ERRORES DE MEDICION Y EL USO DE VARIABLES INSTRUMENTALES

I Casos de Errores de Medición

En esta sección analizaremos la existencia de errores de medición en la variable dependiente y en el regresor, bajo el contexto del modelo lineal simple. Veremos que el error de medición en la variable dependiente sólo causa una pérdida de eficiencia, manteniéndose la propiedad de insesgamiento de los estimadores. En contraste, el error de medición en el regresor tiene consecuencias mucho más serias, puesto que conduce a estimadores inconsistentes.

1.1 Error de Medición en la Variable Dependiente

Supongamos que el modelo verdadero es:

$$Y_t^* = \beta X_t^* + u_t \qquad t=1, 2,..., T \qquad (1)$$

donde $E(u_t)=0$, $E(u_t^2)=\sigma_u^2$.

(Por simplicidad, se asume un modelo sin constante).

✓ Supongamos que sólo contamos con una medida imperfecta de Y_t^*:

$$Y_t = Y_t^* + v_t \qquad (2)$$

donde $E(v_t)=0$, $E(v_t^2)=\sigma_v^2$, $E(u_t v_t)=0$.

De (1) y (2):

$$Y_t = \beta X_t^* + u_t + v_t \qquad (3)$$

$$= \beta X_t^* + \xi_t$$

donde $\xi_t = u_t + v_t$. Por lo tanto, $E(\xi_t)=0$, $E(\xi_t^2) = \sigma_u^2 + \sigma_v^2$.

El estimador MICO de β proveniente del modelo (3), $\tilde{\beta}=\left(\sum_{t=1}^{T} X_t^* Y_t\right)\left(\sum_{t=1}^{T} X_t^{*2}\right)^{-1}$ es insesgado pero **menos eficiente** que aquel proveniente del modelo (1), $\hat{\beta}=\left(\sum_{t=1}^{T} X_t^* Y_t^*\right)\left(\sum_{t=1}^{T} X_t^{*2}\right)^{-1}$:

$$\mathrm{Var}(\tilde{\beta})=\frac{\sigma_u^2+\sigma_v^2}{\sum_{i=1}^{n} X_i^2} > \frac{\sigma_u^2}{\sum_{i=1}^{n} X_i^2}=\mathrm{Var}(\hat{\beta}) \qquad (4)$$

1.2 Error de Medición en el Regresor

Supongamos que el modelo verdadero es:

$$Y_t^* = \beta X_t^* + u_t \qquad (5)$$

donde $E(u_t)=0$, $E(u_t^2)=\sigma_u^2$.

Desafortunadamente, sólo contamos con una medida imperfecta de X_t^*:

$$X_t = X_t^* + \varepsilon_t \qquad (6)$$

donde $E(\varepsilon_t)=0$, $E(\varepsilon_t^2)=\sigma_\varepsilon^2$, $E(u_t\varepsilon_t) = E(X_t^* u_t) = E(X_t^*\varepsilon_t) = 0$

Entonces:

$$\begin{aligned} Y_t^* &= \beta(X_t - \varepsilon_t) + u_t \\ &= \beta X_t + u_t - \beta\varepsilon_t \\ &= \beta X_t + \omega_t \qquad \text{con } \omega_t \equiv u_t - \beta\varepsilon_t \end{aligned}$$

✓ $\hat{\beta}$ será sesgado porque $E(X_t\omega_t)\neq 0$. Ello, porque se viola uno de los supuestos claves del modelo lineal:

$$E(X_t\,\omega_t)=E(X_t^*+\varepsilon_t)(u_t-\beta\varepsilon_t)$$

$$= E(X_t^*u_t) - \beta E(X_t^*\varepsilon_t) + E(\varepsilon_t u_t) - \beta E(\varepsilon_t^2)$$

$$= -\beta\,\sigma_\varepsilon^2 \tag{7}$$

Se puede demostrar que aún cuando el tamaño de la muestra sea suficientemente grande ($T\to\infty$), $\tilde{\beta}$ será sesgado. El **sesgo** estará dado por $-\dfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_{x*}^2}$. Veamos por qué:

$$\tilde{\beta} = \frac{\sum_{t=1}^{T} Y_t^* X_t}{\sum_{t=1}^{T} X_t^2} = \frac{\sum_{t=1}^{T} (\beta X_t^* + u_t)(X_t^* + \varepsilon_t)}{\sum_{t=1}^{T} (X_t^* + \varepsilon_t)^2}$$

$$= \frac{\beta\left(\frac{1}{T}\sum_{t=1}^{T} X_t^{2^*} + \frac{1}{T}\sum_{t=1}^{T} X_t^*\varepsilon_t\right) + \frac{1}{T}\sum_{t=1}^{T} u_t X_t^* + \frac{1}{T}\sum_{t=1}^{T} u_t\varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} X_t^{2^*} + 2\frac{1}{T}\sum_{t=1}^{T} X_t^*\varepsilon_t + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2}$$

Por propiedades de probabilidad límite, se tiene que:

$$\tilde{\beta} \xrightarrow{p} \frac{\beta(E(X_t^{*2}) + E(X_t^*\varepsilon_t)) + E(X_t^* u_t) + E(u_t\varepsilon_t)}{E(X_t^{*2}) + E(X_t^*\varepsilon_t) + E(\varepsilon_t^2)} = \frac{\beta\sigma_{x*}^2}{\sigma_{x*}^2 + \sigma_\varepsilon^2}$$

$$= \beta - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_{x*}^2} \tag{8}$$

donde $E(X_t^{*2})=\sigma_{x*}^2$ (por simplicidad se asume que $E(X_t^*)=0$) y, además, dados los supuestos, $E(X_t^*\varepsilon_t)=E(X_t^* u_t)=E(u_t\varepsilon_t) = 0$.

De lo anterior, se deduce que $\tilde{\beta}$ estará sesgado hacia cero. Entre mayor sea la variabilidad del error de medida, σ_{ε}^{2}, mayor será el sesgo.

3.1 Una Aplicación: Función de Consumo

Se tiene la siguiente observación sobre gasto efectivo en consumo (C^*), ingreso efectivo (Y^*), gasto medido en consumo (C) e ingreso medido (Y), para una muestra de 10 familias. Todas las cifras están medidas en dólares, y corresponden a promedios semanales:

Observación	C*	I*	C	I
1	75.47	80	67.60	80.09
2	74.98	100	75.44	91.57
3	102.82	120	109.70	112.14
4	125.77	140	129.42	145.60
5	106.50	160	104.24	168.56
6	131.43	180	125.83	171.47
7	149.37	200	153.99	203.54
8	143.86	220	152.92	222.85
9	177.52	240	176.33	232.98
10	182.28	260	174.52	261.18

Se plantea el siguiente modelo de consumo para las cifras efectivas:

$$C_i^* = \beta_1 + \beta_2 I_i^* + u_i$$

donde $E(u_i)=0$, $E(u_i^2) = \sigma_u^2$.

En base a la información proporcionada, la función de consumo verdadera (estimada) vendría dada por:

Tabla 1 Función Consumo Verdadera Estimada

Variable dependiente: C^*
Número de observaciones: 10

Variable	Coeficiente	Error Estándar	Estadígrafo t	Valor p
Constante	25.000	10.477	2.386	0.044
I^*	0.600	0.058	10.276	0.000

R^2	0.929	Media variable dependiente	127.000
R^2 ajustado	0.921	Desv. est. variable dependiente	37.683
error estándar regresión	10.606	Estadígrafo F	105.599
Estadígrafo Durbin-Watson	2.816	Valor p estadígrafo F	0.000

Se sabe, además, que $C_i = C_i^* + \varepsilon_i$, $I_i = I_i^* + \omega_i$, con ε_i y ω_i, errores de medición que satisfacen lo siguiente:

$$E(\varepsilon_i) = 0,\ E(\varepsilon_i^2) = \sigma_\varepsilon^2;\ E(\omega_i) = 0,\ E(\omega_i^2) = \sigma_\omega^2;\ E(\varepsilon_i\omega_i) = E(u_i\omega_i) = 0$$

$$E(I_i^*\varepsilon_i) = E(I_i^*\omega_i) = E(I_i^*u_i) = 0$$

a) Supongamos que usted observa el ingreso efectivo, I^*, pero sólo cuenta con el consumo medido, C. Con dicha información usted estima la siguiente función consumo:

Tabla 2 Función Consumo cuando sólo I^* es Observable

Variable dependiente: C
Número de observaciones: 10

Variable	Coeficiente	Error Estándar	Estadígrafo t	Valor p
Constante	24.999	12.218	2.046	0.075
I^*	0.600	0.068	8.811	0.000

R^2	0.906594	Media variable dependiente	127.000
R^2 ajustado	0.894918	Desv. est. variable dependiente	38.158
error estándar regresión	12.36931	Estadígrafo F	77.647
Estadígrafo Durbin-Watson	2.286928	Valor p estadígrafo F	0.000

- Este es el caso en que los estimadores continúan siendo insesgados y consistentes. (Las magnitudes de los coeficientes son prácticamente idénticas a las de la Tabla 1). El error de medición se refleja en mayores errores estándar.

b) Supongamos que usted observa el consumo efectivo, C^*, pero sólo cuenta con el ingreso medido, I. Con dicha información usted estima la siguiente función consumo:

Tabla 3 Función Consumo cuando sólo C^* es Observable

Variable Dependiente: C*
Número de observaciones: 10

Variable	Coeficiente	Error Estándar	Estadígrafo t	Valor p
Constante	28.457	11.282	2.522	0.036
I	0.583	0.063	9.246	0.000
R^2	0.914	Media variable dependiente		127.000
R^2 ajustado	0.904	Desv. est. variable dependiente		37.683
error estándar regresión	11.692	Estadígrafo F		85.481
Estadígrafo Durbin-Watson	2.8421	Valor p estadígrafo F		0.000

- En este caso, los estimadores son sesgados. Como vemos, las magnitudes de los coeficientes de la Tabla 2 se alejan de aquellas reportadas en la Tabla 1, especialmente en lo que concierne al término de la constante.

III Variables Instrumentales

3.1 Conceptos Básicos

Supóngase el modelo Y=Xβ + u, donde $\text{plim}\left(\frac{\mathbf{X'u}}{T}\right) \neq \mathbf{0}$. En este escenario el estimador MICO es **inconsistente**. Se desea encontrar una matriz Z_{nxk} de "**instrumentos**", que satisfaga las siguientes condiciones:

$$\text{plim}\left(\frac{\mathbf{Z'X}}{T}\right) = \mathbf{Q}_{zx}, \ \text{plim}\left(\frac{\mathbf{Z'Z}}{T}\right) = \mathbf{Q}_{zz}, \ \text{plim}\left(\frac{\mathbf{Z'u}}{T}\right) = \mathbf{0} \qquad (9)$$

donde $\mathbf{Q}_{zx}$ es una matriz de rango k y $\mathbf{Q}_{zz}$ es una matriz positiva definida.

La condición (9) indica que **Z** está correlacionada con **X** pero no con el error. El **estimador de variables instrumentales (VI)** se define como:

$$\boxed{\hat{\beta}_{VI} = (\mathbf{Z'X})^{-1}\mathbf{Z'Y}} \qquad (10)$$

De la expresión (10) se tiene:

$$\hat{\beta}_{VI} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Y} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\mathbf{X}\beta + \mathbf{u}) = \beta + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u}$$

De ello:

$$\text{plim}(\hat{\beta}_{VI}) = \beta + \text{plim}\left(\frac{\mathbf{Z}'\mathbf{X}}{T}\right)^{-1} \text{plim}\left(\frac{\mathbf{Z}'\mathbf{u}}{T}\right) = \beta$$

La varianza (asintótica) de $\hat{\beta}_{VI}$ viene dada por:

$$\text{AVar}(\hat{\beta}_{VI}) = \text{E}(\hat{\beta}_{VI} - \text{plim}(\hat{\beta}_{VI}))(\hat{\beta}_{VI} - \text{plim}(\hat{\beta}_{VI}))'$$

$$= \text{E}(\hat{\beta}_{VI} - \beta)(\hat{\beta}_{VI} - \beta)' = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\text{E}(\mathbf{uu}')\mathbf{Z}(\mathbf{Z}'\mathbf{X})^{-1}$$

Esto es, $\boxed{\text{AVar}(\hat{\beta}_{VI}) = \sigma_u^2(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{X})^{-1}}$ (11)

2.2 Una Aplicación a la Existencia de Error de Medición

Supongamos que X_t^* sigue la siguiente dinámica:

$X_t^* = \rho X_{t-1}^* + \eta_t$ (12)

donde η_t es ruido blanco.

Sin embargo, X_t^* no es observable. Lo que se observa es una medida imperfecta:

$X_t = X_t^* + \varepsilon_t$ con $E(\eta_t\varepsilon_t)=0$ (13)

De (12) y (13), se tiene:

$X_t - \varepsilon_t = \rho(X_{t-1} - \varepsilon_t) + \eta_t$

$\Leftrightarrow$ $Xt = \rho X_{t-1} + \eta_t + \varepsilon_t(1-\rho)$

El modelo en términos de Y_t^* y X_t es:

$Y_t^* = \beta X_t + \omega_t$ $\omega_t = u_t - \beta\varepsilon_t$

Propongamos como instrumento X_{t-1}:

$E(X_t X_{t-1}) = \rho E(X_{t-1}^2) \neq 0$

$E(\omega_t X_{t-1}) = (u_t - \beta\varepsilon_t)(X_{t-1}^* + \varepsilon_{t-1}) = 0$

asumiendo que $E(\varepsilon_t \varepsilon_{t-1}) = 0$.

Entonces es X_{t-1} una buena elección. De lo anterior,

$$\mathbf{Z} = \begin{pmatrix} X_1 \\ X_2 \\ \ldots \\ X_{T-1} \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} X_2 \\ X_3 \\ \ldots \\ X_T \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_T \end{pmatrix}$$

con t= 2, ..., T.

$$\hat{\beta}_{VI} = \frac{\sum_{t=2}^{T} X_{t-1} Y_t^*}{\sum_{t=2}^{T} X_{t-1} X_t} \qquad \mathrm{Var}(\hat{\beta}_{VI}) = \frac{\sigma_u^2 \sum_{t=2}^{T} X_{t-1}^2}{\left(\sum_{t=2}^{T} X_{t-1} X_t\right)^2}$$

Se tiene que $\hat{\beta}_{VI}$ es consistente:

$$\hat{\beta}_{VI} = \frac{\sum_{t=2}^{T} (X_{t-1}^* + \varepsilon_{t-1})(\beta X_t^* + u_t)}{\sum_{t=2}^{T} (X_{t-1}^* + \varepsilon_{t-1})(X_t^* + \varepsilon_t)}$$

$$= \frac{\left(\sum_{t=2}^{T} (X_{t-1}^* + \varepsilon_{t-1})(\beta\rho X_{t-1}^* + \beta\eta_t + u_t)\right) \Big/ T}{\left(\sum_{t=2}^{T} (X_{t-1}^* + \varepsilon_{t-1})(\rho X_{t-1}^* + \eta_t + \varepsilon_t)\right) \Big/ T}$$

$$\xrightarrow{p} \frac{\beta\rho E(X_{t-1}^{*\,2})}{\rho E(X_{t-1}^{*\,2})} = \beta$$

3 Test de Hipótesis para Detectar Errores de Medición

Hausman propuso el siguiente estadígrafo para detectar errores de medición:

$$\boxed{W = (\hat{\beta}_{Mico} - \hat{\beta}_{VI})'(\hat{Var}(\hat{\beta}_{VI}) - \hat{Var}(\hat{\beta}_{Mico}))^{-1}(\hat{\beta}_{Mico} - \hat{\beta}_{VI}) \xrightarrow{d} \chi^2(k)}$$

con $\hat{Var}(\hat{\beta}_{VI}) = \hat{\sigma}_u^2 (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{X})^{-1}$, $\hat{Var}(\hat{\beta}_{Mico}) = \hat{\sigma}_u^2 (\mathbf{X}'\mathbf{X})^{-1}$

Bajo la hipótesis nula, NO existe error de medición. Por lo tanto, tanto el estimador de MICO como el de VI son consistentes, pero MICO es más eficiente. Bajo la hipótesis alternativa, sólo el estimador de VI es consistente. Si W supera al valor crítico de la chi-cuadrado con k grados de libertad (la dimensión del vector β) para un (1–ε)% de confianza, se rechaza H_0 en favor de H_1.

Nota: A fin de asegurar que $\hat{Var}(\hat{\beta}_{VI}) - \hat{Var}(\hat{\beta}_{Mico})$ sea una matriz positiva definida, se utiliza el mismo estimador de σ_u^2, $\hat{\sigma}_u^2 = \frac{(\mathbf{Y} - \hat{\beta}_{VI})'(\mathbf{Y} - \hat{\beta}_{VI})}{T - K}$.