{i0} y ϵ_{i1} son errores de medición. En notación matricial podemos escribir:

$$\mathbf{Z}_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} 0 \\ \alpha \end{pmatrix} + \begin{pmatrix} 1 \\ \beta \end{pmatrix} x_i + \boldsymbol{\epsilon}_i \quad i = 1, \dots, n. \quad (3.9)$$

Como en Bolfarine y Arellano-Valle (1994), suponemos que los vectores aleatorios no observables $(x_i, \epsilon_{i0}, \epsilon_{i1})^\top$ son iid como una distribución t trivariada con vector de posición $(\mu_x, 0, 0)^\top$ y matriz de escala diagonal $\text{diag}(\phi_x, \phi_0, \phi_1)$. Además suponemos, al igual que en el capítulo 1, que la razón de varianzas $\lambda = \phi_0/\phi_1$ es conocida, y sin pérdida de generalidad podemos tomar $\phi_0 = \phi_1 = \phi$, ($\lambda = 1$). Luego la distribución de $\mathbf{Z}_i = (X_i, Y_i)^\top$ es una distribución t bivariada con vector de posición $\boldsymbol{\mu} = (\mu_x, \alpha + \beta\mu_x)^\top$ y matriz de escala

$$\mathbf{V} = \begin{pmatrix} \phi_x + \phi & \beta\phi_x \\ \beta\phi_x & \beta^2\phi_x + \phi \end{pmatrix}, \quad (3.10)$$

es decir, $\mathbf{Z}_i \stackrel{\text{iid}}{\sim} t_2(\boldsymbol{\mu}, \mathbf{V}; \nu)$, $\nu > 0$, $i = 1, \dots, n$. La función de densidad de \mathbf{Z}_i esta dada por:

$$f(\mathbf{z}_i; \boldsymbol{\theta}) = \frac{1}{2\pi} |\mathbf{V}|^{-1/2} \left(1 + \frac{1}{\nu} \delta_i(\boldsymbol{\theta})\right)^{-\frac{1}{2}(\nu+2)}, \quad (3.11)$$

donde $\delta_i(\boldsymbol{\theta}) = (\mathbf{z}_i - \boldsymbol{\mu})^\top \mathbf{V}^{-1} (\mathbf{z}_i - \boldsymbol{\mu})$ y $\boldsymbol{\theta} = (\alpha, \beta, \phi, \mu_x, \phi_x)^\top$. La distribución t incorpora un parámetro adicional, ν , denominado grados de libertad, el cual permite ajustar la curtosis de la distribución. Este parámetro puede ser fijado a priori y Lange et al. (1989) recomiendan tomar $\nu = 4$. Alternativamente, puede ser estimado desde los datos. Para algunas dificultades en la estimación de ν , en el contexto de regresión, ver Fernández y Steel (1999).

Para obtener el EMV del vector $\boldsymbol{\theta}$, usamos el algoritmo EM, ver Bolfarine y Arellano-Valle (1994). En efecto las etapas del algoritmo son,

Etapla E: Dado $\boldsymbol{\theta}^{(m)}$, en la iteración m , calcule $v_i = (\nu + 2)/(\nu + \delta_i(\boldsymbol{\theta}^{(m)}))$ $i = 1, \dots, n$.

Etapla M: Obtenga un nuevo estimador $\boldsymbol{\theta}^{(m+1)}$, usando las ecuaciones,

$$\mu_x^{(m+1)} = \bar{X}(v), \quad \beta^{(m+1)} = \frac{-(S_{XX}(v) - S_{YY}(v)) + \sqrt{(S_{XX}(v) - S_{YY}(v))^2 + 4(S_{XY}(v))^2}}{2S_{XY}(v)} \quad (3.12)$$

y

$$\alpha^{(m+1)} = \bar{Y}(v) - \beta^{(m+1)} \bar{X}(v), \quad \phi^{(m+1)} = \frac{1}{n} (S_{YY}(v) - \beta^{(m+1)} S_{XY}(v)), \quad \phi_x^{(m+1)} = \frac{S_{XY}(v)}{n\beta^{(m+1)}},$$

donde,

$$S_{XX}(v) = \frac{1}{n} \sum_{i=1}^n v_i (X_i - \bar{X}(v))^2, \quad S_{XY}(v) = \frac{1}{n} \sum_{i=1}^n v_i (X_i - \bar{X}(v))(Y_i - \bar{Y}(v)),$$

$$S_{YY}(v) = \frac{1}{n} \sum_{i=1}^n v_i (Y_i - \bar{Y}(v))^2, \quad \bar{X}(v) = \sum_{i=1}^n v_i X_i / \sum_{i=1}^n v_i, \quad \bar{Y}(v) = \sum_{i=1}^n v_i Y_i / \sum_{i=1}^n v_i.$$

Repita el proceso para $m = 0, 1, \dots$

Para estimar la varianza de los EMV podemos usar la inversa de la matriz de información esperada. Como discutido en Arellano-Valle y Bolfarine (1996b), el cálculo de la matriz de información es más simple si usamos la siguiente parametrización ortogonal, en el sentido de Cox y Reid (1987), transformando $\boldsymbol{\theta}$ en $\boldsymbol{\gamma} = (\mu_1, \mu_2, \phi_1, \phi_2, \phi_3)^T$, donde,

$$\mu_1 = \alpha + \beta\mu_x, \quad \mu_2 = \mu_x, \quad \phi_1 = (1 + \beta^2)\phi_x + \phi, \quad \phi_2 = \phi, \quad \text{y} \quad \phi_3 = \beta. \quad (3.13)$$

Con esta parametrización tenemos que,

$$\mathbf{Z}_i \text{ iid } t_2(\boldsymbol{\mu}, \mathbf{V}, \nu), \quad i = 1, \dots, n, \quad (3.14)$$

donde

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{y} \quad \mathbf{V} = (1 + \beta^2)^{-1} \begin{pmatrix} \phi_1 + \beta^2\phi_2 & \beta(\phi_1 - \phi_2) \\ \beta(\phi_1 - \phi_2) & \phi_2 + \beta^2\phi_1 \end{pmatrix}. \quad (3.15)$$

Usando la expresión dada por Lange et al. (1989), Arellano-Valle y Bolfarine (1996b) muestran que la matriz de información esperada para el modelo estructural t bajo la parametrización ortogonal toma la forma,

$$\mathbf{K} = \text{Diag}(\mathbf{K}_\mu, \mathbf{K}_\phi, k_\beta), \quad (3.16)$$

donde

$$\mathbf{K}_\mu = n \left(\frac{\nu+2}{\nu+4} \right) \mathbf{V}^{-1}, \quad \mathbf{K}_\phi = \begin{pmatrix} \left(\frac{\nu+1}{\nu+4} \right) \left(\frac{2\phi_1^2}{n} \right)^{-1} & - \left(\frac{1}{\nu+4} \right) \left(\frac{2\phi_1\phi_2}{n} \right)^{-1} \\ - \left(\frac{1}{\nu+4} \right) \left(\frac{2\phi_1\phi_2}{n} \right)^{-1} & \left(\frac{\nu+1}{\nu+4} \right) \left(\frac{2\phi_2^2}{n} \right)^{-1} \end{pmatrix},$$

y $k_\beta = \left(\frac{\nu+2}{\nu+4} \right) \left(\frac{\sigma_\beta^2}{n} \right)^{-1}$, con $\sigma_\beta^2 = \left(\frac{1+\beta^2}{\phi_1-\phi_2} \right)^2 \phi_1\phi_2$. Luego, para n grande, $\hat{\boldsymbol{\gamma}} \sim AN_5(\boldsymbol{\gamma}, \mathbf{K}^{-1})$. En particular, $\hat{\beta} \sim AN\left(\beta, \frac{1}{n} \left(\frac{\nu+4}{\nu+2} \right) \sigma_\beta^2\right)$.

3.4. Test de hipótesis

En esta sección estamos interesados en probar la hipótesis $H_0 : \beta = \beta_0$, con β_0 conocido. Usando la parametrización ortogonal (3.13), Arellano-Valle y Bolfarine (1996b) muestran que el test de la razón de verosimilitud para probar H_0 , con ν conocido, esta dado por,

$$G_t = 2\{l(\hat{\boldsymbol{\gamma}}) - l(\hat{\boldsymbol{\gamma}}_0)\} = n\left\{ \log\left(\frac{\tilde{\phi}_1 \tilde{\phi}_2}{\hat{\phi}_1 \hat{\phi}_2} \right) - (\nu+2) \log\left(\frac{\tilde{m}_g}{\hat{m}_g} \right) \right\}, \quad (3.17)$$

donde, $\tilde{\phi}_j$ denota el EMV de ϕ_j bajo H_0 , $j = 1, 2$, y $m_g = (\prod_{i=1}^n v_i)^{1/n}$, $v_i = (\nu + 2)/(\nu + \delta_i(\gamma))$, $i = 1, \dots, n$; \tilde{m}_g y \hat{m}_g denotan los EMV de m_g bajo H_0 y sin restricciones, respectivamente. El estadístico (3.17) tiene bajo H_0 , para n grande, aproximadamente una distribución chi cuadrado con 1 grado de libertad. Como es bien conocido, la performance de este estadístico puede ser mejorada si consideramos la modificación $(1+b)^{-1}G_t$, donde b es un factor de corrección de Bartlett. En efecto, Arellano-Valle y Bolfarine (1996b) muestran que el TRV corregido toma la forma,

$$G_t^* = (1 + b_\nu)^{-1}G_t, \quad (3.18)$$

con $b_\nu = b_\infty(1 + o_\nu)$, donde $b_\infty = 5/2n$ es el factor de corrección del modelo normal, y

$$o_\nu = -\frac{4\nu^3 + 25\nu^2 + 88\nu - 24}{5(\nu + 2)(\nu + 6)^2(\nu + 8)}. \quad (3.19)$$

Note que $b_\nu \rightarrow b_\infty$ cuando $\nu \rightarrow \infty$.

3.5. Bondad de ajuste

Sabemos, ver propiedad (3.2), que $\delta_i(\boldsymbol{\theta})/2$ tiene una distribución F con 2 y ν grados de libertad. Además, $F_i = \delta_i(\hat{\boldsymbol{\theta}})/2$ tiene asintóticamente la misma distribución F , $i = 1, 2, \dots, n$. Como sugerido por Lange et al. (1989), un método práctico para evaluar el ajuste del modelo estructural t es el gráfico de la distancia transformada. Utilizando la aproximación de Wilson-Hilferty (Johnson y Kotz, 1970), tenemos que

$$z_i = \frac{(1 - \frac{2}{9\nu}) F_i^{1/3} - (\frac{8}{9})}{\left\{ (\frac{2}{9\nu}) F_i^{2/3} + (\frac{1}{9}) \right\}^{1/2}}, \quad (3.20)$$

$i = 1, 2, \dots, n$, tiene aproximadamente una distribución normal estándar. Así, un $Q - Q$ plot (estadísticos de orden esperados versus observados) de las distancias transformadas z_i , $i = 1, 2, \dots, n$, puede ser usado para evaluar el ajuste de la distribución t multivariada.

Ilustración: Datos de Kelly Los datos de la tabla 3.1, tomados de Kelly (1984), corresponden a mediciones de niveles de suero *kanamycin* hechas por dos métodos en muestras de sangre de 20 recién nacidos. Con el método estándar (X) la muestra de sangre se toma desde el talón del bebé mientras que con el método nuevo (Y) la muestra se toma través de un catéter umbilical.

La figura 3.1 muestra el diagrama de dispersión de los datos. Se observa cierta tendencia lineal y un par de observaciones potencialmente atípicas. Las figuras 3.2 y 3.3 muestran

Tabla 3.1: Medidas de suero hechas por 2 instrumentos

Obs.	X	Y
1	23.0	25.5
2	33.2	26.0
3	16.6	16.3
4	26.3	27.2
5	20.0	23.2
6	20.0	18.1
7	20.6	22.2
8	18.9	17.2
9	17.8	18.8
10	20.0	16.4
11	26.4	24.8
12	21.8	26.8
13	14.9	15.4
14	17.4	14.9
15	20.0	18.1
16	13.2	16.3
17	28.4	31.3
18	25.9	31.2
19	18.9	18.0
20	13.8	15.6

Fuente: Kelly, 1984

Figura 3.1: Gráfico de dispersión

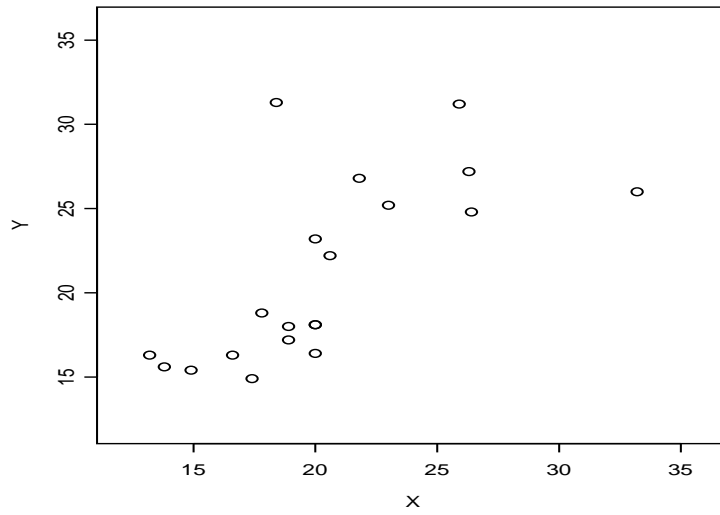


Figura 3.2: Gráfico de la distancia transformada para el modelo normal

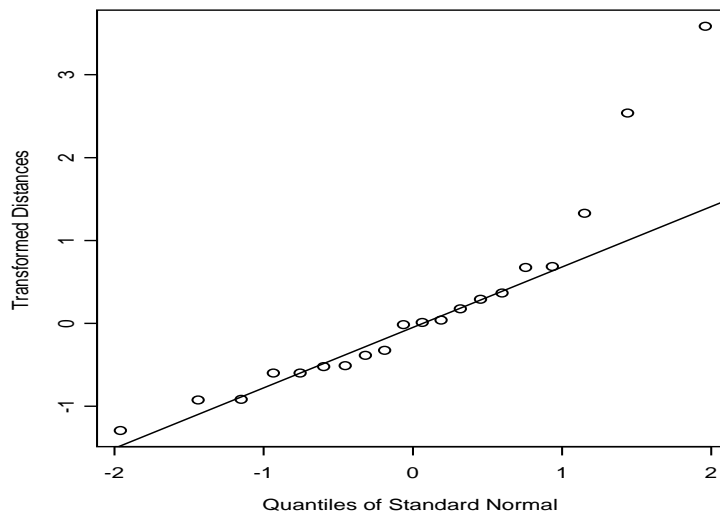
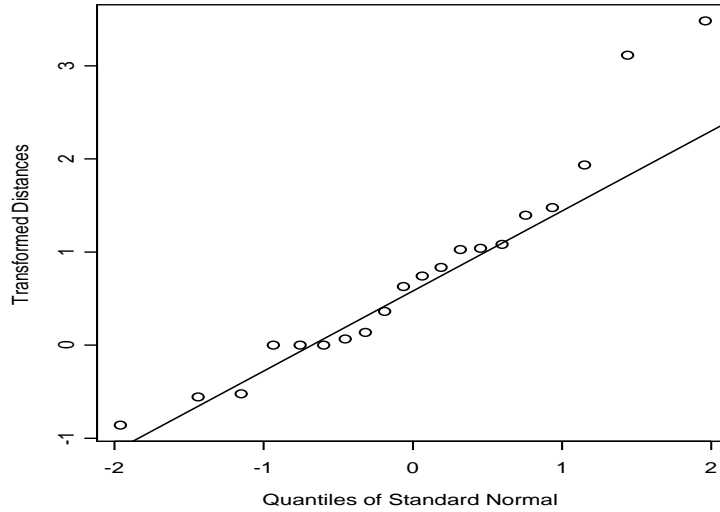


Figura 3.3: Gráfico de la distancia transformada para el modelo t con $\nu = 10$



los gráficos de las distancias transformada para los modelos normal y t con $\nu = 10$. Como podemos observar el modelo estructural t presenta un mejor ajuste que el modelo normal.

La Tabla 3.2, tomada de Bolfarine y Arellano-Valle (1994), muestra los valores de $l(\hat{\gamma})$ y $\hat{\beta}$ para varios valores de ν . Vemos que $\nu = 10$ presenta el mejor ajuste.

Tabla 3.2: Valores de $l(\hat{\gamma})$ y $\hat{\beta}$ para varios valores de ν

ν	$\hat{\beta}$	$l(\hat{\gamma})$
5000	1.07	-73.4701
100	1.08	-73.4195
50	1.09	-73.3718
25	1.11	-73.2945
11	1.15	-73.1925
10	1.16	-73.1905
09	1.17	-73.1938

Con $\nu = 10$, las estimas de los parámetros son,

$$\hat{\mu}_x = 20,36, \quad \hat{\alpha} = -2,89, \quad \hat{\phi}_x = 16,29, \quad \hat{\phi} = 3,30, \quad \text{y} \quad \hat{\beta} = 1,16. \quad (3.21)$$

Algunas estimas de los errores estándar de $\hat{\alpha}$ y $\hat{\beta}$ se presentan en la tabla 3.3. Para más detalle ver Kelly (1984) y Bolfarine y Arellano-Valle (1994). La hipótesis $H_0 : \alpha = 0, \beta = 1$ en la parametrización original es equivalente a $H_0 : \mu_1 = \mu_2, \beta = 1$ en la parametrización

ortogonal. En efecto usando el TVR (3.17) tenemos que $G_t = 2(73,76 - 73,19) = 1,14$, de modo que no podemos rechazar $H_0 : \alpha = 0, \beta = 1$.

Tabla 3.3: Errores estándar de los estimadores $\hat{\alpha}$ y $\hat{\beta}$

Estimador	normal	t con $\nu = 10$	bootstrap, B=200
$\hat{\alpha}$	3.39	3.60	4.58
$\hat{\beta}$	0.16	0.17	0.23

Capítulo 4

Diagnósticos de Influencia

4.1. Introducción

La detección de datos atípicos (outliers) y/o influyentes, es una etapa importante en cualquier análisis estadístico. Esto es esencial para evaluar la sensibilidad (robustez) de los resultados obtenidos, con el conjunto de datos disponibles, ya que observaciones atípicas pueden distorsionar las estimas de los parámetros, conduciendo en algunos casos a decisiones erróneas.

Existen varias alternativas, para evaluar la influencia de perturbaciones en los datos y/o en los supuestos del modelo sobre los estimadores de los parámetros de interés. Ver, por ejemplo, Cook y Weisberg (1982), Chatterjee y Hadi (1988), Cook (1986) y Barnett y Lewis (1994). La eliminación de casos, es una técnica de diagnóstico común para evaluar el efecto de una observación sobre el proceso de estimación y test de hipótesis. Este es un análisis de influencia global, ya que el efecto de la observación es cuantificado eliminándola del conjunto de datos. Alternativamente, Cook (1986) propone un interesante método, denominado influencia local, para evaluar el efecto de pequeñas perturbaciones en los datos y/o en los supuestos del modelo estadístico, sobre los estimadores máximo verosímiles, sin eliminar observaciones. El método fue aplicado por Galea et al. (1997, 2003) en modelos lineales elípticos. Resultados adicionales sobre influencia local y aplicaciones pueden ser encontrados en Escobar y Meeker (1992), Zhao y Lee (1998), Galea et al. (2005) y Lesaffre y Verbeke (1998).

4.2. Influencia local

Sea $l(\boldsymbol{\theta})$ la función de log-verosimilitud del modelo postulado ($\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \phi, \mu_x, \phi_x)^T$) para el modelo estructural) y sea $\boldsymbol{\omega}$ un vector $q \times 1$ de perturbaciones perteneciente a

algún subconjunto de \mathbb{R}^q . Las perturbaciones son hechas en la función de log-verosimilitud que toma la forma $l(\boldsymbol{\theta}|\boldsymbol{\omega})$. Denotando el vector de no perturbación por $\boldsymbol{\omega}_0$, asumimos que $l(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = l(\boldsymbol{\theta})$. Para evaluar la influencia de las perturbaciones sobre el EMV de $\boldsymbol{\theta}$, podemos considerar el desplazamiento de verosimilitudes

$$LD(\boldsymbol{\omega}) = 2 \{l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_\omega)\} ,$$

donde $\widehat{\boldsymbol{\theta}}_\omega$ denota el EMV bajo el modelo perturbado $l(\boldsymbol{\theta}|\boldsymbol{\omega})$.

En algunas situaciones, puede ser de interés evaluar la influencia sobre un subconjunto $\boldsymbol{\theta}_1$ de $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$. Por ejemplo, podemos tener interés en $\boldsymbol{\theta}_1 = \boldsymbol{\beta}$ o $\boldsymbol{\theta}_1 = \boldsymbol{\phi}$. En tales casos, el desplazamiento de verosimilitudes puede definirse como

$$LD_1(\boldsymbol{\omega}) = 2 \{l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_{1\omega}, \widehat{\boldsymbol{\theta}}_2(\widehat{\boldsymbol{\theta}}_{1\omega}))\} ,$$

donde $\widehat{\boldsymbol{\theta}}_{1\omega}$ puede obtenerse desde $\widehat{\boldsymbol{\theta}}_\omega = (\widehat{\boldsymbol{\theta}}_{1\omega}^T, \widehat{\boldsymbol{\theta}}_{2\omega}^T)^T$ y $\widehat{\boldsymbol{\theta}}_2(\widehat{\boldsymbol{\theta}}_{1\omega})$ es el EMV de $\boldsymbol{\theta}_2$ para $\widehat{\boldsymbol{\theta}}_{1\omega}$ fijo en el modelo perturbado.

La idea de influencia local es caracterizar el comportamiento $LD(\boldsymbol{\omega})$ en $\boldsymbol{\omega}_0$. El procedimiento consiste en seleccionar una dirección unitaria \boldsymbol{l} , $\|\boldsymbol{l}\| = 1$ y luego considerar el gráfico de $LD(\boldsymbol{\omega}_0 + a\boldsymbol{l})$ contra a con $a \in \mathbb{R}$. Este gráfico es denominado *línea ajustada*. Note que ya que $LD(\boldsymbol{\omega}_0) = 0$, $LD(\boldsymbol{\omega}_0 + a\boldsymbol{l})$ tiene un mínimo local en $a = 0$. Cada línea ajustada puede caracterizarse considerando la curvatura normal $C_l(\boldsymbol{\theta})$ alrededor de $a = 0$. La sugerencia es tomar la dirección \boldsymbol{l}_{\max} correspondiente a la mayor curvatura $C_{l_{\max}}(\boldsymbol{\theta})$. El gráfico índice de \boldsymbol{l}_{\max} puede revelar aquellos elementos que bajo pequeñas perturbaciones ejercen una notable influencia sobre $LD(\boldsymbol{\omega})$. Cook (1986) muestra que la curvatura normal en la dirección \boldsymbol{l} es de la forma

$$C_l(\boldsymbol{\theta}) = 2 |\boldsymbol{l}^T \boldsymbol{\Delta}^T \ddot{\mathbf{L}}^{-1} \boldsymbol{\Delta} \boldsymbol{l}| , \quad (4.1)$$

donde $-\ddot{\mathbf{L}}$ es la matriz de información observada para el modelo postulado ($\boldsymbol{\omega} = \boldsymbol{\omega}_0$) y $\boldsymbol{\Delta}$ es una matriz $p^* \times q$ con elementos

$$\Delta_{ij} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \theta_i \partial \omega_j} ,$$

evaluada en $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ y $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, $i = 1, \dots, p^*$ y $j = 1, \dots, q$. En el modelo estructural, $p^* = 2p + 3$. Pero, la maximización de (4.1) es equivalente a encontrar el mayor valor propio, en valor absoluto $C_{l_{\max}}$ de la matriz $\mathbf{B} = \boldsymbol{\Delta}^T \ddot{\mathbf{L}}^{-1} \boldsymbol{\Delta}$, y \boldsymbol{l}_{\max} es el correspondiente vector propio. Para un subconjunto $\boldsymbol{\theta}_1$, la curvatura en la dirección \boldsymbol{l} esta dada por

$$C_l(\boldsymbol{\theta}_1) = 2 |\boldsymbol{l}^T \boldsymbol{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta} \boldsymbol{l}| ,$$

donde

$$\mathbf{B}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{22}^{-1} \end{pmatrix},$$

y $\ddot{\mathbf{L}}_{22}$ es obtenida de la partición de $\ddot{\mathbf{L}}$ acorde a la partición de $\boldsymbol{\theta}$. El vector propio \mathbf{l}_{\max} corresponde al mayor valor propio, en valor absoluto, de la matriz $\mathbf{B} = \boldsymbol{\Delta}^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_{22}) \boldsymbol{\Delta}$.

Otra importante dirección, ver Escobar y Meeker (1992) y Verbeke y Molenberghs (2000) es $\mathbf{l} = \mathbf{e}_{in}$, la cual corresponde a la i -ésima posición, donde hay un uno. En ese caso, la curvatura normal, denominada influencia local total, de la unidad experimental i , esta dada por $C_i = 2|\mathbf{e}_{in}^T \mathbf{B} \mathbf{e}_{in}| = 2|b_{ii}|$, donde b_{ii} es el i -ésimo elemento de la diagonal de \mathbf{B} , $i = 1, \dots, n$. Verbeke y Molenberghs (2000) proponen considerar la i -ésima observación como influyente si C_i es mayor que el valor de corte $2 \sum_{i=1}^n C_i/n$. Nosotros usamos \mathbf{l}_{\max} y C_i como diagnóstico para influencia local.

Fung y Kwan (1997), presentan una interesante discusión sobre aplicaciones de influencia local en otras medidas de influencia diferentes al desplazamiento de verosimilitudes. Ellos muestran que una medida de influencia, digamos \hat{T}_ω , es escala invariante si $\dot{\mathbf{I}} = \partial \hat{T}_\omega / \partial \boldsymbol{\omega} |_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \mathbf{0}$. Cuando esta derivada es distinta de cero el ordenamiento entre los componentes de \mathbf{l}_{\max} no es necesariamente preservado bajo cambios de escala. En particular, para el desplazamiento de verosimilitudes, $\dot{\mathbf{I}} = \partial L(\hat{\boldsymbol{\theta}}_\omega) / \partial \boldsymbol{\omega} |_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \mathbf{0}$. Esta propiedad también se cumple, por ejemplo, para la medida de influencia propuesta en Paula (1993). Sin embargo esta propiedad no se satisface en otras medidas de influencia, como es notado por Fung y Kwan (1997).

4.3. Derivación de la curvatura para el MEM Multivariado

En esta sección derivamos la matriz de información observada y la matriz $\boldsymbol{\Delta}$ para deferentes esquemas de perturbación en el MEM multivariado (2.3). Primero derivamos la matriz de información observada.

4.3.1. Matriz de información observada

Sea $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \phi, \mu_x, \phi_x)^T$ el vector de parámetros del MEM. Para el modelo estructural dado en (2.4), la función de log-verosimilitud puede ser escrita como

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n l_i(\boldsymbol{\theta}) , \quad (4.2)$$

donde

$$l_i(\boldsymbol{\theta}) = -\frac{r}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} Q_i(\boldsymbol{\theta}) , \quad (4.3)$$

con

$$Q_i(\boldsymbol{\theta}) = (\mathbf{Y}_i - \boldsymbol{\mu}_Y)^T \mathbf{V}^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_Y) = \frac{1}{\phi} (q_{1i} - f q_{2i}^2) ,$$

$i = 1, \dots, n$, ya que

$$\mathbf{V}^{-1} = \frac{1}{\phi} (\mathbf{I}_r - f \mathbf{b}\mathbf{b}^T) \quad y \quad f = \frac{\phi_x}{\phi + \phi_x c} ,$$

con

$$q_{1i} = (X_i - \mu_x)^2 + (\mathbf{Y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\mu_x)^T (\mathbf{Y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\mu_x)$$

y

$$q_{2i} = (X_i - \mu_x) + (\mathbf{Y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\mu_x)^T \boldsymbol{\beta} ,$$

$i = 1, \dots, n$. Luego la matriz $\ddot{\mathbf{L}}_S$ esta dada por

$$\ddot{\mathbf{L}}_S = \begin{pmatrix} \mathbf{L}_{\alpha\alpha} & \mathbf{L}_{\alpha\beta} & \mathbf{L}_{\alpha\phi} & \mathbf{L}_{\alpha\mu_x} & \mathbf{L}_{\alpha\phi_x} \\ & \mathbf{L}_{\beta\beta} & \mathbf{L}_{\beta\phi} & \mathbf{L}_{\beta\mu_x} & \mathbf{L}_{\beta\phi_x} \\ & & \mathbf{L}_{\phi\phi} & \mathbf{L}_{\phi\mu_x} & \mathbf{L}_{\phi\phi_x} \\ & & & \mathbf{L}_{\mu_x\mu_x} & \mathbf{L}_{\mu_x\phi_x} \\ & & & & \mathbf{L}_{\phi_x\phi_x} \end{pmatrix} .$$

Los elementos de la matriz $\ddot{\mathbf{L}}_S$ son presentados en el apéndice.

4.3.2. Perturbación de casos

Considere el vector de pesos $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ de modo que la función de log-verosimilitud perturbada es

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \omega_i l_i(\boldsymbol{\theta}) , \quad (4.4)$$

donde $l_i(\boldsymbol{\theta})$ está dada por (4.3). El vector de no perturbación es denotado por $\boldsymbol{\omega}_0 = \mathbf{1}_n = (1, \dots, 1)^T$. Sea $\boldsymbol{\Delta}_\gamma = (\boldsymbol{\Delta}_{\gamma 1}, \dots, \boldsymbol{\Delta}_{\gamma n})$ la sub-matriz de $\boldsymbol{\Delta}$ in (4.1), correspondiente al

parámetro γ . Así, $\Delta_{\gamma i}$, $i = 1, \dots, n$ es la i -ésima columna de Δ_γ con $\gamma = \alpha, \beta, \phi, \mu_x, \phi_x$. Luego, después de algunas manipulaciones algebraicas, tenemos que

$$\Delta_{\alpha i} = \frac{1}{\phi} [(\mathbf{Y}_i - \bar{\mathbf{Y}}) - f q_{2i} \boldsymbol{\beta}], \quad (4.5)$$

$$\Delta_{\beta i} = -f \boldsymbol{\beta} + (\bar{X} + f q_{2i}) \Delta_{\alpha i} \quad (4.6)$$

$$\Delta_{\phi i} = -\frac{1}{2} \left(\frac{p}{\phi} + \frac{f}{\phi_x} \right) - \frac{1}{2\phi} \left[\frac{f^2}{\phi_x} q_{2i}^2 - Q_i(\boldsymbol{\theta}) \right] \quad (4.7)$$

$$\Delta_{\mu_x i} = \left(\frac{1 - fc}{\phi} \right) q_{2i}, \quad (4.8)$$

$$\Delta_{\phi_x i} = -\frac{f}{2\phi_x} (c - q_{2i} \Delta_{\mu_x i}), \quad (4.9)$$

$i = 1, \dots, n$. Las expresiones (4.5) a (4.9) deben ser evaluadas en los correspondientes EMV y la matriz Δ en (4.1) toma la forma

$$\Delta_S = \begin{pmatrix} \Delta_\alpha \\ \Delta_\beta \\ \Delta_\phi \\ \Delta_{\mu_x} \\ \Delta_{\phi_x} \end{pmatrix}$$

, la cual es de dimension $(2p + 3) \times n$.

4.3.3. Perturbación del predictor

En esta sección modificamos el predictor observado mediante esquemas de perturbación aditivo y multiplicativo. El modelo perturbado nos queda

$$\mathbf{Z}_i(\boldsymbol{\omega}) = \begin{pmatrix} 0 \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix} x_i + \boldsymbol{\epsilon}_i,$$

con $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n$ iid $N_r(\mathbf{0}, \phi \mathbf{I}_r)$, donde

$$\mathbf{Z}_i(\boldsymbol{\omega}) = \begin{cases} \begin{pmatrix} X_i + \omega_i \\ \mathbf{Y}_i \end{pmatrix}, & \text{con perturbación aditiva,} \\ \begin{pmatrix} X_i \omega_i \\ \mathbf{Y}_i \end{pmatrix}, & \text{con perturbación multiplicativa,} \end{cases}$$

$i = 1, \dots, n$ y $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$. El vector de perturbación nula es $\boldsymbol{\omega}_0 = \mathbf{0}$ en el caso aditivo y $\boldsymbol{\omega}_0 = \mathbf{1}_n$ para el caso multiplicativo. La función de log-verosimilitud perturbada

sigue desde (4.2) con $\mathbf{Z}_i(\boldsymbol{\omega})$ reemplazado por \mathbf{Z}_i , $i = 1, \dots, n$. Así, usando la misma notación de la sección 4.3.2, sea $\Delta_{\gamma i}^a$ ($\Delta_{\gamma i}^m$) la i -ésima columna de Δ_γ cuando consideramos el esquema aditivo (multiplicativo). Luego nos queda

$$\begin{aligned}\Delta_{\alpha i}^a &= -\left(\frac{f}{\phi}\right) \boldsymbol{\beta}, \\ \Delta_{\beta i}^a &= \frac{f}{\phi} [(\mathbf{Y}_i - \bar{\mathbf{Y}} - \bar{X} \boldsymbol{\beta}) - 2f q_{2i} \boldsymbol{\beta}], \\ \Delta_{\phi i}^a &= \frac{1}{\phi} [q_{2i} \frac{\partial f}{\partial \phi} + \frac{1}{\phi} (X_i - \bar{X} - f q_{2i})], \\ \Delta_{\mu_x i}^a &= \frac{1 - fc}{\phi} \text{ y} \\ \Delta_{\phi_x i}^a &= \frac{1}{\phi} \frac{\partial f}{\partial \phi_x} q_{2i},\end{aligned}$$

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

$$\frac{\partial f}{\partial \phi} = -\frac{f^2}{\phi_x} \text{ y } \frac{\partial f}{\partial \phi_x} = \frac{f(1 - fc)}{\phi_x}. \quad (4.10)$$

Para el esquema multiplicativo tenemos

$$\Delta_{\gamma i}^m = X_i \Delta_{\gamma i}^a,$$

con $\gamma = \alpha, \beta, \phi, \mu_x, \phi_x$, $i = 1, \dots, n$.

4.3.4. Perturbación de una respuesta

En esta sección perturbamos una variable de respuesta utilizando los mismos esquemas de la sección anterior. Sin pérdida de generalidad, perturbaremos la primera respuesta. De este modo el vector de respuesta perturbado es

$$\mathbf{Z}_i(\boldsymbol{\omega}) = \begin{pmatrix} X_i \\ \mathbf{Y}_i(\boldsymbol{\omega}) \end{pmatrix},$$

donde

$$\mathbf{Y}_i(\boldsymbol{\omega}) = \begin{cases} \mathbf{Y}_i + \omega_i \mathbf{e}_1, & \text{additive perturbation,} \\ \mathbf{Y}_i \odot \mathbf{1}_p(\omega_i), & \text{multiplicative perturbation,} \end{cases}$$

con $\mathbf{e}_1 = (1, 0, \dots, 0)^T$, $\mathbf{1}_p(\omega_i) = (\omega_i, 1, \dots, 1)^T$, vectores $p \times 1$, $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ y \odot denota el producto Hadamard. Similarmente a la sección 4.3.3 obtenemos:

$$\Delta_{\alpha i}^a = \frac{1}{\phi} [\mathbf{e}_1 - f \beta_1 \boldsymbol{\beta}],$$

$$\begin{aligned}\Delta_{\beta i}^a &= \frac{1}{\phi} [\bar{X} (\mathbf{e}_1 - f \beta_1 \boldsymbol{\beta}) + f q_{2i} (\mathbf{e}_1 - 2f \beta_1 \boldsymbol{\beta}) + f \beta_1 (\mathbf{Y}_i - \bar{\mathbf{Y}})] , \\ \Delta_{\phi i}^a &= \frac{1}{\phi} [\beta_1 \frac{\partial f}{\partial \phi} q_{2i} + \frac{1}{\phi} \{(Y_{i1} - \bar{Y}_1) - f q_{2i} \beta_1\}] , \\ \Delta_{\mu_{xi}}^a &= \frac{1}{\phi} \beta_1 (1 - fc) , \\ \Delta_{\phi_{xi}}^a &= \frac{\beta_1}{\phi} \frac{\partial f}{\partial \phi_x} q_{2i} ,\end{aligned}$$

y

$$\Delta_{\gamma i}^m = Y_{i1} \Delta_{\gamma i}^a ,$$

con $\gamma = \alpha, \beta, \phi, \mu_x, \phi_x$, $i = 1, \dots, n$ y $\partial f / \partial \phi$ y $\partial f / \partial \phi_x$ como en (4.10).

4.3.5. Perturbación de varianzas de los errores de medición

En el MEM hemos asumido homocedasticidad, es decir que las varianzas de los errores de medición son iguales. En esta sección, consideramos un modelo heterocedástico donde

$$Var(\epsilon_{ij}) = \frac{\phi}{\omega_j} , \quad \omega_j > 0 , \quad j = 0, 1, \dots, p .$$

Así, el modelo perturbado puede escribirse como

$$\mathbf{Z}_i = \begin{pmatrix} 0 \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix} x_i + \boldsymbol{\epsilon}_i ,$$

con $\boldsymbol{\epsilon}_i$, $i = 1, \dots, n$ vectores aleatorios iid $\mathbf{N}_r(\mathbf{0}, \phi \mathbf{D}^{-1}(\boldsymbol{\omega}^*))$, donde $\mathbf{D}(\boldsymbol{\omega}^*) = \text{diag}(\omega_0, \omega_1, \dots, \omega_p)$ y $\mathbf{D}^{-1}(\boldsymbol{\omega}^*)$ denotando su inversa, $\boldsymbol{\omega}^* = (\omega_0, \boldsymbol{\omega}^T)^T$, con $\boldsymbol{\omega} = (\omega_1, \dots, \omega_p)^T$.

En este modelo perturbado tenemos que \mathbf{Z}_i son vectores aleatorios iid con

$$\mathbf{Z}_i \sim \mathbf{N}_r\left(\begin{pmatrix} 0 \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} 1 \\ \boldsymbol{\beta} \end{pmatrix} \mu_x , \mathbf{V}(\boldsymbol{\omega}^*)\right) , \quad (4.11)$$

$i = 1, \dots, n$, donde $\mathbf{V}(\boldsymbol{\omega}^*) = \phi_x \mathbf{b}\mathbf{b}^T + \phi \mathbf{D}^{-1}(\boldsymbol{\omega}^*)$. Luego, (4.11) implica que la función de log-verosimilitud para el MEM perturbado es

$$l(\boldsymbol{\theta} | \boldsymbol{\omega}^*) = \sum_{i=1}^n l_i(\boldsymbol{\theta} | \boldsymbol{\omega}^*) , \quad (4.12)$$

where

$$l_i(\boldsymbol{\theta} | \boldsymbol{\omega}^*) = -\frac{r}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}(\boldsymbol{\omega}^*)| - \frac{1}{2} Q_i(\boldsymbol{\theta} | \boldsymbol{\omega}^*) ,$$

con

$$Q_i(\boldsymbol{\theta}|\boldsymbol{\omega}^*) = \frac{1}{\phi} [q_{1i}(\boldsymbol{\omega}^*) - f(\boldsymbol{\omega}^*) q_{2i}^2(\boldsymbol{\omega}^*)] ,$$

$$q_{1i}(\boldsymbol{\omega}^*) = \omega_0(X_i - \mu_x)^2 + (\mathbf{Y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\mu_x)^T \mathbf{D}(\boldsymbol{\omega}) (\mathbf{Y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\mu_x) ,$$

$$q_{2i}(\boldsymbol{\omega}^*) = \omega_0(X_i - \mu_x) + \boldsymbol{\beta}^T \mathbf{D}(\boldsymbol{\omega}) (\mathbf{Y}_i - \boldsymbol{\alpha} - \boldsymbol{\beta}\mu_x) ,$$

$i = 1, \dots, n$, y

$$f(\boldsymbol{\omega}^*) = \frac{\phi_x}{\phi + \phi_x c(\boldsymbol{\omega}^*)} ,$$

con $c(\boldsymbol{\omega}^*) = \mathbf{b}^T \mathbf{D}(\boldsymbol{\omega}^*) \mathbf{b}$, como definida anteriormente. Usando (4.12) tenemos:

$$\begin{aligned} \Delta_\alpha &= \mathbf{0} , \\ \Delta_\beta &= \frac{nf}{\phi} \{ -(\phi + f \mathbf{b}^T \mathbf{S} \mathbf{b}) (\mathbf{0} , \mathbf{D}(\boldsymbol{\beta})) + f \phi \boldsymbol{\beta} \mathbf{b}^T \mathbf{D}(\mathbf{b}) + (\mathbf{0} , \mathbf{D}(\mathbf{S}_{10} + \mathbf{S}_{11} \boldsymbol{\beta})) \\ &\quad + [(\mathbf{S}_{10} , \mathbf{S}_{11}) - f(\mathbf{S}_{10} + \mathbf{S}_{11} \boldsymbol{\beta}) \mathbf{b}^T - 2f \boldsymbol{\beta} \mathbf{b}^T \mathbf{S} + 2f^2 (\mathbf{b}^T \mathbf{S} \mathbf{b}) \boldsymbol{\beta} \mathbf{b}^T] \mathbf{D}(\mathbf{b}) \} , \\ \Delta_\phi &= fn \left\{ \frac{f}{\phi_x} \left(\frac{1}{2} + \frac{f \mathbf{b}^T \mathbf{S} \mathbf{b}}{\phi} + \frac{\phi_x \mathbf{b}' \mathbf{S} \mathbf{b}}{2\phi^2} \right) \mathbf{b}' \mathbf{D}(\mathbf{b}) \right. \\ &\quad \left. + \frac{1}{2f\phi^2} \mathbf{1}_r^T \mathbf{D}(\mathbf{v}) - \frac{1}{\phi} \left(\frac{f}{\phi_x} + \frac{1}{\phi} \right) \mathbf{b}^T \mathbf{S} \mathbf{D}(\mathbf{b}) \right\} , \\ \Delta_{\mu_x} &= \mathbf{0} \text{ and} \\ \Delta_{\phi_x} &= -\frac{fn(1-fc)}{2\phi\phi_x} \{ [\phi + 2f \mathbf{b}^T \mathbf{S} \mathbf{b}] \mathbf{b}^T \mathbf{D}(\mathbf{b}) - 2\mathbf{b}^T \mathbf{S} \mathbf{D}(\mathbf{b}) \} , \end{aligned}$$

con $\mathbf{v} = (S_{00}, S_{11}, \dots, S_{pp})^T$ y las matrices \mathbf{S}_{10} y \mathbf{S}_{11} están definidas en el apéndice A.

Si el supuesto de homogeneidad es razonable los elementos del vector $|\mathbf{l}_{\max}|$ deberían ser similares. En otro caso, \mathbf{l}_{\max} puede ser útil para construir un modelo más apropiado.

4.4. Derivación de la curvatura para el modelo estructural t

Como en Kim (2000) consideramos un modelo perturbado en el cual los vectores aleatorios $\mathbf{Z}_i = (X_i, Y_i)^T$ son independientes con distribución $t_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\omega_i; \nu)$, $i = 1, \dots, n$. Aquí $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T$ y $\boldsymbol{\omega}_0 = \mathbf{1}_n = (1, \dots, 1)^T$. Con este esquema de perturbación colocamos pesos a la matriz de escala para cada observación y, en general, provee resultados similares a perturbación de casos.

En este caso la matriz Δ esta dada por,

$$\Delta = D_\theta D(\mathbf{a}), \tag{4.13}$$

donde $D_\theta = [\mathbf{d}_{1\theta} \dots \mathbf{d}_{n\theta}]$ y $D(\mathbf{a}) = \text{diag}(a_1, \dots, a_n)$, con $a_i = -\nu(\nu + 2)/2(\nu + \delta_i(\boldsymbol{\theta}))$, $\mathbf{d}_{i\theta} = (d_{i\alpha}, d_{i\beta}, d_{i\phi}, d_{i\mu_x}, d_{i\phi_\gamma})^T$, $d_{i\gamma} = \frac{\partial \delta_i(\boldsymbol{\theta})}{\partial \gamma}$ como en el apéndice $i = 1, \dots, n$, evaluada en $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Note que para el modelo normal ($\nu \rightarrow \infty$), $\boldsymbol{\Delta} = (-1/2)D_\theta$.

Para el esquema de perturbación de casos, la correspondiente matriz $\boldsymbol{\Delta}$ es

$$\boldsymbol{\Delta} = \mathbf{C} - \frac{1}{2}D_\theta D(\mathbf{b}), \quad (4.14)$$

donde $\mathbf{C} = \mathbf{c}(\hat{\boldsymbol{\theta}}) \otimes \mathbf{1}_n^T$, $\mathbf{c}(\hat{\boldsymbol{\theta}}) = -\frac{1}{2} \frac{\partial \log |\mathbf{V}|}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, ver apéndice, y $\mathbf{b} = (b_1, \dots, b_n)^T$ con $b_i = \frac{\nu + 2}{\nu + \delta_i(\boldsymbol{\theta})}$, $i = 1, \dots, n$. Para el modelo normal, $\boldsymbol{\Delta} = \mathbf{C} - \frac{1}{2}D_\theta$.

Para más detalles sobre este capítulo, ver Galea et al. (2002a, 2002b)

4.5. Aplicaciones

En esta sección aplicamos la metodología a tres conjuntos de datos aparecidos en la literatura.

Aplicación 1: Datos de Chipkevitch Estos datos corresponden a mediciones del volumen testicular de 42 adolescentes usando cinco técnicas diferentes: ultrasonido (US, X), método gráfico (I, Y_1), medición dimensional (II, Y_2), Prader orchidometer (III, Y_3) y anillo orchidometer (IV, Y_4), con el ultrasonido considerado como instrumento estándar. Los volúmenes testiculares derechos son presentados en la tabla 4.1. Una transformación raíz cúbica fue usada para alcanzar normalidad. Asumiremos el modelo homocedástico y nos centraremos en el vector de parámetros $\boldsymbol{\theta}$ y sobre el subconjunto $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$, que corresponden a los sesgos aditivos y multiplicativos, respectivamente, de los instrumentos de medición.

Tabla 4.1: Datos volumen testicular (en ml)

Adolescente número	Métodos				
	US	I	II	III	IV
1	5.0	7.5	5.9	8.0	9.0
2	5.7	5.0	4.8	6.0	10.0
3	7.4	5.0	6.8	9.0	12.0
4	2.6	3.5	3.1	4.0	4.0
5	5.7	5.0	5.0	6.0	7.0
6	6.1	5.0	4.4	7.0	8.0
7	6.2	5.0	6.0	8.0	9.0
8	10.4	10.0	8.8	10.0	10.0
9	9.1	7.5	7.9	10.0	11.0
10	14.8	10.0	13.0	12.0	15.0
11	16.4	12.5	10.3	17.5	17.5
12	9.6	7.5	8.2	10.0	11.0
13	15.7	15.0	19.8	20.0	20.0
14	3.0	2.0	2.0	3.0	4.0
15	16.4	15.0	17.3	20.0	20.0
16	17.6	15.0	17.3	20.0	22.5
17	10.0	7.5	7.9	12.0	12.0
18	4.1	3.5	4.4	4.0	6.0
19	12.7	10.0	11.4	12.0	12.0
20	2.7	3.5	4.1	2.5	6.0
21	10.2	10.0	11.1	12.0	13.5
22	16.5	10.0	15.3	15.0	15.0
23	4.5	3.5	3.9	6.0	7.0
24	5.6	5.0	4.5	4.5	6.0
25	11.0	7.5	9.7	9.0	11.0
26	9.2	10.0	11.3	12.0	13.5
27	8.5	7.5	8.8	12.0	12.0
28	5.4	5.0	6.1	8.0	8.0
29	6.7	7.5	7.2	10.0	8.0
30	5.3	5.0	5.9	8.0	10.0
31	20.0	20.0	16.3	25.0	22.5
32	18.8	15.0	16.3	20.0	25.0
33	13.9	12.5	12.2	15.0	17.5
34	9.4	10.0	10.3	12.0	13.5
35	9.1	7.5	10.8	12.0	12.0
36	14.1	15.0	13.0	13.5	15.0
37	9.3	10.0	8.4	10.0	10.0
38	20.9	20.0	22.1	25.0	25.0
39	11.5	10.0	10.6	15.0	13.5
40	9.7	10.0	9.7	11.0	12.0
41	13.7	12.5	11.6	17.5	15.0
42	8.9	10.0	8.1	12.0	12.0

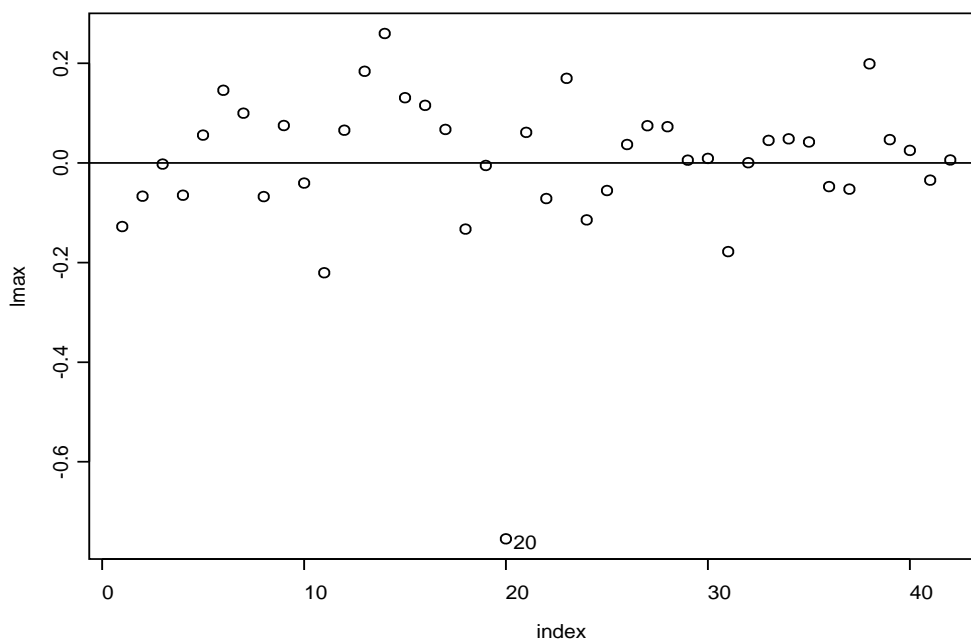


Figura 4.1: Gráfico índice de l_{\max} para perturbación de casos–Aplicación 1.

La suposición de homogeneidad parece sensible si nos centramos en θ , ya que

$$|l_{\max}| = (0,43504, 0,51309, 0,43625, 0,42440, 0,42078)^T .$$

Las estimas de los parámetros son $\widehat{\alpha}_1 = 0,068428$, $\widehat{\alpha}_2 = 0,030502$, $\widehat{\alpha}_3 = 0,033242$, $\widehat{\alpha}_4 = 0,38755$, $\widehat{\beta}_1 = 0,93198$, $\widehat{\beta}_2 = 0,96856$, $\widehat{\beta}_3 = 1,0319$, $\widehat{\beta}_4 = 0,89725$, $\widehat{\phi} = 0,0054155$, $\widehat{\mu}_x = 2,1005$ y $\widehat{\phi}_x = 0,12333$.

En el esquema de perturbación de casos tenemos $C_{l_{\max}} = 4,582$ y se destaca el adolescente 20, como se observa en la figura 4.1.

Siguiendo el enfoque de Wu y Luo (1993) presentamos en la figura 4.2 gráficos del desplazamiento de verosimilitudes $LD(\omega(a))$ versus a para las direcciones l_{\max} y e_{20} , donde $\omega(a) = \mathbf{1}_{42} + a l_{\max}$ y $\omega(a) = \mathbf{1}_{42} - a e_{20}$, respectivamente, con e_{20} denotando un vector nulo teniendo con el 20^{ésimo} componente reemplazado por 1. Las curvas presentan comportamientos similares. Tomando $a > 0$, ellas presentan cambios substanciales, indicando que el adolescente 20 podría ser influyente.

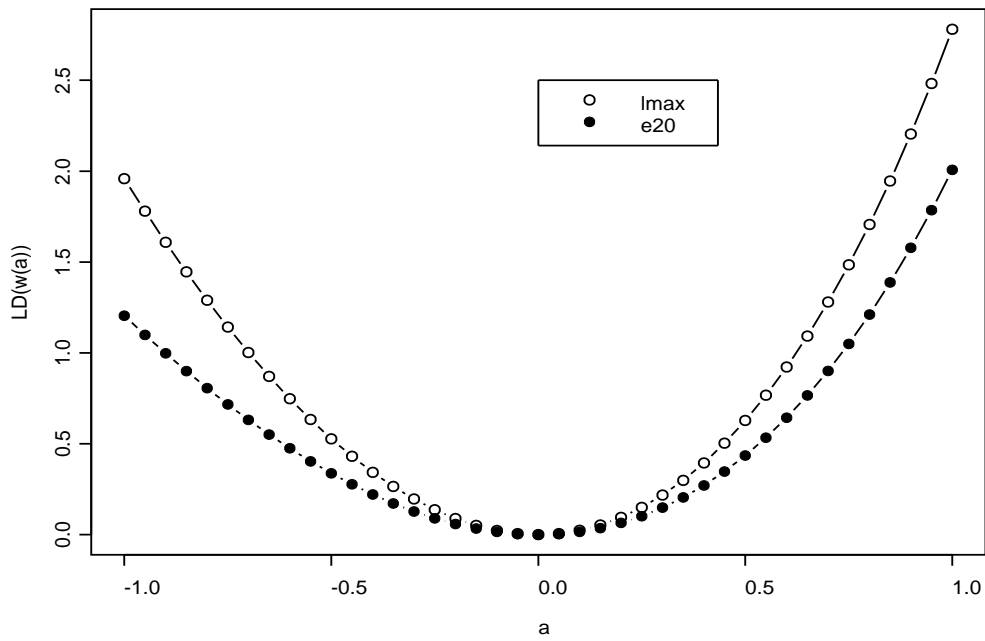


Figura 4.2: Gráficos del desplazamiento de verosimilitudes $LD(\omega(a))$ versus a para las direcciones l_{\max} y e_{20} – Aplicación 1.

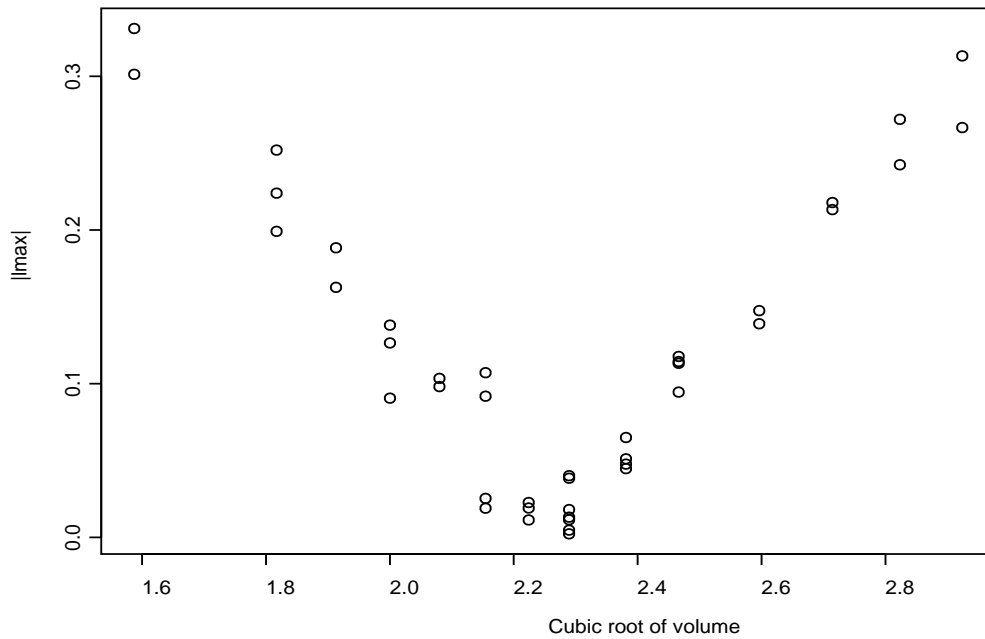


Figura 4.3: Gráfico de $|l_{\max}|$ versus raíz cúbica de las mediciones tomadas por la técnica IV, perturbaciones aditivas – Aplicación 1.

A continuación analizamos los efectos de perturbar las mediciones hechas por los cinco instrumentos. Los valores de $C_{l_{\max}}$ para las perturbaciones aditivas son 235.14 (US), 243.83 (I), 238.8 (II), 231.20 (III) y 246.50 (IV), mientras que para las perturbaciones multiplicativas los valores de $C_{l_{\max}}$ son 1412.6 (US), 1367.6 (I), 1400.1 (II), 1515.2 (III) y 1680.9 (IV). Estos valores revelan que el desplazamiento de verosimilitudes es más sensible a perturbaciones multiplicativas que aditivas. También, el desplazamiento de verosimilitudes es más sensible a la perturbación de las mediciones que al esquema de ponderación de casos. Las figuras 4.3 a 4.5 ilustran las diferencias entre las perturbaciones. La técnica IV es usada debido a que presenta un mayor valor de $C_{l_{\max}}$.

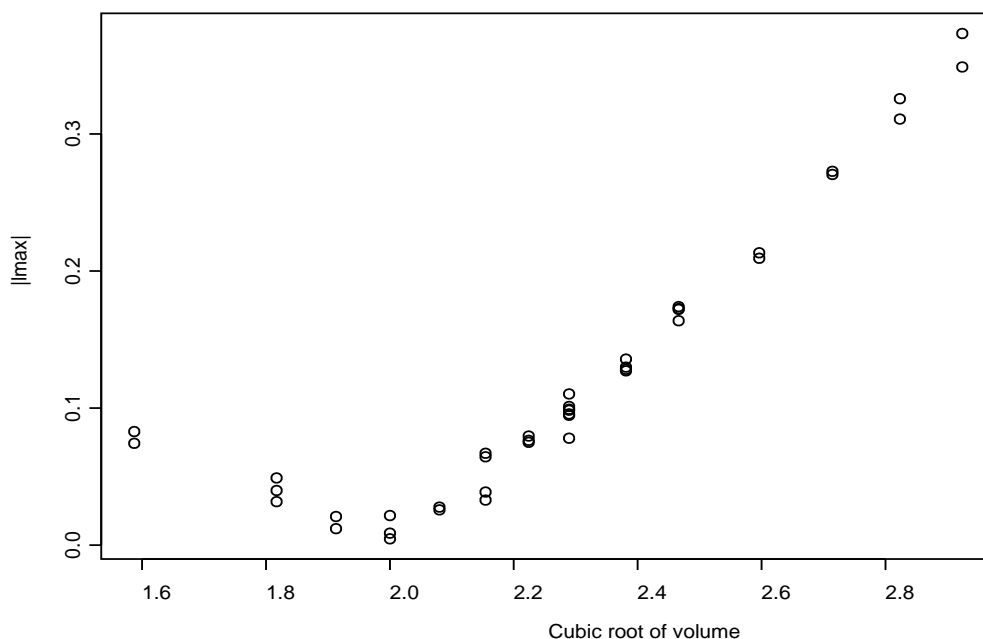


Figura 4.4: Gráfico de $|\mathbf{l}_{\max}|$ versus raíz cúbica de las mediciones tomadas por la técnica IV, perturbaciones multiplicativas – Aplicación 1.

Aplicación 2: Datos de Barnett Consideremos nuevamente los datos de Barnett (1969), presentados en la tabla 2.1. Adoptando el modelo homocedástico, y de acuerdo al esquema de perturbación de varianzas tenemos que $C_{l_{\max}}(\boldsymbol{\theta}) = 59,35$,

$$\mathbf{l}_{\max} = (0,70839, 0,30780, 0,38823, 0,50271)^T ,$$

$$C_{l_{\max}}(\boldsymbol{\theta}_1) = 1,221 \text{ y}$$

$$\mathbf{l}_{\max} = (-0,73253, -0,17966, 0,36928, 0,54291)^T . \quad (4.15)$$

Los elementos del vector $|\mathbf{l}_{\max}|$ son una indicación de que la varianza de los errores de medición son diferentes. Nosotros nos centraremos en $\boldsymbol{\theta}_1$. El vector \mathbf{l}_{\max} puede ser usado para ajustar las varianzas de los errores mediante

$$Var(\epsilon_{ij}) = \frac{\phi}{1 + al_{\max_j}} ,$$

a es tal que el denominador es positivo, $j = 0, 1, \dots, p$. Transformamos los datos \mathbf{Z}_i resultando

$$\left((1 + al_{max_0})^{1/2} X_i, (1 + al_{max_1})^{1/2} Y_{i1}, \dots, (1 + al_{max_p})^{1/2} Y_{ip} \right)^T ,$$

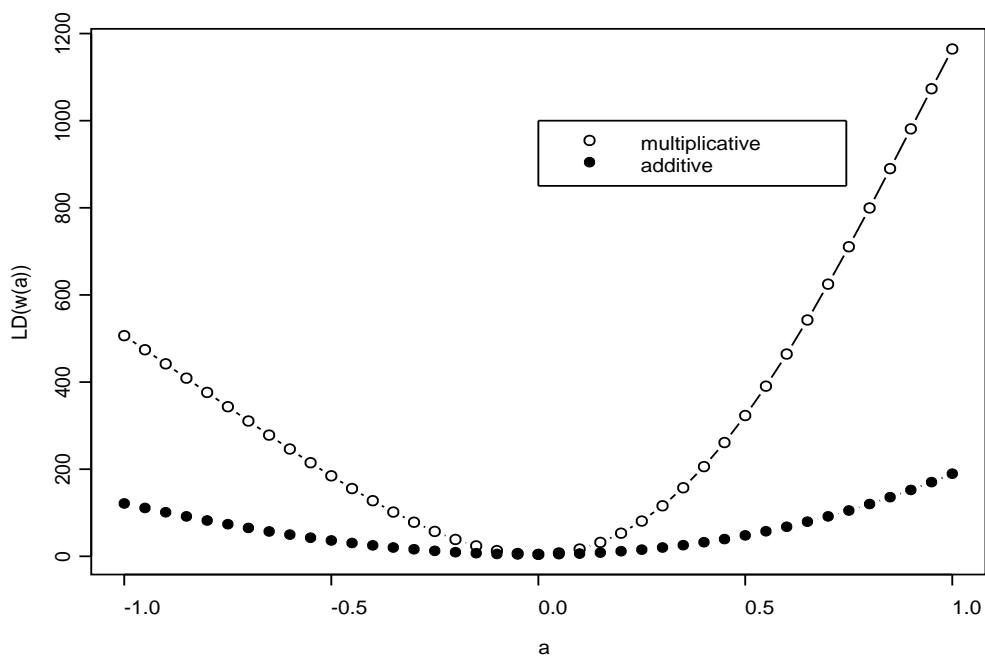


Figura 4.5: Gráficos del desplazamiento de verosimilitudes $LD(\omega(a))$ versus a para la dirección \mathbf{l}_{\max} usando perturbaciones aditivas y multiplicativas sobre las mediciones tomadas por la técnica IV – Aplicación 1.

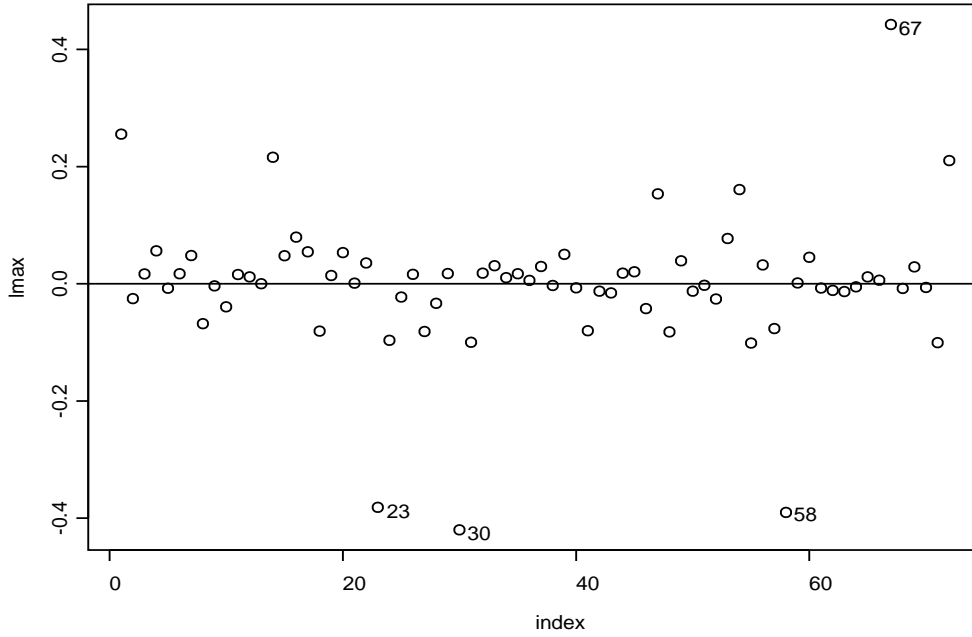


Figura 4.6: Gráfico índice de \mathbf{l}_{\max} para perturbación de casos – Aplicación 2.

$i = 1, \dots, n$, y proporcionando un nuevo vector \mathbf{l}_{\max} (\mathbf{l}_{\max}^* , digamos). Graficando

$$\delta_a = \max_{j=0,1,\dots,p} |l_{\max_j}^*| - \min_{j=0,1,\dots,p} |l_{\max_j}^*|$$

versus a seleccionamos a tal que δ_a sea mínimo. Comenzando desde \mathbf{l}_{\max} en (4.15) con $\delta_a = 0,55287$ obtenemos $a = 0,2295$ y $\delta_a = 0,36783$. Así, podemos considerar un modelo revisado con varianzas de los errores

$$\frac{\phi}{0,83188}, \frac{\phi}{0,95877}, \frac{\phi}{1,0848} \text{ y } \frac{\phi}{1,1246}.$$

Las estimas de los parámetros en este modelo son are $\widehat{\alpha}_1 = -190,95$, $\widehat{\alpha}_2 = -528,78$, $\widehat{\alpha}_3 = -447,39$, $\widehat{\beta}_1 = 1,0537$, $\widehat{\beta}_2 = 1,1920$, $\widehat{\beta}_3 = 1,1351$, $\widehat{\phi} = 33964$, $\widehat{\mu}_x = 2246,1$ y $\widehat{\phi}_x = 534,031$. Perturbación de casos destaca los pacientes 23, 30, 58 y 67, al igual que lo observado en la figura 4.6.

En lo que sigue comparamos influencia local y eliminación de casos. Siguiendo a Zhao y Lee (1998), la distancia de Cook puede definirse por

$$D2_i = (\widehat{\boldsymbol{\theta}}_{1(i)} - \widehat{\boldsymbol{\theta}}_1)^T [\mathbf{V}(\widehat{\boldsymbol{\theta}}_1)]^{-1} (\widehat{\boldsymbol{\theta}}_{1(i)} - \widehat{\boldsymbol{\theta}}_1) / 2p,$$

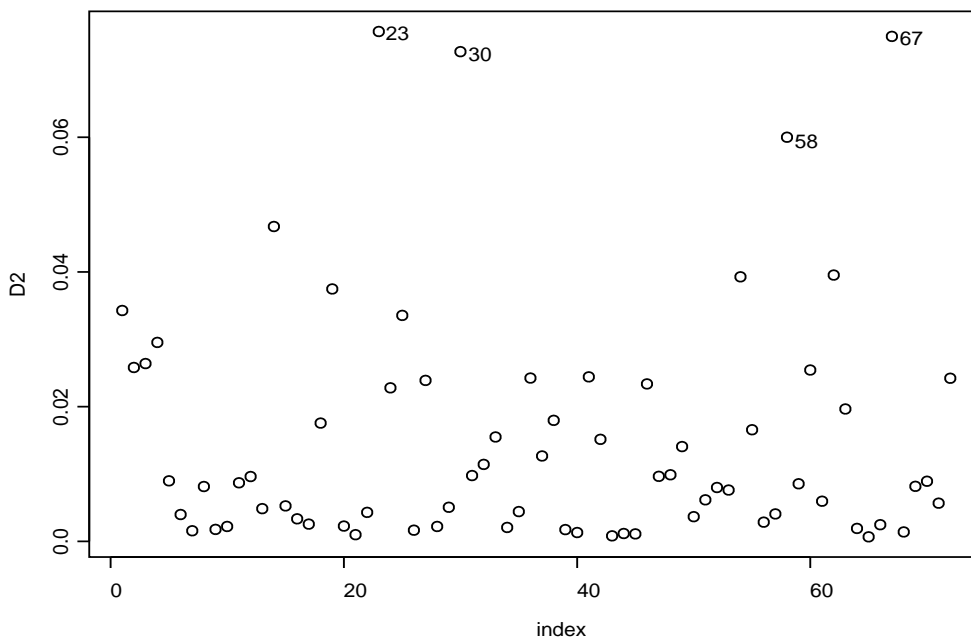


Figura 4.7: Gráfico índice de la distancia de Cook ($D2_i$) – Aplicación 2.

$i = 1, \dots, n$, donde $\hat{\theta}_{1(i)}$ denota el estimador del parámetro sin el caso i , y $\mathbf{V}(\hat{\theta}_1)$ es un estimador consistente de $\text{Var}(\hat{\theta}_1)$, y en este caso usamos $(-\ddot{\mathbf{L}}_S)^{-1}$. La figura 4.7 nos muestra el gráfico índice de $D2_i, i = 1, \dots, n$. De nuevo se destacan los pacientes 23, 30, 58 y 67.

La figura 4.8 muestra gráficos del desplazamiento de verosimilitudes $LD_1(\omega(a))$ versus a para las direcciones \mathbf{l}_{\max} y \mathbf{e}_{67} . Note que la curva LD_1 en la dirección \mathbf{e}_{67} tiene mucho menos cambio que la curva en la dirección \mathbf{l}_{\max} (En la figura 4.6 el caso 67 tiene mayor componente en $|\mathbf{l}_{\max}|$). Ya que \mathbf{l}_{\max} es dominado por los pacientes 23, 30, 58 y 67, tenemos indicios de influencia conjunta global. Ver también las figuras 4.8 y 4.2.

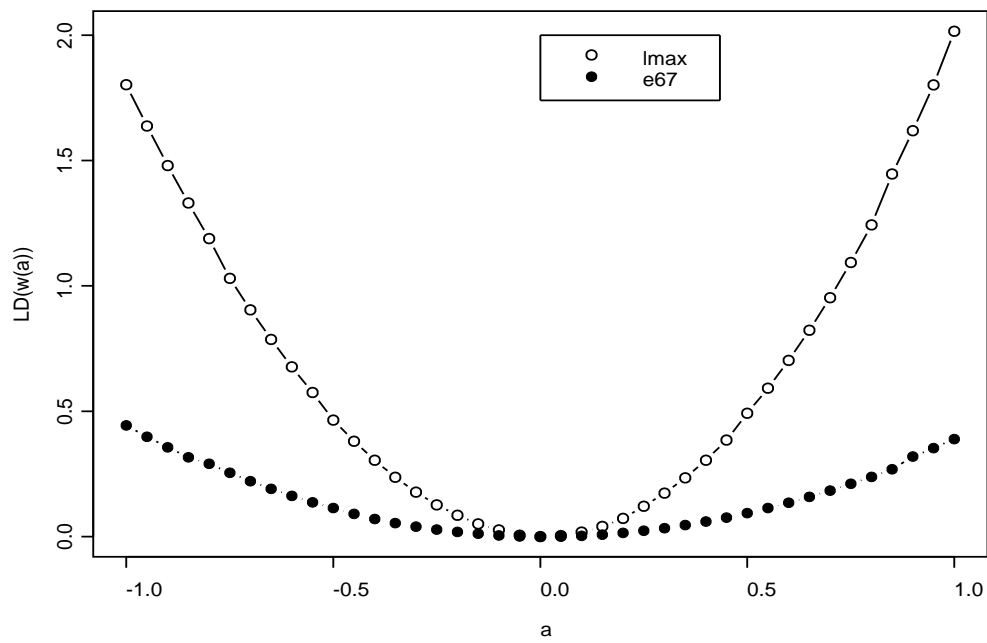


Figura 4.8: Gráficos del desplazamiento de verosimilitudes $LD_1(\omega(a))$ versus a para las direcciones \mathbf{l}_{\max} y \mathbf{e}_{67} – Aplicación 2.

Aplicación 3: Datos de Kelly Consideremos nuevamente el conjunto de datos reportados por Kelly (1984) y modificados por Kim (2000). Los datos consisten de mediciones pareadas de niveles de serum kanamycin en muestras de sangre de 20 recién nacidos. Kim (2000), considerando un modelo normal, detecta las observaciones 2, 17 y 18 como más influyentes en $\hat{\beta}$, el EMV de β .

Siguiendo el enfoque de Bolfarine y Arellano-Valle (1994), ajustamos un modelo t a los datos considerando varios valores de ν . La tabla 4.2 abajo presenta las estimas de los parámetros y los valores de la función de log-verosimilitud para algunos valores de ν . Como podemos ver desde la tabla, cuando $\nu = 1$ obtenemos el mejor ajuste.

Tabla 4.2: Estimaciones de los parámetros para algunos valores de ν

ν	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\phi}$	$\hat{\mu}$	$\hat{\phi}_x$	$l(\hat{\theta})$
1	2.922	0.803	1.767	18.872	7.674	-339.06
10	6.078	0.669	4.647	20.900	27.83	-2148.38
1000	7.027	0.624	5.373	21.583	33.277	-2829.59

La figura 4.9 muestra gráficos índice de l_{max} para $\nu = 1$, $\nu = 10$ y $\nu = 1000$ grados de libertad. Las observaciones 2, 17 y 18 parecen tener una moderada influencia en la estimación de θ en el modelo normal ($\nu \rightarrow \infty$). Por otro lado, la figura 4.10 presenta gráficos índice para C_i , los cuales indican que la observación 2 no es influyente en el modelo normal, sólo lo son las observaciones 17 y 18.

En el caso Cauchy ($\nu = 1$) y en el modelo t con bajo grados de libertad, esta influencia se reduce sustancialmente.

En las figuras 4.11 y 4.12 consideramos un estudio de influencia local con interés en β , también analizado en Kim (2000) para el modelo normal. Los resultados de Kim (2000) son similares a los de la figura 4.9, para $\nu = 1000$. Sin embargo, en el modelo t con bajo grados de libertad las observaciones tienen poca influencia, como se puede apreciar en las figuras 4.11 y 4.12.

Figura 4.9: Gráfico índice de l_{max} correspondiente a θ con $\nu = 1, 10, 1000$

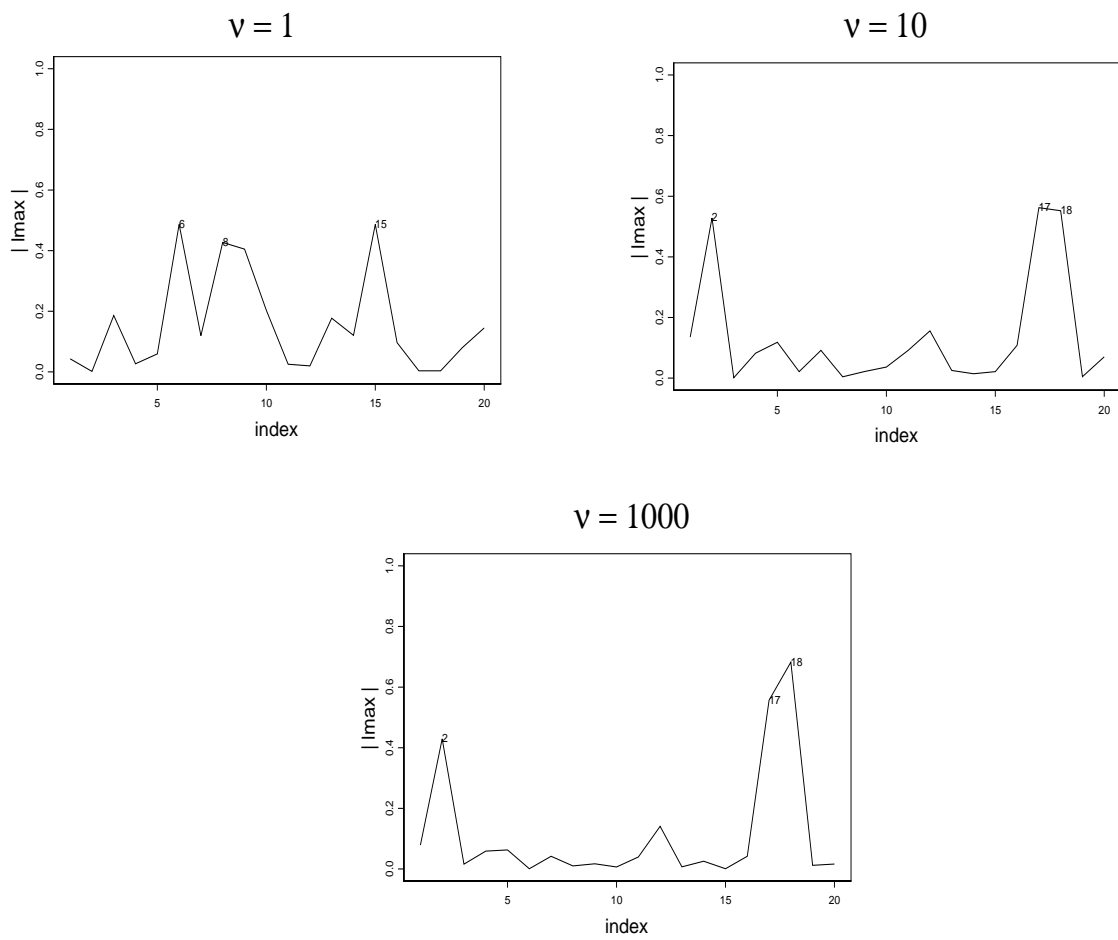


Figura 4.10: Gráfico índice para C_i correspondiente a θ , con $\nu = 1, 10, 1000$

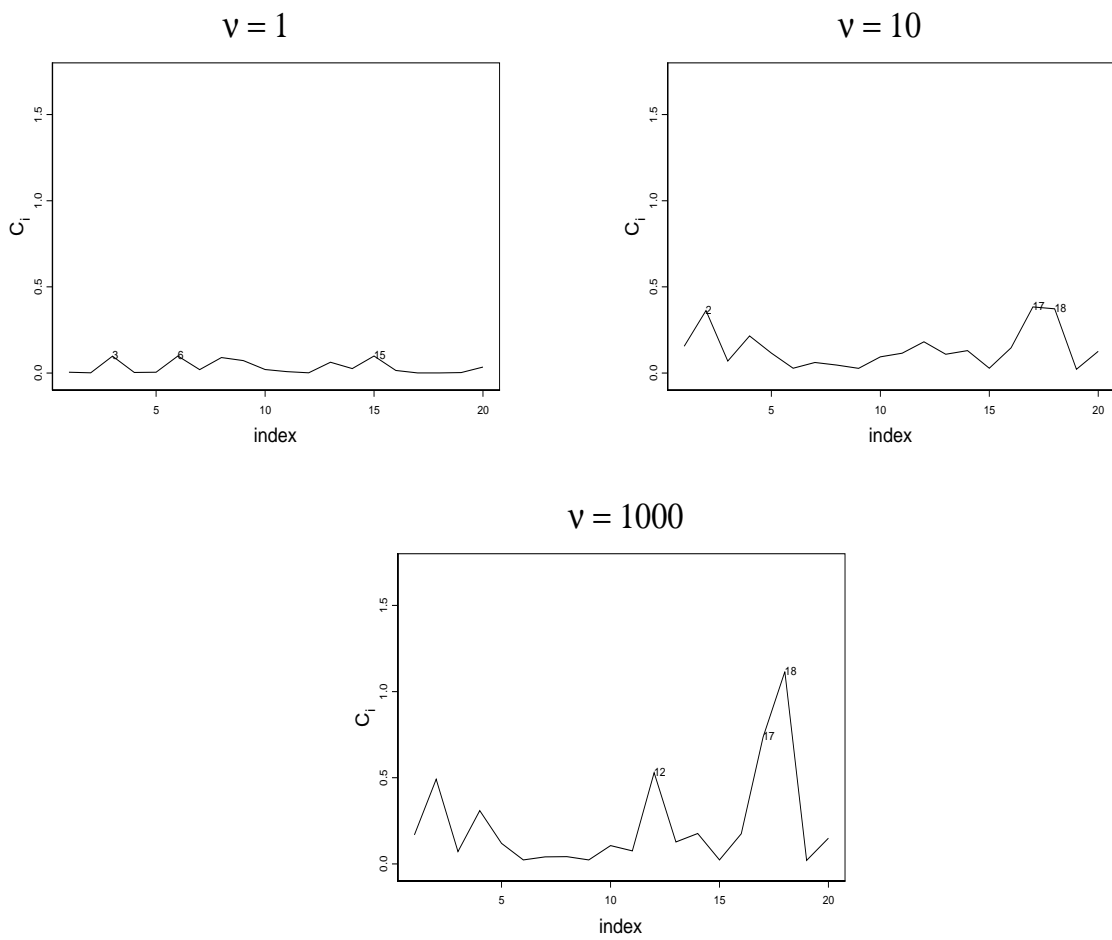


Figura 4.11: Gráfico índice para l_{max} correspondiente a β con $\nu = 1, 10, 1000$

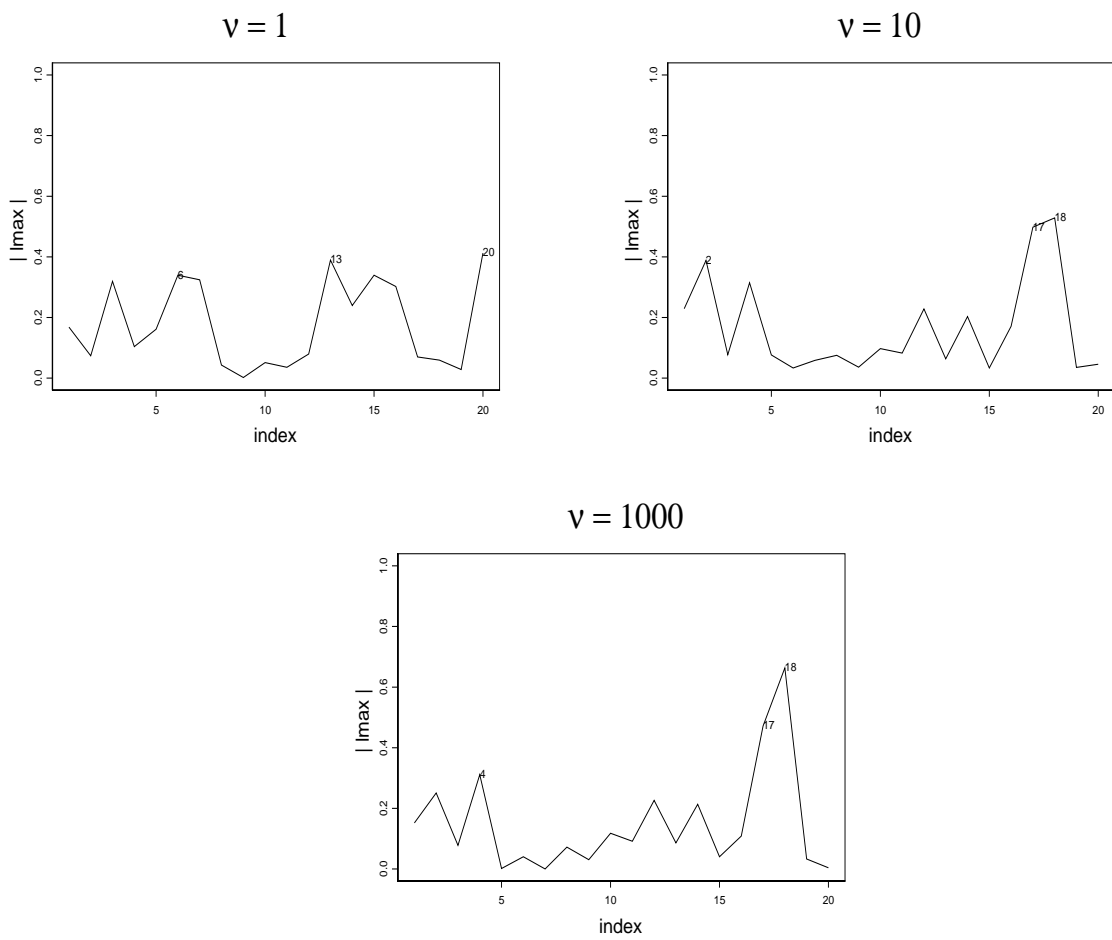
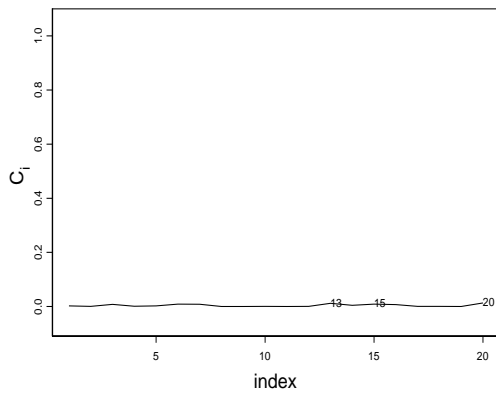
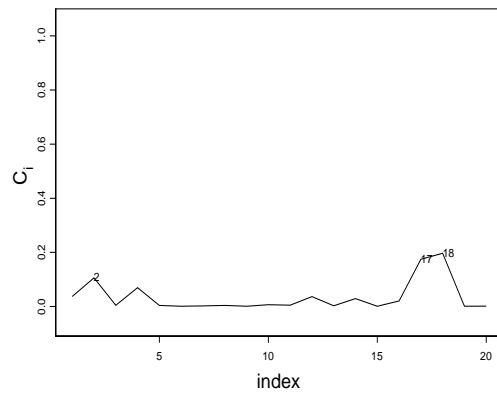


Figura 4.12: Gráfico índice para C_i correspondiente a β con $\nu = 1, 10, 1000$

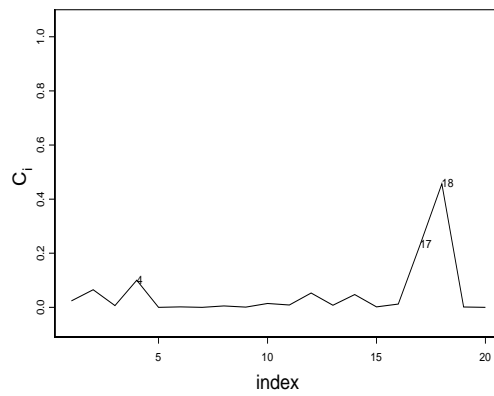
$\nu = 1$



$\nu = 10$



$\nu = 1000$



La tabla 4.3 considera influencia global para $\hat{\beta}$ usando el cambio relativo, $RC = |(\hat{\beta} - \hat{\beta}_{(I)})/\hat{\beta}|$, como considerado en Kim (2000), donde $\hat{\beta}_{(I)}$ denota el EMV de β después de eliminar las observaciones pertenecientes al conjunto de índices I . Al contrario de Kim (2000), al eliminar cualquiera de las observaciones 2, 17 y 18, el cambio relativo permanece prácticamente igual. Lo mismo ocurre cuando grupos de dos de las tres observaciones son eliminadas. Kim (2000) también señala que las observaciones 17 y 18 enmascaran la influencia de la observación 2 en el modelo estructural normal. Podemos notar que esto no parece ser el caso en el modelo estructural t con grados de libertad pequeños.

Tabla 4.3: Influencia global para $\hat{\beta}$

I	$\nu=1$		$\nu=10$		$\nu=1000$	
	$\hat{\beta}_{(I)}$	RC	$\hat{\beta}_{(I)}$	RC	$\hat{\beta}_{(I)}$	RC
2	0.92	0.14	0.75	0.13	0.66	0.06
17	0.93	0.15	0.78	0.17	0.69	0.11
18	0.93	0.15	0.80	0.19	0.73	0.16
2, 17	1.07	0.33	0.95	0.42	0.78	0.25
2, 18	1.07	0.33	0.96	0.43	0.83	0.33
17, 18	1.07	0.34	0.97	0.45	0.88	0.41
2, 17, 18	1.24	0.55	1.19	0.77	1.17	0.87

Del análisis de los residuos, denotados por e_i , basados en los EMV y definidos por Kim (2000), vemos que las observaciones 18, 12 y 17 presentan residuos grandes en valor absoluto, ver tabla 4.4. También podemos notar que el residuo, en valor absoluto, de la observación 2 es más grande que en el modelo normal. Sin embargo, eliminando las tres observaciones, y denotando los nuevos residuos por \hat{e}_i , vemos claramente que las observaciones 2, 17 y 18 son outliers, como fue también notado por Kim (2000). Para más detalles ver Galea et al. (2002b).

Tabla 4.4: Residuos para los modelos Cauchy y Normal

Obs	Cauchy		Normal	
	e_i	\hat{e}_i	e_i	\hat{e}_i
1	3.81	1.85	3.82	1.31
2	-3.58	-10.02	-1.75	-9.80
3	0.05	0.90	-1.09	-0.12
4	3.16	-0.25	3.76	-0.54
5	4.22	3.57	3.69	2.81
6	-0.88	-1.53	-1.41	-2.29
7	2.74	1.83	2.32	1.11
8	-0.90	-1.06	-1.62	-1.91
9	1.59	1.91	0.67	0.98
10	-2.58	-3.23	-3.11	-3.99
11	0.68	-2.77	1.30	-3.06
12	6.38	4.94	6.17	4.31
13	0.52	2.11	-0.93	0.96
14	-1.99	-1.50	-2.99	-2.45
15	-0.88	-1.53	-1.41	-2.29
16	2.78	5.12	1.04	3.85
17	-5.22	-12.01	-3.24	-11.73
18	-6.52	-13.75	-4.37	-13.40
19	-0.10	-0.26	-0.82	-1.11
20	1.60	3.67	-0.04	2.45

Apéndice A: Matriz de información observada para el modelo homocedástico

En este apéndice presentamos los elementos de la matriz de información observada $-\ddot{\mathbf{L}}_S$ dada en la sección 4.3.1. Sea

$$\mathbf{S}^* = \mathbf{S}_{10}\boldsymbol{\beta}' + \boldsymbol{\beta}\mathbf{S}_{01} + \mathbf{S}_{11}\boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{S}_{11} \quad (\text{A.1})$$

donde \mathbf{S}_{01} , \mathbf{S}_{10} y \mathbf{S}_{11} son sub-matrices correspondientes a la partición de

$$\mathbf{S} = \begin{pmatrix} s_{00} & \mathbf{S}_{01} \\ \mathbf{S}_{10} & \mathbf{S}_{11} \end{pmatrix},$$

con s_{00} un escalar y \mathbf{S}_{11} de dimension $p \times p$. La matriz $\ddot{\mathbf{L}}_S$ tiene los siguientes elementos :

$$\begin{aligned} \mathbf{L}_{\alpha\alpha} &= -\frac{n}{\phi} (\mathbf{I}_p - f\boldsymbol{\beta}\boldsymbol{\beta}'), \\ \mathbf{L}_{\alpha\beta} &= \bar{X}\mathbf{L}_{\alpha\alpha}, \quad \mathbf{L}_{\alpha\phi} = \mathbf{0}, \\ \mathbf{L}_{\alpha\mu_x} &= -\frac{n(1-fc)}{\phi}\boldsymbol{\beta}, \quad \mathbf{L}_{\alpha\phi_x} = \mathbf{0}, \\ \mathbf{L}_{\beta\beta} &= -\frac{nf}{\phi} [\phi(\mathbf{I}_p - 2f\boldsymbol{\beta}\boldsymbol{\beta}') + \frac{\bar{X}^2}{f}(\mathbf{I}_p - f\boldsymbol{\beta}\boldsymbol{\beta}')] \\ &\quad + f\mathbf{b}'\mathbf{S}\mathbf{b}(\mathbf{I}_p - 4f\boldsymbol{\beta}\boldsymbol{\beta}') - \mathbf{S}_{11} + 2f\mathbf{S}^*], \\ \mathbf{L}_{\beta\phi} &= \frac{nf^2}{\phi}\boldsymbol{\beta} - \frac{n}{\phi^2} [f(\mathbf{S}_{10} + \mathbf{S}_{11}\boldsymbol{\beta})(1 + \frac{f\phi}{\phi_x}) - f^2\mathbf{b}'\mathbf{S}\mathbf{b}(1 + \frac{2f\phi}{\phi_x})\boldsymbol{\beta}], \\ \mathbf{L}_{\beta\mu_x} &= \bar{X}\mathbf{L}_{\alpha\mu_x}, \\ \mathbf{L}_{\beta\phi_x} &= -\frac{fn(1-fc)}{\phi_x} [(1 + \frac{2f\mathbf{b}'\mathbf{S}\mathbf{b}}{\phi})\boldsymbol{\beta} - \frac{1}{\phi}(\mathbf{S}_{10} + \mathbf{S}_{11}\boldsymbol{\beta})], \\ \mathbf{L}_{\phi\phi} &= \frac{n}{\phi^2} [\frac{1}{2}(p + (1-fc)^2) + \frac{f^2}{\phi_x}\mathbf{b}'\mathbf{S}\mathbf{b}(1 + \frac{\phi f}{\phi_x}) - \frac{1}{\phi}(\text{tr}(\mathbf{S}) - f\mathbf{b}'\mathbf{S}\mathbf{b})], \\ \mathbf{L}_{\phi\mu_x} &= \mathbf{0}, \\ \mathbf{L}_{\phi\phi_x} &= \frac{nf}{2\phi_x} [\frac{fc}{\phi_x} - \frac{\mathbf{b}'\mathbf{S}\mathbf{b}}{\phi} (\frac{1-fc}{\phi} - \frac{f(2fc-1)}{\phi_x})], \\ \mathbf{L}_{\mu_x\mu_x} &= -\frac{nc}{\phi}(1-fc), \quad \mathbf{L}_{\mu_x\phi_x} = \mathbf{0} \text{ y} \\ \mathbf{L}_{\phi_x\phi_x} &= n(\frac{fc}{\phi_x})^2 [\frac{1}{2} + \frac{\mathbf{b}'\mathbf{S}\mathbf{b}}{c\phi}(fc-1)], \end{aligned}$$

donde \mathbf{S}^* como en (A.1) y todas las expresiones deben ser evaluadas en los EMV.

Apéndice B: Matriz de información observada para el modelo estructural t

En este apéndice presentamos los elementos de la matriz de información observada para el modelo estructura t . De (3.11), sigue que

$$\frac{\partial l_i(\boldsymbol{\theta})}{\partial \gamma} = -\frac{1}{2} \frac{\partial \log|\mathbf{V}|}{\partial \gamma} - \frac{1}{2} \frac{\nu + 2}{\nu + \delta_i(\boldsymbol{\theta})} d_{i\gamma}, \quad (\text{B.1})$$

con $d_{i\gamma} = \frac{\partial \delta_i(\boldsymbol{\theta})}{\partial \gamma}$, $\gamma = \alpha, \beta, \phi, \mu_x$ y ϕ_x , $\delta_i(\boldsymbol{\theta})$ como en (3.11), $i = 1, \dots, n$. Luego, tenemos que

$$\begin{aligned} \frac{\partial \log|\mathbf{V}|}{\partial \beta} &= 2\beta a, & \frac{\partial \log|\mathbf{V}|}{\partial \phi} &= \frac{1}{\phi} + \frac{a}{\phi_x}, \\ \frac{\partial \log|\mathbf{V}|}{\partial \phi_x} &= \frac{ac}{\phi_x}, & \frac{\partial \log|\mathbf{V}|}{\partial \gamma} &= 0, \quad \gamma = \alpha, \mu_x, \end{aligned} \quad (\text{B.2})$$

$$d_{i\alpha} = \frac{2}{\phi} \{a\beta q_{2i} - (Y_i - \alpha - \beta\mu_x)\}, \quad (\text{B.3})$$

$$d_{i\beta} = \frac{2}{\phi} \{a^2\beta q_{2i}^2 - aq_{2i}(Y_i - \alpha - 2\beta\mu_x) - \mu_x(Y_i - \alpha - \beta\mu_x)\}, \quad (\text{B.4})$$

$$d_{i\phi} = \frac{a}{\phi} q_{2i}^2 \left(\frac{1}{\phi} + \frac{a}{\phi_x} \right) - \frac{q_{1i}}{\phi^2}, \quad (\text{B.5})$$

$$d_{i\mu_x} = \frac{2}{\phi} q_{2i}(ac - 1), \quad (\text{B.6})$$

$$d_{i\phi_x} = \frac{aq_{2i}^2}{\phi\phi_x}(ac - 1), \quad (\text{B.7})$$

donde $c = 1 + \beta^2$, $a = \phi_x/(\phi + c\phi_x)$, $q_{1i} = (Y_i - \alpha - \beta\mu_x)^2 + (X_i - \mu_x)^2$ y $q_{2i} = \beta(Y_i - \alpha - \beta\mu_x) + (X_i - \mu_x)$, $i = 1, \dots, n$.

Desde (B.1) vemos que la matriz de información observada, por elemento, esta dada por

$$\ddot{\mathbf{L}}_i(\boldsymbol{\theta}) = - \left[\left(\frac{\partial^2 l_i}{\partial \gamma \partial \tau} \right) \right], \quad (\text{B.8})$$

donde

$$\frac{\partial^2 l_i}{\partial \gamma \partial \tau} = -\frac{1}{2} \frac{\partial^2 \log|\mathbf{V}|}{\partial \gamma \partial \tau} - \frac{1}{2} \frac{\nu + 2}{\nu + \delta_i(\boldsymbol{\theta})} \left\{ d_{i\gamma\tau} - \frac{d_{i\gamma} d_{i\tau}}{\nu + \delta_i(\boldsymbol{\theta})} \right\},$$

con $d_{i\gamma}$, $\gamma = \alpha, \beta, \phi, \mu_x, \phi_x$ como en (B.3)-(B.7) y $d_{i\gamma\tau} = \frac{\partial^2 \delta_i(\boldsymbol{\theta})}{\partial \gamma \partial \tau}$ $\gamma, \tau = \alpha, \beta, \phi, \mu_x, \phi_x$, donde

$$\frac{\partial^2 \log|\mathbf{V}|}{\partial \alpha \partial \gamma} = \frac{\partial^2 \log|\mathbf{V}|}{\partial \mu_x \partial \gamma} = 0, \quad \gamma = \alpha, \beta, \phi, \mu_x, \phi_x, \quad (\text{B.9})$$

$$\frac{\partial^2 \log|\mathbf{V}|}{\partial \beta \partial \beta} = 2a(1 - 2a\beta^2), \quad (\text{B.10})$$

$$\frac{\partial^2 \log|\mathbf{V}|}{\partial \beta \partial \phi} = -2\beta a^2 / \phi_x, \quad (\text{B.11})$$

$$\frac{\partial^2 \log|\mathbf{V}|}{\partial \beta \partial \phi_x} = 2 \frac{a\beta}{\phi_x} (1 - ac), \quad (\text{B.12})$$

$$\frac{\partial^2 \log|\mathbf{V}|}{\partial \phi \partial \phi} = -\left(\frac{1}{\phi^2} + \frac{a^2}{\phi_x^2}\right), \quad (\text{B.13})$$

$$\frac{\partial^2 \log|\mathbf{V}|}{\partial \phi \partial \phi_x} = -\frac{a^2 c}{\phi_x^2}, \quad (\text{B.14})$$

$$\frac{\partial^2 \log|\mathbf{V}|}{\partial \phi_x \partial \phi_x} = -\left(\frac{ac}{\phi_x}\right)^2, \quad (\text{B.15})$$

$$d_{i\alpha\alpha} = \frac{2}{\phi}(1 - a\beta^2), \quad (\text{B.16})$$

$$d_{i\alpha\beta} = \frac{2}{\phi}\{a\beta(Y_i - \alpha - 2\beta\mu_x) + aq_{2i}(1 - 2a\beta^2) + \mu_x\}, \quad (\text{B.17})$$

$$d_{i\alpha\phi} = \frac{2}{\phi}\left\{\frac{1}{\phi}(Y_i - \alpha - \beta\mu_x) - a\beta q_{2i}\left(\frac{1}{\phi} + \frac{a}{\phi_x}\right)\right\}, \quad (\text{B.18})$$

$$d_{i\alpha\mu_x} = \frac{2\beta}{\phi}(1 - ac), \quad (\text{B.19})$$

$$d_{i\alpha\phi_x} = \frac{2a\beta}{\phi\phi_x}q_{2i}(1 - ac), \quad (\text{B.20})$$

$$d_{i\beta\beta} = \frac{2}{\phi}\{\mu_x^2 + 2a\mu_x q_{2i} - a(Y_i - \alpha - 2\beta\mu_x)^2 \quad (\text{B.21})$$

$$+ 4a^2\beta q_{2i}(Y_i - \alpha - 2\beta\mu_x) + a^2(1 - 4a\beta^2)q_{2i}^2\},$$

$$d_{i\beta\phi} = \frac{2}{\phi}\{aq_{2i}(Y_i - \alpha - 2\beta\mu_x)\left(\frac{1}{\phi} + \frac{a}{\phi_x}\right) - a^2\beta q_{2i}^2\left(\frac{1}{\phi} + \frac{2a}{\phi_x}\right) \quad (\text{B.22})$$

$$+ \frac{\mu_x}{\phi}(Y_i - \alpha - \beta\mu_x)\},$$

$$d_{i\beta\mu_x} = \frac{2(1 - ac)}{\phi}\{2a\beta q_{2i} - (Y_i - \alpha - 2\beta\mu_x)\}, \quad (\text{B.23})$$

$$d_{i\beta\phi_x} = \frac{2a(1 - ac)}{\phi\phi_x}\{2a\beta q_{2i}^2 - q_{2i}(Y_i - \alpha - 2\beta\mu_x)\}, \quad (\text{B.24})$$

$$d_{i\phi\phi} = \frac{2}{\phi^2}d_i(\boldsymbol{\theta}) - \frac{2a^2}{\phi\phi_x}q_{2i}^2\left(\frac{1}{\phi} + \frac{a}{\phi_x}\right), \quad (\text{B.25})$$

$$d_{i\phi\mu_x} = \frac{2}{\phi^2}\{q_{2i} - ac(1 + \frac{a\phi}{\phi_x})q_{2i}\}, \quad (\text{B.26})$$

$$d_{i\phi\phi_x} = \frac{a}{\phi\phi_x}\left(\frac{1 - ac}{\phi} + \frac{a - 2a^2c}{\phi_x}\right)q_{2i}^2, \quad (\text{B.27})$$

$$d_{i\mu_x\mu_x} = 2c(1 - ac)/\phi, \quad (\text{B.28})$$

$$d_{i\mu_x\phi_x} = \frac{2ac(1 - ac)}{\phi\phi_x}q_{2i}, \quad (\text{B.29})$$

$$d_{i\phi_x\phi_x} = \frac{2a^2c(1 - ac)}{\phi\phi_x^2}q_{2i}^2, \quad (\text{B.30})$$

$i = 1, \dots, n$. Finalmente, la matriz de información observada es $\ddot{\mathbf{L}} = \sum_{i=1}^n \ddot{\mathbf{L}}_i(\boldsymbol{\theta})$.

Apéndice C: Matriz de información esperada para el modelo heterocedástico

Los elementos de la matriz de información esperada para el modelo heterocedástico están dados por:

$$I_{\mu_x \mu_x} = n(1 - 1/v\sigma_x^2)/\sigma_x^2, \quad I_{\mu_x \alpha_k} = \frac{n\beta_k}{v\sigma_k^2\sigma_x^2}, \quad I_{\mu_x \beta_k} = \frac{n\beta_k\mu_k}{v\sigma_k^2\sigma_x^2}, \quad I_{\mu_x \sigma_x^2} = I_{\mu_x \sigma_j^2} = 0, \quad (\text{C.1})$$

con $k = 1, \dots, p$, y $j = 0, \dots, p$.

$$I_{\alpha_k \alpha_k} = \frac{n}{\sigma_k^2} \left(1 - \frac{\beta_k^2}{v\sigma_k^2}\right), \quad I_{\alpha_k \alpha_l} = -\frac{n\beta_k\beta_l}{v\sigma_k^2\sigma_l^2}, \quad I_{\alpha_k \beta_k} = -\frac{n\beta_k^2\mu_k}{v\sigma_k^4} + \frac{n\mu_x}{\sigma_x^2}, \quad I_{\alpha_k \beta_l} = -\frac{n\beta_k\beta_l\mu_k}{v\sigma_k^2\sigma_l^2}, \quad (\text{C.2})$$

$$I_{\alpha_k \sigma_x^2} = I_{\alpha_k \sigma_j^2} = I_{\alpha_k \sigma_l^2}, \quad k = 1, \dots, p; \quad j = 0, \dots, p.$$

$$I_{\beta_k \beta_k} = \frac{n}{\sigma_k^2} (\mu_x^2 + \sigma_x^2 - 1/v) + \frac{n\beta_k^2}{v\sigma_k^4} (2/v - \sigma_x^2 - \mu_x^2), \quad I_{\beta_k \beta_l} = \frac{n\beta_k\beta_l}{v\sigma_k^2\sigma_l^2} \{2/v - (\mu_x^2 + \sigma_x^2)\}, \quad (\text{C.3})$$

$$I_{\beta_k \sigma_x^2} = \frac{n\beta_k}{v\sigma_k^2\sigma_x^4} (\sigma_x^2 - 1/v), \quad I_{\beta_k \sigma_0^2} = -\frac{n\beta_k}{v^2\sigma_k^2\sigma_0^4}, \quad I_{\beta_k \sigma_k^2} = \frac{n\beta_k}{v\sigma_k^4} - \frac{n\beta_k^3}{v^2\sigma_k^6},$$

$$I_{\beta_k \sigma_l^2} = -\frac{n\beta_k\beta_l^2}{v^2\sigma_k^2\sigma_l^4}, \quad k \neq l = 1, \dots, p.$$

$$I_{\sigma_x^2 \sigma_x^2} = \frac{n}{2\sigma_x^4} - \frac{n}{v\sigma_x^6} + \frac{n}{2v\sigma_x^8}, \quad I_{\sigma_x^2 \sigma_0^2} = \frac{n}{2v^2\sigma_x^4\sigma_0^4}, \quad I_{\sigma_x^2 \sigma_k^2} = \frac{n\beta_k^2}{2v^2\sigma_x^4\sigma_k^4}, \quad k = 1, \dots, p; \quad (\text{C.4})$$

$$I_{\sigma_0^2 \sigma_0^2} = \frac{n}{2\sigma_0^4} - \frac{n}{v\sigma_0^6} + \frac{n}{2v^2\sigma_0^8}, \quad I_{\sigma_0^2 \sigma_k^2} = \frac{n\beta_k^2}{2v^2\sigma_0^4\sigma_k^4}, \quad k = 1, \dots, p; \quad (\text{C.5})$$

$$I_{\sigma_k^2 \sigma_k^2} = \frac{n}{2\sigma_k^4} - \frac{n\beta_k^2}{v\sigma_k^6} + \frac{n\beta_k^4}{2v^2\sigma_k^8}, \quad I_{\sigma_k^2 \sigma_l^2} = \frac{n\beta_k^2\beta_l^2}{2v^2\sigma_k^4\sigma_l^4}, \quad k \neq l = 1, \dots, p. \quad (\text{C.6})$$

Agradecimientos

Agradezco a la Comisión Organizadora Local del XXXV Coloquio de la Sociedad Argentina de Estadística por la invitación a dictar este Minicurso. Parte del financiamiento para la elaboración de estas notas proviene del Proyecto Fondecyt 1070919.

Referencias

- Abdullah, M. B. (1995). Detection of influential observations in functional errors-in-variables model. *Communications in Statistics: Theory and Method*, 24, 1585-1595.
- Arellano-Valle, R. (1994). *Distribuições Elípticas: Propriedades, Inferência e Aplicações a Modelos de Regressão*, Tese, IME-USP.
- Arellano-Valle, R. y Bolfarine, H. (1996a). A note on the simple structural regression model. *Annals of the Institute of Statistical Mathematics*, 48, 111-125.
- Arellano-Valle, R. y Bolfarine, H. (1996b). Elliptical structural models. *Communications in Statistics: Theory and Method*, 25, 2319-2341.
- Arellano-Valle, R. y Bolfarine, H. (1998). On score tests in structural regression models. *Statistics*, 131-149.
- Barnett, V. D. (1969). Simultaneous pairwise linear structural relationships. *Biometrics*, 25, 129-142.
- Barnett, V. y Lewis, T. (1994). *Outliers in statistical data*. 3rd Edition. John Wiley.
- Belsley, D.A., Kuh, E. y Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley. New York.
- Bennett, B.M. (1961). On a certain multivariate non-normal distribution. *Proceedings of the Cambridge Philosophical Society*, 57, 434-436.
- Berkane, M. y Bentler, P.M. (1987). Characterizing parameters of multivariate elliptical distributions. *Communications in Statistics- Simulation and Computation*, 16, 193-198.
- Berkane, M., Kano, Y. y Bentler, P.M. (1994). Pseudo Maximum Likelihood Estimation in Elliptical Theory: Effects of misspecification. *Computational Statistics and Data Analysis*, 18, 255-267.
- Bolfarine, H. y Arellano-Valle, R. B. (1994). Robust Modeling in measurement error models using the Student-t distributions. *Brazilian Journal of Probability and Statistics*, 8, 67-84.

- Bolfarine, H. y Galea, M. (1995). Maximum Likelihood Estimation of Simultaneous Pair-wise Linear Structural Relationships. *Biometrical Journal*, 37, 673-689.
- Bolfarine, H. y Galea, M. (1996). One Structural Comparative Calibration under a t-Models. *Computational Statistics*, 11, 63-85.
- Chatterjee, S. y Hadi, A.S. (1988). *Sensitivity Analysis in Linear Regression*. John Wiley. New York.
- Cheng, C.L. y Van Ness, J.W. (1990). Bounded influence errors- in-variables regression. *Contemporary Mathematics*, 112, 227-241.
- Cheng, C.-L. y VanNess, J. (1999). *Statistical Regression With Measurement Error*. Arnold, London.
- Chipkevitch, E., Nishimura, R. T., Tu, D. G. S. y Galea, M. (1996). Clinical Measurement of Testicular Volume in Adolescents: Comparison of the Reliability of 5 Methods. *Journal of Urology*, 156, 2050-2053.
- Cochran, W.G. (1968). Errors of measurement in statistics. *Technometrics*, 10, 637-666.
- Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, B*, 48, 133-169.
- Cook, R. D. y Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall. London.
- Dempster, A., Laird, N., y Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, B*, 39, 1-38.
- Dolby, G.R. (1976). The ultrastructural relation: A synthesis of the functional and structural relations. *Biometrika*, 63, 39-50.
- Draper, N. y Guttman, I. (1975). Two simultaneous measurement procedures: A bayesian approach. *Journal of the American Statistical Association*, 70, 43-46.
- Draper, N. y Guttman, I. (1976). K simultaneous measurement procedures: A bayesian approach. *Journal of the American Statistical Association*, 71, 605-607.
- Dunn, G. (1992). *Design and Analysis of Reliability: The statistical evaluation of measurement errors*. Edward Arnold. New York.
- Dunn, G. (2004). *Statistical Evaluation of Measurement Errors: Design and Analysis of Reliability Studies*. Edward Arnold. New York.

- Dunnett, C. W. y Sobel, M. (1954). A bivariate generalization of student's t-distribution, with tables for certain special cases. *Biometrika*, 41, 153-169.
- Escobar, L. y Meeker, W. (1992). Assessing influence in regression analysis with censored data. *Biometrics*, 48, 507-528.
- Fang, K.T. y Anderson, T.W. (1990). *Statistical Inference in Elliptical Contoured and Related Distributions*. Allerton Press Inc.. New York.
- Fang, K.T., Kotz, S. y Ng, K.W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall. London.
- Fang, K.T. y Zhang, Y.T. (1990). *Generalized Multivariate Analysis*. Springer-Verlag. London.
- Fernández, C. y Steel, M. (1999). Multivariate Student-t Regression Models: Pitfalls and Inference. *Biometrika*, 86, 156-167.
- Fuller, W. A. (1987). *Measurement error models*. Wiley, New York.
- Fung, W. K. y Kwan, C. W. (1997). A Note on Local Influence Based on Normal Curvature. *Journal of the Royal Statistical Society - Series B*, 59, 839-843.
- Gabrielsen, A. (1978). Consistency and identifiability. *Journal of Econometrics*, 8, 261-263.
- Galea, M., Paula, G. A. y Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician*, 46, 71-79.
- Galea, M., Bolfarine, H. y de Castro, M. (2002a). Local influence in comparative calibrations models. *Biometrical Journal*, 44, 59-81.
- Galea, M., Bolfarine, H. y Vilca, F. (2002b). Influence diagnostics for the structural errors-in-variables model under the Student-t distribution. *Journal of Applied Statistics*, 29, 1191-1204.
- Galea, M., Paula, G. A. y Uribe, M. (2003). Influence diagnostics in univariate linear elliptical regression model. *Statistical Paper*, 44, 23-45.
- Galea, M., Bolfarine H. y Vilca, F. (2005). Local influence in comparative calibration models under elliptical-t distributions. *Biometrical Journal*, 47, 691-706.
- Gleser, L.J. (1981). Estimation in a multivariate "error in variables" regression model: Large sample results. *The Annals of Statistics*, 9, 24-44.
- Graybill, F.A. (1983). *Matrices with applications in statistics*. Second Edition, Wadsworth.

- Grubbs, F. E. (1948). On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association*, 43, 243-264.
- Grubbs, F. E. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics* 15, 53-66.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. y Stahel, W.A. (1986). *Robust Statistics*. John Wiley. New York.
- Hawkins, D.M. (1980). *Identification of Outliers*, Chapman and Hall. London.
- Huber, P.J. (1981). *Robust Statistics*. John Wiley. New York.
- Jaech, J. L. (1985). *Statistical analysis of measurement errors*. Exxon Monographs. John Wiley, New York.
- Johnson, N. y Kotz, S. (1970). *Distributions in statistics: continuous univariate distributions - 2*. New York. Houghton Mifflin.
- Kaaks, R., Riboli, E., Estive, J., van Kappel y van Staveren (1994). Estimating the accuracy of questionnaire assessments: Validation in terms of structural equation models. *Statistics in Medicine*, 13, 127-142.
- Kelly, G. (1984). The influence function in the errors-in-variables problem. *The Annals of Statistics*, 12, 87-100.
- Kelly, G. (1985). Use of the structural equations model in assessing the reliability of a new measurement technique. *Applied Statistics*, 34, 258-263.
- Kendall, M.G. y Stuart, A. (1979). *The Advanced Theory of Statistics*, 2, 4th ed. Hafner. New York.
- Kent, J.T., Tyler, D.E. y Vardi, Y. (1994). A curious likelihood identity for the multivariate t distribution. *Communications in Statistics - Simulation and Computation*, 23, 441-453.
- Kim, M. G. (2000). Outliers and influential observations in the structural errors-in-variables model. *Journal of Applied Statistics*, 24, 461-473.
- Kimura, D.K. (1992). Functional comparative calibration using an EM algorithm. *Biometrics*, 48, 1263-1271.
- Lange, K. L., Little, R.J. y Taylor, J. (1989). Robust statistical modelling using the t -distribution. *Journal of the American Statistical Association*, 84, 881-896.

- Lee, A. H. y Zhao, Y. (1996). Assessing local influence in measurement error models. *Biometrical Journal*, 38(7), 829-841.
- Lesaffre, E. y Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, 54, 570-582.
- Leurgans, S. (1980). Evaluating laboratory measurement techniques. *In Biostatistics Casebook*. Edited by Miller, R., Efron, B., Brown, B. and Moses, L. John Wiley.
- Little, R.J. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics*, 37,23-38.
- Little, R.J. y Rubin, D.B. (1987). *Statistical Analysis with Missing Data*, John Wiley. New York.
- Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519-530.
- Maronna, R.A. (1976). Robust M-estimadors of multivariate location and scatter. *The Annals of Statistics*, 4, 51-67.
- Maronna, R.A., Martin, R.D. y Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*. John Wiley. New York.
- Paula, G. A. (1993). Assessing Local Influence in Restricted Regression Models. *Computational Statistics & Data Analysis*, 16, 63-79.
- Peña, D. y Yohai, V. (1995). The detection of influential subsets in linear regression by using an influence matrix. *Journal of the Royal Statistical Society, B*, 57, 145-156.
- Pitman, E.G. (1939). A note on normal correlation. *Biometrika*, 31, 9-12.
- Rousseeuw, P.J. y Leroy, A.M. (1987). *Robust Regression and Outlier Detection*, John Wiley. New York.
- Rubin, D.B. (1983). Iteratively reweighted least squares. *Encyclopedia of the Statistical Sciences*, 4, 272-275.
- Shyr, J. (1984). Comparative precision in linear structural relationships. *Technical Report 84-30*. Department of Statistics Purdue University, USA.
- Shyr, J.Y. y Gleser, L.J. (1986). Inference about comparative precision in linear structural relationships. *Journal of the Statistical Planning and Inference*, 14, 339-358.
- Staudte, R.G. y Sheather, S.J. (1990). *Robust Estimation and Testing*. John Wiley. New York.

- Sutradhar, B.C. (1993). Score tests for the covariance matrix of the Elliptical t -distribution. *Journal of Multivariate Analysis*, 46, 1-12.
- Sutradhar, B.C. y Ali, M.M. (1986). Estimation of the parameter of a regression model with a multivariate t error variable. *Communications in Statistics - Theory and Methods*, 15, 429-450.
- Taylor, J. (1992). Properties of modelling the error distribution with an extra shape parameter. *Computational Statistics & Data Analysis*, 13, 33-46.
- Theobald, C. M. y Mallison, J. R. (1978). Comparative calibration, linear structural relationship and congeneric measurements. *Biometrics*, 34, 39-45.
- Tong, Y.L. (1990). *The multivariate Normal Distribution*. Springer-Verlag. New York.
- Verbeke, G. y Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer, New York.
- Wellman, M. J. y Gunst, R.F. (1991). Influence diagnostics for linear measurement error models. *Biometrika*, 78, 373-380.
- Wong, M.Y. (1989). Likelihood estimation of a simple linear regression model when both variables have error. *Biometrika*, 76, 141-148.
- Wu, X. y Luo, Z. (1993). Second-order Approach to Local Influence. *Journal of the Royal Statistical Society - Series B*, 55, 929-936.
- Zamar, R.H. (1989). Robust estimation in the errors-in-variables model. *Biometrika*, 76, 149-160.
- Zhao, Y. y Lee, A. H. (1998). Influence diagnostics for simultaneous equations models. *Australian and New Zealand Journal of Statistics*, 40, 345-357.