

## TEMA: REGRESORES ESTOCÁSTICOS EN EL MODELO DE REGRESIÓN LINEAL GENERAL (MRLG).

### 1. Introducción.

La hipótesis de que los regresores del MRLG son deterministas es poco realista y restrictiva. Veamos 4 casos en los que no se cumple este supuesto.

**A) Simultaneidad.** Sea el modelo de regresión:

$$y_{1t} = \beta_0 + \beta_1 y_{2t} + \beta_2 x_t + \varepsilon_t \quad (1)$$

donde ignoramos que  $y_{2t}$  es una variable endógena que sigue el modelo:

$$y_{2t} = \delta_0 + \delta_1 x_t + v_t \quad (2)$$

Pueden plantearse distintas situaciones si:

- A.1)  $y_{1t}$  es también una variable explicativa de (2) o no.
- A.2) La variable determinista  $x_t$  de (1) y (2) es la misma o no.
- A.3) Los términos de error  $\varepsilon_t$  y  $v_t$  están correlacionados o no.

Pero sin entrar en cualquiera de esas 3 situaciones, nos fijamos sólo en (1). Es evidente que la presencia de la variable  $y_{2t}$  en (1) invalida el supuesto de regresores fijos o deterministas.

**B) Modelos dinámicos.** En el modelo de regresión:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + \varepsilon_t \quad (3)$$

La presencia de  $y_{t-1}$  como regresor en (3) invalida el supuesto, ya que  $y_{t-1} = \beta_0 + \beta_1 y_{t-2} + \beta_2 x_{t-1} + \varepsilon_{t-1}$  y por tanto,  $y_{t-1}$  es un regresor estocástico.

**C) Errores de medida.** Se quiere estimar el modelo:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (4)$$

pero no se dispone de observaciones de  $x_t$  sino de una variable distinta  $x_t^* = x_t + v_t$ , donde  $x_t^*$  es la variable realmente observada y  $v_t$  es un error de medida. Por tanto, sólo es posible estimar el modelo:

$$y_t = \delta_0 + \delta_1 x_t^* + \varepsilon_t^* \quad (5)$$

que no es exactamente igual a (4). No obstante, la presencia de  $x_t^*$  en (5) invalida el supuesto de regresores fijos.

**D) Modelos con expectativas.** Se quiere estimar el siguiente modelo lineal:

$$r_t = \beta_0 + \beta_1 m_t + \beta_2 E_t(\pi_{t+1}) + \varepsilon_t \quad (6)$$

donde  $r_t$  es un tipo de interés,  $m_t$  es una variable sobre cantidad de dinero (un agregado monetario) y  $E_t(\pi_{t+1})$  es la expectativa que se tiene en  $t$  sobre la tasa de inflación futura  $\pi_{t+1}$ . Sabemos que  $E_t(\pi_{t+1})$  no es más que la predicción a horizonte un período de la tasa de inflación y por tanto,  $E_t(\pi_{t+1}) = \pi_{t+1} - e_t(1)$ , donde  $e_t(1)$  denota el error de previsión a horizonte uno. Es claro que la presencia de la variable expectativa en (6) invalida el supuesto del MRLG que estamos analizando.

## 2. Propiedades del estimador MCO en presencia de regresores estocásticos.

Sea el modelo sencillo  $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ , donde  $x_t$  es estocástica. Si estimamos por MCO en desviaciones con respecto a la media de los datos, se tiene que:

$$\hat{\beta}_1 = \frac{\sum y_t x_t}{\sum x_t^2} = \frac{\sum (\beta_1 x_t + \varepsilon_t) x_t}{\sum x_t^2} = \beta_1 + \frac{\sum \varepsilon_t x_t}{\sum x_t^2} \quad (7)$$

Es conveniente recordar que:

- 1) Si  $x_t$  es una variable determinista,  $\hat{\beta}_1$  es un estimador insesgado de  $\beta_1$ , ya que  $E[\varepsilon_t] = 0$ .
- 2) Si  $x_t$  es estocástica, entonces:

$$E[\hat{\beta}_1] = \beta_1 + E\left[\frac{\sum \varepsilon_t x_t}{\sum x_t^2}\right] \quad (8)$$

donde la esperanza del segundo término del lado derecho de (8) es una función no lineal de las variables  $\varepsilon_t$  y  $x_t$ .

**Conclusión:** Cuando no puede asegurarse la insesgaredad de un estimador, hay que confiar en sus propiedades asintóticas, tratando de ver si el sesgo del estimador tiende a desaparecer a medida que aumenta el tamaño de la muestra. Es decir, en lugar de centrarnos en el sesgo, nos centraremos en la *consistencia* del estimador.

**ANEXO 1. Definición:** Un estimador  $\hat{\beta}_N$  obtenido con una muestra de tamaño  $N$ , es un estimador consistente de  $\beta$  si:

$$\lim_{N \rightarrow \infty} Pr[|\hat{\beta}_N - \beta| < \varepsilon] = 1 \quad (9)$$

cuando  $\varepsilon$  es una cantidad positiva, tan pequeña como queramos. Es decir, un estimador es consistente si, en el límite (cuando  $N$  es muy grande), la probabilidad de que el estimador tome valores tan cercanos como queramos al verdadero valor  $\beta$ , tiende a uno. En otras palabras, el estimador converge en probabilidad a  $\beta$ :

$$\hat{\beta}_N \xrightarrow{p} \beta \quad (10)$$

o bien:  $plim \hat{\beta}_N = \beta$ . Tanto (9) como (10) indican la consistencia de  $\hat{\beta}_N$ . De una forma gráfica, y suponiendo que la distribución de  $\hat{\beta}_N$  sea simétrica y distintos tamaños muestrales  $n_1 < n_2 < n_3, \dots$ , la consistencia implica que la distribución muestral es menos dispersa (más plana) a medida que crece  $n$  y cuando  $n \equiv \infty$ , la distribución colapsará con  $\beta$ .

**Lema:** Para que  $\hat{\beta}_N$  sea un estimador consistente de  $\beta$ , son condiciones suficientes:

a)  $\lim_{N \rightarrow \infty} E[\hat{\beta}_N] = \beta$

b)  $\lim_{N \rightarrow \infty} VAR[\hat{\beta}_N] = 0$

Nótese que a) y b) son condiciones suficientes de consistencia, pero no necesarias.

**ANEXO 2. Propiedades del operador  $plim$  (límite en probabilidad).** Si tenemos dos estimadores consistentes ( $\hat{\beta}_N$  y  $\hat{\alpha}_N$ ), se cumplen las siguientes propiedades:

a)  $plim(\hat{\alpha}_N \pm \hat{\beta}_N) = plim \hat{\alpha}_N \pm plim \hat{\beta}_N = \alpha \pm \beta$

b)  $plim(\hat{\alpha}_N \cdot \hat{\beta}_N) = plim \hat{\alpha}_N \cdot plim \hat{\beta}_N = \alpha \cdot \beta$

c)  $plim(\hat{\alpha}_N / \hat{\beta}_N) = plim \hat{\alpha}_N / plim \hat{\beta}_N = \alpha / \beta$ , siempre que  $\beta \neq 0$ .

d)  $plim f(\hat{\alpha}_N) = f(\alpha)$  siendo  $f$  una función continua cualesquiera.

e)  $plim \alpha = \alpha$ , puesto que  $\alpha$  es un parámetro constante.

Volvamos al ejemplo inicial para estudiar si  $\hat{\beta}_1$  es un estimador consistente cuando  $x_t$  es una variable estocástica. Tomando límites en probabilidad en (7):

$$plim \hat{\beta}_1 = \beta_1 + \frac{plim \sum \varepsilon_t x_t}{plim \sum x_t^2} = \beta_1 + \frac{plim \frac{\sum \varepsilon_t x_t}{N}}{plim \frac{\sum x_t^2}{N}} \quad (11)$$

donde usamos que  $\frac{\sum x_t^2}{N} = \frac{\sum (x_t - \bar{x})^2}{N}$  es un estimador consistente de la  $var(x_t)$ , por lo que:

$$plim \frac{\sum x_t^2}{N} = var(x_t)$$

También sabemos que  $\frac{\sum \varepsilon_t x_t}{N}$  es un estimador consistente de la  $cov(\varepsilon_t, x_t)$ , por lo que volviendo a (11):

$$plím \hat{\beta}_1 = \beta_1 + \frac{cov(\varepsilon_t, x_t)}{var(x_t)} \quad (12)$$

Concluyendo, para que  $\hat{\beta}_1$  sea consistente, tiene que anularse la covarianza entre  $\varepsilon_t$  y  $x_t$ . En el caso de que el regresor, aún siendo estocástico, no esté correlacionado con el error  $\varepsilon_t$ , el estimador MCO seguirá siendo consistente. Pero si el regresor está correlacionado con el error del modelo, el estimador MCO ni siquiera será consistente. En esta última situación, habrá que buscar un procedimiento distinto de estimación.

### 3. Ejemplos: Modelos dinámicos.

Vamos a estudiar la consistencia del estimador MCO en distintos modelos dinámicos.

CASO 1: Supongamos el siguiente modelo dinámico:

$$y_t = \beta y_{t-1} + \varepsilon_t \quad (13)$$

donde  $|\beta| < 1$ ;  $\varepsilon_t$  es un proceso de ruido blanco y las variables  $y_t$  e  $y_{t-1}$  están medidas en desviaciones con respecto a su media. Cabe señalar que:

- Se incumple el supuesto de regresores fijos, porque  $y_{t-1} = f(\varepsilon_{t-1})$  y por tanto, el regresor de (13) es estocástico.
- El estimador MCO de  $\beta$  en (13) es consistente, ya que  $\varepsilon_t$  es ruido blanco. Es decir,  $y_{t-1} = f(\varepsilon_{t-1}) \neq f(g(\varepsilon_t))$ , ya que  $\varepsilon_t$  tiene covarianza nula con  $\varepsilon_{t-1}$ . Se ha visto que esta condición es la que se tiene que cumplir para garantizar la consistencia del estimador MCO.

Demostración: El estimador MCO de  $\beta$  en (13) tiene la expresión:

$$\hat{\beta}_{MCO} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} = \beta + \frac{\sum \varepsilon_t y_{t-1}}{\sum y_{t-1}^2} \quad (14)$$

Estudiamos consistencia tomando límites en probabilidad a (14):

$$plím \hat{\beta}_{MCO} = \beta + \frac{plím \frac{\sum \varepsilon_t y_{t-1}}{N}}{plím \frac{\sum y_{t-1}^2}{N}} = \beta + \frac{cov(\varepsilon_t, y_{t-1})}{var(y_{t-1})} \quad (15)$$

Para calcular la  $cov(\varepsilon_t, y_{t-1})$ , sabemos que (13) se puede escribir como:

$$(1 - \beta B)y_{t-1} = \varepsilon_t, \text{ o bien, } y_t = \varepsilon_t + \beta \varepsilon_{t-1} + \beta^2 \varepsilon_{t-2} + \dots = \sum_{s=0}^{\infty} \beta^s \varepsilon_{t-s} \quad (16)$$

A partir de (16) se tiene que:

$$cov(\varepsilon_t, y_{t-1}) = E[(\sum_{s=0}^{\infty} \beta^s \varepsilon_{t-s-1}) \varepsilon_t] = 0 \quad (17)$$

porque  $E[\varepsilon_t \varepsilon_{t-s}] = 0$ ,  $\forall t, s$ .

**CASO 2:** Supongamos el siguiente modelo dinámico:

$$y_t = \beta y_{t-1} + \varepsilon_t \quad (18)$$

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + a_t \quad (19)$$

donde  $|\beta| < 1$ ,  $|\phi_1| < 1$  y  $a_t$  es un proceso de ruido blanco. En el modelo (18) los datos están medidos en desviaciones con respecto a la media. En primer lugar, por el mismo motivo que en el Caso 1, sabemos que existe un regresor estocástico en (18). En segundo lugar, sabemos que el estimador MCO de (18) es inconsistente porque el error sigue una estructura AR(1), en donde la covarianza entre  $\varepsilon_t$  y  $\varepsilon_{t-1}$  es distinta de cero. Por ello, la covarianza entre  $y_{t-1}$  y  $\varepsilon_t$  será también distinta de cero.

**Demostración:** El error de (18) según (19) sigue una estructura AR(1). Ello implica que  $\gamma_j = cov(\varepsilon_t, \varepsilon_{t-j}) = \phi_1^j \gamma_0$  donde  $\gamma_0 = var(\varepsilon_t) = \sigma_a^2 / (1 - \phi_1^2)$ . Además, igual que en el Caso 1:

$$plím \hat{\beta}_{MCO} = \beta + \frac{plím \frac{\sum \varepsilon_t y_{t-1}}{N}}{plím \frac{\sum y_{t-1}^2}{N}} = \beta + \frac{cov(\varepsilon_t, y_{t-1})}{var(y_{t-1})} \quad (20)$$

$$cov(\varepsilon_t, y_{t-1}) = E[(\sum_{s=0}^{\infty} \beta^s \varepsilon_{t-s-1}) \varepsilon_t] = E[\varepsilon_{t-1} \varepsilon_t] + E[\beta \varepsilon_{t-2} \varepsilon_t] + \dots = \phi_1 \gamma_0 + \beta \phi_1^2 \gamma_0 \dots \neq 0$$

La conclusión es que en este caso es necesario buscar un estimador distinto al MCO que, al menos, sea consistente.

#### 4. Estimador de Variables Instrumentales (VI).

En el caso del modelo (18)-(19) para estimar consistentemente por VI, hay que buscar una variable instrumental, denotada por  $z_t$ , que sustituya al regresor  $y_{t-1}$ , de manera que cumpla dos condiciones:

- 1) Un instrumento válido para  $y_{t-1}$  ha de ser una variable distinta pero correlacionada con ella. Es decir, la  $cov(y_{t-1}, z_t) \neq 0$ .
- 2) Un instrumento válido no puede estar correlacionado con el término de error,  $\varepsilon_t$ . Es decir, la  $cov(\varepsilon_t, z_t) = 0$ .

### Búsqueda de instrumentos válidos:

En el caso que nos ocupa hay que buscar instrumento(s) válido(s) para la variable  $y_{t-1}$ . Se comienza siempre probando con el pasado de esta variable, es decir, con  $y_{t-2}$ ,  $y_{t-3}$ , ... etc. Escogemos  $z_t = y_{t-2}$  y vemos si cumple las dos condiciones:

- 1) El instrumento está correlacionado con  $y_{t-1}$ , de acuerdo con el propio modelo.
- 2) Esta condición no se cumple ya que  $y_{t-2} = f(\varepsilon_{t-2}) = f(g(\varepsilon_t))$ . En concreto, al seguir  $\varepsilon_t$  un AR(1), la  $cov(\varepsilon_t, \varepsilon_{t-2}) = \gamma_0 \phi_1^2$ . La conclusión es que no sirven como instrumentos ningún retardo de  $y_{t-1}$ , al no cumplir ninguno la condición 2).

Sin embargo, si en el modelo bajo estudio se incluye además un regresor determinista, denotado por  $x_t$ :

$$y_t = \beta y_{t-1} + \delta x_t + \varepsilon_t \quad (21)$$

siguiendo el error un proceso AR(1), es fácil encontrar un instrumento adecuado que sería  $x_{t-1}$ . Comprobando las dos condiciones:

- 1)  $x_{t-1}$  está correlacionado con  $y_{t-1}$ . Esto lo indica el propio modelo:  $y_{t-1} = \beta y_{t-2} + \delta x_{t-1} + \varepsilon_{t-1}$
- 2)  $x_{t-1}$  no está correlacionado con  $\varepsilon_t$ . Esto es así, porque si  $x_t$  es determinista, sus retardos también lo serán y por tanto, la  $cov(x_{t-1}, \varepsilon_t) = 0$ .

### **5. Estimador de Variables Instrumentales (VI).**

#### Expresión del estimador de VI:

Este estimador no surge de un criterio de optimalidad, como ocurre con el estimador MCO que minimiza la suma de los cuadrados de los residuos. Simplemente es un estimador consistente cuya expresión matricial es la siguiente:

$$\hat{\beta}_{VI} = (Z^T X)^{-1} Z^T Y \quad (22)$$

donde  $Z$  es una matriz de tamaño  $(n \times k)$  en la que los regresores estocásticos del modelo se sustituyen por sus instrumentos y los deterministas se dejan en el modelo tal como aparecen. La matriz  $X$  recoge los regresores del modelo original, siendo también de tamaño  $(n \times k)$ . Por último,  $Y$  es un vector de tamaño  $(n \times 1)$  que recoge las observaciones de la variable endógena del modelo. En el ejemplo anterior, donde  $y_t = \beta y_{t-1} + \delta x_t + \varepsilon_t$ , la expresión matricial del estimador por VI se ofrece a continuación. Escribiendo las columnas de las matrices  $Z$  y  $X$ , en este caso:

$$Z = [x_{t-1} \quad x_t] \quad ; \quad X = [y_{t-1} \quad x_t]$$
$$\begin{bmatrix} \hat{\beta} \\ \hat{\delta}_{VI} \end{bmatrix} = \begin{bmatrix} \sum x_{t-1} y_{t-1} & \sum x_{t-1} x_t \\ \sum x_t y_{t-1} & \sum x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_t x_{t-1} \\ \sum y_t x_t \end{bmatrix} \quad (23)$$

Nótese que la matriz  $Z^T X$  no es simétrica, siendo esto importante para calcular su inversa.

Propiedades del estimador de VI:

1) Es un estimador consistente. La demostración es sencilla. Tomando límites en probabilidad a la expresión (22):

$$\begin{aligned} \text{plím } \hat{\beta}_{VI} &= \text{plím}[(Z^T X)^{-1} Z^T Y] = \text{plím}[(Z^T X)^{-1} Z^T (X\beta + \varepsilon)] = \\ &= \text{plím}[\beta + (Z^T X)^{-1} Z^T \varepsilon] = \beta + \text{plím}[(Z^T X)^{-1} Z^T \varepsilon] = \\ &= \beta + \text{plím} \left[ \frac{Z^T X}{N} \right]^{-1} \text{plím} \left[ \frac{Z^T \varepsilon}{N} \right] = \beta \end{aligned} \quad (24)$$

donde  $\text{plím} [Z^T \varepsilon]/N = 0$ , ya que las variables instrumentales siempre se escogen para que no tengan correlación con el término de error de la ecuación.

2) No es un estimador eficiente. La matriz de varianzas-covarianzas del estimador por VI no es necesariamente la varianza asintótica mínima. La razón es que, en general, existen múltiples instrumentos no correlacionados con la perturbación y sí con las variables explicativas estocásticas. Por tanto, escogiendo un único instrumento, el estimador puede que no sea eficiente. No obstante, al no tener otra alternativa, el estimador por VI es, al menos, consistente.

Ejercicios para el alumno: 1) Estudiar la consistencia del estimador MCO en el modelo (13), cuando el error sigue una estructura de tipo MA(1) y de tipo MA(2). En el caso de inconsistencia, proponer un estimador de VI adecuado.

2) Estudiar la consistencia del estimador MCO en el modelo (21), cuando el término de error tiene una estructura de tipo ARMA(1,1). En el caso de inconsistencia, proponer un estimador de VI adecuado.