

CAPÍTULO 3

MODELOS DE SERIES DE TIEMPO

ESTACIONARIAS MULTIVARIADOS

La técnica de series de tiempo univariadas que revisamos en el Capítulo 2 provee una importante herramienta de pronóstico sobre la evolución futura de alguna variable de la economía y, de hecho, es ampliamente utilizada. Sin embargo, ella es limitada para entender y modelar el funcionamiento de la economía en general, pues excluye la interacción entre variables económicas que, razonablemente, es lo que caracteriza la estructura de la economía. El instrumental de los modelos ARMA puede ser extendido directamente para hacer modelos multivariados, es decir, aquellos en que se modela un vector de variables de interés.

En este capítulo nos concentraremos en modelos de vectores autoregresivos (VAR) excluyendo del análisis los vectores de medias móviles (VMA). Existen dos razones para excluir estos últimos. Primero, los modelos VMA son altamente no lineales, lo que complica su estimación y hace muy difícil su interpretación. Segundo, como hemos visto, cualquier modelo MA(q) puede ser aproximado de manera arbitrariamente cercana usando un modelo AR de orden suficientemente grande.

3.01 Vectores Autoregresivos¹

Consideremos el siguiente modelo de series de tiempo multivariadas que llamaremos Vector Autoregresivo de orden 1 o, más sintéticamente, VAR(1):

$$\begin{aligned} y_t &= \beta_{10} - \beta_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \epsilon_t^y \\ z_t &= \beta_{20} - \beta_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \epsilon_t^z \end{aligned} \quad (3.1)$$

donde ϵ^y y ϵ^z son dos procesos ruido blanco independientes entre sí, es decir, con $cov(\epsilon^y, \epsilon^z) = 0$.

La ecuación (3.1), que a lo largo de este capítulo llamaremos *VAR primitivo*, presenta tres características interesantes:

- Ambas variables z_t e y_t son endógenas, en el sentido que su nivel y su trayectoria se determina por la interacción con la otra variable al interior del modelo.

¹ Los modelos VAR fueron impulsados por el célebre trabajo de C. Sims (1980), "Macroeconomics and Reality," *Econometrica*, 48:1-48, donde se hace una lúcida crítica de los modelos econométricos tradicionales para el análisis de políticas económicas.

- En ambas variables hay dinámica, en el sentido que las innovaciones se propagan en el sistema con rezagos afectando a las variables en el tiempo.
- Si estas ecuaciones se estimasen por separado usando mínimos cuadrados ordinarios, los estimadores de los parámetros incurrirían en sesgo de simultaneidad si β_{12} o β_{21} no son cero.

Aunque no podemos estimar la forma primitiva del modelo, podemos estimar su forma reducida. Considere:

$$\begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^y \\ \epsilon_t^z \end{bmatrix} \quad (3.2)$$

La ecuación (3.2) puede ser escrita de manera abreviada como:

$$B x_t = \Gamma_0 + \Gamma_1 x_{t-1} + \epsilon_t \quad (3.3)$$

Premultiplicando por B^{-1} , obtenemos

$$x_t = A_0 + A_1 x_{t-1} + e_t \quad (3.4)$$

donde con $A_0 = B^{-1} \Gamma_0$, $A_1 = B^{-1} \Gamma_1$, y $e_t = B^{-1} \epsilon_t$. Como resulta evidente, el modelo (3.4) no tiene problemas de simultaneidad, por lo cual es posible estimarlo directamente usando mínimos cuadrados. A la vez, nótese que mantiene la estructura dinámica del modelo primitivo. Sin embargo, las innovaciones de la ecuación (3.4) –que llamaremos *forma reducida o estimable*– no son las mismas que las del modelo primitivo, por cuanto si β_{12} o β_{21} no son cero, habrá correlación entre los componentes de e_t .

Equivalentemente, podemos escribir el modelo de forma reducida como:

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix} \quad (3.5)$$

3.02 El Problema de Identificación

Debido a que el modelo que se puede estimar sin problemas econométricos es la forma reducida –que no es de interés en muchas ocasiones–, resulta importante preguntarse si podemos recuperar el VAR primitivo a partir de los parámetros estimados en la forma reducida. Una cuenta rápida de incógnitas y ecuaciones sugiere que esto no es posible:²

$$\begin{aligned} \text{Desconocidos: } & \beta_{10}, \beta_{20}, \beta_{21}, \beta_{12}, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \sigma_{\tilde{z}}^2, \sigma_y^2 \\ \text{Conocidos: } & a_{10}, a_{20}, a_{11}, a_{12}, a_{21}, a_{22}, \sigma_1^2, \sigma_2^2, \text{cov}(e_t^1, e_t^2) \end{aligned}$$

El número de parámetros desconocidos en el modelo primitivo es mayor que el número de ecuaciones que determinan los parámetros de la forma reducida para la cual tenemos estimadores. Por lo tanto, no se puede identificar el modelo primitivo a partir del modelo reducido, al menos que se imponga una restricción $\beta_{12}=0$ o $\beta_{21}=0$. De hecho, este resultado es general, no importando el número de variables incluidas en el VAR ni su orden (número de rezagos). La principal implicación de este resultado es que todo modelo de ecuaciones simultáneas está sub-identificado. El corolario de este resultado es que para identificar un modelo primitivo, obligatoriamente habrá que imponer restricciones de identificación. El problema de estimar modelos de ecuaciones simultáneas se transforma, entonces, en un problema de tipos y formas de restricciones de identificación que cada científico impone y la justificación que se le da a dichas restricciones.

Veamos qué es lo que hace una restricción de identificación en este contexto. ¿Qué es lo que se logra al imponer una (o más) restricción? Supongamos que forzamos $\beta_{21}=0$. Entonces, el VAR primitivo se convierte en

$$\begin{aligned} y_t &= \beta_{10} - \beta_{12} \tilde{z}_t + \gamma_{11} y_{t-1} + \gamma_{12} \tilde{z}_{t-1} + \epsilon_t^y \\ \tilde{z}_t &= \beta_{20} + \gamma_{21} y_{t-1} + \gamma_{22} \tilde{z}_{t-1} + \epsilon_t^{\tilde{z}} \end{aligned} \quad (3.6)$$

es decir, ahora es un modelo recursivo. Una innovación en ϵ^z en el instante t impacta a z_t y éste, instantáneamente, afecta a y_t . El efecto de la innovación de z_t es simultáneo sobre ambas variables. Pero una innovación en ϵ^y afecta a z_t sólo en el tiempo, a través de los rezagos de y_t , pero no contemporáneamente.

Al incluir una restricción es posible recuperar los coeficientes de la forma primitiva a partir de los coeficientes estimados en la forma reducida porque:

² En estricto rigor, lo que debe verificarse es la condición de rango del sistema de ecuaciones y no la condición de orden (ver Apéndice A). Para fines didácticos, el uso de la condición de orden es suficiente.

$$\begin{aligned}
a_{10} &= \beta_{10} - \beta_{12} \beta_{20} & a_{11} &= \gamma_{11} - \beta_{12} \gamma_{21} & a_{12} &= \gamma_{12} - \beta_{12} \gamma_{22} \\
a_{20} &= \beta_{20} & a_{21} &= \gamma_{21} & a_{22} &= \gamma_{22} \\
\sigma_1^2 &= \sigma_y^2 + \beta_{12} \sigma_z^2 & \sigma_2^2 &= \sigma_z^2 & cov(e^1, e^2) &= -\beta_{12} \sigma_z^2
\end{aligned} \tag{3.7}$$

Esta manera de identificar un modelo primitivo se llama descomposición de Cholesky.³ Esencialmente, es una descomposición triangular (en nuestro caso *lower triangular*). En palabras simples lo que hace esta descomposición es ordenar el modo como las innovaciones se propagan en el sistema dinámico.

La descomposición de Cholesky no es única. De hecho, podríamos haber supuesto $\beta_{12}=0$ y haber obtenido otro modelo primitivo a partir de la misma forma reducida. Esto implica necesariamente que toda identificación es arbitraria. Para hacer un análisis con VARs de manera seria debemos presentar y estudiar las posibles descomposiciones alternadamente.

La descomposición de Choleski no es la única forma de identificar. Hay una línea de análisis basada en imponer ceros ex-ante con pretextos (buenos y malos) teóricos con el nombre de VAR estructural, que discutiremos más adelante. Hay otra posibilidad de identificar la forma primitiva del VAR no con restricciones a innovaciones contemporáneas, sino mediante la imposición de restricciones sobre los efectos acumulados de las innovaciones de alguna variable sobre el sistema.⁴ En cualquier caso, sea que se ocupe la descomposición de Cholesky o algún otro modo de identificación, el econométrista debe incluir información exógena al sistema en el momento de identificar la conexión entre forma primitiva y forma reducida.

3.03 Ecuaciones Simultáneas Tradicionales vs. VARs

La descomposición de Cholesky tiene una ventaja que los modelos de ecuaciones simultáneas tradicionales no tiene: es explícita. El econométrista debe señalar la o las restricciones que impuso y efectuar sus análisis condicionales en dichas restricciones. En los modelos de ecuaciones simultáneas tradicionales⁵ la identificación es frecuentemente

3 Nadie sabe con certeza quién descubrió la descomposición de Cholesky. La historia oficial es que Andre-Louis Cholesky (1875-1918) fue un militar francés que, dedicado a labores geodésicas en Creta y Noráfrica, desarrolló la descomposición para computar las soluciones de las ecuaciones normales en problemas de cuadrados mínimos aplicados a la confección de mapas. Su trabajo se publicó póstumamente (en 1924) en el *Bulletin Geodesique*. Sin embargo Rudolf Zurmuhl afirma que el método habría sido usado por H.M. Doolittle para fines similares en 1878 (*Matrizen*, Springer, p.64, 1961).

4 Ver Blanchard, J.O. y D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances". *American Economic Review*, 79(4):655-673 y R. Soto (1990) "Dinámica del Crecimiento y la Inflación en Chile: 1957-1988, *Revista de Análisis Económico*, 5(1):89-112. Para identificar la forma primitiva del modelo, en estos trabajos se impone la restricción que variaciones de demanda agregada tengan solo efectos transitorios sobre el desempleo y la inflación, respectivamente.

5 Una descripción de los modelos de ecuaciones simultáneas tradicionales se presenta en el Apéndice B.

confusa y no es reconocida por el econometrista como un elemento que condicione su análisis. Como un ejemplo de este problema consideremos el siguiente modelo típico de ecuaciones simultáneas tradicionales:

$$\begin{array}{ll} \text{Demanda} & q_d = \alpha_1 p_t + \alpha_2 y_t + \epsilon_t^d \\ \text{Oferta} & q_s = \beta_1 p_t + \epsilon_t^s \\ \text{Equilibrio} & q_s = q_d \end{array} \quad (3.8)$$

El folklore de los modelos de ecuaciones simultáneas tradicionales señala que (1) p_t y q_t son variables endógenas, (2) y_t es una variable exógena, y (3) que las ecuaciones reflejan la estructura de la “economía”, en particular su dinámica. De la ecuación (3.8) es posible obtener la siguiente forma reducida por reemplazo:

$$\begin{aligned} q_t &= \frac{\beta_1 \alpha_2 y_t}{\beta_1 - \alpha_1} + \frac{\epsilon_t^d - \epsilon_t^s}{\beta_1 - \alpha_1} \\ p_t &= \frac{\alpha_2 y_t}{\beta_1 - \alpha_1} + \frac{\beta_1 \epsilon_t^d - \alpha_1 \epsilon_t^s}{\beta_1 - \alpha_1} \end{aligned} \quad (3.9)$$

la que puede ser estimada usando, p.e., mínimos cuadrados. Ello permite recuperar la ecuación de “oferta” en (3.8) –que está *exactamente* identificada– pero no se podría parametrizar la función de “demanda” porque está *subidentificada*.

Las diferencias de enfoque entre la metodología de series de tiempo multivariadas y las ecuaciones simultáneas tradicionales son notorias y muy profundas. En particular,

- En VARs no se clasifica las variables en “endógenas” o “exógenas” a priori. Todas las variables son potencialmente endógenas y serán los datos los que sugieran si es razonable una u otra especificación como la más adecuada.
- En VARs no se impone una determinada estructura de rezagos a priori. Se determina la dinámica en función de los datos. En particular, nótese la arbitrariedad implícita en (3.8) pues el econometrista eliminó ex-ante cualquier rezago de las variables. Es decir, sin siquiera estudiar los datos se eliminó cualquier forma de dinámica.
- En VARs se impone el número mínimo de restricciones de identificación, se reconoce a priori éstas, y se estudia cómo cambian los resultados con distintas maneras de identificar. En particular, note la arbitrariedad de la identificación impuesta por el econometrista en (3.8).⁶

⁶ Al respecto se aplica la célebre opinión de Sherlock Holmes: *It is a capital mistake to theorize before one has the data. Insensibly one begins to twist the facts to suit theories instead of theories to suit facts.* Arthur Conan Doyle, *A scandal in Bohemia*, en *The Complete Sherlock Holmes*, Barnes and Noble, pp. 163.

- En VARs los parámetros y su significancia estadística no son un fetiche. En las ecuaciones simultáneas tradicionales, la “estructura del modelo” es frecuentemente escogida sobre la base de determinar si los coeficientes son individualmente significativos con algún nivel de confianza, en un ambiente donde el sesgo de simultaneidad y la colinealidad son evidentes.

3.04 Causalidad en Series de Tiempo

El tema de causalidad en economía y, en particular en series de tiempo, ocupa un lugar central en la modelación económica. Sin embargo, no existe consenso respecto de qué es exactamente “causalidad” (ver LeRoy, 2004 para una discusión del tema)⁷. Siguiendo a Hendry (2004)⁸, una causa puede ser definida como un proceso quantitativo que induce cambios en el tiempo a través de una estructura. La estructura permanece invariable y caracteriza las relaciones entre variables (es decir, refleja la realidad). La relación entre causa y efecto es asimétrica, pues la segunda no induce la primera. Esta noción es congruente con Simon (1957).⁹

Usualmente en economía se postula que y es causado por x al escribir $y=f(x)$. Esta simple expresión tiene algunos problemas: (1) la descripción es esencialmente circular (el modelo no estaría bien definido si $f(.)$ no describiera la relación), (2) la asimetría de causalidad está fuera del problema porque si $f(.)$ es invertible, entonces x podría ser causado por y , (3) la teoría $f(.)$ podría estar bien definida pero ser irrelevante para el fin del análisis económico, y (4) una definición de causalidad a partir de un modelo puede ser insuficiente para caracterizar todos los canales de causalidad. El punto central sobre causalidad es que ésta es una propiedad de la realidad y no de un modelo teórico o estadístico.

Por ello, los economistas tendemos a usar una definición de causalidad basada en los datos y no en teorías. Como las relaciones económicas no son leyes, la formulación es estocástica, lo que nos lleva al análisis de tests de causalidad y, más adelante, a la idea de evaluar causalidad en términos de los cambios en la distribución conjunta de las variables.

A continuación estudiamos los tests más comunes de causalidad por dos razones. Primero, porque ellos son útiles para determinar una forma particular de causalidad en economía. Segundo, porque existe la tentación de identificar los modelos VAR (y a veces los modelos estructurales) sobre la base de la causalidad en el tiempo, lo

7 Stephen F. LeRoy, “Causality in Economics”, Technical Report 20/04, Centre for Philosophy of Natural and Social Science, London School of Economics, 2004.

8 David F. Hendry, “Causality and Exogeneity in Non-stationary Economic Time-Series”, Technical Report 18/04, Centre for Philosophy of Natural and Social Science, London School of Economics, 2004.

9 Simon, H. A. (1957). *Models of Man*. New York: John Wiley & Sons.

que es incorrecto. Para entender esto estudiaremos primero algunas definiciones de causalidad y luego su relación con el problema de identificación.

Causalidad de Granger (1969)¹⁰

El concepto de causalidad más ampliamente usado en series de tiempo es el desarrollado por Granger (1969). Analíticamente:

$$\begin{aligned} \text{Si } \theta(L) \neq 0 \text{ en } y_t = \phi(L)y_{t-1} + \theta(L)z_{t-1} + \epsilon_t \\ \Rightarrow z_t \text{ causa-a-la-Granger a } y_t \end{aligned} \quad (3.10)$$

es decir, z_t causa-a-la-Granger a y_t si ésta contiene información útil para predecir y_t que no está contenida en la historia de y_t .

Es importante hacer varias observaciones con respecto a este test:

- Primero, este no es un test de causalidad en el sentido profundo de “causa y efecto”, sino sólo un test de “precedencia temporal”.¹¹ La mecánica mediante la cual z_t afecta a y_t es desconocida y se debe tener cuidado con la interpretación de los resultados pues éstos dependen del modelo que se esté estudiando. Ver Box 1.
- Segundo, debido a la probable existencia de colinealidad, el test se hace sobre el bloque de rezagos de z_t y no sobre parámetros individuales, es decir se hace un test F sobre la hipótesis nula: $\theta(L)=0$. Dado que existe la posibilidad que haya correlación espúrea, el test se hace sobre la hipótesis nula z_t no causa a la Granger a y_t .
- Tercero, es posible que z_t cause-a-la-Granger a y_t y simultáneamente que y_t cause-a-la-Granger a z_t . Ello puede suceder tanto porque las variables efectivamente son simultáneas –por ejemplo cuando ambas son generadas por una tercera variable, w_t – como cuando son observacionalmente simultáneas por problemas de medición (p.e., hay causalidad mensual pero los datos observados trimestralmente).
- Cuarto, un VAR es el entorno óptimo para testear causalidad a la Granger. Esto se deduce al comparar el test de causalidad de la ecuación (3.10) con la especificación del VAR de forma reducida de la ecuación (3.5).

¹⁰ Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Methods and Cross-Spectral Methods," *Econometrica*, 34, 424-438. Como Granger se basó en trabajos previos de N. Wiener ("The Theory of Prediction", en E.F. Breckenbrack, *The Theory of Production*, McGraw-Hill, NY, 1956), a veces se le llama causalidad de Wiener-Granger.

¹¹ Una discusión profunda sobre el problema de hacer equivalente causalidad y precedencia temporal se encuentra en J. Geweke (1984), "Inference and Causality in Economic Time Series Models", *Handbook of Econometrics*, vol. II, pp. 1101-1144

Otra manera de pensar el mismo test es en términos del error cuadrático medio: z_t no causa-a-la-Granger a y_t si

$$ECM[y_t^f | y_{t-1}, \dots] = ECM[y_t^f | y_{t-1}, \dots, z_{t-1}, \dots] \quad (3.11)$$

Causalidad de Sims (1972)¹²

Sims (1972) plantea que el test de causalidad debiese ser hacia el futuro (*forward looking*), no hacia el pasado. Por ello,

$$\begin{aligned} \text{si } \theta(L) \neq 0 \text{ en } y_t &= c + \sum_{i=0}^{\infty} \phi_i z_{t-i} + \sum_{i=1}^{\infty} \theta_i z_{t+i} + \epsilon_t \\ \Rightarrow z_t &\text{ causa a-la-Sims a } y_t \end{aligned} \quad (3.12)$$

Box 1 Interpretación de la Causalidad de Granger¹³

Hay que tener cuidado con la interpretación del test de causalidad de Granger con variables *forward-looking*. De acuerdo a Fama (1970) el precio de las acciones refleja el valor presente de los dividendos esperados de dichas acciones. Es decir

$$p_t = E_t \sum_{i=1}^{\infty} \left[\frac{1}{1+r} \right]^i d_{t+i}.$$

Si los dividendos tienen la siguiente ley de movimientos $d_t = d + \mu_t + \delta \mu_{t-1} + v_t$, entonces el valor esperado de d_t es

$$E_t(d_{t+j}) = \begin{cases} d + \delta \mu_t & \text{si } j=1 \\ d & \text{si } j>1 \end{cases}.$$

Entonces el precio p_t es $p_t = \frac{d}{r} + \frac{\delta \mu_t}{1+r}$, es decir es ruido blanco. Por lo tanto debe concluirse que ninguna variable causa a-la-Granger al precio.

Más aún se obtiene fácilmente que dado que d_t depende de p_{t-1} hay causalidad a-la-Granger de precios a dividendos.

¹² Sims, C. (1972). "Money, Income and Causality," *American Economic Review*, 62:540-552.

¹³ Se debe ser siempre cuidadoso con la interpretación de la causalidad: *El matrimonio es la causa número uno de divorcios. Estadísticamente, el 100 por ciento de los divorcios comenzó con un matrimonio.*

Causalidad de Geweke, Meese y Dent (1983)¹⁴

Un problema del test de Sims es que el residuo es típicamente autocorrelacionado y, por lo tanto, un test F no funciona. Una solución simple consiste en controlar la autocorrelación usando rezagos, es decir:

$$\begin{aligned}
 \text{si } \theta(L) \neq 0 \text{ en } y_t &= c - \sum_{i=1}^{\infty} \lambda_i y_{t-i} + \sum_{i=0}^{\infty} \phi_i z_{t-i} + \sum_{i=1}^{\infty} \theta_i z_{t+i} + \epsilon_t \\
 &\Rightarrow z_t \text{ causa a } y_t \text{ a } l\text{-}GMD
 \end{aligned}
 \quad (3.13)$$

Así, la causalidad determinada usando tests de Granger, Sims o Geweke et al. nos informa sobre cómo se afecta una variable en el instante t frente a innovaciones en otras variables en $t-i$ o en $t+i$, pero no nos dice cómo se afecta ésta frente a innovaciones contemporáneas. Por ello, estos tests no pueden ser usados para “identificar” un modelo, ya que la identificación requiere conocer la correlación contemporánea de shocks.

El Cuadro 3.1 presenta los resultados de tests de causalidad aplicados a datos trimestrales de dinero y producto de la economía chilena en el periodo 1977-2002.¹⁵ Los tests usan 6 rezagos (más adelante volveremos sobre el tema de rezagos óptimos).

Cuadro 3.1
Tests de Causalidad entre Dinero y Producto

	Granger	Sims	Geweke et al.
Dinero no causa producto	5.06	4.29	2.77
Producto no causa dinero	3.11	2.46	1.88

Nota: el valor crítico del F(6,80) es 2.22.

14 Geweke, J., Meese, R., and Dent, W. (1983), "Comparing Alternative Tests of Causality in Temporal Systems: Analytic Results and Experimental Evidence," *Journal of Econometrics*, 21, 161-194.

15 En el Apéndice A se encuentra el código Gauss usado para computar estos resultados.

3.05 Estimación de un VAR(p) por Máxima Verosimilitud

Al igual que en los modelos AR(p), vamos a formar la densidad condicional en las primeras p observaciones del modelo:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (3.15)$$

donde y_t es un vector con n variables y ϕ_0 es una constante. Es decir, lo que queremos es obtener:

$$f(y_t, y_{t-1}, \dots \mid y_0, y_1, \dots, y_{p-1}; \theta) \quad (3.16)$$

y maximizar con respecto a $\theta = [\phi_0, \phi_1, \phi_2, \dots, \phi_p, \Omega]$, donde Ω es la matriz de varianzas y covarianzas de los residuos de las n ecuaciones del modelo.

De nuevo, para y_t están dados c y los primeros p rezagos de y_t , por ello la única estocasticidad proviene de ϵ_t . Entonces:

$$y_t \mid y_{t-1} \dots y_{t-p} \rightarrow N[\phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p}, \Omega] \quad (3.17)$$

Sea $x_t = [1, y_{t-1}, y_{t-2}, \dots, y_{t-p}]'$ y $\Phi = [\phi_0, \phi_1, \phi_2, \dots, \phi_p]$. Escribimos entonces la esperanza condicional de y_t como $\Phi' x_t$. Escribimos la densidad condicional de y_t como

$$y_t \mid y_{t-1} \dots y_{t-p} \rightarrow N[\Phi' x_t, \Omega] \quad (3.18)$$

y la densidad de la muestra como

$$f(y_t, y_{t-1}, \dots \mid y_0, \dots, y_{p-1}; \theta) = \prod_{t=1}^{T-p} f(y_t \mid y_{t-1}, \dots, y_{t-p}; \theta) \quad (3.19)$$

por lo que la función de verosimilitud en logaritmos es:

$$\begin{aligned} \log \mathcal{L}(\theta) &= \sum_{t=1}^{T-p} \log(f(y_t \mid y_{t-1}, \dots, y_{t-p}; \theta)) \\ &= \frac{-(T-p)n}{2} \log(2\pi) + \frac{T-p}{2} \log |\Omega^{-1}| - \frac{1}{2} \sum_{t=1}^{T-p} [(y_t - \Phi' x_t)' \Omega^{-1} (y_t - \Phi' x_t)] \end{aligned} \quad (3.20)$$

Consideremos el último término, ya que los otros son constantes. Al igual que en el caso de los modelos AR(p), maximizar la función de verosimilitud equivale a minimizar el tercer término, es decir a minimizar la suma de residuos al cuadrado del

sistema de ecuaciones. Así, se puede parametrizar todo el modelo VAR, usando mínimos cuadrados ordinarios ecuación por ecuación.

¿Qué pasa cuando Ω es desconocido? Podemos obtener un estimador usando la misma función de verosimilitud. Consideremos

$$\log \mathcal{L}(\Omega, \hat{\pi}) = \frac{-(T-p)n}{2} \log(2\pi) + \frac{T-p}{2} \log |\Omega^{-1}| - \frac{1}{2} \sum_{t=1}^{T-p} \hat{\epsilon}_t' \Omega^{-1} \hat{\epsilon}_t \quad (3.21)$$

donde $\hat{\epsilon}$ son los residuos muestrales. Podemos derivar la función de verosimilitud con respecto a Ω para encontrar un estimador adecuado:

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \Omega^{-1}} &= \frac{T-p}{2} \frac{\partial \log |\Omega^{-1}|}{\partial \Omega^{-1}} - \frac{1}{2} \frac{\sum_{t=1}^{T-p} \partial \hat{\epsilon}_t' \Omega^{-1} \hat{\epsilon}_t}{\partial \Omega^{-1}} \\ &= \frac{T-p}{2} \Omega' - \frac{1}{2} \sum_{t=1}^{T-p} \hat{\epsilon}_t \hat{\epsilon}_t' = 0 \end{aligned} \quad (3.22)$$

Por lo que el estimador asintótico de Ω es:

$$\hat{\Omega} = \frac{1}{T-p} \sum_{t=1}^{T-p} \hat{\epsilon}_t \hat{\epsilon}_t' \quad (3.23)$$

Introduciendo (3.23) en (3.20) podemos estimar el VAR(p), aún si Ω es desconocido.

$$\begin{aligned} \log \mathcal{L}(\hat{\Omega}, \hat{\pi}) &= \frac{-(T-p)n}{2} \log(2\pi) + \frac{T-p}{2} \log |\hat{\Omega}^{-1}| - \frac{1}{2} \sum_{t=1}^{T-p} \hat{\epsilon}_t' \hat{\Omega}^{-1} \hat{\epsilon}_t \\ &= \frac{-(T-p)n}{2} \log(2\pi) + \frac{T-p}{2} \log |\hat{\Omega}^{-1}| - \frac{(T-p)n}{2} \end{aligned} \quad (3.24)$$

La inclusión del estimador de Ω , como se muestra en (3.24) produce la llamada *función de verosimilitud condensada*, que es la que usualmente se utiliza para el análisis empírico. La estimación se hace por medio de métodos numéricos.

Esta función nos permite pensar en otra aplicación útil. Si además del VAR(p) estimamos además un VAR con p^* rezagos, podemos desarrollar un test de largo de rezagos óptimos.

- La H_0 es que el número óptimo de rezagos es p y se computa $\log \mathcal{L}(\hat{\Omega}_p)$.
- La H_1 es que el número óptimo de rezagos es p^* y se calcula $\log \mathcal{L}(\hat{\Omega}_{p^*})$

obviamente,

$$2\left[\log \mathcal{L}(\hat{\Omega}_{p^*}) - \log \mathcal{L}(\hat{\Omega}_p)\right] = T\left[\log |\hat{\Omega}_{p^*}| - \log |\hat{\Omega}_p|\right] \rightarrow \chi^2_{n^2(p^*-p)} \quad (3.25)$$

donde los grados de libertad vienen de n ecuaciones con (p^*-p) restricciones.

Este test es de uso común pero no necesariamente es óptimo en términos de poder. Alternativas frecuentemente usadas son los tests de contenido informacional – como los de Akaike y Schwartz descrito en el Capítulo 2– y aplicados a la matriz de varianza y covarianza de los residuos del sistema de ecuaciones.

En el ejemplo descrito en la Sección 3.04 se usó, arbitrariamente, seis rezagos. El procedimiento *Optlag* del Apéndice B aplicado a estos datos señala, sin embargo, que el número óptimo es 7 rezagos.

3.06 Funciones Impulso-Respuesta del VAR

Al igual que en el caso de los modelos univariados, existe interés en estudiar las respuestas de las variables frente a impulsos dados en forma de innovaciones.¹⁶ Recordemos que la respuesta a un impulso queda descrita por la representación de media móvil de orden infinito del sistema. Tomemos el modelo de forma reducida (ecuación 3.5).

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{11} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix} \quad (3.26)$$

por lo tanto, la representación de media móvil es

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-i} \begin{bmatrix} e_{t-i}^1 \\ e_{t-i}^2 \end{bmatrix} \quad (3.27)$$

La expresión (3.26) no es útil. La interpretación económica de las funciones impulso-respuestas descritas en el capítulo anterior no puede hacerse usando las perturbaciones de la forma reducida del modelo sino respecto de las innovaciones de la

16 El “impulso” que se le da a un sistema dinámico para perturbarlo corresponde a una función tipo *delta* de L. Kronecker en tiempo discreto o de P. Dirac si se trata de tiempo continuo. Al respecto de la relación entre modelos y realidad, Dirac señala: *This result is too beautiful to be false; it is more important to have beauty in one's equations than to have them fit experiment.* “The evolution of the Physicist's Picture of Nature”, *Scientific American*, 208(5), 1963.

forma primitiva. Lo que queremos es la respuesta de y_t o z_t frente a las innovaciones (ϵ^y y ϵ^z) y no a frente a perturbaciones del tipo e^1 o e^2 .

Por ello, tenemos que imponer la condición de identificación. Así,

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-i} \begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{t-i}^y \\ \epsilon_{t-i}^z \end{bmatrix} \quad (3.28)$$

De nuevo vemos que la respuesta dinámica del sistema es su representación VMA(∞). Basta llamar ϕ a los términos a la derecha de la suma (con $\bar{\mu} = [\bar{y}, \bar{z}]'$).

$$x_t = \bar{\mu} + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} \quad (3.29)$$

En la jerga de sistemas dinámicos se suele llamar “multiplicadores” a los ϕ_i . Así, $\phi_i(0)$ es llamado el multiplicador de impacto, $\phi_i(T)$ es el multiplicador de largo plazo, etc.

Se debe hacer notar varias características de las respuestas:

- Dado que se cumple sumabilidad absoluta, las respuestas desaparecen en el largo plazo.
- La morfología de las respuestas no depende del tamaño de las innovaciones, pero el nivel sí. Es decir, una innovación de tamaño 1 tiene la misma forma de disipación que una de tamaño 2, pero a un nivel que es $1/2$ del otro.
- Las respuestas dependen de la identificación impuesta, por lo que de nuevo se deben presentar varias descomposiciones. Este es el concepto de ordenamiento.
- Podemos dar una innovación y ver respuesta de todas las variables y/o podemos comparar las respuestas de cada variable frente a varios shocks.

3.07 Funciones Impulso-Respuesta usando Técnica Espacio-Estado

Obtener las funciones impulso-respuesta –que describen la dinámica del sistema de ecuaciones del VAR– podría ser bastante difícil si el número de ecuaciones y rezagos es alto. No obstante, existe una técnica que hace este paso trivial. A modo de ejemplo, vamos a transformar un VARMA(2,1) en un modelo AR(1):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} \quad (3.30)$$

podemos mapear este proceso en esta otra representación:

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \epsilon_t \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \theta_1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \epsilon_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \epsilon_t \quad (3.31)$$

si ahora definimos

$$x_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \epsilon_t \end{bmatrix} \quad (3.32)$$

El modelo de la ecuación (3.30) puede ser re-escrito directamente como un AR(1):

$$x_t = Ax_{t-1} + C \epsilon_t \quad (3.33)$$

Entonces podemos aplicar directamente la función de impulso-respuesta que obtuvimos en el caso del AR(1) univariado, sólo que sin olvidar que las respuestas son multivariadas.

En este caso, si damos un impulso (shock o innovación) de tamaño 1 en el instante cero y luego no perturbamos el sistema, la respuesta del mismo queda descrita en el Cuadro 3.2:

Cuadro 3.2
Impulso y Respuesta

	0	1	2	3	...	j
ϵ_t	0	1	0	0	...	0
x_t	0	C	AC	A ² C	...	A ^j C

La función de impulso-respuesta es, naturalmente, la primera fila de la secuencia AⁱC para la forma reducida y B⁻¹AⁱC para la forma primitiva. Nótese que la función de impulso-respuesta es exactamente lo mismo que la representación media móvil del sistema.

3.08 Errores Estándares de la Función Impulso-Respuesta

Recordemos que la función impulso-respuesta se construye sobre variables aleatorias (los coeficientes estimados) y, por lo tanto, es también una variable aleatoria. Por ello, queremos conocer su distribución.

Usamos los parámetros estimados \hat{A} y el teorema central del límite para escribir:

$$\sqrt{T}(\hat{A}_T - A) \rightarrow X \quad \text{donde} \quad X \rightarrow N(0, \Omega \otimes Q^{-1}) \quad (3.34)$$

donde Q es el valor esperado de la matriz de momentos de los regresores, i.e., $E(y_t y_t')$. Como lo que queremos es la distribución de la función impulso-respuesta que, a su vez, es una función $f(\cdot)$ de los parámetros, usamos el teorema de Slutsky para obtener,

$$\sqrt{T}(f(\hat{A}_T) - f(A)) \rightarrow G_T X \quad (3.35)$$

donde $G_T = \frac{\partial f(\cdot)}{\partial A}$. Entonces

$$\sqrt{T}[f(\hat{A}_T) - f(A)] \rightarrow N(0, G_T(\Omega \otimes Q^{-1})G_T') \quad (3.36)$$

el error estándar que queremos es directamente la raíz de la diagonal de $G_T(\Omega \otimes Q^{-1})G_T'$. El error estándar asintóticamente estimado, naturalmente, usará el estimador $\hat{\Omega}$.

3.09 Funciones Impulso-Respuesta Generalizadas

Una de las debilidades de la técnica de Sims para analizar VARs es la arbitrariedad de los ordenamientos y la necesidad de presentar muchas alternativas de identificación si se quiere ser riguroso. Pesaran y Shin (1998)¹⁷ proveen una alternativa a la ortogonalización de innovaciones para calcular la función de impulso-respuesta que no depende del ordenamiento de las innovaciones y, en ese sentido, es una función “generalizada”.

¹⁷ Pesaran, H. and Y. Shin (1998): “Generalised Impulse Response Analysis in Linear Multivariate Models”, *Economics Letters*, 58:17-29.

Re-ordenemos lo que hemos aprendido. Supongamos un VAR del tipo:

$$x_t = \sum_{i=0}^p \phi_i x_{t-i} + e_t \quad t=1, T \quad (3.37)$$

donde x_t es un vector de n -variables y Φ es una matriz de parámetros. Los siguientes supuestos caracterizan los errores:

$$\text{Supuesto 1: } E[e_t] = 0 \quad E[e_t, e_t'] = \Omega$$

Supuesto 2: Las raíces características de $|I_n - \sum \phi_i \lambda^i| = 0$ caen fuera del círculo unitario. Es decir las variables son estacionarias.

La representación media móvil de (3.37) es:

$$x_t = \sum_{i=0}^{\infty} A^i e_{t-i} \quad t=1, T \quad (3.38)$$

Las matrices A se obtienen recursivamente como:

$$A^i = \phi_1 A^{i-1} + \phi_2 A^{i-2} + \phi_3 A^{i-3} + \dots + \phi_p A^{i-p} \quad (3.39)$$

La función impulso respuesta es, conceptualmente, equivalente al siguiente experimento: ¿cómo habría reaccionado la economía si hubiese recibido una innovación versus cómo efectivamente evolucionó? Esto depende primero del perfil de las innovaciones, δ , segundo de la historia de la economía, I_{t-1} , y tercero del horizonte de simulación, τ .

La función generalizada de impulso –GI– se define como:

$$GI_x(\tau, \delta, I_{t-1}) = E(x_{t+\tau} | e_t = \delta, I_{t-1}) - E(x_{t+\tau}, I_{t-1}) \quad (3.40)$$

Usando (3.40) en (3.38), obtenemos $GI(\tau, \delta, I_{t-1}) = A^\tau \delta$. Es decir, la función impulso respuesta es independiente de I_{t-1} pero no de δ . Por ello, cómo se modelen e identifiquen las innovaciones resulta clave en el resultado del análisis.

Como hemos discutido, Sims (1980) sugiere utilizar la descomposición de Cholesky, es decir, escoger una matriz triangular tal que:

$$PP' = \Omega \quad (3.41)$$

Así, el VAR en representación media móvil con la restricción de identificación se escribe como:

$$\begin{aligned}
 x_t &= \sum_{i=0}^{\infty} (A^i P) (P^{-1} e_{t-i}) \quad t=1, T \\
 &= \sum_{i=0}^{\infty} (A^i P) \xi_{t-i}
 \end{aligned} \tag{3.42}$$

donde $\xi_t = P^{-1} e_t$, es decir la innovación ortogonalizada. Note que $E(\xi_t \xi_t') = I_n$.

Por ello, el vector de impulsos-respuestas Ψ ortogonalizado de una innovación en la ecuación j en el instante $t + \tau$ es:

$$\Psi_j^0(\tau) = A^\tau P \cdot I(j) \tag{3.43}$$

donde $I(j)$ es un vector “indicador” que selecciona la ecuación j del sistema (1 en j , 0 en cualquier otro lugar).

Una propuesta diferente sería usar (3.40) pero, en vez de darle una innovación a todos los componentes de e_t , darle una innovación sólo a un elemento y desconectarle del efecto de las otras innovaciones –observadas o simuladas sobre la base de alguna distribución–. Es decir,

$$GI_x(n, \delta, I_{t-1}) = E(x_{t+n} | e_t = \delta_j, I_{t-1}) - E(x_{t+n}, I_{t-1}) \tag{3.44}$$

Koop et al. (1996)¹⁸ demuestran que, si las innovaciones se distribuyen normal, entonces:

$$E(e_t | e_{jt} = \delta_j) = \eta_j \sigma_{jj}^{-1} \delta_j \tag{3.45}$$

donde $\eta_j = E[e_t, e_{jt}]$ –es decir la covarianza temporal– y $\sigma_j = E[e_{jt}, e_{jt}]$ es la varianza temporal. Por ello, el vector de impulsos-respuestas generalizadas (no escaladas) de una innovación en la ecuación j en el instante $t+n$ es:

$$\left(\frac{A^\tau \Omega e_j}{\sqrt{\sigma_{jj}}} \right) \left(\frac{\delta_j}{\sqrt{\sigma_{jj}}} \right) \tag{3.46}$$

la respuesta normalizada se obtiene si $\delta_j \sqrt{\sigma_{jj}}$. Es decir,

$$\Psi_j^g(\tau) = \sigma_{jj}^{1/2} A^\tau \Sigma I(j) \tag{3.47}$$

En resumen, las diferencias entre las funciones de impulso-respuesta estándares y las generalizadas son:

18 Koop, G., M. H. Pesaran, S. M. Potter (1996): “Impulse response analysis in nonlinear multivariate models”, *Journal of Econometrics* 74: 19-147.

- Ψ_j^0 es sensible al ordenamiento (ortogonalización) de los innovaciones, en cambio Ψ_j^s no es sensible.
- Ψ_j^0 y Ψ_j^s coinciden si Σ es diagonal.
- Si Σ no es diagonal, Ψ_j^0 y Ψ_j^s coinciden sólo en $j=1$ (es decir, para la primera variable del VAR).
- mientras las funciones impulso-respuesta ordinarias están normalizadas a uno, $\sum_{j=1}^{\infty} \theta_{jj}^0 = 1$, las funciones impulso-respuesta generalizadas no lo están, $\sum_{j=1}^{\infty} \theta_{jj}^s \neq 1$, perdiendo comparabilidad.

3.10 Descomposicion de la Varianza del Error de Pronóstico

Consideremos la ecuación (3.36), el error de pronóstico un paso adelante es: $x_{t+1} - E_t(x_{t+1}) = \epsilon_{t+1}$. El error de pronóstico τ pasos adelante es

$$x_{t+\tau} - E_t(x_{t+\tau}) = \sum_{i=0}^{\tau-1} \phi^i \epsilon_{t+\tau-i} \quad (3.48)$$

Tomemos el error de pronóstico τ pasos adelante sólo para el componente y_t pero en función de sus componentes estructurales.

$$\begin{aligned} y_{t+\tau} - E_t(y_{t+\tau}) &= \phi_{11}(0)\epsilon_{t+\tau}^y + \phi_{11}(1)\epsilon_{t+\tau-1}^y + \dots + \phi_{11}(\tau-1)\epsilon_{t+1}^y \\ &\quad + \phi_{12}(0)\epsilon_{t+\tau}^z + \phi_{12}(1)\epsilon_{t+\tau-1}^z + \dots + \phi_{12}(\tau-1)\epsilon_{t+1}^z \end{aligned} \quad (3.50)$$

Tomemos ahora la varianza de este error de predicción sin olvidar que las innovaciones son no correlacionadas

$$\begin{aligned} \sigma_y^2(\tau) &= \sigma_y^2[\phi_{11}(0)^2 + \phi_{11}(1)^2 + \dots + \phi_{11}(\tau-1)^2] + \sigma_z^2[\phi_{12}(0)^2 + \phi_{12}(1)^2 + \dots + \phi_{12}(\tau-1)^2] \\ \sigma_z^2(\tau) &= \sigma_y^2[\phi_{21}(0)^2 + \phi_{21}(1)^2 + \dots + \phi_{21}(\tau-1)^2] + \sigma_z^2[\phi_{22}(0)^2 + \phi_{22}(1)^2 + \dots + \phi_{22}(\tau-1)^2] \end{aligned} \quad (3.51)$$

La ecuación (3.51) es la descomposición de la varianza del error de pronóstico. ¿Cómo entendemos esta descomposición? Lo que tenemos es la suma de varianzas de las innovaciones “estructurales” (ϵ_t) ponderadas por su importancia en la estructura dinámica (memoria) de cada variable (ϕ_i). Es decir, la incertidumbre de la predicción es una combinación variable de la incertidumbre de las innovaciones originales condicional en el proceso de identificación.

Una consecuencia de lo anterior es que, si la mayor parte de la incertidumbre de la predicción de x_t proviene de innovaciones propias, entonces la variable será más bien “exógena”.

El análisis que utilizamos para la función de impulso-respuesta generalizada en la sección anterior es perfectamente válido en este contexto. Las funciones Ψ_j^o y Ψ_j^s pueden ser utilizadas también para obtener las funciones de descomposición de la varianza del error de pronóstico (θ):

$$\theta_{jj}^o(\tau) = \frac{\sum_{i=0}^{\tau} (e_i' A^i P e_j)^2}{\sum_{i=0}^{\tau} e_i' A^i \Omega A^{i'} e_j} \quad \theta_{jj}^s(\tau) = \frac{\sigma_{jj}^{-1} \sum_{i=0}^{\tau} (e_i' A^i \Omega e_j)^2}{\sum_{i=0}^{\tau} e_i' A^i \Omega A^{i'} e_j} \quad (3.52)$$

3.11 VAR Estructurales

El objetivo de un VAR estructural (SVAR) es obtener una ortogonalización no recursiva de los residuos para la función impulso-respuesta basada no en la descomposición de Cholesky sino en la teoría económica. Ello, naturalmente, requiere imponer restricciones de modo de identificar la forma primitiva desde la reducida.

Dado x_t de orden k se busca realizar el *mapping* $Ae_t = B\epsilon_t$ donde ϵ_t son las innovaciones primitivas y e_t los residuos de la forma reducida. Sea Ω la matriz de covarianza de los residuos. Asumiendo (por comodidad) que las innovaciones son ortonormales (es decir, con varianza 1), entonces:

$$A\Omega A' = BB' \quad (3.53)$$

Como las matrices de (3.53) son simétricas el supuesto de ortonormalidad provee $k(k+1)/2$ restricciones sobre los $2k^2$ elementos de A y B. Se requiere proveer $2k^2 - k(k+1)/2$ restricciones adicionales.

La especificación de las restricciones de identificación es frecuentemente de dos tipos: “corto plazo” y “largo plazo”. Para aquellas de corto plazo, usualmente basta con poner ceros.

Más generalmente, se pueden imponer restricciones del tipo:

$$\begin{aligned} \epsilon_1 &= c(1) \cdot e_1 \\ \epsilon_2 &= -c(2) \cdot e_1 + c(3) \cdot e_2 \end{aligned} \quad (3.54)$$

Otra manera de incorporar “estructura” en el VAR es imponer restricciones de largo plazo. Blanchard y Quah (1989)¹⁹ proponen identificar sobre la base de restringir las propiedades de largo plazo de la función impulso-respuesta. La respuesta de largo plazo (acumulada) de las innovaciones primitivas es:

$$C = \Psi_{\infty} A^{-1} B \quad (3.55)$$

donde $\Psi_{\infty} = (I - \hat{A}_1 - \hat{A}_2 - \dots - \hat{A}_p)^{-1}$ es decir la respuesta acumulada estimada de la forma reducida. La restricción de largo plazo afecta los elementos de la matriz C –por ejemplo, $C_{ij}=0$ – lo que permite la identificación. La restricción anterior dice que la respuesta acumulada de la i -ésima variable a la innovación “estructural” j -ésimo es cero.

Existen algunas consideraciones en el uso de un SVAR:

- Las matrices A y B deben ser cuadradas y no-singulares, i.e., debe haber tantas ecuaciones como variables endógenas.
- Las restricciones deben ser lineales.
- La restricción que las innovaciones tengan varianza unitaria es innecesaria, pero ellas deben ser no correlacionadas entre ecuaciones y en el tiempo.
- Es posible que las restricciones no provean “signo” para un coeficiente (ej, $\sqrt{a_{11}} = b_{23}$). Ver Christiano, Eichenbaum, and Evans (1998)²⁰ para una descripción de este problema.
- A diferencia de un VAR estándar, se puede testear el modelo por el método de “over-identifying restrictions”. Es frecuente que una “teoría” provea muchas restricciones para los parámetros o entre ecuaciones. De hecho, podrían sobre-identificar. Ello permite testear el modelo con restricciones y sin restricciones usando un test χ^2 .

3.12 Cuasi VARs

Como es obvio, los modelos VAR son intensivos en parámetros. En ocasiones ello no es problema, pero en muestra pequeña puede ser costoso. Una extensión de los modelos VARs son los cuasi-VAR, que pueden ser vistos como un híbrido entre un modelo VAR puro y un modelo EST. Es decir, aquellos VARs donde algunas variables son consideradas exógenas ex-ante pero su dinámica y efecto sobre las variables endógenas es determinado al interior del modelo. Es decir, excluimos del VAR la ecuación para la variable considerada exógena y la incluimos como una variable determinística.

19 Blanchard, O.J. and D. Quah (1989), “The Dynamic Effects of Aggregate Supply and Demand Disturbances”, *American Economic Review*, 79:655-673.

20 Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1998. “Modeling Money.” Working Paper no. 6371 (January), NBER, Cambridge, MA.

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11}(L) & a_{12}(L) \\ a_{21}(L) & a_{22}(L) \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix} + \begin{bmatrix} b_{11}(L) & b_{12}(L) \\ b_{21}(L) & b_{22}(L) \end{bmatrix} \begin{bmatrix} m_{t-1} \\ s_{t-1} \end{bmatrix} \quad (3.56)$$

En muchos casos, el uso de un modelo cuasi-VARs es justificado (p.e., la tasa de interés internacional en un modelo para una economía pequeña, la demanda de bienes exportables, etc.). La justificación de la especificación, no obstante, debe estar bien hecha.

Apéndice A: Modelos de Ecuaciones Simultáneas Tradicionales

Los modelos de ecuaciones simultáneas han sido la base de muchos desarrollos econométricos en los últimos 50 años y proveen una herramienta útil para modelar problemas económicos. En este apéndice presentamos los dos aspectos más destacados de esta literatura: la derivación formal de las condiciones de identificación y las técnicas de estimación de sistema completo.

Una manera práctica de describir un modelo de ecuaciones simultáneas es:

$$\begin{aligned} \gamma_{11}y_t^1 + \gamma_{12}y_t^2 + \dots + \gamma_{1m}y_t^m + \beta_{11}x_t^1 + \beta_{12}x_t^2 + \dots + \beta_{1k}x_t^k &= \epsilon_t^1 \\ \gamma_{21}y_t^1 + \gamma_{22}y_t^2 + \dots + \gamma_{2m}y_t^m + \beta_{21}x_t^1 + \beta_{22}x_t^2 + \dots + \beta_{2k}x_t^k &= \epsilon_t^2 \\ &\vdots \\ \gamma_{m1}y_t^1 + \gamma_{m2}y_t^2 + \dots + \gamma_{mm}y_t^m + \beta_{m1}x_t^1 + \beta_{m1}x_t^2 + \dots + \beta_{mk}x_t^k &= \epsilon_t^m \end{aligned} \quad (A.1)$$

donde las variables y son llamadas tradicionalmente “endógenas” y las variables x “exógenas”. Note que hay m variables endógenas y k exógenas. Además hay m ecuaciones, x_1 es una columna de unos (constante) y ϵ son innovaciones.

Podemos escribir el modelo de manera más simple usando matrices, de modo que :

$$\begin{bmatrix} y_t^1 & y_t^2 & \dots & y_t^m \end{bmatrix} \begin{bmatrix} \gamma_{11} & \dots & \gamma_{1m} \\ \gamma_{21} & \dots & \gamma_{2m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \dots & \gamma_{mm} \end{bmatrix} + \begin{bmatrix} x_t^1 & x_t^2 & \dots & x_t^k \end{bmatrix} \begin{bmatrix} \beta_{11} & \dots & \beta_{1k} \\ \gamma_{21} & \dots & \gamma_{2k} \\ \vdots & \ddots & \vdots \\ \gamma_{k1} & \dots & \gamma_{kk} \end{bmatrix} = \begin{bmatrix} \epsilon_t^1 & \epsilon_t^2 & \dots & \epsilon_t^k \end{bmatrix} \quad (A.2)$$

De manera más compacta:

$$y_t' \Gamma + x_t' B = \epsilon_t' \quad (A.3)$$

A.1 El problema de identificación

Antes de derivar la técnica general de identificación, estudiemos un caso particular para entender este problema en el contexto de ecuaciones simultáneas. Sea el siguiente modelo

$$\begin{array}{ll}
\text{Demanda} & q_d = \alpha_1 p_t + \alpha_2 y_t + \epsilon_t^d \\
\text{Oferta} & q_s = \beta_1 p_t + \epsilon_t^s \\
\text{Equilibrio} & q_s = q_d
\end{array} \tag{A.4}$$

con $E[\epsilon^d] = E[\epsilon^s] = 0$ y $E[\epsilon^d, y] = E[\epsilon^s, y] = E[\epsilon^d, \epsilon^s] = 0$.

Resulta evidente que p y q son endógenos, y es exógeno y las dos ecuaciones no son independientes. Podemos resolver el sistema de dos incógnitas y dos ecuaciones:

$$\begin{aligned}
q_t &= \frac{\beta_1 \alpha_2 y_t}{\beta_1 - \alpha_1} + \frac{\epsilon_t^d - \epsilon_t^s}{\beta_1 - \alpha_1} \\
p_t &= \frac{\alpha_2 y_t}{\beta_1 - \alpha_1} + \frac{\beta_1 \epsilon_t^d - \alpha_1 \epsilon_t^s}{\beta_1 - \alpha_1}
\end{aligned} \tag{A.5}$$

Note que hay problemas derivados de $Cov[p_t, \epsilon_t^d] = \frac{\sigma_{\epsilon^d}^2}{\beta_1 - \alpha_1}$. Naturalmente, debemos restringir $\alpha_1 \neq \beta_1$ pero ese no es el problema más grave. Si quisiéramos estimar la ecuación de demanda directamente, el estimador $\hat{\alpha}_1$ sería sesgado. Este es el llamado sesgo de simultaneidad: cada vez que hay variables endógenas en el lado derecho de la ecuación, los parámetros están sesgados. La ecuación (A.4) es *mal llamada* la representación “estructural”.

Reescribamos la ecuación (A.5) de modo que conforme la llamada “forma reducida”

$$\begin{aligned}
q_t &= \pi_1 y_t + v_t^1 \\
p_t &= \pi_2 y_t + v_t^2
\end{aligned} \tag{A.6}$$

la que sí puede ser estimada directamente porque no hay covarianza entre los regresores y los residuos v_1 o v_2 .

Una vez obtenidos los estimadores $\hat{\pi}_1$ y $\hat{\pi}_2$ notemos que:

$$\pi_1 = \frac{\alpha_2}{\beta_1 - \alpha_1} \quad \pi_2 = \frac{\beta_1 \alpha_2}{\beta_1 - \alpha_1} \tag{A.7}$$

por lo tanto, podemos obtener el estimador $\hat{\beta}_1 = \frac{\hat{\pi}_2}{\hat{\pi}_1}$ y, naturalmente, la ecuación de oferta. Sin embargo no podemos obtener la ecuación de demanda de la misma forma. En

la jerga de ecuaciones simultáneas se dice que esta ecuación de demanda está sub-identificada, en tanto que la ecuación de oferta está identificada.

Si volvemos al modelo general de la ecuación (A.2), su forma reducida será del tipo:

$$\begin{bmatrix} y_t^1 & y_t^2 & \dots & y_t^m \end{bmatrix} = \begin{bmatrix} x_t^1 & x_t^2 & \dots & x_t^k \end{bmatrix} \begin{bmatrix} \pi_{11} & \dots & \pi_{1k} \\ \pi_{21} & \dots & \pi_{2k} \\ \dots & \dots & \dots \\ \pi_{k1} & \dots & \pi_{kk} \end{bmatrix} + \begin{bmatrix} v_t^1 & v_t^2 & \dots & v_t^m \end{bmatrix} \quad (\text{A.8})$$

Frente al modelo general cabe preguntarse cómo podemos saber cuándo una ecuación está identificada de modo que podamos estimar los parámetros de la forma reducida y obtener los parámetros de la forma estructural.

Para ello, estudiemos la ecuación (A.2)

- en la matriz Γ hay m^2 parámetros desconocidos (podemos hacer que los parámetros del tipo γ_{ii} sean 1, por lo que hay m^2-m incógnitas).
- en la matriz B hay mk parámetros desconocidos
- en la matriz de varianzas y covarianzas de los errores hay $\frac{1}{2}m(m-1)$ incógnitas (las covarianzas) porque la matriz es simétrica.

Estudiamos, ahora, la ecuación (A.8):

- en la matriz Π hay mk parámetros estimados
- en la matriz de varianzas y covarianzas de los residuos hay $\frac{1}{2}m(m-1)$ parámetros estimados (las covarianzas) porque la matriz es simétrica.

Por lo tanto, no es posible resolver el sistema de ecuaciones porque hay más incógnitas que parámetros estimados. Es por ello, que los modelos de ecuaciones simultáneas deben imponer restricciones de identificación, las que toman diferentes formas:

- En ocasiones algunas ecuaciones son identidades, es decir, ecuaciones con parámetros conocidos (por ejemplo, $Y \equiv C + I + G$).
- En otros casos, se imponen restricciones cruzadas entre parámetros (por ejemplo, de una función de producción Cobb-Douglas con retornos constantes de escala obtenemos $\log Y = \beta \log K + (1-\beta) \log L$).
- Se restringen innovaciones porque se sabe que no covarían ($E[\varepsilon^d, \varepsilon^s] = 0$).

En realidad, en la mayoría de los casos es el econometrista el que impone las restricciones necesarias para poder identificar. Algunas veces éstas se derivan de la teoría, pero ello no es frecuente. Típicamente, éstas son arbitrarias.

Método general de determinar el grado de identificación

Para evaluar si una ecuación está identificada, separaremos las variables endógenas y exógenas en grupos:

Cuadro Apéndice 1

	Incluidas en la ecuación	Excluidas de la ecuación
Endógenas	y_j	x_j
Exógenas	y_j^*	x_j^*

Por lo tanto la j -ésima ecuación estructural puede ser escrita como:

$$y_j = Y_j' \Gamma_j + Y_j^{*'} \Gamma_j^* + x_j' B_j + x_j^{*'} B_j^* + \epsilon_j \quad (\text{A9})$$

en tanto que la j -ésima ecuación reducida es:

$$[y_j \ Y_j' \ Y_j^{*'}] = [x_j \ x_j^{*'}] \begin{bmatrix} \pi_j & \Pi_j & \bar{\Pi}_j \\ \pi_j^* & \Pi_j^* & \bar{\Pi}_j^* \end{bmatrix} + \epsilon_j \quad (\text{A10})$$

Al buscar la equivalencia para volver de (A.10) a (A.9) encontramos las siguientes condiciones de identificación:

$$\begin{aligned} \Pi_j - \Pi_j \gamma_j &= \beta_j \\ \Pi_j^* - \Pi_j^* \gamma_j &= 0 \end{aligned} \quad (\text{A11})$$

De la primera condición no podemos derivar mucho, pues depende de β que es desconocida. De la segunda condición tenemos K_j^* ecuaciones y M_j incógnitas. Si $K_j^* \geq M_j$, el problema tiene solución. Esta es la más simple regla de identificación, llamada condición de orden:

Si el número de variables exógenas excluidas de la ecuación es mayor o igual que el número de endógenas incluidas, entonces la ecuación está identificada.

El problema de la condición de orden radica en que es posible que alguna o algunas de las columnas de K_j^* o de M_j Sean linealmente dependientes de las otras. Por ello, la condición de orden no es suficiente. Lo que queremos es que *el rango de la matriz K_j^* sea mayor o igual al rango de la matriz M_j* . No muy sorprendentemente, esta condición es llamada condición de rango.

Habr , entonces, tres casos

- $K_j^* = M_j$ la ecuaci n est  exactamente identificada
- $K_j^* \geq M_j$ la ecuaci n est  sobre identificada
- $K_j^* < M_j$ la ecuaci n est  sub identificada

Tanto en el primero como segundo caso, la estimaci n es posible. En el tercero, no es posible identificar los par metros estructurales a partir de la forma reducida.

A.2 M todos y T cnicas de Estimaci n

Antes de revisar los m todos de estimaci n es conveniente notar que s lo se necesita identificar si es que se va a hacer algun test sobre los par metros estructurales. Si se desea predecir exclusivamente, basta con estimar la ecuaci n reducida (3.66).

Vamos a separar los m todos de estimaci n en dos familias: (1) estimaci n ecuaci n por ecuaci n y (2) estimaci n del sistema completo. Tambi n vamos a separar por t cnica de estimaci n: (1) m nimos cuadrados y (2) m xima verosimilitud. Es decir, hay cuatro casos para estudiar:

Cuadro Ap ndice 2

M�todo de estimaci�n	T�cnica de estimaci�n	
	M�nimos cuadrados	M�xima verosimilitud
Ecuaci�n por ecuaci�n	A	B
Sistema completo	C	D

Métodos de Estimación Ecuación por Ecuación

Caso A: Técnica de mínimos cuadrados

Es importante preguntarse cuándo es legítimo estimar ecuación por ecuación. Existe un tipo de modelo de ecuaciones simultáneas que es posible estimar eficientemente ecuación por ecuación, el llamado modelo recursivo.

$$\begin{aligned} y_t^1 &= \beta_1 x_t^1 + \epsilon_t^1 \\ y_t^2 &= \beta_2 x_t^1 + \gamma_{21} y_t^1 + \epsilon_t^2 \\ &\vdots \\ y_t^m &= \beta_m x_t^1 + \gamma_{m1} y_t^1 + \gamma_{m2} y_t^2 + \dots + \gamma_{ml} y_t^l + \epsilon_t^m \end{aligned} \quad (A12)$$

Note que este modelo se puede estimar de la siguiente manera: (1) estime la primera ecuación por OLS (no hay problema por que no hay sesgo de simultaneidad), (2) prediga y_t^1 usando la ecuación anterior y reemplaze \hat{y}_t^1 por y_t^1 en la segunda ecuación, (3) estime la segunda ecuación, (4) prediga \hat{y}_t^2 usando la ecuación anterior, etc.

Un segundo caso sucede cuando es lícito usar mínimos cuadrados indirectos de modo similar al usado en el ejemplo de oferta y demanda. Si en la condición de identificación -ecuación (A.11)- el rango de la matriz K_j^* es igual al rango de la matriz M_j entonces

$$\Pi_j^* - \Pi_j^* \gamma_j = 0 \Rightarrow \hat{\gamma}_j = \Pi_j^{*-1} \Pi_j^* \quad (A13)$$

En el caso en que el rango de la matriz K_j^* es mayor que el rango de la matriz M_j la pregunta más interesante es ¿cómo usar este “exceso” de información de manera creativa? Obviamente podríamos botar alguna columna, pero eso raramente es óptimo. Una alternativa interesante es utilizar lo que aprendimos en los métodos de variables instrumentales.

En primer lugar, note que si la ecuación está exactamente identificada entonces hay una variable exógena excluida por cada endógena incluida, es decir un instrumento por cada variable de lado dercho que tiene covarianza con el residuo. En segundo lugar, si rango de la matriz K_j^* es mayor que el rango de la matriz M_j habrá más instrumentos de lo estrictamente necesario. Pero este problema ya sabemos enfrentarlo usando un estimador en dos etapas: (1) busquemos la combinación lineal de X_j^* que mejor represente (prediga) Y_j usando mínimos cuadrados, y (2) estimemos la j -ésima ecuación usando la predicción hecha en la primera etapa. Este es el llamado estimador de mínimos cuadrados en dos etapas (2SLS).

Caso B: Técnica de máxima verosimilitud

Una alternativa es hacer la estimación por máxima verosimilitud que, a diferencia de 2SLS, es asintóticamente normal independientemente de la distribución de los residuos. En principio, la función de verosimilitud de la j -ésima ecuación es:

$$\max_{\beta} = \frac{-n}{2} [(m_j + 1) \log 2\pi + \log |\Omega_j|] - \frac{1}{2} \sum_{i=1}^n [\gamma_j - x_j' \Pi_j] \Omega_j^{-1} [\gamma_j - x_j' \Pi_j]' \quad (A14)$$

sujeto a

$$\begin{aligned} \Pi_j - \Pi_j \gamma_j &= \beta_j \\ \Pi_j^* - \Pi_j^* \gamma_j &= 0 \end{aligned} \quad (A15)$$

Existe un algoritmo que soluciona este problema de maximización y que es más simple que tratar de estimar los parámetros directamente usando la función de verosimilitud.

- Sea $W_j^0 = E_j^0' E_j^0$ donde $E_j^0 = M_j Y_j^0 = [I - X_j (X_j' X_j)^{-1} X_j'] Y_j^0$. Es decir E_j^0 son los residuos de una regresión de Y_j^0 en X_j .
- Definimos, además, $W_j^1 = E_j^1' E_j^1$ donde $E_j^1 = [I - X (X' X)^{-1} X'] Y_j^0$. Es decir, los residuos de una regresión de Y_j^0 en X .
- Computamos $\lambda = (W_j^1)^{-1} W_j^0$, la raíz característica más pequeña. Sea $W_j^0 = \begin{bmatrix} w_{jj}^0 & w_j^{0'} \\ w_j^0 & W_{jj}^{0'} \end{bmatrix}$ y $W_j^1 = \begin{bmatrix} w_{jj}^1 & w_j^{1'} \\ w_j^1 & W_{jj}^{1'} \end{bmatrix}$.
- Entonces los estimadores de máxima verosimilitud son $\hat{\gamma}_{j.LML} = [W_{jj}^0 - \lambda W_{jj}^1]^{-1} (w_j^0 - \lambda w_j^1)$ y $\hat{\beta}_{j.LML} = [X_j' X_j]^{-1} [X_j' (\gamma_j - Y_j \hat{\gamma}_{j.LML})]$

Métodos de Estimación del Sistema Completo

La principal diferencia entre la estimación de un modelos como sistema en vez de ecuación por ecuación es aprovechar la información que pueda haber sobre la correlación de residuos de modo de hacer más eficiente la estimación. Es decir, los métodos uniecuacionales dan cuenta del problema de inconsistencia, en tanto que los métodos de sistema dan cuenta, además, del problema de eficiencia. Es decir, estos últimos hacen una corrección tipo mínimos cuadrados generalizados junto con variables instrumentales.

Caso C: Técnica de mínimos cuadrados

Este método, llamado mínimos cuadrados en tres etapas, consiste en:

1. Obtenga \hat{Y}_j para cada ecuación
2. Compute el estimador de mínimos cuadrados en dos etapas, $\hat{\delta}_{2SLS}$.
3. Compute el estimador de la covarianza de los residuos entre ecuaciones,

$$\sigma_{ij} = \frac{(y_j - \hat{\delta}_j' Z_j)' (y_i - \hat{\delta}_i' Z_i)}{T}$$
4. Obtenga la varianza asintótica de los estimadores, $As.Var(\hat{\delta}_j) = (Z' \hat{\Sigma}^{-1} Z)^{-1}$.

Caso D: Técnica de máxima verosimilitud

La última alternativa es usar el estimador de máxima verosimilitud, que en este caso corresponde a:

$$\begin{aligned} \max \log L &= \frac{-T}{2} [M \log 2\pi + \log |\Omega| + tr(\Omega^{-1} N)] \\ \text{sueto a} & \\ \Pi &= -B \Gamma^{-1} \\ \Omega &= (\Gamma^{-1})' \Sigma \Gamma^{-1} \end{aligned} \tag{A16}$$

donde los elementos de Ω son $\omega_{ij} = \frac{(y_i - x_i' \Pi)' (y_j - x_j' \Pi)}{T}$.

La solución de (A.16), el llamado estimador de máxima verosimilitud con información completa o FIML, requiere del uso de métodos de cálculo numérico.

Comparación de Métodos

- El único método que usa toda la información eficientemente es FIML (asintótico).
- Si los residuos se distribuyen normales, entonces 3SLS es eficiente.
- Si la ecuación está exactamente identificada, FIML equivale a OLS ecuación por ecuación.
- Si la muestra es grande ($n \rightarrow \infty$), los métodos FIML y 3SLS dominan a los uniecuacionales.
- Si hay problemas de especificación, los métodos uniecuacionales son preferidos a multiecuacionales porque no se “contagian”.

Simulación de Modelos de Ecuaciones Simultáneas

El principal objetivo de los modelos de ecuaciones simultáneas es obtener predicción de las variables endógenas, condicional en distintos supuestos sobre la evolución futura de las variables exógenas. La jerga de la simulación de estos modelos distingue entre efectos (llamados “multiplicadores”) de corto plazo (llamado “impacto”) de los de largo plazo. También distingue entre simulaciones estáticas (sin retroalimentación de las variables predeterminadas) y simulaciones dinámicas (con retroalimentación).

Sea el modelo $y_t' \gamma + x_t' \beta + y_{t-1}' \phi = \epsilon_t'$. Su forma reducida es:

$$y_t' = x_t' \Pi + y_{t-1}' \Delta + \epsilon_t' \quad (\text{A17})$$

Multiplicador Impacto. El efecto instantáneo de la variables x_k en la ecuación m -ésima es simplemente:

$$\frac{\partial y_t^m}{\partial x_t^k} = \Pi_{km} \quad (\text{A18})$$

Multiplicadores de Largo Plazo. Recuerde que si $y_t' \gamma + x_t' \beta + y_{t-1}' \phi = \epsilon_t'$, entonces $y_{t-1}' \gamma + x_{t-1}' \beta + y_{t-2}' \phi = \epsilon_{t-1}'$. Por ello, cada vez que cambia una variable exógena, las variables endógenas siguen una trayectoria, es decir tienen una evolución en el tiempo.

$$\frac{\partial y_t^m}{\partial x_{t-s}^k} = (\Pi \Delta^s)_{km} \quad (\text{A19})$$

Simulación Dinámica vs. Estática Note que cuando se simula el modelo $y_t' = x_t' \Pi + y_{t-1}' \Delta + \epsilon_t'$ dentro de la muestra hay dos maneras de tratar el valor de los rezagos de las variables endógenas: el predicho por el modelo (simulación dinámica) o el efectivamente observado (simulación estática).

Apéndice B: Ejercicios

1. ¿A qué se refieren los econométristas cuando señalan que la descomposición de Cholesky es una forma de identificación de mínima restricción en un modelo? Use como ejemplo del rol que juega una restricción de identificación en el siguiente modelo propuesto por Sims. En particular responda, ¿cuántas incógnitas hay en la forma estructural? ¿Cuántos parámetros se obtienen de la forma reducida? ¿Cuántas restricciones se necesitan? ¿Cuáles restricciones impone la técnica de Cholesky?

$$\begin{aligned}m_t &= b_{11}y_t + b_{12}i_t + a_{11}(L)m_{t-1} + a_{12}(L)y_{t-1} + a_{13}(L)i_{t-1} + \epsilon_t^m \\ y_t &= b_{21}m_t + b_{22}i_t + a_{21}(L)m_{t-1} + a_{22}(L)y_{t-1} + a_{23}(L)i_{t-1} + \epsilon_t^y \\ i_t &= b_{31}m_t + b_{32}y_t + a_{31}(L)m_{t-1} + a_{32}(L)y_{t-1} + a_{33}(L)i_{t-1} + \epsilon_t^{i2}.\end{aligned}$$

2. Suponga que Ud. desea obtener la función impulso-respuesta del siguiente modelo: $y_t = 0.7y_{t-1} + 0.2y_{t-2} + \epsilon_t$ con $\epsilon_t \sim N(0, \sigma^2)$. Utilice la técnica de la representación espacio-estado (es decir, convierta el modelo en un modelo tipo AR(1)) y obtenga la función impulso-respuesta generalizada.

3. Considere el siguiente modelo VAR. Le proponen el siguiente algoritmo para testear la causalidad entre y y z : (1) estimar cada modelo por OLS y (2) testear si γ_{ij} es cero. Señale, al menos, tres limitaciones que invalidan el algoritmo.

$$\begin{aligned}y_t &= b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \epsilon_t^y \\ z_t &= b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \epsilon_t^z\end{aligned}$$

4. Describa de manera precisa cómo se deriva la varianza del error de predicción un paso adelante de un VAR(p) bivariado. ¿Cuáles son sus componentes? ¿Cómo se interpretan?
5. Considere dos variables estacionarias. Cada una tiene su correlograma. Suponga que se arma un VAR con ellas y que éstas no son independientes. Entonces habrá un correlograma cruzado (*cross correlogram*) que mira las covariaciones entre variables. ¿qué información tiene este correlograma cruzado? ¿qué relación guarda con un VAR? ¿es posible estimar los coeficientes del VAR a partir de las estimaciones del correlograma cruzado?

Índice

Capítulo 3.....	1
Modelos de Series de Tiempo	1
Estacionarias Multivariados.....	1
3.01 Vectores Autoregresivos.....	1
3.02 El Problema de Identificación.....	3
3.03 Ecuaciones Simultáneas Tradicionales vs. VARs.....	4
3.04 Causalidad en Series de Tiempo.....	6
3.05 Estimación de un VAR(p) por Máxima Verosimilitud.....	10
3.06 Funciones Impulso-Respuesta del VAR.....	12
3.07 Funciones Impulso-Respuesta usando Técnica Espacio-Estado.....	13
3.08 Errores Estándares de la Función Impulso-Respuesta.....	15
3.09 Funciones Impulso-Respuesta Generalizadas.....	15
3.10 Descomposicion de la Varianza del Error de Pronóstico.....	18
3.11 VAR Estructurales.....	19
3.12 Cuasi VARs.....	20
Apéndice A: Modelos de Ecuaciones Simultáneas Tradicionales	22
Apéndice B: Ejercicios.....	31

Última corrección: 06/03/10