Tema 5

Regresores Estocásticos

5.1. Introducción

En los temas anteriores hemos mantenido el supuesto básico sobre los regresores, de que X era una matriz de variables explicativas no estocásticas o fijas. Este supuesto es apropiado para experimentos controlados o de laboratorio, para variables como la tendencia o variables ficticias, o cuando trabajamos condicionando a una muestra dada. En este tema relajaremos dicho supuesto para adecuarnos a situaciones en las que este supuesto no sea razonable.

Objetivo del tema: Relajar la hipótesis básica de que X es una matriz de variables no estocásticas considerando que X es una matriz estocástica. Para ello basta con que uno de los regresores incluidos en X sea estocástico.

En este tema analizaremos si los métodos de estimación e inferencia vistos hasta ahora son válidos cuando X es estocástica. En caso de que no sea así analizaremos qué métodos alternativos están disponibles.

Si X es una matriz de regresores estocásticos, el estimador MCO de β, es una función estocástica no lineal de X y u y por tanto, sus propiedades dependen de la distribución conjunta de ambas variables. Esto dificultará, en general, el conocimiento de las propiedades en muestras finitas de este estimador y sus propiedades asintóticas dependerán de la relación entre X y u. En consecuencia también se verá afectada la inferencia.

Para ilustrar alguna de estas situaciones vamos a considerar los siguientes ejemplos:

Ejemplo 5.1 Sea el siguiente modelo de regresión para explicar la productividad de un trabajador, Y_t, medida como número de piezas realizadas en el día:

$$Y_t = \beta_1 + \beta_2 X_t^\star + u_t^\star \qquad u_t^\star \sim N(0, \sigma^2_{u^\star}) \qquad t = 1, 2, \ldots, T$$

siendo $X_t^\star$ la habilidad de un trabajador, variable no observable ya que es difícil de cuantificar. En su lugar observamos la variable X_t años de experiencia del trabajador en el puesto de trabajo, y vamos a utilizarla para aproximar la habilidad del trabajador[1]. Es probable que cuanto más antigüedad

[1] A una variable que aproxima a otra en la literatura Econométrica se le denomina variable "proxy".

tenga en el puesto el trabajador, más hábil sea y viceversa, así que parece sensato suponer que existe relación entre ambas variables. Sin embargo, la aproximación no será exacta, conllevará un error de medida. Será tal que:

$$X_t = X_t^\star + v_t \qquad v_t \sim N(0, \sigma_v^2) \qquad Cov(u_t^\star, v_t) = 0 \quad \forall t \qquad t = 1, 2, \ldots, T$$

donde v_t es una v.a. que recoge el error de medida en t y se le supone independiente de $u_t^\star$. En esta situación, X_t es una v.a. aunque consideremos a $X_t^\star$ como no estocástica. El modelo estimable en términos de la variable observable sería:

$$\begin{aligned} Y_t &= \beta_1 + \beta_2(X_t - v_t) + u_t^\star \\ Y_t &= \beta_1 + \beta_2 X_t + (u_t^\star - \beta_2 v_t) \\ Y_t &= \beta_1 + \beta_2 X_t + u_t \end{aligned}$$

llamamos $u_t = (u_t^\star - \beta_2 v_t)$, en este caso u_t es una función de $u_t^\star$ y v_t, luego de $X_t^\star$; por ello, no tendría sentido realizar un análisis condicionado a unos valores fijos de X matriz de regresores. En el modelo estimable

$$E(X_t u_t) = E[(X_t^\star + v_t)(u_t^\star - \beta_2\, v_t)] = -\beta_2 \sigma_v^2 \neq 0$$

por lo que, el Teorema de Mann y Wald no se cumple. El estimador MCO será inconsistente y es necesario encontrar un método de estimación consistente alternativo. Este caso lo analizaremos en profundidad más adelante y es conocido en la literatura como errores en variables. En este supuesto la variable medida con error es la variable explicativa.

Ejemplo 5.2 Supongamos que se quiere estimar los coeficientes de la siguiente ecuación de demanda de un bien:

$$Q_t = \beta_1 + \beta_2 P_t + u_t \qquad t = 1, 2, \ldots, T$$

donde Q es la cantidad demandada y P es el precio. Dado que en el momento t observamos la cantidad y precio de equilibrio, ambas variables se determinan simultáneamente en el mercado. Luego, tanto Q como P son variables endógenas. Si en t se produce un shock en la demanda del bien debido, por ejemplo, a un cambio en los gustos de los consumidores, al ser recogido por u_t, se genera un cambio tanto en la cantidad demandada como en el precio. En este contexto dado que las variables se determinan simultáneamente ambas son aleatorias. Este es otro ejemplo donde la matriz de regresores es estocástica y no tiene sentido realizar el análisis condicionado a valores fijos de $P_t \quad t = 1, 2, \ldots, T$, dado que P_t se determina simultáneamente a Q_t. Volveremos a retomar este ejemplo más adelante.

Ejemplo 5.3 Supongamos un modelo con dinámica en la parte sistemática porque incluye como regresor a la variable endógena retardada. Por ejemplo, si estudiamos la función de consumo es muy posible que estemos dispuestos a aceptar que, además de la renta actual, R_t un regresor, posiblemente relevante, para explicar el consumo actual, C_t sea el consumo del período anterior, C_{t-1}:

$$C_t = \beta_1 + \beta_2 C_{t-1} + \beta_3 R_t + u_t \qquad t = 2, \ldots, T$$

En este caso la matriz de regresores del modelo es estocástica ya que incluye como regresor a un retardo de la variable endógena, variable estocástica y no tiene sentido realizar el análisis condicionado a valores fijos.

Es necesario revisar los resultados obtenidos para el MRLG bajo el supuesto de que X es una matriz de regresores estocásticos. Comenzaremos revisando la validez del estimador MCO bajo diferentes supuestos de relación entre X estocástica y u.

5.2. Propiedades del estimador MCO

En presencia de regresores estocásticos el método de estimación y las propiedades de los estimadores dependen de la relación entre u y X. Estudiaremos las propiedades del estimador MCO en las siguientes situaciones:

1. Independencia entre regresor y error.
2. Incorrelación contemporánea entre regresor y error.
3. Correlación entre regresor y error.

5.2.1. Independencia entre regresor y error

Sea $Y = X\beta + u$ donde:

- X es una matriz estocástica, (alguno de sus regresores es una v.a.) con una determinada función de densidad $f(X)$, es decir, X toma diferentes valores con diferentes probabilidades.
- $u \sim N(0, \sigma_u^2 I_T)$
- X y u se distribuyen independientemente, es decir, $E(X'u) = E(X')E(u) = 0$.

Vamos a buscar las propiedades del estimador MCO cuando X y u son independientes.

- Propiedades en muestras finitas para valores de X no condicionados:

  - **Linealidad**:
    $\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u$ función no lineal de X y u.
    El estimador $\hat{\beta}_{MCO}$ ya no es una combinación lineal de las perturbaciones, sino que es una función estocástica no lineal de X y u, por lo que, sus propiedades dependen de la distribución conjunta de éstas.

  - **Insesgadez**:
    Dado que X y u son independientes y $E(u) = 0$ por hipótesis básica, tenemos:

$$E(\hat{\beta}_{MCO}) = \beta + E[(X'X)^{-1}X'u] = \beta + E[(X'X)^{-1}X']E[u] = \beta$$

por tanto, $\hat{\beta}_{MCO}$ es insesgado si X y u son independientes y $E[(X'X)^{-1}X']$ existe y es finito[2].

- **Matriz de varianzas y covarianzas**:

$$\begin{aligned} V(\hat{\beta}_{MCO}) = E(\hat{\beta}_{MCO}-\beta)(\hat{\beta}_{MCO}-\beta)' &= E[\,(X'X)^{-1}X'\,uu'\,X(X'X)^{-1}\,] = \\ &= E_X\{E_{u|X}[\,(X'X)^{-1}X'\,uu'\,X(X'X)^{-1}\,]\} = \\ &= E_X\{(X'X)^{-1}X'\,E_{u|X}(uu')\,X(X'X)^{-1}\} = \\ &= E_X\{(X'X)^{-1}X'\,\sigma_u^2 I_T\,X(X'X)^{-1}\} = \\ &= \sigma_u^2\,E_X\{(X'X)^{-1}\} \end{aligned}$$

Luego si $\exists\, E_X\{(X'X)^{-1}\}$, $\hat{\beta}_{MCO}$ sigue siendo el estimador insesgado de mínima varianza[3].

Un estimador insesgado de $V(\hat{\beta}_{MCO})$ donde σ_u^2 y $E_X(X'X)^{-1}$ son los dos elementos desconocidos es:

$$\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}_u^2(X'X)^{-1}$$

- **Distribución e inferencia en muestras finitas**:
El estimador $\hat{\beta}_{MCO}$ es una combinación no lineal de X y u y por tanto, no tiene porqué tener una distribución normal, incluso aunque X e u la tengan. Como consecuencia no tenemos garantizado que $\hat{\beta}_{MCO}$ siga una distribución normal y por tanto, los estadísticos t y F no tienen una distribución exacta conocida por ello, la inferencia para tamaños de muestra pequeños no es válida.

- **Conclusión**: La eliminación del supuesto de que X es no estocástica sustituyéndolo por X estocástica, pero independiente de u no altera las propiedades deseables ni la variabilidad de la estimación mínimo cuadrática.

• Propiedades en muestras grandes:
Bajo los supuestos habituales, y si además, se satisface que $\text{plim}\left(\frac{1}{T}X'X\right) = Q$ finita y no singular el estimador de MCO es consistente y es posible derivar sus propiedades asintóticas utilizando el Teorema de Mann y Wald y el Teorema de Cramer. Las condiciones del Teorema de Mann y Wald son:

i) $u_1, u_2, \ldots, u_T$ v. a tal que $u_t \sim iid(0, \sigma_u^2)$

ii) $E(X'u) = 0$

ii) $\text{plim}\left(\frac{1}{T}X'X\right) = Q$ finita, simétrica y no singular

[2] Para demostrar esta propiedad hemos utilizado los siguientes resultados estadísticos: $E(a) = E_b[E_{a|b}]$ y $E_{WX} = E_X[E_{W|X}]$ ya que $E(\hat{\beta}_{MCO}) = E_X[E(\hat{\beta}_{MCO}|X)] = E_X[\beta + (X'X)^{-1}X'E(u|X)] = \beta$.

[3] Siendo:

E_X el valor esperado de la distribución marginal de X.

$E_{u|X}$ el valor esperado de la distribución condicional de u dado X.

$E_X(X'X)^{-1}$ la matriz de covarianzas poblacional de los regresores calculada en la distribución marginal de X.

Verificamos que se cumplen. La condición iii) se cumple siempre, i) se cumple en este contexto ya que estamos suponiendo perturbaciones esféricas y ii) también se cumple ya X y u son independientes luego $E(X'u) = E(X')E(u) = 0$.

Dado que se cumplen i) + ii) + iii) el Teorema de Mann y Wald produce dos resultados:

a) $\text{plim}\left(\frac{1}{T}X'u\right) = 0$

b) $\frac{1}{\sqrt{T}}X'u \xrightarrow{d} N(0, \sigma_u^2 Q)$

El primer resultado nos garantiza la consistencia del estimador MCO y el segundo nos sirve para encontrar su distribución asintótica. Por lo tanto:

- **Consistencia**: $\hat{\beta}_{MCO}$ es un estimador consistente del parámetro β.

$$\text{plim}\,\hat{\beta}_{MCO} = \text{plim}\,\beta + \text{plim}\left(\frac{1}{T}X'X\right)^{-1} \text{plim}\left(\frac{1}{T}X'u\right) = \beta + Q^{-1}\cdot 0 = \beta$$

- **Distribución asintótica**:
  Dado que
$$\text{plim}\,\hat{\beta}_{MCO} = \beta \Longrightarrow \hat{\beta}_{MCO} \xrightarrow{p} \beta \Longrightarrow (\hat{\beta}_{MCO} - \beta) \xrightarrow{p} 0$$
  el estimador tiene una distribución degenerada en el límite, por lo que, buscamos la distribución asintótica para $\sqrt{T}(\hat{\beta}_{MCO} - \beta)$ tal que
$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) = \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right)$$
  Utilizando el Teorema de Mann y Wald y el Teorema de Cramer obtenemos[4]:
$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) = \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right) \xrightarrow{d} N(0, \sigma_u^2 Q^{-1})$$
  de donde
$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N(0, \sigma_u^2 Q^{-1})$$

- **Inferencia asintótica**: Dado que se cumple el Teorema de Mann y Wald las propiedades asintóticas se mantienen y tiene sentido la inferencia asintótica. Bajo $H_0 : R\beta = r$, los estadísticos t y F se distribuyen asintóticamente como $N(0,1)$ y χ^2_q respectivamente, si el tamaño de la muestra es suficientemente grande. Por lo tanto, podemos utilizar estas distribuciones asintóticas para aproximar la distribución exacta de los estadísticos. Así, para contrastar q restricciones lineales de la forma:
$$H_0 : R\beta = r$$
$$H_a : R\beta \neq r$$

[4]El Teorema de Cramer dice: Si $Y_T = A_T \cdot Z_T$, plim$A_T = A$ y $Z_T \overset{a}{\sim} N(\mu, \sum)$ entonces $A_T Z_T \overset{a}{\sim} N(A_\mu, A \sum A')$. Del Teorema de Mann y Wald hacemos uso del resultado $\frac{X'u}{\sqrt{T}} \xrightarrow{d} N(0, \sigma^2 Q)$.

utilizamos el siguiente estadístico con distribución asintótica:

$$(R\hat{\beta} - r)' \, [R \, \widehat{V}(\hat{\beta}) \, R' \,]^{-1} \, (R\hat{\beta} - r) \; \xrightarrow{d,H_0} \chi^2_q$$

luego

$$(R\hat{\beta}_{MCO} - r)'[R \, \hat{\sigma}^2_u \, (X'X)^{-1}R']^{-1}(R\hat{\beta}_{MCO} - r) \xrightarrow{d,H_0} \chi^2_q$$

Si el estadístico calculado es menor que el valor en tablas de $\chi^2_{q|\alpha}$ no rechazamos la H_0 para un nivel de significación α.

Si q=1 podemos utilizar el siguiente estadístico y distribución asintótica asociada:

$$\frac{R\hat{\beta}_{MCO} - r}{\sqrt{R \, \hat{\sigma}^2_u \, (X'X)^{-1}R'}} \xrightarrow{d,H_0} N(0,1)$$

en cuyo caso no rechazamos H_0 si el valor muestral del estadístico es menor que $N(0,1)_{|\frac{\alpha}{2}}$ para un nivel de significatividad α.

Ejemplo 5.4 Por ejemplo si queremos contrastar la significatividad de una de las variables explicativas del modelo contrastamos: $H_0 : \beta_i = 0$ versus $H_a : \beta_i \neq 0$. En este caso q=1 y podemos escribir el siguiente estadístico a utilizar junto con su distribución asintótica,

$$\frac{\hat{\beta}_{i,MCO}}{\widehat{des}(\hat{\beta}_{i,MCO})} \xrightarrow{d,H_0} N(0,1)$$

Donde $\widehat{des}(\hat{\beta}_{i,MCO})$ es la raíz cuadrada del elemento i-ésimo de la matriz $\hat{\sigma}^2(X'X)^{-1}$. Si el valor muestral del estadístico es menor que $N(0,1)_{|\frac{\alpha}{2}}$ no rechazamos la H_0 para un nivel de significación α.

Ejercicio 5.1 En el modelo lineal simple:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad u_t \sim iid(0, \sigma^2_u) \qquad t = 1, 2, \ldots, T$$

donde X_t es una variable estocástica, pero independiente de la perturbación obtener el estimador de MCO y sus propiedades en muestras finitas.

5.2.2. Incorrelación contemporánea entre regresores y error

Sea $Y = X\beta + u$ con $u \sim N(0, \sigma^2_u I)$, X estocástica tal que X y u no son independientes, pero son incorreladas contemporáneamente, es decir, mantenemos que $E(X_{it}u_t) = 0 \quad i = 1, 2, \ldots, K$, pero no la independencia entre X_{it} y u_t.

En este caso no podemos derivar analíticamente las propiedades en muestras finitas del estimador, pero se sigue cumpliendo el Teorema de Mann y Wald y por tanto, podemos mantener las propiedades asintóticas.

- $\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u$ es una función no lineal de X y u, ambas estocásticas.
- En general es un estimador sesgado, $E(\hat{\beta}_{MCO}) = \beta + E[(X'X)^{-1}X'u]$ y $E[(X'X)^{-1}X'u]$ puede ser distinto de cero ya que X y u no son independientes.
- El cálculo analítico de su matriz de varianzas y covarianzas es complicado dada la no linealidad del estimador en X y u. También lo es el cálculo analítico de $E(\hat{\beta}_{MCO})$.
- Dado que el estimador es no lineal no conocemos su distribución exacta, no sigue una distribución normal aún en el caso de que X_{it} $\forall i$ $\forall t$ y u la sigan. Como consecuencia los estadísticos t y F no tienen distribución exacta conocida.
- Las propiedades asintóticas de consistencia y distribución asintótica se mantienen ya que podemos aplicar el Teorema de Mann y Wald porque se satisfacen las condiciones de este teorema, $u \sim (0, \sigma_u^2 I)$ y $E(X'u) = 0$, junto con $\text{plim}(\frac{1}{T}X'X) = Q$.

Ejercicio 5.2 En el modelo:

$$Y_t = \beta X_t + u_t \qquad t = 1, 2, \ldots, T \qquad u_t \sim N(0, \sigma_u^2)$$

donde X_t es una v.a. no independiente de u_t, pero tal que $E(X_t u_t) = 0$. Demostrar las propiedades en muestras finitas y asintóticas del estimador MCO del coeficiente β.

5.2.3. Correlación entre regresores y error

En ocasiones la hipótesis $E(X_{it}, u_t) = 0$ $\forall i, \forall t$ no es válida. En este caso los estimadores MCO no son ni siquiera consistentes. Hay cuatro puntos a solucionar en relación a este tema:

1. ¿Cómo aparecen las correlaciones entre X y u? Por ejemplo en las siguientes situaciones:
   a) Cuando la variable explicativa está medida con error.
   b) Si el modelo tiene un problema de omisión de variable relevante y la variable omitida está correlada con los regresores. Esta correlación aparecerá vía la perturbación ya que la perturbación recoge las variables omitidas.
   c) En modelos de ecuaciones simultáneas, por ejemplo en el modelo de oferta y demanda donde P y Q se determinan simultáneamente.
   d) En modelos con variable endógena retardada como regresor y perturbación autocorrelada.
2. ¿Qué importancia puede tener este problema? En realidad depende del caso concreto, pero de forma general podemos decir que el estimador de MCO pierde las propiedades en muestras pequeñas y grandes.
3. ¿Cómo podemos detectar que existe el problema? Usando test de contraste que sean capaces de detectar la correlación entre X y u. Por ejemplo, el contraste de Hausman está diseñado para ello.

4. ¿Cómo podemos solucionar el problema? Si la existencia de correlación entre regresores y perturbación se debe a un problema de omisión de variable relevante debemos intentar especificar correctamente el modelo. En el resto de casos tendremos que buscar un método de estimación alternativo a MCO que tenga buenas propiedades, aunque éstas se logren sólo en muestras grandes.

Es claro que este es el caso relevante de los tres descritos. Comenzaremos viendo algunos ejemplos en los que X y u están correlados.

Ejemplo 5.5 Variable exógena medida con error.
Vamos a profundizar en el Ejemplo 5.1 donde analizábamos la productividad de un trabajador, Y, utilizando como variable explicativa su habilidad, X. Sea el modelo de regresión en términos de la variable no observable $X_t^\star \equiv$ productividad:

$$Y_t = \beta_1 + \beta_2 X_t^\star + u_t^\star \qquad t = 1, 2, \ldots, T \qquad u_t^\star \sim iid(0, \sigma^2_{u^\star})$$

donde la variable $X_t^\star$ es una variable no estocástica y no observable, ya que no se puede cuantificar. Sin embargo, disponemos de X_t los años de antigüedad en la empresa, variable observable, y vamos a utilizarla para aproximar a $X_t^\star$. Definimos $X_t = X_t^\star + v_t$, tal que X_t es una variable observable, pero aleatoria ya que incorpora el efecto de v_t, y lo es, aún en el caso de que $X_t^\star$ sea no estocástica o fija. Además, hacemos las siguientes hipótesis:

$$v_t \sim iid(0, \sigma_v^2) \qquad Cov(u_t^\star, v_t) = Cov(u_t^\star, v_s) = Cov(u_s^\star, v_t) = 0$$

En esta situación el modelo que efectivamente se estima es el siguiente:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad \text{con} \qquad u_t = u_t^\star - \beta_2 v_t$$

- Propiedades de la nueva perturbación $u_t^\star$:

$$\begin{aligned}
E(u_t) &= E(u_t^\star - \beta_2 v_t) = E(u_t^\star) - \beta_2 E(v_t) = 0 \\
Var(u_t) &= E(u_t - E(u_t))^2 = E(u_t)^2 = E(u_t^\star - \beta_2 v_t)^2 \\
&= E(u_t^{\star 2}) + \beta_2^2 E(v_t^2) - 2\beta_2 E(u_t^\star v_t) = \sigma^2_{u^\star} + \beta_2^2 \sigma_v^2 - 2\beta_2 \cdot 0 \\
&= \sigma^2_{u^\star} + \beta_2^2 \sigma_v^2 \quad \text{luego homocedástica} \\
Cov(u_t, u_s) &= E[(u_t - E(u_t))(u_s - E(u_s))] = E(u_t^\star - \beta_2 v_t, u_s^\star - \beta_2 v_s) = \\
&= E(u_t^\star u_s^\star) - \beta_2 E(v_t u_s^\star) - \beta_2 E(u_t^\star v_s) + \beta_2^2 E(v_t v_s) = 0, \text{ no autocorrelada}
\end{aligned}$$

- Además, necesitamos conocer la relación entre el regresor y el error, es decir, entre X_t y u_t:

$$\begin{aligned}
Cov(X_t, u_t) &= E(X_t u_t) = E((X_t^\star + v_t)(u_t^\star - \beta_2 v_t)) = \\
&= E(X_t^\star u_t^\star) + E(v_t u_t^\star) - \beta_2 E(X_t^\star v_t) - \beta_2 E(v_t^2) = -\beta_2 \sigma_v^2
\end{aligned}$$

  ya que al ser $X_t^\star$ una variable no estocástica $E(X_t^\star u_t) = E(X_t^\star v_t) = 0$. Al ser la covarianza entre X_t y u_t distinta de cero existe correlación contemporánea entre la variable exógena y la perturbación. Vemos que la correlación depende de la varianza del error de medida, cuanto peor es la proxy o variable que utilizamos para aproximar la variable no observable, mayor es en términos absolutos la correlación. El signo depende del signo del coeficiente β_2. El estimador MCO de β_2 no es consistente, además la diferencia entre $plim\hat{\beta}$ y β depende directamente de esta correlación.

Conclusión: un error de medida en la variable exógena tal que $E(X_t u_t) \neq 0$ implica que los estimadores MCO son sesgados e inconsistentes[5]. El Teorema de Mann y Wald no se satisface. Los coeficientes del modelo deberían ser estimados por un método de estimación que produzca estimadores consistentes. Para conseguirlo propondremos el Método de Variables Instrumentales. Lo veremos en la sección siguiente.

Ejemplo 5.6 Omisión de variable relevante.
Sea el modelo correctamente especificado:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t \qquad t = 1, 2, \ldots, T$$

donde $v_t \sim iid(0, \sigma_v^2)$, y suponemos $E(X_{2t}v_t) = 0$ y $E(X_{3t}v_t) = 0$. Pero el modelo que se estima es el siguiente:

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t \qquad t = 1, 2, \ldots, T$$

tal que $u_t = \beta_3 X_{3t} + v_t$ y donde:

$$\begin{array}{l} E(X_{2t}u_t) = E[X_{2t}\,(\beta_3 X_{3t} + v_t)] = \beta_3 E(X_{2t}X_{3t}) + E(X_{2t}v_t) = \beta_3 E(X_{2t}X_{3t}) \\ E(u_t) = \beta_3 E(X_{3t}) + E(v_t) = \beta_3 E(X_{3t}) \end{array}$$

dado que X_{3t} es una variable relevante, $\beta_3 \neq 0$ luego:

$$\begin{array}{lll} E(X_{2t}u_t) \neq 0 & \text{si} & E(X_{2t}X_{3t}) \neq 0 \\ E(u_t) \neq 0 & \text{si} & E(X_{3t}) \neq 0 \end{array}$$

Por tanto, X y u son independientes y $E(u|X) = E(u) = 0$ si y solo si X_{2t} y X_{3t} son variables aleatorias independientes y $E(X_{3t}) = 0\ \forall t$. En el resto de los posibles supuestos no lo son. Por tanto, en el caso contemplado, donde se omite la variable X_{3t} correlada con X_{2t}, en el modelo especificado a estimar $E(X_{2t}u_t) \neq 0$ y el estimador MCO es inconsistente.

Por ejemplo, supongamos que estamos estudiando la función de salarios y que proponemos como variable explicativa el nivel de educación del individuo medido en años de educación:

$$WAGE_i = \beta_1 + \beta_2 EDUC_i + u_i$$

Si este es nuestro modelo a estimar es claro que estamos omitiendo factores que determinan el salario aparte del nivel educativo como la experiencia, la habilidad, la motivación. Podemos esperar que los individuos más motivados y/o con más talento tengan a su vez mayor formación medida en más años de educación y por lo tanto, sería lógico pensar que $E(EDUC_i\, u_i) \neq 0$, luego el estimador MCO de β_1 y β_2 sería sesgado e inconsistente.

En términos de la especificación del modelo es importante remarcar que el modelo debe estar correctamente especificado antes de proponer un método de estimación alternativo a MCO. En el ejemplo que nos ocupa medir la experiencia es sencillo, pero variables como la habilidad o la motivación no son sencillas de medir. En general y en el mejor de los casos seremos capaces de

[5]En el Anexo 5.2. Errores de medida en las variables, se han desarrollado con detalle las consecuencias en los estimadores MCO de los coeficientes de un modelo, de que el error de medida se produzca en la variable endógena y simultáneamente, en la variable endógena y en la exógena.

aproximar dichas variables con un error de medida, que como ya se ha apuntado anteriormente, y se desarrollará en esta sección, conllevan problemas de correlación entre regresor y perturbación y por tanto, la inconsistencia del estimador de MCO.

Ejemplo 5.7 Simultaneidad.
Sea el modelo formado por las siguientes dos ecuaciones:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad t = 1, 2, \ldots, T \tag{5.1}$$

$$X_t = \gamma_1 + \gamma_2 Y_t + v_t \qquad t = 1, 2, \ldots, T \tag{5.2}$$

tal que:

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right)$$

Es decir, u_t y v_t son v.a. normales. Estamos interesados en estimar β_1 y β_2 en (5.2), para ello queremos saber si X y u son independientes y/o incorreladas. Dado que $E(u_t) = 0 \quad \forall t$:

$$\begin{aligned} E(X_t u_t) &= E[(\gamma_1 + \gamma_2(\beta_1 + \beta_2 X_t + u_t) + v_t)u_t] = \\ &= \gamma_1 E(u_t) + \gamma_2\beta_1 E(u_t) + \gamma_2\beta_2 E(X_t u_t) + \gamma_2 E(u_t^2) + E(v_t u_t) = \\ &= \gamma_2\beta_2 E(X_t u_t) + \gamma_2 \sigma_u^2 + \sigma_{uv} \end{aligned}$$

resolviendo:

$$E(X_t u_t) = \frac{1}{1 - \gamma_2\beta_2}(\sigma_{uv} + \gamma_2 \sigma_u^2)$$

con lo que, $E(X_t u_t) \neq 0$ si $\gamma_2 \neq 0$ y/o $\sigma_{uv} \neq 0$. En este ejemplo el estimador MCO es inconsistente. Además, el Teorema de Mann y Wald no se satisface. Deberíamos buscar un estimador alternativo a MCO y que sí fuese consistente.

Ejemplo 5.8 Variable endógena retardada como regresor y perturbación autocorrelada.

Supongamos un modelo con dinámica en la parte sistemática porque incluye como regresor a la variable endógena retardada y con dinámica en la perturbación ya que esta sigue un proceso autorregresivo de primer orden.

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_t + u_t \qquad u_t = \rho u_{t-1} + \epsilon_t \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2)$$

En este caso $E(Y_{t-1}u_t) \neq 0$ luego $E(X'u) \neq 0$ y el estimador MCO no es consistente. Este caso se analizará en profundidad en el tema siguiente, Modelos Dinámicos.

En los ejemplos anteriores el estimador MCO es inconsistente ya que la matriz de regresores está correlada con la perturbación. Un estimador alternativo y consistente es el estimador de Variables Instrumentales.

5.3. Estimador de Variables Instrumentales

En modelos donde existe correlación entre regresores y error el estimador MCO es inconsistente y sesgado ya que $E(X'u) \neq 0$. El procedimiento para obtener estimadores consistentes en un modelo de este tipo es el Método de Variables Instrumentales, VI. El método de Variables Instrumentales se basa en encontrar un conjunto de K variables, $Z_{1t}, Z_{2t}, \ldots, Z_{Kt}$, que se llaman instrumentos o variables instrumentales tales que, satisfacen las siguientes condiciones:

1. Cada uno de los instrumentos está incorrelacionado con el término de error de la ecuación de interés, $E(Z_{jt}u_t) = 0 \quad \forall\, t, \forall\, j,\ j = 1, 2, \ldots, K$.
2. Los instrumentos están correlacionados con la variable para la cual hacen de instrumento, $E(Z_{jt}X_{it}) \neq 0$. En cuanto a esta correlación, debe existir, pero no puede ser muy importante pues en este caso, si lo fuera, $E(Z_{jt}u_t) \neq 0$ y Z_{jt} no serviría de instrumento.

Junto con las dos anteriores deben de cumplirse dos condiciones más:

3. $(Z'X)$ es una matriz no singular, es decir, es invertible.
4. $\text{plim}(\frac{1}{T}Z'Z) = Q_{ZZ}$ matriz finita.

El estimador de variables instrumentales, dada una matriz de instrumentos Z se define[6]:

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

Una forma intuitiva de pensar en cómo obtener este estimador podemos encontrarla a partir de las ecuaciones normales del estimador de MCO. Cualquiera de los modelos anteriores podemos escribirlos de la forma matricial habitual $Y = X\beta + u$. Premultiplicando el modelo por X' tenemos:

$$X'Y = X'X\beta + X'u$$

Si $E(X'u) = 0$ entonces $E(X'Y) = E(X'X)\beta + 0$ luego $E(X'Y) - E(X'X)\beta = 0$ que sería la expresión de las ecuaciones normales en términos poblacionales. En términos muestrales podemos escribir $X'Y - X'X\hat{\beta} = 0$ luego $X'(Y - X\hat{\beta}) = 0$ y $X'\hat{u} = 0$.

Si $\text{plim}\left(\frac{1}{T}X'u\right) = 0$ los estimadores así obtenidos son consistentes. Por tanto, parece que sugiere que cuando $\text{plim}\left(\frac{1}{T}X'u\right) \neq 0$ en vez de premultiplicar por X' lo hagamos por una matriz Z' que satisface $\text{plim}\left(\frac{1}{T}Z'u\right) = 0$ tal que, podemos definir el estimador consistente $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$ dada la matriz Z que cumple $E(Z'u) = 0$. Llamamos a Z matriz de instrumentos. De las condiciones citadas anteriormente notar que las condiciones 1 y 2 garantizan la consistencia del estimador, mientras que las condiciones 3 y 4 garantizan que el estimador β_{VI} definido como $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$ se pueda evaluar en la muestra.

En general en la matriz de regresores X sólo habrá unas variables que no satisfagan la condición $E(X_{it}u_t) = 0$ y son estas variables las que necesitan de variables instrumentales. Para el resto de variables, las no correladas con la perturbación, el mejor instrumento es la propia variable. Es

[6] En el Anexo 5.3. Estimador de Variables Instrumentales, aparece derivado formalmente este estimador de VI.

decir, las matrices Z y X tendrán en común aquellas columnas correspondientes a las variables incorrelacionadas con el término de error. Notar que necesariamente Z y X deben tener el mismo número de columnas ya que en otro caso $(Z'X)$ no sería cuadrada.

Antes de ver ejemplos de cómo buscar los instrumentos vamos a ver las propiedades del estimador.

5.3.1. Propiedades del estimador de Variables Instrumentales

- **Linealidad:** Dado que X es una matriz estocástica el estimador de VI no es lineal en u,

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

  En la expresión del estimador aparecen X y u ambas variables aleatorias, por tanto, $\hat{\beta}_{VI}$ es una función no lineal de X y u.

- **Insesgadez:** En general el estimador será sesgado ya que $E[(Z'X)^{-1}(Z'u)] \neq 0$. En la expresión $E[(Z'X)^{-1}(Z'u)]$ hemos de notar dos cosas. Por un lado, el requisito impuesto para obtener el estimador de Variables Instrumentales, $\hat{\beta}_{VI}$, es $\text{plim}\left(\frac{1}{T}Z'u\right) = 0$ y no la independencia entre Z y u. Por otro lado, aunque Z y u fuesen independientes la expresión anterior participa de X, matriz estocástica no independiente de u. Luego $E[(Z'X)^{-1}(Z'u)] \neq 0$ y

$$E(\hat{\beta}_{VI}) = \beta + E[(Z'X)^{-1}(Z'u)] \neq \beta$$

  Por tanto, en muestras finitas el estimador de Variables Instrumentales es un estimador no lineal y en general sesgado. Su distribución exacta es desconocida. En muestras grandes el estimador es consistente y tiene distribución asintótica conocida.

- **Consistencia del estimador de Variables Instrumentales y Distribución asintótica:** Vamos a demostrar la consistencia del estimador y a buscar su distribución asintótica utilizando el Teorema de Mann y Wald aplicado a Z y u. Sea X matriz de variables explicativas de orden $(T \times K)$ y Z una matriz de instrumentos de orden $(T \times K)$. Si se cumplen las siguientes condiciones:

  i) $u \sim (0, \sigma_u^2 I)$.
  ii) $E(Z'u) = 0$
  iii) $\text{plim}\left(\frac{1}{T}Z'Z\right) = Q_{ZZ}$ finita, simétrica y definida positiva.
  iv) $\text{plim}\left(\frac{1}{T}Z'X\right) = Q_{ZX}$ finita, no singular

  entonces se tiene por el Teorema de Mann y Wald los siguientes resultados:

  a) $\text{plim}\left(\frac{1}{T}Z'u\right) = 0$
  b) $\frac{1}{\sqrt{T}}Z'u \xrightarrow{d} N(0, \sigma_u^2 Q_{ZZ})$

  Utilizando el resultado a) podemos demostrar la consistencia del estimador $\hat{\beta}_{VI}$:

$$\text{plim}\,\hat{\beta}_{VI} = \beta + \text{plim}\left(\frac{1}{T}Z'X\right)^{-1} \text{plim}\left(\frac{1}{T}Z'u\right) = \beta + Q_{ZX}^{-1} \cdot 0 = \beta$$

El estimador de Variables Instrumentales es consistente y la condición importante es la ausencia de correlación entre los instrumentos y el término de error del modelo de interés.

Utilizando el resultado b) junto con el Teorema de Cramer obtenemos su distribución asintótica[7]:

$$\sqrt{T}(\hat{\beta}_{VI} - \beta) = \left(\frac{1}{T}Z'X\right)^{-1}\left(\frac{1}{\sqrt{T}}Z'u\right) \xrightarrow{d} N\left(0\,,\,\sigma_u^2 \cdot Q_{ZX}^{-1} \cdot Q_{ZZ} \cdot (Q_{ZX}^{-1})'\right)$$

El resultado anterior justifica que en muestras grandes se utilice como matriz de covarianzas asintótica del estimador de variables instrumentales a:

$$\begin{aligned}\widehat{V}(\hat{\beta}_{VI}) &= \frac{\hat{\sigma}_u^2}{T}\left(\frac{Z'X}{T}\right)^{-1}\frac{Z'Z}{T}\left[\left(\frac{Z'X}{T}\right)^{-1}\right]' = \\ &= \hat{\sigma}_{VI}^2\,(Z'X)^{-1}\,Z'Z\,((Z'X)^{-1})'\end{aligned}$$

donde se utilizan las matrices de momentos muestrales $\frac{Z'X}{T}$ y $\frac{Z'Z}{T}$ para aproximar sus límites respectivos Q_{ZX} y Q_{ZZ} y

$$\hat{\sigma}_{VI}^2 = \frac{(Y - X\hat{\beta}_{VI})'\,(Y - X\hat{\beta}_{VI})}{T - K}$$

es un estimador consistente de σ_u^2. Notar que el denominador de $\hat{\sigma}_{VI}^2$ es $T-K$, pero podríamos haber considerado T dado que el estimador es asintótico y por lo tanto, invariante a que el denominador sea T ó $T-K$.

5.3.2. Cómo buscar los instrumentos

Con respecto a los instrumentos sabemos que éstos tienen que cumplir las siguientes condiciones:

1. Estar incorrelacionados con la perturbación del modelo de interés, $E(Z_{jt}u_t) = 0$.

2. Estar correlacionados con la variable para la cuál hacen de instrumento, $E(Z_{jt}X_{it}) \neq 0$.

Además, la matriz $(Z'X)$ debe ser una matriz no singular con $rg(Z) = K < T$, es decir, $\exists (Z'X)^{-1}$. En la práctica existen dos situaciones diferentes en la búsqueda de instrumentos:

1. Que el número de instrumentos disponibles coincida con el número de variables que necesiten instrumento.

2. Que el número de instrumentos disponibles sea mayor que el número de variables que necesiten instrumento.

[7]Este resultado puede generalizarse al caso en el que la perturbación tiene autocorrelación, pero en ese caso la matriz de varianzas y covarianzas del estimador VI no es la derivada anteriormente.

Número de instrumentos disponibles igual al número de variables explicativas que lo necesitan

Supongamos que el número de instrumentos de que se dispone es igual al número de variables explicativas que necesitan instrumento. En este caso cada instrumento constituye una variable instrumental para su correspondiente variable explicativa correlada con la perturbación. Para el resto de variables, para las que no necesitan instrumento, su mejor instrumento es ella misma. Los instrumentos deben cumplir los requisitos mencionados anteriormente, no deben estar correlados con la perturbación del modelo de interés y deben de estar correlados con la variable explicativa para la que actúan de instrumento. Además, debe existir $(Z'X)^{-1}$. A continuación veremos un ejemplo teórico que nos servirá para mostrar el desarrollo matricial.

Ejemplo 5.9 Sea el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad u_t \sim iid(0, \sigma_u^2) \quad t = 1, 2, \ldots, T$$

tal que $\quad X_t = 0,7X_{t-1} + v_t \quad v_t \sim iid(0, \sigma_v^2)$
$\qquad E(u_t v_s) = 5 \quad \text{si} \quad t = s$
$\qquad E(u_t v_s) = 0 \quad \text{si} \quad t \neq s$

En este ejercicio la matriz $X = [1 \quad \mathbf{X}_t\,]$ es estocástica ya que el regresor X_t es un regresor estocástico[8]. La constante no crea problemas de correlación con u_t ya que es un regresor no estocástico, $E(1\, u_t) = E(u_t) = 0$. Sin embargo, el regresor estocástico X_t está correlado contemporáneamente con la perturbación:

$$E(X_t u_t) = E[(0,7X_{t-1} + v_t)\, u_t] = 0,7E(X_{t-1}u_t) + E(v_t u_t) = 0,7 \cdot 0 + 5 = 5$$

Luego $E(X'u) \neq 0$ y el estimador MCO será no lineal y sesgado. En muestras grandes además, será inconsistente. Deberíamos estimar los coeficientes β_1 y β_2 por el Método de Variables Instrumentales. Para ello buscamos un instrumento para X_t, el único regresor correlado con la perturbación. Podemos pensar en $Z_t = X_{t-1}$ dado que X_{t-1} está incorrelado con la perturbación, $E(X_{t-1}u_t) = 0$ y correlado con X_t ya que ésta se genera por un proceso autorregresivo, $E(X_t X_{t-j}) \neq 0 \quad \forall j > 0$. Además, con $Z_t = X_{t-1} \ \rightarrow rg(Z) = 2 = K$, luego $\exists (Z'X)^{-1}$.

Aplicamos para $Z_t = X_{t-1}$ el estimador de Variables Instrumentales[9]:

[8] En adelante la notación utilizada $X = [1 \quad \mathbf{X}_t\,]$ hace referencia a la siguiente matriz:

$$X = [1 \quad \mathbf{X}_t\,] = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_T \end{bmatrix}$$

[9] Notar en el orden de las matrices que para $Z_t = X_{t-1}$ perdemos una observación al definir el instrumento. Si el instrumento hubiera sido $Z_t = X_{t-2}$ perderíamos dos observaciones y en ese caso las matrices Y, X y Z serían:

$$\underset{((T-2)\times 1)}{Y} = \begin{bmatrix} Y_3 \\ Y_4 \\ \vdots \\ Y_T \end{bmatrix} \qquad \underset{((T-2)\times 2)}{X} = \begin{bmatrix} 1 & X_3 \\ 1 & X_4 \\ \vdots & \vdots \\ 1 & X_T \end{bmatrix} \qquad \underset{((T-2)\times 2)}{Z} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_{T-2} \end{bmatrix}$$

$$\underset{((T-1)\times 1)}{Y} = \begin{bmatrix} Y_2 \\ Y_3 \\ \vdots \\ Y_T \end{bmatrix} \qquad \underset{((T-1)\times 2)}{X} = \begin{bmatrix} 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_T \end{bmatrix} \qquad \underset{((T-1)\times 2)}{Z} = \begin{bmatrix} 1 & X_1 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_{T-1} \end{bmatrix}$$

El estimador $\hat{\beta}_{VI} = (Z'X)^{-1}(Z'Y)$ es el siguiente[10]:

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{VI} = \begin{bmatrix} (T-1) & \sum_2^T X_t \\ \sum_2^T X_{t-1} & \sum_2^T X_t X_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} \sum_2^T Y_t \\ \sum_2^T X_{t-1} Y_t \end{bmatrix}$$

Su matriz de varianzas y covarianzas estimada se define:

$$\widehat{V}(\hat{\beta}_{VI}) = \hat{\sigma}^2_{VI}(Z'X)^{-1}(Z'Z)((Z'X)^{-1})' =$$

$$\hat{\sigma}^2_{VI} \begin{bmatrix} (T-1) & \sum_2^T X_t \\ \sum_2^T X_{t-1} & \sum_2^T X_t X_{t-1} \end{bmatrix}^{-1} \begin{bmatrix} (T-1) & \sum_2^T X_{t-1} \\ \sum_2^T X_{t-1} & \sum_2^T X^2_{t-1} \end{bmatrix} \left[\begin{bmatrix} (T-1) & \sum_2^T X_t \\ \sum_2^T X_{t-1} & \sum_2^T X_t X_{t-1} \end{bmatrix}^{-1} \right]'$$

siendo:

$$\hat{\sigma}^2_{VI} = \frac{(Y - X\hat{\beta}_{VI})'(Y - X\hat{\beta}_{VI})}{T-K}$$

Ejemplo 5.10 En este ejemplo vamos a volver sobre la estimación de un modelo con variable exógena medida con error y nos va a permitir ilustrar cómo buscar instrumentos de manera adecuada:

Supongamos que queremos estimar el modelo $Y_t = \beta X_t^\star + u_t^\star \quad u_t^\star \sim iid(0, \sigma^2_{u^\star})$ siendo $X_t^\star$ una variable no estocástica e inobservable. Sin embargo, se dispone de la variable observable $X_t = X_t^\star + \epsilon_t$ que nos va a permitir aproximar a la variable inobservable. Además, en cuanto al error de medida, se sabe que $\epsilon_t \sim iid(0, \sigma^2_\epsilon)$ y $Cov(u_t^\star, \epsilon_t) = Cov(u_t^\star, \epsilon_s) = Cov(u_s^\star, \epsilon_t) = 0$. El modelo considerado para estimar en términos de la variable observable X_t es:

$$Y_t = \beta X_t + u_t \qquad t = 1, \ldots, T$$

donde $u_t = (u_t^\star - \beta\epsilon_t)$ y tal que $E(X_t u_t) = -\beta\sigma^2_\epsilon \neq 0$. El estimador de MCO es inconsistente. El estimador consistente de β lo podemos obtener utilizando el estimador de VI, $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$ tal que:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix}; \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{bmatrix}; \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_T \end{bmatrix};$$

$$\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y = \frac{\sum Z_t Y_t}{\sum Z_t X_t}$$

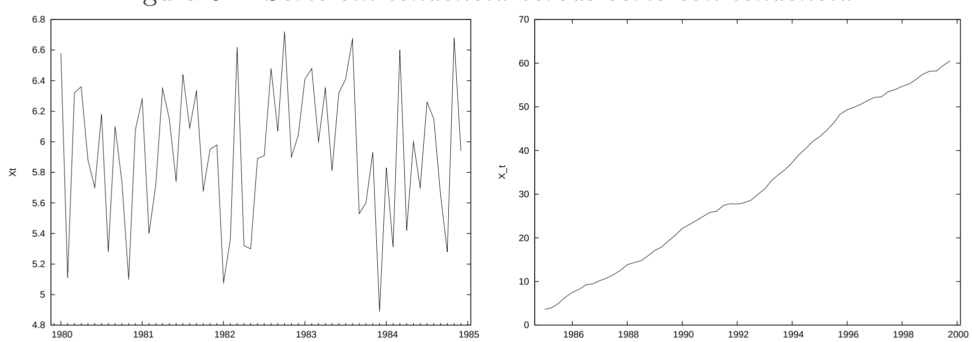
[10] A la hora de trabajar con el estimador de VI es importante notar que la matriz a invertir, $(Z'X)$, no es simétrica.

Es fundamental en el estimador de VI la elección del instrumento y ha de cumplir las condiciones explicitadas anteriormente. Vamos a proponer algunos posibles instrumentos y ver sus posibilidades como instrumentos adecuados:

- Supongamos que $Z_t = 1 \; \forall t$. En este caso $E(Z_t u_t) = 1E(u_t) = 0$ y $E(Z_t X_t) = E(X_t) = E(X_t^\star + \epsilon_t) = E(X_t^\star)$. Notar que X_t no tiene que estar en desviaciones a la media ya que en ese caso $\sum X_t = 0$ y $\hat{\beta}_{VI}$ no está definido.

- Otro posible instrumento sería $Z_t = t$, luego la variable instrumental es una tendencia determinista y por lo tanto, variable no estocástica incorrelada con la perturbación u_t, $E(Z_t u_t) = Z_t E(u_t) = tE(u_t) = 0$. Ahora bien, $Z_t = t$ tiene que estar correlacionada con X_t tal que $\sum Z_t X_t \neq 0$. Por lo tanto, X_t tiene que presentar cierta tendencia en el tiempo. Por ejemplo, en el gráfico de la izquierda de la Figura 5.1 la serie no muestra tendencia temporal luego el instrumento $Z_t = t$ no sería adecuado. Sin embargo, en el gráfico de la derecha la serie dibujada si muestra tendencia temporal, en este caso creciente, con lo cual el instrumento $Z_t = t$ tiene sentido.

Figura 5.1: *Serie sin tendencia versus serie con tendencia*



En el caso de $Z_t = t$ la matriz Z es:

$$Z = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ T \end{bmatrix}$$

- Otra variable instrumental puede ser otra variable medida con error, y distinta de X_t, que aproxime a $X_t^\star$, por ejemplo, $Z_t = X_t^\star + \omega_t$, tal que $\omega_t \sim iid(0, \sigma_\omega^2)$ y $E(\omega_t \, u_t^\star) = 0$. Además, los errores de medida tienen que estar incorrelados, $E(\epsilon_t \, \omega_t) = 0$, entonces,

$$E(Z_t u_t) = E((X_t^\star + \omega_t)(u_t^\star - \beta\epsilon_t)) = X_t^\star E(u_t^\star) + E(\omega_t \, u_t^\star) - \beta X_t^\star E(\epsilon_t) - \beta E(\omega_t \, \epsilon_t) = 0$$

En este caso la única dificultad es encontrar la proxy alternativa. Una posibilidad es que dos instituciones independientes midan la misma variable de forma que los errores de medida se mantengan independientes.

- Actuando como en el ejercicio anterior podemos pensar en $Z_t = X_{t-1}$. En este caso si $E(\epsilon_t\, u_s^\star) = 0$, $E(X_{t-1}u_t) = 0$ ya que no existe correlación no contemporánea:

$$\begin{aligned} E(X_{t-1}u_t) &= E((X_{t-1}^\star + \epsilon_{t-1})u_t) = E((X_{t-1}^\star + \epsilon_{t-1})(u_t^\star - \beta\epsilon_t)) = \\ &= E(X_{t-1}^\star\, u_t^\star) + E(\epsilon_{t-1}\, u_t^\star) - \beta E(X_{t-1}^\star\, \epsilon_t) - \beta E(\epsilon_t\, \epsilon_{t-1}) = 0 \end{aligned}$$

  y además, $E(X_{t-1}X_t) \neq 0$. Por tanto, también es un instrumento adecuado.

- Si generalizamos el modelo para incluir un término constante:

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1, 2, \ldots, T$$

  entonces:

$$\underset{(T\times 2)}{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_T \end{bmatrix} \qquad \underset{(T\times 2)}{Z} = \begin{bmatrix} Z_{11} & Z_{21} \\ Z_{12} & Z_{22} \\ \vdots & \vdots \\ Z_{1T} & Z_{2T} \end{bmatrix}$$

  Al menos necesitamos dos instrumentos. Para la constante 1 podemos usar $Z_{1t} = 1\ \forall\, t$ ya que $E(Z_{1t}u_t) = 1E(u_t) = 0$, es decir, el mejor instrumento es la misma variable. Para X_t hemos de buscar un instrumento Z_{2t} distinto de la constante 1 ya que si repetimos el instrumento $rg(Z) = 1 < 2$ y $\not\exists (Z'X)^{-1}$, con lo que el estimador $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$ no se puede obtener. No está definido. Además, tampoco sirven otras constantes como $Z_{2t} = 5$ por el mismo problema.

Ejemplo 5.11 Queremos estimar la relación:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t}^\star + u_t^\star \qquad u_t \sim iid(0, \sigma_{u^\star}^2) \qquad t = 1, 2, \ldots, T$$

donde X_{2t} es un regresor no estocástico y $X_{3t}^\star$ es una variable no estocástica e inobservable. Pero se dispone de una variable observable X_{3t} tal que $X_{3t} = X_{3t}^\star + \epsilon_t$ y

$$\epsilon_t \sim iid(0, \sigma_\epsilon^2) \qquad Cov(u_t^\star, \epsilon_t) = Cov(u_t^\star, \epsilon_s) = Cov(u_s^\star, \epsilon_t) = 0$$

luego el modelo a estimar en términos de las variables observables es:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1, 2, \ldots, T$$

con $u_t = u_t^\star - \beta_3\epsilon_t$, $E(X_{2t}u_t) = 0$ y $E(X_{3t}u_t) = -\beta_3\sigma_\epsilon^2 \neq 0$. En este contexto el estimador de MCO es inconsistente y para obtener consistencia los coeficientes β_1, β_2 y β_3 deberían ser estimados por Variables Instrumentales.

Supongamos que disponemos de otra proxy Z_t para aproximar a $X_{3t}^\star$ tal que $Z_t = X_{3t}^\star + \eta_t$ con $\eta_t \sim iid(0, \sigma_\eta^2)$, $Cov(u_t^\star, \eta_t) = Cov(u_t^\star, \eta_s) = 0$ y $E(\epsilon_t\eta_t) = 0$, es decir, los dos errores de medida están incorrelacionados.

Podemos usar esta proxy Z_t como instrumento para X_{3t} ya que:

$$\begin{aligned} E(Z_t u_t) &= E((X_{3t}^\star + \eta_t)(u_t^\star - \beta_3\epsilon_t)) = \\ &= E(X_{3t}^\star u_t^\star) + E(\eta_t u_t^\star) - \beta_3 E(X_{3t}^\star\epsilon_t) - \beta_3 E(\eta_t\epsilon_t) = 0 + 0 - \beta_3 \cdot 0 - \beta_3 \cdot 0 = 0 \end{aligned}$$

además, $E(Z_t X_{3t}) \neq 0$ ya que ambas son medidas de la variable $X_{3t}^{\star}$ no observable. Para la constante 1 y X_{2t} el mejor instrumento son ellas mismas. En este caso las matrices involucradas en el estimador de Variables Instrumentales son:

$$\underset{(T\times 1)}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} \qquad \underset{(T\times 3)}{X} = \begin{bmatrix} 1 & X_{21} & X_{31} \\ 1 & X_{22} & X_{32} \\ \vdots & \vdots & \vdots \\ 1 & X_{2T} & X_{3T} \end{bmatrix} \qquad \underset{(T\times 3)}{Z} = \begin{bmatrix} 1 & X_{21} & Z_1 \\ 1 & X_{22} & Z_2 \\ \vdots & \vdots & \vdots \\ 1 & X_{2T} & Z_T \end{bmatrix}$$

con $rg(Z) = 3$, luego $\exists (Z'X)^{-1}$ y el estimador $\hat{\beta}_{VI} = (Z'X)^{-1}(Z'Y)$ está bien definido.

En este caso no podemos contemplar dos posibilidades para definir el instrumento Z_t, este no puede ser ni la constante 1 ni el regresor X_{2t} ya que la matriz no sería de rango completo y $\nexists (Z'X)^{-1}$.

Número de instrumentos disponibles mayor al número de variables explicativas que lo necesitan. Método de Mínimos Cuadrados en dos Etapas

Generalmente, se dispondrá de un número de instrumentos mayor que de variables explicativas que lo necesiten, en este caso habría muchas formas de construir las variables instrumentales que precisamos para obtener un estimador consistente. Pero dado que la matriz de covarianzas del estimador de VI depende de los valores de éstas, el modo en que se combinan los instrumentos para generar variables instrumentales influye sobre la eficiencia del estimador de VI respecto a otro estimador de VI de su misma clase. De ahí que en ocasiones se hable de la eficiencia relativa de los estimadores de Variables Instrumentales. A continuación vamos a proponer algunos ejemplos:

Ejemplo 5.12 En el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \qquad t = 1, 2, \ldots, T$$

donde:

$$\begin{array}{ll} u_t \sim iid(0, \sigma_u^2) & \\ X_{3t}, X_{4t} \quad \text{no estocásticas} & \\ X_{2t} = 0,5 X_{2,t-1} + u_t & v_t \sim iid(0, \sigma_v^2) \\ E(u_t v_t) = 5 \quad \forall t = s & E(u_t v_s) = 0 \quad \forall t \neq s \end{array}$$

En este ejercicio la matriz de regresores $X = [1 \quad \mathbf{X}_{2t} \ \mathbf{X}_{3t} \ \mathbf{X}_{4t}]$ es estocástica ya que incluye al regresor estocástico X_{2t}. Necesitamos ver como es $E(X'u)$. Dado que X_{3t} y X_{4t} son no estocásticos $E(X_{3t}u_t) = 0$ y $E(X_{4t}u_t) = 0$, pero la variable explicativa X_{2t} y la perturbación están contemporáneamente correladas,

$$E(X_{2t}u_t) = E[(0,5X_{2,t-1} + v_t)\, u_t] = 0,5E(X_{2,t-1}u_t) + E(v_t u_t) = 0,5 \cdot 0 + 5 = 5$$

Por tanto, $E(X'u) \neq 0$ y el estimador MCO será no lineal y sesgado. En muestras grandes además, será inconsistente. Deberíamos estimar por el Método de Variables Instrumentales, pero sólo la variable X_{2t} necesita instrumento, para X_{3t} y X_{4t} el mejor instrumento es la propia variable. En este caso hay varios instrumentos que cumplen los requisitos. Los retardos de los dos regresores no estocásticos X_{3t} y X_{4t}, así como el retardo del regresor estocástico X_{2t} están incorrelados con la perturbación:

$$E(X_{2,t-1}\, u_t) = 0; \quad E(X_{3,t-1}\, u_t) = 0; \quad E(X_{4,t-1}\, u_t) = 0$$

Además, las variables $X_{2,t-1}$, $X_{3,t-1}$ y $X_{4,t-1}$ están correladas con X_{2t} por la propia relación del modelo. Por tanto, $X_{2,t-1}$, $X_{3,t-1}$ y $X_{4,t-1}$ serían buenos instrumentos. También lo serían combinaciones lineales de los mismos. La cuestión es que si utilizamos todos los posibles instrumentos, supongamos que su número es por ejemplo l, y los introducimos en $Z_{(T\times l)}$, la matriz de instrumentos tiene más columnas que $X_{(T\times K)}$. En esta situación en la que hay más hay más instrumentos que los que se necesitan, $l > K$, la matriz $(Z'X)$ no sería cuadrada y no existiría su inversa, $\nexists (Z'X)^{-1}$, con lo que no podemos definir el estimador de VI como $\hat{\beta}_{VI} = (Z'X)^{-1}(Z'Y)$.

Dado que en esta situación el número de instrumentos supera al de variables explicativas que lo precisan, se trata de buscar cuál de todos los posibles instrumentos minimiza la varianza del estimador resultante. Una posibilidad consiste en generar la variable instrumental con mayor correlación con X_{2t}. Para ello se estima por MCO una regresión auxiliar de esta variable sobre todos los posibles instrumentos. Para el ejemplo que nos ocupa, si suponemos que los únicos instrumentos de que disponemos son los primeros retardos de los regresores originales la regresión auxiliar sería[11]:

$$X_{2t} = \gamma_1 + \gamma_2 X_{2,t-1} + \gamma_3 X_{3,t-1} + \gamma_4 X_{4,t-1} + \gamma_5 X_{3t} + \gamma_6 X_{4t} + \eta_t \qquad t = 2, 3, \ldots, T$$

así $\widehat{X}_{2t} = \hat{\gamma}_1 + \hat{\gamma}_2 X_{2,t-1} + \hat{\gamma}_3 X_{3,t-1} + \hat{\gamma}_4 X_{4,t-1} + \hat{\gamma}_5 X_{3t} + \hat{\gamma}_6 X_{4t}$ es una combinación lineal de todos los posibles instrumentos que se incluyen en la matriz de instrumentos Z, cada uno ponderado por su correlación con X_{2t}, la variable a instrumentalizar, siendo Z:

$$Z = \begin{bmatrix} 1 & X_{21} & X_{31} & X_{41} & X_{32} & X_{42} \\ 1 & X_{23} & X_{32} & X_{42} & X_{33} & X_{43} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{2T-1} & X_{3T-1} & X_{4T-1} & X_{3T} & X_{4T} \end{bmatrix}$$

A continuación se estima el modelo por VI con:

$$\widehat{X} = [\,1 \quad \widehat{\mathbf{X}}_{2t} \quad \mathbf{X}_{3t} \quad \mathbf{X}_{4t}\,]$$

$$X = [\,1 \quad \mathbf{X}_{2t} \quad \mathbf{X}_{3t} \quad \mathbf{X}_{4t}\,]$$

Al estimador de VI así obtenido se le llama estimador de **Mínimos Cuadrados en dos Etapas**, MC2E, y se define[12]:

$$\hat{\beta}_{MC2E} = (\widehat{X}'X)^{-1}\widehat{X}'Y$$

donde

[11] Notar que en este ejemplo perdemos una observación al construir $\widehat{X}_{2t} \quad t = 2, 3, \ldots, T$.

[12] En el Anexo 5.4. Estimador de Mínimos Cuadrados en dos Etapas se desarrolla formalmente el estimador.

$$\widehat{X} = \begin{bmatrix} 1 & \widehat{X}_{22} & X_{32} & X_{42} \\ 1 & \widehat{X}_{23} & X_{33} & X_{43} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \widehat{X}_{2T} & X_{3T} & X_{4T} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_{22} & X_{32} & X_{42} \\ 1 & X_{23} & X_{33} & X_{43} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{2T} & X_{3T} & X_{4T} \end{bmatrix}$$

El estimador de MC2E es un estimador de VI que utiliza todos los instrumentos de forma óptima en el sentido de que es un estimador consistente y es el más eficiente asintóticamente dentro de la clase de estimadores de VI que toman solamente un subconjunto de todos los posibles instrumentos.

Ejemplo 5.13 Sea el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 W_{1t} + u_t \qquad t = 1, 2, \ldots, T$$

donde

$$X_t = \gamma_1 + \gamma_2 Y_t + \gamma_3 W_{2t} + \gamma_4 W_{3t} + v_t$$

$$u_t \sim NID(0, \sigma_u^2) \qquad v_t \sim NID(0, \sigma_v^2) \qquad Cov(u_t, v_t) = \sigma_{uv}$$

$$W_{1t}, W_{2t} \text{ y } W_{3t} \text{ son no estocásticas}$$

Dado que $\gamma_2 \neq 0$ y/o $\sigma_{uv} \neq 0$ existe correlación entre X_t y u_t, $E(X_t u_t) \neq 0$. El estimador adecuado para obtener al menos la propiedad de consistencia es el estimador de Variables Instrumentales. Necesitamos buscar un instrumento para la variable X_t. Tenemos dos posibles instrumentos disponibles W_{2t} y W_{3t}. Para combinarlos de forma óptima podemos realizar la siguiente regresión:

$$X_t = \alpha_0 + \alpha_1 W_{1t} + \alpha_2 W_{2t} + \alpha_3 W_{3t} + \eta_t \qquad t = 1, 2, \ldots, T$$

Así el instrumento adecuado para X_t es $\widehat{X}_t$ obtenido de la estimación por MCO de la regresión anterior tal que:

$$Z = \begin{bmatrix} 1 & W_{11} & W_{21} & W_{31} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & W_{1T} & W_{2T} & W_{3T} \end{bmatrix}$$

Utilizamos el instrumento para fijar la matriz de instrumentos $\widehat{X}$. Aplicamos el estimador de Variables Instrumentales para el cuál:

$$\widehat{X} = \begin{bmatrix} 1 & \widehat{X}_1 & W_{11} \\ 1 & \widehat{X}_2 & W_{12} \\ \vdots & \vdots & \vdots \\ 1 & \widehat{X}_T & W_{1T} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & X_1 & W_{11} \\ 1 & X_2 & W_{12} \\ \vdots & \vdots & \vdots \\ 1 & X_T & W_{1T} \end{bmatrix} \qquad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix}$$

Sobre el estimador de VI podemos hacer dos observaciones:

- Hay que notar que en el estimador de MC2E se regresan **todas** las variables explicativas sobre los posibles instrumentos. Para aquellas variables incorreladas con la perturbación el mejor instrumento son ellas mismas.

- La inclusión de retardos de las variables exógenas como instrumentos aumentaría el conjunto de información utilizado en la construcción del estimador de MC2E. Así, hay un estimador de MC2E para cada conjunto de instrumentos que se considere. Al utilizar más información, el estimador MC2E resultante sería más eficiente asintóticamente relativamente a otro estimador de VI que utilizase un subconjunto de los posibles instrumentos. Sin embargo, el uso de retardos obliga a prescindir de algunas observaciones muestrales, lo que disminuye el número de grados de libertad y puede influir en la eficiencia.

5.3.3. Contraste de hipótesis con el estimador de VI

Con el estimador de VI podemos utilizar el siguiente estadístico para hacer contrastes de restricciones lineales de la forma:

$H_0 : R\beta = r$
$H_a : R\beta \neq r$

$$(R\hat{\beta}_{VI} - r)'[R\,(\widehat{V}(\hat{\beta}_{VI})\,R']^{-1}(R\hat{\beta}_{VI} - r) \xrightarrow{d,H_0} \chi^2_q$$

donde q es el número de restricciones que se contrastan.

Podemos escribir el estadístico anterior como:

$$(R\hat{\beta}_{VI} - r)'[\,R\,\hat{\sigma}^2_{VI}(Z'X)^{-1}\,Z'Z\,((Z'X)^{-1})'\,R'\,]^{-1}(R\hat{\beta}_{VI} - r) \xrightarrow{d,H_0} \chi^2_q$$

Las reglas de aceptación y rechazo son las usuales. No rechazaremos la hipótesis nula si el valor muestral del estadístico es menor que el valor crítico $\chi^2_{q|\alpha}$ para un valor de significación elegido α.

Ejemplo 5.14 Para contrastar la significatividad individual de un regresor, $H_0 : \beta_i = 0$ versus $H_a : \beta_i \neq 0$ dado que $q = 1$ podemos utilizar:

$$\frac{\hat{\beta}_{i,VI}}{\widehat{des}(\hat{\beta}_{i,VI})} \xrightarrow{d,H_0} N(0,1)$$

Donde $\widehat{des}(\hat{\beta}_{i,VI})$ es la raíz cuadrada del elemento i-ésimo de la matriz $\hat{\sigma}^2(Z'X)^{-1}\,Z'Z\,((Z'X)^{-1})'$. Si el valor muestral del estadístico es mayor que $N(0,1)_{|\frac{\alpha}{2}}$ rechazamos la hipótesis nula para un nivel de significatividad α.

5.4. Contraste de Hausman

Necesitamos derivar un test de contraste que sea capaz de juzgar la incorrelación entre X y u. Este test es el contraste de Hausman. Sirve para contrastar si los regresores están o no correlacionados con la perturbación. Las hipótesis de contraste son:

$$\begin{array}{lcl} H_0\text{: } X \text{ y } u \text{ no están correlacionadas} & \Leftrightarrow & E(X'u) = 0 \\ H_a\text{: } X \text{ y } u \text{ están correlacionadas} & \Leftrightarrow & E(X'u) \neq 0 \end{array}$$

Consideremos el modelo de regresión lineal general en forma matricial:

$$Y = X\beta + u \Longleftrightarrow Y = \underset{(T\times K_1)}{X_1}\ \beta_1 + \underset{(T\times K_2)}{X_2}\ \beta_2 + u \qquad u \sim (0, \sigma^2 I)$$

donde K_1 variables explicativas en la matriz X no están correlacionadas con u, $E(X_1'u) = 0$, pero se cree que K_2 variables pueden estarlo[13]. El contraste de Hausman considera, en este contexto, contrastar:

$$\begin{array}{l} H_0 : E(X_2'u) = 0 \\ H_a : E(X_2'u) \neq 0 \end{array}$$

El estadístico de contraste considera la diferencia de dos estimadores de los coeficientes β, el estimador de MCO y un estimador de VI tal que:

$\hat{\beta}_{MCO}$ bajo H_0 es consistente y eficiente asintóticamente, pero inconsistente bajo H_a.
$\hat{\beta}_{VI}$ consistente bajo H_0 y H_a, pero menos eficiente asintóticamente que MCO bajo H_0.

El estadístico de contraste podemos escribirlo como:

$$(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})'\,[\widehat{V}(\hat{\beta}_{2,IV}) - \widehat{V}(\hat{\beta}_{2,MCO})]^{-1}\,(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO}) \xrightarrow{d,H_0} \chi^2_p$$

con las reglas de decisión habituales y $p = K_2$.

Supongamos que queremos contrastar una única restricción, $H_0 : E(X_{it}, u_t) = 0$ versus $H_a : E(X_{it}u_t) \neq 0$, y $p = K_2 = 1$. En este caso podemos escribir el estadístico de contraste como:

$$\frac{(\hat{\beta}_{i,VI} - \hat{\beta}_{i,MCO})^2}{\widehat{V}(\hat{\beta}_{i,IV}) - \widehat{V}(\hat{\beta}_{i,MCO})} \xrightarrow{d,H_0} \chi^2_1$$

Si el estadístico calculado es menor que el valor $\chi^2_{1|\alpha}$ no rechazamos $H_0 : E(X_{it}u_t) = 0$ para un nivel de significatividad α y concluiríamos que no existe correlación estadísticamente significativa entre el regresor X_{it} y u_t. En este caso si las perturbaciones del modelo son esféricas el estimador adecuado sería MCO. Este estimador a pesar de ser no lineal y sesgado sería consistente.

Si el estadístico calculado es mayor que el valor $\chi^2_{1|\alpha}$ rechazamos $H_0 : E(X_{it}u_t) = 0$ para un nivel de significatividad α, y concluimos que el regresor X_{it} y u_t están correlacionados. El estimador MCO es inconsistente. En este caso si las perturbaciones del modelo son esféricas el estimador adecuado sería VI, estimador no lineal y sesgado en general, pero consistente y asintóticamente eficiente si se buscan correctamente el instrumento o instrumentos necesarios.

Al computar el estadístico de contraste es necesario calcular $\widehat{V}(\hat{\beta}_{VI})$ y $\widehat{V}(\hat{\beta}_{MCO})$, ambas expresiones participan de un estimador consistente de σ^2. Suponiendo bajo H_0 que $u \sim (0, \sigma^2_u I)$, podemos utilizar:

$$\begin{array}{l} \widehat{V}(\hat{\beta}_{VI}) = \hat{\sigma}^2_{VI}(Z'X)^{-1}Z'Z\,((Z'X)^{-1})' \\ \widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2_{MCO}(X'X)^{-1} \end{array}$$

[13]Un caso particular sería que en X_1 se incluyera únicamente a la columna de unos y en X_2 el resto de variables explicativas.

con:

$$\hat{\sigma}^2_{VI} = \frac{(Y - X\hat{\beta}_{VI})'(Y - X\hat{\beta}_{VI})}{T - K}$$

$$\hat{\sigma}^2_{MCO} = \frac{(Y - X\hat{\beta}_{MCO})'(Y - X\hat{\beta}_{MCO})}{T - K}$$

Asintóticamente ambos coinciden y el test es asintótico. Sin embargo, se puede demostrar que la potencia del contraste aumenta si utilizamos como estimador consistente de σ^2_u al estimador de VI, $\hat{\sigma}^2_{VI}$. Por otro lado, y con respecto al denominador de ambos estimadores, dado que son asintóticos es indiferente que éste sea $T - K$ ó T.

Ejemplo 5.15 Se propone la siguiente especificación para la función de demanda de vino de un país:

$$Q_t = \beta_1 + \beta_2 P_t + u_t \qquad t = 1, 2, \ldots, T$$

donde $u_t \sim (0\,,\,0{,}0921)$. Dado que el precio se determina simultáneamente con la cantidad Q_t, se sospecha que P_t pueda estar correlacionada con u_t. Se dispone de datos de un índice de costes de almacenamiento, S_t que se determina exógenamente, por lo que se considera independiente de u_t. Dados los siguientes datos para los años de 1955-1975:

$$\begin{array}{ll}
\sum p_t q_t = \sum (P_t - \bar{P})(Q_t - \bar{Q}) = 1,78037 & \sum s_t^2 = \sum (S_t - \bar{S})^2 = 2,1417 \\
\sum p_t s_t = \sum (P_t - \bar{P})(S_t - \bar{S}) = 0,500484 & \sum p_t^2 = \sum (P_t - \bar{P})^2 = 0,507434 \\
\sum s_t q_t = \sum (S_t - \bar{S})(Q_t - \bar{Q}) = 2,75474 &
\end{array}$$

Se pide proponer un estimador consistente de los coeficientes β_1 y β_2 del modelo.

Para proponer un estimador consistente de los coeficientes de la relación es necesario saber si existe o no correlación entre P_t y u_t. Lo contrastamos:

$$H_0 : E(P_t u_t) = 0$$

$$H_a : E(P_t u_t) \neq 0$$

con

$$\frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})^2}{\widehat{V}(\hat{\beta}_{2,IV}) - \widehat{V}(\hat{\beta}_{2,MCO})} \xrightarrow{d, H_0} \chi^2_1$$

Calculamos los estimadores MCO y VI utilizando en este último como instrumento para P_t al el índice de coste de existencias de almacén S_t:

$$\hat{\beta}_{2,MCO} = \frac{\sum p_t q_t}{\sum p_t^2} = \frac{1,78037}{0,507434} = 3,5085$$

$$\hat{\beta}_{2,VI} = \frac{\sum s_t q_t}{\sum s_t p_t} = \frac{2,75474}{0,500484} = 5,4862$$

A continuación calculamos las correspondientes varianzas estimadas utilizando el valor $\sigma^2 = 0,09217$, conocido por el enunciado,

$$\widehat{V}(\hat{\beta}_{2,MCO}) = \frac{\sigma_u^2}{\sum p_t^2} = \frac{0,09217}{0,507434} = 0,18164$$

$$\widehat{V}(\hat{\beta}_{2,VI}) = \sigma_u^2 \frac{\sum s_t^2}{(\sum s_t p_t)^2} = \frac{0,0921 \cdot 2,1417}{(0,500484)^2} = 0,78747$$

y finalmente, computamos el estadístico:

$$\text{H} = \frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})^2}{\widehat{V}(\hat{\beta}_{2,VI}) - \widehat{V}(\hat{\beta}_{2,MCO})} = \frac{(5,4862 - 3,5085)^2}{0,78747 - 0,18164} = 6,5721$$

H $= 6,5721 > 3,841 = \chi^2_{1|0,05}$ por tanto, rechazo la hipótesis nula para $\alpha = 5\,\%$ y $E(P_t u_t) \neq 0$, P_t y u_t no son incorreladas.

Este ejercicio recoge un ejemplo práctico de la dificultad de estimar relaciones simultáneas. En un contexto donde la variable endógena y y el regresor se determinan simultáneamente como es éste, donde, lo que se especifica es la función de demanda del vino, la matriz de regresores es estocástica, como ya se apuntó en el Ejemplo 5.2. Es una de las situaciones en que existe correlación muestral significativa entre regresor y perturbación, como se ha detectado utilizando el estadístico de Hausman. El estimador MCO es inconsistente, los coeficientes del modelo deben ser estimados por VI, estimador consistente y asintóticamente eficiente. Además, existe un único instrumento, S_t y una única variable que lo necesita, P_t. La FRM podemos escribirla:

$$\widehat{Q}_t = \hat{\beta}_{1,VI} + 5,4862\, P_t$$

Por otra parte, el estimador de VI bajo perturbaciones esféricas permite hacer inferencia asintótica válida. Como podemos comprobar, la variable precio es significativa individualmente. Para ello contrastamos $H_0 : \beta_2 = 0$ versus $H_a : \beta_2 \neq 0$ con el estadístico y distribución asintótica:

$$\frac{\hat{\beta}_{2,VI}}{\widehat{des}(\hat{\beta}_{2,VI})} \xrightarrow{d,H_0} N(0,1)$$

Dado que el valor muestral del estadístico es t $= \frac{5,4862}{\sqrt{0,78747}} = 6,1823 > 1,96 = N(0,1)_{|0,025}$ rechazamos H_0 para un nivel de significatividad del 5 % y el precio es una variable individualmente significativa.

A continuación vamos a ver un ejemplo, utilizando el programa gretl, de cómo trabajar con el estimador de Variables Instrumentales. En el ejemplo se resume todo el tema. Las instrucciones de gretl necesarias para conseguir los resultados que se muestran las encontraréis en el Anexo 5.1: Instrucciones de gretl para el estimador de Variables Instrumentales. Estas instrucciones serán mostradas en la Práctica de Ordenador propuesta en este tema. También serán de aplicabilidad en el tema siguiente.

Ejemplo 5.16 En este ejemplo vamos a analizar los determinantes de la oferta laboral de las mujeres casadas[14]. Para ello vamos a utilizar el archivo *mroz87.gdt* incluido en el programa gretl, en la carpeta de archivos de muestra "Gretl". En él se dispone de observaciones de las siguientes variables, entre otras[15]:

- LFP: Variable ficticia que toma valor 1 si la mujer ha trabajado en 1975 y cero en otro caso
- $HOURS$: Número de horas trabajadas por la esposa en 1975, ($WHRS$).
- $KL6$: Número de hijos menores de seis años, en la familia.
- $K618$: Número de hijos entre seis y dieciocho años, en la familia.
- AGE: Edad de la esposa, (WA).
- $EDUC$: Años de educación recibidos por la esposa, (WE).
- $WAGE$: Salario de la esposa en el momento de la encuesta, 1976, ($RPWG$).
- $FAMINC$: Renta familiar en dólares de 1975.
- $EXPER$: Años de experiencia en la actualidad, (AX).

Se trata de una muestra de sección cruzada con 428 observaciones de mujeres trabajadoras donde el término trabajadoras implica que tienen un salario monetario[16].

Con la muestra anterior creamos las siguientes variables:

$$
\begin{aligned}
&l\_WAGE_i = ln(WAGE)_i,\\
&sq\_EXPER_i = EXPER_i^2,\\
&NWIFEINC_i = FAMINC_i - (WAGE_i \times HOURS_i)
\end{aligned}
$$

Dado nuestro objetivo de estudiar los determinantes de la oferta laboral de la población femenina casada, la variable a estudiar es $HOURS$. Como determinantes de la misma podemos pensar en incluir en el modelo el salario en logaritmos, l_WAGE, los años de educación recibidos, $EDUC$, la edad, AGE, el número de hijos de la familia, $KL6$ y $K618$, y la variable $NWIFEINC$, que de alguna manera mide la importancia de las rentas familiares que no dependen de los ingresos de la esposa. Así, el modelo a estimar sería:

$$
\begin{aligned}
HOURS_i = \beta_1 + \beta_2\, l\_WAGE_i + \\
\beta_3 EDUC_i + \beta_4 AGE_i + \beta_5 KL6_i + \beta_6 K618_i + \beta_7\, NWIFEINC_i + u_i \qquad (5.3)
\end{aligned}
$$

[14] Fichero mrox87.gdt, disponible en gretl pestaña Gretl. Fuente: "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions", *Econometrica* 55, 765-799.

[15] Algunas de las variables han sido renombradas para que el ejercicio resulte más cómodo. Entre paréntesis aparece el nombre original utilizado en el fichero para que las podáis reconocer fácilmente. A la hora de reproducir el ejercicio en vuestro trabajo personal es aconsejable hacer el cambio de nombre sugerido.

[16] Este archivo es en realidad una submuestra de otro archivo original con un total de 753 observaciones de las cuales solo trabajan 428 mujeres. El archivo original es *mroz.gdt*, está incluido en la carpeta Wooldridge de gretl. Sin embargo, como a lo largo del ejercicio de verá, para todos los individuos de la muestra todas las observaciones no están completas por lo que finalmente, el número utilizado de observaciones completas en el ejercicio es de 326.

A priori esperaríamos que los coeficientes de las variables l_WAGE, $EDUC$ y $K618$ fuesen positivos, ceteris paribus. Es de esperar que si se tiene un sueldo alto la estimulación a seguir trabajando sea mayor. También es lógico pensar que cuando la preparación (estudios) es mayor, se tenga un empleo mejor y mejor remunerado, luego es más probable que la mujer trabaje fuera de casa. Familias con mayor número de hijos necesitan de una mayor renta, por lo que es probable que la esposa decida trabajar. Por otro lado, cuando hay niños pequeños podemos pensar que la esposa se retire del mercado de trabajo para cuidarlos, por lo que esperaríamos un signo negativo en el coeficiente que acompaña a $KL6$ manteniendo el resto de factoes constante. A mayor edad, también es posible que no se quiera trabajar fuera de casa, por lo que, esperamos signo negativo para la variable AGE ceteris paribus. Finalmente, manteniendo el resto de factores constantes también esperaríamos signo negativo para el coeficiente de la variable $NWIFEINC$ ya que, si la renta disponible que no depende de la remuneración de la mujer es alta, es probable que eso desincentive a trabajar fuera de casa a la esposa. Es importante tener en cuenta la fecha de la recogida de datos, 1975, cuando todavía la participación en el mercado de trabajo de la mujer no estaba consolidada.

Los resultados de la estimación MCO de la ecuación (5.3) son:

Estimaciones MCO utilizando 326 observaciones desde 1–428
Se han quitado las observaciones ausentes o incompletas: 102
Variable dependiente: HOURS

	Coeficiente	Desv. típica	estadístico t	valor p
const	1909,20	344,1980	5,5468	0,0000
l_WAGE	311,7460	95,1025	3,2780	0,0012
EDUC	−18,6620	18,6519	−1,0006	0,3178
AGE	−8,2063	5,5617	−1,4755	0,1411
KL6	−296,62	107,6080	−2,7565	0,0062
K618	−80,759	31,8324	−2,5370	0,0117
NWIFEINC	−0,0098	0,0037	−2,6651	0,0081

Media de la var. dep.	1424,990	R^2	0,09628
D.T. de la var. dep.	687,015	$\bar{R}^2$ corregido	0,07928
SCR	1,38627e+08	$F(6, 319)$	5,66458
Desv. típica regresión ($\hat{\sigma}$)	659,217	valor p para $F()$	1,31042e-05

Lo primero que llama la atención es que las variables $EDUC$ y $K618$ no tienen los signos esperados. Por otro lado $EDUC$ y AGE no son significativas. ¿Es posible que nuestro modelo no esté correctamente estimado? ¿Hay razones para sospechar que l_WAGE esté correlada con la perturbación del modelo? $HOURS$ y l_WAGE se determinan conjuntamente en equilibrio entre la oferta y demanda. Luego efectos recogidos en u_i que afectan al nivel de horas trabajadas también afectaran al nivel de salarios de equilibrio. Se hace indispensable saber si el nivel de salarios l_WAGE_i está o no correlada con la perturbación. Es necesario contrastar:

$$H_0 : E(l\_WAGE_i \, u_i) = 0$$
$$H_a : E(l\_WAGE_i \, u_i) \neq 0$$

El contraste de Hausman implica estimar el modelo además de por MCO, por VI.

Vamos a estimar por VI utilizando como instrumento para la variable l_WAGE a la variable $EXPER$. En principio la variable $EXPER$ no tiene porque estar correlada con la perturbación del modelo, así $E(EXPER_i\ u_i) = 0$. Por otro lado, la correlación muestral entre estas dos variables es baja $\mathrm{r}_{l\_WAGE,EXPER} = 0,216$. Sin embargo, si regresamos l_WAGE sobre $EXPER$ incluyendo una constante la variable es significativa, el valor muestral estadístico es 3,985. Por lo que podemos concluir que el instrumento está significativamente correlado con la variable para la cual hace de instrumento, $E(l\_WAGE_i\ EXPER_i) \neq 0$. La variable cumple las características de una variable instrumental.

Los resultados de la estimación por VI utilizando como instrumento para l_WAGE a la variable $EXPER$ son los siguientes:

Estimaciones MC2E utilizando 326 observaciones desde 1–428
Se han quitado las observaciones ausentes o incompletas: 102
Variable dependiente: HOURS
Instrumented: l_WAGE
Instrumentos: const EXPER EDUC AGE KL6 K618 NWIFEINC

	Coeficiente	Desv. típica	z-stat	valor p
const	1334,86	527,6410	2,5299	0,0114
l_WAGE	2099,23	545,3280	3,8495	0,0001
EDUC	−162,5070	50,3543	−3,2273	0,0012
AGE	−9,4465	8,0821	−1,1688	0,2425
KL6	−372,3520	157,8050	−2,3596	0,0183
K618	−3,1880	51,5713	−0,0618	0,9507
NWIFEINC	−0,0124	0,0054	−2,2833	0,0224

Media de la var. dependiente	1424,988	Desv. tip. var. dependiente.	687,0155
Suma de cuadrados de los residuos	2,92e+08	Desv. tip. regresión	956,9807
R^2	0,050719	Adjusted R^2	0,032864
$F(6,319)$	4,307881	P-value(F)	0,000340

Si nos fijamos en el output que muestra gretl este nos indica cual es la variable que estamos instrumentalizando, l_WAGE. Si observamos la lista de instrumentos veremos que la matriz de instrumentos Z consta de siete columnas, las cuales coinciden con las columnas de la matriz de regresores X salvo la segunda donde en lugar de la variable l_WAGE aparecerá el instrumento $EXPER$. La referencia a MC2E de gretl se corresponde con el estimador de VI, ya que en este caso hay un instrumento y una única variable que lo necesita.

Ya podemos computar el contraste de Hausman. Recordemos que el estadístico de contraste y distribución son:

$$\frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})^2}{\widehat{V}(\hat{\beta}_{2,IV}) - \widehat{V}(\hat{\beta}_{2,MCO})} \xrightarrow{d,H_0} \chi^2_1$$

Aplicado a la muestra obtenemos:

$$H = \frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})^2}{\widehat{V}(\hat{\beta}_{2,IV}) - \widehat{V}(\hat{\beta}_{2,MCO})} = \frac{(2099,23 - 311,746)^2}{(545,328)^2 - (95,1025)^2} = \frac{3195099,05}{288338,1421} = 11,081$$

dado que el valor muestral del estadístico es mayor que $3,84 = \chi^2_{1|0,05}$ rechazamos la hipótesis nula y concluimos que $E(l\_WAGE_i \; u_i) \neq 0$. Luego el estimador apropiado es VI ya que MCO es inconsistente.

Los signos obtenidos de la estimación VI se mantienen con respecto a MCO. Con respecto a la significatividad de las variables en este caso la variable $EDUC$ es individualmente significativa, al contrario que con la estimación MCO. La no significatividad de AGE se mantiene, pero ahora obtenemos que la variable $K618$ no es significativa, cosa que no ocurría con el estimador MCO.

Podemos pensar en mejorar nuestro instrumento. ¿Que pasaría si utilizáramos como instrumento el cuadrado de la experiencia, sq_EXPER? La correlación muestral entre estas dos variables es baja $r_{l\_WAGE,sq\_EXPER} = 0,1853$, algo más baja que en el caso anterior. Si regresamos l_WAGE sobre sq_EXPER incluyendo una constante la variable es significativa, el valor muestral del estadístico es 3,395. Concluimos que el instrumento está significativamente correlado con la variable para la cual hace de instrumento, $E(l\_WAGE_i \; sq\_EXPER_i) \neq 0$. La variable es un instrumento adecuado.

Los resultados de la estimación por VI utilizando como instrumento para l_WAGE a la variable sq_EXPER son los siguientes:

Estimaciones MC2E utilizando 326 observaciones desde 1–428
Se han quitado las observaciones ausentes o incompletas: 102
Variable dependiente: HOURS
Instrumented: l_WAGE
Instrumentos: const sq_EXPER EDUC AGE KL6 K618 NWIFEINC

	Coeficiente	Desv. típica	z-stat	valor p
const	1356,310	524,6080	2,5854	0,0097
l_WAGE	2032,460	599,0830	3,3926	0,0007
EDUC	−157,1340	53,9486	−2,9127	0,0036
AGE	−9,4002	7,9271	−1,1858	0,2357
KL6	−369,5240	155,1580	−2,3816	0,0172
K618	−6,0855	51,9093	−0,1172	0,9067
NWIFEINC	−0,0123	0,0053	−2,3050	0,0212

Media de la var. dependiente	1424,988	Desv. tip. var. dependiente.	687,0155
Suma de cuadrados de los residuos	2,81e+08	Desv. tip. regresión	938,3666
R^2	0,051376	Adjusted R^2	0,033534
$F(6,319)$	3,830092	P-value(F)	0,001056

Los resultados de la estimación por VI utilizando como instrumento a sq_EXPER no varían con respecto a utilizar como instrumento a la variable $EXPER$. Ni en cuanto a los signos obtenidos, ni en cuanto a significatividad[17]. Sin embargo, hemos de notar que si ambas son instrumentos adecuados podemos ampliar el conjunto de instrumentos incluyendo a ambas variables y considerar el estimador de MC2E utilizando todos los instrumentos para mejorar en eficiencia asintótica.

Vamos a mostrar ahora los resultados de la estimación por MC2E utilizando ambas variables como instrumento junto con los regresores originales no correlados con la perturbación:

Estimaciones MC2E utilizando 326 observaciones desde 1–428
Se han quitado las observaciones ausentes o incompletas: 102
Variable dependiente: HOURS
Instrumented: l_WAGE
Instrumentos: const EDUC AGE KL6 K618 NWIFEINC EXPER sq_EXPER

	Coeficiente	Desv. típica	z-stat	valor p
const	1328,65	530,2510	2,5057	0,0122
l_WAGE	2118,57	544,7060	3,8894	0,0001
EDUC	−164,0630	50,3799	−3,2565	0,0011
AGE	−9,4599	8,1281	−1,1639	0,2445
KL6	−373,1720	158,6810	−2,3517	0,0187
K618	−2,3488	51,7912	−0,0454	0,9638
NWIFEINC	−0,0124	0,0054	−2,2757	0,0229

Media de la var. dependiente	1424,988	Desv. tip. de la var. dependiente.	687,0155
Suma de cuadrados de los residuos	2,95e+08	Desv. tip. regresión	962,4351
R^2	0,050536	Adjusted R^2	0,032678
$F(6, 319)$	4,338567	P-value(F)	0,000316

Si analizamos el output que muestra gretl la variable que estamos instrumentalizando coincide con los casos anteriores, pero ahora en la lista de instrumentos hay ocho variables. Se incluyen $EXPER$ y sq_EXPER y los regresores originales. En este caso, el término MC2E se corresponde totalmente con el estimador que hemos visto en la teoría. Hay más de un instrumento y una única variable que lo necesita, gretl ha computado el instrumento utilizando todas las variables indicadas para obtener el estimador de VI.

Los resultados de la estimación por MC2E no cambian frente a las anteriores regresiones de VI, ni en cuanto a los signos obtenidos, ni en cuanto a significatividad de las variables. Los coeficientes de determinación no deben ser utilizados como medida de bondad del ajuste ya que en este caso no tienen la interpretación habitual.

[17] Ya hemos determinado que l_WAGE es una variable correlada con la perturbación y no es necesario computar de nuevo el contraste de Hausman con el nuevo instrumento propuesto. Aún así, si lo calculamos, el estadístico de Hausman aplicado a la muestra es:

$$H = \frac{(\hat{\beta}_{2,VI} - \hat{\beta}_{2,MCO})^2}{\widehat{V}(\hat{\beta}_{2,IV}) - \widehat{V}(\hat{\beta}_{2,MCO})} = \frac{(2032,23 - 311,746)^2}{(599,083)^2 - (95,1025)^2} = \frac{2960065,194}{349855,9554} = 8,46$$

dado que el valor muestral del estadístico es mayor que $3,84 = \chi^2_{1|0,05}$ rechazamos la hipótesis nula y concluimos que $H_a : E(l\_WAGE_i\ u_i) \neq 0$. Luego el estimador apropiado es VI ya que MCO es inconsistente.

Ejercicio 5.3 Para afianzar la comprensión del ejemplo anterior.

1) Repetir todas las estimaciones y contrastes.

2) Calcular el estimador de MC2E paso a paso. Es decir:

   i) Estimar por MCO la regresión:

$$\begin{aligned} & l\_WAGE_i = \alpha_1 + \alpha_2 EXPER_i + \alpha_3 sq\_EXPER_i + \\ & \alpha_4 EDUC_i + \alpha_5 AGE_i + \alpha_6 KL6_i + \alpha_7 K618_i + \alpha_8 NWIFEINC_i + w_i \end{aligned}$$

   y guardar las estimaciones obtenidas, $l\_\widehat{WAGE}_i$.

   ii) Computar el estimador de VI utilizando como instrumento para l_WAGE_i a la variable $Z_i = l\_\widehat{WAGE}_i$.

   iii) Comparar las estimaciones obtenidas con las del ejemplo anterior.

3) Contrastar que las variables $KL6$ y $K618$ no son conjuntamente significativas. Razona la estimación elegida para realizar el contraste.

5.5. Anexo 5.1: Instrucciones básicas de gretl para regresores estocásticos

En este anexo se van a recoger las instrucciones básicas que incluye gretl para poder computar el estimador de VI con el fin de que podáis llevar a cabo la práctica de ordenador con cierta soltura. Instrucciones de carácter básico ya han sido recogidas en temas anteriores.

• **Para computar el estimador de Variables Instrumentales:**

Pulsar sucesivamente en

Modelo $\longrightarrow$ *Otros modelos lineales* $\longrightarrow$ *Mínimos cuadrados en dos etapas*

Seleccionar la variable endógena y las variables independientes de la manera habitual.

A continuación seleccionar los instrumentos.

Hay que tener en cuenta que le estamos indicando a gretl como es la matriz de instrumentos Z y que debe tener tantas columnas como la matriz X que recoge a las variables independientes. Luego debemos incluir como mínimo tantos instrumentos como regresores tenga el modelo. Para aquellas variables no estocásticas o que no estén correlacionadas con la perturbación el mejor instrumento son ellas mismas. Para aquellas variables correlacionadas con la perturbación se selecciona el instrumento-(s) adecuado-(s).

No es necesario respetar el mismo orden que al incluir las variables independientes.

En ocasiones los instrumentos a incluir son retardos de los regresores. En este caso en la ventana gretl: *especificar modelo* tenemos que elegir la variable dependiente, las variables explicativas en t o corrientes y los instrumentos que estarán en la lista en t.

A continuación se pulsa en retardos y aparece una ventana donde se tienen que seleccionar los retardos de las variables que aparecen en el modelo como regresores, tanto de variables exógenas como de la dependiente y también las variables que vamos a usar como instrumentos.

5.6. Anexo 5.2. Errores de medida en las variables

Hasta ahora hemos supuesto que las variables utilizadas en el proceso de estimación se medían sin error. En la práctica es muy posible que existan errores de medida en las variables o que simplemente las variables a utilizar no sean, sino estimaciones de conceptos teóricos que no se observan en la realidad, por ejemplo el stock de capital, el PIB, o las variables de Contabilidad Nacional. Estas situaciones alterarán las propiedades de los estimadores de los coeficientes en concreto introduciendo sesgos en las estimaciones y generando estimadores de MCO inconsistentes. A continuación vamos a repasar esas propiedades. Podemos distinguir tres situaciones:

1. Variable endógena medida con error.

2. Variable exógena medida con error. Analizada en el tema.

3. Variable exógena y endógena medidas con error.

5.6.1. Variable endógena medida con error

Sea el verdadero modelo:

$$Y_t^\star = \alpha + \beta X_t + u_t^\star \qquad u_t^\star \sim iid(0, \sigma^2_{u^\star}) \qquad t = 1, 2, \ldots, T$$

donde X_t es no estocástica. Por alguna razón la variable endógena disponible no es $Y_t^\star$, sino $Y_t = Y_t^\star + \epsilon_t$ con $\epsilon_t \sim iid(0, \sigma^2_\epsilon)$, $Cov(X_t, \epsilon_t) = 0$, $Cov(u_t^\star, \epsilon_t) = 0$.

El modelo estimable en términos de las variables observables es:

$$\begin{aligned} Y_t - \epsilon_t = \alpha + \beta X_t + u_t^\star \ &\Rightarrow Y_t = \alpha + \beta X_t + (u_t^\star + \epsilon_t) \\ &\Rightarrow Y_t = \alpha + \beta X_t + u_t \end{aligned}$$

El error de medida en la variable endógena se acumula en la perturbación original. Las propiedades de la nueva perturbación $u_t^\star$ son:

$$\begin{array}{ll} E(u_t) & = E(u_t^\star + \epsilon_t) = E(u_t^\star) + E(\epsilon_t) = 0 \\ Var(u_t) & = E(u_t - E(u_t))^2 = E(u_t)^2 = E(u_t^\star + \epsilon_t)^2 \\ & = E(u_t^{\star 2}) + E(\epsilon_t^2) + 2E(u_t^\star \epsilon_t) = \sigma^2_{u^\star} + \sigma^2_\epsilon \qquad \text{homocedástica} \\ Cov(u_t, u_s) & = E[(u_t - E(u_t))(u_s - E(u_s))] = \\ & = E(u_t u_s) = E(u_t^\star + \epsilon_t, u_s^\star + \epsilon_s) = 0 \qquad \text{no autocorrelada} \end{array}$$

$$u_t \sim iid(0, \sigma^2_{u^\star} + \sigma^2_\epsilon)$$

En este caso, en presencia de errores de medida en la variable endógena y cuando el error de medida es incorrelado con la perturbación del modelo original el error de medida se traslada a la perturbación del modelo estimable, incrementándola, pero se mantienen sus propiedades esféricas. Por tanto, los MCO son apropiados y tienen buenas propiedades.

5.6.2. Variable exógena y variable endógena medidas con error

Sea el verdadero modelo:

$$Y_t^\star = \alpha + \beta X_t^\star + u_t^\star \qquad t = 1, 2, \ldots, T$$

$$\begin{array}{ll} \text{donde :} & Y_t = Y_t^\star + \epsilon_t \quad \text{es la variable endógena disponible.} \\ & X_t = X_t^\star + v_t \quad \text{es la variable exógena disponible, y } X_t^\star \text{ es no estocástica} \\ & u_t^\star \sim iid(0, \sigma^2_{u^\star}) \quad v_t \sim iid(0, \sigma^2_v) \quad \epsilon_t \sim N(0, \sigma^2_\epsilon) \\ & Cov(u_t^\star, \epsilon_t) = Cov(u_t^\star, v_t) = Cov(\epsilon_t, v_t) = 0 \end{array}$$

El modelo estimable en términos de las variables observables es:

$$Y_t - \epsilon_t = \alpha + \beta(X_t - v_t) + u_t^\star \qquad t = 1, 2, \ldots, T$$

$$Y_t = \alpha + \beta X_t + (u_t^\star + \epsilon_t - \beta v_t) \qquad t = 1, 2, \ldots, T$$

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1, 2, \ldots, T$$

La perturbación del modelo estimable es homocedástica y no autocorrelada

$$u_t \sim iid(0, \sigma^2_{u^\star} + \sigma^2_\epsilon + \beta^2 \sigma^2_v)$$

Existe correlación contemporánea ya que:

$$\begin{aligned} E(X_t u_t) &= E(X_t^\star + v_t)(u_t^\star + \epsilon_t - \beta v_t) = \\ &= E(X_t^\star u_t^\star) + E(v_t u_t^\star) + E(X_t^\star \epsilon_t) + E(v_t \epsilon_t) - \beta E(X_t^\star v_t) - \beta E(v_t^2) = \\ &= -\beta \sigma^2_v \end{aligned}$$

No existe correlación no contemporánea ya que $E(X_s u_t) = 0$

En este caso existen dos fuentes de error, el error de medida en la variable endógena y el error de medida en la variable exógena. El error de medida en la variable endógena implica un incremento en la varianza de la perturbación del modelo estimable mientras que el error de medida en la variable exógena implica que los estimadores MCO de α y β serán sesgados e inconsistentes. El modelo estimable, bajo todos los supuestos realizados anteriormente, debería ser estimado por el Método de Variables Instrumentales.

5.7. Anexo 5.3. Estimador de Variables Instrumentales

A continuación se va a mostrar, a partir de la función objetivo, cómo obtener el estimador de Variables Instrumentales. Demostración:

$$\begin{aligned} Y &= X\beta + u \\ Z'Y &= Z'X\beta + Z'u \end{aligned}$$

Denotamos por $u_\star = Z'u$

$$\begin{aligned} u_\star &= Z'u = Z'(Y - X\beta) \\ \hat{u}_\star &= Z'\hat{u} = Z'(Y - X\hat{\beta}) \\ \hat{u}'_\star \hat{u}_\star &= \left[Z'(Y - X\hat{\beta})\right]' \left[Z'(Y - X\hat{\beta})\right] = (Y - X\hat{\beta})' ZZ'(Y - X\hat{\beta}) \end{aligned}$$

La función objetivo podemos escribirla como:

$$\underset{\hat{\beta}}{Min}\ \hat{u}'_\star \hat{u}_\star = \underset{\hat{\beta}}{Min}\ \left[Y'ZZ'Y - 2\hat{\beta}'X'ZZ'Y + \hat{\beta}'X'ZZ'X\hat{\beta}\right]$$

por las condiciones de primer orden obtenemos:

$$\begin{aligned} \frac{\partial(\hat{u}'_\star \hat{u}_\star)}{\partial\hat{\beta}} &= -2X'ZZ'Y + 2X'ZZ'X\hat{\beta} = 0 \\ &= X'ZZ'(Y - X\hat{\beta}) = 0 \end{aligned}$$

De donde las ecuaciones normales del modelo serían:

$$X'ZZ'X\hat{\beta} = X'ZZ'Y$$

tal que:

$$\begin{aligned} \hat{\beta}_{VI} &= (X'ZZ'X)^{-1}(X'ZZ'Y) \\ \hat{\beta}_{VI} &= (Z'X)^{-1}(X'Z)^{-1}(X'Z)(Z'Y) \\ \hat{\beta}_{VI} &= (Z'X)^{-1}(Z'Y) \end{aligned}$$

5.8. Anexo 5.4. Estimador de Mínimos Cuadrados en dos etapas

A continuación vamos a mostrar como derivar el estimador de MC2E. Sea el modelo en forma matricial:

$$\underset{(T\times 1)}{Y} = \underset{(T\times K)}{X}\ \underset{(K\times 1)}{\beta} + \underset{(T\times 1)}{u} \iff \underset{(T\times 1)}{Y} = \underset{(T\times K_1)}{X_1}\ \underset{(K_1\times 1)}{\beta_1} + \underset{(T\times K_2)}{X_2}\ \underset{(K_2\times 1)}{\beta_2} + \underset{(T\times 1)}{u} \qquad u \sim (0, \sigma^2 I)$$

La matriz $X = [\,\mathbf{X}_1 \vdots \mathbf{X}_2\,]$ es estocástica y tal que $E(\mathbf{X}_1 u) = 0$ y $E(\mathbf{X}_2 u) \neq 0$. Para la matriz de instrumentos Z necesitaremos al menos K_2 instrumentos para las variables X_2 además, de poder utilizar a X_1 como instrumentos para ellas mismas.

$$\underset{(T\times l)}{Z} = [\mathbf{X}_1 \vdots \mathbf{X}_2] \quad \text{donde} \quad l \geq K$$

- Si $l = K$ entonces tenemos exactamente el mismo número de instrumentos que necesitamos. $(Z'X)$ es una matriz cuadrada de orden $(K \times K)$ y no singular tal que existe $(Z'X)^{-1}$ y podemos definir el estimador de Variables Instrumentales de la forma $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$.
- Si $l > K$ entonces tenemos más instrumentos que los que necesitamos $(Z'X)$ es de orden $(l \times K)$ y no es cuadrada.

El proceso de estimación que comprende el estimador de Mínimos Cuadrados en dos Etapas es el siguiente:

En la primera etapa se lleva a cabo la regresión de cada una de las columnas de la matriz X sobre Z para obtener una matriz de valores ajustados $\widehat{X}$.

$$\underset{(T\times K)}{\widehat{X}} = \underset{(T\times l)}{Z}\ \underbrace{\underset{(l\times l)}{(Z'Z)^{-1}}\ \underset{(l\times T)}{Z'}\ \underset{(T\times K)}{X}}_{\text{matriz de coeficientes de las K-regresiones}} = P_z X \quad \text{donde} \quad P_z = Z(Z'Z)^{-1}Z'$$

P_z es idempotente y simétrica, $P_z' = P_z$ y $P_z P_z = P_z$.

En la segunda etapa se realiza la regresión de Y sobre $\widehat{X}$ para obtener el estimador de los coeficientes β por Mínimos Cuadrados en 2 Etapas.

$$\begin{aligned}
\hat{\beta}_{MC2E} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'Y = \\
&= (X' \underbrace{P_z P_z}_{P_z} X)^{-1} X' P_z Y = (X' P_z X)^{-1} X' P_z Y = \\
&= (\widehat{X}'X)^{-1}\widehat{X}'Y = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y
\end{aligned}$$

luego el estimador de MC2E es un estimador de Variables Instrumentales que utiliza como matriz de instrumentos a $\widehat{X} = P_z X$ donde combina todos los instrumentos de forma óptima en el sentido de que es un estimador consistente y es el más eficiente asintóticamente dentro de la clase de estimadores de VI que toman solamente un subconjunto de todos los posibles instrumentos Z.

Hay que notar que, si $l = K$, entonces $(X'Z)$ y $(Z'X)$ son cuadradas y $(X'Z(Z'Z)^{-1}Z'X)^{-1} = (Z'X)^{-1}Z'Z(X'Z)^{-1}$ por lo que en ese caso:

$$\hat{\beta}_{MC2E} = (Z'X)^{-1}Z'Z(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'Y = (Z'X)^{-1}Z'Y$$

Para estimar la matriz de varianzas y covarianzas asintótica de $\hat{\beta}_{MC2E}$ se utiliza el estimador consistente:

$$\widehat{V}(\hat{\beta}_{MC2E}) = \frac{(Y - X\hat{\beta}_{MC2E})'(Y - X\hat{\beta}_{MC2E})}{T}(\widehat{X}'\widehat{X})^{-1}$$