



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Matemática

Tesis de Licenciatura

**Estimadores para modelos con errores en las variables que describen
un problema metrológico**

Manuel Eduardo Benjamín

Directora: Dra. Daniela Rodríguez

Junio 2014

Índice general

Índice general	I
1. Introducción	1
1.1. Problema metrológico	2
1.2. Formulación matemática	3
2. Conceptos previos	5
2.1. Matrices y vectores aleatorios	5
2.2. Matrices definidas positivas	7
2.3. Error cuadrático medio	9
2.3.1. Caso univariado	9
2.3.2. Caso multivariado	9
2.4. Modelo lineal clásico	10
2.4.1. Cuadrados mínimos generalizados	12
2.4.2. Cuadrados mínimos factibles	14
3. Errores en las variables	15
3.1. Modelos con errores en las variables	15
3.2. Regresión lineal con errores en las variables	17
3.3. Problema real como un problema de MEV	20
3.4. Problema a estudiar	22
3.5. Estimador de sesgo corregido	23
3.6. Estimación por cuadrados mínimos factibles	25
3.7. Estimadores de cuadrados mínimos	26
3.7.1. Cuadrados mínimos bajo homoestabilidad	27
3.7.2. Cuadrados mínimos bajo heteroscedasticidad	29
3.7.3. Estimador de cuadrados mínimos generalizado	31
3.7.4. Caso general e implementación del estimador	33
4. Bootstrap	37
4.1. Estimación de Media, varianza, distribución y cuantiles	38
4.2. Estimación del ECM e intervalos de confianza	38
4.3. Bootstrap para estimación multivariada	40
4.4. Estimación de una región de confianza	41
4.5. Estimación de la distribución de una muestra	41
4.6. Aplicación a los modelos lineales con errores en las variables	42
5. Mediciones reales y simulaciones	44

5.1. Simulación de un modelo lineal	44
5.2. Resultados de las mediciones	45
5.3. Validación de los resultados	46

Capítulo 1

Introducción

La metrología es la ciencia de las mediciones y plantea, desde un punto de vista matemático, que los elementos que componen una medición son

- un mesurando z ,
- argumentos x_1, \dots, x_n ,
- un modelo que relaciona determinísticamente a los anteriores
- observaciones X_1, \dots, X_n de los argumentos bajo un error aleatorio
- un valor estimado Z del mesurando
- una medida de la incertidumbre de la estimación.

El documento “Guide to the Expression of Uncertainty in Measurement ” [5] también conocido como GUM es la principal referencia internacional en cuanto a la evaluación y expresión de incertidumbres de medición, y ha dotado a la comunidad metrológica de una metodología común para encarar esta tarea. La guía hace especial hincapié en lo que llama mediciones directas que posiblemente, sean el tipo de mediciones con las que se trabaja más frecuentemente. En éstas, el modelo es determinado por una función f conocida

$$z = f(x_1, \dots, x_n) \quad (1.1)$$

la estimación de Z se obtiene reemplazando las observaciones de los x_i en el modelo

$$Z = f(X_1, \dots, X_n) \quad (1.2)$$

y la medida de la incertidumbre de Z se expresa con su desvío estándar o su varianza. Como el cálculo analítico de éstas puede ser muy difícil o hasta imposible la GUM sugiere el uso de lo que llama el método de propagación de la incertidumbre que en la literatura

de estadística se conoce como el método delta.

Para una medición escalar, bajo la hipótesis de que las observaciones de los argumentos se realizan bajo un error aleatorio aditivo

$$(X_1, \dots, X_n) = (x_1, \dots, x_n) + (\xi_1, \dots, \xi_n) \quad (1.3)$$

el método de propagación de la incertidumbre dice que es una buena aproximación tomar

$$\text{Var}[Z] = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \text{Cov}[\xi_i, \xi_j] \quad (1.4)$$

donde las derivadas parciales están evaluadas en (X_1, \dots, X_n) .

Para el caso particular en que los errores no están correlacionados la ecuación anterior queda

$$\text{Var}[Z] = \sum_{i=1}^n \frac{\partial f}{\partial x_i}^2 \text{Var}[\xi_i]. \quad (1.5)$$

Para justificar el método basta calcular la varianza del el polinomio de Taylor de orden 1 centrado en (x_1, \dots, x_n) y evaluado en (X_1, \dots, X_n) . Las derivadas de la función están evaluadas en (x_1, \dots, x_n) pero como estos valores son desconocidos se aproximan evaluandolas en (X_1, \dots, X_n) .

El método de la propagación de la incertidumbre es generalizable a mesurandos multivariados y se desarrolla detalladamente en el suplemento 2 de la GUM [6].

1.1. Problema metrológico

El laboratorio de Radio Frecuencia y Microondas del Instituto Nacional de Tecnología Industrial es un laboratorio metrológico donde actualmente se trabaja en el desarrollo de una técnica para la medición del coeficiente de reflexión (Γ_g) de un generador de radiofrecuencia.

El coeficiente de reflexión es un valor complejo que caracteriza el cociente de la energía que incide sobre la que refleja el generador. Una correcta estimación de su valor tiene aplicaciones concretas como, por ejemplo, optimizar el rendimiento de una antena.

Salvo en generadores específicos no existen técnicas de medición directa de Γ_g . El método desarrollado en el laboratorio es una técnica que determina el valor buscado como la solución de un sistema de ecuaciones.



FIGURA 1.1: Antena emisora-receptora ARSAT

1.2. Formulación matemática

Argumentos físicos establecen que el valor de $\Gamma_g \in \mathbb{C}$ de un generador se relaciona con ciertos valores $p_i \in \mathbb{R}_{>0}$, $c_i \in \mathbb{C}$ con $1 \leq i \leq 7$ mediante las igualdades

$$\frac{p_i}{p_k} = \frac{|1 - c_k \Gamma_g|^2}{|1 - c_i \Gamma_g|^2} \quad i \neq k. \quad (1.6)$$

Si se piensa el valor de Γ_g como incógnita, esto da lugar a $21 = \binom{7}{2}$ ecuaciones distintas que comparten como única solución común al valor que se desea determinar. Los valores de c_i y p_i , no son conocidos y sólo se cuenta con mediciones directas de estos. Como consecuencia de esto el sistema de ecuaciones resultante tiene conjunto solución vacío cuando se reemplazan los valores verdaderos por los medidos.

El problema a resolver es: **Estimar Γ_g a partir de las estimaciones C_i, P_i .**

Para trabajar con un modelo mas sencillo se simplifican algunos términos de la igualdad con un argumento utilizado frecuentemente en metrología de alta frecuencia.

De la siguiente igualdad

$$|1 - c_i \Gamma_g|^2 = 1 - 2\text{Re}[c_i] \text{Re}[\Gamma_g] + 2\text{Im}[c_i] \text{Im}[\Gamma_g] + |c_i|^2 |\Gamma_g|^2 \quad (1.7)$$

se desprecia el termino cuadrático ya que se sabe por consideraciones físicas que $|c_i|$ y $|\Gamma_g|$ tiene módulos mucho mas chicos que 1. Luego puede utilizarse el siguiente modelo en vez del (1.6)

$$\frac{p_i}{p_k} = \frac{1 - 2\text{Re}[c_k] \text{Re}[\Gamma_g] + 2\text{Im}[c_k] \text{Im}[\Gamma_g]}{1 - 2\text{Re}[c_i] \text{Re}[\Gamma_g] + 2\text{Im}[c_i] \text{Im}[\Gamma_g]} \quad (1.8)$$

a partir de éste puede llegarse a un conjunto de ecuaciones que definen de manera implícita la parte real e imaginaria de Γ_g

$$z(i, k) = \operatorname{Re} [\Gamma_g] x(i, k) + \operatorname{Im} [\Gamma_g] y(i, k), \quad (1.9)$$

donde los $z(i, k)$, $x(i, k)$, $y(i, k)$ son el resultado de evaluar las siguientes funciones escalares

$$\begin{aligned} z(i, k) &= z(p_i, p_k, c_i, c_k) = \frac{1}{2} \left(1 - \frac{p_i}{p_k} \right) \\ x(i, k) &= x(p_i, p_k, c_i, c_k) = \operatorname{Re} \left[\frac{p_i}{p_k} c_i - c_k \right] \\ y(i, k) &= y(p_i, p_k, c_i, c_k) = \operatorname{Im} \left[c_k - \frac{p_i}{p_k} c_i \right]. \end{aligned} \quad (1.10)$$

Una dificultad adicional del problema es que las mediciones de los p_i c_i no son independientes. Esto es consecuencia de que están definidos por argumentos compartidos y que son observados una sola vez para todos. Esto es, para los argumentos $s \in \mathbb{R}^{18}$, $m_i \in \mathbb{R}^5$ con $1 \leq i \leq 7$

$$\begin{aligned} c_i &= f(s, m_i) \\ p_i &= h(s, m_i) \end{aligned} \quad (1.11)$$

donde s es el mismo para todo i . Luego, cuando se observan estos argumentos bajo sus respectivos errores resulta que las mediciones no son independientes y posiblemente estén correlacionadas.

El objetivo de esta tesis es desarrollar un método que permita resolver el problema planteado. Para esto, en el Capítulo 2 introducimos conceptos necesarios y estudiamos algunos aspectos de regresión lineal clásica cuyo conocimiento ayudan a una mejor comprensión de las ideas que presentamos en el Capítulo 3 donde estudiamos los modelos con errores en las variables con especial atención en los modelos lineales con errores en las variables. Para estos últimos presentamos tres estimadores que desarrollamos para un caso general sobre el que no encontramos información en la literatura. En el Capítulo 4 presentamos las técnicas bootstrap y contamos como pueden aplicarse a nuestros estimadores. En el Capítulo 5 mostramos los resultados obtenidos de datos reales y realizamos simulaciones para evaluar el comportamiento de los estimadores y de la técnica bootstrap.

Capítulo 2

Conceptos previos

2.1. Matrices y vectores aleatorios

Una matriz aleatoria, (o vector aleatorio) es una matriz (vector) cuyos elementos son variables aleatorias. Su distribución está caracterizada por la distribución conjunta de sus elementos. La esperanza de una matriz aleatoria \mathbf{Y} de $m \times n$ es una matriz cuyos elementos son (si existen) las esperanzas de los elementos correspondientes de \mathbf{Y} . Si

$$\mathbf{Y} = \begin{pmatrix} Y_{11} & \dots & Y_{1n} \\ \vdots & \ddots & \vdots \\ Y_{m1} & \dots & Y_{mn} \end{pmatrix} \quad (2.1)$$

entonces

$$\mathbf{E}[\mathbf{Y}] = \begin{pmatrix} \mathbf{E}[Y_{11}] & \dots & \mathbf{E}[Y_{1n}] \\ \vdots & \ddots & \vdots \\ \mathbf{E}[Y_{m1}] & \dots & \mathbf{E}[Y_{mn}] \end{pmatrix}. \quad (2.2)$$

Algunas propiedades de la esperanza de matrices son

Propiedades 2.1.

- $\mathbf{E}[\mathbf{Y}'] = \mathbf{E}[\mathbf{Y}]'$.
- Si a es una constante, $\mathbf{E}[a\mathbf{Y}] = a\mathbf{E}[\mathbf{Y}]$.
- Si A y B son matrices constantes, $\mathbf{E}[A\mathbf{Y}B] = A\mathbf{E}[\mathbf{Y}]B$.
- $\mathbf{E}[\mathbf{Y}_1 + \mathbf{Y}_2] = \mathbf{E}[\mathbf{Y}_1] + \mathbf{E}[\mathbf{Y}_2]$.
- Si \mathbf{Y}_1 y \mathbf{Y}_2 son independientes, $\mathbf{E}[\mathbf{Y}_1\mathbf{Y}_2] = \mathbf{E}[\mathbf{Y}_1]\mathbf{E}[\mathbf{Y}_2]$.

La varianza de un vector aleatorio Y de $m \times 1$ se define como

$$\text{Var}[Y] = \mathbf{E}[(Y - \mathbf{E}[Y])(Y - \mathbf{E}[Y])']. \quad (2.3)$$

Si

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad (2.4)$$

entonces

$$\text{Var}[Y] = \begin{pmatrix} \text{Var}[Y_1] & \text{Cov}[Y_1, Y_2] & \dots & \text{Cov}[Y_1, Y_m] \\ \text{Cov}[Y_2, Y_1] & \text{Var}[Y_2] & & \vdots \\ \vdots & & \ddots & \vdots \\ \text{Cov}[Y_m, Y_1] & \dots & \dots & \text{Var}[Y_m] \end{pmatrix}. \quad (2.5)$$

Algunas propiedades de la varianza para vectores aleatorios son,

Propiedades 2.2.

- $E[YY'] = E[Y]E[Y]' + \text{Var}[Y]$.
- Si k es una constante escalar, $\text{Var}[Y] = k^2\text{Var}[Y]$.
- Si A es una matriz constante de $r \times m$ y b un vector constante, $\text{Var}[AY + b] = A\text{Var}[Y]A'$.
- La varianza es simétrica, $\text{Var}[Y] = \text{Var}[Y]'$.
- $\text{Var}[Y]$ es una matriz semidefinida positiva.
Si a es un vector constante, $a\text{Var}[Y]a' = \text{Var}[aY] \geq 0$.

Se puede definir la covarianza entre dos vectores aleatorios Y_1 de $m \times 1$ e Y_2 de $n \times 1$ como la matriz de $m \times n$

$$\text{Cov}[Y_1, Y_2] = E[(Y_1 - E[Y_1])(Y_2 - E[Y_2])']. \quad (2.6)$$

Si

$$Y_1 = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1m} \end{pmatrix} \quad e \quad Y_2 = \begin{pmatrix} Y_{21} \\ \vdots \\ Y_{2n} \end{pmatrix} \quad (2.7)$$

entonces

$$\text{Cov}[Y_1, Y_2] = \begin{pmatrix} \text{Cov}[Y_{11}, Y_{21}] & \dots & \text{Cov}[Y_{11}, Y_{2n}] \\ \vdots & \ddots & \vdots \\ \text{Cov}[Y_{1m}, Y_{21}] & \dots & \text{Cov}[Y_{1m}, Y_{2n}] \end{pmatrix}. \quad (2.8)$$

Algunas propiedades de la covarianza de vectores aleatorios son,

Propiedades 2.3.

- $\text{Cov}[Y, Y] = \text{Var}[Y]$.
- $\text{Cov}[Y_1, Y_2] = E[Y_1Y_2'] - E[Y_1]E[Y_2]$.
- Si Y_1 e Y_2 son independientes, $\text{Cov}[Y_1, Y_2] = \mathbf{0}$

- $\text{Cov}[Y_1, Y_2] = \text{Cov}[Y_2, Y_1]'$.
- Si k_1 y k_2 son constantes escalares, $\text{Cov}[k_1 Y_1, k_2 Y_2] = k_1 k_2 \text{Cov}[Y_1, Y_2]$.
- Si A y B son matrices constantes de $r \times m$ y $q \times n$ respectivamente, c y d vectores constantes, $\text{Cov}[AY_1 + c, BY_2 + d] = A \text{Cov}[Y_1, Y_2] B'$.
- Si Y_1, Y_2, Y_3 tiene dimensiones apropiadas,

$$\text{Cov}[Y_1 + Y_2, Y_3] = \text{Cov}[Y_1, Y_3] + \text{Cov}[Y_2, Y_3]$$

$$\text{Cov}[Y_1, Y_2 + Y_3] = \text{Cov}[Y_1, Y_2] + \text{Cov}[Y_1, Y_3]$$
- $\text{Var}[Y_1 + Y_2] = \text{Cov}[Y_1 + Y_2, Y_1 + Y_2] = \text{Var}[Y_1] + \text{Var}[Y_2] + \text{Cov}[Y_1, Y_2] + \text{Cov}[Y_2, Y_1]$.
- En particular, si son independientes,

$$\text{Var}[Y_1 + Y_2] = \text{Var}[Y_1] + \text{Var}[Y_2].$$

La notación matricial permite una notación en bloque muy cómoda para trabajar. Definiendo

$$Z = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \quad (2.9)$$

con Y_1 e Y_2 definidas como en (2.7), entonces

$$E[Z] = \begin{pmatrix} \overbrace{E[Y_1]}^{m \times 1} \\ \overbrace{E[Y_2]}^{n \times 1} \end{pmatrix} \quad \text{y} \quad \text{Var}[Z] = \begin{pmatrix} \overbrace{\text{Var}[Y_1]}^{m \times m} & \overbrace{\text{Cov}[Y_1, Y_2]}^{m \times n} \\ \underbrace{\text{Cov}[Y_2, Y_1]}_{n \times m} & \underbrace{\text{Var}[Y_2]}_{n \times n} \end{pmatrix} \quad (2.10)$$

2.2. Matrices definidas positivas

Las matrices definidas positivas y semidefinidas positivas juegan un rol muy importante en la estadística. Enunciamos las definiciones y propiedades que utilizaremos más adelante.

Definición 2.4. La matriz A simétrica, real, de $m \times m$ se dice definida positiva ($A \succ 0$) si $\forall v \neq 0 : vAv' > 0$.

Definición 2.5. La matriz A simétrica, real, de $m \times m$ se dice semi definida positiva ($A \succeq 0$) si $\forall v : vAv' \geq 0$.

Propiedad 2.6. Si $A \succ 0$, entonces $A \succeq 0$.

Propiedad 2.7. Sea C una matriz real, entonces $C'C \succeq 0$.

Propiedad 2.8. (Descomposición espectral)

A es (semi) definida positiva, si y solo si, existe una base ortonormal de autovectores de A , $\{v_1, \dots, v_m\}$ cuyos autovalores asociados $\lambda_1, \dots, \lambda_m$ son (no negativos) positivos. Se puede expresar A como

$$A = \lambda_1 v_1' v_1 + \dots + \lambda_m v_m' v_m. \quad (2.11)$$

Propiedad 2.9. Si $A \succ 0$ entonces es inversible y su inversa puede expresarse

$$A^{-1} = \frac{1}{\lambda_1} v_1' v_1 + \dots + \frac{1}{\lambda_m} v_m' v_m \quad (2.12)$$

donde v_1, \dots, v_m y $\lambda_1, \dots, \lambda_m$ son los autovectores y autovalores de su descomposición espectral.

Propiedad 2.10. (Descomposición de Cholesky)

$A \succ 0$ si y solo si existe una única R triangular superior con elementos positivos en la diagonal tal que $A = R'R$.

Además $A^{-1} = U'U$ donde $U = R^{-1}$, $U' = (R')^{-1}$, y U es triangular superior.

Propiedad 2.11. (Producto interno en \mathbb{R}^m)

Si $A \succ 0$ induce un producto interno en \mathbb{R}^m

$$\langle v, w \rangle_A = \langle vA, w \rangle. \quad (2.13)$$

Definición 2.12. (Orden de matrices definidas positivas)

Sean $A \succeq 0$ y $B \succeq 0$, se dice que $A \succ B$ si $A - B \succ 0$. Es decir que A difiere de B en una matriz definida positiva.

Análogamente, $A \succeq B$ si $A - B \succeq 0$.

Propiedad 2.13. Sean $A \succ 0$, $B \succ 0$, $C \succeq 0$, $D \succeq 0$ y $k \in \mathbb{R}_{>0}$ entonces

- $A + B \succ 0$
- $C + D \succeq 0$
- $A + C \succ 0$
- $kA \succ 0$
- $kC \succeq 0$.

Propiedad 2.14. (Minimización de una variedad lineal)

Si $A \succ 0$, $A \in \mathbb{R}^{m \times m}$, $F \in \mathbb{R}^{m \times k}$, $N \in \mathbb{R}^{m \times 1}$, la función $f : \mathbb{R}^{k \times 1} \rightarrow \mathbb{R}_{\geq 0}$

$$f(x) = (Fx - N)'A(Fx - N), \quad (2.14)$$

es una función concava que alcanza mínimo absoluto en

$$x_{\min} = (F'AF)^{-1} F'AN. \quad (2.15)$$

2.3. Error cuadrático medio

2.3.1. Caso univariado

Sea $\theta \in \Theta$ con $\Theta \subset \mathbb{R}$ un parámetro que se desea estimar y $\hat{\theta}$ un estimador. Para cada θ se define el Error Cuadrático Medio (ECM) como

$$\text{ECM}_\theta [\hat{\theta}] = \text{E}_\theta \left[(\hat{\theta} - \theta)^2 \right]. \quad (2.16)$$

Utilizando las propiedades de la esperanza puede descomponerse en dos partes no negativas. El sesgo cuadrado y la varianza.

$$\text{E}_\theta \left[(\hat{\theta} - \theta)^2 \right] = \text{Var}_\theta [\hat{\theta}] + \text{E}_\theta [\hat{\theta} - \theta]^2. \quad (2.17)$$

El ECM sirve como criterio de comparación entre dos estimadores.

El estimador $\hat{\theta}$ se dice inadmisiblesi existe un estimador $\tilde{\theta}$ que cumple simultaneamente

$$\begin{aligned} \exists \theta \in \Theta \quad \text{ECM}_\theta [\tilde{\theta}] &< \text{ECM}_\theta [\hat{\theta}] \\ \forall \theta \in \Theta \quad \text{ECM}_\theta [\tilde{\theta}] &\leq \text{ECM}_\theta [\hat{\theta}]. \end{aligned} \quad (2.18)$$

El estimador $\hat{\theta}$ se dice óptimo si para cualquier estimador $\tilde{\theta}$ vale que para todo $\theta \in \Theta$

$$\text{ECM}_\theta [\hat{\theta}] \leq \text{ECM}_\theta [\tilde{\theta}]. \quad (2.19)$$

Esto implica que cualquier otro estimador resulta inadmisibles. Con esta definición, salvo en casos triviales, no existen estimadores óptimos. Por este motivo es habitual restringir la clases de estimadores para que sí existan. Un subconjunto de estimadores muy habitual es el de los insesgados. Un estimador se dice insesgado si

$$\text{E}_\theta [\hat{\theta}] = \theta. \quad (2.20)$$

El ECM de un estimador insesgado resulta su varianza, por lo qué encontrar un estimador óptimo entre los insesgados equivale a buscar el estimador de mínima varianza uniforme (IMVU).

2.3.2. Caso multivariado

Sea $\boldsymbol{\theta} \in \Theta$, con $\Theta \subset \mathbb{R}^{p \times 1}$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^t$ y $\hat{\boldsymbol{\theta}}$ un estimador. Para cada $\boldsymbol{\theta} \in \Theta$ se define el error cuadrático medio de $\hat{\boldsymbol{\theta}}$ como

$$\text{ECM}_\theta [\hat{\boldsymbol{\theta}}] = \text{E}_\theta \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \right]. \quad (2.21)$$

El ECM es una matriz de $p \times p$, que por las propiedades mostradas en (2.2) se escribe como suma de su varianza y el análogo multivariado del sesgo cuadrado

$$\text{ECM}_{\theta}[\hat{\theta}] = \text{Var}_{\theta} [\hat{\theta}] + \text{E}_{\theta} [\theta - \hat{\theta}] \text{E}_{\theta} [\theta - \hat{\theta}]'. \quad (2.22)$$

Ambas partes son semidefinidas positivas y por lo tanto, también lo es el ECM.

Para un valor de θ se dice que $\tilde{\theta}$ tiene mejor error cuadrático medio si

$$\text{ECM}_{\theta}[\tilde{\theta}] \succeq \text{ECM}_{\theta}[\hat{\theta}]. \quad (2.23)$$

Este criterio de comparación es equivalente a pedir que cualquier combinación lineal de los parametros estimados $\tilde{\theta}$ tenga ECM univariado menor que la misma combinación de $\hat{\theta}$. Esto es cierto porque que para todo $a \in \mathbb{R}^{1 \times p}$ constante,

$$\text{ECM}_{\theta} [a\tilde{\theta}] = a \left(\text{ECM}_{\theta} [\tilde{\theta}] \right) a' \geq a \left(\text{ECM}_{\theta} [\hat{\theta}] \right) a' = \text{ECM}_{\theta} [a\hat{\theta}]. \quad (2.24)$$

Al igual que en el caso univariado, un estimador $\hat{\theta}$ se dice inadmisibile si existe $\tilde{\theta}$ que cumple simultáneamente

$$\begin{aligned} \exists \theta \in \Theta \quad \text{ECM}_{\theta} [\tilde{\theta}] \succ \text{ECM}_{\theta} [\hat{\theta}], \\ \forall \theta \in \Theta \quad \text{ECM}_{\theta} [\tilde{\theta}] \succeq \text{ECM}_{\theta} [\hat{\theta}]. \end{aligned} \quad (2.25)$$

El criterio de comparación entre estimadores multivariados es restrictivo ya que para un θ fijo establece un orden parcial. Esto quiere decir que existen estimadores que no resultan comparables. Hay otros criterios que si dan un orden total, es decir, permiten comparar cualquier par de estimadores, en general estos son funciones escalares de los autovalores de la matriz del error cuadrático medio. Ejemplos de estas funciones son la traza y el determinante.

2.4. Modelo lineal clásico

El modelo lineal clásico puede formularse matricialmente de la siguiente forma

$$\begin{cases} Z = X\beta + \varepsilon \\ \text{E}[\varepsilon] = \mathbf{0} \\ \text{Var}[\varepsilon] = \sigma^2 \mathbf{I} \end{cases} \quad (2.26)$$

donde Z es la variable respuesta, X la matriz de diseño o de variables dependientes que se consideran fijas, β y σ^2 parámetros desconocidos y ε el vector de errores aleatorio.

El estimador de cuadrados mínimos $\hat{\beta}_{\text{CM}}$ se define como el que minimiza la suma de

los cuadrados de los errores.

$$\hat{\beta}_{\text{CM}} = \arg \min_{\beta} \|Z - X\beta\|^2. \quad (2.27)$$

Utilizando la propiedad (2.14), la solución explícita del problema es

$$\hat{\beta}_{\text{CM}} = (X'X)^{-1} X'Z \quad (2.28)$$

A partir de su forma explícita se puede notar que es insesgado y calcular su varianza.

Definiendo $H = (X'X)^{-1} X'$

$$\begin{aligned} E \left[(X'X)^{-1} X'Z \right] &= \underbrace{E[H]}_{=H} E[X\beta + \varepsilon] \\ &= E[\beta] + E[H] \underbrace{E[\varepsilon]}_{=0} \\ &= \beta \\ \text{Var} \left[(X'X)^{-1} X'Z \right] &= H \underbrace{\text{Var}[X\beta]}_{=0} H' + H \text{Var}[\varepsilon] H' \\ &= \sigma^2 (X'X)^{-1}. \end{aligned} \quad (2.29)$$

Definición 2.15. El estimador $\hat{\beta}$ se dice lineal si es una transformación lineal de Z . Es decir, para alguna matriz C que no depende de Z ni de β

$$\hat{\beta} = CZ. \quad (2.30)$$

Observación 2.16. El estimador de cuadrados mínimos es un estimador lineal con $C = (X'X)^{-1} X'$.

El siguiente teorema asegura que el estimador de cuadrados mínimos es óptimo entre los estimadores lineales e insesgados.

Teorema 2.17. (Gauss-Markov)

Si $\hat{\beta}$ es lineal e insesgado para el modelo (2.26), entonces $\text{ECM}_{\beta}[\hat{\beta}] \succeq \text{ECM}_{\beta}[\hat{\beta}_{\text{CM}}]$.

Dem 2.18. Sea $\hat{\beta} = CZ$ un estimador lineal de β . La matriz C depende de la matriz X , pero no de Z ni de β . Para alguna matriz D , $C = (X'X)^{-1} X' + D$

$$\begin{aligned} E_{\beta}[CZ] &= E_{\beta} \left[\left((X'X)^{-1} X' + D \right) (X\beta + \varepsilon) \right] \\ &= \left((X'X)^{-1} X' + D \right) (X\beta + E[\varepsilon]) \\ &= \left((X'X)^{-1} X' + D \right) (X\beta) \\ &= (\mathbf{I} + DX) \beta \end{aligned} \quad (2.31)$$

Por lo tanto, $\hat{\beta}$ es insesgado si y solo si $DX = \mathbf{0}$.

Como ambos estimadores son insesgados, para comparar el ECM basta comparar su varianza.

$$\begin{aligned}
\text{Var}_{\beta} [CZ] &= C \text{Var}_{\beta} [Z] C' \\
&= \sigma^2 C C' \\
&= \sigma^2 \left((X'X)^{-1} X' + D \right) \left(X (X'X)^{-1} + D' \right) \\
&= \sigma^2 (X'X)^{-1} + \sigma^2 (X'X)^{-1} \underbrace{(DX)'}_{=0} + \sigma^2 \underbrace{DX}_{=0} (X'X)^{-1} + \sigma^2 D D' \\
&= \text{Var}_{\beta} \left[\hat{\beta}_{\text{CM}} \right] + \sigma^2 D D'.
\end{aligned} \tag{2.32}$$

Por la propiedad (2.7) $\text{Var}_{\beta} \left[\hat{\beta} \right] - \text{Var}_{\beta} \left[\hat{\beta}_{\text{CM}} \right] = \sigma^2 D D' \succeq 0$ y por ende

$$\text{Var} \left[\hat{\beta} \right] \succeq \text{Var} \left[\hat{\beta}_{\text{CM}} \right]. \tag{2.33}$$

En ningún paso de la demostración se utiliza la distribución de los errores ni que sean independientes (sólo que no están correlacionados). Las hipótesis de linealidad e insesgado son necesarias para la optimalidad. Más aún, el estimador de cuadrados mínimos es inadmisibles si se quita la linealidad (Estimador de James) o si no se pide insesgado (Ridge Regression). Aún así es un estimador muy utilizado por su simplicidad y su historia. Se puede leer sobre otros estimadores para el modelo lineal clásico y sus propiedades el libro de Hastie, Tibshirani, y Friedman [11].

2.4.1. Cuadrados mínimos generalizados

El modelo generalizado se obtiene de relajar las hipótesis del modelo (2.26)

$$\begin{cases} Z = X\beta + \varepsilon \\ E[\varepsilon] = \mathbf{0} \\ \text{Var}[\varepsilon] = \sigma^2 D \end{cases} \tag{2.34}$$

Ahora la matriz de covarianza no es necesariamente un múltiplo de la identidad y se la conoce salvo la constante σ^2 .

En este caso $\hat{\beta}_{\text{CM}}$ sigue siendo insesgado pero resulta inadmisibles entre los lineales insesgados.

Para encontrar el estimador óptimo se transforman las observaciones para que el problema caiga sobre las hipótesis de Gauss-Markov.

Teorema 2.19. Bajo las hipótesis del modelo (2.34), el estimador

$$\hat{\beta}_{\text{CMG}} = (X'D^{-1}X)^{-1} X'D^{-1}Z \tag{2.35}$$

es óptimo entre los lineales insesgados.

Dem 2.20. Como D es definida positiva, se puede escribir $D = R'R$ y $D^{-1} = U'U$ donde $U = R^{-1}$. Multiplicando por U' a ambos lados de la igualdad de (2.34), queda

$$U'Z = U'X\beta + U'\varepsilon. \quad (2.36)$$

Se define, $Y = U'Z$, $W = U'X$ y $\eta = U'\varepsilon$. La esperanza y varianza de η se pueden calcular usando las propiedades (2.2).

$$\begin{aligned} E[\eta] &= U'E[\varepsilon] \\ &= \mathbf{0} \end{aligned} \quad (2.37)$$

$$\begin{aligned} \text{Var}[\eta] &= U'\text{Var}[\varepsilon]U \\ &= \sigma^2 U' \underbrace{R'R}_D U \\ &= \sigma^2 \mathbf{I}. \end{aligned} \quad (2.38)$$

Luego β satisface el siguiente modelo lineal que cumple con las hipótesis de Gauss Markov

$$\begin{cases} Y = W\beta + \eta \\ E[\eta] = \mathbf{0} \\ \text{Var}[\eta] = \sigma^2 \mathbf{I} \end{cases} \quad (2.39)$$

y el estimador óptimo para β resulta

$$\begin{aligned} (W'W)^{-1}W'Y &= (X'U'UX)^{-1}X'U'UZ \\ &= (X'D^{-1}X)^{-1}X'D^{-1}Z \\ &= \hat{\beta}_{\text{CMG}}. \end{aligned} \quad (2.40)$$

La varianza queda

$$\text{Var}[\hat{\beta}_{\text{CMG}}] = \sigma^2 (X'D^{-1}X)^{-1} \quad (2.41)$$

De la propiedad (2.14), $\hat{\beta}_{\text{CMG}}$ se puede interpretar geoméricamente como el estimador que resuelve

$$\hat{\beta}_{\text{CMG}} = \arg \min_{\beta} \|Z - X\beta\|_{D^{-1}}^2 \quad (2.42)$$

donde $\|\cdot\|_{D^{-1}}$ es la norma asociada al producto interno inducido por D^{-1} (2.11). Es interesante notar que todas las propiedades se deducen de su forma explícita y no de su definición geométrica.

2.4.2. Cuadrados mínimos factibles

Cuando la matriz de varianza de los errores D no es conocida, cuadrados mínimos generalizados no puede ser calculado. El estimador de cuadrados mínimos factibles trata de resolver esta situación. Consiste en un procedimiento general de dos pasos que varía según la información que se tiene sobre D . Los pasos son,

1. Estimar D por \hat{D}
2. Calcular $\hat{\beta}_{\text{CMF}} = \left(X' \hat{D}^{-1} X\right)^{-1} X' \hat{D}^{-1} Z$.

Las propiedades de $\hat{\beta}_{\text{CMF}}$ dependen de como fue estimado D . Mientras más información se tenga sobre la estructura de la varianza, mejor será la calidad del estimador.

El siguiente es un ejemplo de como obtener un estimador para β en el problema (2.34) cuando se asume que la matriz de covarianza es una matriz diagonal $D = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$ con σ_i^2 desconocidos.

1. Obtener estimación inicial para D .

$$\begin{aligned}\hat{\beta}_0 &= \hat{\beta}_{\text{CM}} \\ U_0 &= Z - X\hat{\beta}_0 \\ \hat{D}_0 &= \text{diag}(U_{1,0}^2, \dots, U_{m,0}^2) \\ n &= 0\end{aligned}$$

2. Obtener una estimación para D

Hasta criterio de terminación
 $(\hat{D}_n \approx \hat{D}_{n-1} \text{ o } n > N)$:

$$\begin{aligned}n &= n + 1 \\ \hat{\beta}_n &= \left(X' \hat{D}_{n-1}^{-1} X\right)^{-1} X' \hat{D}_{n-1}^{-1} Z \\ U_n &= Z - X\hat{\beta}_n \\ \hat{D}_n &= \text{diag}(U_{1,n}^2, \dots, U_{m,n}^2)\end{aligned}$$

3. Calcular $\hat{\beta}_{\text{CMF}}$

$$\hat{\beta}_{\text{CMF}} = \left(X' \hat{D}_n^{-1} X\right)^{-1} X' \hat{D}_n^{-1} Z$$

Capítulo 3

Errores en las variables

3.1. Modelos con errores en las variables

La formulación matemática que describe las mediciones es conocida en la literatura estadística como modelos con errores en las variables (MEV) o modelos de medición con errores. A continuación los presentamos en una forma muy general ya que de esta manera las ideas resultan más claras. Más adelante veremos que el problema de estimar el coeficiente de reflexión cae naturalmente en esta clase de problemas.

En un Modelo con Error en las Variables existen dos parámetros desconocidos $\boldsymbol{\mu}^* \in \mathbb{R}^q$ (argumento) y $\boldsymbol{\beta} \in \mathbb{R}^p$ (mesurando) que satisfacen un modelo, en el sentido que forman un cero de una función conocida $\Upsilon : \mathbb{R}^{q+p} \rightarrow \mathbb{R}^m$

$$\Upsilon(\boldsymbol{\mu}^*, \boldsymbol{\beta}) = \mathbf{0}. \quad (3.1)$$

Se observa $\boldsymbol{\mu}$ bajo un error aditivo $\boldsymbol{\xi}$ del que se asume conocida su distribución o sus momentos

$$\boldsymbol{\mu} = \boldsymbol{\mu}^* + \boldsymbol{\xi} \quad (3.2)$$

y se busca una estimación de $\boldsymbol{\beta}$ a partir de $\boldsymbol{\mu}$.

Observación 3.1. Que el error de la observación sea aditivo es una hipótesis con la que nos interesa trabajar. Existen otras formas de errores como los multiplicativos donde, como el nombre lo indica, el error multiplica al valor verdadero.

$$\boldsymbol{\mu} = \boldsymbol{\xi}\boldsymbol{\mu}^* \quad (3.3)$$

Observación 3.2. Las mediciones directas para mesurandos escalares desarrolladas en el primer capítulo son un problema MEV con

$$\Upsilon(\boldsymbol{\mu}^*, \boldsymbol{\beta}) = \boldsymbol{\beta} - f(\boldsymbol{\mu}^*). \quad (3.4)$$

Observación 3.3. El suplemento 2 de la GUM [6] contempla lo que llama modelos de medición implícita. Estos asumen que hay tantas ecuaciones como dimensión del mesurando ($m = p$). Para obtener una estimación del mesurando se busca $\hat{\beta}$ que cumpla

$$\Upsilon(\boldsymbol{\mu}, \hat{\boldsymbol{\beta}}) = \mathbf{0}. \quad (3.5)$$

Es decir, que satisfaga el modelo evaluado en las observaciones. La matriz de varianza de la medición se obtiene a partir de las derivadas de la función implícita.

La GUM no contempla casos donde $m > p$, es decir cuando hay mas ecuaciones que las necesarias. La dificultad radica en como definir el estimador ya que si $\boldsymbol{\mu} \neq \boldsymbol{\mu}^*$ puede no existir $\hat{\boldsymbol{\beta}}$ que satisfaga (3.5).

Es cierto que si se descartan ecuaciones hasta quedarse con p se puede aplicar la técnica descripta, pero esta solución no resulta satisfactoria ya que es arbitraria y se está perdiendo información.

Observación 3.4. La literatura de MEV es muy amplia y sin embargo, toda la que hemos estudiado trabaja con ciertas hipótesis que no se adaptan del todo al problema que nos interesa resolver.

Cuando hay mas ecuaciones que incógnitas ($m > p$) suele asumirse una hipótesis de independencia o “no correlación” de las observaciones para cada ecuación. En los términos antes planteados, con $\Upsilon = (\Upsilon_1, \dots, \Upsilon_m)$, considerando $\boldsymbol{\mu}^* = (\boldsymbol{\mu}_1^*, \dots, \boldsymbol{\mu}_m^*)$, con $\boldsymbol{\mu}_i^* \in \mathbb{R}^{q_i}$ y $\sum_{i=1}^m q_i = q$, los modelos de la literatura cumplen que

$$\begin{cases} \Upsilon_1(\boldsymbol{\mu}_1^*, \boldsymbol{\beta}) = 0 \\ \vdots \\ \Upsilon_m(\boldsymbol{\mu}_m^*, \boldsymbol{\beta}) = 0 \end{cases} \quad (3.6)$$

es decir, cada ecuación depende de argumentos observados bajo errores aditivos $\boldsymbol{\xi}_i$ independientes o no correlacionados. Esto es, para $1 \leq i \leq m$ se observa

$$\boldsymbol{\mu}_i = \boldsymbol{\mu}_i^* + \boldsymbol{\xi}_i \quad (3.7)$$

y se asume

$$\text{Cov}[\boldsymbol{\xi}_i, \boldsymbol{\xi}_j] = \mathbf{0} \quad j \neq i. \quad (3.8)$$

Bajo ciertas hipótesis adicionales Amemiya y Fuller [15] proponen un estimador de máxima verosimilitud al cual derivan sus propiedades asintóticas. El estimador se define como los $\boldsymbol{\mu}_1^*, \dots, \boldsymbol{\mu}_m^*, \boldsymbol{\beta}$ más verosímiles que satisfacen (3.6). Sin embargo la hipótesis de independencia no aplica a nuestro problema. Es por esto que nos hemos esforzado en conseguir formular una manera que su ausencia resulte natural.

Libros de referencia para el modelo (3.6) son el de Fuller [1] y el de Carol et al. [2].

3.2. Regresión lineal con errores en las variables

Dentro de los problemas donde la cantidad de ecuaciones es mayor que la dimensión del parámetro ($m > p$) los modelos más sencillos y mas estudiados en la literatura MEV son los llamados modelos de regresión lineal con errores en las variables. Estos modelos están extensamente desarrollados por Fuller [1] y pueden formularse como

$$\begin{cases} z_t^* = \beta_0 + \beta_1 x_{t1}^* + \dots + \beta_p x_{tp}^* \\ X_{tk} = x_{tk}^* + \eta_{tk} \\ Z_t = z_t^* + \varepsilon_t \\ 1 \leq t \leq m \quad 1 \leq k \leq p. \end{cases} \quad (3.9)$$

Es sencillo ver que este problema es un modelo con error en las variables definidos en la sección anterior, basta considerar

$$\begin{aligned} \boldsymbol{\mu}^* &= (z_1^*, \dots, z_m^*, x_{11}^*, \dots, x_{1p}^*, \dots, x_{m1}^*, \dots, x_{mp}^*) \\ \boldsymbol{\mu} &= (Z_1, \dots, Z_m, X_{11}, \dots, X_{1p}, \dots, X_{m1}, \dots, X_{mp}) \\ \boldsymbol{\xi} &= (\varepsilon_1, \dots, \varepsilon_m, \eta_{11}, \dots, \eta_{1p}, \dots, \eta_{m1}, \dots, \eta_{mp}) \end{aligned} \quad (3.10)$$

y la función Υ definida para cualquier valor de $\boldsymbol{\mu}$,

$$\Upsilon(\boldsymbol{\mu}, \boldsymbol{\beta}) = \begin{pmatrix} Z_1 \\ \vdots \\ Z_m \end{pmatrix} - \begin{pmatrix} 1 & X_{11} & \dots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{mp} & \dots & X_{mp} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}. \quad (3.11)$$

Una forma de entender estos modelos es pensar que $\boldsymbol{\beta}$ satisface un modelo de regresión lineal clásico del cual no se conocen con exactitud el valor de sus variables explicativas, si no que se los observa bajo un error. Pensarlos de esta forma sugiere que estos problemas son más “difíciles” que los de regresión lineal clásica.

Existe una dificultad que excede lo conceptual y que es causada por la notación utilizada en la literatura. Esta resulta muy similar a la de regresión lineal clásica. Como esta última está tan estudiada y apropiada por la comunidad científica termina generando confusiones. En un ejemplo a continuación mostramos los graves problemas que puede ocasionar no notar esta diferencia y tratar al problema como uno clásico. Para $\boldsymbol{\beta} \in \mathbb{R}^2$ el modelo (3.9) puede reescribirse

$$\begin{cases} Z_t = x_t^* \beta_1 + \beta_0 + \varepsilon_t \\ X_t = x_t^* + \eta_t \\ 1 \leq t \leq m. \end{cases} \quad (3.12)$$

Esta formulación resulta muy similar a la de regresión lineal clásica con variables explicativas aleatorias con media desconocida

$$\begin{cases} Z_t = \underline{X}_t \beta_1 + \beta_0 + \varepsilon_t \\ X_t = x_t^* + \eta_t \\ 1 \leq t \leq m. \end{cases} \quad (3.13)$$

Lo que parece una “sutil” diferencia de notación esconde que los estimadores habituales del modelo (3.13) no resultan aplicables para el modelo (3.12) ya que pierden propiedades deseables como la insesgadez o la consistencia.

Para mostrar este efecto y dejar en claro que los problemas de regresión lineal con errores en las variables no pueden ser tratados como problemas de regresión clásica, realizamos una simulación de Montecarlo en la que estimamos las propiedades del estimador de cuadrados mínimos de β_1 que es el mismo del expuesto en (2.28)

Para la simulación se establece $\beta_1 = 1$, $\beta_0 = 0$ y los x_1^*, \dots, x_m^* equiespaciados en el intervalo $[-2, 2]$ y con errores con distribuciones normales bivariadas independientes

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}\right) \quad (3.14)$$

Para cada m y σ_η^2 de la tabla se realizaron $N = 1000$ replicaciones con las que se estimó $E[\hat{\beta}_1]$ y $P(|\hat{\beta}_1 - \beta_1| > \delta)$ con $\delta = 0,2$. Los resultados obtenidos fueron

$E[\hat{\beta}_1]$	$\sigma_\eta^2 = 0$	$\sigma_\eta^2 = 0,25$	$\sigma_\eta^2 = 0,5$	$\sigma_\eta^2 = 1$	$\sigma_\eta^2 = 1,5$	$\sigma_\eta^2 = 2$
$m = 40$	1.01	0.86	0.75	0.60	0.49	0.42
$m = 80$	1.00	0.85	0.74	0.59	0.48	0.41
$m = 150$	1.00	0.85	0.73	0.58	0.48	0.41
$m = 300$	1.00	0.84	0.73	0.57	0.47	0.40
$m = 1000$	1.00	0.84	0.73	0.57	0.47	0.40

$P(\hat{\beta}_1 - \beta_1 > 0,2)$	$\sigma_\eta^2 = 0$	$\sigma_\eta^2 = 0,25$	$\sigma_\eta^2 = 0,5$	$\sigma_\eta^2 = 1$	$\sigma_\eta^2 = 1,5$	$\sigma_\eta^2 = 2$
$m = 40$	0.12	0.35	0.65	0.95	0.99	1.00
$m = 80$	0.03	0.28	0.73	0.99	1.00	1.00
$m = 150$	0.00	0.25	0.86	1.00	1.00	1.00
$m = 300$	0.00	0.17	0.93	1.00	1.00	1.00
$m = 1000$	0.00	0.05	1.00	1.00	1.00	1.00

De la primera tabla puede apreciarse que, independientemente del tamaño de la muestra, la esperanza del estimador se distancia del valor verdadero $\beta = 1$ a medida que aumenta σ_η^2 . Este fenómeno está bien estudiado en la literatura de errores en las variables y se dice que el estimador de cuadrados mínimos está sesgado hacia el cero.

De la segunda tabla, para $\sigma^2 \geq 0,5$, se aprecia que con tamaño de muestra grandes, con probabilidad estimada en 1, el estimador se encuentra a una distancia mayor de $\delta = 0,2$ del valor verdadero, mostrando que el estimador no es consistente. Esto puede verse para cualquier $\sigma^2 > 0$, pero para esto es necesario utilizar valores δ mas chicos.

En el siguiente gráfico puede observarse el efecto de “sesgo hacia el cero” del estimador de cuadrados mínimos a medida que aumenta σ_η^2 y con para un $m = 40$.

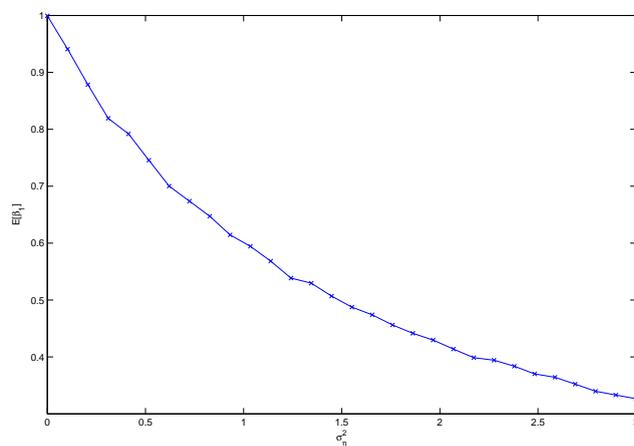


FIGURA 3.1: Esperanza de $\hat{\beta}_1$ en función de σ_η^2 .

Los siguientes gráficos corresponden a simulaciones con distinto σ_η^2 y $m = 40$. Los puntos corresponden a las observaciones (X_t, Z_t) , La recta con rayas es la recta del modelo “ $y = 1x + 0$ ”, y la recta restante es la que se obtuvo con el estimador de cuadrados mínimos.

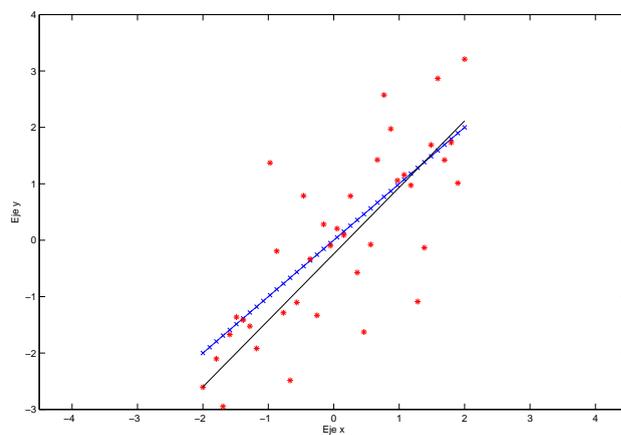


FIGURA 3.2: Observaciones y recta predicha con $\sigma_\eta^2 = 0$.

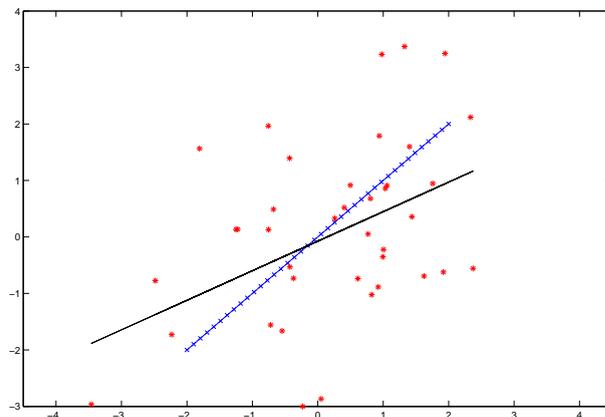


FIGURA 3.3: Observaciones y recta predicha con $\sigma_{\eta}^2 = 1$.

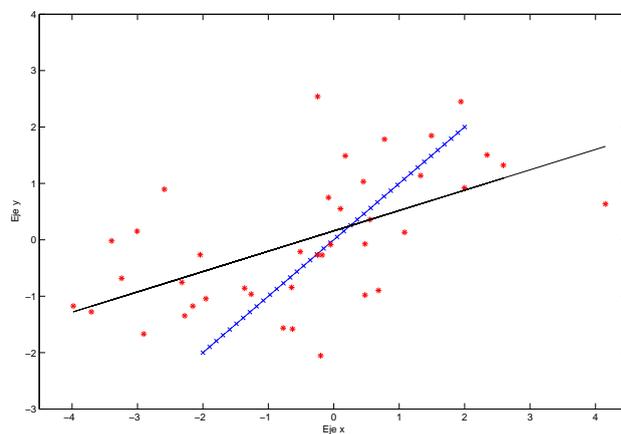


FIGURA 3.4: Observaciones y recta predicha con $\sigma_{\eta}^2 = 2$.

3.3. Problema real como un problema de MEV

Nos interesa resolver un tipo particular de modelos con errores en las variables que incluye al problema de medición del coeficiente de reflexión. En este, los valores $\boldsymbol{\mu}^*$ y $\boldsymbol{\beta}$ satisfacen un modelo en el sentido de que se conocen $\Phi : \mathbb{R}^q \rightarrow \mathbb{R}^{m \times 1}$ y $\Psi : \mathbb{R}^q \rightarrow \mathbb{R}^{m \times p}$ tales que

$$\begin{pmatrix} \Phi_1(\boldsymbol{\mu}^*) \\ \vdots \\ \Phi_m(\boldsymbol{\mu}^*) \end{pmatrix} = \begin{pmatrix} \Psi_{11}(\boldsymbol{\mu}^*) & \dots & \Psi_{1p}(\boldsymbol{\mu}^*) \\ \vdots & \ddots & \vdots \\ \Psi_{m1}(\boldsymbol{\mu}^*) & \dots & \Psi_{mp}(\boldsymbol{\mu}^*) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}. \quad (3.15)$$

Como antes, la observación de $\boldsymbol{\mu}^*$ está perturbada por un error aditivo $\boldsymbol{\xi}$ del que se asumen conocida su distribución o sus momentos

$$\boldsymbol{\mu} = \boldsymbol{\mu}^* + \boldsymbol{\xi}. \quad (3.16)$$

Queremos poder estimar β y asignarle una región de confianza a partir de la observación de μ .

Es sencillo plantear el problema de estimar Γ_g en estos términos, basta considerar $\beta = (\text{Re}[\Gamma_g], \text{Im}[\Gamma_g])$, $\mu^* = (s, m_1, \dots, m_7)$ y obtener las funciones Φ , Ψ a partir de composiciones de las z , x , y con las f y h definidas en el primer capítulo.

El problema no es fácil de resolver en forma general. La propuesta en este trabajo es linealizar las componentes de Ψ , Φ obteniendo un problema de regresión lineal con errores en las variables y estimar y asignar regiones de confianza a partir de este.

En lo que sigue mostramos como obtener el problema lineal asociado. Si las Ψ_{tk} y Φ_t son \mathcal{C}^2 podemos aplicar el método Delta. El polinomio de Taylor de primer orden centrado en μ^* y evaluado en μ queda

$$\begin{aligned}\Psi_{tk}(\mu) &= \Psi_{tk}(\mu^*) + \nabla\Psi_{tk}(\mu^*) \underbrace{(\mu - \mu^*)}_{\xi} + R_{tk} \\ \Phi_t(\mu) &= \Phi_t(\mu^*) + \nabla\Phi_t(\mu^*) \underbrace{(\mu - \mu^*)}_{\xi} + R_t.\end{aligned}\tag{3.17}$$

Considerando como errores aleatorios

$$\varepsilon_t = \nabla\Phi_t(\mu^*)\xi \quad y \quad \eta_{tk} = \nabla\Psi_{tk}(\mu^*)\xi,\tag{3.18}$$

y que los restos de orden 2 son despreciables, se obtiene como una buena aproximación del modelo (3.15)

$$\begin{cases} \Phi_t(\mu^*) &= \beta_0 + \Psi_{t1}(\mu^*)\beta_1 + \dots + \Psi_{tp}(\mu^*)\beta_p \\ \Phi_t(\mu) &\approx \Phi_t(\mu^*) + \varepsilon_t \\ \Psi_{tk}(\mu) &\approx \Psi_{tk}(\mu^*) + \eta_{tk}.\end{cases}\tag{3.19}$$

Si definimos

$$\begin{aligned}Z_t &= \Phi_t(\mu^*) + \varepsilon_t & X_{tk} &= \Psi_{tk}(\mu^*) + \eta_{tk} \\ z_t^* &= \Phi_t(\mu^*) & x_{tk}^* &= \Psi_{tk}(\mu^*)\end{aligned}\tag{3.20}$$

la aproximación resulta un problema de regresión lineal con errores en las variables como en (3.9).

Los momentos de los errores dependen de μ^* . Para calcularlos se utilizan las propiedades de esperanza, varianza y covarianza de vectores aleatorios. Llamando Σ a la matriz de

varianza de ξ queda

$$\begin{aligned}
E[\varepsilon_t] &= \nabla\Phi_t(\boldsymbol{\mu}^*)E[\xi] = \mathbf{0} \\
E[\eta_{tk}] &= \nabla\Psi_{tk}(\boldsymbol{\mu}^*)E[\xi] = \mathbf{0} \\
\text{Var}[\varepsilon_t] &= [\nabla\Phi_t(\boldsymbol{\mu}^*)]\Sigma[\nabla\Phi_t(\boldsymbol{\mu}^*)]' \\
\text{Var}[\eta_{tk}] &= [\nabla\Psi_{tk}(\boldsymbol{\mu}^*)]\Sigma[\nabla\Psi_{tk}(\boldsymbol{\mu}^*)]' \\
\text{Cov}[\varepsilon_t, \varepsilon_i] &= [\nabla\Phi_t(\boldsymbol{\mu}^*)]\Sigma[\nabla\Phi_i(\boldsymbol{\mu}^*)]' \\
\text{Cov}[\eta_{tk}, \eta_{ij}] &= [\nabla\Psi_{tk}(\boldsymbol{\mu}^*)]\Sigma[\nabla\Psi_{ij}(\boldsymbol{\mu}^*)]' \\
\text{Cov}[\varepsilon_t, \eta_{ij}] &= [\nabla\Phi_t(\boldsymbol{\mu}^*)]\Sigma[\nabla\Psi_{ij}(\boldsymbol{\mu}^*)]'.
\end{aligned} \tag{3.21}$$

Si aproximamos los elementos de (3.21) utilizando $\nabla\Phi_t(\boldsymbol{\mu})$ y $\nabla\Psi_{tk}(\boldsymbol{\mu})$ y asumimos estas estimaciones como exactas, tenemos un modelo lineal con errores en las variables en el que conocemos los momentos de los errores. Estimaremos $\boldsymbol{\beta}$ y asignaremos regiones de confianza a partir de este modelo.

3.4. Problema a estudiar

Motivados por la linearización del problema original, queremos determinar un estimador y una región de confianza para

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \tag{3.22}$$

que satisface un modelo lineal con errores en las variables en el que la matriz de varianza de los errores es conocida. Los parámetros $\boldsymbol{\beta}$ satisfacen el siguiente conjunto de ecuaciones

$$\begin{cases} Z_t = \beta_0 + \beta_1 x_{t1}^* + \dots + \beta_p x_{tp}^* + \varepsilon_t & 1 \leq t \leq m \\ X_{tk} = x_{tk}^* + \eta_{tk} & 1 \leq k \leq p. \end{cases} \tag{3.23}$$

Si llamamos observación t y error de la observación t a los vectores

$$Y_t = \begin{pmatrix} X_{t1} \\ \vdots \\ X_{tp} \\ Z_t \end{pmatrix} \quad \boldsymbol{\epsilon}_t = \begin{pmatrix} \eta_{t1} \\ \vdots \\ \eta_{tp} \\ \varepsilon_t \end{pmatrix} \tag{3.24}$$

donde cada X_{tk} es una observación de x_{tk}^* (un valor desconocido) y consideramos el vector de errores $\boldsymbol{\xi} \in \mathbb{R}^{m(p+1) \times 1}$ que se obtiene apilando los errores de las observaciones

$$\boldsymbol{\xi} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{pmatrix}, \tag{3.25}$$

entonces contamos como hipótesis que

$$\begin{aligned} E[\boldsymbol{\xi}] &= \mathbf{0} \\ \text{Var}[\boldsymbol{\xi}] &= D, \end{aligned} \quad (3.26)$$

donde la varianza de los errores D es una matriz de $m(p+1) \times m(p+1)$ **conocida e inversible**. Si llamamos

$$\begin{aligned} D_{\epsilon_t \epsilon_t} &= \text{Var}[\epsilon_t] \implies \\ D_{\epsilon_t \epsilon_k} &= \text{Cov}[\epsilon_t, \epsilon_k], \end{aligned}$$

la matriz D puede expresarse utilizando la notación de (2.10) como la siguiente matriz de bloques

$$D = \begin{pmatrix} D_{\epsilon_1 \epsilon_1} & D_{\epsilon_1 \epsilon_2} & \cdots & D_{\epsilon_1 \epsilon_m} \\ D_{\epsilon_2 \epsilon_1} & D_{\epsilon_2 \epsilon_2} & \cdots & D_{\epsilon_2 \epsilon_m} \\ \vdots & \vdots & \ddots & \vdots \\ D_{\epsilon_m \epsilon_1} & D_{\epsilon_m \epsilon_2} & \cdots & D_{\epsilon_m \epsilon_m} \end{pmatrix}. \quad (3.27)$$

El problema también puede enunciarse en notación matricial

$$\begin{cases} Z = X^* \boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ X = X^* + \boldsymbol{\eta} \end{cases} \quad (3.28)$$

donde

$$X^* = \begin{pmatrix} 1 & x_{11}^* & \cdots & x_{1p}^* \\ \vdots & \vdots & & \vdots \\ 1 & x_{m1}^* & \cdots & x_{mp}^* \end{pmatrix} \quad \boldsymbol{\eta} = \begin{pmatrix} 0 & \eta_{11} & \cdots & \eta_{1p} \\ \vdots & \vdots & & \vdots \\ 0 & \eta_{m1} & \cdots & \eta_{mp} \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix} \quad Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_m \end{pmatrix}. \quad (3.29)$$

Por supuesto utilizaremos cada notación cuando resulte más conveniente.

3.5. Estimador de sesgo corregido

Una estrategia es estimar el modelo por cuadrados mínimos. Calcular su sesgo y corregirlo. Utilizando la notación matricial de (3.28),

$$\begin{aligned} Z &= X^* \boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= X \boldsymbol{\beta} - (X - X^*) \boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= X \boldsymbol{\beta} - \boldsymbol{\eta} \boldsymbol{\beta} + \boldsymbol{\varepsilon}. \end{aligned} \quad (3.30)$$

Multiplicando a ambos lados de la igualdad por $(X'X)^{-1} X'$ y utilizando que $X = X^* + \eta$ se obtiene:

$$\begin{aligned} (X'X)^{-1} X'Z &= \beta - (X'X)^{-1} X'\eta\beta + (X'X)^{-1} X'\varepsilon \\ \hat{\beta}_{\text{CM}} &= \beta - \underbrace{(X'X)^{-1} \eta'\eta\beta}_{\text{I}} + \underbrace{(X'X)^{-1} \eta'\varepsilon}_{\text{II}} - \underbrace{(X'X)^{-1} X^*\eta\beta}_{\text{III}} + \underbrace{(X'X)^{-1} X^*\varepsilon}_{\text{IV}} \end{aligned} \quad (3.31)$$

Si existe una sucesión de observaciones que garanticen la existencia de los siguientes límites en probabilidad

$$\begin{aligned} \frac{X'X}{m} &\xrightarrow{\text{P}} Q \\ \frac{\eta'\eta}{m} &\xrightarrow{\text{P}} C_{\eta\eta} \\ \frac{\eta'\varepsilon}{m} &\xrightarrow{\text{P}} C_{\eta\varepsilon} \end{aligned} \quad (3.32)$$

donde $C_{\eta\eta} \in \mathbb{R}^{(p+1) \times (p+1)}$ y $C_{\eta\varepsilon} \in \mathbb{R}^{(p+1) \times 1}$. Es posible ver que

$$\begin{aligned} \text{I} &= \left(\frac{X'X}{m} \right)^{-1} \frac{\eta'\eta}{m} \xrightarrow{\text{P}} Q^{-1} C_{\eta\eta} \\ \text{II} &= \left(\frac{X'X}{m} \right)^{-1} \frac{\eta'\varepsilon}{m} \xrightarrow{\text{P}} Q^{-1} C_{\eta\varepsilon}. \end{aligned} \quad (3.33)$$

Los terminos **III** y **IV** tienden a cero en probabilidad ya que X^* es constante, $E[\eta] = \mathbf{0}$ y $E[\varepsilon] = \mathbf{0}$

$$\begin{aligned} \text{III} &= \left(\frac{X'X}{m} \right)^{-1} \frac{X^*\eta}{m} \xrightarrow{\text{P}} 0 \\ \text{IV} &= \left(\frac{X'X}{m} \right)^{-1} \frac{X^*\varepsilon}{m} \xrightarrow{\text{P}} 0 \end{aligned} \quad (3.34)$$

por lo tanto

$$\hat{\beta}_{\text{CM}} \xrightarrow{\text{P}} \beta + Q^{-1} C_{\eta\eta} \beta + Q^{-1} C_{\eta\varepsilon} \quad (3.35)$$

Luego

$$(I - Q^{-1} C_{\eta\eta})^{-1} (\hat{\beta}_{\text{CM}} - Q^{-1} C_{\eta\varepsilon}) \xrightarrow{\text{P}} \beta \quad (3.36)$$

Es posible estimar Q^{-1} con \hat{Q}^{-1} donde

$$\hat{Q} = \frac{X'X}{m} \quad (3.37)$$

el conocimiento de la matriz D se utiliza para estimar $C_{\eta\eta}$ y $C_{\eta\varepsilon}$ por su esperanza

$$\hat{C}_{\eta\eta} = E \left[\frac{\eta'\eta}{m} \right] \quad \text{y} \quad \hat{C}_{\eta\varepsilon} = E \left[\frac{\eta'\varepsilon}{m} \right]. \quad (3.38)$$

El estimador de sesgo corregido se define como

$$\hat{\beta}_{\text{SC}} = \left(I - \hat{Q}^{-1} \hat{C}_{\eta\eta} \right)^{-1} \left(\hat{\beta}_{\text{CM}} - \hat{Q}^{-1} \hat{C}_{\eta\varepsilon} \right). \quad (3.39)$$

Procedimientos parecidos para modelos lineales con distintas hipótesis están desarrollados por Heckman y Leamer en [4]

3.6. Estimación por cuadrados mínimos factibles

Como se observó en la sección anterior, el modelo puede expresarse

$$\begin{aligned} Z &= X\boldsymbol{\beta} + \underbrace{\boldsymbol{\varepsilon} - \boldsymbol{\eta}\boldsymbol{\beta}}_{\boldsymbol{\zeta}(\boldsymbol{\beta})} \\ X &= X^* + \boldsymbol{\eta} \end{aligned} \quad (3.40)$$

donde $\boldsymbol{\zeta}$ es un un término de error que depende del valor verdadero de $\boldsymbol{\beta}$. Su matriz de varianza puede calcularse usando las propiedades (2.3)

$$\text{Var}_{\boldsymbol{\beta}} [\boldsymbol{\zeta}] = \text{Var} [\boldsymbol{\varepsilon}] + \text{Var} [\boldsymbol{\eta}\boldsymbol{\beta}] - \text{Cov} [\boldsymbol{\eta}\boldsymbol{\beta}, \boldsymbol{\varepsilon}] - \text{Cov} [\boldsymbol{\varepsilon}, \boldsymbol{\eta}\boldsymbol{\beta}]. \quad (3.41)$$

Si bien $\boldsymbol{\beta}$ no es conocido $\text{Var}_{\boldsymbol{\beta}} [\boldsymbol{\zeta}]$ puede expresarse en función de él. Notemos que el término

$$\boldsymbol{\eta}\boldsymbol{\beta} = \begin{pmatrix} 0 & \eta_{11} & \dots & \eta_{1p} \\ \vdots & \vdots & & \vdots \\ 0 & \eta_{m1} & \dots & \eta_{mp} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} \quad (3.42)$$

puede re-expresarse para utilizar las propiedades de la varianza como

$$\boldsymbol{\eta}\boldsymbol{\beta} = \underbrace{\begin{pmatrix} \beta_1 \dots \beta_p & 0 \dots 0 & 0 \dots 0 & 0 \dots 0 \\ 0 \dots 0 & \beta_1 \dots \beta_p & 0 \dots 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots \\ 0 \dots 0 & 0 \dots 0 & 0 \dots 0 & \beta_1 \dots \beta_p \end{pmatrix}}_B \begin{pmatrix} \eta_{11} \\ \vdots \\ \eta_{1p} \\ \vdots \\ \eta_{m1} \\ \vdots \\ \eta_{mp} \end{pmatrix} \Bigg\} v(\boldsymbol{\eta}) \quad (3.43)$$

donde $B = B(\boldsymbol{\beta})$ es una matriz de $(m \times mp)$ y $v(\boldsymbol{\eta})$ un vector de $(mp \times 1)$ quedando

$$\text{Var}_{\boldsymbol{\beta}} [\boldsymbol{\zeta}] = D_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} + BD_{\boldsymbol{\eta}\boldsymbol{\eta}}B' + BD_{\boldsymbol{\eta}\boldsymbol{\varepsilon}} + D_{\boldsymbol{\varepsilon}\boldsymbol{\eta}}B' \quad (3.44)$$

donde

$$D_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} = \text{Var} [\boldsymbol{\varepsilon}] \quad D_{\boldsymbol{\eta}\boldsymbol{\eta}} = \text{Var} [v(\boldsymbol{\eta})] \quad D_{\boldsymbol{\eta}\boldsymbol{\varepsilon}} = \text{Cov} [v(\boldsymbol{\eta}), \boldsymbol{\varepsilon}] \quad D_{\boldsymbol{\varepsilon}\boldsymbol{\eta}} = \text{Cov} [\boldsymbol{\varepsilon}, v(\boldsymbol{\eta})]. \quad (3.45)$$

El estimador de cuadrados mínimos factible se define mediante el siguiente procedimiento:

Paso 1: Obtener estimación inicial para $\Sigma = \text{Var}_\beta [\zeta]$.
(utilizando $\hat{\beta}_{\text{SC}}$ definido en (3.39)).

$$\begin{aligned}\hat{\beta}_0 &= \hat{\beta}_{\text{SC}} \\ \hat{B}_0 &= B(\hat{\beta}_0) \\ \hat{\Sigma}_0 &= D_{\varepsilon\varepsilon} + \hat{B}_0 D_{\eta\eta} \hat{B}_0' + \hat{B}_0 D_{\eta\varepsilon} + D_{\varepsilon\eta} \hat{B}_0' \\ n &= 0.\end{aligned}$$

Paso 2: Obtener una estimación de Σ .

Hasta Criterio de terminación ($\hat{\Sigma}_n \approx \hat{\Sigma}_{n-1}$ o $n > N$):

$$\begin{aligned}n &= n + 1 \\ \hat{\beta}_n &= \left(X' \hat{\Sigma}_{n-1}^{-1} X \right)^{-1} X' \hat{\Sigma}_{n-1}^{-1} Z \\ \hat{B}_n &= B(\hat{\beta}_n) \\ \hat{\Sigma}_n &= D_{\varepsilon\varepsilon} + \hat{B}_n D_{\eta\eta} \hat{B}_n' + \hat{B}_n D_{\eta\varepsilon} + D_{\varepsilon\eta} \hat{B}_n' .\end{aligned}$$

Paso 3: Calcular $\hat{\beta}_{\text{CMF}}$.

$$\hat{\beta}_{\text{CMF}} = \left(X' \hat{\Sigma}_n^{-1} X \right)^{-1} X' \hat{\Sigma}_n^{-1} Z.$$

Observación 3.5. El estimador de cuadrados mínimos generalizado no es eficiente para el modelo (3.40) aún si conociéramos $\text{Var} [\zeta]$. El modelo sufre de lo que se conoce como “endogeneidad”. Esto significa que las variables observadas X , que en este caso son aleatorias, están correlacionadas con el error ζ y esto viola las hipótesis que se utilizan para probar la optimalidad de cuadrados mínimos cuando las variables explicativas son aleatorias. De todas formas veremos en las simulaciones que el estimador de cuadrados mínimos factibles obtiene buenos resultados en los modelos lineales.

3.7. Estimadores de cuadrados mínimos

En esta sección exhibimos los estimadores de cuadrados mínimos encontrados en la literatura que utilizan distintas hipótesis sobre la varianza de los errores. Además proponemos y definimos un estimador para el caso más general donde sólo se asume que D es inversible.

En ambos casos de la literatura se asume no correlación entre errores de distintas observaciones. Esto es

$$D_{\varepsilon_t \varepsilon_k} = \mathbf{0} \quad t \neq k. \quad (3.46)$$

Con el fin de facilitar la notación y exponer mas claramente las ideas trabajaremos con $\beta = (\beta_0, \beta_1)^t$. Los procedimientos para $\beta \in \mathbb{R}^{p+1}$ son análogos.

3.7.1. Cuadrados mínimos bajo homoestabilidad

Con $\beta \in \mathbb{R}^2$ el modelo estudiado queda

$$\begin{cases} Z_t = \beta_0 + \beta_1 x_t^* + \varepsilon_t \\ X_t = x_t^* + \eta_t \\ 1 \leq t \leq m. \end{cases} \quad (3.47)$$

En este caso la observación t y el error de la observación t son

$$Y_t = \begin{pmatrix} X_t \\ Z_t \end{pmatrix} = \begin{pmatrix} x_t^* + \eta_t \\ \beta_0 + \beta_1 x_t^* + \varepsilon_t \end{pmatrix} \quad y \quad \epsilon_t = \begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix}. \quad (3.48)$$

La condición de homoestabilidad asume que los errores de las observaciones tienen mismos momentos. Esto es

$$\begin{cases} \text{Var}[\epsilon_t] = D_{\epsilon\epsilon} = \begin{pmatrix} \sigma_{\eta\eta} & \sigma_{\eta\varepsilon} \\ \sigma_{\eta\varepsilon} & \sigma_{\varepsilon\varepsilon} \end{pmatrix} \\ 1 \leq t \leq m. \end{cases} \quad (3.49)$$

Considerando la independencia entre errores de distintas ecuaciones (3.46), la matriz de covarianza de los errores (3.27) queda

$$D = \begin{pmatrix} D_{\epsilon\epsilon} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & D_{\epsilon\epsilon} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & D_{\epsilon\epsilon} \end{pmatrix} \quad (3.50)$$

Notemos que a diferencia de regresión lineal clásica, una observación se puede desviar del punto de la recta que lo originó tanto en dirección vertical como horizontal.

Más precisamente, el punto que originó la observación (X_t, Z_t) es el $(x_t^*, \beta_0 + \beta_1 x_t^*)$ y el cuadrado de la distancia euclídea entre ambos

$$(X_t - x_t^*)^2 + (Z_t - \beta_0 - \beta_1 x_t^*)^2 = \eta_t^2 + \varepsilon_t^2. \quad (3.51)$$

Esta distancia es apropiada como distancia estadística sólo si $\sigma_{\eta\varepsilon} = 0$ y $\sigma_{\eta\eta} = \sigma_{\varepsilon\varepsilon}$.

En cambio, cuando $\sigma_{\eta\varepsilon} \neq 0$ y $\sigma_{\eta\eta} \neq \sigma_{\varepsilon\varepsilon}$ resulta más apropiado considerar otra distancia que tenga esto en cuenta. Por ejemplo,

$$[\text{distancia estadística}]^2 = \frac{\varepsilon_t^2}{\sigma_{\varepsilon\varepsilon}} + \frac{\eta_t^2}{\sigma_{\eta\eta}} \quad (3.52)$$

En general, si $D_{\epsilon\epsilon}$ es inversible, el cuadrado de la distancia estadística de (X_t, Z_t) al punto de la recta que lo generó se obtiene como

$$[\text{distancia estadística}]^2 = \frac{\sigma_{\epsilon\epsilon}\eta_t^2 + \sigma_{\eta\eta}\varepsilon_t^2 - 2\sigma_{\eta\epsilon}\eta_t\varepsilon_t}{\sigma_{\epsilon\epsilon}\sigma_{\eta\eta} - \sigma_{\epsilon\eta}^2} \quad (3.53)$$

que es equivalente a la siguiente expresión que se conoce como distancia de Mahalanobis

$$(X_t - x_t^*, Z_t - \beta_0 - \beta_1 x_t^*) D_{\epsilon\epsilon}^{-1} (X_t - x_t^*, Z_t - \beta_0 - \beta_1 x_t^*)^t. \quad (3.54)$$

El estimador de cuadrados mínimos $(\hat{\beta}_0, \hat{\beta}_1)$ será quien minimice la suma de los cuadrados de las distancias estadísticas de las observaciones a la recta definida por $z = \beta_0 + \beta_1 x$.

Para esto se debe encontrar $(\hat{\beta}_0, \hat{\beta}_1, \hat{x}^*) = (\hat{\beta}_0, \hat{\beta}_1, \hat{x}_1, \dots, \hat{x}_m)$ que minimicen

$$S(\beta_0, \beta_1, x^*) = \sum_{t=1}^m (X_t - x_t^*, Z_t - \beta_0 - \beta_1 x_t^*) D_{\epsilon\epsilon}^{-1} (X_t - x_t^*, Z_t - \beta_0 - \beta_1 x_t^*)^t. \quad (3.55)$$

Como cada término de la suma es positivo, pues las $D_{\epsilon\epsilon}^{-1}$ son definidas positivas, basta minimizar cada uno de ellos para encontrar x^* .

Se puede expresar el valor x_t^* que minimiza (3.54) como una función de (β_0, β_1) . De la propiedad (2.14) para un (β_0, β_1) esta queda

$$\ddot{x}_t(\beta_0, \beta_1) = [(1, \beta_1) D_{\epsilon\epsilon}^{-1} (1, \beta_1)^t]^{-1} (1, \beta_1) D_{\epsilon\epsilon}^{-1} (X_t, Z_t - \beta_0)^t \quad (3.56)$$

luego \ddot{x}_t es una función que para cada (β_0, β_1) devuelve el estimador de cuadrados mínimos generalizado del problema

$$\left\{ \begin{array}{l} \begin{pmatrix} X_t \\ Z_t - \beta_0 \end{pmatrix} = \begin{pmatrix} 1 \\ \beta_1 \end{pmatrix} x_t^* + \begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \\ \mathbb{E} \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} = \mathbf{0} \quad \text{Var} \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} = D_{\epsilon\epsilon} \end{array} \right. \quad (3.57)$$

Siempre que el contexto resulte claro omitiremos el valor donde se evalúa la función \ddot{x}_t . Observemos que si se considera \mathbb{R}^2 con la norma inducida por $D_{\epsilon\epsilon}^{-1}$, el punto $(\ddot{x}_t, \beta_1 \ddot{x}_t + \beta_0)$ es la proyección de (X_t, Z_t) sobre la recta $z = \beta_1 x + \beta_0$.

Luego minimizar (3.55) equivale a encontrar $(\hat{\beta}_0, \hat{\beta}_1)$ que minimicen

$$\tilde{S}(\beta_0, \beta_1) = \left[\sum_{t=1}^m (X_t - \ddot{x}_t, Z_t - \beta_0 - \beta_1 \ddot{x}_t) D_{\epsilon\epsilon}^{-1} (X_t - \ddot{x}_t, Z_t - \beta_0 - \beta_1 \ddot{x}_t)^t \right]. \quad (3.58)$$

Este problema tiene solución cerrada y su desarrollo puede encontrarse en el capítulo 1 de [1]. En el caso de $\sigma_{\eta\epsilon} = 0$ y $\sigma_{\epsilon\epsilon} = \sigma_{\eta\eta}$ la recta encontrada es la que minimiza la suma de los cuadrados de las distancias ortogonales.

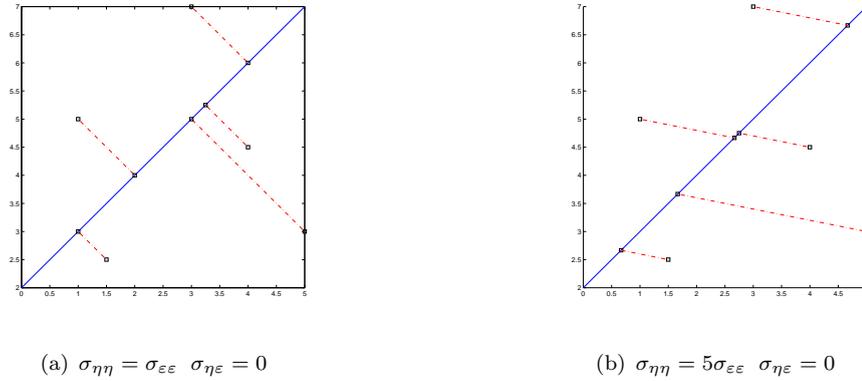


FIGURA 3.5: Proyección de observaciones sobre una recta con distintas distancias estadísticas en cada figura

3.7.2. Cuadrados mínimos bajo heteroscedasticidad

Un conjunto de variables aleatorias se dice que sufre de heteroscedasticidad si existen subconjuntos con distinta variabilidad.

La presencia de heteroscedasticidad es un problema que no puede pasarse por alto ya que, en general, estimadores óptimos bajo homocedasticidad resultan inadmisibles.

Se estudia el modelo de (3.47) manteniendo la independencia entre errores de ecuación (3.46), pero cambiando la hipótesis (3.49) por

$$\begin{cases} \text{Var}[\epsilon_t] = D_{\epsilon_t\epsilon_t} = \begin{pmatrix} \sigma_{\eta_t\eta_t} & \sigma_{\eta_t\epsilon_t} \\ \sigma_{\eta_t\epsilon_t} & \sigma_{\epsilon_t\epsilon_t} \end{pmatrix} \\ 1 \leq t \leq n \end{cases} \quad (3.59)$$

esto es, permitir que cada observación tenga matriz de varianza distinta, entonces la matriz de varianza de los errores queda

$$D = \begin{pmatrix} D_{\epsilon_1\epsilon_1} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & D_{\epsilon_2\epsilon_2} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & D_{\epsilon_m\epsilon_m} \end{pmatrix} \quad (3.60)$$

Chan y Mak [12] definen el estimador de cuadrados mínimos como los valores $(\hat{\beta}_0, \hat{\beta}_1)$ que minimizan la suma de los cuadrados de las distancias estadísticas de las observaciones a la recta definida por $z = \beta_0 + \beta_1 x$. Para esto se debe encontrar $(\hat{\beta}_0, \hat{\beta}_1, \hat{x}_1, \dots, \hat{x}_m)$

que minimicen

$$S(\beta_0, \beta_1, x^*) = \sum_{t=1}^m (X_t - x_t^*, Z_t - \beta_0 - \beta_1 x_t^*) D_{\epsilon_t \epsilon_t}^{-1} (X_t - x_t^*, Z_t - \beta_0 - \beta_1 x_t^*)^t \quad (3.61)$$

Matei en su tesis [14] afirma que, en general, no existe una fórmula cerrada para minimizar (3.61) y resuelve el problema numéricamente.

La complejidad del problema se debe a que bajo heteroscedasticidad los errores de las ecuaciones sufren de asintropía y de no homogeneidad. Esto hace que las proyecciones de una observación sobre la recta que la generó tengan distinto ángulo.

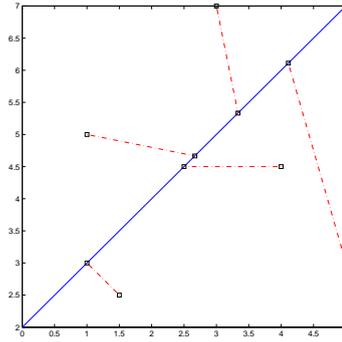


FIGURA 3.6: Proyección de observaciones sobre una recta con distintas $D_{\epsilon_t \epsilon_t}$.

Como en el caso de homocedasticidad, fijado el (β_0, β_1) el término de la suma que depende de t se minimiza en

$$\ddot{x}_t(\beta_0, \beta_1) = [(1, \beta_1) D_{\epsilon_t \epsilon_t}^{-1} (1, \beta_1)^t]^{-1} (1, \beta_1) D_{\epsilon_t \epsilon_t}^{-1} (X_t, Z_t - \beta_0)^t. \quad (3.62)$$

Luego optimizar (3.61) como función de $n + 2$ parámetros es equivalente a minimizar

$$\tilde{S}(\beta_0, \beta_1) = \sum_{t=1}^m (\ddot{x}_t - X_t, Z_t - \beta_0 - \beta_1 \ddot{x}_t) D_{\epsilon_t \epsilon_t}^{-1} (\ddot{x}_t - X_t, Z_t - \beta_0 - \beta_1 \ddot{x}_t)^t. \quad (3.63)$$

Como las matrices $D_{\epsilon_t \epsilon_t}^{-1}$ son definidas positivas, cada uno de los términos de la suma es no negativo y S resulta una suma de cuadrados. Esta propiedad permite usar algoritmos numéricos más eficientes que el descenso del gradiente.

3.7.3. Estimador de cuadrados mínimos generalizado

Vamos a proponer un estimador de cuadrados mínimos para el problema general

$$\begin{cases} Z_t = \beta_0 + \beta_1 x_t^* + \varepsilon_t \\ X_t = x_t^* + \eta_t \\ 1 \leq t \leq m \end{cases} \quad (3.64)$$

con el vector de errores cumpliendo las hipótesis planteadas al comienzo del capítulo. Esto es,

$$\boldsymbol{\xi} = \begin{pmatrix} \eta_1 \\ \varepsilon_1 \\ \vdots \\ \eta_m \\ \varepsilon_m \end{pmatrix} \quad (3.65)$$

cumple que

$$E[\boldsymbol{\xi}] = \mathbf{0} \quad \text{y} \quad \text{Var}[\boldsymbol{\xi}] = D \quad (3.66)$$

y asumimos que la matriz D es **conocida e inversible**.

Si llamamos

$$M = \begin{pmatrix} X_1 - x_1^* \\ Z_1 - \beta_0 - \beta_1 x_1^* \\ \vdots \\ X_m - x_m^* \\ Z_m - \beta_0 - \beta_1 x_m^* \end{pmatrix} \quad (3.67)$$

definimos el estimador de $(\hat{\beta}_0, \hat{\beta}_1)$ que se obtienen de $(\hat{\beta}_0, \hat{\beta}_1, \hat{x}_1, \dots, \hat{x}_m)$ que minimicen

$$S(\beta_0, \beta_1, x^*) = M^t D^{-1} M \quad (3.68)$$

Esto no es otra cosa que minimizar la distancia Mahalanobis de los errores a su media cero y veremos en la siguiente sección que es una suma de cuadrados. Como se comentó antes, este estimador es una generalización de los antes presentados.

Bajo las hipótesis de la sección anterior, con

$$D = \begin{pmatrix} D_{\varepsilon_1 \varepsilon_1} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & D_{\varepsilon_2 \varepsilon_2} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & D_{\varepsilon_m \varepsilon_m} \end{pmatrix} \quad (3.69)$$

es posible reescribir la sumatoria a minimizar de (3.61) como

$$M^t \begin{pmatrix} D_{\epsilon_1 \epsilon_1}^{-1} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & D_{\epsilon_2 \epsilon_2}^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & D_{\epsilon_m \epsilon_m}^{-1} \end{pmatrix} M. \quad (3.70)$$

La generalización queda clara cuando se nota que

$$\begin{pmatrix} D_{\epsilon_1 \epsilon_1}^{-1} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & D_{\epsilon_2 \epsilon_2}^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & D_{\epsilon_m \epsilon_m}^{-1} \end{pmatrix} = \begin{pmatrix} D_{\epsilon_1 \epsilon_1} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & D_{\epsilon_2 \epsilon_2} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & D_{\epsilon_m \epsilon_m} \end{pmatrix}^{-1}. \quad (3.71)$$

Además el estimador tiene la propiedad de ser el de máxima verosimilitud cuando los errores tienen distribución normal multivariada como formalizamos para una familia de distribuciones aún mas amplia.

Definición 3.6. Decimos que una densidad g es elíptica si existe Σ definida positiva, $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ tales que

$$g(x) = f((x - \mu)' \Sigma (x - \mu)) \quad (3.72)$$

donde μ es la mediana de la densidad.

Definición 3.7. Decimos que una densidad g es elíptica decreciente si la f es decreciente.

Propiedad 3.8. Si una densidad elíptica tiene esperanza finita coincide con su mediana. Si tiene matriz de covarianza D es proporcional a Σ . Es decir, $D = k\Sigma$ con $k \in \mathbb{R}$

Teorema 3.9. Si el vector de errores tiene una densidad elíptica decreciente entonces el estimador propuesto bajo las hipótesis planteadas al principio de la sección es el de máxima verosimilitud.

Dem 3.10. Llamando

$$Y = \begin{pmatrix} X_1 \\ Z_1 \\ \vdots \\ X_m \\ Z_m \end{pmatrix} \quad \text{y} \quad w = \begin{pmatrix} x_1^* \\ \beta_0 + \beta_1 x_1^* \\ \vdots \\ x_m^* \\ \beta_0 + \beta_1 x_m^* \end{pmatrix} \quad (3.73)$$

si B un boreliano de \mathbb{R}^{2n} entonces

$$\begin{aligned} P(Y \in B \mid (\beta_0, \beta_1, x_1^*, \dots, x_m^*)) &= P(Y - w \in (B - w) \mid (\beta_0, \beta_1, x_1^*, \dots, x_m^*)) \\ &= P(\xi \in (B - w)) \end{aligned} \quad (3.74)$$

luego la función de verosimilitud queda

$$\mathcal{L}(\beta_0, \beta_1 x_1^*, \dots, x_m^* | Y) = f(M^t D^{-1} M) \quad (3.75)$$

que por hipótesis se maximiza cuando $M^t D^{-1} M$ es mínimo.

Observación 3.11. La densidad de la normal multivariada con media μ y varianza Σ

$$g(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu)\right) \quad (3.76)$$

es elíptica decreciente.

Observación 3.12. La densidad de la distribución “ t-multivariada ” con media μ , matriz de varianza Σ y ν grados de libertad

$$g(x) = \frac{\Gamma[(\nu + d)/2]}{\Gamma(\nu/2) |\Sigma|^{\frac{1}{2}}} \left(1 + \frac{1}{\nu} (x - \mu)' \Sigma^{-1} (x - \mu)\right)^{-\frac{\nu+d}{2}} \quad (3.77)$$

es elíptica decreciente. Esta distribución es la generalización de la t de student para dimensión d .

3.7.4. Caso general e implementación del estimador

Para el caso general, con $\beta \in \mathbb{R}^{p+1}$ es necesario definir nueva notación

$$Y_t = \begin{pmatrix} X_{t1} \\ \vdots \\ X_{tp} \\ Z_t \end{pmatrix} \quad \text{y} \quad w_t = \begin{pmatrix} x_{t1}^* \\ \vdots \\ x_{tp}^* \\ z_t^* \end{pmatrix} \quad (3.78)$$

$$z_t^* = \beta_0 + \beta_1 x_{t1}^* + \dots + \beta_p x_{tp}^* \quad (3.79)$$

se definen también Y y w de $m(p+1) \times 1$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix} \quad \text{y} \quad w = \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}. \quad (3.80)$$

Con $M = w - Y$, se puede formular el estimador de cuadrados mínimos para $\beta \in \mathbb{R}^{p+1}$ como el $\hat{\beta}$ de $(\hat{\beta}, \hat{\mathbf{x}}^*)$ que minimiza

$$S(\beta, \mathbf{x}^*) = M^t D^{-1} M \quad (3.81)$$

donde

$$(\beta, \mathbf{x}^*) = (\beta_0, \beta_1, \dots, \beta_p, x_{11}^*, \dots, x_{1p}^*, \dots, x_{m1}^*, \dots, x_{mp}^*). \quad (3.82)$$

La función objetivo tiene $p(m+1)$ variables, pero de manera similar a (3.61), su optimización resulta equivalente a minimizar otra función que depende solo de β .

Para un valor de β fijo el vector M es una variedad lineal de \mathbf{x}^* .

$$M = \tilde{B}\mathbf{x}^* - N \quad (3.83)$$

donde \tilde{B} es una matriz de bloques de tamaño $m(p+1) \times mp$ y N de $m(p+1) \times 1$

$$\tilde{B} = \tilde{B}(\beta) = \begin{pmatrix} B & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & B & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & B \end{pmatrix} \quad \text{y} \quad N = \begin{pmatrix} N_1 \\ N_2 \\ \vdots \\ \vdots \\ N_m \end{pmatrix} \quad (3.84)$$

con cada bloque B de $(p+1) \times p$ y N_t de $p+1$

$$B = B(\beta) = \begin{pmatrix} 1 & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & 1 \\ \beta_1 & \beta_2 & \dots & \dots & \beta_p \end{pmatrix} \quad \text{y} \quad N_t = N_t(\beta) = \begin{pmatrix} X_1 \\ \vdots \\ X_m \\ Z_t - \beta_0 \end{pmatrix}. \quad (3.85)$$

Por la propiedad (2.14) de las matrices definida positivas, el valor de \mathbf{x}^* donde (3.81) alcanza mínimo es una función de β

$$\ddot{\mathbf{x}}(\beta) = (F'D^{-1}F)^{-1} F'D^{-1}N \quad (3.86)$$

Con todo esto, optimizar $S(\beta, \mathbf{x}^*)$ equivale a optimizar

$$\tilde{S}(\beta) = S(\beta, \ddot{\mathbf{x}}(\beta)). \quad (3.87)$$

Observación 3.13. Trabajando un poco más con la estructura matricial de \tilde{B} se puede conseguir una mejor implementación numérica y obtener más entendimiento de lo que sucede.

Utilizando las propiedades del producto de matrices por bloques

$$\tilde{B}'D^{-1} = \begin{pmatrix} B'D_{\epsilon_1\epsilon_1}^{-1} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & B'D_{\epsilon_2\epsilon_2}^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & B'D_{\epsilon_m\epsilon_m}^{-1} \end{pmatrix} \quad (3.88)$$

y utilizando la propiedad de la inversa de una matriz diagonal por bloques

$$\left(\tilde{B}'D^{-1}\tilde{B}\right)^{-1} = \begin{pmatrix} (B'D_{\epsilon_1\epsilon_1}^{-1}B)^{-1} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & (B'D_{\epsilon_2\epsilon_2}^{-1}B)^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & (B'D_{\epsilon_m\epsilon_m}^{-1}B)^{-1} \end{pmatrix}. \quad (3.89)$$

La fórmula (3.86) puede expresarse

$$\ddot{\mathbf{x}} = \begin{pmatrix} (B'D_{\epsilon_1\epsilon_1}^{-1}B)^{-1}B'D_{\epsilon_1\epsilon_1}^{-1} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & (B'D_{\epsilon_2\epsilon_2}^{-1}B)^{-1}B'D_{\epsilon_2\epsilon_2}^{-1} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & (B'D_{\epsilon_m\epsilon_m}^{-1}B)^{-1}B'D_{\epsilon_m\epsilon_m}^{-1} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \\ \vdots \\ \vdots \\ N_m \end{pmatrix} \quad (3.90)$$

Dando un resultado análogo al de (3.62) ya que para cada observación t

$$\ddot{\mathbf{x}}_t(\boldsymbol{\beta}) = (B'D_{\epsilon_t\epsilon_t}^{-1}B)^{-1}B'D_{\epsilon_t\epsilon_t}^{-1}N_t \quad (3.91)$$

donde

$$\ddot{\mathbf{x}}_t(\boldsymbol{\beta}) = \begin{pmatrix} \ddot{x}_{t1}(\boldsymbol{\beta}) \\ \vdots \\ \ddot{x}_{tp}(\boldsymbol{\beta}) \end{pmatrix} \quad \text{y} \quad \ddot{\mathbf{x}} = \begin{pmatrix} \ddot{\mathbf{x}}_1 \\ \vdots \\ \ddot{\mathbf{x}}_m \end{pmatrix}. \quad (3.92)$$

Esto nos dice que para un $\boldsymbol{\beta}$ dado la estimación de los valores de la observación t no se ve afectada por las otras observaciones.

La función objetivo \tilde{S} tiene la importante propiedad de ser una suma de cuadrados. Si $D^{-1} = U'U$ donde U es la descomposición de Cholesky (2.10)

$$\begin{aligned} M'D^{-1}M &= (UM)'UM \\ &= \|UM\|^2 \end{aligned} \quad (3.93)$$

luego

$$\tilde{S}(\boldsymbol{\beta}) = \|(s_1(\boldsymbol{\beta}), \dots, s_m(\boldsymbol{\beta}))\|^2 = (s_1(\boldsymbol{\beta}))^2 \dots + (s_m(\boldsymbol{\beta}))^2 \quad (3.94)$$

con

$$s(\boldsymbol{\beta}) = \begin{pmatrix} s_1(\boldsymbol{\beta}) \\ \vdots \\ s_m(\boldsymbol{\beta}) \end{pmatrix} = U(\tilde{B}(\boldsymbol{\beta})\ddot{\mathbf{x}}(\boldsymbol{\beta}) - N(\boldsymbol{\beta})). \quad (3.95)$$

Para obtener la matriz diferencial de s basta conocer el vector de derivadas respecto β_k

$$\frac{\partial s}{\partial \beta_k} = U \left(\frac{\partial \tilde{B}}{\partial \beta_k} \ddot{\mathbf{x}} + \tilde{B} \frac{\partial \ddot{\mathbf{x}}}{\partial \beta_k} - \frac{\partial N}{\partial \beta_k} \right) \quad (3.96)$$

Las derivadas de F y N resultan triviales. Para obtener las de $\ddot{\mathbf{x}}$ utilizamos que

$$\left(\tilde{B}'D^{-1}\tilde{B}\right)\ddot{\mathbf{x}} = \tilde{B}'D^{-1}N \quad (3.97)$$

y derivando en ambos lados de la igualdad se obtiene la siguiente ecuación

$$\begin{aligned} \left(\tilde{B}'D^{-1}\tilde{B}\right)\frac{\partial\ddot{\mathbf{x}}}{\partial\beta_k} &= \frac{\partial\left(\tilde{B}'D^{-1}N\right)}{\partial\beta_k} - \frac{\partial\left(\tilde{B}'D^{-1}\tilde{B}\right)}{\partial\beta_k}\ddot{\mathbf{x}} \\ &= \frac{\partial\tilde{B}'}{\partial\beta_k}D^{-1}N + \tilde{B}'D^{-1}\frac{\partial N}{\partial\beta_k} - \left(\frac{\partial\tilde{B}'}{\partial\beta_k}\left(D^{-1}\tilde{B}\right) + \left(\tilde{B}'D^{-1}\right)\frac{\partial\tilde{B}}{\partial\beta_k}\right)\ddot{\mathbf{x}}. \end{aligned} \quad (3.98)$$

Implementación numérica

Para la optimización numérica de la función objetivo se pueden utilizar dos algoritmos ampliamente estudiados en la literatura para problemas de suma de cuadrados sin restricciones en las variables. Estos son el algoritmo de Newton-Raphson y el algoritmo de Levenverg-Marquardt. Ambos resultan más eficientes que el descenso del gradiente. Son algoritmos de búsqueda local por lo que es necesario proveer una solución inicial. Un problema habitual de estos métodos locales es que en funciones con múltiples mínimos locales no se puede garantizar convergencia al mínimo global. Esta convergencia depende de la calidad de la solución inicial.

Las diferencias entre ambos algoritmos se ven reflejadas en que el de Newton-Raphson tiene mayor velocidad de convergencia cerca del óptimo y en cambio el Levenverg-Marquardt es más robusto respecto a la solución inicial. Existen numerosos softwares que tienen implementados estos métodos y que para utilizarlos sólo hay que proveerles las funciones s_i y sus respectivas funciones de gradientes. En el caso concreto del problema de variables con error, la solución inicial puede obtenerse a partir de otro estimador de β como el de sesgo corregido o incluso haciendo cuadrados mínimos como si no hubiera errores en las variables. Si no se confía en la calidad de la solución inicial Levenverg-Marquardt es más aconsejable.

La implementación del cálculo de $\ddot{\mathbf{x}}$ y $\frac{\partial\ddot{\mathbf{x}}}{\partial\beta_k}$ deben hacerse con cuidado por problemas de estabilidad numérica. En general es más conveniente resolver sistemas de ecuaciones que calcular inversas y utilizar la descomposición de Cholesky en vez de calcular D^{-1} .

Capítulo 4

Bootstrap

La técnica de bootstrap se utiliza para calcular propiedades de estimadores cuya distribución resulta difícil o imposible de calcular. Fue introducida por Efron [9] y sin duda constituye uno de los principales avances dentro de la estadística. La idea principal en la cual se basa esta técnica puede describirse como sigue.

Sea Y_1, \dots, Y_n una muestra aleatoria de una distribución F que tiene un parámetro de interés $\theta = T(F)$, donde T es un funcional sobre un espacio de funciones de distribución. Sea $\hat{\theta} = h(Y_1, \dots, Y_n)$ un estimador de θ que tiene una función de distribución G que está determinada por h y F . Es decir $G = G(h, F)$.

El bootstrap permite estimar la distribución de $\hat{\theta}$ con un procedimiento que puede describirse en dos pasos:

Paso 1: Estimar F por \tilde{F} .

Paso 2: Estimar G por $\tilde{G} = G(h, \tilde{F})$.

A \tilde{G} se la conoce como distribución bootstrap, en ciertas ocasiones puede calcularse analíticamente, pero en la mayoría de las aplicaciones se aproxima mediante el método de Montecarlo. Este método permite obtener estimaciones de ciertas propiedades de $\hat{\theta}$ a partir de la distribución \tilde{G} . Es decir, las propiedades bajo la distribución G se estimarán por las mismas propiedades bajo la distribución \tilde{G} . Por ejemplo, $E_G[\hat{\theta}]$ se estimará por $E_{\tilde{G}}[\hat{\theta}]$.

Si W_1, \dots, W_n es una muestra aleatoria de la distribución \tilde{F} , entonces $\hat{\theta}^b = h(W_1, \dots, W_n)$ tendrá distribución \tilde{G} . Luego, es suficiente saber generar muestras \tilde{F} para generar muestras con distribución \tilde{G} .

Se llama muestra bootstrap de B replicaciones a $\hat{\theta}_1^b, \dots, \hat{\theta}_B^b$ una muestra aleatoria generada por el proceso recién descrito. La muestra bootstrap sigue la distribución \tilde{G} .

4.1. Estimación de Media, varianza, distribución y cuantiles

Asumiendo resuelto cómo estimar F y cómo obtener muestras de su estimación, y por ende también, cómo generar $\hat{\theta}_1^b, \dots, \hat{\theta}_B^b$ se pueden estimar

la media de $\hat{\theta}$

$$\mu_{bt} = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i^b, \quad (4.1)$$

la varianza de $\hat{\theta}$,

$$\sigma_{bt}^2 = \frac{1}{B-1} \sum_{i=1}^B (\hat{\theta}_i^b - \mu_{bt})^2, \quad (4.2)$$

valores de la distribución $G(t) = P(\hat{\theta} \leq t)$,

$$G_{bt}(t) = \frac{1}{B} \sum_{i=1}^B \mathbb{1}_{\hat{\theta}_i^b \leq t} \quad (4.3)$$

y cuantiles de la distribución, $G^{-1}(\alpha) = \inf\{k \in \mathbb{R} : G(k) \leq \alpha\}$,

$$G_{bt}^{-1}(\alpha) = \text{Cuantil } \alpha \text{ de } \{\hat{\theta}_1^b, \dots, \hat{\theta}_B^b\}. \quad (4.4)$$

Como las estimaciones se obtienen a partir de una muestra de \tilde{G} no es cierto, por ejemplo, que $\mu_{bt} = E_{\tilde{G}}[\hat{\theta}]$. Sin embargo, por la ley de los grandes números, $\mu_{bt} \xrightarrow{B \rightarrow \infty} E_{\tilde{G}}[\hat{\theta}]$. Lo mismo sucede con las otras estimaciones y es por esto que es importante tomar B suficientemente grande.

4.2. Estimación del ECM e intervalos de confianza

Para determinar el error cuadrático medio (ECM) e intervalos de confianza es necesario conocer la media y ciertos cuantiles de la distribución de $\hat{\theta} - \theta$.

Recordemos que el ECM de $\hat{\theta}$ es,

$$\text{ECM}[\hat{\theta}] = E[\hat{\theta} - \theta]^2 + \text{Var}[\hat{\theta}] \quad (4.5)$$

y que un intervalo de confianza de nivel $1 - \alpha$ para θ es un intervalo $IC_\alpha(\hat{\theta})$ tal que $P(\theta \in IC_\alpha(\hat{\theta})) = 1 - \alpha$.

Si k_1, k_2 satisfacen

$$\begin{cases} P(\hat{\theta} - \theta < k_1) = \alpha/2 \\ P(\hat{\theta} - \theta \leq k_2) = 1 - \alpha/2 \end{cases} \quad (4.6)$$

el siguiente intervalo tiene la propiedad deseada

$$\text{IC}_\alpha(\hat{\theta}) = [\hat{\theta} - k_2, \hat{\theta} - k_1]. \quad (4.7)$$

Notemos que, si H es la distribución de $\hat{\theta} - \theta$, entonces podemos estimarla por $\tilde{H}(t) = \tilde{G}(t + \hat{\theta})$, ya que

$$H(t) = \text{P}(\hat{\theta} - \theta \leq t) = \text{P}(\hat{\theta} \leq t + \theta) = G(t + \theta). \quad (4.8)$$

Luego, $\hat{\theta}_1^b - \hat{\theta}, \dots, \hat{\theta}_B^b - \hat{\theta}$ resulta una muestra aleatoria de la distribución \tilde{H} y por lo tanto podemos estimar la esperanza de $\hat{\theta} - \theta$ a partir de

$$\hat{\text{E}}[\hat{\theta} - \theta] = \frac{1}{B} \sum_{i=1}^B (\hat{\theta}_i^b - \hat{\theta}) = \mu_{\text{bt}} - \hat{\theta}. \quad (4.9)$$

El error cuadrático medio de $\hat{\theta}$ puede estimarse utilizando la muestra bootstrap como

$$\text{ECM}_{\text{bt}} = (\mu_{\text{bt}} - \hat{\theta})^2 + \sigma_{\text{bt}}^2. \quad (4.10)$$

Para estimar k_1 y k_2 se calculan los cuantiles correspondientes de $\hat{\theta}_1^b - \hat{\theta}, \dots, \hat{\theta}_B^b - \hat{\theta}$

$$\begin{aligned} \hat{k}_1 &= G_{\text{bt}}^{-1}(\alpha/2) - \hat{\theta} \\ \hat{k}_2 &= G_{\text{bt}}^{-1}(1 - \alpha/2) - \hat{\theta} \end{aligned} \quad (4.11)$$

luego, la estimación del intervalo de confianza resulta

$$\text{IC}_\alpha^{\text{bt}}[\hat{\theta}] = [2\hat{\theta} - G_{\text{bt}}^{-1}(1 - \alpha/2), 2\hat{\theta} - G_{\text{bt}}^{-1}(\alpha/2)]. \quad (4.12)$$

Intervalo de confianza a partir de un pivot

Existen otras formas de determinar intervalos de confianza para un estimador. A continuación describiremos brevemente cómo hacerlo a partir de la estimación de la distribución de un pivot.

Sea k_α que satisface $\text{P}\left(\frac{(\hat{\theta} - \theta)^2}{\text{Var}[\hat{\theta}]} \leq k_\alpha\right) = 1 - \alpha$, luego, el intervalo aleatorio

$$\text{IC}_\alpha(\hat{\theta}) = \left[\hat{\theta} - \sqrt{k_\alpha \text{Var}[\hat{\theta}]}, \hat{\theta} + \sqrt{k_\alpha \text{Var}[\hat{\theta}]} \right] \quad (4.13)$$

cumple que $P\left(\theta \in \text{IC}_\alpha(\hat{\theta})\right) = 1 - \alpha$.

Notemos que k_α es un cuantil de la distribución de $\frac{(\hat{\theta} - \theta)^2}{\text{Var}[\hat{\theta}]}$ y puede ser estimado por

$$\hat{k}_\alpha = \text{Cuantil } 1 - \alpha \text{ de } \left\{ \frac{(\hat{\theta}_1^b - \hat{\theta})^2}{\sigma_{\text{bt}}^2}, \dots, \frac{(\hat{\theta}_B^b - \hat{\theta})^2}{\sigma_{\text{bt}}^2} \right\} \quad (4.14)$$

por lo tanto, una posible estimación de un intervalo de confianza puede definirse como

$$\text{IC}_\alpha(\hat{\theta}) = \left[\hat{\theta} - \sqrt{\hat{k}_\alpha \sigma_{\text{bt}}^2}, \hat{\theta} + \sqrt{\hat{k}_\alpha \sigma_{\text{bt}}^2} \right] \quad (4.15)$$

4.3. Bootstrap para estimación multivariada

En el caso de estimadores multivariados también puede utilizarse técnicas bootstrap para estimar media, varianza, ECM, y regiones de confianza. Análogo al caso univariado, Y_1, \dots, Y_n es una muestra aleatoria de distribución F y $\hat{\theta} = h(Y_1, \dots, Y_n)$ un estimador de $\theta \in \mathbb{R}^p$ un parámetro de interés de la distribución F con distribución $G = G(h, F)$. Se estima F con \tilde{F} y G con $\tilde{G} = G(h, \tilde{F})$. Luego puede generarse una muestra bootstrap $\hat{\theta}_1^b, \dots, \hat{\theta}_B^b$ de la misma forma que en el caso univariado y estimar las cantidades de interés.

la media de $\hat{\theta}$

$$\mu_{\text{bt}} = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i^b \quad (4.16)$$

la matriz de varianza de $\hat{\theta}$

$$S_{\text{bt}} = \frac{1}{B-1} \sum_{i=1}^B (\hat{\theta}_i^b - \mu_{\text{bt}})(\hat{\theta}_i^b - \mu_{\text{bt}})' \quad (4.17)$$

valores de la distribución G

$$G_{\text{bt}}(t_1, \dots, t_p) = \frac{1}{B} \sum_{i=1}^B \left(\prod_{k=1}^p \mathbb{1}_{\hat{\theta}_{ik}^b \leq t_k} \right) \quad (4.18)$$

el error cuadrático medio de $\hat{\theta}$

$$\text{ECM}_{\text{bt}} = (\hat{\theta} - \mu_{\text{bt}})(\hat{\theta} - \mu_{\text{bt}})' + S_{\text{bt}}. \quad (4.19)$$

4.4. Estimación de una región de confianza

Como en el caso univariado, recordamos la idea de una región de confianza. Para ello consideremos k_α tal que

$$P \left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \left(\text{Var} \left[\hat{\boldsymbol{\theta}} \right] \right)^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq k_\alpha \right) = 1 - \alpha \quad (4.20)$$

luego, una región de confianza de nivel $1 - \alpha$ para $\boldsymbol{\theta}$ es

$$R_\alpha(\hat{\boldsymbol{\theta}}) = \left\{ \boldsymbol{\mu} \in \Theta : (\hat{\boldsymbol{\theta}} - \boldsymbol{\mu})' \left(\text{Var} \left[\hat{\boldsymbol{\theta}} \right] \right)^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}) \leq k_\alpha \right\}. \quad (4.21)$$

Observemos que esta región es un p-elipsoide en \mathbb{R}^p y se puede estimar utilizando la técnica bootstrap por

$$R_\alpha^{\text{bt}}(\hat{\boldsymbol{\theta}}) = \left\{ \boldsymbol{\mu} \in \Theta : (\hat{\boldsymbol{\theta}} - \boldsymbol{\mu})' (S_{\text{bt}})^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}) \leq \hat{k}_\alpha \right\} \quad (4.22)$$

donde \hat{k}_α es la estimación de k_α

$$\hat{k}_\alpha = \text{Cuantil } 1-\alpha \text{ de } \left\{ (\hat{\boldsymbol{\theta}}_1^b - \boldsymbol{\mu}_{\text{bt}})' (S_{\text{bt}})^{-1} (\hat{\boldsymbol{\theta}}_1^b - \boldsymbol{\mu}_{\text{bt}}), \dots, (\hat{\boldsymbol{\theta}}_B^b - \boldsymbol{\mu}_{\text{bt}})' (S_{\text{bt}})^{-1} (\hat{\boldsymbol{\theta}}_B^b - \boldsymbol{\mu}_{\text{bt}}) \right\}. \quad (4.23)$$

4.5. Estimación de la distribución de una muestra

El paso más delicado en el bootstrap es el de estimar la distribución F . Para eso existen diversas técnicas que dependen del conocimiento que se tenga sobre la distribución. Cada técnica da un método bootstrap distinto. Es clave comprender que la calidad de las estimaciones bootstrap depende directamente de la calidad de la estimación \tilde{F} . Si se asume que la distribución F vive en una familia paramétrica F_η donde η es un parámetro que se estima con $\hat{\eta}$, entonces la estimación de F puede hacerse como

$$\tilde{F} = F_{\hat{\eta}}. \quad (4.24)$$

Esto es conocido como bootstrap paramétrico.

Existen otras variantes al bootstrap paramétrico, como el bootstrap por resamplero o el wild bootstrap o el bootstrap noparamétrico o el bootstrap suavizado entre otros que no desarrollaremos aquí. Algunas referencias que pueden consultarse son los libros de Chernick [7], Davison y Hinkley [8] o Efron y Tibshirani [10].

4.6. Aplicación a los modelos lineales con errores en las variables

Hasta ahora solo conocíamos la covarianza de los errores. Ahora vamos a asumir conocida su distribución ya que condice con nuestro problema de interés. Por lo tanto, podemos aplicar bootstrap paramétrico para cualquiera de los estimadores propuestos. Con el fin de mantener sencilla la notación trabajaremos con $\beta = (\beta_0, \beta_1)$. Como las observaciones no son independientes, es necesario considerar Y como única muestra ($n = 1$) donde

$$Y = \begin{pmatrix} X_1 \\ Z_1 \\ \vdots \\ X_m \\ Z_m \end{pmatrix} = \begin{pmatrix} x_1^* \\ \beta_1 x_1^* + \beta_0 \\ \vdots \\ x_m^* \\ \beta_1 x_m^* + \beta_0 \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \varepsilon_1 \\ \vdots \\ \eta_m \\ \varepsilon_m \end{pmatrix}. \quad (4.25)$$

Por lo tanto, la distribución F de la muestra Y , está caracterizada por la distribución del vector de errores y los parámetros $\beta_1, \beta_0, x_1^*, \dots, x_m^*$. Como la primera se asume conocida F pertenece a una familia paramétrica $F_{\beta_1, \beta_0, x_1^*, \dots, x_m^*}$.

Los estimadores propuestos en (3.5), (3.6) y (3.7.3) son estimadores de β , por lo tanto para estimar F faltan considerar estimaciones de los valores de x_1^*, \dots, x_m^* . Estas se pueden obtener resolviendo el problema de regresión lineal

$$\begin{pmatrix} X_1 \\ Z_1 - \hat{\beta}_0 \\ \vdots \\ X_m \\ Z_m - \hat{\beta}_0 \end{pmatrix} = \overbrace{\begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ \hat{\beta}_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & \hat{\beta}_1 & 0 & \dots & 0 & 0 \\ \ddots & \ddots & \ddots & \dots & \ddots & \ddots \\ 0 & 0 & 0 & \dots & 0 & \hat{\beta}_1 \end{pmatrix}}^{2m \times m} \begin{pmatrix} x_1^* \\ \vdots \\ x_m^* \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \varepsilon_1 \\ \vdots \\ \eta_m \\ \varepsilon_m \end{pmatrix}. \quad (4.26)$$

Como se describió en el capítulo anterior, aquí es necesario utilizar cuadrados mínimos generalizados (2.4.1) ya que se conoce la matriz de covarianza de los errores.

Por lo tanto, el bootstrap paramétrico puede ser descrito a partir de los siguientes pasos:

Paso 1: Obtener el estimador de $\hat{\beta} = h(Y)$.

Paso 2: Aplicar cuadrados mínimos generalizados y obtener $\hat{x}_1^*, \dots, \hat{x}_m^*$ resolviendo (4.26)

Paso 3: Generar Y^b de la siguiente manera

$$Y^b = \begin{pmatrix} \hat{x}_1^* \\ \hat{\beta}_1 \hat{x}_1^* + \hat{\beta}_0 \\ \vdots \\ \hat{x}_m^* \\ \hat{\beta}_1 \hat{x}_m^* + \hat{\beta}_0 \end{pmatrix} + \begin{pmatrix} \eta_1 \\ \varepsilon_1 \\ \vdots \\ \eta_m \\ \varepsilon_m \end{pmatrix} \quad (4.27)$$

donde el vector de los errores $\begin{pmatrix} \eta_1 \\ \varepsilon_1 \\ \vdots \\ \eta_m \\ \varepsilon_m \end{pmatrix}$ será generado con la distribución conocida.

Paso 4: A partir de Y^b obtener $\hat{\beta}^b = h(Y^b)$.

Paso 5: Repetir los pasos 3 y 4, B veces.

Finalmente, una vez obtenida una muestra bootstrap de tamaño B de $\hat{\beta}$ calcularemos regiones de confianza como describimos al comienzo del capítulo.

Capítulo 5

Mediciones reales y simulaciones

5.1. Simulación de un modelo lineal

Con el objetivo de constatar las regiones de confianza obtenidas con el método bootstrap, generamos una simulación bajo el modelo lineal

$$\begin{cases} z_t^* = 2x_{t1}^* + x_{t2}^* \\ Z_t = z_t^* + \varepsilon_t \\ X_{ti} = x_{ti}^* + \eta_{ti} \\ 1 \leq t \leq 21 \end{cases}$$

Aquí $\beta = (2, 1)$ y $1 \leq x_{ti}^* \leq 8$. Los errores (η, ε) se generaron bajo una distribución normal con covarianzas relativamente chicas y con autovalores que cumplen $\frac{\lambda_{\max}}{\lambda_{\min}} = 2,5$ y $\lambda_{\max} = 0,0025$. Se eligieron estos valores de λ porque están en el orden de los que se obtienen en metrología de radio frecuencia. En cada replicación se tomó una muestra bootstrap con $B = 1000$. Los resultados de cobertura de las elipses se presentan en la siguiente tabla para los distintos niveles de confianza y para cada estimador. Utilizamos las abreviaciones SC para sesgo corregido, CMG para cuadrados mínimos generalizados y CMF cuadrados mínimos factibles.

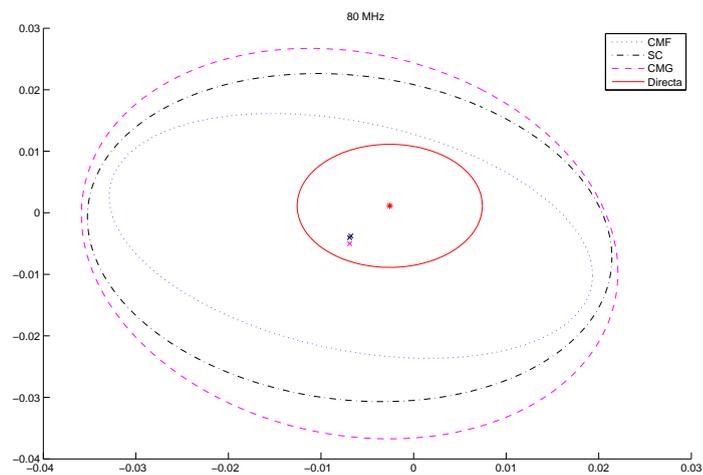
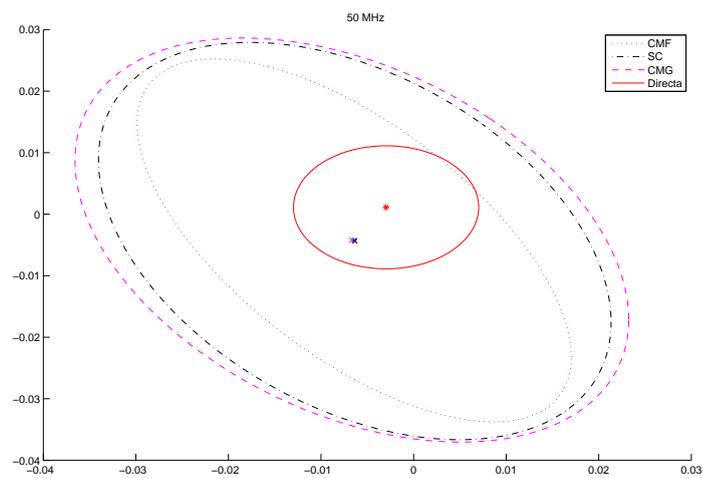
	$1 - \alpha$				
	0.65	0.8	0.9	0.95	0.99
	Proporción				
$\hat{\beta}_{SC}$	0.633	0.802	0.908	0.955	0.987
$\hat{\beta}_{CMG}$	0.642	0.799	0.907	0.955	0.988
$\hat{\beta}_{CMF}$	0.649	0.796	0.897	0.955	0.988

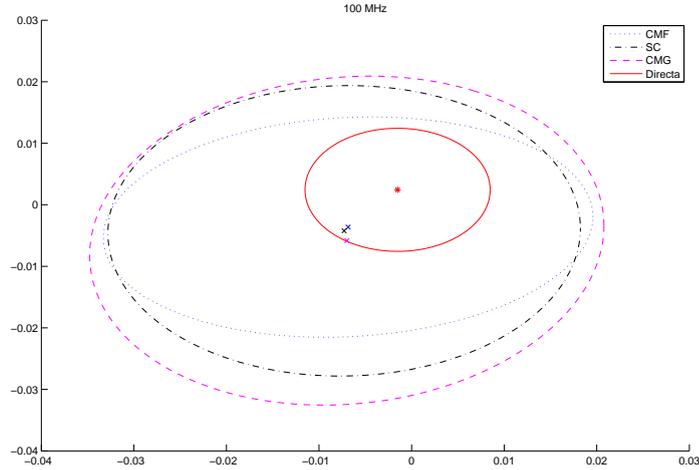
En todos los niveles y para los tres estimadores las proporciones de cobertura se encuentran muy cerca de su valor nominal.

5.2. Resultados de las mediciones

Un generador de radiofrecuencias es capaz de generar señales a distintas frecuencias. Por cada una de éstas, le corresponde un Γ_g distinto. Presentamos los resultados de la medición del coeficiente de reflexión de un generador en tres frecuencias distintas.

Debido a las particularidades del generador, existe un método alternativo que permite medir de forma directa su Γ_g . Esta medición la utilizamos par juzgar los resultados de la técnica desarrollada. Los resultados los presentamos en forma gráfica con las respectivas elipses de 95 % de confianza. La única diferencia entre la implementación y lo desarrollado en los Capítulos anteriores, es que la matriz de varianza de los errores D se obtuvo por simulación de Montecarlo.





En todas los casos, las estimaciones del nuevo método están dentro de la región de confianza obtenida por el método de medición directa. Ésto, en cierta forma, valida la técnica de medición desarrollada ya que se obtienen valores parecidos a los conocidos. Por otro lado, la región de confianza de la medición directa está incluida en las otras y es considerablemente mas chica. Ésto se debe a que depende de seis argumentos cuando la otra técnica depende de cincuenta y tres.

5.3. Validación de los resultados

Desarrollamos un método para constatar cómo se desempeñan las regiones bootstrap en un problema no lineal como el de Γ_g . Para ésto, necesitaríamos generar replicas del modelo, pero no es posible debido a que los valores que lo determinan son desconocidos.

Los siguientes pasos permiten generar N replicas de un modelo similar al que generó los datos. La descripción la hacemos en los términos generales de (3.15) ya que ésta comprobación debe realizarse para cualquier modelo con el que se trabaje.

Paso 1: Fijar un valor $\tilde{\beta}$ cercano a β . Esto puede ser a partir de una misma estimación.

Paso 2: Buscar un valor $\tilde{\mu}^*$ de manera que sea “parecido” al valor observado μ y que $\tilde{\mu}^*$ y $\tilde{\beta}$ sean soluciones del sistema de ecuaciones (3.15). (Notar que no se puede utilizar $\tilde{\mu}^* = \mu$ y $\tilde{\beta} = \hat{\beta}$ simultáneamente ya que no cumplen las ecuaciones.)

Paso 3: Generar N observaciones independientes $\tilde{\mu} = \tilde{\mu}^* + \xi$. Donde ξ es un error aleatorio con la misma distribución del error con que se hicieron las observaciones reales.

Paso 4: Encontrar estimaciones y regiones de confianza del sistema lineal asociado que bajo la notación de (3.9) tiene como observaciones

$$Z_t = \Phi_t(\tilde{\boldsymbol{\mu}}) \quad X_{tk} = \Psi_{tk}(\tilde{\boldsymbol{\mu}}). \quad (5.1)$$

En nuestro caso de interés se fijó $\tilde{\boldsymbol{\beta}}$ del resultado de la medición directa. El valor de S se mantuvo igual al observado y se ajustaron los m_i para que valiera la igualdad (1.8). A diferencia de la simulación de la sección anterior, en cada replicación se resuelve un modelo lineal distinto. Esto se debe a que los valores de $\tilde{\boldsymbol{\mu}}$ varían y por lo tanto también lo hace la matriz de varianza de los errores. Esto difiere de la simulación de la sección anterior donde la matriz D se mantiene fija.

Para nuestro problema generamos $N = 1000$ replicaciones y calculamos los estimadores y sus regiones de confianza bootstrap con $B = 1000$. Las proporciones de cobertura obtenidas fueron

	$1 - \alpha$				
	0.65	0.8	0.9	0.95	0.99
	Proporción				
$\hat{\boldsymbol{\beta}}_{SC}$	0.63	0.799	0.903	0.952	0.99
$\hat{\boldsymbol{\beta}}_{CMG}$	0.64	0.8	0.885	0.935	0.99
$\hat{\boldsymbol{\beta}}_{CMF}$	0.255	0.376	0.537	0.678	0.873

Los estimadores de cuadrados mínimos generalizados y de sesgo corregido reportaron valores de cobertura muy cercanos a los nominales. En cambio, el estimador de cuadrados mínimos factibles dio proporciones muy por debajo de los niveles fijados. Esto sugiere que el estimador es sensible a pequeñas variaciones del modelo lineal ya que en la sección anterior vimos que las regiones de confianza obtenidas eran adecuadas.

Con el objetivo de testear esta hipótesis realizamos una simulación siguiendo los pasos descritos a continuación

Paso 1: Obtener $\tilde{\boldsymbol{\beta}}$ y $\tilde{\boldsymbol{\mu}}^*$ como en los pasos 1 y 2 anteriores.

Paso 2: Generar $N = 1000$ valores independientes $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^* + \boldsymbol{\xi}$. Por cada uno calcular la matriz de varianza D y generar una replicación del modelo lineal asociado. Esto es, en la notación que se utiliza en (3.9) y a diferencia de (5.1), considerar como observaciones

$$Z_t = \Phi_t(\tilde{\boldsymbol{\mu}}^*) + \varepsilon_t \quad X_{tk} = \Psi_{tk}(\tilde{\boldsymbol{\mu}}^*) + \eta_{tk} \quad (5.2)$$

donde los η_{tk} y ε_t tienen distribución normal y sus varianzas y covarianzas se tomaron de D .

Paso 3: Obtener las regiones de confianza y estimaciones de cada replicación

Notar que en esta simulación las observaciones fueron generadas bajo un modelo lineal a diferencia de la simulación anterior donde seguían el modelo completo.

Para quitar la aleatoriedad de la comparación se volvieron a calcular regiones de confianza para el modelo completo. La siguiente tabla muestra las proporciones de cobertura para ambas simulaciones obtenidas con $B = 1000$.

Modelo	$1 - \alpha$				
	0.65	0.8	0.9	0.95	0.99
Completo	0,269	0,396	0,565	0,677	0,868
Lineal	0,697	0,834	0,926	0,966	0,993

Como puede apreciarse los resultados obtenidos bajo modelos lineales son cercanos a los valores nominales. Dando credibilidad a la hipótesis formulada.

Comentarios y trabajo a futuro

El trabajo se desarrolló de forma progresiva. En primer lugar hubo que comprender el problema que se quería resolver y lograr darle una formulación matemática adecuada. En la búsqueda de un forma de resolverlo fue que llegamos a investigar los modelos lineales con errores en las variables. Aquí nos encontramos con que no había estimadores en la literatura para los casos en que los errores de las distintas observaciones estuvieran correlacionados. Los estimadores propuestos fueron fruto de generalizar los ya existentes y de utilizar ideas de la regresión lineal clásica. Consideramos que una parte muy importante del trabajo, aunque tal vez no tan visible, fue la implementación numérica de los estimadores, en particular el de cuadrados mínimos generalizados. Las técnicas bootstrap. Estas, generalmente explicadas para estimar parámetros escalares, requirieron ser adaptadas al caso multivariado e implementadas para nuestro modelo. Por último logramos encontrar un marco general en los modelos con errores en las variables donde la ausencia de independencia entre las observaciones resulta natural y que permite apreciar los alcances del trabajo.

Como trabajo a futuro, consideramos importante estudiar el comportamiento de los estimadores en función de la cantidad de ecuaciones y la variabilidad en el error de la observación. Estudiar propiedades asintóticas para los modelos lineales y también queda por probar como se comportan los estimadores ante la presencia de datos atípicos. Algunas ideas para la estimación robusta pueden surgir de minimizar una función $\tau(M^t D^{-1} M)$ que sea poco sensible a outliers o también, utilizando la notación de (3.93), minimizar $\|SM\|_1^2$ en vez de $\|SM\|_2^2$. El estimador de cuadrados mínimos factibles también podría adaptarse a distintas variantes robustas a partir de realizar el proceso iterativo descrito en (3.40) utilizando distintos estimadores robustos en vez del de cuadrados mínimos.

Bibliografía

- [1] Fuller, W. Measurement Error Models: Wiley Series in Probability and Statistics, (2006)
- [2] Raymond, J. et al. Measurement Error in Nonlinear Models. Chapman & Hall (2006)
- [3] Johnson, R. and Wichern, D. Applied Multivariate Statistical Analysis, Prentice-Hall 6ta edición, (2007).
- [4] Heckman, J. Leamer, E. Handbook of Econometrics, volume 5, chapter 59, Elseiver, (2001).
- [5] BIPM, IEC, IFCC, ISO, IUPAC, UPAP y OIML. Guide to the expression of uncertainty in measurement, International organization for standardization, Geneva (1993)
- [6] BIPM, IEC, IFCC, ISO, IUPAC, UPAP y OIML. Evaluation of Measurment Data - supplement 2 to the 'Guide to the Expression of Uncertainty in Measurement'. Extension to any number of output quantities. (2011)
- [7] Chernick, M. Bootstrap Methods: A Practitioner's Guide. Wiley Series in Probability and Statistics, 16. (1999)
- [8] Davison, A. C. and Hinkley, D. V. Bootstrap Methods and their Applications, Cambridge University Press, Cambridge, 1997.
- [9] Efron, B. Bootstrap methods: Another look at the jackknife, The Annals of Statistics 7 (1979), 1–26.
- [10] Efron, B. and Tibshirani, R. J. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Statistical Science 1 (1986), 54–77.
- [11] Hastie, T. Tibshirani, R. Friedman, J. The Elements of Statistical Learning. Springer Series in Statistics, (2009)
- [12] Chan, N. Mak, T. Heteroscedastic Errors in a Linear Functional Relationship. Biometrika Vol 71 No 1 (1984) 212-215

- [13] Gillard, J. A historical Overview of Linear Regression with Errors in Both Variables. Cardiff University, (2006)
- [14] Matei, B. Heteroscedastic Errors-in-Variables Models in Computer Vision. (2001)
- [15] Amemiya, Y. Fuller A. Estimation for the Nonlinear Functional Relationship *Annals of Statistics* Volume 16, Number 1 (1988), 147-160.