CAPíTULO 12

Regresores estocásticos

El supuesto de regresores fijos en muestras repetidas juega un papel fundamental en el establecimiento de las propiedades estadísticas del estimador de mínimos cuadrados, así como en la derivación de su distribución muestral. En este tema relajamos este supuesto permitiendo que los regresores sean estocásticos. Vamos a distinguir tres tipos de regresores dependiendo de la forma de correlación que guarden con la perturbación estocástica. El caso más problemático se presenta cuando existe correlación contemporánea entre los regresores y la perturbación, que justifica la búsqueda de un estimador alternativo basado en variables instrumentales.

12.1. Introducción

La interpretación del modelo lineal general $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ como un proceso generador de datos nos permite entender el significado del supuesto de regresores fijos en muestras repetidas. Podemos generar distintas muestras de la variable aleatoria n-dimensional $\mathbf{y}$ sumando a la parte sistemática $\mathbf{X}\boldsymbol{\beta}$ diferentes muestras de la perturbación aletoria n-dimensional $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$. Bajo este supuesto, derivamos fácilmente la distribución muestral del estimador de mínimos cuadrados $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Expresando $\hat{\boldsymbol{\beta}}$ en términos de $\mathbf{u}$

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

vemos que $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma_u^2\mathbf{I}_n)$. La normalidad del estimador se justifica porque $\hat{\boldsymbol{\beta}}$ es una transformación lineal de una variable aleatoria con distribución normal multivariante, y sus dos primeros momentos son

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underbrace{E[\mathbf{u}]}_{\mathbf{0}} \\ \underbrace{E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})']}_{V(\boldsymbol{\beta})} &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underbrace{E(\mathbf{u}\mathbf{u}')}_{\sigma_u^2\mathbf{I}_n}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Si la matriz $\mathbf{X}$ es estocástica, entonces la aleatoriedad de $\hat{\boldsymbol{\beta}}$ proviene no sólo de $\mathbf{u}$ sino también de $\mathbf{X}$. Ahora el estimador es una función no lineal de las variables aleatorias $\mathbf{X}$ y $\mathbf{u}$ y, por tanto, su distribución muestral es desconocida incluso si suponemos una determinada distribución para $\mathbf{X}$. Por otro lado, no podemos sacar la matriz $\mathbf{X}$ del operador esperanda, $E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] \neq (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{u}]$ y $E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \neq (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$, y la derivación de los dos primeros momentos del estimador no es tan inmediata.

Cuando no conocemos las propiedades estadísticas de un estimador en muestras pequeñas, recurrimos a la teoría asintótica para ver como se comporta el estimador en muestras grandes. En nuestro caso, la distribución asintótica de $\hat{\boldsymbol{\beta}}$ va a depender de los supuestos que hagamos sobre sus dos componentes estocásticos $\mathbf{X}'\mathbf{X}$ y $\mathbf{X}'\mathbf{u}$. Para el primero, vamos a suponer que (1) los momentos poblacionales de segundo orden de las variables explicativas existen, (2) los momentos muestrales convergen a los poblacionales

$$\lim_{n\to\infty} \text{prob}\left(\frac{\mathbf{X}'\mathbf{X}}{n} = \mathbf{\Sigma}_{xx}\right) = 1 \quad \Longleftrightarrow \quad \text{plim}\,\frac{\mathbf{X}'\mathbf{X}}{n} = \mathbf{\Sigma}_{xx}$$

y (3) la matriz de segundos momentos poblacionales $\mathbf{\Sigma}_{xx}$ tiene rango pleno, $\rho(\mathbf{\Sigma}_{xx}) = k$.

Observación 69. *El límite de la probabilidad, el operador* plim, *tiene la propiedad fundamental de ser intercambiable con cualquier función f contínua. Así, el límite probabilístico de un producto (cociente) de secuencias aleatorias es el producto (cociente) de los límites probabilísticos de cada secuencia. Además, se cumple que*

$$\text{plim}\left(\sum_{t=1}^{n} \frac{\mathbf{x}_t\mathbf{x}_t'}{n}\right)^{-1} = \left(\text{plim}\sum_{t=1}^{n} \frac{\mathbf{x}_t\mathbf{x}_t'}{n}\right)^{-1} = \mathbf{\Sigma}_{xx}^{-1}$$

o, equivalentemente, en forma matricial

$$\boxed{\text{plim}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} = \left(\text{plim}\,\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} = \mathbf{\Sigma}_{xx}^{-1}}$$

Para el segundo componente, $\mathbf{X}'\mathbf{u}$, que estudiaremos en la siguiente sección, nesecitamos caracterizar su convergencia en probabilidad y distribución. Aquí, serán de interés las siguientes proposiciones.

Proposición 101. *(Teorema central del límite)Si $\{z_t\}_{t=1}^n$ es una secuencia de vectores idéntica e independientemente distribuidos con vector medias $\boldsymbol{\mu}$ y matriz de varianzas y covarianzas $\mathbf{\Sigma}_t$, entonces*

$$n^{-1/2}\sum_{t=1}^{n}(\mathbf{z}_t - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma})$$

en donde

$$\mathbf{\Sigma} = \lim \frac{\sum_{t=1}^{n} \mathbf{\Sigma}_t}{n}$$

Proposición 102. *(Ley de las esperanzas iteradas) Sea $f(X,Y)$ una función de las variables aleatorias X e Y. Entonces, las esperanza incondicional*

$$E[f(X,Y)] = E_X\{E_{Y|X}[f(X,Y)]\}$$

en donde E_X es la esperanza en la función de densidad marginal de X y $E_{X|Y}$ es la esperanza en la función de densidad condional de Y dado X.

Proposición 103. *(Ley de la varianza total) La varianza incondicional*

$$V[f(X,Y)] = V_X\{E_{Y|X}[f(X,Y)]\} + E_X\{V_{Y|X}[f(X,Y)]\}$$

en donde V_X es la varianza en la función de densidad marginal de X y $V_{X|Y}$ es la varianza en la función de densidad condional de Y dado X.

12.2. Exogeneidad y endogeneidad

Vamos a considerar tres tipos de regresores estocásticos:

1. Regresores fuertemente exógenos

$$E(u_t|\mathbf{x}_{t-k}) = 0\ \forall k \quad \Longleftrightarrow \quad E(\mathbf{u}|\mathbf{X}) = 0$$

2. Regresores débilmente exógenos (también llamados predeterminados)

$$E(u_t|\mathbf{x}_t) = 0 \quad \not\Rightarrow \quad E(\mathbf{u}|\mathbf{X}) = 0$$

3. Regresores endógenos (contemporáneamente correlacionados con la perturbación estocástica)

$$E(u_t, \mathbf{x}_t) \neq 0$$

Definición 88. *El modelo lineal general presenta endogeneidad si existe correlación contemporánea entre las variables explicativas y el término de error.*

Ejemplo 23. *Sea el modelo $AR(1)$*

$$y_t = \phi_1 y_{t-1} + u_t, \quad u_t iid N(0, \sigma_u^2)$$

cuya forma $MA(\infty)$ es

$$y_t = u_t - \phi_1 u_{t-1} - \phi_1^2 u_{t-2} - \dots$$

Es claro que y_t es una variable aleatoria para todo t. Por tanto, la variable explicativa $x_t = y_{t-1}$ es un regresor estocástico. En concreto, es un regresor débilmente exógeno o predeterminado

$$E(u_t|y_{t-1}) = 0$$

Vemos que no se cumple la condición de exogeidad estricta porque

$$E(u_t|y_{t-k}) = u_t \neq 0 \quad \forall k > 0$$

◁

Proposición 104. *Bajo exogeneidad (débil o fuerte), la secuencia $\mathbf{X}'\mathbf{u}/n$ converge en probabilidad a un vector nulo*

$$\operatorname{plim} \frac{\mathbf{X}'\mathbf{u}}{n} = \mathbf{0}$$

Demostración. Por la ley de las esperanzas iteradas

$$E(\mathbf{x}_t u_t) = E_{\mathbf{x}_t}[E(\mathbf{x}_t u_t|\mathbf{x}_t)] = E_{\mathbf{x}_t}[\mathbf{x}_t \underbrace{E(u_t|\mathbf{x}_t)}_{\mathbf{0}}] = \mathbf{0}$$

en donde $E(u_t|\mathbf{x}_t) = 0$ por ser $\mathbf{x}_t$ exógeno. De aquí,

$$\operatorname{plim} \sum_{t=1}^{n} \frac{\mathbf{x}_t u_t}{n} = \operatorname{plim} \frac{\mathbf{X}'\mathbf{u}}{n} = \mathbf{0}$$

□

Proposición 105. *Bajo exogeneidad (débil o fuerte), la secuencia $\mathbf{X}'\mathbf{u}/\sqrt{n}$ converge en distribución a un normal con vector de medias nulo y matriz de varianzas y covarianzas $\sigma_u^2 \mathbf{\Sigma}_{xx}$*

$$\frac{\mathbf{X}'\mathbf{u}}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \sigma_u^2 \mathbf{\Sigma}_{xx})$$

Demostración. Por la ley de la varianza incondicional

$$V(\mathbf{x}_t u_t) = E_{\mathbf{x}_t}[V(\mathbf{x}_t u_t|\mathbf{x}_t)] + V_{\mathbf{x}_t}[E(u_t|\mathbf{x}_t)] = \sigma_u^2 \mathbf{x}_t \mathbf{x}_t'$$

Así, $\{\mathbf{x}_t u_t\}_{t=1}^n$ es una secuencia de variables aleatorias con vector de medias nulo y matriz de varianzas y covarianzas $\sigma_u^2 \mathbf{x}_t \mathbf{x}_t'$. Por el teorema central del límite dado en la proposición 101

$$\operatorname{plim} \sum_{t=1}^{n} \frac{\mathbf{x}_t u_t}{\sqrt{n}} = \operatorname{plim} \frac{\mathbf{X}'\mathbf{u}}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \sigma_u^2 \boldsymbol{\Sigma}_{xx})$$

□

PROPOSICIÓN 106. *Cuando los regresores estocásticos son fuertemente exógenos, entonces el estimador de mínimos cuadrados*

1. *es una función no lineal de* $\mathbf{X}$ *y* $\mathbf{u}$,
2. *no sigue una distribución normal en muestras finitas,*
3. *es insegado,*
4. *tiene matriz de varianzas y covarianzas* $V(\hat{\boldsymbol{\beta}}) = \sigma_u^2 E[(\mathbf{X}'\mathbf{X})^{-1}]$
5. *es consistente,*
6. *asintóticamente sigue una distribución normal*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N(\mathbf{0}, \sigma_u^2 \boldsymbol{\Sigma}_{xx}^{-1})$$

DEMOSTRACIÓN.

1. Demostrado en la sección anterior.
2. Demostrado en la sección anterior.
3. Aplicando la ley de las esperanzas iteradas

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = \boldsymbol{\beta} + E_{\mathbf{X}}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \underbrace{E(\mathbf{u}|\mathbf{X})}_{\mathbf{0}}] = \boldsymbol{\beta}$$

También conviene notar que la esperanza condicional $E(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \beta$ no depende de $\mathbf{X}$, por lo que la esperanza incondicional $E(\hat{\boldsymbol{\beta}}) = \beta$.

4. Aplicando la ley de la varianza total

$$V(\hat{\boldsymbol{\beta}}) = E_{\mathbf{X}}(V(\hat{\boldsymbol{\beta}}|\mathbf{X})) + V_{\mathbf{X}}(E(\hat{\boldsymbol{\beta}}|\mathbf{X})) = \sigma_u^2 E[(\mathbf{X}'\mathbf{X})^{-1}]$$

También podemos aplicar la ley de las esperanzas iteradas

$$V(\hat{\boldsymbol{\beta}}) = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = E_{\mathbf{X}}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \underbrace{E(\mathbf{u}\mathbf{u}'|\mathbf{X})}_{\sigma_u^2 \mathbf{I}_n} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]$$

5. Es claro que

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} \frac{\mathbf{X}'\mathbf{u}}{n}$$

De aquí, aplicando límites probabilísticos

$$\operatorname{plim} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \operatorname{plim} \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} \times \operatorname{plim} \frac{\mathbf{X}'\mathbf{u}}{n} = \boldsymbol{\beta} + \boldsymbol{\Sigma}_{xx} \times 0 = \boldsymbol{\beta}$$

6. Escribiendo

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} \frac{\mathbf{X}'\mathbf{u}}{\sqrt{n}}$$

tenemos, asintóticamente, el producto de la matriz $\boldsymbol{\Sigma}_{xx}^{1}$ por el vector de variables normales $N(\mathbf{0}, \sigma_u^2 \boldsymbol{\Sigma}_{xx})$.

También conviene notar que la función de densidad conjunta puede factorizarse como

$$f(\mathbf{y}, \mathbf{X}) = f(\mathbf{y}|\mathbf{X}) f(\mathbf{X})$$

Si la distrubución de $\mathbf{X}$ no depende de los parámetros del modelo, entonces los estimadores de máxima verosimilitud se obtienen maximizando $f(\mathbf{y}|\mathbf{X})$, que es la función de densidad estudiada para el caso de regresores fijos. Bajo ciertas condiciones de regularidad, los estimadores de máxima verosimilitud siguen asintóticamente una distribución normal.

□

PROPOSICIÓN 107. *Cuando los regresores estocásticos son débilmente estocásticos, el estimador de mínimos cuadrados es sesgado, pero mantiene sus propiedades asintóticas.*

DEMOSTRACIÓN. El estimador es sesgado porque $E(\mathbf{u}|\mathbf{X}) \neq 0$. Las propiedades asintóticas solo requieren exogeneidad débil. □

PROPOSICIÓN 108. *Cuando el modelo lineal general presenta endogeneidad, el estimador de mínimos cuadrados ordinarios es inconsistente.*

DEMOSTRACIÓN. Por la definición de endogeneidad

$$\operatorname{plim} \frac{\mathbf{X}'\mathbf{u}}{n} \neq 0$$

□

12.3. Variables instrumentales

Cuando $\operatorname{plim} \mathbf{X}'\mathbf{X}/n \neq \mathbf{0}$, podríamos intentar evitar la inconsistencia del estimador de mínimos cuadrados ordinarios "sustituyendo"los regresores endógenos por otras variables correlacionadas con éstos, pero que no estén correlacionadas con la perturbación aleatoria. Tales variables se denomian instrumentos o variables instrumentales.

DEFINICIÓN 89. *Se dice que una variable Z_j es un instrumento para la variable explicativa X_j si cumple las siguientes propiedades*

1. $\operatorname{plim} \sum_{i=1}^{n} Z_{ji}u_i/n = 0$,
2. $\operatorname{plim} \sum_{i=1}^{n} X_{ji}Z_{ji}/n = \sigma_{X_jZ_j} < \infty$,
3. $\operatorname{plim} \sum_{i=1}^{n} Z_{ji}^2/n = \sigma_{Z_jZ_j} < \infty$,

las cuales garantizan que la variable instrumental Z_j (1) esté asintóticamente incorrelacionada con la perturbación aleatoria, (2) esté asintóticamente correlacionada con la variable instrumentada X_j y (3) tenga varianza finita.

Observación 70. *Si el regresor X_j no está correlacionado con la perturbación u, entonces X_j es su propio instrumento. De aquí, podemos definir una matriz $\mathbf{Z}$ de orden $n \times k$ cuyas columnas contienen los instrumentos para cada una de las variables explicativas.*

DEFINICIÓN 90. *Se dice que $\mathbf{Z}$ es una matriz $n \times k$ de variables instrumentales para $\mathbf{X}$ si cumple las siguientes propiedades*

1. $\operatorname{plim} \mathbf{Z}'\mathbf{u}/n = \mathbf{0}$,
2. $\operatorname{plim} \mathbf{Z}'\mathbf{X}/n = \mathbf{\Sigma}_{zx}$,
3. $\operatorname{plim} \mathbf{Z}'\mathbf{Z}/n = \mathbf{\Sigma}_{zz}$,

siendo $\mathbf{\Sigma}_{zx}$ y $\mathbf{\Sigma}_{zz}$ matrices cuadradas no estocásticas y de rango pleno.

Observación 71. *El requisito de rango pleno, $\rho(\boldsymbol{\Sigma}_{zx}) = \rho(\boldsymbol{\Sigma}_{zz}) = k$, garantiza que no hay instrumentos repetidos ni redundantes (ausencia de multicolinealidad) y, por tanto, que las matrices de segundos momentos son no singulares (invertibles).*

Podemos usar la propiedad de ortogonalidad entre instrumentos y perturbación para proponer un estimador alternativo $\hat{\boldsymbol{\beta}}_{VI}$ basado en el método de momentos. La contrapartida muestral de esta condición es

$$\sum_{i=1}^{n} \mathbf{z}_i \hat{u}_i = \sum_{i=1}^{n} \mathbf{z}_i (Y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_{VI}) = 0$$

en donde $\hat{u}_i \, (i = 1, \; 2, \; \ldots, \; n)$ son los residuos derivados del nuevo estimador, $\mathbf{x}_i = (1 \; X_{2i} \; \ldots \; X_{ki})'$ es el vector $k \times 1$ que contiene la observación i-ésima de cada variable explicativa, y $\mathbf{z}_i = (1 \; Z_{2i} \; \ldots \; Z_{ki})'$ es el vector $k \times 1$ que contiene la observación i-ésima de cada variable instrumental. Alternativamente, usando la notación matricial compacta,

$$\mathbf{Z}'\hat{\mathbf{u}} = \mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{VI}) = \mathbf{0}$$

En ambos casos obtenemos el sistema de "ecuaciones normales"

$$\sum_{i=1}^{n} \mathbf{z}_i Y_i = \mathbf{z}_i \mathbf{x}_i' \hat{\boldsymbol{\beta}}_{VI} \quad \Longleftrightarrow \quad \mathbf{Z}'\mathbf{y} = \mathbf{Z}'\mathbf{X}\hat{\boldsymbol{\beta}}_{VI}$$

Definición 91. *El estimador de variable instrumental simple es*

$$\hat{\boldsymbol{\beta}}_{VI} = (\mathbf{z}_i \mathbf{x}_i')^{-1} \sum_{i=1}^{n} \mathbf{z}_i Y_i \quad \Longleftrightarrow \quad \hat{\boldsymbol{\beta}}_{VI} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$$

Observación 72. *Si todos los regresores están incorrelacionados con la perturbación, entonces $\mathbf{Z} = \mathbf{X}$ y $\hat{\boldsymbol{\beta}}_{VI} = \hat{\boldsymbol{\beta}}_{MCO}$.*

Proposición 109. *El estimador de variables instrumentales es consistente.*

Demostración. Reemplazando $\mathbf{y}$ por su expresión en el modelo

$$\hat{\boldsymbol{\beta}}_{VI} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u} = \boldsymbol{\beta} + \left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)^{-1} \frac{\mathbf{Z}'\mathbf{u}}{n}$$

y tomado límites probabilísticos

$$\operatorname{plim} \hat{\boldsymbol{\beta}}_{VI} = \boldsymbol{\beta} + \operatorname{plim} \left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)^{-1} \operatorname{plim} \frac{\mathbf{Z}'\mathbf{u}}{n} = \boldsymbol{\beta} + \boldsymbol{\Sigma}_{zx}\mathbf{0} = \boldsymbol{\beta}$$

□

Proposición 110. *La distribución asintótica de $\sqrt{n}(\hat{\boldsymbol{\beta}}_{VI} - \boldsymbol{\beta})$ es $N(\mathbf{0}, \sigma_u^2 \boldsymbol{\Sigma}_{zx}^{-1} \boldsymbol{\Sigma}_{zz} \boldsymbol{\Sigma}_{xz}^{-1})$.*

Demostración. Como en la proposición 105, $\mathbf{Z}'\mathbf{u}/\sqrt{n}$ converge en distribución a $N(\mathbf{0}, \sigma_u^2 \boldsymbol{\Sigma}_{zz})$. □

En muestras finitas, parece razonable estimar la matriz de varianzas y covarianzas del estimador de variables instrumentales como

$$\hat{V}(\hat{\boldsymbol{\beta}}_{VI}) = \hat{\sigma}_u^2 (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1} = \hat{\sigma}_u^2 (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1} = \hat{\sigma}_u^2 (\mathbf{X}'\mathbf{P}_z\mathbf{Z})^{-1}$$

en donde $\mathbf{P}_z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}$ y

$$\hat{\sigma}_u^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{VI})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{VI})/n$$

El estimador de variable instrumentales simple utiliza un instrumento para cada variable explicativa. La matriz $\mathbf{Z}$ se diseña como la matriz $\mathbf{X}$ reemplazando las columnas de regresores esotcásticos por los instrumentos. Sin embargo, cuando disponemos de varios instrumentos para una variable explicativa, podemos diseñar una matriz $\mathbf{Z}$ de orden $n \times r$ con $r > k$. Ahora la ordenación de los instrumentos en la matriz $\mathbf{Z}$ es irrelevante porque vamos a utilizar como instrumentos la matriz de valores ajustados en la regresión de $\mathbf{X}$ sobre $\mathbf{Z}$.

Definición 92. *El estimador de variables instrumentales generalizado es*

$$\hat{\boldsymbol{\beta}}_{VI} = (\mathbf{X}'\mathbf{P}_z\mathbf{X}')^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y}$$

en donde $\mathbf{P}_z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

Observación 73. Cuando $\mathbf{Z}$ es de orden $n \times k$, el estimador de variables instrumentales generalizado coincide con el simple. Si $r = k$, la matriz $\mathbf{X}'\mathbf{Z}$ es cuadrada y no singular. Así,

$$(\mathbf{X}'\mathbf{P}_Z\mathbf{X}')^{-1} = (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X})^{-1} = (\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{X}'\mathbf{Z})^{-1}$$

en donde las dos últimas matrices se compensan con las dos primeras matrices de

$$\mathbf{X}'\mathbf{P}_Z\mathbf{y} = (\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{y})$$

resultando que

$$\hat{\boldsymbol{\beta}}_{VI} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X}')^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$$

Observación 74. El estimador de variables instrumentales puede interpretarse como el estimador de mínimos cuadrados en dos etapas

$$\hat{\boldsymbol{\beta}}_{MC2E} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}$$

en donde $\hat{\mathbf{X}} = \mathbf{P}_z\mathbf{X}$ es la matriz de valores ajustados en la regresión de $\mathbf{X}$ y $\mathbf{Z}$ estimada en la primera etapa, y $\hat{\boldsymbol{\beta}}_{MC2E}$ es el estimador de mínimos cuadrados en la regresión de $\mathbf{y}$ sobre $\mathbf{X}$ estimada en la segunda etapa.

12.4. Modelos econométricos con endogeneidad

La solución al problema de la inconsistencia del estimador de mínimos cuadrados en modelos con endogeneidad requiere el uso instrumentos. Ahora bien, esto no significa que tales instrumentos existan. Hay, no obstante, tres tipos de modelos con endogeneidad para los cuales podemos dar unas reglas para la elección de instrumentos:

1. el modelo $ARMA(p, q)$ como $p > 1$ y $q > 1$,
2. el modelo con variables medidas con error,
3. el modelo de ecuaciones simultáneas.

12.4.1. Modelos ARMA. Sea el modelo $ARMA(1, 1)$

$$\begin{aligned} Y_t =& \delta + \phi_1 Y_{t-1} + u_t, \quad (t = 1,\ 2,\ \dots,\ n) \\ u_t =& \theta u_{t-1} + a_t, \quad a_t \sim iidN(0, \sigma_a^2) \end{aligned}$$

Proposición 111. *El estimador de mínimos cuadrados de la "pendiente"*

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n}(Y_{t-1} - \bar{Y})(Y_t - \bar{Y})}{\sum_{t=2}^{n}(Y_{t-1} - \bar{Y})^2}$$

es inconsistente.

Demostración. Tomando límites probabilisticos

$$\text{plim}\,\hat{\phi}_1 = \frac{\gamma_1}{\gamma_0} = \phi_1 + \frac{\theta\sigma_a^2}{\gamma_0}$$

□

Proposición 112. *Usando Y_{t-2} como instrumento de Y_{t-1}, el estimador de variable instrumental*

$$\tilde{\phi}_1 = \frac{\sum_{t=2}^{n}(Y_{t-2} - \bar{Y})(Y_t - \bar{Y})}{\sum_{t=2}^{n}(Y_{t-2} - \bar{Y})(Y_{t-1} - \bar{Y})}$$

es consistente.

Demostración. Tomando límites probabilisticos

$$\text{plim}\,\tilde{\phi}_1 = \frac{\gamma_2}{\gamma_1} = \phi_1$$

□

Observación 75. *En un modelo $ARMA(p,q)$ podemos usar como instrumentos de Y_{t-1}, ..., Y_{t-k} los retardos Y_{t-1-q}, ..., Y_{t-k-q}.*

12.4.2. Errores en las variables. Supongamos que la relación verdadera entre dos variables Y_t y X_t es

$$Y_t = \beta_1 + \beta X_t + u_t, \quad u_t \sim iidN(0, \sigma_u^2)$$

y que la variable $X_t \sim N(\mu, \sigma_X^2)$ se mide con error

$$X_t^* = X_t + w_t, \quad w_t \sim iidN(0, \sigma_w^2)$$

El modelo que vamos a estimar es

$$Y_t = \beta_1 + \beta_2 X_t^* + v_t$$

en donde

$$v_t = u_t - \beta w_t$$

Proposición 113. *Si u_t, v_t, y X_t son mútuamente independientes, el estimador de mínimos cuadrados de la pendiente es inconsistente.*

Demostración. El estimador de mínimos cuadrados de la pendiente

$$\hat{\beta}_2 = \frac{\sum_{t=1}^{n}(X_t^* - \bar{X}^*)(Y_t - \bar{Y})}{\sum_{t=1}^{n}(X_t^* - \bar{X}^*)}$$

converge en probabilidad a

$$\text{plim}\,\hat{\beta}_2 = \frac{cov(X_t^*, Y_t)}{var(X_t^*)} = \frac{\beta_2\sigma_X^2}{\sigma_X^2 + \sigma_w^2} \le \beta_2$$

□

Observación 76. *Conviene recordar que la causa de la inconsistencia es la correlación entre el regresor y la perturbación estocástica. Los errores de medición en la variable dependiente no causan inconsistencia, salvo que estén relacionados con el regresor.*

Proposición 114. *Un estimador consistente de* β_2 *es*

$$\beta_2 = \frac{\bar{Y}_2 - \bar{Y}_1}{\bar{X}_2^* - \bar{X}_1^*}$$

en donde $\bar{X}_1$ *y* $\bar{X}_2$ *son las medias de los valores de* X^* *que son menores o mayores que la mediana* $\tilde{X}$*, esto es,*

$$\bar{X}_1 = \frac{n}{2}\sum_{t=1}^{n} X_t^* \mathcal{I}(X_t^* < \tilde{X}) \quad y \quad \bar{X}_2 = \frac{n}{2}\sum_{t=1}^{n} X_t^* \mathcal{I}(X_t^* > \tilde{X})$$

e $\bar{Y}_1$ *e* $\bar{Y}_2$ *son las medias de los valores de* Y *en esos dos grupos*

$$\bar{Y}_1 = \frac{n}{2}\sum_{t=1}^{n} Y_t^* \mathcal{I}(X_t < \tilde{X}) \quad e \quad \bar{Y}_2 = \frac{n}{2}\sum_{t=1}^{n} Y_t \mathcal{I}(X_t^* > \tilde{X})$$

Demostración. Escribiendo el modelo como

$$Y_t = \underbrace{\beta_1 + \beta_2 \bar{X}^*}_{\beta_1^*} + \beta_2(X_t^* - \bar{X}^*) + v_t$$

el esimador propuesto puede contemplarse como un estimador de variable instrumental usando Z_t como instrumento para $X_t^* - \bar{X}^*$, en donde $Z_t = 1$ para $X_t > \tilde{X}$ y $Z_t = -1$ para $X_t < \tilde{X}$. Así, suponiendo que n es par,

$$\begin{pmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2 \end{pmatrix} = (\mathbf{Z}'\mathbf{X}^*)^{-1}\mathbf{Z}'\mathbf{y} = \begin{pmatrix} n & 0 \\ 0 & \frac{n}{2}(\bar{X}_2^* - \bar{X}_1^*) \end{pmatrix}^{-1} \begin{pmatrix} n\bar{Y} \\ \frac{n}{2}(\bar{Y}_2 - \bar{Y}_1) \end{pmatrix} = \begin{pmatrix} \bar{Y}_1 \\ (\bar{Y}_2 - \bar{Y}_1)/(\bar{X}_2^* - \bar{X}_1^*) \end{pmatrix}$$

□

Observación 77. *Usando los cuantiles del 33 y 66 por ciento, podemos formar tres grupos de* $n/3$ *observaciones. Omitiendo el grupo central, podemos estimar la pendiente como*

$$\beta_2 = \frac{\bar{Y}_3 - \bar{Y}_1}{\bar{X}_3^* - \bar{X}_1^*}$$

Palabras clave

Regresores estocásticos	Variable instrumental
Exogeneidad fuerte	Mínimos cuadrados en dos etapas
Exogeneidad débil	Errores en variables
Endogeneidad	Sesgo de medición
Teoría asintótica	Sesgo de simultaneidad

12.5. Ejercicios

1. Demuestre que el estimador de variables instrumentales generalizado es consistente.
2. Demuestre que el estimador $\hat{\sigma}_u^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{VI})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{VI})/n$ es consistente.
3. En el modelo de regresión lineal simple con endogeneidad, obtenga el estimador de variable instrumental simple para la ordenada y la pendiente usando la notación escalar.
4. Demuestre que el estimador de variables instrumentales generalizado coincide con el estimador de mínimos cuadrados ordinarios cuando el número de instrumentos es igual que el número de observaciones.

5. Usando el modelo de regresión simple, demuestre que los errores en la medición de la variable dependiente no causan inconsistencia.
6. En el modelo de regresión simple con errores en la variable explicativa, demuestre que
7. Imagine que la variable explicativa X_t que se mide con error sigue un proceso $AR(1)$. Comprueba que Y_{t-1} y X^*_{t-1} son dos instrumentos válidos para X_t.