CAPITULO 9
MODELOS DINAMICOS

9.1. INTRODUCCION

En los capítulos anteriores hemos venido suponiendo que las variables que aparecen como explicativas en el modelo econométrico estaban contemporáneamente relacionadas con la variable endógena, por lo que sus índices temporales eran iguales. Sin embargo, la Teoría Económica sugiere que, en la mayoría de los casos, las relaciones entre variables son dinámicas. Ello puede ocurrir porque: *a*) el impacto de una variable sobre otra sólo se deja notar tras un cierto tiempo; *b*) el impacto, aunque sea instantáneo, se deja notar durante un cierto número de períodos, o *c*) ambas cosas a la vez.

Ello conduce, respectivamente, a considerar modelos como los siguientes:

$$y_t = \beta_1 + \beta_2 x_{t-1} + u_t$$
$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \beta_4 x_{t-2} + u_t$$
$$y_t = \beta_1 + \beta_2 x_{t-2} + \beta_3 x_{t-3} + u_t$$

En el primer modelo, un cambio en el valor de x_t no se deja sentir sobre la variable endógena hasta el período siguiente; en el segundo modelo, un cambio en x_t tiene un efecto inmediato sobre y_t, medido por el coeficiente β_2. pero también influirá sobre y_{t+1} e y_{t+2}, con efectos dados por los valores de β_2 y β_3. En el tercer modelo, un cambio en x_t no influye sobre el valor de y_t ni de y_{t+1}, pero sí sobre y_{t+2} e y_{t+3}.

Los ejemplos acerca de este tipo de relaciones son numerosos: *a*) la inversión productiva genera incrementos en el nivel de producción que. generalmente, se extienden a varios años y no comienzan de inmediato; *b*) el crecimiento de la oferta monetaria tiene efectos sobre los precios que, aparte de ser persistentes, se manifiestan sólo tras un retardo inicial.

Por otra parte, las variables económicas tienen bastante inercia, lo que hace que una variable dependa de su propio pasado, además de otras causas.

Así, por ejemplo, para tratar de explicar el comportamiento de la inflación π_t, tendría sentido introducir como variables explicativas, junto con las tasas de crecimiento monetario m_t, retardos de la propia tasa de inflación:

$$\pi_t = \beta_0 + \beta_1 \pi_{t-1} + \beta_2 m_t + u_t$$

Es importante observar que la existencia de una relación dinámica entre variables, así como su mayor o menor persistencia (es decir, el número de retardos precisos para representarla), dependen crucialmente de cual sea la frecuencia de observación de los datos que se emplean en la estimación. Así, si una variable x_t influye sobre otra y_t no sólo contemporáneamente, sino también durante los dos meses siguientes, entonces la relación sería dinámica si el investigador utiliza datos mensuales, pero resultará estática si utilizase datos anuales. En este último caso, tan sólo el valor contemporáneo de x_t aparecería como variable explicativa en el modelo de regresión.

9.1.a. Primeras propiedades

Comencemos examinando las distintas propiedades dinámicas que revisten estos modelos. Consideremos a modo de ejemplo la relación:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \beta_4 x_{t-2}$$

que, por simplificar, consideramos determinista, y supongamos que la economía está en equilibrio, es decir, que las observaciones de las variables x_t e y_t vienen siendo iguales en los últimos períodos a sus valores de equilibrio, x^* e y^*; de acuerdo con el modelo, debe cumplirse que $y^* = \beta_1 + (\beta_2 + \beta_3 + \beta_4)x^*$.

Supongamos que, por alguna razón, en el instante t_0 la variable x_t se desvía de dicho equilibrio en una cantidad Δ, es decir: $x_{t0} = x^* + \Delta$, pero vuelve al equilibrio a partir de entonces $x_t = x^*$ para todo $t > t_0$. Es fácil ver que $y_{t0} = y^* + \beta_2\Delta$, pero no todo acaba aquí; también tendremos $y_{t0+1} = \beta_1 + \beta_2 x_{t0+1} + \beta_3 x_{t0} + \beta_4 x_{t0-1} = \beta_1 + \beta_2 x^* + \beta_3(x^* + \Delta) + \beta_4 x^* = \beta_1 + (\beta_2 + \beta_3 + \beta_4)x^* + \beta_3\Delta = y^* + \beta_3\Delta$. El lector puede comprobar que, de modo análogo, se tiene $y_{t0+2} = y^* + \beta_4\Delta$, así como $y_t = y^*$ para todo $t > t_0 + 2$. *La función de respuesta al impulso* es la sucesión de estos efectos que, en el caso de este modelo particular, resultan ser β_2, β_3, β_4, 0, 0, ..., y resume el efecto que sobre y_t tiene una desviación puramente *transitoria* de x_t respecto de su valor de equilibrio inicial. Este efecto consiste en una desviación de y_t respecto de su equilibrio inicial durante tres períodos, para terminar volviendo al mismo a partir de entonces.

La función de respuesta al escalón es un estadístico que responde a una pregunta similar: ¿qué efecto tendría sobre y_t una desviación *permanente* de x_t respecto de su equilibrio inicial, es decir, si pasa a un nuevo valor de equilibrio? Denotando dicho valor de equilibrio por $x^* + \Delta$, supongamos que $x_t = x^*$ si $t < t_0$, $x_{t0} = x^* + \Delta$, y $x_t = x^* + \Delta$ para $t > t_0$. Un razonamiento similar al anterior muestra que, en nuestro ejemplo, la función de respuesta al escalón es β_1, $\beta_1 + \beta_2$, $\beta_1 + \beta_2 + \beta_3$ y permanece constante en este valor a partir de tres períodos. En este caso, al pasar x_t a su nuevo valor de equilibrio, y_t pasa también a un nuevo valor de equilibrio: $y^* + (\beta_1 + \beta_2 + \beta_3)\Delta$, en el que permanece a partir de entonces. *La función de respuesta al escalón no es sino la sucesión de valores acumulados de la función de respuesta al impulso.*

Si normalizamos los coeficientes $\beta_i' = \beta_i/\Sigma_j\beta_j$, cada uno de los nuevos coeficientes β_i' nos da la proporción del efecto total que se deja sentir j períodos después del cambio en x_t. De este modo, puede responderse a preguntas como: ¿Cuántos períodos deben transcurrir antes de que se deje notar el 90 por 100 del efecto que el cambio en x_t tiene sobre y_t? En particular, el *retardo mediano* es aquel período en que ya se ha dejado sentir sobre y_t el 50 por 100 del efecto producido por la desviación en x_t.

El *retardo medio* es el cociente $\Sigma_i i\beta_i/\Sigma_j\beta_j$. Un valor bajo del retardo medio significa que el ajuste se lleva a cabo rápidamente.

Utilizando el operador de retardos, $Lx_t = x_{t-1}$, podemos escribir un modelo general como:

$$y_t - \alpha_1 y_{t-1} - \cdots - \alpha_p y_{t-p} = \delta + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_s x_{t-s} + u_t$$

en la forma $A(L)y_t = \delta + B(L)x + u_t$, donde $A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p$ y $B(L) = \beta_0 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_s L^s$, y la *ganancia* se define como el cociente $B(1)/A(1)$, donde se ha sustituido el operador L por la unidad en los polinomios $A(L)$ y $B(L)$.

Todos estos conceptos son diferentes cuando hay retardos de la variable endógena. Entonces hay que distinguir entre respuestas a corto y a largo plazo.

Por ejemplo, en el modelo

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 x_t + u_t$$

la respuesta a corto plazo de y_t a un impulso en x_t es simplemente β_2, pero la aparición del retardo de la variable endógena hace dicho modelo equivalente a $(1 - \beta_1 L)y_t = \alpha + \beta_2 x_t + u_t$ que, bajo la *condición de estacionariedad* $-1 < \beta_1 < 1$, puede escribirse:

$$y_t = \frac{\alpha}{1-\beta_1} + \Sigma_0^\infty \beta_2 (\beta_1)^s x_{t-s} + \Sigma_0^\infty (\beta_1)^s u_{t-s} = \frac{\alpha}{1-\beta_1} +$$
$$+ (\beta_2 x_t + \beta_2\beta_1 x_{t-1} + \beta_2\beta_1^2 x_{t-2} + \cdots) + \Sigma_0^\infty (\beta_1)^s u_{t-s}$$

y, prescindiendo del componente aleatorio, nuestro análisis anterior conduce a una función de respuesta al impulso: β_2, $\beta_2\beta_1$, $\beta_2\beta_1^2$, $\beta_2\beta_1^3$, ..., mientras que la función de respuesta al escalón acumula ésta. Como puede verse, dicha respuesta es convergente tan sólo si el parámetro β_2 tiene valor absoluto inferior a la unidad. Esta es *la condición de estacionariedad del modelo, que debe manifestarse en que las funciones de respuesta al impulso tienden a cero, y las de respuesta al escalón se estabilizan.*

El tratamiento estadístico de estos modelos difiere, según que los valores retardados que aparecen como variables explicativas pertenezcan a variables exógenas, o que entre ellos haya algún retardo de la variable endógena.

9.1.b. Cuando todos los retardos corresponden a variables exógenas

Si el modelo es del tipo

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{2t-1} + \cdots + \beta_s x_{2t-s} + u_t \qquad [9.1]$$

no se incumplen las hipótesis básicas del modelo lineal general, por cuanto las distintas variables explicativas del modelo de regresión son todas deterministas. En este modelo aparecen tan sólo dos *posibles dificultades:*

1. Los retardos consecutivos de una variable económica tienden a estar correlacionados entre sí, tanto más cuanto mayor sea la estructura de autocorrelación de dicha variable. Así, por ejemplo, si x_t sigue una estructura autorregresiva de primer orden con coeficiente 0,95:

$$x_t = 0{,}95x_{t-1} + \varepsilon_t$$

la correlación entre x_t y x_{t-s} será muy elevada, incluso para valores altos de s, siendo dicha correlación muy inferior si el coeficiente del modelo de x_{t-1} fuese igual a 0,40.

Cuanto mayor sea la correlación entre los retardos de x_t, más importante será la presencia de *multicolinealidad* en el modelo de regresión, cuyas implicaciones analizaremos en el Capítulo 10.

2. La segunda dificultad surge cuando $s = \infty$, es decir, cuando la estructura de retardos es de orden infinito. Entonces es imposible estimar directamente el modelo [9.1], por cuanto que no tendríamos observaciones suficientes para ello. Como veremos en la Sección 9.3, para estimar este modelo es imprescindible imponer a priori algún tipo de restricción entre los coeficientes, de modo que el modelo pueda transformarse en otro con un número reducido de variables explicativas.

9.1.c. Si aparecen valores retardados de la variable endógena

como variables explicativas, entonces dejaría de cumplirse uno de los supuestos bajo los que desarrollamos las teorías de estimación e inferencia del modelo econométrico, pues algunas de las variables explicativas serían ahora variables aleatorias (ya que y_t lo es). Sin embargo, si el término de error no tiene autocorrelación, el problema de estimación no es muy importante.

Consideremos, por ejemplo, el modelo:

$$y_t = \beta y_{t-1} + u_t, \qquad |\beta| < 1 \tag{9.2}$$

donde u_t es un proceso de ruido blanco. El estimador de mínimos cuadrados del parámetro β es:

$$\hat{\beta}_{\text{MCO}} = \frac{\Sigma_2^T y_t y_{t-1}}{\Sigma_2^T y_{t-1}^2} = \Sigma_2^T \frac{(\beta y_{t-1} + u_t) y_{t-1}}{\Sigma_2^T y_{t-1}^2} = \beta + \frac{\Sigma_2^T y_{t-1} u_t}{\Sigma_2^T y_{t-1}^2}$$

de modo que el estimador será insesgado si y sólo si se cumple:

$$E\left(\frac{\Sigma_2^T y_{t-1} u_t}{\Sigma_2^T y_{t-1}^2}\right) = 0 \tag{9.3}$$

Si la distribución de u_t fuese independiente de y_s para todo par (t, s), entonces se tendría para $s = 2, \ldots, T$

$$E(y_{s-1} u_s / \Sigma_2^T y_{t-1}^2) = E(y_{s-1} / \Sigma_2^T y_{t-1}^2) \cdot E(u_s) = 0 \tag{9.4}$$

por lo que la condición [9.3] se cumpliría y el estimador MCO sería insesgado. Sin embargo, [9.2] muestra que las distribuciones de y_t y u_s no son independientes, puesto que si el valor absoluto de β es inferior a la unidad, entonces:

$$y_t = \Sigma_{s=0}^{\infty} \beta^s u_{t-s}$$

por lo que y_t depende de u_t y de valores retardados de u_t. Por consiguiente, y puesto que u_t es ruido blanco, y_t será independiente de u_{t+s} para $s \geqslant 1$; en particular, $E(y_{t-1} u_t) = 0$, por lo que el numerador de la fracción en [9.3] tendrá esperanza cero. Lo que ocurre es que el denominador incluye valores de y_t posteriores a u_s, por lo que no podemos efectuar el desarrollo en [9.4]. El estimador MCO del modelo [9.2] será, en general, sesgado. Sin embargo, veremos en la Sección 9.4.a que, bajo determinadas condiciones, dicho sesgo tiende a cero, por lo que el estimador MCO de [9.2] es consistente.

El problema se complica sustancialmente cuando *aparecen valores retardados de la variable endógena como variables explicativas y, además, el término de error tiene autocorrelación:*

$$y_t = \beta y_{t-1} + u_t, \qquad |\beta| < 1 \tag{9.5.a}$$

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{9.5.b}$$

En este caso, basta escribir el modelo [9.5.a] en el instante $t - 1$ para darse cuenta de que la variable explicativa y_{t-1} está correlacionada con u_{t-1}, quien, a su vez, está correlacionada con u_t (véase [9.5.b]). En consecuencia, una de las variables explicativas del modelo está correlacionada con el término de error, por lo que ya no se tiene $E(y_{t-1} u_t) = 0$.

En efecto, si u_t tiene una estructura autorregresiva como [9.5.b], entonces $E(u_t u_{t-s}) = \rho^s \sigma_u^2$, y teniendo en cuenta que $|\beta\rho| < 1$, se tiene en el modelo [9.5]:

$$E(y_{t-1} u_t) = E[(\Sigma_{i=0}^{\infty} \beta^i u_{t-1-i}) u_t] = \Sigma_{i=0}^{\infty} \beta^i \rho^{i+1} \sigma_u^2 = \frac{\rho \sigma_u^2}{1 - \beta\rho} \qquad [9.6]$$

y, en general:

$$E(y_{t-s} u_t) = E[(\Sigma_{i=0}^{\infty} \beta^i u_{t-s-i}) u_t] = \Sigma_{i=0}^{\infty} \beta^i \rho^{i+s} \sigma_u^2 = \frac{\rho^s \sigma_u^2}{1 - \beta\rho}$$

por lo que $E(y_{t-s} u_t) \neq 0$ para todo $s \geqslant 1$. Ello hace que, como veremos en secciones posteriores, no podamos garantizar en este caso la consistencia del estimador MCO.

La explicación de la gran diferencia que existe entre estos dos casos es que todo lo que es preciso para que el estimador mínimo cuadrático de los coeficientes del modelo

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t, \qquad t = 1, 2, \ldots, T$$

sea consistente es que se tenga $E[x_{t-s} u_t] = 0$ para todo $s \geqslant 0$ y para todas las variables explicativas del modelo. Una variable que satisface esta condición se llama *predeterminada*, a diferencia de las variables *exógenas*, que son las que satisfacen la condición $E[x_{t-s} u_t] = 0$ para todo s, ya sea positivo, negativo o cero. Toda variable exógena es, claramente, predeterminada. En particular, y_{t-1} juega el papel de una de las variables explicativas x_t, *siendo predeterminada cuando u_t es ruido blanco, pero no cuando u_t tiene autocorrelación*. Por otra parte, y_{t-1} no es exógena en ningún caso.

9.2. JUSTIFICACION TEORICA DE LOS MODELOS ECONOMETRICOS DINAMICOS

Ha sido tradicional en Econometría utilizar tres ideas alternativas para justificar una especificación dinámica. Veamos brevemente dichas especificaciones.

9.2.a. El modelo de expectativas adaptativas

En 1956, P. Cagan propuso un modelo analítico en el que la demanda de saldos monetarios reales se hacía depender del valor esperado de la tasa de inflación futura:

$$\frac{M_t}{P_t} = \beta_1 + \beta_2 E_t \pi_{t+1} + u_t$$

Este modelo encierra, en realidad, distintas especificaciones, según cual sea la hipótesis que se haga acerca del modo en que los agentes económicos forman sus expectativas acerca del valor futuro de la tasa de inflación. El mecanismo de *expectativas adaptativas*, utilizado por Cagan (así como por M. Friedman en su Teoría de Consumo), es:

$$E_t\pi_{t+1} = E_{t-1}\pi_t + \lambda(\pi_t - E_{t-1}\pi_t), \qquad 0 < \lambda < 1$$

que postula que los agentes modifican la expectativa que acerca de π_t formaron el período anterior teniendo en cuenta únicamente el error de predicción cometido. Es decir, si la tasa de inflación está hoy por encima de lo que se esperaba, se corrige al alza dicha expectativa, mientras que se corrige a la baja en caso contrario.

El modelo de expectativas adaptativas puede también escribirse:

$$E_t\pi_{t+1} = \lambda\pi_t + (1-\lambda)E_{t-1}\pi_t$$

mostrando que la expectativa de inflación futura que hoy se forma es una combinación lineal del valor actual de la tasa de inflación, y de la expectativa de inflación que se formó el período anterior.

Nótese que, en el caso extremo en que $\lambda = 0$, entonces $E_t\pi_{t+1} = E_{t-1}\pi_t$, es decir, que las expectativas de inflación son estáticas y no se hacen depender del error de predicción que se haya cometido. En el otro caso extremo en que $\lambda = 1$, se tiene $E_t\pi_{t+1} = \pi_t$, y las expectativas son totalmente adaptativas, ya que se adopta como valor esperado de la inflación futura el valor que la tasa de inflación ha tomado en este período. Se ignora así la información que condujo a formar las expectativas pasadas.

Aun puede obtenerse una nueva representación para la expectativa del valor futuro de la tasa de inflación, si iteramos en la última expresión:

$$E_t\pi_{t+1} = \lambda\pi_t + \lambda(1-\lambda)\pi_{t-1} + \lambda(1-\lambda)^2\pi_{t-2} + \lambda(1-\lambda)^3\pi_{t-3} + \cdots$$

es decir, una combinación lineal del valor actual y de todos los valores pasados de la tasa de inflación con coeficientes que decrecen a una tasa igual a $1-\lambda$.

Si se incorporan las expectativas adaptativas al modelo de regresión, se tiene:

$$\frac{M_t}{P_t} = \lambda\beta_1 + \lambda\beta_2\pi_t + (1-\lambda)\left(\frac{M_{t-1}}{P_{t-1}}\right) + v_t$$

Aunque la estimación de este modelo no está exenta de dificultades (como pronto vamos a ver), una vez que se hubiesen estimado sus coeficientes, el parámetro λ podría obtenerse a partir del coeficiente estimado de $\dfrac{M_{t-1}}{P_{t-1}}$, mientras que el parámetro β_2 se obtendría dividiendo el coeficiente de π_t por la estimación del parámetro λ, y β_1 a partir del término independiente.

9.2.b. El modelo de ajuste parcial de Nerlove

Supongamos que el nivel de capital deseado en la economía, K_t^*, es una función del nivel de producto Y_t:

$$K_t^* = \beta_1 + \beta_2 Y_t + u_t \qquad [9.7]$$

Si un investigador quisiera proceder a estimar cómo varía el stock de capital deseado (u óptimo) según la economía transcurre a través de una época de recesión o de expansión, tendría el grave problema de no disponer de observaciones K_t^*. Para evitar tal dificultad, se añade al modelo anterior una ecuación que describe el mecanismo por el que el stock de capital se ajusta a su nivel deseado. Supongamos que:

$$K_t - K_{t-1} = \delta(K_t^* - K_{t-1}), \qquad 0 < \delta < 1$$

que postula que el stock de capital observado varía de un período a otro en una proporción de su distancia con respecto al stock deseado. Este es el modelo conocido en Econometría como de *ajuste parcial*. Si $\delta = 1$, entonces el stock de capital es igual, en cada período, a su valor deseado, lo que sólo sería razonable en una economía en que (al contrario de lo que ocurre en economías reales) el stock no está sujeto a importantes costes de ajuste. Por otra parte, si $\delta = 0$, entonces $K_t = K_{t-1}$, y el stock de capital no cambia, con independencia de lo lejos que se halle de su valor deseado, lo que, en general, no será óptimo[(1)].

El modelo puede también escribirse:

$$K_t = \delta K_t^* + (1 - \delta) K_{t-1}$$

que postula que, en cada período, el stock de capital es una combinación lineal convexa del valor deseado y de su valor previo. También puede iterarse la expresión anterior, para obtener:

$$K_t = \delta K_t^* + \delta(1 - \delta) K_{t-1}^* + \delta(1 - \delta)^2 K_{1-2}^* + \cdots$$

Al introducir el mecanismo de ajuste parcial en el modelo econométrico se tiene:

$$K_t = \delta\beta_1 + \delta\beta_2 Y_1 + (1 - \delta) K_{t-1} + \delta u_t \qquad [9.8]$$

Una vez estimado el modelo, el parámetro δ se obtiene del coeficiente del K_{t-1}, mientras que β_2 se obtendría dividiendo el coeficiente de Y_t por el valor

(1) En ocasiones, la ecuación [9.7] de demanda a largo plazo se supone determinista, mientras que la ecuación de ajuste se supone aleatoria: $K_t - K_{t-1} = \delta(K_t^* - K_{t-1}) + u_t$. El modelo que se obtiene de este modo es indistinguible del descrito en esta sección.

de δ y β_1 a partir del término independiente estimado. La ecuación [9.8] suele calificarse como demanda de capital a corto plazo, frente a [9.7], que se interpreta como demanda a largo plazo. Esta distinción permite explicar la diferencia entre el valor observado K_t y el valor deseado K_t^*, que puede ahora obtenerse a partir del modelo [9.7].

Por último, en el Problema 9.6 al final del capítulo se pide al lector que combine en un modelo una especificación de costes de ajuste junto con una de expectativas adaptativas.

9.3. MODELOS DE RETARDOS INFINITOS

En ocasiones, especialmente con observaciones frecuentes, el investigador especifica una relación del tipo:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \beta_4 x_{t-2} + \beta_4 x_{t-3} + \cdots + u_t \qquad [9.9]$$

en que la variable endógena se hace depender de un número infinito de retardos de una de las variables exógenas. Es claro que el modelo anterior no puede estimarse, puesto que para estimar los coeficientes de los retardos de x_t hay que perder observaciones iniciales y, en este caso, nos quedaríamos sin grados de libertad. Se hace preciso, por tanto, cambiar la parametrización del modelo de forma que aparezca un número finito (y reducido) de parámetros. Para conseguir esta reducción en el número de parámetros es preciso hacer supuestos acerca de la evolución de los coeficientes de los sucesivos retardos de la variable exógena.

9.3.a. El modelo de Koyck

Un supuesto muy habitual es el de Koyck:

$$\beta_i = \delta\beta_{i-1} \qquad |\delta| < 1 \quad \text{para todo} \quad i \geqslant 3 \qquad [9.10]$$

que convierte el modelo [9.9] en:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_t + \delta\beta_2 x_{t-1} + \delta^2\beta_2 x_{t-2} + \delta^3\beta_2 x_{t-3} + \cdots + u_t = \\ &= \beta_1 + \beta_2(x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \delta^3 x_{t-3} + \cdots) + u_t \end{aligned} \qquad [9.11]$$

que depende de tan sólo tres parámetros: β_1, β_2 y δ. El supuesto $|\delta| < 1$ es necesario para evitar una especificación en que el pasado de x_t tuviese un impacto más importante sobre y_t cuanto más lejano en el tiempo fuese el valor considerado de x_t.

La ganancia del modelo de Koyck es β_2 y su retardo medio es $\frac{1}{(1-\delta)}$, menor cuanto mayor es δ. Para poder estimar [9.11] es preciso truncar polinomio de retardos que aparece dentro del paréntesis en el primer período contenido en la muestra, de modo que se tiene:

$$y_t = \beta_1 + \beta_2(x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \cdots + \delta^{t-1} x_1) + \\ + \delta^t \beta_2(x_0 + \delta x_{-1} + \delta^2 x_{-2} + \cdots) + u_t$$

Si consideramos una nueva variable z_t definida por el contenido del primer paréntesis: $z_t = x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \cdots + \delta^{t-1} x_1$, entonces, dado un valor del parámetro δ, podríamos construir una serie temporal de observaciones de dicha variable, simplemente acumulando las observaciones pasadas de la variable x_t, multiplicadas por potencias de δ, del modo que indica la definición de z_t.

Sin embargo, no puede llevarse a cabo un procedimiento similar para el segundo paréntesis, puesto que no tenemos ninguna información acerca de los valores de las variables que en él aparecen. Dicho paréntesis puede, sin embargo, tratarse como un parámetro desconocido, puesto que no depende del tiempo. Así, definiendo

$$\gamma = \beta_2(x_0 + \delta x_{-1} + \delta x_{-2} + \delta^3 x_{-3} + \cdots)$$

se tiene el modelo transformado:

$$y_t = \beta_1 + \beta_2 z_t + \gamma\delta^t + u_t \qquad [9.12]$$

En cuanto a generar las observaciones muestrales de la variable auxiliar z_t que hemos introducido más arriba, es conveniente notar que se tiene:

$$z_1 = x_1\,; \quad z_2 = x_2 + \delta z_1\,; \quad z_3 = x_3 + \delta z_2\,; \quad \cdots; \quad z_t = x_t + \delta z_{t-1}$$

de modo que se puede hacer recursivamente.

9.3.b. Estimación de máxima verosimilitud del modelo de Koyck

Si suponemos que el término de error u_t sigue una distribución Normal$(0, \sigma_u^2)$, entonces el logaritmo de la función de verosimilitud de [9.12] resulta:

$$\ln L(y_t, z_t/\beta_1, \beta_2, \gamma, \delta) = -\frac{T}{2}\ln 2\pi - \frac{T}{2}\ln \sigma_u^2 - \frac{1}{2\sigma_u^2}\Sigma_1^T(y_t - \beta_1 - \beta_2 z_t - \gamma\delta^t)^2 \qquad [9.13]$$

por lo que maximizar la función de verosimilitud con respecto a β_1, β_2, γ y δ es equivalente a minimizar la suma de los cuadrados de los residuos del modelo [9.12] coincidiendo de este modo el estimador de máxima verosimilitud con el estimador de mínimo cuadrados ordinarios.

Puesto que el parámetro δ debe tomar valores en el intervalo $(-1, 1)$, es posible hacer una partición de dicho intervalo, por ejemplo: -1, $-0{,}90$, ..., 0, 0,10, 0,20, ..., 1, y estimar el modelo [9.12] por mínimos cuadrados ordinarios bajo cada uno de estos valores de δ:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \Sigma_1^T z_t & \dfrac{\delta(1-\delta^T)}{1-\delta} \\ & \Sigma_1^T z_t^2 & \Sigma_1^T \delta^t z_t \\ & & \dfrac{\delta^2(1-\delta^{2T})}{1-\delta^2} \end{pmatrix} \cdot \begin{pmatrix} \Sigma_1^T y_t \\ \Sigma_1^T z_t y_t \\ \Sigma_1^T \delta^t y_t \end{pmatrix} \qquad [9.14]$$

En general, no tiene mucho sentido suponer que los coeficientes β_i del modelo [9.9] alternan en signo, por lo que δ se supone inferior a la unidad en valor absoluto, pero positivo. En tal caso, es el intervalo (0, 1) el que se particiona. Tras estimar el modelo [9.12] suponiendo que δ toma, alternativamente, cada uno de los valores de la partición, se escoge aquel valor de δ que generó una suma residual menor o, equivalentemente, un R^2 más alto. Las estimaciones de β_1 y β_2 son las que se obtuvieron con dicho valor de δ. Si se quiere afinar más en los valores numéricos estimados, puede hacerse una subdivisión de un intervalo alrededor del valor de δ inicialmente estimado, y repetir el proceso.

La matriz de covarianzas obtenida de la estimación de mínimos cuadrados finalmente escogida sólo sería apropiada si el verdadero valor del parámetro δ hubiese sido conocido de antemano, de modo que dicha regresión hubiese sido la única que hubiésemos estimado. Este no ha sido el caso, y la matriz de covarianzas apropiada es la inversa de la matriz de información, pues el estimador que hemos obtenido es, en realidad, el de máxima verosimilitud.

Para ello, habría que obtener la matriz de derivadas segundas de la función de verosimilitud con respecto al vector de parámetros (β_1, β_2, γ, δ, σ_u^2), puesto que ahora se estiman todos simultáneamente. Dicha matriz de covarianzas es:

$$I^{-1}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\gamma} \\ \hat{\delta} \\ \hat{\sigma}_u^2 \end{pmatrix} = \sigma_u^2 \begin{pmatrix} T & \Sigma_1^T z_t & \dfrac{\delta(1-\delta^T)}{1-\delta} & \Sigma_1^T\left(\beta_2 \dfrac{\partial z_t}{\partial \delta} + t\delta^{t-1}\gamma\right) & 0 \\ & \Sigma_1^T z_t^2 & \Sigma_1^T \delta^t z_t & \Sigma_1^T\left(\beta_2 \dfrac{\partial z_t}{\partial \delta} + t\delta^{t-1}\gamma\right) z_t & 0 \\ & & \dfrac{\delta^2(1-\delta^{2T})}{1-\delta^2} & \Sigma_1^T\left(\beta_2 \dfrac{\partial z_t}{\partial \delta} + t\delta^{t-1}\gamma\right)\delta^t & 0 \\ & & & \Sigma_1^T\left(\beta_2 \dfrac{\partial z_t}{\partial \delta} + t\delta^{t-1}\gamma\right)^2 & 0 \\ & & & & \dfrac{T}{2\sigma_u^2} \end{pmatrix}^{-1}$$

acerca de la cual cabe hacer varias observaciones:

1. La varianza del estimador $\hat{\sigma}_u^2$ es igual a $\frac{2\sigma_u^4}{T}$, y es independiente de $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\gamma}$ y $\hat{\delta}$.

2. La submatriz superior de orden 3×3 coincide con la matriz que se utilizó en [9.14] para obtener el estimador MCO. Sin embargo, la matriz de covarianzas del estimador de máxima verosimilitud del vector $(\beta_1, \beta_2, \gamma)$ es la submatriz 3×3 que aparecería *en la inversa de la matriz de información*, y no coincide con la inversa de la matriz en [9.14]. Esta última ignoraría el hecho de que la última variable explicativa del modelo depende del parámetro desconocido δ.

9.4. ESTIMACION CON RETARDOS DE LA VARIABLE ENDOGENA

Como mencionamos antes, en la estimación de un modelo de este tipo es fundamental distinguir entre dos casos, según que el término de error del modelo tenga o no autocorrelación.

9.4.a. El término de error no tiene autocorrelación

Consideremos el modelo:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 x_t + u_t, \qquad |\beta_2| < 1$$

cuyas variables explicativas y término de error satisfacen las siguientes propiedades:

a) $E(\mathbf{u}) = \mathbf{0}_T$, $E(\mathbf{uu}') = \sigma_u^2 \mathbf{I}_t$, es decir, no existe autocorrelación.

b) $E(x_t u_t) = 0$ para todo t, ya que x_t es determinista.

c) $E(y_{t-1} u_t) = 0$, pues aunque y_{t-1} es estocástica, como se vio en la Sección 9.1, si $|\beta_2| < 1$, y_{t-1} depende de $u_{t-1}, u_{t-2}, \ldots$, pero no de u_t, y si este proceso es un ruido blanco, entonces se tiene el resultado citado.

d) $plim \left(\frac{\mathbf{X}'\mathbf{X}}{T}\right) = \mathbf{\Sigma}_{xx}$, matriz simétrica, definida positiva, **donde:**

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} T-1 & \Sigma_2^T y_{t-1} & \Sigma_2^T x_t \\ & \Sigma_2^T y_{t-1}^2 & \Sigma_2^T y_{t-1} x_t \\ & & \Sigma_2^T x_t^2 \end{pmatrix}$$

que a pesar de ser denotada por $\mathbf{X}'\mathbf{X}$ incluye también, en este caso, valores de la variable endógena, por aparecer un retardo suyo como variable explicativa. Esta condición *d*) se satisface, en general, bajo el supuesto $|\beta_2| < 1$, siempre que existan las varianzas y covarianzas de las variables explicativas x_t e y_{t-1}. Con estos supuestos, el teorema de Mann-Wald (véase Capítulo 2) asegura que:

$$plím\left(\frac{\mathbf{X}'\mathbf{u}}{T}\right) = \mathbf{0}_k$$

y que:

$$\frac{\mathbf{X}'\mathbf{u}}{\sqrt{T}} \xrightarrow{D} N(\mathbf{0}, \sigma_u^2 \mathbf{\Sigma}_{xx}) \qquad [9.15]$$

de modo que se tiene:

$$plím\, \hat{\boldsymbol{\beta}}_{\text{MCO}} = plím\left[\boldsymbol{\beta} + \left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{u}}{T}\right)\right] =$$
$$= \boldsymbol{\beta} + plím\left[\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1}\right] plím\left(\frac{\mathbf{X}'\mathbf{u}}{T}\right) = \boldsymbol{\beta} + \mathbf{\Sigma}_{xx}^{-1}\mathbf{0}_k = \boldsymbol{\beta}$$

y el estimador de mínimos cuadrados ordinarios es consistente. Además, como

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

entonces:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{\text{MCO}} - \boldsymbol{\beta}) = \left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{u}}{\sqrt{T}}\right)$$

y utilizando [9.15] y la Proposición 2.16:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{\text{MCO}} - \boldsymbol{\beta}) \xrightarrow{D} (\mathbf{\Sigma}_{xx})^{-1} N(\mathbf{0}, \sigma_u^2 \mathbf{\Sigma}_{xx}) = N(\mathbf{0}, \sigma_u^2 (\mathbf{\Sigma}_{xx})^{-1})$$

Aunque esta distribución sólo es rigurosamente válida según tiende el tamaño muestral a infinito, en la práctica suele llevarse a cabo la siguiente aproximación: en primer lugar, pasando el factor $\sqrt{T}$, así como el vector constante $\boldsymbol{\beta}$, al miembro de la derecha se tiene que, aproximadamente, $\hat{\boldsymbol{\beta}}_{\text{MCO}}$ se distribuye en muestras grandes según una $N\left[\boldsymbol{\beta}, \left(\frac{\sigma_u^2}{T}\right)(\mathbf{\Sigma}_{xx})^{-1}\right]$. En segundo lugar, como la matriz $\mathbf{\Sigma}_{xx}$ es el límite de $\frac{\mathbf{X}'\mathbf{X}}{T}$, entonces, si la muestra es suficientemente grande, la matriz $\mathbf{\Sigma}_{xx}$ puede sustituirse por esta última matriz $\frac{\mathbf{X}'\mathbf{X}}{T}$, por lo que, en definitiva, resulta como matriz de covarianzas aproximada de $\hat{\boldsymbol{\beta}}_{MCO}$ la expresión habitual $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, a pesar de la presencia de la variable endógena retardada formando parte de la matriz $\mathbf{X}'\mathbf{X}$.

En resumen, en tanto en cuanto el término de error esté libre de autocorrelación, está justificado el uso de mínimos cuadrados en un modelo que incluye retardos de la variable endógena. Puede utilizarse la matriz de covarianzas habitual de dicho estimador, quien tiene además una distribución Normal en muestras grandes, por lo que los resultados de inferencia estadística del Capítulo 4 son aproximadamente válidos.

Debe notarse que el razonamiento anterior es válido con independencia del número de retardos de la variable endógena que aparecen como variables explicativas. Todo lo que es preciso es que, considerados separadamente del resto de las variables explicativas, los coeficientes de dichos retardos satisfagan las condiciones de estacionariedad que hemos supuesto aquí, y que discutiremos en mayor extensión en el Capítulo 13.

9.4.b. Estimación cuando el término de error tiene autocorrelación

Consideremos el modelo:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 y_{t-1} + \beta_3 x_t + u_t \qquad & |\beta_2| < 1 \\ u_t &= \rho u_{t-1} + \varepsilon_t & |\rho| < 1 \end{aligned} \qquad [9.16]$$

donde ε_t es ruido blanco.

La existencia de autocorrelación en el término de error hace que la propiedad *c*) del modelo discutido en 9.4.a no se satisfaga, ya que $E(y_{t-1}u_t) \neq 0$, como se vio en la Sección 9.1.c: y_{t-1} depende de u_{t-1} a través del modelo, pero u_{t-1} y u_t están relacionados con la estructura autorregresiva del término de error. En consecuencia, y_{t-1} y u_t están correlacionados. El estimador de mínimos cuadrados es sesgado, y además su sesgo no desaparece al aumentar el tamaño muestral.

Por ejemplo, en el caso más sencillo del modelo [9.16], con $\beta_1 = \beta_3 = 0$, se tiene:

$$plím\ \hat{\beta}_{2\,MCO} = \beta_2 + \frac{plím\ 1/T(\Sigma_2^T y_{t-1}u_t)}{plím\ 1/T(\Sigma_2^T y_{t-1}^2)}$$

y si los momentos muestrales convergen en probabilidad a sus análogos poblacionales, como $E(y_{t-1}u_t) = \frac{\rho\sigma_u^2}{1-\beta_2\rho}$ y $E(y_{t-1}^2)$ también es distinto de cero (véase Problema 9.8), entonces se tiene que el estimador mínimo cuadrático no es consistente. Si esto ocurre en un modelo sencillo, no debe esperarse que la situación sea más favorable en modelos más complejos. En efecto, este resultado de inconsistencia es válido, en general, siempre que aparezcan conjuntamente retardos de la variable endógena y autocorrelación del término de error.

9.4.c. El estimador de variables instrumentales

El procedimiento para obtener estimaciones consistentes de un modelo de este tipo se conoce como *estimador de variables instrumentales.* Una variable instrumental es una variable z_t que satisface tres condiciones: *a*) no está incluida en el modelo como variable explicativa, *b*) está incorrelacionada con el término de error $E(z_t u_t) = 0$, y *c*) está correlacionada con la variable para la cual hace de instrumento, en este caso y_{t-1}.

Por ejemplo, en el modelo [9.16], el primer retardo de la variable exógena x_{t-1} satisface estas tres condiciones, pues aparte de que, por ser determinista: $E(x_{t-1}u_{t-1}) = 0$, también se tiene que x_{t-1} influye sobre y_{t-1} a través del propio modelo [9.16] escrito para el período $t - 1$.

En general, en el vector $\mathbf{x}_t$ tan sólo habrá unas variables que no satisfagan la condición $E(x_t u_t) = 0$, y son estas variables las que necesitan de variables instrumentales. Es decir, los vectores $\mathbf{x}_t$ y $\mathbf{z}_t$ tendrán en común aquellas variables que están incorrelacionadas con el término de error. Así, en el ejemplo anterior $\mathbf{x}_t = (1, y_{t-1}, x_t)$, mientras que $\mathbf{z}_t$ podría ser el vector $\mathbf{z}_t = (1, x_{t-1}, x_t)$. Finalmente, el estimador de variables instrumentales viene dado por la expresión matricial:

$$\hat{\boldsymbol{\beta}}_{\text{VI}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} \qquad [9.17]$$

donde $\mathbf{Z}$ denota la matriz $T \times k$ de observaciones muestrales de las variables que componen el vector $\mathbf{z}_t$, y suponemos que $\mathbf{Z}'\mathbf{X}$ es invertible. Volviendo al ejemplo anterior, el estimador de variables instrumentales se obtendría mediante.

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix}_{\text{VI}} = \begin{pmatrix} T-1 & \Sigma_2^T y_{t-1} & \Sigma_2^T x_t \\ \Sigma_2^T x_{t-1} & \Sigma_2^T x_{t-1} y_{t-1} & \Sigma_2^T x_t x_{t-1} \\ \Sigma_2^T x_t & \Sigma_2^T x_t y_{t-1} & \Sigma_2^T x_t^2 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_2^T y_t \\ \Sigma_2^T x_{t-1} y_t \\ \Sigma_2^T x_t y_t \end{pmatrix} \qquad [9.18]$$

donde, como puede verse, la matriz $\mathbf{Z}'\mathbf{X}$ dista de ser simétrica.

En cuanto a la correlación que debe existir entre una variable instrumental y la variable explicativa para la que se utiliza como instrumento, cabe observar lo siguiente: *a*) es importante que dicha correlación exista, puesto que como muestra la expresión [9.18], la variable instrumental x_{t-1} sustituye parcialmente a la variable y_{t-1} en la estimación del modelo econométrico. pero, sin embargo, *b*) dicha correlación no puede ser muy importante, puesto que entonces también existiría una correlación apreciable entre la variable instrumental z_t y el término de error u_t, que es lo que motivó la necesidad de la variable instrumental.

En el modelo [9.16] también podría utilizarse otro retardo, por ejemplo x_{t-2}, como variable instrumental. La diferencia es que la relación entre esta variable e y_{t-1} se hace más indirecta; así, x_{t-2} está directamente correlacionada con y_{t-2}, quien a su vez está correlacionada con y_{t-1} a través del modelo [9.16], y ésta con y_t.

El estimador de variables instrumentales del modelo [9.16] es, en general, sesgado, pues la variable y_{t-1} aparece en la matriz $\mathbf{Z}'\mathbf{X}$. Para su insesgo debería ocurrir que $E[(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u}] = \mathbf{0}_k$, pero ello envuelve en general funciones muy complejas cuya esperanza desconocemos. Afortunadamente, podremos afirmar que el estimador es consistente bajo las condiciones de la siguiente proposición:

Proposición 9.1. Sea $\mathbf{Z}$ una matriz $T \times k$ de observaciones de las variables $z_1, z_2, \ldots, z_k$, quizá aleatorias. Sea $\mathbf{z}'_t$ la fila t de $\mathbf{Z}$ y supongamos que se tiene:

1. $E(\mathbf{z}'_t u_t) = \mathbf{0}_k$ para todo t.
2. $plím \left(\frac{\mathbf{Z}'\mathbf{X}}{T}\right) = \mathbf{\Sigma}_{zx}$, $plím \left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right) = \mathbf{\Sigma}_{zz}$, ambas no singulares y finitas.

Entonces, el estimador de variables instrumentales es consistente.

Demostración:

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

$$= \boldsymbol{\beta} + \left[plím\left(\frac{\mathbf{Z}'\mathbf{X}}{T}\right)\right]^{-1} plím\left(\frac{\mathbf{Z}'\mathbf{u}}{T}\right) = \boldsymbol{\beta} + \mathbf{\Sigma}_{zx}^{-1}\mathbf{0}_k = \boldsymbol{\beta}$$

donde se ha utilizado que los instrumentos satisfacen: $plím \frac{\mathbf{Z}'\mathbf{u}}{T} = \mathbf{0}_k$.

Así pues, la consistencia de $\hat{\boldsymbol{\beta}}_{\text{VI}}$ proviene de la ausencia de correlación entre instrumentos y término de error, con independencia de que éste tenga o no autocorrelación. En ausencia de autocorrelación, podemos caracterizar la distribución asintótica de $\hat{\boldsymbol{\beta}}_{\text{VI}}$:

Proposición 9.2. Dado el modelo $y_t = \mathbf{x}'_t\boldsymbol{\beta} + u_t$, donde $\mathbf{x}_t$ es el vector de variables explicativas, que puede incluir algunos retardos de la variable endógena, y u_t, el término de error es un ruido blanco, sea $\mathbf{Z}$ la matriz $T \times k$ de observaciones de las variables $z_1, z_2, \ldots, z_k$, y supongamos que:

1. $E(\mathbf{z}'_t u_t) = \mathbf{0}_k$ para todo t.
2. $plím \left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right) = \mathbf{\Sigma}_{zz}$, simétrica, definida positiva.
3. $plím \left(\frac{\mathbf{Z}'\mathbf{X}}{T}\right) = \mathbf{\Sigma}_{zx}$, no singular.

Entonces el estimador $\hat{\boldsymbol{\beta}}_{VI} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$ es consistente y se tiene:

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{VI} - \boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}, \sigma_u^2(\boldsymbol{\Sigma}_{zx})^{-1}\boldsymbol{\Sigma}_{zz}(\boldsymbol{\Sigma}_{zx}^{-1})')$$

Demostración. La consistencia del estimador ha sido el objeto de la proposición anterior. En segundo lugar, el teorema de Mann-Wald asegura que bajo los supuestos 1, 2, 3 se tiene:

$$plím\left(\frac{\mathbf{Z}'\mathbf{u}}{T}\right) = \mathbf{0}_k$$

y

$$\frac{\mathbf{Z}'\mathbf{u}}{\sqrt{T}} \xrightarrow{D} N(\mathbf{0}_k, \sigma_u^2\boldsymbol{\Sigma}_{zz})$$

y como $\sqrt{T}(\hat{\boldsymbol{\beta}}_{VI} - \boldsymbol{\beta}) = \left(\frac{\mathbf{Z}'\mathbf{X}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{u}}{\sqrt{T}}\right)$, esta función converge en distribución a $(\boldsymbol{\Sigma}_{zx})^{-1}N(\mathbf{0}_k, \sigma_u^2\boldsymbol{\Sigma}_{zz})$, que es una distribución $N(\mathbf{0}_k, \sigma_u^2\boldsymbol{\Sigma}_{zx}^{-1}\boldsymbol{\Sigma}_{zz}(\boldsymbol{\Sigma}_{zx}^{-1})')$.

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

$$\text{Var}(\hat{\boldsymbol{\beta}}_{VI}) = \frac{\sigma_u^2}{T}(\boldsymbol{\Sigma}_{zx}^{-1})\boldsymbol{\Sigma}_{zz}[(\boldsymbol{\Sigma}_{zx})^{-1}]'$$

y si se utilizan las matrices de momentos muestrales $\frac{\mathbf{Z}'\mathbf{X}}{T}, \frac{\mathbf{Z}'\mathbf{Z}}{T}$ para aproximar sus límites respectivos $\boldsymbol{\Sigma}_{zx}$, $\boldsymbol{\Sigma}_{zz}$, entonces se tiene como matriz de covarianzas:

$$\text{Var}(\hat{\boldsymbol{\beta}}_{VI}) = \left(\frac{\sigma_u^2}{T}\right)\left(\frac{\mathbf{Z}'\mathbf{X}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)\left[\left(\frac{\mathbf{Z}'\mathbf{X}}{T}\right)^{-1}\right]' = \sigma_u^2(\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'\mathbf{Z})\,[(\mathbf{Z}'\mathbf{X})^{-1}]$$

Por último, el parámetro σ_u^2 se estimaría dividiendo la suma residual por el número de grados de libertad, $T-k$. Sin embargo, es importante tener en cuenta que, una vez más, los residuos deben calcularse utilizando las variables originales del modelo, es decir:

$$\hat{\sigma}_u^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{VI})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{VI})}{T-k}$$

El resultado anterior no puede generalizarse fácilmente al caso en que el término de error tiene autocorrelación, por lo que suele utilizarse la matriz de covarianzas anterior incluso en tal caso, aun a sabiendas que no es sino un aproximación. Un modelo de este tipo debe estimarse, a ser posible, por el método de máxima verosimilitud, que describimos en la Sección 9.7.

9.4.d. Contraste de exogeneidad de Hausman y Wu

La aparición de retardos de la variable endógena como variables explicativas en presencia de autocorrelación es un caso muy habitual en que se necesita *utilizar un procedimiento de variables instrumentales* si se pretende obtener estimaciones consistentes. Pero es aconsejable cuestionarse acerca de las propiedades de exogeneidad del resto de las variables explicativas, pues, de no satisfacerse, obtendríamos igualmente estimadores inconsistentes.

Hausman (1978) y Wu (1973) sugieren escribir el modelo a estimar, distinguiendo entre las r variables explicativas $\mathbf{Y}_1$ que *pueden* estar correlacionadas con el término de error de aquellas $k - r$ variables $\mathbf{Z}_1$ cuya ortogonalidad a u_t no se cuestiona:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = \mathbf{Y}_1\boldsymbol{\alpha} + \mathbf{Z}_1\boldsymbol{\delta} + \mathbf{u}$$

y supongamos que se dispone de una lista de instrumentos para $\mathbf{Y}_1$, en caso de que se necesitasen.

El contraste consiste en estimar inicialmente el modelo por MCO y obtener la suma residual SR_0. A continuación se estiman regresiones auxiliares de las variables en $\mathbf{Y}_1$ sobre los instrumentos, y se sustituye $\mathbf{Y}_1$ por la matriz de valores previstos $\hat{\mathbf{Y}}_1$ en el modelo inicial, que se estima por MCO, obteniendo la suma residual SR_1. El estadístico:

$$(\hat{\boldsymbol{\beta}}_{\mathrm{MCO}} - \hat{\boldsymbol{\beta}}_{\mathrm{VI}})'[\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{VI}}) - \mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{MCO}})]^{-1}(\hat{\boldsymbol{\beta}}_{\mathrm{MCO}} - \hat{\boldsymbol{\beta}}_{\mathrm{VI}}) = \frac{\mathrm{SR}_0 - \mathrm{SR}_1}{\hat{\sigma}_u^2} \sim \chi_r^2$$

bajo la hipótesis nula de que *todas* las variables explicativas del modelo original son exógenas. Un valor elevado del estadístico rebatiría tal supuesto y mostraría la necesidad de utilizar un procedimiento de estimación de variables instrumentales.

La lógica de este constraste puede utilizarse en situaciones distintas de la aquí analizada, constituyendo la filosofía general de «un test tipo Hausman». Supongamos que se dispone de dos estimadores: uno, $\hat{\boldsymbol{\beta}}_1$, que bajo H_0 es consistente y eficiente, pero que es inconsistente bajo la hipótesis alternativa y, un segundo estimador, $\hat{\boldsymbol{\beta}}_2$, que es consistente tanto bajo H_0 como bajo H_1, aunque sea, posiblemente, ineficiente. Bajo la hipótesis nula, la diferencia $\hat{\mathbf{q}} = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2$ converge en probabilidad a cero, pero no bajo la hipótesis alternativa. Hausman probó que si la distribución asintótica de $\sqrt{T}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta})$ está bien definida bajo H_0, entonces la covarianza asintótica de $\sqrt{T}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})$ y $\sqrt{T}\hat{\mathbf{q}}$ es igual a cero. Como consecuencia, la matriz de covarianzas del vector $\hat{\mathbf{q}}$ es igual a la diferencia de matrices: $\mathrm{Var}(\hat{\mathbf{q}}) = \mathrm{Var}(\hat{\boldsymbol{\beta}}_2) - \mathrm{Var}(\hat{\boldsymbol{\beta}}_1)$.