

CAPÍTULO 6

RAÍCES UNITARIAS MULTIVARIADAS Y COINTEGRACIÓN

El análisis del capítulo anterior demuestra que una variable estacionaria en diferencias o en tendencia presenta comportamientos que son bastante diferentes, lo que tiene importantes efectos en los procesos de estimación, inferencia y predicción de modelos.

En este capítulo extendemos el análisis para grupos de variables. Más allá de la predicción de la evolución futura de un variable económica, el principal interés de la economía es entender la forma como se relacionan las variable económicas entre sí. Considerando que una parte sustancial de los modelos económicos contienen variables económicas que posiblemente tienen raíces unitarias, resulta fundamental entender cómo la presencia de procesos no-estacionarios afecta a los modelos de ecuaciones simultáneas desarrollados en el Capítulo 3.

Más allá del efecto que las raíces unitarias pueden tener sobre estimadores, tests y predicciones, en este capítulo nos interesa estudiar dos fenómenos extremadamente importantes que ocurren cuando hay variables integradas en un modelo económico. En primer lugar, la *correlación espuria*, es decir, la obtención de coeficientes de correlación parcial altos entre variables que –por construcción– son estadísticamente independientes. En segundo lugar, la *cointegración*, es decir la propiedad que tienen ciertas variables integradas para exhibir combinaciones lineales que son estacionarias.

Para efectos de la modelación econométrica ambos fenómenos son fundamentales. El primer problema nos señala que, si el econometrista no está conciente del tipo de variable con que modela, es posible que sus resultados sean enteramente espurios, pese a que tests estadísticos convencionales sugieran que las variables están relacionadas. El segundo fenómeno nos señala que bajo ciertas circunstancias aún si hay variables integradas es posible que un modelo econométrico una descripción adecuada, pues variables que cointegran efectivamente están relacionadas en el largo plazo.

El análisis en este capítulo, además, se adentra en problemas de modelación conjunta de sistemas de variables, retoma la discusión de problemas de causalidad, y extiende el análisis de identificación de ecuaciones simultáneas desarrollado anteriormente.

6.01 Correlación espuria

El tema de la correlación espuria es muy antiguo, incluso anterior al desarrollo formal de la econometría. Existe el justificado temor que se extraigan conclusiones o se hagan inferencias sobre la base de variables que se mueven en tandem sin que tengan, en realidad, interacción alguna. Un ejemplo ilustrativo, si bien algo grueso, es la correlación espuria entre el crecimiento del PIB y la altura de un niño. La correlación será probablemente alta –al menos en los primeros 18 años–sin que exista, evidentemente, relación alguna entre ellas. Los primeros en denunciar este problema en series de tiempo fueron Yule (1910) y Student (1914), quienes resaltaron el hecho que correlación no es lo mismo que causalidad.¹ Simon (1954) señala que una forma de correlación espuria sucede cuando sobre dos variables no correlacionadas actúa una tercera fuerza que influye sistemáticamente en ambas.² Este es, probablemente, un caso bastante común en el trabajo aplicado. No obstante, es evidente que la fuente de correlación espuria en este caso es más bien un problema de mala especificación del modelo y no necesariamente un fenómeno particular.

El problema de correlación espuria no se limita al caso de series de tiempo. El término “correlación espuria” fue introducido por Pearson en 1897³, en un artículo en el que se demostraba que la correlación de variables medidas como razones puede producir falsamente la ilusión de relación entre variables. Pearson extrajo numerosos grupos de tres tipos de huesos de una huesera (fémur, tibia y húmero) y formó razones entre la longitud de dichos huesos del tipo fémur/húmero y tibia/húmero. Por el simple hecho de usar razones en las que el denominador es común, Pearson demostró que la correlación entre ambas variables podía alcanzar a 0.4-0.5 aún si los huesos habían sido sacado completamente al azar.

En el Gráfico 6.1 se presenta una ilustración de este fenómeno, usando técnicas de Montecarlo. Se generó tres series completamente aleatorias a partir de una distribución normal, x^1 , x^2 y x^3 . Luego se procedió a crear las razones x^1/x^3 y x^2/x^3 , y se corrió una regresión entre ambas razones. Este ejercicio se repite 1.000 veces (llamado “replicaciones” del experimento) y se procede a computar el histograma correspondiente a la bondad de ajuste del modelo. Como se desprende de la Gráfico 6.1, los resultados para el coeficiente de correlación distan mucho de ser cero como uno esperaría si se toma en consideración que, por contrucción no existe ninguna relación

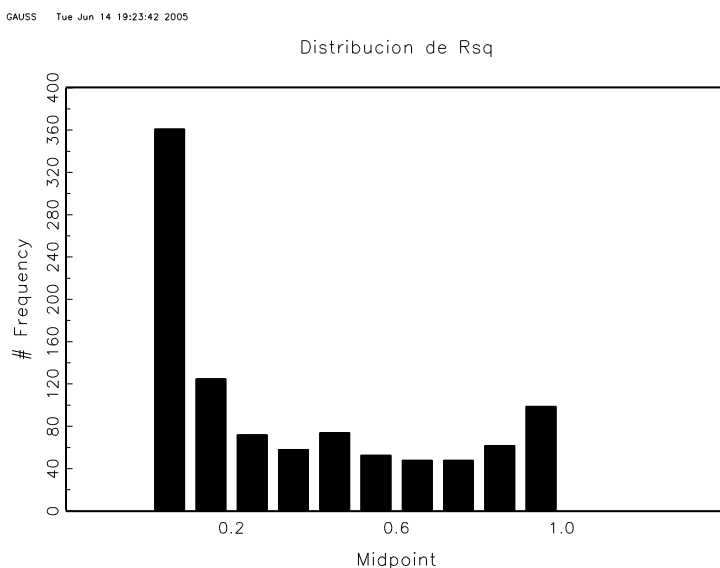
¹ Yule, G. U. (1910). “On the Interpretation of Correlations Between Indices or Ratios”. *Journal of the Royal Statistical Society*, 73, 644-647. Student's (1914) "The Elimination of Spurious Correlation due to Position in Time or Space" *Biometrika*, 10, 179-180.

² H. Simon (1954). "Spurious Correlation: A Causal Interpretation," *Journal of the American Statistical Association*, 49:467-479.

³ K. Pearson (1897). "Mathematical Contributions to the Theory of Evolution - on a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs," *Proc. Royal Society*, 60:489-498.

entre las variables. De hecho, el valor promedio de la correlación en este caso alcanzó a 0.35, valor que es bastante alto en comparación a muchos trabajos de corte transversal.⁴

Gráfico 6.1
Correlación Espuria de Pearson
Tamaño de muestra 100, Replicaciones = 1.000



En el caso de series de tiempo –y bajo el supuesto que el modelo está bien especificado– estamos interesados en una tercera forma de correlación espuria, la que sucede cuando se hacen inferencias erróneas al aplicarse la teoría econométrica para series estacionarias a variables que no son estacionarias. En el Gráfico 6.2 se presenta una ilustración de este fenómeno, siguiendo el trabajo pionero de Granger y Newbold (1974)⁵. Usando técnicas de Montecarlo ellos descubrieron que un modelo del tipo $x_t^1 = \alpha + \beta x_t^2 + \epsilon_t$ estimado con series no correlacionadas por construcción, podía entregar alta correlación y significancia de parámetros un número anormalmente alto de casos (es decir, existen problemas de tamaño de los tests). Se crean dos series completamente aleatorias a partir de una distribución normal, ϵ_t^1 y ϵ_t^2 . Usando estas series como innovaciones y suponiendo un punto de partida igual a cero, se procedió a construir dos variables aleatorias I(1) con deriva. Se corrió una regresión de mínimos cuadrados entre ambas razones. Este ejercicio se repite 1.000 veces y se procede a computar el histograma correspondiente para la correlación y el estadígrafo t asociado al parámetro. Como se desprende del gráfico, los resultados para el coeficiente de

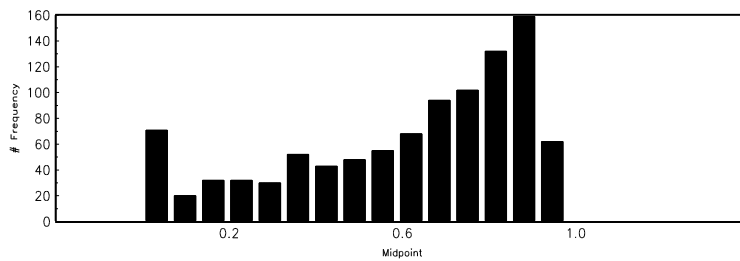
⁴ Ver el código Gauss de ésta y otras simulaciones en el Apéndice A.

⁵ Granger, C.W.J. and P. Newbold (1974). “Spurious regressions in econometrics”, *Journal of Econometrics*, 2(2):111-120.

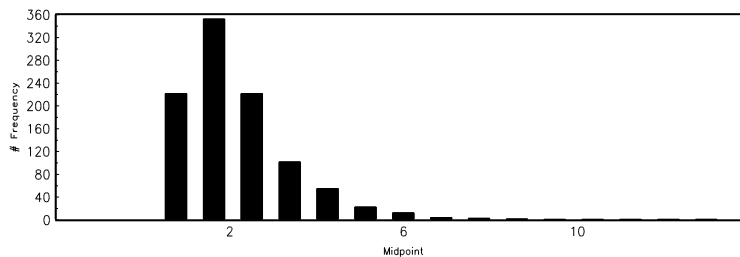
correlación y la significancia del parámetro distan mucho de ser cero como uno esperaría si se toma en consideración que, por construcción no existe ninguna relación alguna entre las variables. De hecho, el valor promedio de la correlación en esta simulación alcanzó a 0.64, valor significativamente distinto de cero. Por otro lado, en un 48% de los casos el test t resultó mayor que 2.

Gráfico 6.2
Correlación Espuria de Granger y Newbold
Tamaño de muestra 100, Replicaciones = 1000

Distribucion de la Bondad de Ajuste



Distribucion de los Tests t



El resultado de Granger y Newbold quedó como una desconcertante regularidad empírica hasta que Phillips⁶ proveyó una teoría que explica por qué razón y en cuáles circunstancias ocurre este problema de correlación espuria. Supongamos que tenemos un vector de $(n \times 1)$ variables, y_t , tal que:

$$(I_n - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t = \alpha + \epsilon_t \quad (6.1)$$

El polinomio en L^p puede ser factorizado como:

$$(I_n - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) = (I_n - \rho L) - (\xi_1 L + \xi_2 L^2 + \dots + \xi_{p-1} L^{p-1})(1 - L) \quad (6.2)$$

6 P.C.B. Phillips (1986). "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33(3): 311-340.

donde

$$\begin{aligned} \rho &= \phi_1 + \phi_2 + \dots + \phi_p \\ \xi_s &= -[\phi_{s+1} + \phi_{s+2} + \dots + \phi_p] \quad \forall s = 1, 2, \dots, p-1 \end{aligned} \quad (6.3)$$

por lo que aplicando la factorización a la ecuación (6.1) se obtiene

$$(I_n - \rho L) y_t - (\xi_1 L + \xi_2 L^2 + \dots + \xi_{p-1} L^{p-1})(1 - L) y_t = \alpha + \epsilon_t \quad (6.4)$$

o, alternativamente,

$$\begin{aligned} y_t &= (\xi_1 L + \xi_2 L^2 + \dots + \xi_{p-1} L^{p-1}) \Delta y_t + \rho y_{t-1} + \alpha + \epsilon_t \\ &= \xi_1 \Delta y_{t-1} + \xi_2 \Delta y_{t-2} + \dots + \xi_{p-1} \Delta y_{t-p+1} + \rho y_{t-1} + \alpha + \epsilon_t \end{aligned} \quad (6.5)$$

esta representación, similar al VAR en primeras diferencias pero en el nivel de las variables, nos servirá para estudiar qué papel juega la cointegración al modelar series de tiempo.

Vamos a suponer que la primera diferencia de y_t es un VAR(p-1). Entonces, se debe cumplir:

$$\rho = I_n = \phi_1 + \phi_2 + \dots + \phi_p \quad (6.6)$$

Por comodidad supongamos $\alpha=0$ y supongamos que estimamos el VAR por mínimos cuadrados. El proceso generador de datos es:

$$(I_n - \xi_1 L - \xi_2 L^2 - \dots - \xi_{p-1} L^{p-1}) \Delta y_t = \epsilon_t \quad (6.7)$$

Si no hay raíces unitarias, entonces los z_t tal que:

$$|I_n - \xi_1 z - \xi_2 z^2 - \dots - \xi_{p-1} z^{p-1}| = 0 \quad (6.8)$$

están fuera del círculo unitario. Así $\Delta y_t = \mu_t$, donde

$$\mu_t = (I_n - \xi_1 L - \xi_2 L^2 - \dots - \xi_{p-1} L^{p-1})^{-1} \epsilon_t \quad (6.9)$$

y, además, $y_t = y_0 + \mu_1 + \mu_2 + \dots + \mu_t$.

Consideremos la i-ésima ecuación de (6.5)

$$y_t^i = \xi_1 \Delta y_{t-1}^i + \xi_2 \Delta y_{t-2}^i + \dots + \xi_{p-1} \Delta y_{t-p+1}^i + \rho^i y_{t-1}^i + \alpha^i + \epsilon_t^i \quad (6.10)$$

Si nuestra suposición es correcta, entonces $\rho_i=1$. Veamos los estimadores de mínimos cuadrados (es decir, del tipo $(x_t' x_t)^{-1} x_t' y_t$) y su desviación respecto del verdadero valor del parámetro:

$$\hat{\beta}_T - \beta = \begin{bmatrix} \hat{\xi}_1^i - \xi_1^i \\ \hat{\xi}_2^i - \xi_2^i \\ \vdots \\ \hat{\xi}_{p-1}^i - \xi_{p-1}^i \\ \hat{\alpha}_i \\ \hat{\rho}_i^i - 1 \end{bmatrix} \quad (6.11)$$

La matriz de ponderación para obtener la distribución asintótica de los estimadores es:

$$Y_T = \begin{bmatrix} T^{1/2} \cdot I_{n(p-1)} & 0 & 0 \\ 0' & T^{1/2} & 0' \\ 0 & 0 & T \cdot I_n \end{bmatrix} \quad (6.12)$$

pre-multiplicando la ecuación (6.11) por la ecuación (6.12) obtenemos que:

$$Y_T(\hat{\beta}_T - \beta) = (Y_T^{-1} \sum x_t' x_t Y_T^{-1})^{-1} (Y_T^{-1} \sum x_t' \epsilon_t) \quad (6.13)$$

El primer término del lado derecho converge a $\begin{bmatrix} V & 0 \\ 0 & Q \end{bmatrix}$ donde

$$V = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \dots & \Gamma_{p-2} \\ \Gamma_{-1} & \Gamma_0 & \dots & \Gamma_{p-3} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{-p-2} & \Gamma_{-p-3} & \dots & \Gamma_0 \end{bmatrix} \quad \text{con} \quad \Gamma_s = E(\Delta y_t' \Delta y_{t-s}) \quad (6.14)$$

$$Q = \begin{bmatrix} 1 & \left[\int W(r) dr \right]' \Lambda \\ \Lambda \int W(r) dr & \Lambda \left[\int W(r) W(r)' dr \right] \Lambda' \end{bmatrix} \quad (6.15)$$

con $\Lambda = (I_n - \xi_1 - \xi_2 - \dots - \xi_{p-1})^{-1} P$ y donde $E(\epsilon_t' \epsilon_t) = P' P$.

El segundo término de la ecuación (6.13) converge a

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad \text{con} \quad b_1 \rightarrow N(0, \sigma_{ii} V) \quad y \quad b_2 = \begin{bmatrix} P W(1) \\ \Lambda \left[\int W(r) dW(r)' \right] P' \end{bmatrix} \quad (6.16)$$

Entonces, el estimador cumple

$$Y_T(\hat{\beta}_T - \beta) \rightarrow \begin{bmatrix} V^{-1}b_1 \\ Q^{-1}b_2 \end{bmatrix} \quad (6.17)$$

La interpretación de este resultado es la siguiente:

- los primeros $p-1$ estimadores convergen a variables Gaussianas. Por ello, se puede usar test de χ^2 o F para estudiar la significancia del bloque de parámetros (largo óptimo de rezagos).
- los otros dos estimadores $[\hat{\alpha}^i, \hat{\rho}^i - 1]$ no tienen una distribución estándar, pero $\hat{\rho}^i - 1$ converge más rápido que el resto de los parámetros (T vs \sqrt{T}).
- los tests de causalidad no se pueden usar porque no tienen distribución límite χ^2 .⁷

Cuando las variables tienen deriva, aparece un término adicional en la regresión asociado a ella, equivalente a $\delta(t-1)=\gamma$

$$Y_T(\hat{\beta}_T - \beta) = \begin{bmatrix} T^{1/2}(\hat{\Omega}_T - \omega) \\ T^{1/2}(\hat{\alpha} - \alpha) \\ T(\hat{\rho}_T - 1) \\ T^{3/2}(\hat{\gamma}_T - \gamma) \end{bmatrix} \rightarrow \begin{bmatrix} V^{-1}b_1 \\ Q^{-1}b_2 \\ Q^{-1}b_3 \\ Q^{-1}b_4 \end{bmatrix} \quad (6.18)$$

es decir, los parámetros de los rezagos de Δy_t son asintóticamente Gaussianos, el resto no lo es. Por ello, tests de bloques de rezagos están bien especificados, pero los otros tests no lo están. En particular, el test de causalidad de Granger no se aplica.

¿Qué relación tiene esto con correlación espuria? Considere la siguiente regresión $y_t = \alpha + x_t \gamma + \mu_t$ donde algunos elementos de x_t pueden ser I(1). Si μ_t no es I(0), entonces como veremos es posible y muy probable que la regresión arroje resultados espurios.

Supongamos que y_t y x_t son I(1) y hacemos la regresión anterior. Los estimadores mínimos cuadrados son:

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\gamma}_T \end{bmatrix} = \begin{bmatrix} T & \sum x_t \\ \sum x_t & \sum x_t' x_t \end{bmatrix}^{-1} \begin{bmatrix} \sum y_t \\ \sum x_t' y_t \end{bmatrix} \quad (6.19)$$

⁷ L. Ohanian ("The Spurious Effects of Unit Roots on Vector Autoregressions: A Monte Carlo Study", *Journal of Econometrics*, 39: 251-266, 1988) demuestra que si se añade un proceso paseo aleatorio a un VAR con variables estacionarias, se encuentra causalidad espuria en 20% de los casos.

Un test F hecho sobre cualquier hipótesis del tipo $R\gamma=r$ tiene la siguiente forma:

$$F_T = (R\hat{\gamma}_T - r)' \left(\hat{\sigma}_T^2 \begin{bmatrix} 0 & R \end{bmatrix} \begin{bmatrix} T & \sum x_t \\ \sum x_t & \sum x_t' x_t \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ R \end{bmatrix} \right)^{-1} (R\hat{\gamma}_T - r) / m \quad (6.20)$$

al menos que exista un valor de γ tal que $y_t - x_t \gamma$ sea $I(0)$, es casi seguro que se va a rechazar H_0 a medida que T crece. Recuerden que si $y_t - x_t \gamma$ es $I(1)$, esta diferencia es una variable que crece en el tiempo, por lo que $R\gamma - r$ va a crecer rechazando H_0 .

Phillips (1986) demuestra que si dicho γ no existe:

- Los parámetros estimados por mínimos cuadrados son inconsistentes (dos muestras aleatorias de la misma regresión entregan parámetros distintos).
- El estimador de la varianza de los parámetros diverge (lo que hace que se vean “significativos”). Esto sucede porque $\hat{\sigma}_T^2 = \mu_t' \mu_t / (n-k)$ y μ tiene deriva. Así, se obtiene espuriamente que los parámetros son significativos.
- El test F (y el t) diverge a medida que T crece (para que no diverja debiese ser dividido por \sqrt{T} y no por T), lo que hace que la regresión se vea “significativa”.

Granger, Heung y Young (2001)⁸ demuestran empíricamente que el problema de correlación espuria es posible que ocurra también en series de tiempo estacionarias, pero por razones diferentes que las que producen el problema en variables integradas. Si las variables tienen correlación serial, las varianzas de los estimadores de una regresión estática estarán subestimadas, lo que distorsionará el tamaño del test, es decir, se van a encontrar parámetros significativamente distintos de cero con excesiva frecuencia.

Si las variables tienen correlación serial, el estimador de la varianza de los parámetros mediante OLS no será consistente. En particular, si la autocorrelación es positiva este estimador estará subestimado y consiguientemente la significancia de los parámetros estará sobrestimada. El resultado es que parámetros significativos serán encontrados con mayor frecuencia a la correcta. El ejemplo siguiente caracteriza el problema que está ocurriendo. Suponga que se estima la ecuación (1) con variables que siguen procesos tipo (2).

1. $y_t = \alpha + \beta x_t + \epsilon_t$, donde ϵ_t es el error de la regresión
2. $x_t = \theta_x x_{t-1} + e_t^x$ y $y_t = \theta_y y_{t-1} + e_t^y$, donde e_t^x, e_t^y son innovaciones i.i.d. con media cero.

⁸ “Spurious Regressions With Stationary Series”, C.W.J. Granger, N. Hyung, y Y. Jeon, *Applied Economics*, 2001, 33:899- 904.

Al estimar la ecuación (1) por OLS, el estadígrafo t de β no convergerá a una distribución normal estándar a medida que T aumenta si el modelo no tiene considerado el problema de correlación serial. Al hacer inferencia, se estará asumiendo que el estadígrafo t converge a (3), cuando en realidad lo está haciendo a (4).

$$3. \quad t_{\hat{\beta}} = \frac{\hat{\beta}}{\hat{\sigma}_{\beta}} \rightarrow N(0,1)$$

$$4. \quad t_{\hat{\beta}} = \frac{\hat{\beta}}{\hat{\sigma}_{\beta}} \rightarrow N\left(0, \frac{1 - \theta_x^2 \theta_y^2}{(1 - \theta_x \theta_y)^2}\right)$$

La ecuación (4) identifica claramente el problema y muestra que la correlación espuria depende directamente del coeficiente de autocorrelación de las series. El cuadro 6.1 presenta una simulación de Montecarlo para dimensionar la importancia del fenómeno.

Correcciones arbitrarias, como por ejemplo la de White, no solucionan el problema (ya que no es un problema de heterocedasticidad). Otra corrección, como la de Bartlett corrige en parte el problema, excepto para coeficientes autorregresivos muy altos (mayor a 0.9).

La gravedad de la correlación espuria depende muy marginalmente del tamaño de muestra, pero no de la distribución de los residuos (normal, exponencial, Laplace, uniforme y t), excepto si éstos se distribuyen Cauchy, en donde el problema de correlación espuria disminuye al aumentar el tamaño de muestra. Naturalmente, el grado de correlación espuria aumenta a medida que aumenta el coeficiente autorregresivo o la persistencia de las series.

Cuadro 6.1
Tamaño real del test de significancia para un tamaño nominal de 5%
Replicaciones = 1.000

Coeficiente de Autocorrelación	T = 50	T = 100	T = 1.000
0.00	6.2	6.6	4.9
0.50	16.0	11.3	12.2
0.90	41.9	52.9	50.2
0.95	54.1	59.3	66.3
0.99	63.1	71.2	84.0

6.02 Noción de cointegración

Supongamos que tenemos un vector de $(n \times 1)$ series que son $I(1)$. Si existe una combinación lineal de éstas que sea $I(0)$, entonces las series cointegran. Por ejemplo, sea

$$\begin{aligned} y_t^1 &= \gamma y_t^2 + \mu_t^1 \\ y_t^2 &= y_{t-1}^2 + \mu_t^2 \end{aligned} \quad (6.21)$$

con μ_t^1 y μ_t^2 i.i.d. ruidos blancos no correlacionados. Obviamente, y_t^2 es un paseo aleatorio. Tomemos diferencia de la ecuación (6.21) para obtener:

$$\begin{aligned} \Delta y_t^1 &= \gamma \Delta y_t^2 + \mu_t^1 - \mu_{t-1}^1 \\ \Delta y_t^2 &= \mu_t^2 \end{aligned} \quad (6.22)$$

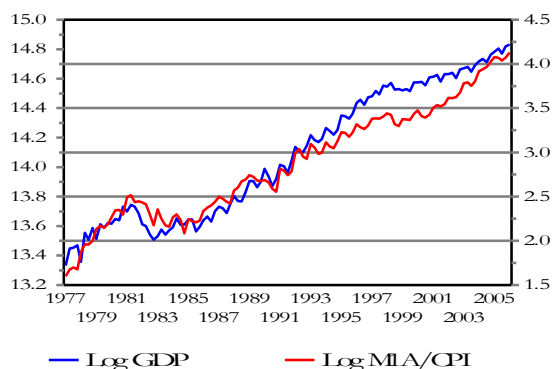
el lado derecho de la primera ecuación en (6.22) puede ser escrito como $\gamma \mu_t^2 + \mu_t^1 - \mu_{t-1}^1$, es decir es un MA(1) estacionario al que llamamos v_t con $\theta \neq -1$. Entonces:

$$\Delta y_t^1 = v_t + \theta v_{t-1} \quad (6.23)$$

Así, mientras cada serie por separado es $I(1)$, una combinación lineal de ellas es $I(0)$ y las variables cointegran.

Este es un resultado excepcional: aún cuando dos series están gobernadas por cambios permanentes, existe una relación de equilibrio entre ellas (γ). Veamos un ejemplo clásico: la relación dinero-producto en Chile.⁹ Ambas series son típicamente caracterizadas como $I(1)$, sin embargo, mantienen su relación de largo plazo.

Figura 6.1
Dinero y Actividad Económica en Chile, 1986-1998



⁹Ver, por ejemplo, R. Soto y M. Tapia (2000), "Cointegración Estacional en la Demanda de Dinero", *Economía Chilena*, 3(3):57-72.

El modelo de la ecuación (6.22) puede ser reescrito en términos matriciales. Defina los errores de predicción de las variables en las ecuaciones (6.21) como $\epsilon_t^2 = \mu_t^2$ y $\epsilon_t^1 = \gamma \mu_t^2 + \mu_t^1$. Entonces, $\gamma \mu_t^2 + \mu_t^1 - \mu_{t-1}^1 = \epsilon_t^1 - (\epsilon_{t-1}^1 - \gamma \epsilon_{t-1}^2) = (1-L)\epsilon_t^1 + \gamma L \epsilon_t^2$. Por lo que:

$$\begin{bmatrix} \Delta y_t^1 \\ \Delta y_t^2 \end{bmatrix} = \psi(L) \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \end{bmatrix} \text{ con } \psi(L) = \begin{bmatrix} 1-L & \gamma L \\ 0 & 1 \end{bmatrix} \quad (6.24)$$

es decir, parece ser un VAR(0). Nótese, sin embargo, que esta representación no es invertible, porque tiene una raíz unitaria en $\psi(1)$, es decir $|\psi(1)|=0$ y $\psi(1)^{-1}$ no existe.

La intuición y conclusión de este resultado es directa. La representación VAR no es una buena representación del modelo en la ecuación (6.23) porque el nivel de y_t^2 contiene información útil para predecir y_t^1 , más allá de la contenida en los cambios de y_t^2 .

Reescribamos el VAR de modo de incluir dicha información.

$$\begin{bmatrix} \Delta y_t^1 \\ \Delta y_t^2 \end{bmatrix} = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1}^1 \\ y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \gamma \mu_t^2 + \mu_t^1 \\ \mu_t^2 \end{bmatrix} \quad (6.25)$$

Esto quiere decir que es importante (necesario) incluir rezagos del nivel de las variables para representar adecuadamente el modelo.

Un ejemplo pintoresco que ayuda a entender la cointegración es el de dos borrachitos que una noche se encuentran a la salida de un bar. Como es bien sabido, la trayectoria de cada borrachito queda adecuadamente descrita por una caminata aleatoria. Si la calle tiene pendiente, la trayectoria tendrá, además, deriva. Dejados a su voluntad, los borrachitos caminarán por la calle de forma impredecible, salvo por la deriva. Suponga ahora que una mano bondadosa ata un resorte de largo γ a los borrachitos, de modo que no se pierdan entre sí. Cuando los borrachitos se separan, el resorte se estira y los atrae; cuando se acercan, el resorte se contrae y los separa. De este modo, observar a la mañana siguiente que uno de los borrachitos está en determinada parte de la ciudad contiene importante información para inferir dónde está el otro borrachito.

Note, en primer lugar que, en valor esperado la distancia entre ellos es fija (γ). Es por ello que, viendo a uno de los borrachitos se puede inferir la ubicación del otro.

6.03 Cointegración

Supongamos que tenemos un vector y_t de $(n \times 1)$ series que son $I(1)$. Si existe una combinación lineal a de éstas que sea $I(0)$, entonces las series cointegran.¹⁰

Nótese que la representación no es única. Si a es el vector de cointegración tal que $a' y_t$ es $I(0)$, entonces $ba' y_t$ es también un vector de cointegración si b es un escalar. Por ello, deberemos normalizar el vector de cointegración.

Es posible que haya más de un vector de cointegración en un sistema con n variables. Es posible encontrar a_1 y a_2 tal que $a_1' y_t$ y $a_2' y_t$ son $I(0)$ y se cumple que a_1 y a_2 son linealmente independientes.

De hecho, podría haber muchos vectores de cointegración, tal que:

$$A = \begin{bmatrix} a_1' \\ a_2' \\ \vdots \\ a_k' \end{bmatrix} \quad (6.26)$$

Si $A' y_t$ es $I(0)$, entonces A es la base del espacio de vectores de cointegración.

Veamos la representación MA de un sistema dinámico del tipo $\Delta y_t = \delta + \mu_t$ y supongamos que μ_t tiene la siguiente descomposición de Wold:

$$\mu_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots = \psi(L) \epsilon_t \quad (6.27)$$

con $E(\epsilon_t) = 0$, $E(\epsilon_t, \epsilon_\tau) = \Omega \quad \forall \quad t = \tau \quad \text{y} \quad 0 \quad \forall \quad t \neq \tau$.

Sea $\psi(1) \equiv I_n + \psi_1 + \psi_2 + \psi_3 \dots$. La cointegración se obtiene si $A' y_t$ es $I(0)$, entonces $A' \psi(1) = 0$. Para verificar esto recordemos que, si $\sum_s \psi_s$ es absolutamente sumable, entonces

$$\begin{aligned} y_t &= y_0 + \delta t + \mu_1 + \mu_2 + \dots + \mu_t \\ &= y_0 + \delta t + \psi(1) \cdot (\epsilon_1 + \epsilon_2 + \dots + \epsilon_t) + \eta_T - \eta_0 \end{aligned} \quad (6.28)$$

con η estacionario.

Multiplicando la ecuación (6.28) por A obtenemos:

¹⁰ Este es el resultado básico del célebre trabajo de R Engle y C. W. Granger ("Co-integration and error correction: representation, estimation, and testing", *Econometrica* 55, 251-76, 1987) que usaremos frecuentemente en este capítulo.

$$A' y_t = A' y_0 + A' \delta t + A' \psi(1) \cdot (\epsilon_1 + \epsilon_2 + \dots + \epsilon_t) + A' \eta_T - A' \eta_0 \quad (6.29)$$

para que y_t sea cointegrado se requiere que $A' \psi(1) = 0$ pero, además, que $A' \delta = 0$. Es decir, A debe ser tal que remueva efectos permanentes $\psi(1)$ y a la vez, si algunas variables tienen deriva, que elimine dichos derivas también (δ).

Otro resultado interesante de la ecuación (6.28) es que un VAR en primeras diferencias finito no es una adecuada descripción de un sistema que cointegra. Ello se deriva de $A' \psi(1) = 0$, porque el determinante $|\psi(z)| = 0$ en $z=1$, por lo que $\psi(L)$ no es invertible.

Hay dos maneras interesantes de representar un sistema co-integrado. La representación triangular de Phillips (1986) y la representación de tendencias comunes de Stock y Watson (1988)¹¹. Ambas nos ayudan a pensar en la naturaleza de la cointegración y por ello las desarrollamos en detalle a continuación.

Representación triangular

Retomemos la matriz A de la ecuación (6.26):

$$A = \begin{bmatrix} a_1' \\ a_2' \\ \vdots \\ a_k' \end{bmatrix} = \begin{bmatrix} 1 & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{h1} & a_{h2} & a_{h3} & \dots & a_{hn} \end{bmatrix} \quad (6.30)$$

multiplicamos la primera fila por a_{21} y la restamos de la segunda fila. Sigue siendo un vector de cointegración. Luego, multiplicamos la primera fila por a_{31} y lo restamos de la tercera fila. Esto se puede hacer para las h filas usando a_{h1} . Obtenemos la siguiente matriz:

$$A_1 = \begin{bmatrix} a_1' \\ a_2' \\ \vdots \\ a_k' \end{bmatrix} = \begin{bmatrix} 1 & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22}^* & a_{23}^* & \dots & a_{2n}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{h2}^* & a_{h3}^* & \dots & a_{hn}^* \end{bmatrix} \quad (6.31)$$

Ahora dividimos la segunda fila por a_{22}^* y la multiplicamos por a_{12} (sigue siendo un vector de cointegración) y se resta de la primera fila. La fila resultante la restamos de la primera fila. Si $a_{22}^* = 0$, mueva la fila al final de la matriz. Ahora

¹¹ Stock, J.H., and M. W. Watson (1988), "Testing for common trends", *Journal of the American Statistical Association*, 83: 1097-107.

multiplique la segunda fila por a_{32}^* y réstela de la tercera fila (sigue siendo un vector de cointegración), lo que se repite con los h filas. Se obtiene la siguiente matriz

$$A_2 = \begin{bmatrix} a_1' \\ a_2' \\ \vdots \\ a_k' \end{bmatrix} = \begin{bmatrix} 1 & 0 & a_{13}^{**} & \dots & a_{1n}^{**} \\ 0 & 1 & a_{23}^{**} & \dots & a_{2n}^{**} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & a_{h3}^{**} & \dots & a_{hn}^{**} \end{bmatrix} \quad (6.32)$$

Podemos repetir el procedimiento con las h filas de A hasta obtener:

$$A = \begin{bmatrix} a_1' \\ a_2' \\ \vdots \\ a_k' \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 & -\gamma_{1,b+1} & \dots & -\gamma_{1,n} \\ 0 & 1 & \dots & 0 & -\gamma_{2,b+1} & \dots & -\gamma_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & -\gamma_{b,b+1} & \dots & -\gamma_{b,n} \end{bmatrix} \quad (6.33)$$

$$= \begin{bmatrix} I_n & -\Gamma' \end{bmatrix}$$

donde Γ' es una matriz ($h \times g$) con $g = n - h$.

Sean $z_t = A' y_t$ los residuos de las relaciones de cointegración. z_t es $(h \times 1)$. Como z_t es estacionario, existe $E[z_t] = \mu_t^*$. Sea $z_t^* = z_t - \mu_t^*$. Particionamos y_t en $[y_t^1 \ y_t^2]$. Entonces:

$$z_t = A' y_t$$

$$z_t^* - \mu_t^* = [I_h - \gamma'] \begin{bmatrix} y_t^1 \\ y_t^2 \end{bmatrix} \quad (6.34)$$

es decir,

$$y_t^1 = \gamma' y_t^2 + \mu_1^* + z_t^* \quad (6.35)$$

en tanto que la representación de y_t^2 corresponde a las g filas de la representación en primeras diferencias de las variables $I(1)$ que no cointegran.

$$\Delta y_t^2 = \delta_2 + \mu_t^2 \quad (6.36)$$

Las ecuaciones (6.35) y (6.36) son la representación triangular de Phillips, donde μ_t^2 y z_t^* son innovaciones con media cero. Nótese que hay h relaciones de cointegración, las variables contenidas en y_t^1 cointegran, en cambio las variables y_t^2 son $I(1)$ y no cointegran.

Un ejemplo de esta representación común en la literatura es la paridad del poder de compra (PPP), donde si ésta se verifica podemos escribir el sistema dinámico como:

$$\begin{aligned} p_t &= \gamma_1 s_t + \gamma_2 p_t^* + \mu_1^* + \varepsilon_t^* \\ \Delta s_t &= \delta_s + \mu_t^s \\ \Delta p_t^* &= \delta_p + \mu_t^{p^*} \end{aligned} \quad (6.37)$$

La hipótesis de PPP en el modelo equivale a que los parámetros γ sean 1.

Representación de tendencias comunes

Acabamos de ver que si y_t contiene h relaciones de cointegración, entonces es posible representarlo como un sistema triangular, con $[\varepsilon_t^*, \mu_t]$ estacionarios con media cero. Supongamos que

$$\begin{bmatrix} \varepsilon_t^* \\ \mu_t^2 \end{bmatrix} = \sum_{s=0}^{\infty} \begin{bmatrix} H_s \varepsilon_{t-s} \\ J_s \varepsilon_{t-s} \end{bmatrix} \quad (6.38)$$

con ε_t ruido blanco y se cumple sumabilidad absoluta. Entonces,

$$\begin{aligned} y_t^2 &= y_0^2 + \delta t^2 + \mu_1^2 + \mu_2^2 + \dots + \mu_t^2 \\ &= y_0^2 + \delta t^2 + J(1) \cdot (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t) + \eta_T^2 - \eta_0^2 \end{aligned} \quad (6.39)$$

con $J(1) = J_0 + J_1 + J_2 + \dots$, $\eta^2 = \sum_s s \alpha s^2 \varepsilon_{t-s}$ y $\alpha s^2 = -J_{s+1} + J_{s+2} + J_{s+3} + \dots$.

Como ε_t es ruido blanco, entonces $J(1)\varepsilon_t$ es ruido blanco pero $\xi_t^2 = J(1)\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t$ es un paseo aleatorio. Sustituyendo en la ecuación (6.39) tenemos:

$$\begin{aligned} y_t^2 &= y_0^2 - \eta_0^2 + \delta t^2 + \xi_t^2 + \eta_t^2 \\ &= \mu_t^2 + \delta t^2 + \xi_t^2 + \eta_t^2 \end{aligned} \quad (6.40)$$

Sustituyendo la ecuación (6.40) en la representación de y_t^1 :

$$y_t^1 = \bar{\mu}_t^1 + \gamma'(\delta t^2 + \xi_t^2) + \bar{\eta}_t^1 \quad (6.41)$$

con $\bar{\mu}_t^1 = \mu_1^* + \gamma' \bar{\mu}_T^2$ y $\bar{\eta}_t^1 = \varepsilon_t^* + \gamma' \eta_t^2$.

Es decir, ambas (vectores) variables tienen g tendencias determinísticas iguales (incluidas en δt^2) y combinaciones lineales de g paseos aleatorios comunes (en ξ_t^2).

Veamos la intuición de “bolsillo” de la representación de tendencias comunes. Suponga que y_t y x_t son dos variables que tiene un componente transitorio, z , y permanente, p , y que ambas cointegran. Entonces,

$$\theta x_t + \phi y_t = \epsilon_t \quad (6.42)$$

con $\epsilon_t \rightarrow I(0)$. Pero $y_t = z_t^y + p_t^y$ y $x_t = z_t^x + p_t^x$ entonces la ecuación (6.42) es

$$\begin{aligned} \phi(z_t^y + p_t^y) + \theta(z_t^x + p_t^x) &= \epsilon_t \\ [\phi z_t^y + \theta z_t^x] + [\phi p_t^y + \theta p_t^x] &= \epsilon_t \end{aligned} \quad (6.43)$$

el primer término es obviamente estacionario; el error es estacionario por definición. Por lo tanto, el segundo término debe ser estacionario. La única forma que se cumpla el modelo es que en el largo plazo el segundo término no “acumule” distancia (brecha creciente). Entonces, cuando la muestra crece, debe converger a una constante

$$\phi p_t^y + \theta p_t^x \rightarrow cte \quad (6.44)$$

y por lo tanto $p_t^y = \frac{cte}{\phi} - \frac{\theta}{\phi} p_t^x$.

Cointegración en la representación VAR

El VAR en diferencias no es consistente con un sistema cointegrado. Sin embargo, un VAR en niveles puede serlo. Retomemos la ecuación (6.1)

$$(I_n - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t = \alpha + \epsilon_t \quad (6.45)$$

o $\phi(L) y_t = \alpha + \epsilon_t$. Supongamos que Δy_t tiene una representación de Wold:

$$(1-L) y_t = \delta + \psi(L) \epsilon_t \quad (6.46)$$

Premultiplicamos por $\phi(L)$ y obtenemos¹²:

$$(1-L) \phi(L) y_t = \phi(1) \delta + \phi(L) \psi(L) \epsilon_t \quad (6.47)$$

sustituyendo por $\phi(L)$ de la ecuación (6.45) tenemos:

$$(1-L) \epsilon_t = \phi(1) \delta + \phi(L) \psi(L) \epsilon_t \quad (6.48)$$

Dos cosas deben cumplirse. Primero, $\phi(1) \delta = 0$. Es decir, el efecto acumulado de la deriva debe ser cero. Segundo, $(1-L) I_n$ corresponde a los polinomios contenidos

¹² Note que $\phi(L) \delta = (I_n - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) \delta = \phi(1) \delta$

en $\phi(L)\Psi(L)$. Entonces, para todo z se cumplirá que $(1-L)I_n$ equivale a $\phi(z)\Psi(z)$ y, más aún, para $z=1$ debe cumplirse que $\phi(1)\Psi(1)=0$. Es decir, que el efecto acumulado de los shocks debe ser cero.

Para ver más fácilmente el rol que juega la cointegración, nos concentramos en una fila cualquiera de $\phi(1)$ a la que llamamos π' . Se cumple, $\pi'\Psi(1)=0'$ y $\pi'\Delta=0$. Luego π' es un vector de cointegración.

Si $A=[a_1, a_2, \dots, a_b]$ era el espacio de los vectores de cointegración, entonces existe una combinación lineal b tal que:

$$\begin{aligned}\pi &= [a_1, a_2, \dots, a_b] b \\ \pi' &= b' A'\end{aligned}\tag{6.49}$$

si aplicamos este razonamiento a cada fila de $\phi(1)$, tendremos que existe una matriz de tamaño $(n \times h)$ que contiene todos los vectores de cointegración:

$$\phi(1) = BA'\tag{6.50}$$

Cointegración en representación de corrección de errores

Usemos la representación del VAR en niveles:

$$y_t = \xi_1 \Delta y_{t-1} + \xi_2 \Delta y_{t-2} + \dots + \xi_{p-1} \Delta y_{t-p+1} + \rho y_{t-1} + \alpha + \epsilon_t\tag{6.51}$$

y recordemos que $\rho = \phi_1 + \phi_2 + \dots + \phi_p$.

Restando y_{t-1} de la ecuación (6.51) tenemos:

$$\Delta y_t = \xi_1 \Delta y_{t-1} + \dots + \xi_{p-1} \Delta y_{t-p+1} + (\rho - I_n) y_{t-1} + \alpha + \epsilon_t\tag{6.52}$$

pero $(\rho - I_n) = -\phi(1) = -BA'$. Entonces:

$$\Delta y_t = \xi_1 \Delta y_{t-1} + \dots + \xi_{p-1} \Delta y_{t-p+1} + \alpha - BA' y_{t-1} + \epsilon_t\tag{6.53}$$

Sea $A' y_{t-1} = z_{t-1}$, entonces:

$$\Delta y_t = \xi_1 \Delta y_{t-1} + \dots + \xi_{p-1} \Delta y_{t-p+1} + \alpha - B z_{t-1} + \epsilon_t\tag{6.54}$$

Esto significa que los cambios en una variable dependen de sus propios cambios pasados, los cambios pasados en las otras variables del VAR, y los niveles de z_t rezagados. ¿qué hay en z_t ?

Si pensamos en el VAR bivariado que usamos antes, en $z_{t-1} = y_{t-1}^1 - \gamma y_{t-1}^2$, es decir, las desviaciones del equilibrio impuesto por la cointegración. De esa manera, llegamos al teorema de representación de Granger, que sintetiza todas las anteriores representaciones:

Teorema de representación de Granger¹³

Considere un vector $(n \times 1)$ y_t donde Δy_t satisface la representación de Wold en la ecuación (6.43) para un ϵ_t que es ruido blanco, con matriz de covarianza positiva definida y absolutamente sumable. Suponga que hay exactamente h relaciones de cointegración entre los componentes de y_t . Entonces, existe una matriz $(h \times n)$ A tal que el vector z_t que cumple $z_t \equiv A' y_t$ es estacionario y la matriz A cumple $A'\Psi(1)=0$. Si el proceso puede ser representado por un VAR(p) en niveles, entonces existe una matriz $(n \times h)$ B tal que $\Phi(1)=BA'$ y existen matrices de $(n \times n)$ tales que $\Delta y_t = \xi_1 \Delta y_{t-1} + \dots + \xi_{p-1} \Delta y_{t-p+1} + \alpha - Bz_{t-1} + \epsilon_t$

6.04 Tests de cointegración

Cuando el vector de cointegración es conocido.

En contadas ocasiones, la teoría impone un vector de cointegración particular (ex-ante), que podemos testear. Por ejemplo, la PPP impone una relación particular entre los precios domésticos (p), internacionales (p^*) y el tipo de cambio (s).

$$z_t = p_t - p_t^* - s_t \quad (6.55)$$

Nótese que el vector de cointegración propuesto por la teoría es $[1, -1, -1]$ y que, si las variables son I(1) y hay cointegración, z_t debe ser I(0).

La estrategia sería:

1. verificar que las variables sean I(1)
2. verificar que z_t es I(0).

Podemos usar tests de raíces unitarias tipo Dickey-Fuller o Phillips-Perron directamente si el vector de cointegración es conocido. El problema de equivalencia observacional que discutimos para los tests de raíces unitarias, obviamente, estará presente en este caso.

¹³ Esta representación se deriva del trabajo de Granger (1983): "Co-Integrated Variables and Error-Correcting Models", mimeo, UCSD Discussion Paper 83-13.

Cuando el vector de cointegración no es conocido y las variables no tienen deriva.

Si el vector de cointegración no es conocido, hay que estimarlo. Esto produce algunas variaciones de lo señalado más arriba. Una alternativa para estimar el vector de cointegración es usar el método de mínimos cuadrados. ¿Es válido este estimador?

Bajo la hipótesis nula que existe un vector de cointegración, entonces el estimador de mínimos cuadrados se refiere a $\tilde{z}_t \rightarrow I(0)$, por lo que:

$$T^{-1} \sum_{i=1}^T \tilde{z}_t^2 = T^{-1} \sum_{i=1}^T (a' y_t)^2 \rightarrow E(\tilde{z}_t^2) \quad (6.56)$$

Bajo la hipótesis alternativa (no existe un vector de cointegración), entonces el estimador de mínimos cuadrados se refiere a $\tilde{z}_t \rightarrow I(1)$, por lo que:

$$T^{-2} \sum_{i=1}^T (a' y_t)^2 \rightarrow \lambda^2 \int_0^1 [W(r)]^2 dr \quad (6.57)$$

donde λ es un parámetro determinado por la función de autocovarianzas de $\Delta \tilde{z}_t$. Así, si a no es un vector de cointegración, el estadígrafo calculado según la ecuación (6.57) diverge hacia $+\infty$.

Entonces, una manera práctica de obtener un estimador consistente de a (de hecho, superconsistente) consiste en minimizar la ecuación (6.57) sujeto a que a esté normalizado (recuerde que si a es el vector de cointegración, ba también lo es).

Una normalización conveniente es que el primer elemento de a sea 1 (el resto es y_2, y_3, \dots, y_n). Así, el objetivo es escoger y_2, y_3, \dots, y_n tal que $a' y_t$ sea mínimo. Pero, ¿qué hay en esta minimización?

$$T^{-1} \sum_{i=1}^T (a' y_t)^2 = T^{-1} \sum_{i=1}^T (y_t^1 - y_2 y_t^2 - \dots - y_n y_t^n)^2 \quad (6.58)$$

Y, naturalmente, esto es equivalente a hacer una estimación de mínimos cuadrados.

En términos generales, si usamos la representación triangular del sistema:

$$\begin{aligned} y_t^1 &= \alpha + y' \cdot y_t^2 + \tilde{z}_t^* \\ \Delta y_t^2 &= \mu_t^2 \end{aligned} \quad (6.59)$$

Nótese que hemos impuesto que las variables no tienen deriva. Sea

$$\begin{bmatrix} \tilde{z}_t^* \\ \mu_t^2 \end{bmatrix} = \Psi^*(L) \epsilon_t \quad (6.60)$$

y supongamos que la secuencia de matrices $\{s\Psi_s^*\}$ es absolute summable. Entonces los estimadores de los parámetros de la ecuación (6.60) son:

$$\begin{bmatrix} \hat{\alpha}_T \\ \hat{\gamma}_T \end{bmatrix} = \begin{bmatrix} T & \sum y_t'^2 \\ \sum y_t^2 & \sum y_t'^2 y_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_t^1 \\ \sum y_t'^2 y_t^1 \end{bmatrix} \quad (6.61)$$

particionamos $\Psi^*(1)$ en $\Psi^*(1)\sqrt{\text{diag}(\epsilon_t'\epsilon_t)} = \begin{bmatrix} \lambda_1^* \\ \lambda_2^* \end{bmatrix}$ con $\lambda^*(1 \times n)$ y $\Lambda^*(g \times n)$. Es decir, entre los parámetros del vector de cointegración y la constante. Entonces,

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\gamma}_T - \gamma) \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \left[\int W(r) dr \right]' \lambda_2^* \\ \lambda_2^* \int W(r) dr & \lambda_2^* \left[\int W(r) W(r)' dr \right] \lambda_2^* \end{bmatrix}^{-1} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (6.62)$$

$$\text{donde } \lambda^* = (I_n - \xi_1 - \xi_2 - \dots - \xi_{p-1})^{-1} P \text{ y } \begin{aligned} b_1 &= \lambda_1^* W(1) \\ b_2 &= \lambda_2^* \left[\int_0^1 W(r)' d[W(r)] \right] \lambda_1^* + \sum_{p=0}^{\infty} E(\mu_2' \tilde{z}_{t+p}^*) \end{aligned}$$

¿Qué aprendemos?

- Los estimadores de mínimos cuadrados del vector de cointegración convergen a h_2 . Su distribución no es asintóticamente normal. Ésta depende de la existencia de correlación entre μ y los Δy^2 . En tal caso, h_2 no tendrá media cero y por lo tanto el estimador será sesgado. No obstante, el estimador será consistente, pues el sesgo desaparece cuando $T \rightarrow \infty$. De hecho son superconsistentes $O(T^{-1})$.
- Se puede demostrar que, si hay cointegración, el error cuadrático medio muestral converge $T^{-1} \sum_{t=1}^{\infty} \hat{\mu}_{t,T}^2 \rightarrow E[\sigma^2]$. Por otro lado, como la variable es integrada $T^{-1} \sum_{t=1}^{\infty} (y_t^1 - \bar{y}^1)^2 \rightarrow \infty$. Por ello, el R^2 converge a 1 cuando la muestra crece.
- Se puede generalizar esto para el caso en que las variables tienen deriva.

- Si no hay tendencia determinística en el verdadero vector de cointegración, estimadores son superconsistentes aún si tratamos de estimar una tendencia determinística.
- Note el rol que juegan los valores fuera de la diagonal en la matriz de momentos de la ecuación (6.62)

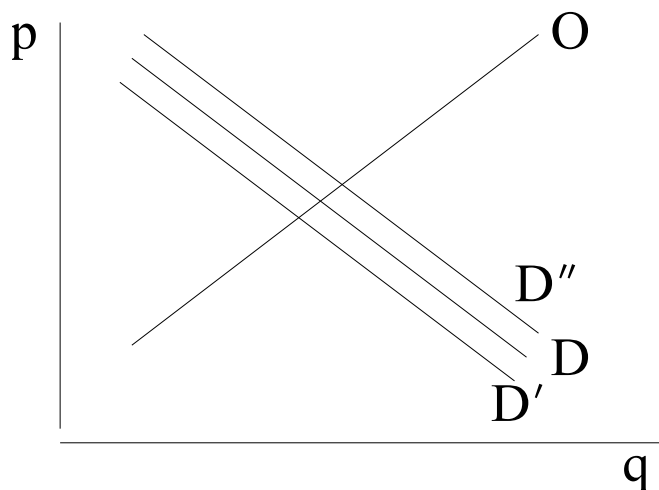
¿Por qué la ecuación de cointegración “funciona”? Supongamos que el modelo es el típico que causa sesgo de simultaneidad:

$$\begin{aligned} q_t^s &= \gamma p_t + \epsilon_t^s \\ q_t^d &= \beta p_t + \epsilon_t^d \end{aligned}$$

Si la varianza de los shocks de oferta $V(\epsilon_t^s)$ es finita y la varianza de los shocks de demanda $V(\epsilon_t^d)$ se va a infinito (como sería el caso si fuese una variable integrada), entonces es posible estimar consistentemente γ por mínimos cuadrados aún si hay sesgo de simultaneidad.

En el caso del ejemplo en la ecuación (6.64), note que si el vector de cointegración es $[1, -\gamma]$, la varianza de ϵ_t^s es finita, porque el residuo es $I(0)$. Y como el vector de cointegración no es $[1, -\beta]$, la varianza de ϵ_t^d se va a infinito. Así, la cointegración actúa como una restricción de identificación.

Figura 6.2
Cointegración y Sesgo de Simultaneidad



6.05 Otras preguntas interesantes

Qué pasa si hay más de un vector de cointegración, pero nosotros asumimos que sólo hay uno? El estimador mínimos cuadrados sigue siendo superconsistente. ¿Pero de cuál vector? De aquel cuyos residuos son ortogonales a todas las otras posibles combinaciones de variables $I(1)$ que hay en el modelo.

¿Qué estima el vector de cointegración cuando no hay vector de cointegración verdaderamente? Obviamente, una correlación espuria (parámetros inconsistentes, residuos no estacionarios).

¿Es irrelevante el orden de normalización del vector de cointegración? Si hacemos la siguiente normalización $y_t^1 = \alpha + \beta y_t^2 + \epsilon_t$ obtenemos lo mismo que si hacemos $y_t^2 = \pi + \gamma y_t^1 + \epsilon_t$? Sólo si el R^2 es 1, se obtiene el mismo vector de cointegración. Dos preguntas, (1) ¿por qué? y (2) ¿si cointegra en una de ellas, cointegra en la otra?

6.06 Test de hipótesis cuando las variables cointegran

Consideremos el modelo

$$\begin{aligned} y_t^1 &= \alpha + \gamma y_t^2 + \varepsilon_t \\ y_t^2 &= y_{t-1}^2 + \mu_t \end{aligned} \quad (6.63)$$

Supongamos que ε_t y μ_t son procesos Gaussianos y no-correlacionados (los dos supuestos son importantes para la distribución de los tests), entonces

$$\begin{bmatrix} \varepsilon_t \\ \mu_t \end{bmatrix} \rightarrow i.i.d.N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & \Omega_{22} \end{bmatrix} \right) \quad (6.64)$$

en este caso, condicional en las y_t^2 , los estimadores de α y γ son asintóticamente normales.

Consideremos ahora un test de restricción tipo $R\beta = r$, donde R_α , R_γ y r son matrices de escalares. Es directo probar que el test de Wald estándar es:

$$\begin{aligned} & R_\alpha \hat{\alpha} + R_\gamma \hat{\gamma} \\ &= r (R_\alpha \hat{\alpha} + R_\gamma \hat{\gamma} - r)' \left[\hat{\sigma}^2 [R_\alpha \ R_\gamma] \begin{bmatrix} \sum y_t^2 \\ \sum y_t^2 y_t^2' \end{bmatrix}^{-1} \begin{bmatrix} R_\alpha' \\ R_\gamma' \end{bmatrix} \right]^{-1} \cdot (R_\alpha \hat{\alpha} + R_\gamma \hat{\gamma} - r) / m \end{aligned} \quad (6.65)$$

y se distribuye como una $F\left[\frac{m}{T-n}\right]$. Es decir, a pesar de lo complejo de la cointegración, la regresión se hace por mínimos cuadrados ordinarios y el test F tiene la misma distribución (no se necesitan otras tablas) y se hace igual que si las variables fuesen $I(0)$. Naturalmente, a priori no sabemos si z_t es Gaussiano.

¿Qué pasa si las variables no son Gaussianas? Si y_t^2 es un paseo aleatorio y z_t es ruido blanco, el test de una transformación estabilizadora de los parámetros es asintóticamente normal.

$$\begin{bmatrix} T^{1/2}(\hat{\alpha} - \alpha) \\ T(\hat{y} - y) \end{bmatrix} \rightarrow N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & P_{22} \int W_2(r) dr \\ P_{22} \int W_2(r) dr & P_{22} \int W_2(r) \cdot W_2(r)' dr P_{22}' \end{bmatrix}\right) \quad (6.66)$$

donde $\Omega_{22} = P_{22} P_{22}'$. Este resultado se obtuvo en el capítulo 4 (ecuación 4.65).

De la ecuación (6.65) se desprende, de nuevo, que el test F se aplica. En particular si el test sólo tiene que ver con el vector de cointegración - $R_\alpha = 0$ -, entonces el test se distribuye con $\chi^2(m)$. Por otro lado, si sólo se refiere a las α , es decir si $R_y = 0$, entonces el test también es $\chi^2(m)$.

El test es inválido si las variables z_t y μ_t son correlacionados, pues las variables no tienden a la ecuación (6.70) sino a distribuciones que dependen de la forma de la correlación entre z_t y μ_t (aparecen términos tipo Λ_1 como en la ecuación 6.62).

Phillips y Loretan (1991)¹⁴ y Stock y Watson (1993)¹⁵ proponen incluir rezagos y adelantos (leads) de Δy_t^2 para controlar la correlación entre z_t y μ_t . Recuerden que ambas variables son $I(0)$, por lo que no es problema añadir los rezagos y adelantos. Como z_t y μ_t están correlacionados, podemos escribir

$$z_t = \sum_{s=-p}^p \beta_s' \mu_{t-s}^2 + \bar{z}_t \quad (6.67)$$

obviamente, \bar{z}_t no está correlacionado con μ_{t-s} para $|s| > p$. Recordemos que μ_t es Δy_t^2 , entonces la ecuación (6.63) se puede escribir como:

$$y_t^1 = \alpha + \gamma y_t^2 + \sum_{s=-p}^p \beta_s' \Delta y_{t-s}^2 + \bar{z}_t \quad (6.68)$$

¹⁴ Phillips, P.C.B. and Loretan, M. (1991) "Estimating Long-Run Economic Equilibria", *Review of Economic Studies*, 58: 407-36.

¹⁵ Stock, J.H. and M.W. Watson (1993) "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems", *Econometrica*, 61: 783-820.

entonces, el test es bastante simple. Estime la ecuación (6.68) por mínimos cuadrados, luego pruebe su hipótesis $R\gamma=r$ usando un test que se distribuye casi exactamente como una F. En realidad, necesitamos el test F calculado de la regresión ajustando por $\hat{\sigma}^2/\hat{\lambda}^2$ donde:

$$\begin{aligned}\hat{\sigma}^2 &= (T-p)^{-1} \sum_{t=p+1}^T \hat{\epsilon}_t^2 \\ \hat{\lambda}^2 &= \hat{\sigma} / (1 - \hat{\phi}_1 - \hat{\phi}_2 \dots - \hat{\phi}_p)\end{aligned}\tag{6.69}$$

con ϕ los coeficientes del modelo AR estimado en μ_t .

6.07 Estimación y Testeo de Cointegración en Sistemas de Ecuaciones

Existen algunas ventajas de estimar los vectores de cointegración como un sistema de ecuaciones simultáneas a la Johansen (1988)¹⁶:

- Es posible testear cuántos vectores de cointegración, h , existen.
- Se evita el problema de mala especificación derivado de imponer la normalización de un vector de cointegración equivocadamente. Recuerden que hicimos $a_{11}=1$ para normalizar, pero ¿qué pasa si la primera variable no forma parte de ningún vector de cointegración, $a_{11}=0$?

Antes de ver directamente la técnica de Johansen, vamos a desarrollar una herramienta práctica, llamada correlaciones canónicas.¹⁷ El objetivo es que, en el proceso de derivar los tests y sus distribuciones, se pueda seguir la pista –separadamente pero de forma integrada– a los dos tipos de variables de interés, aquellas que cointegran y aquellas que no lo hacen.

Correlaciones Canónicas

Sean y_t y x_t dos vectores de variables de orden $n_1 \times 1$ y $n_2 \times 1$, respectivamente, en desviaciones con respecto a la media. La matriz de varianzas y covarianzas entre las variables es:

¹⁶ Johansen, S. (1988) "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12: 231-254.

¹⁷ Correlaciones canónicas fueron utilizadas por primera vez por H. Hotelling en "Relations between Two Sets of Variates," *Biometrika*, 28:321-377, 1936. En la sección 2 del trabajo Hotelling señala que "The word 'canonical' is used in the algebraic theory of invariants with a meaning consistent with that of this paper." Es decir, "A canonical form is required to have two essential properties. Every object under consideration must have exactly one canonical form, and two objects that have the same canonical form must be essentially the same."

$$\begin{bmatrix} Ey_t' y_t & Ey_t' x_t \\ Ex_t' y_t & Ex_t' x_t \end{bmatrix} = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix} \quad (6.70)$$

Definimos dos variables aleatorias de orden $n \times 1$ (con $n = \min(n_1, n_2)$), tales que:

$$\begin{aligned} \eta_t &\equiv \mathfrak{H}' y_t \\ \xi_t &\equiv \wp' x_t \end{aligned} \quad (6.71)$$

Las matrices ξ' ($n \times n_2$) y \mathfrak{H}' ($n \times n_1$) se escogen tales que cumplan:

$$\begin{aligned} E(\eta_t' \eta_t) &= \mathfrak{H}' \Sigma_{YY} \mathfrak{H} = I_n \\ E(\xi_t' \xi_t) &= \wp' \Sigma_{XX} \wp = I_n \\ E(\xi_t' \eta_t) &= \wp' \Sigma_{XY} \mathfrak{H} = R \end{aligned} \quad (6.72)$$

donde $R = \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_n \end{bmatrix}$.

Es decir, cada elemento de ξ y η tienen varianza unitaria y son no correlacionados. Además, el elemento i -ésimo de η no está correlacionado con el j -ésimo elemento de ξ cuando $i \neq j$. Finalmente, se escoge ξ y η tales que $1 \geq r_1 \geq r_2 \geq \cdots \geq r_n \geq 0$

Cada uno de los r_i es la “correlación canónica poblacional” y se obtiene de manera simple usando los valores propios $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ y vectores propios (vectores propios, $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_n$) de:

$$\Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \quad (6.72)$$

Los vectores y valores propios satisfacen:

$$\Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \kappa_i = \lambda_i \kappa_i \quad (6.73)$$

pero también κ_i satisface la ecuación (6.77), por lo que normalizamos (¿les suena conocido?). A diferencia de componentes principales, la normalización no será del tipo $\kappa_i' \kappa_i = 1$, sino:

$$\kappa_i' \Sigma_{YY} \kappa_i = 1 \quad (6.74)$$

Una vez obtenido los $\lambda_1, \lambda_2, \dots, \lambda_n$ la matriz R corresponde a las raíces cuadradas de los λ_i

La intuición es directa: el primer par de *canonical variates*, η_1 y ξ_1 (los componentes de η y ξ) son aquellas combinaciones lineales de x_t e y_t tales que la correlación entre η_1 y ξ_1 es máxima. El segundo par de *canonical variates*, η_2 y ξ_2 son aquellas combinaciones lineales de x_t e y_t no correlacionadas con η_1 y ξ_1 , tales que la correlación entre η_2 y ξ_2 es máxima. Etc.

Derivación del Test de Johansen

Retomemos la siguiente representación VAR del sistema con h relaciones de cointegración.

$$\Delta y_t = \xi_1 \Delta y_{t-1} + \dots + \xi_{p-1} \Delta y_{t-p+1} + \xi_0 y_{t-1} + \alpha + \epsilon_t \quad (6.75)$$

con $E(\epsilon_t) = 0$ y $E(\epsilon_t' \epsilon_t) = \Omega I$.

En la ecuación (6.75) hay h relaciones de cointegración. Entonces, como ya vimos, $\xi_0 = -BA'$, donde B es $(h \times n)$ y A es $(h \times n)$. Si hay h relaciones de cointegración, entonces ξ_0 debe ser de rango h .

Consideremos ahora que si ϵ_t es un proceso Gaussiano, entonces el logaritmo de la función de verosimilitud condicional en la muestra $\{y_{-p+1}, y_{-p+2}, \dots, y_T\}$ es:

$$\begin{aligned} \mathcal{L}(\Omega, \xi_1, \dots, \xi_{p-1}, \alpha, \xi_0) = & (-Tn/2) \log(2\pi) - (T/2) \log|\Omega| \\ & - (1/2) \sum_{t=1}^T (\Delta y_t - \xi_1 \Delta y_{t-1} - \dots - \xi_{p-1} \Delta y_{t-p+1} - \xi_0 y_{t-1} - \alpha)' \\ & \times \Omega^{-1} (\Delta y_t - \xi_1 \Delta y_{t-1} - \dots - \xi_{p-1} \Delta y_{t-p+1} - \xi_0 y_{t-1} - \alpha) \end{aligned} \quad (6.76)$$

El objetivo es escoger $(\Omega, \xi_1, \dots, \xi_{p-1}, \alpha, \xi_0)$ para maximizar la ecuación (6.76) sujeto a que $\xi_0 = -BA'$.

Aunque se ve bastante complicado de resolver existe un algoritmo que permite optimizar la ecuación (6.76) más fácilmente (usando mínimos cuadrados para hacer más manejable el problema), que permite testear el número de vectores de cointegración, h , y que es insensible a la normalización. El algoritmo es el siguiente:

Paso 1: Haga la estimación VAR de Δy_t en $p-1$ rezagos y compute los residuos, $\hat{\mu}_t$

$$\Delta y_t = \hat{\pi}_0 + \hat{\Pi}_1 \Delta y_{t-1} - \dots - \hat{\Pi}_{p-1} \Delta y_{t-p+1} + \hat{\mu}_t \quad (6.77)$$

Paso 2: Haga la estimación VAR de y_{t-1} en $p-1$ rezagos y compute los residuos, \hat{v}_t

$$y_{t-1} = \theta_0 + \hat{\varepsilon}_1 \Delta y_{t-1} - \dots - \hat{\varepsilon}_{p-1} \Delta y_{t-p+1} + \hat{v}_t \quad (6.78)$$

Paso 3: Compute la correlación canónica de las matriz de varianzas y covarianzas de los residuos de la ecuación (6.77) y la ecuación (6.78)

$$\begin{aligned} \hat{\Sigma}_{vv} &= (1/T) \sum_{t=1}^T \hat{v}_t \hat{v}_t' \\ \hat{\Sigma}_{\mu\mu} &= (1/T) \sum_{t=1}^T \hat{\mu}_t \hat{\mu}_t' \\ \hat{\Sigma}_{v\mu} &= (1/T) \sum_{t=1}^T \hat{v}_t \hat{\mu}_t' \end{aligned} \quad (6.79)$$

y compute los valores propios de $\Sigma_{vv}^{-1} \Sigma_{v\mu} \Sigma_{\mu\mu}^{-1} \Sigma_{\mu v}$, que ordenadamente son $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_b$.

La función de verosimilitud se maximiza sujeta a que existan h vectores de cointegración cuando el número de rezagos está bien especificado (esto es importante, como veremos más adelante):

$$\mathcal{L}^* = (-Tn/2) \log(2\pi) - (Tn/2) - (T/2) \log |\Sigma_{\mu\mu}| - (T/2) \sum_{i=1}^b \log(1 - \hat{\lambda}_i) \quad (6.80)$$

Paso 4: Calcule los estimadores de máxima verosimilitud usando los vectores propios $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_h$ de $\Sigma_{vv}^{-1} \Sigma_{v\mu} \Sigma_{\mu\mu}^{-1} \Sigma_{\mu v}$.

Recuerde que los vectores propios proveen el espacio de los vectores de cointegración. El estimador de máxima verosimilitud será aquel vector que puede ser escrito como:

$$a = b_1 \hat{a}_1 + b_2 \hat{a}_2 + \dots + b_b \hat{a}_b \quad (6.81)$$

(escogiendo óptimamente los b). Hay que normalizar, por lo que usamos $\hat{a}_i' \Sigma_{vv} \hat{a}_i = 1$ y luego juntamos los h vectores normalizados en \hat{A} .

Los estimadores de máxima verosimilitud de ξ_0 , α , x_{it} y Ω son:

$$\begin{aligned} \hat{\xi}_0 &= \Sigma_{v\mu} \hat{A} \hat{A}' \\ \hat{\xi}_i &= \hat{\Pi}_i - \hat{\xi}_0 \hat{\varepsilon}_i \\ \hat{\alpha} &= \hat{\pi}_0 - \hat{\xi}_0 \hat{\theta} \\ \hat{\Omega} &= \frac{1}{T} \sum_{t=1}^T [(\hat{\mu} - \hat{\xi}_0 \hat{v}_t)(\hat{\mu} - \hat{\xi}_0 \hat{v}_t)'] \end{aligned} \quad (6.82)$$

La intuición del uso de las regresiones auxiliares es que se necesitan para concentrar la función de verosimilitud. Lo que se hace es suponer inicialmente que ξ_0 , y Ω están dados y se maximiza la función verosimilitud con respecto a α y ξ . Mirando la ecuación (6.79) se deduce que esto es equivalente a hacer regresiones de $\Delta y_t - \xi_0 y_{t-1}$ en una constante y $\Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p+1}$. Obtenemos $\hat{\alpha}^*$ y $\hat{\xi}_i^*$ que son funciones de ξ_0 .

Si hicimos dichas regresiones, entonces los residuos

$$\Delta y_t - \xi_0 y_{t-1} - \hat{\alpha}^*(\xi_0) - \hat{\xi}_1^*(\xi_0) \Delta y_{t-1} - \dots - \hat{\xi}_{p-1}^*(\xi_0) \Delta y_{t-p+1} \quad (6.83)$$

son ortogonales a $\Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p+1}$.

Pero note que los residuos $\hat{v}_t, \hat{\mu}_t$ también satisfacen la condición de ortogonalidad y una combinación lineal de ellos también lo hace, $\hat{v}_t - \hat{\xi}_0 \hat{\mu}_t$ y ¿qué es esta combinación lineal?

$$\begin{aligned} \hat{\mu}_t - \xi_0 \hat{v}_t = & (\Delta y_t - \hat{\pi}_0 - \hat{\Pi}_1 \Delta y_{t-1} - \hat{\Pi}_2 \Delta y_{t-2} - \dots - \hat{\Pi}_{p-1} \Delta y_{t-p+1}) \\ & - \xi_0 (y_{t-1} - \hat{\theta} - \Xi_1 \Delta y_{t-1} - \Xi_2 \Delta y_{t-2} - \dots - \hat{\theta} - \Xi_{t-p} \Delta y_{t-p+1}) \end{aligned} \quad (6.84)$$

Si $\hat{\alpha}^*(\xi_0) = \hat{\pi}_0 - \xi_0 \hat{\theta}$ y $\hat{\xi}_i^*(\xi_0) = \hat{\Pi}_i - \xi_0 \hat{\Xi}_i$, entonces la ecuación (6.87) y la ecuación (6.88) son equivalentes. Reemplazando en la función de verosimilitud la ecuación (6.80), tenemos:

$$\begin{aligned} \mathcal{L}(\Omega, \hat{\xi}_1^*(\xi_0), \dots, \hat{\xi}_{p-1}^*(\xi_0), \hat{\alpha}^*(\xi_0), \xi_0) = & (-Tn/2) \log(2\pi) - (T/2) \log |\Omega| \\ & - (1/2) \sum_{t=1}^T [(\hat{\mu}_t - \xi_0 \hat{v}_t)' \Omega^{-1} (\hat{\mu}_t - \xi_0 \hat{v}_t)] \end{aligned} \quad (6.85)$$

Ahora basta con maximizar con respecto a ξ_0 y Ω . En realidad no. Porque el estimador de Ω es $\hat{\Omega} = \frac{1}{T} \sum_{t=1}^T [(\hat{\mu}_t - \hat{\xi}_0 \hat{v}_t)(\hat{\mu}_t - \hat{\xi}_0 \hat{v}_t)']$ y por lo tanto, la ecuación (6.89) queda:

$$\begin{aligned} \mathcal{L}(\hat{\Omega}^*(\xi_0), \hat{\xi}_1^*(\xi_0), \dots, \hat{\xi}_{p-1}^*(\xi_0), \hat{\alpha}^*(\xi_0), \xi_0) \\ = (-Tn/2) \log(2\pi) - (Tn/2) - (T/2) \log \left(\frac{1}{T} \sum_{t=1}^T [(\hat{\mu}_t - \xi_0 \hat{v}_t)(\hat{\mu}_t - \xi_0 \hat{v}_t)'] \right) \end{aligned} \quad (6.86)$$

así es que para maximizar la función de verosimilitud sólo hay que minimizar el paréntesis final.

¿Para qué queremos las correlaciones canónicas? Para poder expresar el paréntesis de la ecuación (6.90) en términos de los vectores característicos, λ .

Usando correlación canónica podemos demostrar que minimizar la ecuación (6.79) sujeto a que ξ_0 sea BA' es la ecuación (6.84).

Si queremos minimizar la ecuación (6.79) vamos a escoger ξ_0 tal que $\hat{\mu}_t$ sea lo más correlacionado posible con \hat{v}_t . Eso hace pequeña la distancia entre ellos. Pero si hacemos eso, es equivalente a escoger el λ_i más grande (lo que vimos en correlación canónica). Luego, hemos demostrado la ecuación (6.84).

¿Para qué sirve esto? Cuando usamos la técnica mínimos cuadrados ecuación por ecuación (Engle-Granger, 1987) estabamos más preocupados de la existencia de un vector de cointegración vs. no cointegración que de cuántos vectores había.

Ahora podemos usar la ecuación (6.84) para testear el número de vectores de cointegración. Supongamos que hay h vectores de cointegración. La función de verosimilitud se maximiza en la ecuación (6.84):

$$\mathcal{L}^* = (-Tn/2) \log(2\pi) - (Tn/2) - (T/2) \log|\Sigma_{\mu\mu}| - (T/2) \sum_{i=1}^h \log(1 - \hat{\lambda}_i) \quad (6.87)$$

Supongamos alternativamente que hay n vectores de cointegración, es decir que toda combinación lineal de y_t es estacionaria. La función de verosimilitud es ahora:

$$\mathcal{L}_{\mathcal{A}}^* = (-Tn/2) \log(2\pi) - (Tn/2) - (T/2) \log|\Sigma_{\mu\mu}| - (T/2) \sum_{i=1}^n \log(1 - \hat{\lambda}_i) \quad (6.88)$$

Un test hipótesis de la ecuación (6.91) vs la ecuación (6.92) es directo:

$$2(\mathcal{L}_{\mathcal{A}}^* - \mathcal{L}^*) = -T \sum_{i=h+1}^n \log(1 - \hat{\lambda}_i) \quad (6.89)$$

Si las hipótesis fuesen sobre variables $I(0)$, el test en la ecuación (6.93) sería χ^2 . Bajo la nula, hay $g = n - h$ variables $I(1)$, luego la distribución no es estándar, pero está tabulada. Por si no se han dado cuenta, la ecuación (6.93) es un test sobre cuántos vectores de cointegración hay. Se le llama test de traza (porque suma sobre la diagonal de R).

Otra alternativa sería testear:

H0: k vectores de cointegración vs. H1: $k+1$ vectores.

Usando la ecuación (6.91) y la ecuación (6.92) obtenemos:

$$2(\mathcal{L}_{\mathcal{A}}^* - \mathcal{L}^*) = -T \log(1 - \hat{\lambda}_{k+1}) \quad (6.90)$$

Este es llamado test del máximo valor propio (max eigenvalue).

6.08 Cointegración Estacional

La existencia de raíces unitarias estacionales sugiere de modo natural evaluar la presencia de cointegración estacional. Es decir, tendencias estocásticas comunes de largo plazo entre los componentes estacionales de distintas variables. La cointegración estacional puede verse del mismo modo que la cointegración de largo plazo, es decir, como un desplazamiento paralelo entre las variables como resultado de tendencias estocásticas comunes.

El modelo de cointegración para distintas frecuencias se obtiene de una extensión directa que Engle et al. (1992)¹⁸ realizan del popular esquema de cointegración y corrección de errores desarrollado por Engle y Granger (1987). Utilizando la descomposición de las series ya expuesta, nos interesa ver la existencia de vectores de cointegración en distintas frecuencias:

- (1) Si la cointegración se produce sólo en los componentes de largo plazo, tenemos un modelo de corrección de errores clásico, como el planteado por Engle-Granger (1987):

$$\Delta y_t = \alpha \beta y_{t-1} + \epsilon_t \quad (6.91)$$

donde y es un vector que contiene las variables de interés, β es el vector de cointegración y α es el factor de ajuste. Este último señala qué fracción del desequilibrio entre las variables respecto de su relación de largo plazo que se observó en "t-1" se ajusta en el periodo "t".

- (2) Cuando además hay cointegración en los componentes estacionales semestrales, el modelo de corrección de errores puede expresarse como:

$$y_t - y_{t-2} = \frac{1}{2} \alpha_1 \beta_1 (y_{t-1} + y_{t-2}) + \frac{1}{2} \alpha_2 \beta_2 (y_{t-1} - y_{t-2}) + \epsilon_t \quad (6.92)$$

donde α_1 y α_2 son factores de ajuste. El primer término de la ecuación corresponde al promedio anual de la variable, por lo que β_1 corresponde al vector de cointegración de largo plazo clásico expuesto anteriormente. El segundo término, en cambio, mide la variación intra-anual, por lo que β_2 es el vector de parámetros que hace cointegrar la variación intra-anual de estas variables.

- (3) Si existe también cointegración en la frecuencia trimestral, el modelo de corrección de errores se puede expresar como:

$$y_t - y_{t-4} = \frac{1}{4} \alpha_1 \beta_1 (y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4}) + \frac{1}{4} \alpha_2 \beta_2 (y_{t-1} - y_{t-2} + y_{t-3} - y_{t-4}) + \frac{1}{4} (\alpha_R \beta_R + \alpha_I \beta_I) (y_{t-2} - y_{t-4}) + \frac{1}{4} (\alpha_I \beta_R - \alpha_R \beta_I) (y_{t-1} - y_{t-3}) + \epsilon_t \quad (6.93)$$

¹⁸ Engle, R; C.W.J. Granger; S. Hylleberg; y H.S. Lee (1992). "Seasonal Cointegration: The Japanese Consumption Function." *Journal of Econometrics* 55:275-298.

donde los primeros dos términos corresponden a la cointegración no estacional y semestral, respectivamente, y su interpretación es análoga a la realizada anteriormente. Los últimos dos términos tienen una interpretación más compleja ya que corresponden a un caso de cointegración polinomial (los subíndices I y R denotan los componentes imaginarios y reales de la solución).

El problema de la ecuación (9) radica en que estos polinomios no necesariamente tienen rango reducido y los parámetros no están identificados. Lee (1992) propone, como mecanismo para lograr la identificación, deshacerse del segundo término, asumiendo que $\alpha_R \beta_I - \alpha_I \beta_R = 0$. Ello es claramente arbitrario. Un supuesto menos restrictivo sería, como lo plantean Johansen y Schaumburg (1999), asumir que $\beta_I = 0$, con lo cual la ecuación queda como:

$$y_t - y_{t-4} = \frac{1}{4} \alpha_1 \beta_1 (y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4}) + \frac{1}{4} \alpha_2 \beta_2 (y_{t-1} - y_{t-2} + y_{t-3} - y_{t-4}) + \frac{1}{4} (\alpha_R L - \alpha_I) \beta_R' (y_{t-1} - y_{t-3}) + \epsilon_t \quad (6.94)$$

La interpretación del último término es ahora algo más sencilla: el término $\beta_R (y_{t-1} - y_{t-3})$ es directamente estacionario o, en su defecto, cointegra con su propio rezago.

La especificación de la cointegración estacional deja en evidencia el papel que ésta podría cumplir para mejorar nuestro entendimiento de los determinantes de la demanda de dinero, permitiendo a la vez obtener estimaciones más robustas. Si no existen componentes cointegrados en frecuencia semestral o trimestral, la especificación de la ecuación (6.94) corresponde al modelo estándar de corrección de errores empleado en muchas estimaciones para el caso chileno (Soto, 1996; Adam, 2000). No obstante, si existen componentes que cointegren en frecuencia estacional, la ecuación señala que dichos modelos estándares están mal especificados, pues han omitido relaciones de equilibrio que contienen información relevante para determinar los saldos monetarios deseados por los agentes económicos

Apéndice A: Códigos Gauss

```

/*          Regresion Espuria de Pearson          */
new;
cls;
library pgraph;
"Dame el largo de la serie (50+) " ;
len=con(1,1);
i=1;
result=zeros(1000,1);
do while i<=1000;
    data1=rndn(len,1);
    data2=rndn(len,1);
    data3=rndn(len,1);
    ratio1=data1./data3;
    ratio2=data2./data3;
    __output=0;
    {vnam,m,b,skb,vc,stderr,sigma,cx,rsq,resid,dwstat } = ols(0,ratio1,ratio2);
    result[i,1]=rsq;
    i=i+1;
end;
{v,m,f}=hist(result,10);
"El promedio de las correlaciones es: " meanc(result[:,1]);
end;

```

```

/*          Regresion Espuria de Granger y Newbold          */
new;
cls;
library pgraph;
"Dame el largo de la serie (50+) " ;
len=con(1,1);
replic=1000;
result=zeros(replic,2);
i=1;
do while i<=replic;
    shock1=rndn(len,2);
    shock2=rndn(len,2);
    var1=zeros(len,1);
    var2=var1;
    j=2;
    do while j<=len;
        var1[j,]=0.1+var1[j-1,]+shock1[j,1];
        var2[j,]=0.05+var2[j-1,]+shock2[j,2];
        j=j+1;
    end;
    __output=0;
    {vnam,m,b,skb,vc,stderr,sigma,cx,rsq,resid,dwstat } = ols(0,var1,var2);
    result[i,1]=rsq;
    result[i,2]=b[2,1]./skb[2,1];
    i=i+1;
end;
format /rd 5,2;
"Porcentaje de Casos con Rsq mayor que 0.8"; rows(selif(result,result[:,1] .gt 0.8))/replic*100;
"Porcentaje de Casos con 't' mayor que 2.0"; rows(selif(result,abs(result[:,2]) .gt 2.0))/replic*100;
mono(result[:,1],result[:,2]);
end;

```



```

/*          Regresion Espuria de Granger, Heung y Young          */

new;
cls;
library pgraph;
"Dame el largo de la serie (50+)      " ;
len=con(1,1);                               /* pide largo de datos desde la consola */
replic=1000;
result=zeros(replic,2);
i=1;
aut=0.99;
do while i<=replic;
    shock1=rndn(len,2);                      /* crea data white noise */
    shock2=rndn(len,2);
    var1=zeros(len,1);                       /* inicializa variables */
    var2=var1;
    j=2;
    do while j<=len;
        var1[j,.]=0.+aut*var1[j-1,.]+shock1[j,1]; /* crea variable 1 */
        var2[j,.]=0.+aut*var2[j-1,.]+shock2[j,2]; /* crea variable 2 */
        j=j+1;
    endo;
    {bhat,that,covb,sigma2,stde,see,rss,d,akaike,schwarz,r2,r2a} =
myols(var1,ones(rows(var2),1)~var2);
    result[i,2]=that[2,1];
    result[i,1]=r2;
    i=i+1;
endo;

format /rd 5,2;
"Porcentaje de Casos con Rsq mayor que 0.8"; rows(selif(result,result[.,1] .gt 0.8))/replic*100;
"Porcentaje de Casos con 't' mayor que 2.0";  rows(selif(result,abs(result[.,2]) .gt 2.0))/replic*100;

mono(result[.,1],result[.,2]);
end;

```

Apéndice B: Ejercicios

1. Use el programa de regresión espuria de Granger-Newbold para estudiar cómo cambia la presencia de correlación espuria cuando el tamaño de muestra crece (p.e., $T=1000$).
2. Suponga que el modelo es $y_t = \alpha + \delta z_t + \epsilon_t$, donde z_t contiene dos variables: una variable estacionaria (wt) y una variable que crece con una tendencia exactamente lineal (x_t). Demuestre que la distribución de los estimadores de δ degenera cuando $t \rightarrow \infty$. ¿por qué en este modelo $T^{3/2} \hat{\delta}$ es superconsistente?

Table of Contents

6.01 Correlación espuria.....	2
6.02 Noción de cointegración.....	10
6.03 Cointegración.....	12
6.04 Tests de cointegración.....	19
6.05 Otras preguntas interesantes.....	22
6.06 Test de hipótesis cuando las variables cointegran.....	23
6.08 Cointegración Estacional.....	31
Apéndice A: Códigos Gauss.....	33
Apéndice B: Ejercicios.....	35