

Procesos VAR y Cointegración

Marta Regúlez Castillo¹
Departamento de Economía Aplicada III
(Econometría y Estadística) (UPV-EHU)

Octubre 2006

¹Estas notas han sido elaboradas para el curso de Econometría Financiera II dentro del programa de doctorado en Finanzas Cuantitativas, pero no tienen la intención de suplir los libros de texto recomendados. De hecho en gran parte están basadas en ellos. Agradezco a la profesora Ainhoa Zarraga sus comentarios y correcciones a una versión más antigua de estas notas.

Índice general

1. Introducción	1
1.1. Matrices de correlaciones cruzadas	1
1.2. Dependencia lineal	3
1.3. Contrastes de Portmanteau Multivariantes	9
2. Procesos VAR Estables	11
2.1. Condición de estabilidad	11
2.2. Representación de Medias Móviles	14
2.3. Procesos estacionarios. Condición de estacionariedad	17
2.4. Estimación y Contrastes	19
2.5. Predicción	23
2.5.1. Predicción por punto	23
2.5.2. Predicción por intervalo	28
2.5.3. Predicción con procesos estimados	30
2.6. Causalidad en el sentido de Granger	32
2.6.1. Definición	32
2.6.2. No-Causalidad en el sentido de Granger	33
2.7. Análisis Impulso-Respuesta	36
2.8. Selección del orden del VAR	40
3. Ejemplo VAR con Gretl	45
3.1. Estimación de un VAR	45
3.2. Análisis Impulso-Respuesta	52
3.3. Seleccionando el orden del retardo	56
3.4. Otros contrastes de especificación	58
3.5. Causalidad en el sentido de Granger	63
3.6. Predicción	66
4. Procesos integrados y Cointegración	69
4.1. Difference-stationary versus Trend-stationary	69
4.2. Algunos contrastes de raíces unitarias	74
4.2.1. Contrastes de Dickey-Fuller	74
4.2.2. Contrastes de Phillips y Perron	78

4.2.3. Otros	78
4.3. Regresión espuria y Cointegración	79
4.4. Modelo de corrección del error	82
4.5. Contrastes de cointegración y Estimación	86
4.5.1. Contrastes basados en residuos	86
4.5.2. Método en dos etapas de Engle y Granger	88
4.5.3. Método de Johansen	89
4.5.4. Contrastes sobre el rango de cointegración	91
5. Apéndice	93
5.1. Valores y Vectores propios-Raíces características	93
5.2. Descomposición de Choleski	95
5.3. Sumas infinitas de variables aleatorias	95
5.4. Método de Johansen	97

Capítulo 1

Introducción¹

1.1. Matrices de correlaciones cruzadas

Considerar un vector $y_t = (y_{1t}, \dots, y_{kt})'$ k -dimensional de series temporales. Un proceso estocástico es estacionario (en sentido débil) si sus primeros y segundos momentos son invariantes con el tiempo. Esto es si

i) $E(y_t) = \mu$ para todo t .

ii) $E[(y_t - \mu)(y_{t-h} - \mu)'] = \Gamma_y(h) = \Gamma_y'(-h)$ para todo t y $h = 0, 1, \dots$

La media μ es un vector k -dimensional que consiste en la esperanza no condicionada de los componentes de y_t . La matriz de covarianzas $\Gamma_y(0)$ es una matriz $(k \times k)$. El i -ésimo elemento de la diagonal principal de $\Gamma_y(0)$ es la varianza de y_{it} , mientras que el elemento (i, j) -ésimo de $\Gamma_y(0)$ es la covarianza entre y_{it} e y_{jt} . A sus elementos los denotamos por $\Gamma_{ij}(0)$.

Sea D una matriz diagonal $(k \times k)$ donde la diagonal está formada por las desviaciones típicas de y_{it} para $i = 1, \dots, k$. Esto es,

$$D = \text{diag} \left(\sqrt{\Gamma_{11}(0)}, \dots, \sqrt{\Gamma_{kk}(0)} \right)$$

La matriz de correlaciones cruzadas en el retardo h (CCM) para y_t se define como:

$$\rho(h) = D^{-1} \Gamma_y(h) D^{-1}$$

En concreto, el (i, j) -ésimo elemento de ρ_h es

$$\rho_{ij}(h) = \frac{\Gamma_{ij}(h)}{\sqrt{\Gamma_{ii}(h)} \sqrt{\Gamma_{jj}(h)}} = \frac{\text{Cov}(y_{it}, y_{j,t-h})}{\text{std}(y_{it}) \text{std}(y_{jt})}$$

¹Se basa en el capítulo 8 del libro de Tsay.

donde std denota desviación estandar, esto es el coeficiente de correlación entre y_{it} e $y_{j,t-h}$.

Comentarios:

- La matriz de correlaciones cruzadas contemporánea

$$\rho(0) = D^{-1}\Gamma_y(0)D^{-1}$$

es una matriz simétrica con elementos en su diagonal principal igual a la unidad, esto es $\rho_{ij}(0) = \rho_{ji}(0)$, $-1 \leq \rho_{ij}(0) \leq 1$ y $\rho_{ii}(0) = 1$ para $1 \leq i, j \leq k$. Sus elementos nos dan las correlaciones contemporáneas entre los elementos del vector y_t .

- Los elementos de CCM para el retardo $h > 0$ nos muestran la correlaciones entre los diferentes elementos del vector y_t e y_{t-h} , por lo que $\rho_{ij}(h)$ mide la dependencia lineal de y_{it} en $y_{j,t-h}$ que se observó h periodos antes. En consecuencia, si $\rho_{ij}(h) \neq 0$ y $h > 0$, se dice que la serie y_{jt} va por delante o antecede (*leads*) a la serie y_{it} en el retardo h . De igual forma, $\rho_{ji}(h)$ mide la dependencia lineal de y_{jt} en $y_{i,t-h}$. Si $\rho_{ji}(h) \neq 0$ y $h > 0$, se dice que la serie y_{it} va por delante o antecede a la serie y_{jt} en el retardo h . El elemento de la diagonal principal de $\rho_{ii}(h)$ es el coeficiente de autocorrelación con h -retardos de y_{it} .
- En general, para $h > 0$ $\rho_{ij}(h) \neq \rho_{ji}(h)$ para $i \neq j$ porque los dos coeficientes de correlación miden relaciones lineales diferentes entre las series y_{it} y y_{jt} . Por lo tanto, en general $\Gamma_y(h)$ y $\rho(h)$ no son matrices simétricas.
- Utilizando que $Cov(x, z) = Cov(z, x)$ y el supuesto de estacionariedad débil, se tiene que

$$Cov(y_{it}, y_{j,t-h}) = Cov(y_{j,t-h}, y_{it}) = Cov(y_{jt}, y_{i,t+h}) = Cov(y_{jt}, y_{i,t-(-h)})$$

por lo que $\Gamma_{ij}(h) = \Gamma_{ji}(-h)$. Dado que $\Gamma_{ji}(-h)$ es el elemento (j, i) -ésimo de la matriz $\Gamma_y(-h)$, y la anterior igualdad se satisface para $1 \leq i, j \leq k$, tenemos que

$$\Gamma_y(h) = \Gamma_y'(-h) \quad \text{y} \quad \rho(h) = \rho'(-h)$$

En consecuencia, al contrario de lo que ocurre en el caso univariante, $\rho(h) \neq \rho(-h)$ para $h > 0$.

Dado que $\rho(h) = \rho'(-h)$ en la práctica es suficiente con considerar las matrices de correlaciones cruzadas $\rho(h)$ para $h \geq 0$.

1.2. Dependencia lineal

Las matrices de correlaciones cruzadas, $\{\rho(h) \mid h = 0, 1, \dots\}$ de un vector de series temporales estacionario (al menos en sentido débil), contiene la siguiente información:

- Los elementos de la diagonal principal $\{\rho_{ii}(h) \mid h = 0, 1, \dots\}$ son las funciones de autocorrelación de y_{it} .
- El elemento de fuera de la diagonal $\rho_{ij}(0)$ mide la relación lineal contemporánea entre y_{it} e y_{jt}
- Para $h > 0$, los elementos de fuera de la diagonal $\rho_{ij}(h)$ mide la dependencia lineal de y_{it} en el valor pasado $y_{j,t-h}$

En general, las relaciones lineales que pueden existir entre dos series temporales y_{it} y y_{jt} se pueden resumir como sigue:

1. No existe relación lineal entre y_{it} e y_{jt} si $\rho_{ij}(h) = \rho_{ji}(h) = 0$ para todo $h \geq 0$.
2. y_{it} e y_{jt} están correlacionadas contemporáneamente si $\rho_{ij}(0) \neq 0$.
3. y_{it} e y_{jt} no presentan relaciones lineales retardo-adelanto (*lead-lag*) si $\rho_{ij}(h) = 0$ y $\rho_{ji}(h) = 0$ para todo $h > 0$. En este caso se dice que las dos series están desparejadas (*uncoupled*).
4. Existe una relación unidireccional entre y_{it} e y_{jt} si $\rho_{ij}(h) = 0$ para todo $h > 0$ pero $\rho_{ji}(g) \neq 0$ para algún $g > 0$.
5. Existe una relación de retroalimentación (*feedback*) entre y_{it} e y_{jt} si $\rho_{ij}(h) \neq 0$ para algún $h > 0$ pero $\rho_{ji}(g) \neq 0$ para algún $g > 0$.

Una aproximación más informativa para analizar las relaciones entre series temporales es la de construir un modelo multivariante. Más adelante veremos algunos de estos modelos. De todas formas, analizar las matrices de correlaciones cruzadas es informativo y puede servir como análisis previo.

En la práctica, tendremos que utilizar un estimador de $\Gamma_y(h)$ y de $\rho(h)$. Dados los datos a utilizar $y_t \mid t = 1, \dots, T$, estas matrices pueden estimarse de forma consistente, bajo ciertas condiciones, mediante los siguientes estimadores:

$$\hat{\Gamma}_y(h) = \frac{1}{T} \sum_{t=h+1}^T (y_t - \bar{y})(y_{t-h} - \bar{y})', \quad h \geq 0$$

donde $\bar{y} = (\sum_{t=1}^T y_t) / T$ es el vector de medias muestrales. La matriz de correlaciones cruzadas $\rho(h)$ se estima con

$$\hat{\rho}(h) = \hat{D}^{-1} \hat{\Gamma}_y(h) \hat{D}^{-1}, \quad h \geq 0$$

donde \hat{D} es una matriz diagonal ($k \times k$) con las desviaciones típicas muestrales de los componentes del vector y_t en su diagonal.

Ejemplo 8.1 del capítulo 8 del Tsay

Considerar el fichero¹ de datos **m-ibmspln.txt**:

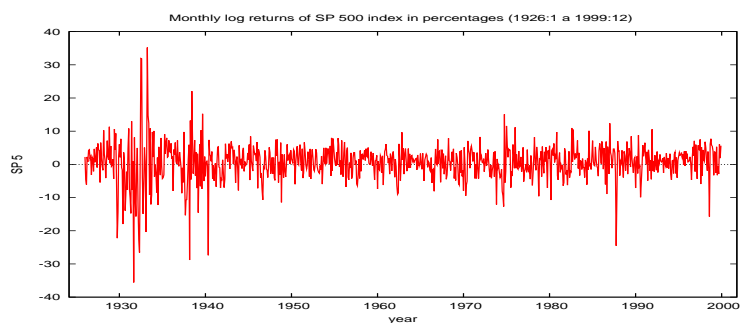
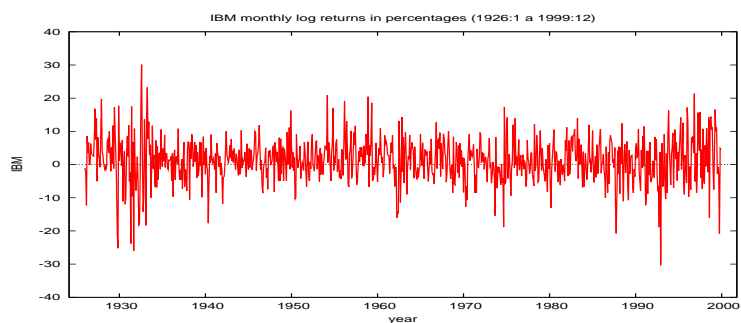
Periodo: 1926:01 - 1999:12 ($T = 888$), frecuencia mensual.

Variables:

IBM: Monthly log returns of IBM stock

SP500: Standard & Poors 500 index.

Vector bi-variante de series temporales $y_t = (IBM_t, SP500_t)$.



¹El fichero IBM-SP.gdt contiene también esos datos en formato Gretl.

Estadísticos principales

de los rendimientos mensuales en logaritmos de IBM y el índice SP500
(en tantos por ciento, Enero 1926 a Diciembre 1999)

Variable	Media	Mediana	Mínimo	Máximo
IBM	1,24023	1,22299	-30,367	30,0971
SP500	0,537164	0,936600	-35,584	35,2219
Variable	Desv. Típ.	C.V.	Asimetría	Exc. de curtosis
IBM	6,72868	5,42535	-0,236855	1,92779
SP500	5,64472	10,5084	-0,522144	8,14155

Matrices de Correlaciones cruzadas

(valor crítico² al 0,05 (a dos colas) = 0,0658)

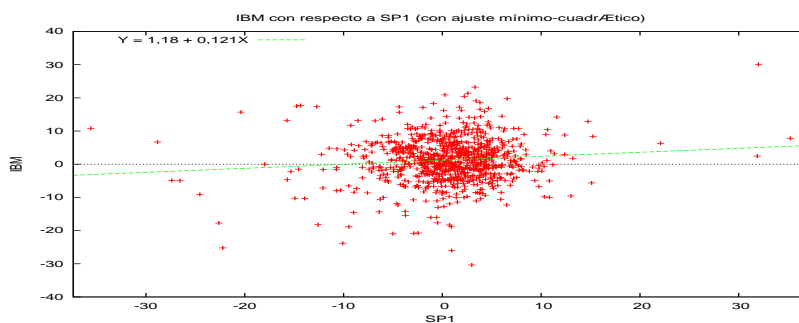
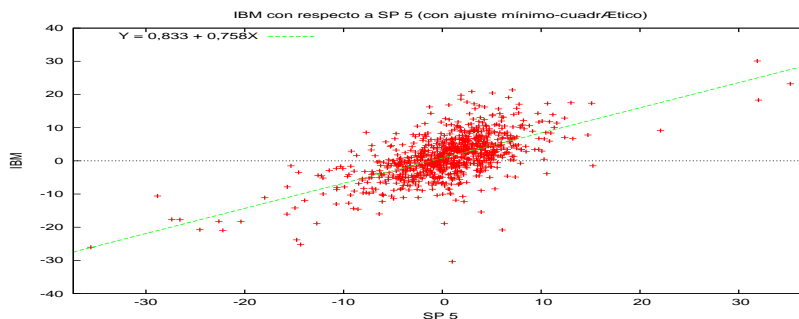
h=0		h=1		
<i>IBM</i>	<i>SP500</i>	<i>IBM</i> - 1	<i>SP500</i> - 1	
1,0000	0,6356	0,0758	0,1012	<i>IBM</i>
	1,0000	0,0445	0,0761	<i>SP500</i>
h=2		h=3		
<i>IBM</i> - 2	<i>SP500</i> - 2	<i>IBM</i> - 3	<i>SP500</i> - 3	
0,0163	-0,0597	-0,0189	-0,0710	<i>IBM</i>
0,0214	-0,0157	-0,0735	-0,1103	<i>SP500</i>
h=4		h=5		
<i>IBM</i> - 4	<i>SP500</i> - 4	<i>IBM</i> - 5	<i>SP500</i> - 5	
-0,0228	-0,0331	0,0044	0,0671	<i>IBM</i>
0,0392	0,0239	0,0024	0,0845	<i>SP500</i>

²Los valores en negrita dentro de la tabla indican una correlación significativamente distinta de cero. El valor crítico corresponde a $2\sqrt{T}$ siendo $T = 888$.

Coefficientes de correlación, usando las observaciones 1926:01 - 1999:12
(se ignoraron los valores perdidos)

valor crítico al 5% (a dos colas) = 0,0658 para T = 888

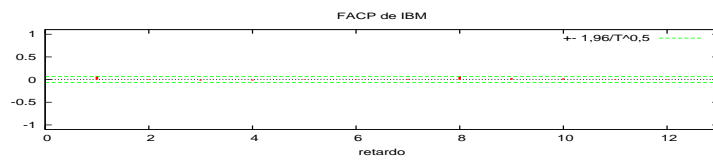
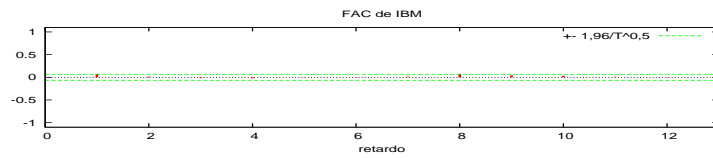
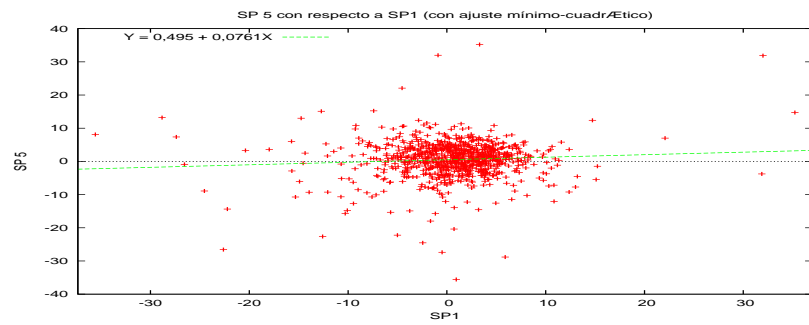
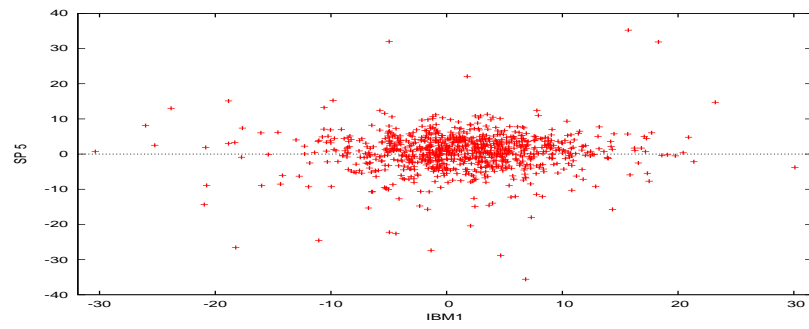
IBM	IBM_1	IBM_2	IBM_3	IBM_4	
1,0000	0,0758	0,0163	-0,0189	-0,0228	IBM
	1,0000	0,0755	0,0182	-0,0186	IBM_1
		1,0000	0,0780	0,0186	IBM_2
			1,0000	0,0762	IBM_3
				1,0000	IBM_4
IBM_5	SP500	SP500_1	SP500_2	SP500_3	
0,0044	0,6356	0,1012	-0,0597	-0,0710	IBM
-0,0226	0,0445	0,6355	0,1011	-0,0603	IBM_1
-0,0184	0,0214	0,0440	0,6355	0,1006	IBM_2
0,0176	-0,0735	0,0249	0,0451	0,6434	IBM_3
0,0760	0,0392	-0,0729	0,0250	0,0458	IBM_4
1,0000	0,0024	0,0395	-0,0729	0,0254	IBM_5
	1,0000	0,0761	-0,0157	-0,1103	SP500
		1,0000	0,0758	-0,0167	SP500_1
			1,0000	0,0756	SP500_2
				1,0000	SP500_3
		SP500_4	SP500_5		
		-0,0331	0,0671		IBM
		-0,0707	-0,0329		IBM_1
		-0,0600	-0,0706		IBM_2
		0,0990	-0,0611		IBM_3
		0,6432	0,0988		IBM_4
		0,0456	0,6432		IBM_5
		0,0239	0,0845		SP500
		-0,1098	0,0242		SP500_1
		-0,0165	-0,1097		SP500_2
		0,0764	-0,0163		SP500_3
		1,0000	0,0762		SP500_4
			1,0000		SP500_5

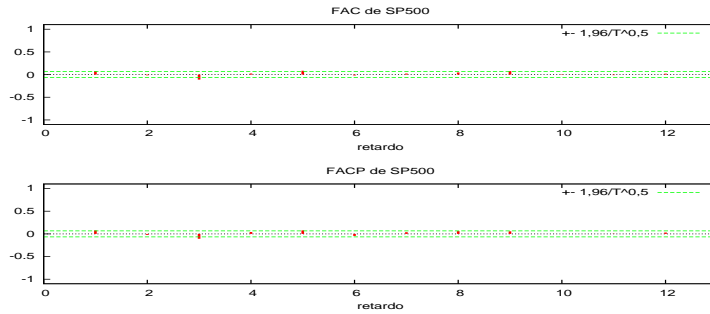


Los gráficos muestran que las dos series están contemporáneamente correlacionadas. De hecho el coeficiente de correlación muestral para $h = 0$ entre las dos series es de 0,64 siendo significativo al 5%. Sin embargo, las correlaciones cruzadas en el retardo $h = 1$ son muy débiles si es que existen.

Las correlaciones cruzadas significativas al 5% aparecen para los retardos 1 y 3.

- SP500 presenta cierta autocorrelación muy marginal en los retardos 1 y 3,
- IBM depende débilmente de los retardos 1 y 3 de SP500, primero positivamente y luego de forma negativa. Esta observación se basa en el elemento (1,2) de las CCM en los retardos $h = 1$ y $h = 3$.





1.3. Contrastes de Portmanteau Multivariantes

El contraste univariante de Ljung-Box $Q(m)$ ha sido generalizado para el caso multivariante por Hosking y Li-McLeod. Para una serie multivariante la hipótesis nula del contraste y la alternativa son:

$$H_0 : \rho(1) = \dots = \rho(m) = 0 \quad \text{vs.} \quad H_a : \rho(i) \neq 0 \quad \text{para algún } i = 1, \dots, m$$

El estadístico se utiliza por tanto para contrastar si no hay ni autocorrelación ni correlaciones cruzadas en el vector de series temporales y_t . El estadístico es:

$$Q_k(m) = T^2 \sum_{h=1}^m \frac{1}{T-h} \text{tr}(\hat{\Gamma}'_y(h) \hat{\Gamma}_y^{-1}(0) \hat{\Gamma}_y(h) \hat{\Gamma}_y^{-1}(0))$$

donde T es el tamaño muestral, k es la dimensión del vector y_t y $\text{tr}(A)$ es la traza de la matriz A , esto es la suma de los elementos de la diagonal principal de A . Bajo la hipótesis nula y ciertas condiciones de regularidad, $Q_k(m)$ sigue una distribución asintótica χ^2 con $k^2 m$ grados de libertad.

Para el ejemplo de *IBM* y *SP500*, se obtienen los siguientes valores del estadístico para diferente elección de m .

$$Q_2(1) = 9,81 > \chi^2(4)_{0,05} = 9,448 \quad Q_2(5) = 47,06 > \chi^2(20)_{0,05} = 31,41$$

$$Q_2(10) = 71,65 > \chi^2(40)_{0,05} = 55,758$$

Así, el contraste de Portmanteau multivariante confirma la existencia de dependencia lineal en el tiempo en el vector bivalente de series de rendimientos *IBM* y *SP500*.

Dado que se rechaza la hipótesis nula, entonces podemos estar interesados en proponer un modelo multivariante que permita analizar las relaciones dinámicas entre estas variables, y que nos pueda servir por ejemplo para predecir. Uno de estos modelos es el Vector Autorregresivo (VAR) que seguidamente pasamos a estudiar.

Capítulo 2

Procesos Vectoriales Autorregresivos Estables²

2.1. Condición de estabilidad

VAR(p)

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \quad t = 0, \pm 1, \pm 2, \dots$$

- $y_t = (y_{1t}, \dots, y_{kt})'$ un vector ($k \times 1$) de variables aleatorias,
- A_i matrices fijas ($k \times k$) de coeficientes,

$$A_i = \begin{bmatrix} a_{11,i} & a_{12,i} & \dots & a_{1k,i} \\ a_{21,i} & a_{22,i} & \dots & a_{2k,i} \\ \vdots & \ddots & \vdots & \vdots \\ a_{k1,i} & a_{k2,i} & \dots & a_{kk,i} \end{bmatrix}$$

- $\nu = (\nu_1, \dots, \nu_k)'$ es un vector de términos de intercepto permitiendo la posibilidad de que $E(y_t)$ sea distinta de cero.
- Finalmente, $u_t = (u_{1t}, \dots, u_{kt})'$ es un vector ($k \times 1$) de innovaciones tal que $E(u_t) = 0$, $E(u_t u_t') = \Sigma_u$, $E(u_t u_s') = 0$ para $t \neq s$ donde

$$\Sigma_u = \begin{bmatrix} E(u_{1t}^2) & E(u_{1t}u_{2t}) & \dots & E(u_{1t}u_{kt}) \\ E(u_{2t}u_{1t}) & E(u_{2t}^2) & \dots & E(u_{2t}u_{kt}) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_{kt}u_{1t}) & E(u_{kt}u_{2t}) & \dots & E(u_{kt}^2) \end{bmatrix}$$

se supone finita y no singular.

²Se basa en los capítulos 2,3 y 4 del libro de Lütkepohl.

VAR(1) estable

$$y_t = \nu + A_1 y_{t-1} + u_t \quad t = 0, \pm 1, \pm 2, \dots$$

Se dice que un VAR(1) es estable si todos los valores propios de A_1 tienen módulo menor que la unidad. Equivalentemente, el VAR(1) es estable si

$$\det(I_k - A_1 z) \neq 0 \text{ para } |z| \leq 1$$

Esto es, todas las raíces del polinomio $\det(I_k - A_1 z)$ caen fuera del círculo unidad.

Ejemplo 1:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} + \begin{bmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

$$\det(I_k - A_1 z) = \left| \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{bmatrix} z \right| =$$

$$\left| \begin{array}{ccc} 1 - 0,5z & 0 & 0 \\ -0,1z & 1 - 0,1z & -0,3z \\ 0 & -0,2z & 1 - 0,3z \end{array} \right| = (1 - 0,5z)(1 - 0,4z - 0,03z^2)$$

Raíces del polinomio:

- De $(1 - 0,5z_1) = 0$ se obtiene $z_1 = 1/(0,5) = 2$
- De $(1 - 0,4z - 0,03z^2) = 0$ se obtienen otras dos raíces:
 $z = \frac{0,4 \pm \sqrt{0,16 + 0,12}}{-0,06} \Rightarrow z_2 = -15,4858 \quad z_3 = 2,1525$

Todas ellas son reales y mayores que uno en valor absoluto. Por lo tanto el proceso VAR(1) es estable.

Ejercicio: ¿Qué hubiera ocurrido si $a_{11} = 1$? ¿Sería el proceso estable en ese caso?

Companion form:

¿Cómo representar un proceso VAR(p) en forma de VAR(1)?

$$\underbrace{Y_t}_{(kpx1)} = \eta + \underbrace{A}_{(kpxkp)} Y_{t-1} + U_t$$

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_k & 0 & \dots & 0 & 0 \\ 0 & I_k & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_k & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Y_t es estable si $\det(I_{kp} - Az) \neq 0$ para $|z| \leq 1$

Dado que:

$$\det(I_{kp} - Az) = \det(I_k - A_1z - \dots - A_pz^p)$$

Entonces se dice que y_t es un proceso VAR(p) estable si

$$\det(I_k - A_1z - \dots - A_pz^p) \neq 0 \text{ para } |z| \leq 1$$

Ejemplo 2:

$$\underbrace{\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}}_{y_t} = \underbrace{\begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}}_{\nu} + \underbrace{\begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix}}_{A_1} \underbrace{\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix}}_{y_{t-1}} + \underbrace{\begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix}}_{A_2} \underbrace{\begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix}}_{y_{t-2}} + \underbrace{\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}}_{u_t}$$

Companion Form:

$$\underbrace{\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{1,t-1} \\ y_{2,t-1} \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \nu_1 \\ \nu_2 \\ 0 \\ 0 \end{bmatrix}}_{\eta} + \underbrace{\begin{bmatrix} 0,5 & 0,1 & 0 & 0 \\ 0,4 & 0,5 & 0,25 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_A \underbrace{\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{1,t-2} \\ y_{2,t-2} \end{bmatrix}}_{Y_{t-1}} + \underbrace{\begin{bmatrix} u_{1t} \\ u_{2t} \\ 0 \\ 0 \end{bmatrix}}_{U_t}$$

$$\begin{aligned} \det(I_4 - Az) &= \begin{vmatrix} 1 - 0,5z & -0,1z & 0 & 0 \\ -0,4z & 1 - 0,5z & -0,25z & 0 \\ -z & 0 & 1 & 0 \\ 0 & -z & 0 & 1 \end{vmatrix} = \begin{vmatrix} (I_2 - A_1z) & -A_2z \\ -zI_2 & I_2 \end{vmatrix} = \\ &= |I_2| |(I_2 - A_1z) - (-A_2z)I_2^{-1}(-zI_2)| = |I_2 - A_1z - A_2z^2| \end{aligned}$$

$$|I_2 - A_1z - A_2z^2| = \left| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} z - \begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} z^2 \right| =$$

$$\begin{vmatrix} 1 - 0,5z & -0,1z \\ -0,4z - 0,25z^2 & 1 - 0,5z \end{vmatrix} = 1 - z + 0,21z^2 - 0,025z^3$$

Las raíces de este polinomio son una real y dos imaginarias,

$$z_1 = 1,3 \quad z_2 = 3,55 + 4,26i \text{ y } z_3 = 3,55 - 4,26i$$

Todas ellas satisfacen ¹ $|z| > 1$. El proceso es estable.

2.2. Representación de Medias Móviles

Haciendo sustituciones repetidas en el proceso $Y_t = \eta + AY_{t-1} + U_t$ se obtiene hasta la j -ésima:

$$Y_t = (I_{kp} + A + \dots + A^j)\eta + A^{j+1}Y_{t-j-1} + \sum_{i=0}^j A^i U_{t-i}$$

Bajo la condición de estabilidad:

Esto se debe a que, dada la condición de estabilidad,

- La secuencia $A^i, i = 0, 1, \dots$, es absolutamente sumable. Por lo tanto $\sum_{i=0}^{\infty} A^i U_{t-i}$ existe en media cuadrática.
- Cuando $j \rightarrow \infty$,

$$(I_{kp} + A + \dots + A^j)\eta \rightarrow (I_{kp} - A)^{-1}\eta$$

- Cuando $j \rightarrow \infty$ entonces A^{j+1} converge a cero rápidamente por lo que podemos ignorar el término $A^{j+1}Y_{t-j-1}$ en

Representación MA para Y_t

$$Y_t = \zeta + \sum_{i=0}^{\infty} A^i U_{t-i}$$

donde Y_t viene expresado en términos de las innovaciones pasadas y presente U_t y del término medio, $\zeta = (I_{kp} - A)^{-1}\eta$

¹En el caso de las raíces imaginarias $|z_2| = |z_3| = \sqrt{(3,55)^2 + (4,26)^2} = 5,545$.

Representación MA para y_t

Utilizando la matriz de selección ($k \times kp$), $J = \begin{bmatrix} I_k & 0 & \dots & 0 \end{bmatrix}$,

$$y_t = JY_t = J\zeta + \sum_{i=0}^{\infty} JA^i J' JU_{t-i}$$

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}$$

donde $\mu = (I_k - A_1 - A_2 - \dots - A_p)^{-1}\nu$, $\Phi_i = JA^i J'$, $U_t = J'JU_t$ y $u_t = JU_t$. Dado que A_i son absolutamente sumables también lo son Φ_i .

Obtención de Φ_i $i = 0, 1, 2, \dots$

Una forma más directa de computar las matrices de coeficientes de la representación de MA para y_t es la siguiente. Utilizando el operador de retardos L tal que $L^j y_t = y_{t-j}$ podemos escribir:

$$y_t = \nu + (A_1 L + \dots + A_p L^p) y_t + u_t$$

$$A(L) y_t = \nu + u_t$$

donde $A(L) = (I_k - A_1 L - \dots - A_p L^p)$. Sea $\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i$ tal que $\Phi(L)A(L) = I_k$. Entonces

$$\Phi(L)A(L)y_t = \Phi(L)\nu + \Phi(L)u_t$$

$$y_t = (\sum_{i=0}^{\infty} \Phi_i)\nu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}$$

Si $\det(A(z)) \neq 0$ para $|z| \leq 1$, se satisface la condición de estabilidad, entonces $A(L)$ es invertible, existe $A(L)^{-1} = \Phi(L)$, sus matrices de coeficientes son absolutamente sumables y por lo tanto el proceso $\Phi(L)u_t = A(L)^{-1}u_t$ está bien definido. Por otro lado,

$$\Phi(1) = \sum_{i=0}^{\infty} \Phi_i = A(1)^{-1} = (I_k - A_1 - \dots - A_p)^{-1}$$

Las matrices de coeficientes Φ_i se pueden obtener de

$$I_k = (\Phi_0 + \Phi_1 L + \Phi_2 L^2 + \dots)(I_k - A_1 L - \dots - A_p L^p)$$

Operando y agrupando términos en función del operador de retardos se tiene que:

$$\Phi_0 = I_k \quad \Phi_i = \sum_{j=1}^i \Phi_{i-j} A_j, i = 1, 2, \dots \text{ con } A_j = 0 \text{ para } j > p$$

Por lo que de estas ecuaciones se pueden ir computando recursivamente Φ_0, Φ_1, \dots

Ejemplos:

Para un VAR(1) estable:

$$\Phi_0 = I_k \quad \Phi_1 = A_1 \quad \Phi_2 = A_1^2, \dots, \Phi_i = A_1^i, \dots$$

Ejemplo 1:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} + \begin{bmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

$$\Phi_1 = \begin{bmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0,25 & 0 & 0 \\ 0,06 & 0,07 & 0,12 \\ 0,02 & 0,08 & 0,15 \end{bmatrix},$$

$$\Phi_3 = \begin{bmatrix} 0,125 & 0 & 0 \\ 0,037 & 0,031 & 0,057 \\ 0,018 & 0,038 & 0,069 \end{bmatrix}, \text{ etc.}$$

Para un VAR(2) estable:

$$\Phi_0 = I_k \quad \Phi_1 = A_1 \quad \Phi_2 = \Phi_1 A_1 + A_2, \dots, \Phi_i = \Phi_{i-1} A_1 + \Phi_{i-2} A_2, \dots$$

Ejemplo 2:

$$\underbrace{\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}}_{y_t} = \underbrace{\begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}}_{\nu} + \underbrace{\begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix}}_{A_1} \underbrace{\begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix}}_{y_{t-1}} + \underbrace{\begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix}}_{A_2} \underbrace{\begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix}}_{y_{t-2}} + \underbrace{\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}}_{u_t}$$

$$\Phi_1 = \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0,29 & 0,1 \\ 0,65 & 0,29 \end{bmatrix}, \quad \Phi_3 = \begin{bmatrix} 0,21 & 0,079 \\ 0,566 & 0,21 \end{bmatrix},$$

etc.

2.3. Procesos Estacionarios

Proposición. *Estabilidad \Rightarrow Estacionariedad*

Un proceso y_t , $t = 0, \pm 1, \pm 2, \dots$, VAR(p) estable es estacionario².
Utilizando la representación de medias móviles:

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}$$

$$E(y_t) = \mu = (I_k - A_1 - \dots - A_p)^{-1} \nu$$

$$\begin{aligned} \Gamma_y(h) &= E[(y_t - \mu)(y_{t-h} - \mu)'] \\ &= E\left(\sum_{i=0}^{h-1} \Phi_i u_{t-i} + \sum_{i=0}^{\infty} \Phi_{h+i} u_{t-h-i}\right) \left(\sum_{i=0}^{\infty} \Phi_i u_{t-h-i}\right)' \\ &= \sum_{i=0}^{\infty} \Phi_{h+i} \Sigma_u \Phi_i' \end{aligned}$$

Para un VAR(1):

- $\Gamma_y(0) = \sum_{i=0}^{\infty} \Phi_i \Sigma_u \Phi_i' = \sum_{i=0}^{\infty} A_1^i \Sigma_u (A_1^i)'$
- $\Gamma_y(h) = A_1 \Gamma_y(h-1)$, $h > 0$
- $\Gamma_y(h) = A_1^h \Gamma_0$, para $h > 0$

Teorema de descomposición de Wold.

Todo proceso estacionario x_t se puede representar como la suma de dos procesos incorrelacionados z_t e y_t , donde z_t es un proceso determinista perfectamente predecible de su propio pasado e y_t es un proceso que tiene una representación MA,

$$y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i}$$

donde $\Phi_0 = I_k$, u_t es un proceso ruido blanco y la suma infinita está definida en media cuadrática aunque las Φ_i no son necesariamente absolutamente sumables.

²Dado que *estabilidad* implica *estacionariedad*, la condición de estabilidad se conoce también como condición de estacionariedad. Un proceso no estable no es necesariamente no estacionario

Si además las Φ_i son absolutamente sumables y existe $A(L)$ tal que $A(L)\Phi(L) = I_k$ con coeficientes absolutamente sumables, entonces $\Phi(L)$ es invertible y

$$y_t = \sum_{i=1}^{\infty} A_i y_{t-i} + u_t \quad VAR(\infty)$$

donde $A(z) = I_k - \sum_{i=1}^{\infty} A_i z^i = (\sum_{i=0}^{\infty} \Phi_i z^i)^{-1}$ para $|z| \leq 1$. Dado que la sumabilidad absoluta de las A_i implica que las matrices de coeficientes del VAR convergen a cero rápidamente esto supone que bajo condiciones bastante generales *todo proceso estacionario sin componentes deterministas puede aproximarse bien por un proceso VAR de orden finito*. La hipótesis crucial es la de estacionariedad.

2.4. Estimación y Contrastes de Restricciones lineales

Supongamos que y_t es un vector de variables aleatorias ($k \times 1$) que sigue un VAR(p) estable,

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$$

donde $u_t \sim i.i.d(0, \Sigma_u)$. Supongamos que disponemos de T observaciones y_1, \dots, y_T y p valores iniciales y_{-p+1}, \dots, y_0 . Consideremos la siguiente notación,

$$\begin{aligned} Y &= \begin{pmatrix} y_1 & \cdots & y_T \end{pmatrix} && (k \times T), \\ B &= \begin{pmatrix} \nu & A_1 & \cdots & A_p \end{pmatrix} && (k \times (kp + 1)), \\ Z_t &= \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} && ((kp + 1) \times 1), \\ Z &= \begin{pmatrix} Z_0 & \cdots & Z_{T-1} \end{pmatrix} && ((kp + 1) \times T), \\ U &= \begin{pmatrix} u_1 & \cdots & u_T \end{pmatrix} && (k \times T) \end{aligned}$$

$$\begin{aligned} y_{(j)} &= \begin{pmatrix} y_{j1} & \cdots & y_{jT} \end{pmatrix}' && (T \times 1) \\ \mathbf{y} &= \text{vec}(Y') = \begin{bmatrix} y_{(1)} \\ \vdots \\ y_{(k)} \end{bmatrix} && (kT \times 1) \\ b_{(j)} &= \begin{pmatrix} \nu_j & a_{j1,1} & \cdots & a_{jk,1} & \cdots & a_{j1,p} & \cdots & a_{jk,p} \end{pmatrix}' && ((kp + 1) \times 1) \\ \mathbf{b} &= \text{vec}(B') = \begin{bmatrix} b_{(1)} \\ \vdots \\ b_{(k)} \end{bmatrix} && ((k^2p + k) \times 1) \\ u_{(j)} &= \begin{pmatrix} u_{j1} & \cdots & u_{jT} \end{pmatrix}' && (T \times 1) \\ \mathbf{u} &= \text{vec}(U') = \begin{bmatrix} u_{(1)} \\ \vdots \\ u_{(k)} \end{bmatrix} && (kT \times 1) \end{aligned}$$

Usando esta notación, para $t = 1, \dots, T$ podemos escribir el modelo VAR(p) de la siguiente forma

$$Y = BZ + U \quad \Leftrightarrow \quad \text{vec}(Y') = \underbrace{\text{vec}(Z'B')}_{(I_k \otimes Z')(\text{vec}(B'))} + \text{vec}(U')$$

o

$$\mathbf{y} = (I_k \otimes Z')\mathbf{b} + \mathbf{u} \quad \mathbf{u} \sim (0, (\Sigma_u \otimes I_T))$$

Ejemplo: VAR(1) bivalente, con la condición inicial $\mathbf{y}_0 = (y_{1,0}, y_{2,0})'$

$$y_{1t} = \nu_1 + a_{11,1}y_{1,t-1} + a_{12,1}y_{2,t-1} + u_{1t} \quad t = 1, \dots, T$$

$$y_{2t} = \nu_2 + a_{21,1}y_{1,t-1} + a_{22,1}y_{2,t-1} + u_{2t} \quad t = 1, \dots, T$$

o equivalentemente

$$\underbrace{\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \end{bmatrix}}_Y = \underbrace{\begin{bmatrix} \nu_1 & a_{11,1} & a_{12,1} \\ \nu_2 & a_{21,1} & a_{22,1} \end{bmatrix}}_B \underbrace{\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ y_{1,0} & y_{1,1} & y_{1,2} & \cdots & y_{1,(T-1)} \\ y_{2,0} & y_{2,1} & y_{2,2} & \cdots & y_{2,(T-1)} \end{bmatrix}}_Z + \underbrace{\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1T} \\ u_{21} & u_{22} & \cdots & u_{2T} \end{bmatrix}}_U$$

$$\underbrace{\begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \end{bmatrix}}_{\text{vec}(Y')} = \underbrace{\begin{bmatrix} 1 & y_{1,0} & y_{2,0} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_{1,T-1} & y_{2,T-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & y_{1,0} & y_{2,0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & y_{1,T-1} & y_{2,T-1} \end{bmatrix}}_{\text{vec}(Z'B')} \underbrace{\begin{bmatrix} \nu_1 \\ a_{11,1} \\ a_{12,1} \\ \nu_2 \\ a_{21,1} \\ a_{22,1} \end{bmatrix}}_{\text{vec}(U')} + \underbrace{\begin{bmatrix} u_{11} \\ \vdots \\ u_{1T} \\ u_{21} \\ \vdots \\ u_{2T} \end{bmatrix}}_{\text{vec}(U')}$$

$$\underbrace{\begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} Z' & 0 \\ 0 & Z' \end{bmatrix}}_{(I_k \otimes Z')} \underbrace{\begin{bmatrix} b^{(1)} \\ b^{(2)} \end{bmatrix}}_{\mathbf{b}} + \underbrace{\begin{bmatrix} u^{(1)} \\ u^{(2)} \end{bmatrix}}_{\mathbf{u}}$$

- Dado que es un **sistema de ecuaciones aparentemente no relacionadas (SURE) con idénticos regresores (Z')** en cada ecuación, el estimador de \mathbf{b} por **Mínimos Cuadrados Generalizados** en el sistema **coincide** con el de **Mínimos Cuadrados Ordinarios** en el sistema, que a su vez coincide con el de Mínimos Cuadrados Ordinarios Ecuación por Ecuación.

Si denotamos $X = (I_k \otimes Z')$ y $\Sigma = (\Sigma_u \otimes I_T)$, haciendo uso de las propiedades del producto Kronecker tenemos

$$\begin{aligned}
\hat{\mathbf{b}} &= (X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}\mathbf{y}) = \\
&= ((I_k \otimes Z)(\Sigma_u \otimes I_T)^{-1}(I_k \otimes Z'))^{-1}((I_k \otimes Z)(\Sigma_u \otimes I_T)^{-1}\mathbf{y}) = \\
&= (\Sigma_u^{-1} \otimes ZZ')^{-1}(\Sigma_u^{-1} \otimes Z)\mathbf{y} = (I_k \otimes (ZZ')^{-1}Z)\mathbf{y} = \\
&= \begin{bmatrix} (ZZ')^{-1}Z & 0 & \cdots & 0 \\ 0 & (ZZ')^{-1}Z & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & (ZZ')^{-1}Z \end{bmatrix} \begin{bmatrix} y_{(1)} \\ \vdots \\ y_{(k)} \end{bmatrix} \\
&= \begin{bmatrix} (ZZ')^{-1}Zy_{(1)} \\ \vdots \\ (ZZ')^{-1}Zy_{(k)} \end{bmatrix} = (X'X)^{-1}X'\mathbf{y}
\end{aligned}$$

- Bajo ciertas condiciones, (ver Lütkepohl(1991), pp. 66) el **estimador MCO es consistente y asintóticamente Normal**.

$$\sqrt{T}(\hat{\mathbf{b}} - \mathbf{b}) \xrightarrow{d} N(0, (\Sigma_u \otimes \Gamma^{-1}))$$

donde $\Gamma = plim(1/T)ZZ'$. Para poder realizar inferencia se necesitan **estimadores consistentes de Γ y Σ_u** :

$$\hat{\Gamma} = (1/T)ZZ'$$

$$\tilde{\Sigma}_u = (1/T) \sum_{t=1}^T \hat{u}_t \hat{u}_t' \quad \text{o bien} \quad \hat{\Sigma}_u = (1/(T - kp - 1)) \sum_{t=1}^T \hat{u}_t \hat{u}_t'$$

donde

$$\begin{aligned}\hat{u}'_t &= [\hat{u}_{1t} \ \hat{u}_{2t} \ \cdots \ \hat{u}_{kt}] \\ \hat{u}_{jt} &= y_{jt} - Z'_{t-1} \hat{b}_{(j)} \\ Z'_{t-1} &= [1 \ y'_{t-1} \ y'_{t-2} \ \cdots \ y'_{t-p}]\end{aligned}$$

Estos últimos son asintóticamente equivalentes, aunque argumentos conservadores tomarían como más adecuado el segundo, que corrige por grados de libertad.

- Si estamos interesados en **contrastar hipótesis sobre los coeficientes de solamente una ecuación del VAR**, podemos utilizar los estadísticos t y F usuales de la regresión MCO en esa ecuación ya que son asintóticamente válidos. Dado que

$$\sqrt{T}(\hat{b}_{(j)} - b_{(j)}) \xrightarrow{d} N(0, (\sigma_j^2 \Gamma^{-1}))$$

donde $\sigma_j^2 = E(u_{jt}^2)$ es la varianza de la innovación de la j-ésima ecuación del VAR. Podemos estimarla consistentemente con el estimador usual $\hat{\sigma}_j^2 = (1/(T - kp - 1)) \sum_{t=1}^T \hat{u}_{jt}^2$.

Por lo tanto, las estimaciones de las varianzas y covarianzas de los coeficientes de la regresión MCO en la ecuación j-ésima vendrían dados por el estimador usual

$$\hat{\sigma}_j^2 (ZZ')^{-1} = \hat{\sigma}_j^2 \left[\sum_{t=1}^T Z_{t-1} Z'_{t-1} \right]^{-1}$$

- **Hipótesis más generales** de la forma $H_0 : R\mathbf{b} = r$ que impliquen coeficientes de distintas ecuaciones del VAR se pueden contrastar utilizando el estadístico de Wald,

$$(R\hat{\mathbf{b}} - r)' \left\{ R \left[\hat{\Sigma}_u \otimes (ZZ')^{-1} \right] R' \right\}^{-1} (R\hat{\mathbf{b}} - r)$$

que bajo la hipótesis nula se distribuye asintóticamente $\chi^2(m)$ donde m es el número de restricciones bajo la hipótesis nula o el número de filas de la matriz R . Tanto $\hat{\mathbf{b}}$ como $\hat{\Sigma}_u$ son los estimadores correspondientes al modelo sin restringir.

- Bajo la hipótesis de que el proceso VAR(p) es gaussiano (Normalidad de u_t) el estimador por Máxima Verosimilitud (MV) de los coeficientes del sistema coincide con el de MCO. En este sentido, si las hipótesis son ciertas (correcta elección de p, u_t ruido blanco gaussiano,...) es consistente y eficiente asintóticamente.

2.5. Predicción

Predicción es uno de los principales objetivos del análisis multivariante de series temporales. Primeramente veremos predicción en el marco de procesos VAR estables suponiendo conocidos los valores de los parámetros del proceso y posteriormente cuando previamente han sido estimados.

Proceso generador de los datos $y_t = (y_{1t}, \dots, y_{kt})' \rightarrow \text{VAR}(p)$
 Conjunto de información: $\Omega_t = \{y_s | s \leq t\}$ donde $y_s = (y_{1s}, \dots, y_{ks})'$
 Origen de predicción $\rightarrow t$
 Horizonte de predicción $\rightarrow h$
 Predictor h periodos hacia adelante $\rightarrow \bar{y}_t(h)$
 Función de pérdida: Error cuadrático medio de la predicción,

$$MSE[\bar{y}_t(h)] = E(y_{t+h} - \bar{y}_t(h))(y_{t+h} - \bar{y}_t(h))'$$

En el contexto de modelos VAR, los predictores más utilizados son aquellos que minimizan el MSE. Algunos autores como Granger y Newbold demuestran que los predictores que minimizan el MSE también minimizan otras funciones de pérdida y para muchas de ellas los predictores óptimos son simples funciones de los predictores que minimizan el MSE. Además, la varianza de un predictor insesgado es el MSE lo cual es importante a la hora de construir intervalos de predicción.

2.5.1. Predicción por punto

El predictor h periodos hacia adelante que minimiza el Error cuadrático medio de la predicción es el valor esperado condicionado al conjunto de información en el origen de predicción,

$$E_t(y_{t+h}) = E(y_{t+h} | \Omega_t) = E(y_{t+h} | \{y_s | s \leq t\})$$

Este predictor minimiza el MSE de cada componente de y_t .

- Si u_t es un ruido blanco *independiente* de u_s para todo $t \neq s$ (por lo que $E_t(u_{t+h}) = 0$ para $h > 0$) el predictor óptimo h periodos hacia adelante de un proceso y_t que sigue un VAR(p) es,

$$E_t(y_{t+h}) = \nu + A_1 E_t(y_{t+h-1}) + \dots + A_p E_t(y_{t+h-p})$$

Propiedades:

1. Es un predictor insesgado, $E[y_{t+h} - E_t(y_{t+h})] = 0$.
2. Si u_t es un ruido blanco independientemente distribuido, $MSE[E_t(y_{t+h})] = MSE[E_t(y_{t+h})|y_t, y_{t-1}, \dots]$.

Comenzando en $h = 1$ podemos ir obteniendo recursivamente las predicciones,

$$\begin{aligned} E_t(y_{t+1}) &= \nu + A_1 y_t + \dots + A_p y_{t-p} \\ E_t(y_{t+2}) &= \nu + A_1 E_t(y_{t+1}) + A_2 y_t + \dots + A_p y_{t-p+2} \\ &\vdots \end{aligned}$$

En general obtenemos **para un VAR(1)**,

$$E_t(y_{t+h}) = (I_k + A_1 + \dots + A_1^{h-1})\nu + A_1^h y_t$$

Ejemplo 1:

VAR(1) tridimensional donde $y_t = (-6, 3, 5)'$ y $\nu = (0, 2, 1)'$

$$E_t(y_{t+1}) = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{bmatrix} \begin{bmatrix} -6 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} -3,0 \\ 3,2 \\ 3,1 \end{bmatrix}$$

$$E(y_{t+2}) = (I_3 + A_1)\nu + A_1^2 y_t = \begin{bmatrix} -1,50 \\ 2,95 \\ 2,57 \end{bmatrix} \text{ etc.}$$

Ejemplo 2:

VAR(2) bidimensional donde $y_t = (0,06; 0,03)'$ $y_{t-1} = (0,055; 0,03)'$ y $\nu = (0,02; 0,03)'$

$$\begin{aligned} E_t(y_{t+1}) &= \nu + A_1 y_t + A_2 y_{t-1} = \begin{bmatrix} 0,02 \\ 0,03 \end{bmatrix} + \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} \begin{bmatrix} 0,06 \\ 0,03 \end{bmatrix} + \\ &\begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} \begin{bmatrix} 0,055 \\ 0,03 \end{bmatrix} = \begin{bmatrix} 0,053 \\ 0,08275 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} E(y_{t+2}) &= \nu + A_1 E_t(y_{t+1}) + A_2 y_t = \begin{bmatrix} 0,02 \\ 0,03 \end{bmatrix} + \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} \begin{bmatrix} 0,053 \\ 0,08275 \end{bmatrix} + \\ &\begin{bmatrix} 0 & 0 \\ 0,25 & 0 \end{bmatrix} \begin{bmatrix} 0,06 \\ 0,03 \end{bmatrix} = \begin{bmatrix} 0,0548 \\ 0,1076 \end{bmatrix} \end{aligned}$$

etc.

- Si u_t no es un ruido blanco independientemente distribuido, normalmente se requieren supuestos adicionales para encontrar el predictor óptimo de un proceso VAR en términos de la esperanza condicionada. Sin estos supuestos se puede considerar el objetivo menos ambicioso de encontrar el predictor que minimiza el MSE dentro de la clase de predictores que son función lineal de y_t, y_{t-1}, \dots

Proceso VAR(1) de media cero,

$$y_t = A_1 y_{t-1} + u_t$$

Sustituyendo recursivamente tenemos que

$$y_{t+h} = A_1^h y_t + \sum_{i=0}^{h-1} A_1^i u_{t+h-i}$$

Así, dado un predictor lineal

$$y_t(h) = B_0 y_t + B_1 y_{t-1} + \dots$$

donde las B_i son matrices de coeficientes ($k \times k$), obtenemos el **error de predicción**,

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} A_1^i u_{t+h-i} + (A_1^h - B_0) y_t - \sum_{i=1}^{\infty} B_i y_{t-i}$$

Dado que u_{t+j} , para $j > 0$, está incorrelacionado con y_{t-i} , para $i \geq 0$,

$$\begin{aligned} MSE[y_t(h)] &= E \left(\sum_{i=0}^{h-1} A_1^i u_{t+h-i} \right) \left(\sum_{i=0}^{h-1} A_1^i u_{t+h-i} \right)' \\ &\quad E \left[(A_1^h - B_0) y_t - \sum_{i=1}^{\infty} B_i y_{t-i} \right] \\ &\quad \times \left[(A_1^h - B_0) y_t - \sum_{i=1}^{\infty} B_i y_{t-i} \right]' \end{aligned}$$

El MSE mínimo se obtiene para $B_0 = A_1^h$ y $B_i = 0$ para $i > 0$.

Predictor lineal óptimo:

$$y_t(h) = A_1^h y_t = A_1 y_t(h-1)$$

El **error de predicción** es en este caso

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} A_1^i u_{t+h-i} + (A_1^h - B_0) y_t - \sum_{i=1}^{\infty} B_i y_{t-i} = \sum_{i=0}^{h-1} A_1^i u_{t+h-i}$$

y el **MSE o matriz de varianzas y covarianzas del error de predicción** es,

$$\begin{aligned} MSE[y_t(h)] &= E \left(\sum_{i=0}^{h-1} A_1^i u_{t+h-i} \right) \left(\sum_{i=0}^{h-1} A_1^i u_{t+h-i} \right)' \\ &= \sum_{i=0}^{h-1} A_1^i \Sigma_u (A_1^i)' = MSE[y_t(h-1)] + A_1^{h-1} \Sigma_u (A_1^{h-1})' \end{aligned}$$

Para el caso general de un **VAR(p) con media cero**

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t$$

que en forma de VAR(1) se puede escribir como

$$Y_t = AY_{t-1} + U_t$$

Predictor lineal óptimo,

$$Y_t(h) = A^h Y_t = AY_t(h-1)$$

donde

$$Y_t(h) = \begin{bmatrix} y_t(h) \\ y_t(h-1) \\ \vdots \\ y_t(h-p+1) \end{bmatrix}$$

e $y_t(j) = y_{t+j}$ para $j \leq 0$.

Utilizando la matriz de selección ($k \times kp$), $J = [I_k 0 \dots 0]$, podemos obtener el **predictor lineal óptimo h periodos hacia adelante del proceso y_t en el origen t** como

$$\begin{aligned} y_t(h) &= JAY_t(h-1) = [A_1 \dots A_p]Y_t(h-1) \\ &= A_1 y_t(h-1) + \dots + A_p y_t(h-p) \end{aligned}$$

Podemos utilizar esta fórmula para ir computando recursivamente las predicciones³. Notar que si u_t es un ruido blanco independientemente distribuido entonces $y_t(h) = E_t(y_{t+h})$.

³Si el proceso y_t tiene media distinta de cero, esto es

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t$$

definimos $x_t = y_t - \mu$ donde $\mu = E(y_t) = (I - A_1 - \dots - A_p)^{-1} \nu$. El proceso x_t tiene media cero y por lo tanto el predictor óptimo h periodos hacia adelante es

$$x_t(h) = A_1 x_t(h-1) + \dots + A_p x_t(h-p)$$

Sumando a ambas ecuaciones μ obtenemos el predictor lineal óptimo para y_t

$$\begin{aligned} y_t(h) &= x_t(h) + \mu = \mu + A_1(y_t(h-1) - \mu) + \dots + A_p(y_t(h-p) - \mu) \\ &= \nu + A_1 y_t(h-1) + \dots + A_p y_t(h-p) \end{aligned}$$

Para un proceso de media cero, utilizando

$$Y_{t+h} = A^h Y_t + \sum_{i=0}^{h-1} A^i U_{t+h-i}$$

obtenemos el error de predicción

$$\begin{aligned} y_{t+h} - y_t(h) &= J[Y_{t+h} - Y_t(h)] = J \left[\sum_{i=0}^{h-1} A^i U_{t+h-i} \right] \\ &= \sum_{i=0}^{h-1} J A^i J' J U_{t+h-i} = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} \end{aligned}$$

donde las Φ_i son las matrices de coeficientes de la representación MA. El error de predicción no cambia si y_t tiene media distinta de cero ya que el término medio se cancela.

De esta representación del error de predicción podemos obtener que **el predictor $y_t(h)$ puede expresarse en términos de la representación MA:** como $y_{t+h} = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t+h-i}$ tenemos

$$\begin{aligned} y_t(h) &= y_{t+h} - \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} \\ &= \mu + \sum_{i=0}^{\infty} \Phi_i u_{t+h-i} - \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} \\ &= \mu + \sum_{i=h}^{\infty} \Phi_i u_{t+h-i} \\ &= \mu + \sum_{i=0}^{\infty} \Phi_{h+i} u_{t-i} \end{aligned}$$

El MSE o matriz de varianzas y covarianzas del error de predicción es

$$\Sigma_y(h) = MSE[y_t(h)] = \sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i' = \Sigma_y(h-1) + \Phi_{h-1} \Sigma_u \Phi_{h-1}'$$

A medida que $h \rightarrow \infty$, $\Sigma_y(h) = MSE[y_t(h)]$ tiende a la matriz de varianzas y covarianzas de y_t , $\Gamma_y(0) = \Sigma_y = \sum_{i=0}^{\infty} \Phi_i \Sigma_u \Phi_i'$.

La razón está en la estabilidad del proceso. Dado que los valores propios de A tienen todos módulo menor que la unidad, entonces $\Phi_i = J A^i J' \rightarrow 0$ cuando $i \rightarrow \infty$. Esto implica que el predictor óptimo en el largo plazo es la media del proceso. En otras palabras, el pasado del proceso no contiene información en el desarrollo del proceso en un futuro lejano.

Ejemplo 1

Utilizando las matrices de coeficientes de la representación MA del proceso, calculadas previamente, y Σ_u como sigue:

$$\Sigma_u = \begin{bmatrix} 2,25 & 0 & 0 \\ 0 & 1 & 0,5 \\ 0 & 0,5 & 0,74 \end{bmatrix}$$

se obtienen las siguientes matrices de MSE de predicción

$$\Sigma_y(1) = \Sigma_u$$

$$\Sigma_y(2) = \Sigma_u + \Phi_1 \Sigma_u \Phi_1' = \begin{bmatrix} 2,813 & 0,113 & 0 \\ 0,113 & 1,129 & 0,632 \\ 0 & 0,632 & 0,907 \end{bmatrix}$$

$$\Sigma_y(3) = \Sigma_y(2) + \Phi_2 \Sigma_u \Phi_2' = \begin{bmatrix} 2,953 & 0,146 & 0,011 \\ 0,146 & 1,161 & 0,663 \\ 0,011 & 0,663 & 0,943 \end{bmatrix}$$

Ejemplo 2

Utilizando las matrices de coeficientes de la representación MA del proceso, calculadas previamente, y Σ_u como sigue:

$$\Sigma_u = \begin{bmatrix} 0,09 & 0 \\ 0 & 0,04 \end{bmatrix}$$

se obtienen las siguientes matrices de MSE de predicción

$$\Sigma_y(1) = \Sigma_u$$

$$\Sigma_y(2) = \Sigma_u + \Phi_1 \Sigma_u \Phi_1' = \begin{bmatrix} 0,1129 & 0,02 \\ 0,022 & 0,0644 \end{bmatrix}$$

2.5.2. Predicción por intervalo

Para establecer intervalos de predicción necesitamos hacer supuestos sobre la distribución de y_t o de u_t . Normalmente se suponen procesos Gaussianos, esto es $u_t \sim N(0, \Sigma_u)$ y u_t y u_s independientes para $s \neq t$. Bajo estas condiciones los errores de predicción también tienen distribución Normal,

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} \sim N(0, \Sigma_y(h))$$

Si denotamos $y_{k,t+h}$ e $y_{k,t}(h)$ al k -ésimo componente de los vectores y_{t+h} e $y_t(h)$ respectivamente entonces se tiene que

$$\frac{y_{k,t+h} - y_{k,t}(h)}{\sigma_k(h)} \sim N(0, 1)$$

donde $\sigma_k(h)$ es la raíz cuadrada del k -ésimo elemento de la diagonal principal de $\Sigma_y(h)$. Denotemos por $z_{(\alpha)}$ tal que $Pr[Z \leq z_{(\alpha)}] = (1 - \alpha)$ donde $Z \sim N(0, 1)$.

Entonces un **intervalo de predicción h periodos hacia adelante** del $(1 - \alpha)100\%$ de confianza para $y_{k,t+h}$ es

$$y_{k,t}(h) \pm z_{(\alpha/2)}\sigma_k(h)$$

Si se computaran repetidamente intervalos de predicción para un gran número de realizaciones del proceso considerado entonces sobre un $(1 - \alpha)100\%$ de los intervalos contendrían el verdadero valor de $y_{k,t+h}$.

Ejemplo 1

Intervalos de predicción del 95 % de confianza para los componentes del VAR(1)⁴:

$$y_{1,t}(1) \pm 1,96\sqrt{2,25} = -3 \pm 2,94$$

$$y_{2,t}(1) \pm 1,96\sqrt{1,0} = 3,23 \pm 1,96$$

$$y_{3,t}(1) \pm 1,96\sqrt{0,74} = 3,1 \pm 1,96$$

$$y_{1,t}(2) \pm 1,96\sqrt{2,813} = -1,5 \pm 3,29$$

$$y_{2,t}(2) \pm 1,96\sqrt{1,129} = 2,95 \pm 2,08$$

$$y_{3,t}(2) \pm 1,96\sqrt{0,907} = 2,57 \pm 1,87$$

Para la construcción de **regiones de predicción conjuntas** para N componentes del vector y_{t+h} se utiliza la siguiente aproximación (basada en el método de Bonferroni). Si elegimos un intervalo de predicción h periodos hacia adelante de $(1 - \alpha)100\%$ para cada uno de las N variables, la región conjunta de predicción contendría todas ellas con una probabilidad de al menos $(1 - N\alpha)100\%$.

⁴Utilizando para este ejemplo las predicciones realizadas previamente y la matriz Σ_u

Por ejemplo, si los intervalos de predicción un periodo hacia adelante del $(1 - 0,05)100\% = 95\%$ para $y_{1,t+1}$ e $y_{2,t+1}$ son respectivamente

$$y_{1,t}(1) \pm 1,96\sigma_1(1) \text{ e } y_{2,t}(1) \pm 1,96\sigma_2(1)$$

entonces

$$\{y_{1,t}(1) \pm 1,96\sigma_1(1), y_{2,t}(1) \pm 1,96\sigma_2(1)\}$$

es una región de predicción conjunta de al menos $(1 - 0,05 \times 2)100\% = 90\%$ de confianza.

Ejemplo 1

$$\{y_{1,t}(1) \pm 1,96\sigma_1(1), y_{2,t}(1) \pm 1,96\sigma_2(1)\} =$$

$$\{-3 - 2,94 \leq y_{1,t+1} \leq -3 + 2,94; 3,2 - 1,96 \leq y_{2,t+1} \leq 3,2 + 1,96\}$$

es una región de predicción conjunta de al menos 90% de confianza.

2.5.3. Predicción con procesos estimados

En la sección anterior hemos visto que el predictor óptimo h periodos hacia adelante del proceso VAR(p) es

$$y_t(h) = \nu + A_1 y_t(h-1) + \dots + A_p y_t(h-p)$$

donde $y_t(j) = y_{t+j}$ para $j \leq 0$. Si en lugar de los verdaderos valores de los parámetros $B = \begin{pmatrix} \nu & A_1 & \dots & A_p \end{pmatrix}$ utilizamos el estimador $\hat{B} = \begin{pmatrix} \hat{\nu} & \hat{A}_1 & \dots & \hat{A}_p \end{pmatrix}$ obtenemos un predictor

$$\hat{y}_t(h) = \hat{\nu} + \hat{A}_1 \hat{y}_t(h-1) + \dots + \hat{A}_p \hat{y}_t(h-p)$$

donde $\hat{y}_t(j) = y_{t+j}$ para $j \leq 0$.

El **error de predicción** es

$$\begin{aligned} y_{t+h} - \hat{y}_t(h) &= [y_{t+h} - y_t(h)] + [y_t(h) - \hat{y}_t(h)] \\ &= \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} + [y_t(h) - \hat{y}_t(h)] \end{aligned}$$

Los dos términos del lado derecho de la igualdad están incorrelados porque el primero es función de periodos de tiempo $s > t$ mientras que el segundo depende de y_s con $s \leq t$. Por lo tanto el **MSE del predictor** $\hat{y}_t(h)$ es

$$\begin{aligned} \Sigma_{\hat{y}}(h) &= MSE[y_t(h)] + MSE[y_t(h) - \hat{y}_t(h)] \\ &= \Sigma_y(h) + MSE[y_t(h) - \hat{y}_t(h)] \\ &= \sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i' + MSE[y_t(h) - \hat{y}_t(h)] \end{aligned}$$

A la hora de evaluar $MSE[y_t(h) - \hat{y}_t(h)]$ se necesita la distribución del estimador \hat{B} . Dado que solamente conocemos la distribución asintótica de este estimador solamente es posible obtener una aproximación asintótica de $MSE[y_t(h) - \hat{y}_t(h)]$. Lütkepohl deriva, bajo ciertos supuestos, la siguiente aproximación para el MSE de $\hat{y}_t(h)$

$$\Sigma_{\hat{y}}(h) = \Sigma_y(h) + \frac{1}{T}\Omega(h)$$

donde

$$\Omega(h) = E \left[\frac{\partial y_t(h)}{\partial \beta'} \Sigma_{\hat{\beta}} \frac{\partial y_t(h)}{\partial \beta} \right]$$

siendo $\beta = \text{vec}(B)$ y $\Sigma_{\hat{\beta}}$ la matriz de varianzas y covarianzas asintótica. A mayor eficiencia a la hora de estimar B menor incertidumbre en la predicción.

Lütkepohl (1991, pp. 88) obtiene una expresión explícita para $\Omega(h)$ utilizando el estimador de MCO de $\hat{\beta}$, que para $h = 1$ es simplemente

$$\Omega(1) = (kp + 1)\Sigma_u$$

Así, la aproximación para el MSE de la predicción un periodo hacia adelante a partir del origen de predicción T (se supone que el periodo utilizado para la estimación es hasta T) es

$$\begin{aligned} \Sigma_{\hat{y}}(1) &= \Sigma_y(1) + (1/T)\Omega(1) \\ &= \Sigma_u + \frac{kp+1}{T}\Sigma_u = \frac{T+kp+1}{T}\Sigma_u \end{aligned}$$

Podemos ver que la variabilidad que se introduce a la hora de predecir en términos del MSE de la predicción utilizando estimaciones de los coeficientes del VAR depende de la dimensión del proceso (k), del orden del VAR, p , y del tamaño muestral T utilizado en la estimación. Si T es pequeño, puede ser bastante importante. A medida que el tamaño muestral crece, ese factor tiende a uno. Esto se debe a que estimamos consistentemente los coeficientes del VAR.

Para $h > 1$ también depende de B, Γ, Φ_i además de Σ_u . En la práctica se reemplazarán estas matrices por sus correspondientes estimadores consistentes para obtener un estimador consistente de $\Sigma_{\hat{y}}(h)$. Esto es importante a la hora de construir intervalos de confianza para las predicciones que serán aproximaciones asintóticas.

Suponiendo y_t gaussiano, un intervalo de predicción h periodos hacia adelante para el componente k -ésimo de y_t de aproximadamente $(1 - \alpha)100\%$ de confianza es

$$\hat{y}_{k,t}(h) \pm z_{(\alpha/2)} \hat{\sigma}_k(h)$$

donde $\hat{\sigma}_k(h)$ es la raíz cuadrada del elemento k -ésimo de la diagonal principal del estimador consistente de $\Sigma_{\hat{y}}(h)$.

2.6. Causalidad en el sentido de Granger

2.6.1. Definición

Es un concepto⁵ que está muy relacionado con predicción y que se analiza fácilmente en un contexto VAR por lo que es muy popular aunque como veremos tiene sus limitaciones.

Idea: **la causa no puede venir después del efecto**. Así, si una variable x afecta a otra variable z la primera debe de ayudar a predecir la segunda.

- Ω_t : Conjunto de información que contiene toda la información relevante en el universo disponible hasta el momento t incluido.
- $z(h|\Omega_t)$: predictor óptimo (en el sentido de minimizar el MSE) h periodos hacia adelante del proceso z_t en el origen de predicción t condicionado al conjunto de información Ω_t .

El proceso x_t se dice que **causa z_t en el sentido de Granger** si

$$\Sigma_z(h|\Omega_t) < \Sigma_z(h|\Omega_t \setminus \{x_s | s \leq t\})$$

para al menos un $h = 1, 2, \dots$ donde $\Omega_t \setminus \{x_s | s \leq t\}$ es el conjunto de toda la información relevante en el universo excepto por la información pasada y presente del proceso x_t .

En otras palabras, si z_t se puede predecir más eficientemente (con un menor MSE) teniendo en cuenta la información pasada y presente del proceso x_t además del resto de información relevante en el universo, entonces x_t *causa en el sentido de Granger* a z_t .

⁵Este concepto de causalidad fue introducido por Granger en su artículo publicado en *Econometrica* en 1969, *Investigating causal relations by econometric models and cross spectral methods*.

Generalización al caso donde z_t y x_t son procesos vectoriales. En ese caso $\Sigma_z(h|\Omega_t)$ y $\Sigma_z(h|\Omega_t \setminus \{x_s | s \leq t\})$ son matrices, por lo que se requiere para decir que x_t causa en el sentido de Granger a z_t que las dos matrices no sean idénticas y que la diferencia entre la primera y la segunda sea una matriz semidefinida positiva para algún t y h .

Si x_t causa a z_t y z_t causa a x_t el proceso $(z'_t, x'_t)'$ recibe el nombre de *feedback system*.

Problema: elección del conjunto de información Ω_t . Normalmente en la práctica no se dispone de toda la información relevante en el universo. Por lo tanto, en la práctica se considera como relevante solamente la información pasada y presente del proceso que se estudia y Ω_t se reemplaza por $\{z_s, x_s | s \leq t\}$. Más aún, en lugar de comparar predictores óptimos se comparan predictores *lineales* óptimos.

2.6.2. No-Causalidad en el sentido de Granger

Consideremos la representación MA del proceso $y_t = (z'_t, x'_t)'$ donde z_t es un vector $(m \times 1)$ y x_t es un vector $((k - m) \times 1)$

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i} = \mu + \Phi(L)u_t, \quad \Phi_0 = I_k$$

particionando de acuerdo con $y_t = (z'_t, x'_t)'$,

$$y_t = \begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

Proposición. x_t **no causa en el sentido de Granger** a z_t si y solo si $\Phi_{12}(L) = 0$, o equivalentemente,

$$z_t(h|\{y_s | s \leq t\}) = z_t(h|\{z_s | s \leq t\}) \text{ para } h = 1, 2, \dots$$

\Leftrightarrow

$$\Phi_{12,i} = 0 \text{ para } i = 1, 2, \dots$$

Alternativamente, z_t **no causa en el sentido de Granger** a x_t si y solo si $\Phi_{21}(L) = 0$, o equivalentemente,

$$x_t(h|\{y_s | s \leq t\}) = x_t(h|\{x_s | s \leq t\}) \text{ para } h = 1, 2, \dots$$

\Leftrightarrow

$$\Phi_{21,i} = 0 \text{ para } i = 1, 2, \dots$$

Teniendo en cuenta que $(A(L))^{-1} = \Phi(L)$ podemos ver que implica esta condición en términos de las matrices de coeficientes del VAR(p),

$$y_t = \begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} z_t \\ x_t \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

Corolario. Si y_t es un proceso VAR(p) estable entonces x_t no causa en el sentido de Granger a z_t si y solo si $A_{12}(L) = 0$, o equivalentemente,

$$z_t(h|\{y_s|s \leq t\}) = z_t(h|\{z_s|s \leq t\}) \text{ para } h = 1, 2, \dots$$

\Leftrightarrow

$$A_{12,i} = 0 \text{ para } i = 1, 2, \dots, p.$$

Alternativamente, z_t no causa en el sentido de Granger a x_t si y solo si $A_{21}(L) = 0$, o equivalentemente,

$$x_t(h|\{y_s|s \leq t\}) = x_t(h|\{x_s|s \leq t\}) \text{ para } h = 1, 2, \dots$$

\Leftrightarrow

$$A_{21,i} = 0 \text{ para } i = 1, 2, \dots, p.$$

Contraste de causalidad en el sentido de Granger:

Podemos contrastar la hipótesis nula de no-Causalidad en el sentido de Granger de un subconjunto de variables z_t en el resto de variables x_t del sistema considerado y_t en términos de contrastes de restricciones cero sobre cierto subconjunto de coeficientes de las matrices A_1, \dots, A_p del VAR.

Se pueden llevar a cabo con el estadístico de Wald como hemos descrito en la sección de estimación y contrastes de restricciones, o con un contraste de Razón de verosimilitudes

$$\lambda_{LR} = T(\ln|\tilde{\Sigma}_u^R| - \ln|\tilde{\Sigma}_u|) \xrightarrow{d, H_0} \chi^2(q)$$

donde q es el número de restricciones bajo H_0 , $\tilde{\Sigma}_u^R$ es el estimador MV de Σ_u bajo la hipótesis nula de no causalidad y $\tilde{\Sigma}_u$ la correspondiente al no restringido.

Ejemplo: consideremos el siguiente sistema bivalente VAR(2):

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{11,2} & a_{12,2} \\ a_{21,2} & a_{22,2} \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

$H_0 : y_{2t}$ no causa en el sentido de Granger a $y_{1t} \Leftrightarrow H_0 : a_{12,1} = a_{12,2} = 0.$

Dado que esta restricción solo afecta a los coeficientes de una sola ecuación, este contraste se puede llevar a cabo con el estadístico tipo F usual y comparar el valor del estadístico multiplicado por el número de restricciones, en este caso 2, con el valor crítico en la distribución Chi-cuadrado con 2 grados de libertad.

Comentarios:

1. En general hay que ser cautelosos a la hora de interpretar la existencia o no de causalidad en el sentido de Granger con una implicación sobre la dirección de causalidad en términos más generales de que una variable determine otra.

Por ejemplo, series temporales que reflejan un comportamiento *forward-looking* de los agentes económicos, como por ejemplo los precios de los activos o tipos de interés, en general se encuentran que son excelentes predictores de otras importantes series económicas. Esto no significa que estas series *causen* el PNB, la inflación o los dividendos a subir o bajar sino que estas variables incorporan la mejor predicción del mercado sobre la futura evolución de estas otras variables. En este sentido, contrastar causalidad en el sentido de Granger puede ser útil para analizar la hipótesis de eficiencia de los mercados pero no para inferir una dirección de causalidad. (ver algunos ejemplos en el Hamilton pp: 306-308)

2. Otro problema está en que en la práctica nos restringimos a un conjunto de información concreto que incluye solamente los valores pasados y presente de las variables del sistema considerado. El considerar otras variables relevantes en el sistema y por lo tanto en el conjunto de información pueden cambiar las relaciones de causalidad en el sentido de Granger encontradas en un sistema que las ha omitido.
3. La frecuencia de los datos considerados, esto es que el conjunto de información contenga datos mensuales, trimestrales o anuales puede influir en que se encuentren relaciones causales a la Granger para una frecuencia y no para otras. Por otro lado, considerar series ajustadas estacionalmente o no en el conjunto de información también puede tener su influencia.

Aún así, el análisis de causalidad en el sentido de Granger es un instrumento útil para al menos decirnos algo sobre si un conjunto de variables contienen información útil para mejorar nuestras predicciones sobre otro conjunto de variables.

2.7. Análisis Impulso-Respuesta

El interés de este análisis se centra en ver la reacción o respuesta de una variable a una innovación o shock exógeno (impulso) en otra variable del sistema que puede constar de otras variables además de estas.

Ejemplo 1: (VAR(1) tridimensional donde $\nu = 0$)

Supongamos un sistema de tres variables $y_t = (y_{1t} \ y_{2t} \ y_{3t})'$ que sigue un proceso estable VAR(1) dado por

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

Supongamos que las tres variables antes del shock en $t = 0$ toman su valor medio que hemos supuesto igual a cero (ya que lo que nos interesa es ver la variación de las variables alrededor de su valor medio).

¿Cuál es la respuesta del sistema a un incremento unitario en la primera variable, esto es, $u_{1,0} = 1$ durante los siguientes periodos posteriores al shock $t = 1, 2, \dots$, si no hay otros shocks en el sistema, es decir $u_{2,0} = 0, u_{3,0} = 0, u_1 = 0, u_2 = 0, \dots$?

$$y_0 = \begin{bmatrix} y_{1,0} \\ y_{2,0} \\ y_{3,0} \end{bmatrix} = \begin{bmatrix} u_{1,0} \\ u_{2,0} \\ u_{3,0} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$y_1 = \begin{bmatrix} y_{1,1} \\ y_{2,1} \\ y_{3,1} \end{bmatrix} = A_1 y_0 = \begin{bmatrix} 0,5 \\ 0,1 \\ 0 \end{bmatrix}$$

$$y_2 = \begin{bmatrix} y_{1,2} \\ y_{2,2} \\ y_{3,2} \end{bmatrix} = A_1 y_1 = A_1^2 y_0 = \begin{bmatrix} 0,25 \\ 0,06 \\ 0,02 \end{bmatrix}$$

Continuando con este proceso resulta que $y_i = (y_{1,i}, y_{2,i}, y_{3,i})'$ es la primera columna de A_1^i . De la misma forma, un shock unitario en y_{2t} (y_{3t}) en $t = 0$ después de i periodos resulta en un vector y_i igual a la segunda (tercera) columna de A_1^i . Así, **los elementos de A_1^i representan los efectos de shocks unitarios en las variables del sistema después de i periodos**. Por ello se les conoce con el nombre de **multiplicadores dinámicos o respuestas al impulso**.

Recordar que $A_1^i = \Phi_i$ donde Φ_i es la i -ésima matriz de coeficientes de la representación de medias móviles (MA) del proceso VAR(1). Por lo tanto, **las matrices de coeficientes de la representación MA contienen las respuestas al impulso del sistema.**

Generalización a un VAR(p)

Si y_t sigue un proceso estable VAR(p), supongamos de media cero, se puede escribir en forma de un VAR(1) (companion form)

$$Y_t = AY_{t-1} + U_t$$

Siguiendo con los mismos supuestos que en el ejemplo anterior, $y_t = 0$ para $t < 0$, $u_t = 0$ para $t > 0$ y $y_0 = u_0 = e_k$ donde e_k es un vector ($k \times 1$) con un uno en la k -ésima posición y cero en el resto,

$$Y_0 = \begin{pmatrix} e_k' & 0 & \dots & 0 \end{pmatrix}'$$

Por lo tanto,

$$Y_i = A^i Y_0$$

Así tenemos que las respuestas-impulso son los elementos de la matriz ($k \times k$), $\Phi_i = JA^iJ'$ donde $J = \begin{bmatrix} I_k & 0 & \dots & 0 \end{bmatrix}$ es la matriz de selección ($k \times kp$).

El coeficiente $\phi_{jk,i}$ de la matriz Φ_i representa la reacción de la variable j -ésima del sistema a un shock unitario de la variable k -ésima, i periodos después del shock, manteniendo que no hay otros shocks en el sistema del resto de variables ni otros shocks posteriores al impulso. La **función impulso-respuesta** sería la representación gráfica de los coeficientes $\phi_{jk,i}$ para $i = 1, 2, \dots$

1. Si las variables están medidas en escalas diferentes se suelen considerar innovaciones iguales a su desviación típica ($u_{j,0} = \sqrt{\text{var}(u_j)}$) en lugar de shocks unitarios. Esto no sería más que reescalar las funciones impulso-respuesta.
2. Si una variable no causa en el sentido de Granger al resto de variables en el sistema entonces las respuestas al impulso en esa variable sobre el resto serán cero.
3. Lütkepohl (1991, pp. 45) demuestra que, para un VAR(p) estable k -dimensional, si las primeras $(pk - p)$ respuestas de la variable j -ésima a un impulso en la variable k -ésima son cero, entonces todas las siguientes respuestas tienen que ser cero. Por ejemplo si tenemos un VAR(2) trivariante ($k=3, p=2$) si $\phi_{12,i} = 0$ para $i = 1, 2, 3, 4$ entonces $\phi_{12,i} = 0$ para $i = 1, 2, \dots$

4. La j -ésima columna de la matriz $\Psi_m = \sum_{i=0}^m \Phi_i$ contiene las respuestas acumuladas hasta m periodos de cada una de las variables del sistema a un shock unitario en la variable j -ésima. Se conoce con el nombre de m -ésimo **multiplicador intermedio** ("interim multiplier").
5. La matriz $\Psi_\infty = \sum_{i=0}^{\infty} \Phi_i = \Phi(1) = (I_k - A_1 - \dots - A_p)^{-1}$ es la **matriz de los efectos de largo plazo** o multiplicadores totales.

Ejemplo 1: VAR(1) tridimensional con $\nu = 0$.

Multiplicadores dinámicos o respuestas al impulso, $\Phi_i = A_1^i$

$$\Phi_0 = I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\Phi_1 = A_1 = \begin{bmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{bmatrix}$$

$$\Phi_2 = A_1^2 = \begin{bmatrix} 0,25 & 0 & 0 \\ 0,06 & 0,07 & 0,12 \\ 0,02 & 0,08 & 0,15 \end{bmatrix}$$

$$\Phi_3 = A_1^3 = \begin{bmatrix} 0,125 & 0 & 0 \\ 0,037 & 0,031 & 0,057 \\ 0,018 & 0,038 & 0,069 \end{bmatrix}$$

etc.

Gráficos (Ver Lütkepohl, p. 46 y 49).

Ortogonalización:

Un supuesto problemático en el análisis impulso-respuesta es que los shocks solamente se dan en una variable cada vez. Si los errores están correlacionados un shock en una variable puede venir acompañado de un shock en otra variable por lo que ignorarlo puede distorsionar la verdadera relación dinámica entre las variables.

Por esta razón normalmente el análisis impulso-respuesta se lleva a cabo en términos de las matrices de coeficientes de la representación MA

$$y_t = \sum_{i=0}^{\infty} \Theta_i w_{t-i} = \sum_{i=0}^{\infty} \Phi_i P P^{-1} u_{t-i} = \sum_{i=0}^{\infty} \Phi_i u_{t-i}$$

donde $\Theta_i = \Phi_i P$, $w_t = P^{-1} u_t$ y P es una matriz triangular con ceros por arriba de la diagonal tal que $\Sigma_u = P P'$ utilizando la descomposición de Cholesky. De esta forma los componentes de w_t están incorrelacionados y tienen varianza unitaria $\Sigma_w = P^{-1} \Sigma_u (P^{-1})' = I_k$.

A diferencia de $\Phi_0 = I_k$ en términos de esta representación $\Theta_0 = P$ por lo que, a no ser que Σ_u sea diagonal, esta matriz no será diagonal y sus elementos recogerán las respuestas inmediatas del sistema a shocks unitarios. Por ello se conocen con el nombre de **multiplicadores de impacto**.

El hecho de que $\Theta_0 = P$ sea triangular por abajo implica que el orden de las variables en el vector es importante⁶. La primera variable será la única que pueda tener un impacto inmediato en todo el resto de variables. La segunda puede tener un impacto inmediato en las $(k-2)$ restantes variables pero no en la primera, y así sucesivamente. Esta elección puede determinar las respuestas al impulso y por lo tanto es crítico a la hora de su interpretación.

Ejemplo 1:

$$\text{VAR}(1) \text{ tridimensional con } \nu = 0 \text{ y } \Sigma_u = \begin{bmatrix} 2,25 & 0 & 0 \\ 0 & 1 & 0,5 \\ 0 & 0,5 & 0,74 \end{bmatrix}$$

Respuestas a impulsos ortogonales, $\Theta_i = \Phi_i P$

$$\Theta_0 = P = \begin{bmatrix} 1,5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0,5 & 0,7 \end{bmatrix}$$

⁶Diferente orden puede originar diferentes funciones de respuesta al impulso.

$$\Theta_1 = \Phi_1 P = \begin{bmatrix} 0,75 & 0 & 0 \\ 0,15 & 0,25 & 0,21 \\ 0 & 0,35 & 0,21 \end{bmatrix}$$

$$\Theta_2 = \Phi_2 P = \begin{bmatrix} 0,375 & 0 & 0 \\ 0,09 & 0,13 & 0,084 \\ 0,03 & 0,055 & 0,105 \end{bmatrix}$$

etc. Gráficos (Lütkepohl, p.53).

Normalmente tenemos que **estimar previamente el VAR** para obtener las funciones impulso-respuesta. Una vez estimadas las matrices de coeficientes A_1, \dots, A_p las Φ_i se pueden obtener recursivamente de

$$\Phi_i = \sum_{j=1}^i \Phi_{i-j} A_j \quad i = 1, 2, \dots$$

comenzando con $\Phi_0 = I_k$ y haciendo $A_j = 0$ para $j > p$. Ejemplo:
 $y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + u_t$

Entonces $\Phi_0 = I_k$, $\Phi_1 = \Phi_0 A_1$, $\Phi_2 = \Phi_1 A_1 + \Phi_0 A_2$, $\Phi_3 = \Phi_2 A_1 + \Phi_1 A_2, \dots, \Phi_n = \Phi_{n-1} A_1 + \Phi_{n-2} A_2$

En el caso de las funciones impulso-respuesta ortogonalizadas las matrices Θ_i se estimarían de $\Theta_i = \Phi_i P$ una vez obtenidas las Φ_i como antes y la matriz P de la descomposición de Cholesky $\Sigma_u = P P'$ dada la estimación de Σ_u .

2.8. Selección del orden del VAR

En la práctica el orden del VAR, p , será desconocido. Elegir un orden p innecesariamente largo reduce la precisión de la predicción del modelo VAR(p) estimado. También la precisión en la estimación de las funciones impulso-respuesta depende de la precisión en la estimación de los parámetros del VAR. Por lo tanto, es útil disponer de procedimientos o criterios para elegir un orden del VAR adecuado.

A) Contrastes secuenciales: Likelihood ratio (LR) test.

Suponiendo que M es la cota superior del orden del VAR y que es conocida, se puede utilizar el siguiente procedimiento secuencial utilizando contrastes de Razón de verosimilitud (LR).

1. Elegir un orden de VAR como cota superior M .
2. Contrastar H_0 : $M-1$ retardos frente a H_a : M retardos. Si no se rechaza la hipótesis nula,

3. Contrastar H_0 : (M-2) retardos frente a H_a : (M-1) retardos, continuando el proceso hasta rechazar la hipótesis nula en favor de la alternativa, eligiendo el orden de retardos de la alternativa. En términos de restricciones en las matrices de coeficientes del VAR, este proceso sería el siguiente:

$$\begin{array}{ll}
H_0^1 : A_M = 0 & H_1^1 : A_M \neq 0 \\
H_0^2 : A_{M-1} = 0 & H_1^2 : A_{M-1} \neq 0 \mid A_M = 0 \\
\vdots & \\
H_0^i : A_{M-i+1} = 0 & H_1^i : A_{M-i+1} \neq 0 \mid A_M = \dots = A_{M-i+2} = 0 \\
\vdots & \\
H_0^M : A_1 = 0 & H_1^M : A_1 \neq 0 \mid A_M = \dots = A_2 = 0
\end{array}$$

En este proceso secuencial cada hipótesis nula es contrastada condicionando a que las hipótesis nulas previas sean ciertas.

Si $H_0^i : A_{M-i+1} = 0$ se rechaza entonces el orden del VAR elegido es $\hat{p} = M - i + 1$.

El estadístico de Razón de verosimilitudes (LR) para contrastar la i -ésima hipótesis nula es

$$\lambda_{LR}(i) = T(\ln|\widetilde{\Sigma}_u(M-i)| - \ln|\widetilde{\Sigma}_u(M-i+1)|)$$

donde $\widetilde{\Sigma}_u(m)$ es el estimador por Máxima Verosimilitud de Σ_u cuando se ajusta un VAR(m) a la serie temporal de T observaciones (disponiendo de m valores iniciales). Este estadístico tiene una distribución asintótica $\chi^2(k^2)$ si H_0^i y todas las hipótesis nulas anteriores son ciertas.

- Este resultado se mantiene aunque y_t no sea gaussiano. Solamente se necesita que u_t sea un ruido blanco independientemente distribuido con momentos de cuarto orden acotados (standard white noise).
- El estadístico LR también se puede expresar como el estadístico de Wald pero donde el estimador de Σ_u considerado es el del modelo restringido, esto es bajo H_0^i .
- Dado que es un contraste secuencial el nivel de significación del contraste no coincide con el nivel de significación de cada uno de los contrastes llevados a cabo en la secuencia. Por ejemplo si el nivel de significación en cada paso es $\alpha = 0,05$ entonces $\alpha_1 = 1 - (1 - 0,05)$, $\alpha_2 = 1 - (1 - 0,05)^2 \dots \alpha_i = 1 - (1 - 0,05)^i \dots \alpha_M = 1 - (1 - 0,05)^M$. Aunque esto es una

aproximación del error de tipo I dado que la distribución es asintótica.

Para obtener este resultado se ha utilizado el hecho de que para $m \neq j$ y $m, j, \leq i$, $\lambda_{LR}(m)$ y $\lambda_{LR}(j)$ son estadísticos asintóticamente independientes si H_0^1, \dots, H_0^i son ciertas.

B) Criterios de selección del orden del VAR

B.1) Minimizar el MSE de predicción.

Si el objetivo es predecir tiene sentido elegir el orden del VAR tal que se minimice una medida de la precisión de la predicción. Esta medida es el error cuadrático medio (MSE) de la predicción. En estos términos los siguientes dos criterios están diseñados para minimizar la varianza del error de predicción.

B.1.1) Final Prediction Error (FPE).

Akaike(1969,1971) sugiere el siguiente criterio basado en el MSE aproximado de la predicción 1 periodo hacia adelante.

$$FPE(m) = \left[\frac{T + km + 1}{T - km - 1} \right]^k \det(\widetilde{\Sigma}_u(m))$$

Basado en este criterio el orden del VAR elegido $\hat{p}(FPE)$ de p es tal que

$$FPE[\hat{p}(FPE)] = \min\{FPE(m) | m = 0, 1, \dots, M\}$$

Esto es, se estiman modelos VAR de órdenes $m = 0, 1, \dots, M$ y se calculan los correspondientes valores de $FPE(m)$. El orden que obtiene un menor FPE es el orden \hat{p} elegido para p .

B.1.2) Criterio de Información de Akaike (AIC)

También derivado por Akaike(1973,1974), para un VAR(m) se define como

$$AIC(m) = \ln|\widetilde{\Sigma}_u(m)| + \frac{2}{T}(mk^2)$$

Basado en este criterio el orden del VAR elegido $\hat{p}(AIC)$ de p es tal que

$$AIC[\hat{p}(AIC)] = \min\{AIC(m) | m = 0, 1, \dots, M\}$$

Lütkepohl (1991) demuestra que

$$\ln FPE(m) = AIC(m) + 2k/T + O(T^{-2})$$

Dado que el término $2k/T$ no depende de m , $AIC(m)$ y $AIC(m) + 2k/T$ se minimizan para el mismo valor de m . Por lo tanto, $\ln FPE$ y AIC difieren por un término de orden $O(T^{-2})$ por lo que ambos criterios serán casi equivalentes para tamaños muestrales moderados y grandes.

B.2) Selección del orden consistente.

Un estimador \hat{p} del orden del VAR p es consistente si

$$\text{plim } \hat{p} = p \text{ cuando } T \rightarrow \infty$$

Lütkepohl prueba condiciones necesarias y suficientes para que \hat{p} sea consistente y estas condiciones no son satisfechas por $\hat{p}(AIC)$ ni por $\hat{p}(FPE)$. AIC y FPE sobreestiman o eligen un orden mayor que el verdadero valor de p asintóticamente con probabilidad positiva. También demuestra que la probabilidad de que estos dos criterios elijan un orden del VAR menor que el verdadero es cero asintóticamente.

Criterios consistentes:

B.2.1) Criterio de Hannan-Quinn (HQ).

$$HQ(m) = \ln|\widetilde{\Sigma}_u(m)| + 2\frac{\ln \ln T}{T}(mk^2)$$

Basado en este criterio el orden del VAR elegido $\hat{p}(HQ)$ de p es tal que

$$HQ[\hat{p}(HQ)] = \min\{HQ(m) | m = 0, 1, \dots, M\}$$

B.2.2) Criterio de Schwarz (SC) Utilizando argumentos Bayesianos Schwarz (1978) deriva el siguiente criterio:

$$SC(m) = \ln|\widetilde{\Sigma}_u(m)| + \frac{\ln T}{T}(mk^2)$$

De nuevo el orden del VAR elegido por este criterio es el que minimiza el valor del SC.

Lütkepohl establece la siguiente comparación⁷ para muestras pequeñas de los criterios AIC, HQ y SC.

$$\begin{array}{ll} \hat{p}(SC) \leq \hat{p}(AIC) & \text{si } T \geq 8; \\ \hat{p}(SC) \leq \hat{p}(HQ) & \text{para todo } T; \\ \hat{p}(HQ) \leq \hat{p}(AIC) & \text{si } T \geq 16. \end{array}$$

⁷Para este último resultado no se requiere la estacionariedad de y_t .

Contrastes de Autocorrelación en los residuos del VAR:

También se pueden llevar a cabo contrastes a nivel de ecuación por ecuación o en el sistema para ver si los residuos son ruido blanco (Ljung-Box test o Portmanteau test).

Para contrastar si las matrices de autocorrelaciones R de los errores son cero hasta un número de retardos h , el estadístico del contraste de Portmanteau es

$$H_0 : R_h = (r_1, \dots, r_h) = 0 \text{ frente a } H_1 : R_h \neq 0$$

Estadístico

$$P_h = T \sum_{i=1}^h tr(\hat{C}'_i \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1})$$

donde tr indica la traza de una matriz, $\hat{C}_i = 1/T \sum_{t=i+1}^T \hat{u}_t \hat{u}'_{t-i}$. La distribución asintótica de este estadístico bajo la hipótesis nula es $\chi^2(k^2(h-p))$.

En estudios de Montecarlo se ha visto que este contraste tiene poca potencia frente a muchas alternativas. Una modificación a este estadístico para mejorar este problema es

$$\bar{P}_h = T^2 \sum_{i=1}^h (T-i)^{-1} tr(\hat{C}'_i \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1})$$

manteniendo la misma distribución asintótica.

Capítulo 3

Ejemplo VAR con Gretl

3.1. Estimación de un VAR

Los datos para este ejemplo están en el fichero `datos_VAR.xls`.
Periodo muestral: Son datos trimestrales de 1959 : 1 a 2005 : 3.
Variables: `gdplev`, `pricelev`, `fedfunds`, `M2`.

1. Leemos los datos con Gretl.

Archivo → *Abrir datos* → *Importar* → *excel*

Estructura conjunto de datos → Serie temporal → Trimestral → observación inicial 1959:1

2. Definimos las siguientes variables en tasas de crecimiento:

$gdpch = 400 * \log(gdplev(t)/gdplev(t - 1))$

$inflation = 400 * \log(pricelev(t)/pricelev(t - 1))$

$mgrow = 400 * \log(M2(t)/M2(t - 1))$

Para ello en Gretl una vez seleccionamos con el cursor las variables en niveles `gdplev`, `pricelev` y `M2`, vamos a

Añadir → *diferencias de logaritmos de las variables seleccionadas*

Se añaden al listado de series estas variables

`ld_gdplev`, `ld_pricelev`, `ld_M2`

y finalmente se obtiene `gdpch`, `inflation`, `mgrow`, con

Añadir → *definir nueva variable*

multiplicando por 400 las diferencias de logaritmos.

3. Gráficos de las series

Ver → *graficos* → *series temporales*

Figura 3.1: GDP growth rate

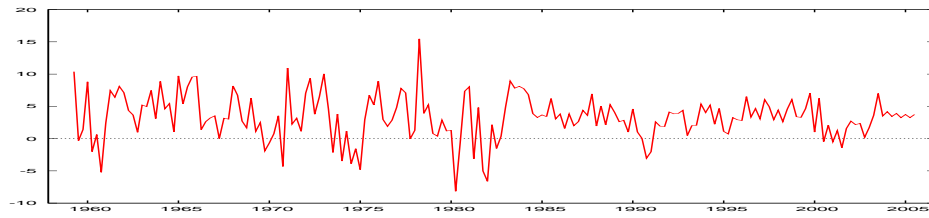


Figura 3.2: Federal Funds Rate

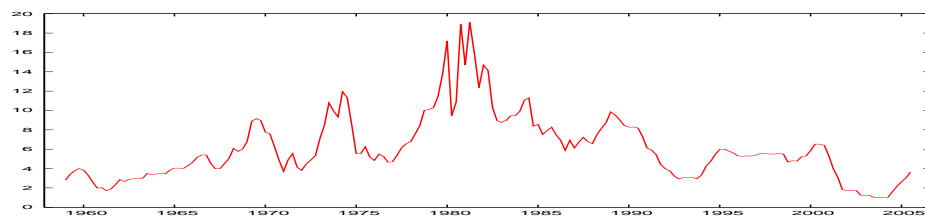


Figura 3.3: Inflation

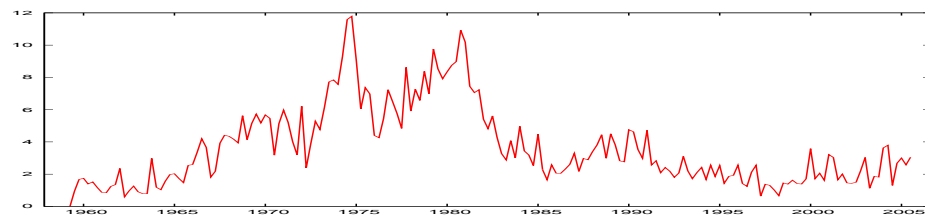
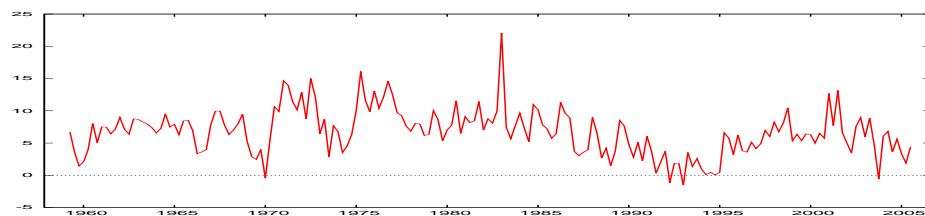


Figura 3.4: M2 growth rate



Seguidamente vamos a estimar un VAR(4) para el vector de variables endógenas $y_t = (gdpch_t, inflation_t, fedfunds_t, mgrow_t)$:

Modelo \rightarrow *Series temporales* \rightarrow *Autorregresion vectorial (VAR)*

Los resultados se muestran en otra ventana en la que hay otras opciones asociadas a la estimación de este VAR y son los siguientes¹:

Sistema VAR, orden del retardo 4
estimaciones MCO, observaciones 1960:2–2005:3 ($T = 182$)

Ecuación 1: gdpch

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	2,08223	0,811822	2,5649	0,0112
$gdpch_{t-1}$	0,118068	0,0790540	1,4935	0,1372
$gdpch_{t-2}$	0,203099	0,0808896	2,5108	0,0130
$gdpch_{t-3}$	-0,0398229	0,0744585	-0,5348	0,5935
$gdpch_{t-4}$	0,0276374	0,0695558	0,3973	0,6916
$inflation_{t-1}$	0,171477	0,211778	0,8097	0,4193
$inflation_{t-2}$	-0,00603636	0,233561	-0,0258	0,9794
$inflation_{t-3}$	-0,313410	0,239204	-1,3102	0,1919
$inflation_{t-4}$	0,0274599	0,217168	0,1264	0,8995
$fedfunds_{t-1}$	0,141392	0,206818	0,6837	0,4952
$fedfunds_{t-2}$	-1,0120	0,228993	-4,4196	0,0000
$fedfunds_{t-3}$	0,424542	0,240408	1,7659	0,0793
$fedfunds_{t-4}$	0,247841	0,219215	1,1306	0,2599
$mgrow_{t-1}$	0,0777511	0,0898896	0,8650	0,3883
$mgrow_{t-2}$	0,0865794	0,0946574	0,9147	0,3617
$mgrow_{t-3}$	0,103284	0,0916783	1,1266	0,2615
$mgrow_{t-4}$	0,00249869	0,0852544	0,0293	0,9767

$$R^2 = 0,325262 \quad \bar{R}^2 \text{ corregido} = 0,259833 \quad F(16, 165) = 4,97121$$

Contrastes F de restricciones cero

Todos los retardos de gdpch	$F(4, 165) = 2,33167$	valor p 0,0580
Todos los retardos de inflation	$F(4, 165) = 0,672294$	valor p 0,6121
Todos los retardos de fedfunds	$F(4, 165) = 6,06702$	valor p 0,0001
Todos los retardos de mgrow	$F(4, 165) = 2,19343$	valor p 0,0719
Todas las variables, retardo 4	$F(4, 165) = 0,362595$	valor p 0,8350

¹No se muestran algunos resultados que muestra Gretl en la estimación de cada ecuación, por ejemplo, Suma de cuadrados de los residuos, Desviación típica de los residuos, Estadístico de Durbin–Watson, Coef. de autocorr. de primer orden, Criterio de información de Akaike, Criterio de información Bayesiano de Schwarz, Criterio de Hannan–Quinn.

Ecuación 2: inflation

Variable	Coefficiente	Desv. típica	Estadístico t	valor p
const	0,174894	0,295202	0,5925	0,5544
gdpch _{$t-1$}	-0,0196801	0,0287463	-0,6846	0,4945
gdpch _{$t-2$}	-0,0220202	0,0294138	-0,7486	0,4551
gdpch _{$t-3$}	0,0154169	0,0270752	0,5694	0,5699
gdpch _{$t-4$}	0,0475021	0,0252925	1,8781	0,0621
inflation _{$t-1$}	0,463886	0,0770085	6,0238	0,0000
inflation _{$t-2$}	0,123791	0,0849294	1,4576	0,1469
inflation _{$t-3$}	0,110048	0,0869814	1,2652	0,2076
inflation _{$t-4$}	0,256896	0,0789685	3,2531	0,0014
fedfunds _{$t-1$}	0,211492	0,0752049	2,8122	0,0055
fedfunds _{$t-2$}	-0,0863222	0,0832683	-1,0367	0,3014
fedfunds _{$t-3$}	-0,0622490	0,0874193	-0,7121	0,4774
fedfunds _{$t-4$}	-0,0966751	0,0797128	-1,2128	0,2269
mgrow _{$t-1$}	0,00421371	0,0326865	0,1289	0,8976
mgrow _{$t-2$}	-0,00278709	0,0344202	-0,0810	0,9356
mgrow _{$t-3$}	0,00736823	0,0333369	0,2210	0,8253
mgrow _{$t-4$}	0,0133597	0,0310010	0,4309	0,6671

$$R^2 = 0,825659 \quad \bar{R}^2 \text{ corregido} = 0,808754 \quad F(16, 165) = 48,8390$$

Contrastes F de restricciones cero

Todos los retardos de gdpch	$F(4, 165) = 1,30313$	valor p 0,2710
Todos los retardos de inflation	$F(4, 165) = 71,6211$	valor p 0,0000
Todos los retardos de fedfunds	$F(4, 165) = 2,5083$	valor p 0,0440
Todos los retardos de mgrow	$F(4, 165) = 0,135041$	valor p 0,9692
Todas las variables, retardo 4	$F(4, 165) = 3,46483$	valor p 0,0095

Ecuación 3: fedfunds

Variable	Coefficiente	Desv. típica	Estadístico t	valor p
const	-0,637819	0,333410	-1,9130	0,0575
gdpch _{$t-1$}	0,108564	0,0324670	3,3438	0,0010
gdpch _{$t-2$}	0,0640776	0,0332209	1,9288	0,0555
gdpch _{$t-3$}	0,0122803	0,0305796	0,4016	0,6885
gdpch _{$t-4$}	0,0154435	0,0285661	0,5406	0,5895
inflation _{$t-1$}	0,161860	0,0869758	1,8610	0,0645
inflation _{$t-2$}	0,167678	0,0959219	1,7481	0,0823
inflation _{$t-3$}	-0,0807241	0,0982396	-0,8217	0,4124
inflation _{$t-4$}	-0,0977973	0,0891895	-1,0965	0,2745
fedfunds _{$t-1$}	0,684688	0,0849388	8,0610	0,0000
fedfunds _{$t-2$}	-0,0616138	0,0940458	-0,6551	0,5133
fedfunds _{$t-3$}	0,456272	0,0987341	4,6212	0,0000
fedfunds _{$t-4$}	-0,166401	0,0900302	-1,8483	0,0664
mgrow _{$t-1$}	-0,0148429	0,0369171	-0,4021	0,6882
mgrow _{$t-2$}	0,0319402	0,0388752	0,8216	0,4125
mgrow _{$t-3$}	0,0260492	0,0376517	0,6918	0,4900
mgrow _{$t-4$}	-0,0507698	0,0350134	-1,4500	0,1490

$$R^2 = 0,883551 \quad \bar{R}^2 \text{ corregido} = 0,872258 \quad F(16, 165) = 78,2452$$

Contrastes F de restricciones cero

Todos los retardos de gdpch	$F(4, 165) = 4,11082$	valor p 0,0033
Todos los retardos de inflation	$F(4, 165) = 3,59141$	valor p 0,0078
Todos los retardos de fedfunds	$F(4, 165) = 131,702$	valor p 0,0000
Todos los retardos de mgrow	$F(4, 165) = 0,70968$	valor p 0,5864
Todas las variables, retardo 4	$F(4, 165) = 1,83359$	valor p 0,1247

Ecuación 4: mgrow

Variable	Coefficiente	Desv. típica	Estadístico t	valor p
const	1,50897	0,748718	2,0154	0,0455
gdpch _{$t-1$}	-0,112665	0,0729090	-1,5453	0,1242
gdpch _{$t-2$}	0,0626401	0,0746020	0,8397	0,4023
gdpch _{$t-3$}	0,111972	0,0686707	1,6306	0,1049
gdpch _{$t-4$}	-0,0395615	0,0641491	-0,6167	0,5383
inflation _{$t-1$}	-0,230408	0,195316	-1,1797	0,2398
inflation _{$t-2$}	0,399855	0,215406	1,8563	0,0652
inflation _{$t-3$}	-0,0539127	0,220610	-0,2444	0,8072
inflation _{$t-4$}	0,0894866	0,200287	0,4468	0,6556
fedfunds _{$t-1$}	-0,459479	0,190742	-2,4089	0,0171
fedfunds _{$t-2$}	0,341644	0,211193	1,6177	0,1076
fedfunds _{$t-3$}	-0,111895	0,221721	-0,5047	0,6145
fedfunds _{$t-4$}	0,250718	0,202175	1,2401	0,2167
mgrow _{$t-1$}	0,410030	0,0829024	4,9459	0,0000
mgrow _{$t-2$}	0,0872904	0,0872996	0,9999	0,3188
mgrow _{$t-3$}	0,149464	0,0845520	1,7677	0,0790
mgrow _{$t-4$}	-0,00831347	0,0786274	-0,1057	0,9159

$$R^2 = 0,460204 \quad \bar{R}^2 \text{ corregido} = 0,407860 \quad F(16, 165) = 8,79193$$

Contrastes F de restricciones cero

Todos los retardos de gdpch	$F(4, 165) = 1,30819$	valor p 0,2691
Todos los retardos de inflation	$F(4, 165) = 1,3427$	valor p 0,2563
Todos los retardos de fedfunds	$F(4, 165) = 1,64895$	valor p 0,1644
Todos los retardos de mgrow	$F(4, 165) = 15,4099$	valor p 0,0000
Todas las variables, retardo 4	$F(4, 165) = 0,70923$	valor p 0,5867

Sistema VAR, orden del retardo 4
observaciones 1960:2–2005:3 ($T = 182$)

Log-verosimilitud = -1397,47

Determinante de la matriz de covarianzas = 54,8884

AIC = 16,1041

BIC = 17,3012

HQC = 16,5894

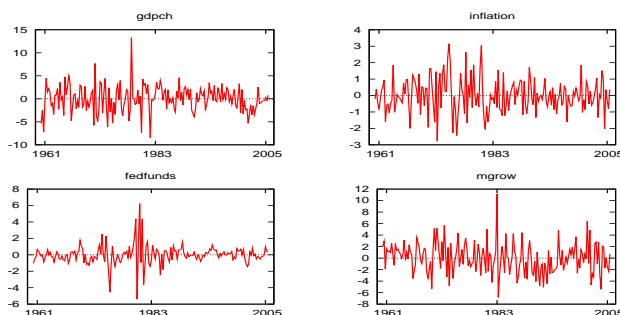
Para el sistema en conjunto —

Hipótesis nula: el retardo más largo es 3

Hipótesis alternativa: el retardo más largo es 4

Contraste de razón de verosimilitudes: $\chi_{16}^2 = 29,251$ (valor p 0,022295)

Figura 3.5: Residuos de cada ecuación del VAR(4)



Podemos ver los gráficos de los residuos de cada ecuación. En esa misma ventana con *Gráficos* → *Graficos de los residuos*.

También podemos guardar los residuos de cada ecuación e incorporarlos a la base de datos mediante *Guardar* → *Residuos de la ecuación*.

Así mismo se puede obtener la estimación de la matriz de varianzas y covarianzas de las perturbaciones²:

$$\tilde{\Sigma}_u = (1/T) \sum_{t=1}^T \hat{u}_t \hat{u}_t'$$

y el logaritmo de su determinante $\ln|\tilde{\Sigma}_u|$

Análisis → *Matriz de covarianzas cruzada entre ecuaciones*

Matriz de covarianzas cruzada residual (correlaciones por encima de la diagonal principal)

7,7983	(-0,099)	(0,187)	(0,069)
-0,28117	1,0311	(0,218)	(-0,041)
0,59733	0,25399	1,3153	(-0,332)
0,49691	-0,10606	-0,97948	6,6331

logaritmo del determinante = 4,0053

²En el resultado se muestran las estimaciones de las varianzas en la diagonal principal y las covarianzas por debajo de ésta. Los valores por encima de la diagonal principal corresponden a las correlaciones, obtenidas una vez se dividida cada covarianza por el producto de las desviaciones típicas correspondientes a cada par de variables

3.2. Análisis Impulso-Respuesta

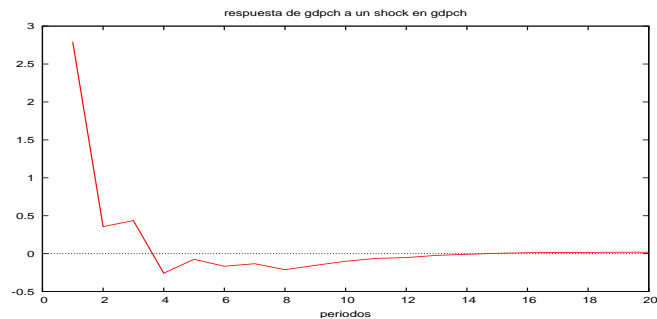
En la misma ventana de resultados de estimación del VAR(4) podemos elegir:

Análisis → Respuestas al impulso

Se muestran las respuestas a lo largo del periodo establecido (en este caso 20 periodos) de cada una de las variables a un shock de tamaño una desviación típica en una variable en concreto del vector. Por ejemplo para un shock en gdpch de tamaño igual a la raíz cuadrada del elemento (1,1) de la matriz de covarianzas cruzada residual se obtiene:

Respuestas en todas las variables del sistema a un shock de tamaño una desviación típica en gdpch

periodo	gdpch	inflation	fedfunds	mgrow
1	2,79255	-0,100686	0,213901	0,177939
2	0,356524	-0,0556762	0,430687	-0,316747
3	0,435514	-0,0360084	0,483843	-0,131831
4	-0,257615	0,0454291	0,496200	0,123198
5	-0,0740569	0,134869	0,510624	-0,112195
6	-0,166243	0,0680873	0,510496	-0,0526597
7	-0,132180	0,0511093	0,485318	-0,00561626
8	-0,210757	0,0312931	0,414946	0,0126127
9	-0,154178	0,0268645	0,366533	0,0651959
10	-0,100185	-0,00117601	0,328224	0,0800721
11	-0,0621131	-0,0196054	0,283630	0,0872735
12	-0,0514274	-0,0376900	0,247576	0,0944442
13	-0,0227227	-0,0460012	0,215871	0,0961805
14	-0,00838101	-0,0558498	0,187003	0,0888364
15	0,00527650	-0,0638128	0,164536	0,0804193
16	0,0123073	-0,0706277	0,144456	0,0703728
17	0,0153921	-0,0743434	0,125223	0,0599459
18	0,0170083	-0,0779615	0,108869	0,0495831
19	0,0195340	-0,0808159	0,0942096	0,0398109
20	0,0193197	-0,0832247	0,0802208	0,0303757

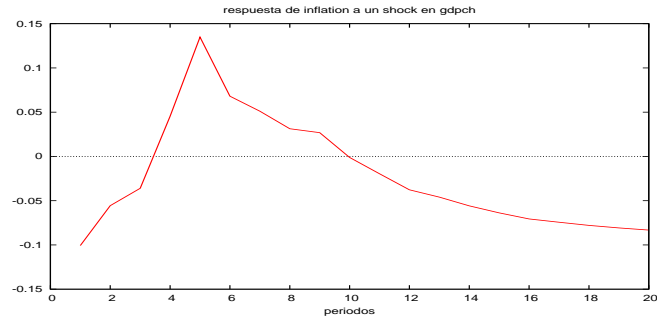


El gráfico con la **respuesta de gdpch a a un shock de tamaño una desviación típica en gdpch** se obtiene de:

Gráficos → *Respuesta de gdpch* → *a gdpch*

El gráfico con la **respuesta de inflation a un shock de tamaño una desviación típica en gdpch** se obtiene de:

Gráficos → *Respuesta de inflation* → *a gdpch*



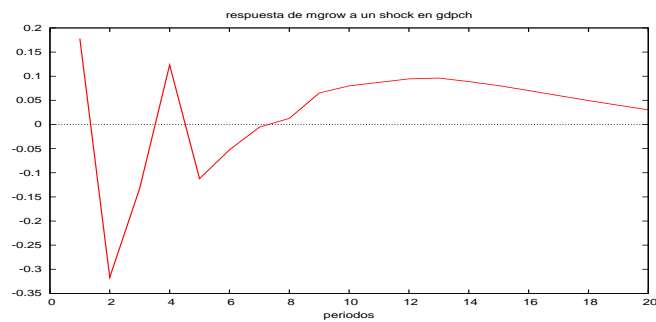
El gráfico con la **respuesta de fedfunds a un shock de tamaño una desviación típica en gdpch** se obtiene de:

Gráficos → *Respuesta de fedfunds* → *a gdpch*



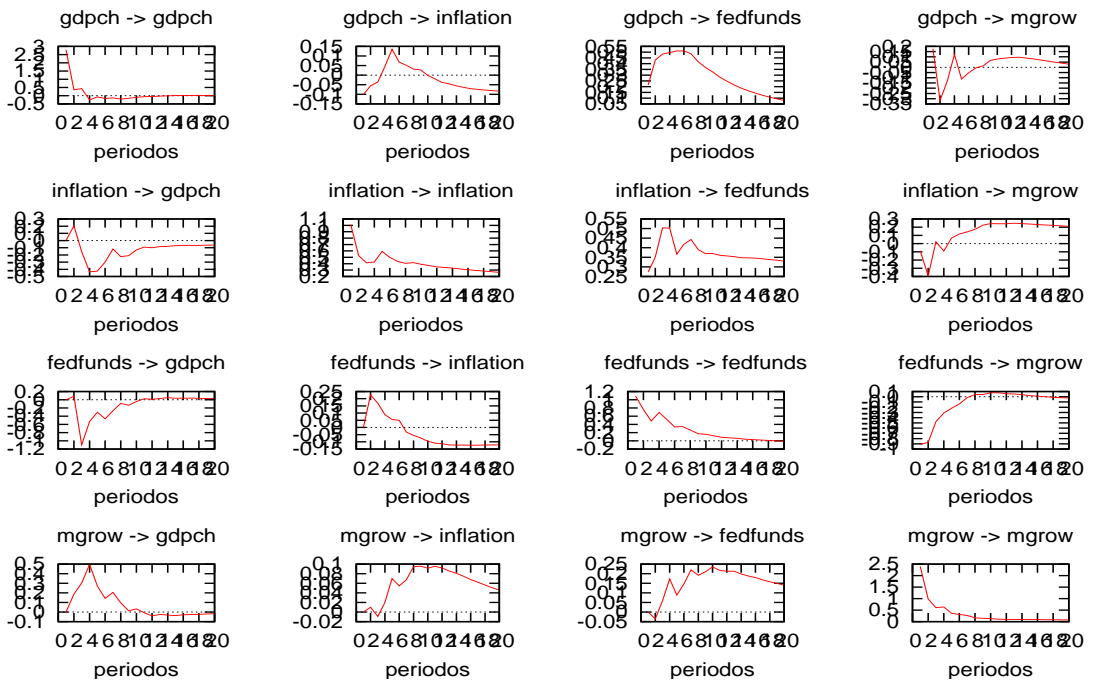
El gráfico con la **respuesta de mgrow a un shock de tamaño una desviación típica en gdpch** se obtiene de:

Gráficos → *Respuesta de mgrowth* → *a gdpch*



También podemos obtener todos los gráficos de las funciones impulso-respuesta a la vez *Gráficos* → *Respuestas al impulso (combinada)*

Figura 3.6: Respuestas al Impulso VAR(4)



3.3. Seleccionando el orden del retardo

En la ventana principal de gretl, elegir

Modelo → *series temporales* → *Selección del orden del VAR*

Elegimos orden máximo igual a 4 y obtenemos:

Sistema VAR, máximo orden de retardos 4

retardos	log.veros	p(RV)	AIC	BIC	HQC
1	-1477,18918		16,452628	16,804717*	16,595360
2	-1438,48708	0,00000	16,203155	16,836914	16,460072
3	-1412,09544	0,00001	16,088961*	17,004391	16,460063*
4	-1397,46977	0,02230	16,104063	17,301165	16,589351

Los asteriscos indican el valor más pequeño del respectivo criterio de información, AIC = Akaike, BIC = Schwarz bayesiano y por último HQC = Hannan-Quinn.

Además en la primera columna se ofrece el logaritmo de la verosimilitud evaluada en las estimaciones obtenidas con cada uno de los retardos³. En la segunda se muestra el valor p asociado al estadístico de ratio de verosimilitudes para contrastar la H_0 : m-1 retardos versus H_a : m retardos, comenzando con m=4.

En los resultados de la **estimación del VAR(4)** también se ofrecían los siguientes resultados que coinciden con los de la línea para 4 retardos: Log-verosimilitud = -1397,47

$$AIC = 16,1041 \quad BIC = 17,3012 \quad HQC = 16,5894$$

Contraste de razón de verosimilitudes: H_0 : 3 retardos versus H_a : 4 retardos:

$$\chi_{16}^2 = 29,251 \quad (\text{valor p } 0,022295)$$

En este caso además del valor p del contraste de razón de verosimilitudes que coincide con el mostrado arriba $p(RV)=0,02230$, se muestra el valor del estadístico que bajo H_0 sigue una distribución asintótica χ_{16}^2 .

Ahora, si hubiéramos elegido un orden de retardos máximo igual a 8, los valores mostrados para 4,3,2,1 no coinciden con los anteriores, ni tampoco con los que se obtendrían de la estimación de cada VAR(p) para cada p. ¿Porqué ocurre esto?

³El número de observaciones en todos los casos es igual al total disponible menos el número de condiciones iniciales para estimar el VAR de mayor orden, en este caso 4

Sistema VAR, máximo orden de retardos 8

retardos	log.veros	p(RV)	AIC	BIC	HQC
1	-1444,04566		16,449951	16,807455*	16,594929
2	-1406,58756	0,00000	16,208849	16,852356	16,469808
3	-1380,01514	0,00001	16,090058	17,019568	16,466999*
4	-1365,37147	0,02207	16,105297	17,320810	16,598220
5	-1347,21398	0,00262	16,081056	17,582572	16,689961
6	-1330,53891	0,00664	16,073471*	17,860990	16,798358
7	-1321,21255	0,28708	16,148456	18,221978	16,989324
8	-1304,99319	0,00876	16,145991	18,505516	17,102842

Esto ocurre porque cada VAR en esa secuencia se estima con el mismo número de observaciones dejando tantas iniciales como se necesitan para estimar el VAR de mayor orden. Por lo tanto, en este caso se está estimando cada VAR con un número menor de observaciones que antes.

Selección:

Comenzando tanto con un orden de retardos máximo 4 como con 8, el AIC elige como orden del VAR un mayor número de retardos (en el primer caso 3 y en el segundo 6) y en cambio el BIC elige el menor número de retardos (en ambos casos 1 retardo). En ambos casos el HQC elige 3 retardos.

Comenzando con un orden máximo $M=4$, el contraste secuencial utilizando el ratio de verosimilitudes elige 4 retardos, ya que el valor del estadístico para contrastar H_0 : 3 retardos versus H_a : 4 retardos, $\chi^2_{16} = 29,251$ es mayor que el valor crítico⁴ al 0,05 de significación. Equivalentemente mirando el valor $p = 0,022295$ como es menor que el nivel de significación 0,05 se rechaza la hipótesis nula.

Por lo tanto parece que un orden de cuatro retardos puede ser adecuado⁵.

⁴Utilizando *Herramientas* → *Tablas estadísticas* → *chi-cuadrado con 16 gl* se obtienen los valores críticos de la distribución Chi-cuadrado para distintos niveles de significación. Las cabeceras de columna muestran el nivel de significación para un contraste a una

cola.	0,10	0,05	0,025	0,01	0,001
gl = 16	23,542	26,296	28,845	32,000	39,252

⁵Si comenzamos con un orden máximo de 8, el contraste secuencial de RV elegiría 6 retardos ya que se no se rechaza la hipótesis nula hasta llegar a la de 5 retardos. Ahora bien, el nivel de significación no sería del 0,05, sino del $1 - (1 - 0,5)^4 = 0,9375$

3.4. Otros contrastes de especificación

En la misma ventana de estimación del VAR, en este caso seguimos con 4 retardos, podemos realizar contrastes sobre las hipótesis de ruido blanco y Normalidad además de efectos ARCH en el vector u_t de perturbaciones del VAR.

a) Contrastes \rightarrow Autocorrelación

Equation 1: Contraste LM de autocorrelación hasta el orden 8 -
 Hipótesis nula: no hay autocorrelación
 Estadístico de contraste: LMF = 1,18259
 con valor p = $P(F(8,149) > 1,18259) = 0,31322$

Equation 2: Contraste LM de autocorrelación hasta el orden 8 -
 Hipótesis nula: no hay autocorrelación
 Estadístico de contraste: LMF = 1,32283
 con valor p = $P(F(8,149) > 1,32283) = 0,236306$

Equation 3: Contraste LM de autocorrelación hasta el orden 8 -
 Hipótesis nula: no hay autocorrelación
 Estadístico de contraste: LMF = 2,28298
 con valor p = $P(F(8,149) > 2,28298) = 0,0246923$

Equation 4: Contraste LM de autocorrelación hasta el orden 8 -
 Hipótesis nula: no hay autocorrelación
 Estadístico de contraste: LMF = 2,13366
 con valor p = $P(F(8,149) > 2,13366) = 0,0359909$

Se muestra un contraste LM en cada una de las ecuaciones de la hipótesis nula de no autocorrelación versus una alternativa general de autocorrelación hasta de orden 8, que es el retardo máximo que hemos elegido. Estos son contrastes univariantes en cada uno de los residuos de cada ecuación. Parece que hay evidencia⁶ de autocorrelación en los residuos de las ecuaciones del VAR para fedfunds y para mgrowth (tercera y cuarta ecuaciones). Esto puede indicar que igual tendríamos que haber elegido un orden mayor del VAR para que todos ellos fueran ruido blanco, o bien que haya cierta inestabilidad en los coeficientes (cambio estructural) de esas ecuaciones que no se haya recogido y sea la causa de detectar autocorrelación.

⁶Probando con menos retardos esta evidencia es menor para la tercera ecuación.

Si queremos hacer el contraste multivariante considerando todo el vector de residuos a la vez con el contraste de Portmanteau Multivariante, tenemos que hacerlo de otra forma. Aquí no tenemos esa opción.

b) **Contrastes** → **Normalidad**

Aparece una ventana con los siguientes resultados:

Residual correlation matrix, C

1,0000	-0,099154	0,18651	0,069090
-0,099154	1,0000	0,21809	-0,040555
0,18651	0,21809	1,0000	-0,33160
0,069090	-0,040555	-0,33160	1,0000

Eigenvalues of the correlation matrix:

0,525501
 0,947579
 1,10472
 1,4222

Test for multivariate normality of residuals Doornik-Hansen
 Chi-square(8) = 246,488, with p-value = 9,56215e-049

Gretl ofrece un contraste multivariante de la hipótesis nula de Normalidad que se debe a Doornik y Hansen. En este caso se rechaza de forma contundente.

En principio, la no-normalidad no afecta a muchos de los resultados obtenidos, en cuanto a consistencia y normalidad asintótica de los estimadores del VAR siempre y cuando se siga satisfaciendo que u_t sea un *standard white noise*⁷.

c) **Contrastes** → **ARCH**

Una vez se elige el orden de retardos que se considera bajo la hipótesis alternativa sobre el proceso ARCH, aparece una ventana con los siguientes resultados:

⁷El supuesto de normalidad garantiza estas condiciones pero no es una condición necesaria para que se satisfagan.

Equation 1: Contraste de ARCH de orden 8 -
 Hipótesis nula: no hay efecto ARCH
 Estadístico de contraste: $TR^2 = 7,64017$
 con valor p = $P(\text{Chi-Square}(8) > 7,64017) = 0,469385$

Equation 2: Contraste de ARCH de orden 8 -
 Hipótesis nula: no hay efecto ARCH
 Estadístico de contraste: $TR^2 = 16,2133$
 con valor p = $P(\text{Chi-Square}(8) > 16,2133) = 0,0394264$

Equation 3: Contraste de ARCH de orden 8 -
 Hipótesis nula: no hay efecto ARCH
 Estadístico de contraste: $TR^2 = 65,8643$
 con valor p = $P(\text{Chi-Square}(8) > 65,8643) = 3,25518e-011$

Equation 4: Contraste de ARCH de orden 8 -
 Hipótesis nula: no hay efecto ARCH
 Estadístico de contraste: $TR^2 = 4,18429$
 con valor p = $P(\text{Chi-Square}(8) > 4,18429) = 0,840125$

De nuevo Gretl ofrece contrastes univariantes en los residuos de cada ecuación para detectar efectos ARCH. Parece haber evidencia de estos efectos en los términos de error de la segunda y tercera ecuación del VAR, ya que los valores p son menores que 0,05, especialmente en la ecuación 3 asociada a fedfunds.

d) **Contrastes de estabilidad en los parámetros de cada ecuación del VAR.**

Consideramos estimar por MCO cada ecuación con

Modelo → *Mínimos Cuadrados Ordinarios*

y en cada caso especificamos una de las ecuaciones del VAR, por ejemplo especificamos⁸ la primera ecuación del VAR(4) para *gdpch* en la ventana que aparece:

Variable dependiente → *Elegir gdpch*

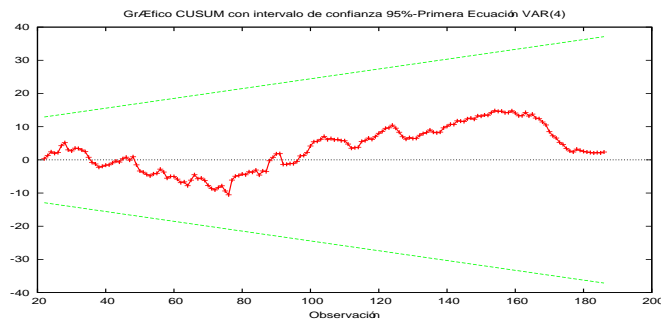
Variables independientes → *Añadir* todos los retardos de *gdpch* y del resto de variables, además de la constante que por defecto se incluye.

⁸Primeramente hemos de definir como nuevas variables los retardos hasta de orden cuatro de todas las variables implicadas. Esto se puede hacer con *Añadir* → *definir nueva variable*, y en la ventana que surge escribir por ejemplo $gdpch1 = gdpch(-1)$ para definir el primer retardo de *gdpch*.

En la ventana donde aparecen los resultados de la estimación (que son los mismos que los obtenidos anteriormente) tenemos una serie de opciones. En especial en **Contrastes** tenemos, aparte de otros, varios contrastes de estabilidad paramétrica, por ejemplo el contraste de Chow, el contraste CUSUM. Veamos este últimos aplicado a cada ecuación del VAR.

1. **Contraste CUSUM - Primera ecuación (gdpch):**

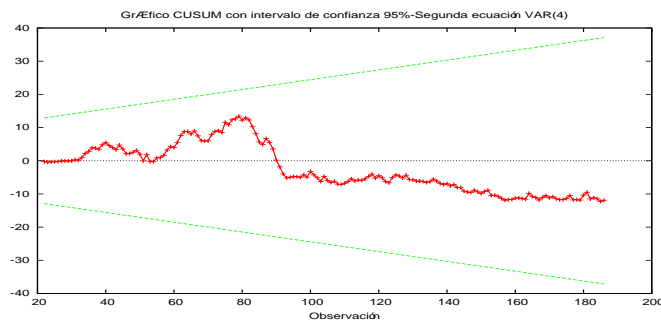
Harvey-Collier $t(164) = 0,184001$ con valor p 0,8542 (No se rechaza la hipótesis nula de estabilidad)



En el gráfico no se sale fuera de las bandas de confianza la suma acumulada de los residuos escalados en ningún punto. Luego no hay evidencia de inestabilidad paramétrica en esta ecuación.

2. **Contraste CUSUM - Segunda ecuación (inflation):**

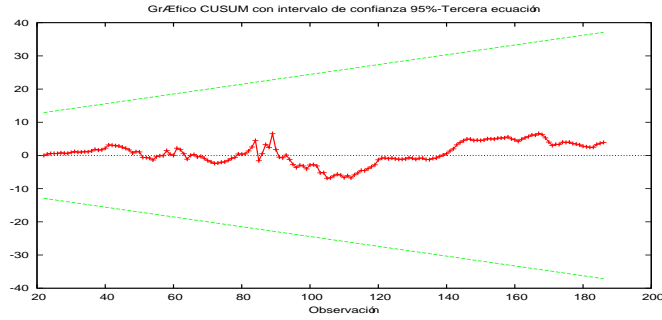
Harvey-Collier $t(164) = -0,926527$ con valor p 0,3555



No hay evidencia de inestabilidad paramétrica en los parámetros de esta ecuación.

3. Contraste CUSUM - Tercera ecuación (fedfunds):

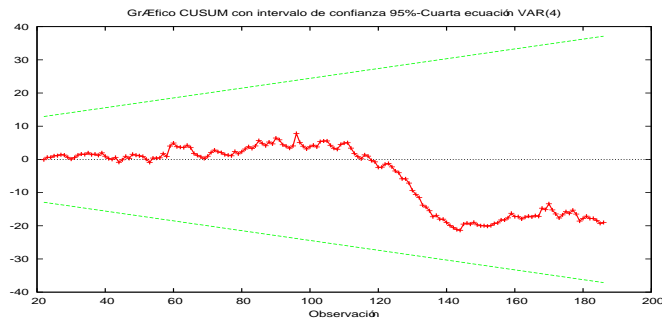
Harvey-Collier $t(164) = 0,306201$ con valor p 0,7598



Tampoco hay evidencia de inestabilidad paramétrica en los parámetros de esta ecuación.

4. Contraste CUSUM - Cuarta ecuación (mgrow):

Harvey-Collier $t(164) = -1,47974$ con valor p 0,1409



Tampoco hay evidencia de inestabilidad paramétrica en los parámetros de esta ecuación.

3.5. Causalidad en el sentido de Granger

El sistema está compuesto por cuatro variables, por tanto, contrastar que **un subconjunto de variables** Z_{1t} del vector y_t **no causa en el sentido de Granger a otro subconjunto de variables** Z_{2t} , implica en general realizar un contraste **conjunto** de que todos los retardos del subvector Z_{2t} no entran en las ecuaciones del VAR para Z_{1t} . Son restricciones de exclusión que puede afectar a la vez a distintas ecuaciones del sistema. Cuando estimamos el VAR mediante

Modelo → *Series temporales* → *Autorregresion vectorial (VAR)*

en la ventana asociada a los resultados no tenemos la opción de hacer estos contrastes de omisión de retardos. En los resultados mostrados para cada ecuación se obtenían contrastes F de restricciones cero. Por ejemplo en el caso de la primera ecuación, para *gdpch*:

Contrastes F de restricciones cero

Todos los retardos de inflation	$F(4, 165) = 1, 3427$	valor p 0,2563
Todos los retardos de fedfunds	$F(4, 165) = 1, 64895$	valor p 0,1644
Todos los retardos de mgrow	$F(4, 165) = 15, 4099$	valor p 0,0000

Estos contrastes no son válidos para contrastar que la inflación no cause en el sentido de Granger a *gdpch*, o que *fedfunds* no cause en el sentido de Granger a *gdpch*, o que *mgrow* no cause en el sentido de Granger a *gdpch*⁹. Tendría que realizarse **conjuntamente** el contraste de que todos los coeficientes de todos los retardos de *inflation*, *fedfunds* y *mgrow* en esa ecuación son cero. Esa hipótesis nula implicaría que la **inflación, fedfunds y mgrow no causan en el sentido de Granger a gdpch**. Una forma de llevar a a cabo este contraste es el siguiente:

Consideramos estimar por MCO cada ecuación con

Modelo → *Mínimos Cuadrados Ordinarios*

y en cada caso especificamos una de las ecuaciones del VAR, por ejemplo¹⁰ la primera ecuación del VAR(4) para *gdpch* en la ventana que aparece:

Variable dependiente → *Elegir gdpch*

⁹Esto sería válido si el sistema fuera bivalente y solamente hubiera en el vector y_t por ejemplo las variables *gdpch* e *inflación*.

¹⁰Primeramente hemos de definir como nuevas variables los retardos hasta de orden cuatro de todas las variables implicadas. Esto se puede hacer con *Añadir* → *definir nueva variable*, y en la ventana que surge escribir por ejemplo $gdpch1 = gdpch(-1)$ para definir el primer retardo de *gdpch*.

Variables independientes → Añadir todos los retardos de gdpch y del resto de variables, además de la constante que por defecto se incluye.

En la ventana correspondiente en cada caso donde aparecen los resultados de la estimación (que son los mismos que los obtenidos anteriormente) en

Contrastes → omitir variables

y en la ventana que aparece *gretl: contrastes del modelo* elegir omitir todos los retardos del resto de variables del sistema, en este caso los cuatro retardos de inflation, fedfunds y mgrow.

Aparece otra ventana con los resultados de estimar el modelo restringido bajo la hipótesis nula. Entre otros resultados aparece

Hipótesis nula: los parámetros de regresión son cero para las variables inflation_1, inflation_2, inflation_3, inflation_4
fedfunds_1, fedfunds_2, fedfunds_3, fedfunds_4
mgrow_1, mgrow_2, mgrow_3, mgrow_4

Estadístico de contraste:

$F(12, 165) = 4,45282$, con valor $p = 3,59731e-006$

Dado el valor del estadístico y el valor p se rechaza esa hipótesis nula.

Así podríamos proceder en cada ecuación del sistema. Por ejemplo **para la ecuación de inflation, contrastar la hipótesis nula de que el gdpch, el fedfunds y mgrow no causan en el sentido de Granger a inflation:**

En la ventana correspondiente a la estimación por MCO de la segunda ecuación del VAR, utilizando

Contrastes → omitir variables

y en la ventana que aparece *gretl: contrastes del modelo* elegir omitir todos los retardos del resto de variables del sistema, en este caso los cuatro retardos de gdpch, fedfunds y mgrow, se obtiene:

Hipótesis nula: los parámetros de regresión son cero para las variables gdpch_1, gdpch_2, gdpch_3, gdpch_4, fedfunds_1, fedfunds_2, fedfunds_3, fedfunds_4, mgrow_1, mgrow_2, mgrow_3, mgrow_4

Estadístico de contraste:

$F(12, 165) = 1,81175$, con valor $p = 0,0499213$

Dado el valor del estadístico y el valor p, se rechaza la hipótesis nula al 0,05 pero no al 0,01 de significación. Parece que en este caso las variables gdpch, fedfunds y mgrow no ayudan a predecir la inflación.

Para la ecuación de fedfunds, contrastar la hipótesis nula de que el gdpch, inflation y mgrow no causan en el sentido de Granger a fedfunds:

En la ventana correspondiente a la estimación por MCO de la tercera ecuación del VAR, utilizando

Contrastes → omitir variables

y en la ventana que aparece *gretl: contrastes del modelo* elegir omitir todos los retardos del resto de variables del sistema, en este caso los cuatro retardos de gdpch, inflation y mgrow, se obtiene:

Hipótesis nula: los parámetros de regresión son cero para las variables gdpch_1, gdpch_2, gdpch_3, gdpch_4, inflation_1, inflation_2, inflation_3, inflation_4, mgrow_1, mgrow_2, mgrow_3, mgrow_4

Estadístico de contraste:

$F(12, 165) = 2,77808$, con valor p = 0,00179576

Se rechaza la hipótesis nula al 0,05 y 0,01 de significación.

Para la ecuación de mgrow, contrastar la hipótesis nula de que el gdpch, inflation y fedfunds no causan en el sentido de Granger a mgrow:

En la ventana correspondiente a la estimación por MCO de la cuarta ecuación del VAR, utilizando

Contrastes → omitir variables

y en la ventana que aparece *gretl: contrastes del modelo* elegir omitir todos los retardos del resto de variables del sistema, en este caso los cuatro retardos de gdpch, inflation y fedfunds, se obtiene:

Hipótesis nula: los parámetros de regresión son cero para las variables gdpch_1, gdpch_2, gdpch_3, gdpch_4, inflation_1, inflation_2, inflation_3, inflation_4, fedfunds_1, fedfunds_2, fedfunds_3, fedfunds_4

Estadístico de contraste:

$F(12, 165) = 1,88122$, con valor p = 0,0399646

Dado el valor del estadístico y el valor p, se rechaza la hipótesis nula al 0,05 pero no al 0,01 de significación. Parece que en este caso las variables *gdpch*, *fedfunds* y la inflación no ayudan a predecir *mgrow*.

3.6. Predicción

Vamos a considerar el periodo de estimación de 1959:01 a 1999:4, dejando el resto como periodo de predicción de 2000:1 a 2005:3.

- Primero establecemos el rango muestral para la estimación en

Muestra → *Establecer rango* → *Inicio: 1959:1 Fin: 1999:4*

- Estimamos el VAR de orden 4 con *Modelo* → *Series temporales* → *Autorregresion vectorial (VAR)*
- En la ventana asociada a los resultados de estimación del VAR(4) elegimos *Análisis* → *Predicciones* y elegimos la variable de interés a predecir, por ejemplo *gdpch*. Aparece una ventana donde se muestra el dominio de predicción de 2000:1 a 2005:3 y se tienen varias opciones:

1. Predicción automática (dinámica fuera de la muestra)

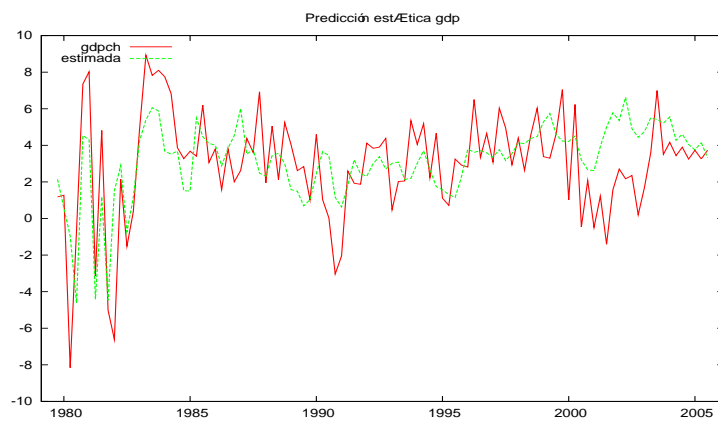
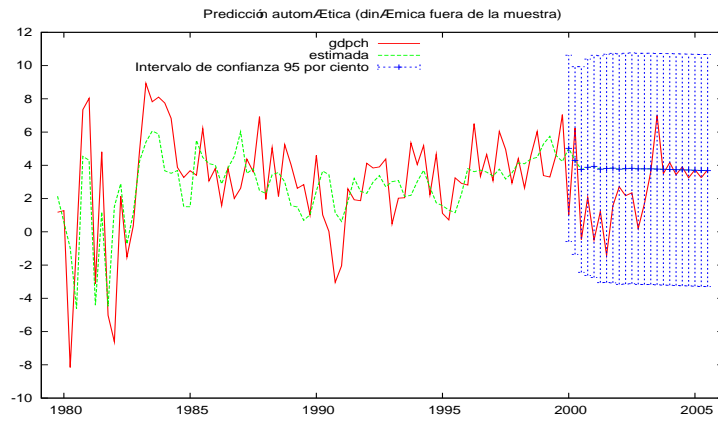
Para intervalos de confianza 95 %, $t(142, .025) = 1,977$

Observaciones	<i>gdpch</i>	predicción	desv. típica	Intervalo de confianza	
				low	hi
2000:1	1,012047	5,018080	2,835507	-0,587182	10
2000:2	6,234424	4,287793	2,856317	-1,358606	9
2000:3	-0,459245	3,747259	3,128152	-2,436505	9
2000:4	2,072575	3,881234	3,291679	-2,625793	10
2001:1	-0,489797	3,935454	3,380310	-2,746781	10
2001:2	1,225388	3,769156	3,451785	-3,054370	10
2001:3	-1,407697	3,814336	3,473223	-3,051568	10
2001:4	1,573221	3,830084	3,490804	-3,070576	10
2002:1	2,707266	3,770231	3,505922	-3,160314	10
2002:2	2,171039	3,799752	3,510221	-3,139292	10
2002:3	2,349639	3,819696	3,512363	-3,123581	10
2002:4	0,202115	3,787283	3,515000	-3,161208	10
2003:1	1,692170	3,785752	3,516583	-3,165867	10
2003:2	3,605503	3,788316	3,518359	-3,166815	10
2003:3	6,995862	3,771731	3,519971	-3,186585	10

2003:4	3,507805	3,761080	3,521312	-3,199888	10,722048
2004:1	4,163882	3,753004	3,522660	-3,210628	10,716636
2004:2	3,437718	3,738976	3,523994	-3,227294	10,705245
2004:3	3,897209	3,727173	3,525198	-3,241477	10,695824
2004:4	3,250732	3,716688	3,526364	-3,254267	10,687644
2005:1	3,733974	3,704326	3,527489	-3,268854	10,677506
2005:2	3,256011	3,692765	3,528556	-3,282524	10,668053
2005:3	3,733916	3,682188	3,529578	-3,295120	10,659497

2. Predicción estática.

Observaciones	gdpch	predicción
2000:1	1,012047	4,214531
2000:2	6,234424	4,506009
2000:3	-0,459245	3,199512
2000:4	2,072575	2,662188
2001:1	-0,489797	2,618423
2001:2	1,225388	3,937957
2001:3	-1,407697	5,004939
2001:4	1,573221	5,784514
2002:1	2,707266	5,360317
2002:2	2,171039	6,622375
2002:3	2,349639	4,923816
2002:4	0,202115	4,439035
2003:1	1,692170	4,744298
2003:2	3,605503	5,509609
2003:3	6,995862	5,415142
2003:4	3,507805	5,213690
2004:1	4,163882	5,556802
2004:2	3,437718	4,281259
2004:3	3,897209	4,613377
2004:4	3,250732	4,045614
2005:1	3,733974	3,760423
2005:2	3,256011	4,142415
2005:3	3,733916	3,327776



Capítulo 4

Procesos integrados y Cointegración

4.1. Difference-stationary versus Trend-stationary

En el capítulo anterior hemos visto que un **proceso VAR(p)** es estable y por lo tanto **estacionario** si

$$\det(I_k - A_1z - \dots - A_pz^p) \neq 0 \text{ para } |z| \leq 1$$

Proceso AR(1) univariante estacionario $y_t = \nu + \alpha y_{t-1} + u_t$ donde u_t es un ruido blanco, esto **significa que**

$$(1 - \alpha z) \neq 0 \text{ para } |z| \leq 1$$

o equivalentemente $|\alpha| < 1$, dado que la raíz del polinomio $(1 - \alpha z)$ es $1/\alpha$. Comenzando en un valor inicial y_0 podemos expresar

$$y_t = \nu \sum_{j=0}^{t-1} \alpha^j + \alpha^t y_0 + \sum_{j=0}^{t-1} \alpha^j u_{t-j}$$

Dado que $|\alpha| < 1$ a medida que $t \rightarrow \infty$,

- $\nu \sum_{j=0}^{t-1} \alpha^j \rightarrow (1 - \alpha)^{-1} \nu$
- $\alpha^t \rightarrow 0$ por lo que podemos ignorar el término $\alpha^t y_0$ en el límite. Esto quiere decir que la condición inicial cada vez tiene menos importancia.
- $\sum_{j=0}^{t-1} \alpha^j u_{t-j}$ converge en media cuadrática.

Por lo tanto, $y_t = \mu + \sum_{j=0}^{\infty} \alpha^j u_{t-j}$ tal que:

$$\begin{aligned} E(y_t) &= \mu = (1 - \alpha)^{-1} \nu \\ \text{var}(y_t) &= \text{var}(u_t)(1 - \alpha)^{-1} = \sigma_u^2(1 - \alpha)^{-1} \\ \text{cov}(y_t, y_{t+h}) &= \alpha^h \text{var}(y_t) \\ \text{corr}(y_t, y_{t+h}) &= \alpha^h \end{aligned}$$

Raíz unitaria

¿Qué ocurre si el polinomio $(1 - \alpha z)$ tiene una raíz unitaria, esto es $\alpha = 1$? En este caso tan simple estaríamos ante un paseo aleatorio. Comenzando en un valor inicial y_0 podemos expresar

$$y_t = t\nu + y_0 + \sum_{j=0}^{t-1} u_{t-j}$$

$$y_t = t\nu + y_0 + \sum_{s=1}^t u_s$$

- El valor presente de y_t depende de su valor inicial y esta influencia no desaparece a medida que $t \rightarrow \infty$.
- La deriva o el *drift*, ν , recoge una tendencia lineal determinista en el proceso.
- $\sum_{s=1}^t u_s$ no converge en media cuadrática a medida que $t \rightarrow \infty$.

y_t es la suma acumulada de shocks o innovaciones pasadas y presente cuyo efecto no desaparece a medida que nos alejamos en el futuro. Esto recoge lo que se llama una tendencia estocástica.

En este caso

$$\begin{aligned} E(y_t) &= t\nu + E(y_0) \\ \text{var}(y_t) &= t\text{var}(u_t) = t\sigma_u^2 \\ \text{cov}(y_t, y_{t+h}) &= t\text{var}(u_t) = t\sigma_u^2 \\ \text{corr}(y_t, y_{t+h}) &= \frac{t}{(t^2+th)^{1/2}} \end{aligned}$$

A medida que $t \rightarrow \infty$ la media tiende a infinito (si $\nu \neq 0$), la varianza y covarianza tienden a infinito y la correlación tiende a uno.

Una serie **estacionaria** tiende a retornar a su valor medio y fluctuará alrededor de éste dentro de un rango más o menos constante (esto es, tiene una varianza finita).

En el caso de que la serie presente una tendencia lineal determinista pero alrededor de ésta fluctúa dentro de un rango más o menos constante, esta serie se dice que es **trend-stationary**, esto es estacionaria alrededor de una tendencia determinista. Por ejemplo

$$y_t = \omega + \nu t + u_t \text{ donde } u_t \sim iid(0, 1)$$

En este caso y_t presenta una tendencia determinista $\omega + \nu t$ pero dado que u_t es estacionario, las desviaciones alrededor de esa tendencia determinista son estacionarias.

En cambio, el **paseo aleatorio**

$$y_t = y_{t-1} + \nu + u_t \text{ donde } u_t \sim iid(0, 1)$$

comenzando en un valor inicial dado y_0 podemos reescribir la serie como

$$y_t = y_0 + \nu t + \sum_{s=1}^t u_s$$

En este caso y_t no fluctúa dentro de un rango más o menos constante alrededor de una tendencia lineal determinista $y_0 + \nu t$ ya que $\sum_{s=1}^t u_s$ tiene una varianza que crece con t .

Esta serie se convierte en estacionaria después de tomar primeras diferencias.

$$y_t - y_{t-1} = \Delta y_t = \nu + u_t$$

Por ello se le llama **difference-stationary** o estacionaria en diferencias. Esto es así, incluso si $\nu = 0$. En ese caso solamente tenemos la tendencia estocástica y no una tendencia lineal determinista.

Diferencias entre un proceso *trend-stationary* y uno con raíz unitaria en términos de predicción¹.

▪ **Predictor óptimo**

1. Proceso **trend-stationary** como es,

$$y_t = \omega + \nu t + u_t \text{ donde } u_t \sim iid(0, 1)$$

El predictor óptimo en t , s periodos hacia adelante es:

$$E_t(y_{t+s}) = \omega + \nu(t + s)$$

2. **Proceso con raíz unitaria** como por ejemplo el paseo aleatorio con deriva o drift

$$y_t = y_{t-1} + \nu + u_t \text{ donde } u_t \sim iid(0, 1)$$

Dado que $\Delta y_t = \nu + u_t$ es un proceso estacionario y que u_t es un ruido blanco independiente se tiene que:

$$E_t(\Delta y_{t+s}) = \nu$$

El nivel de la serie en $t + s$ es simplemente la suma de los cambios entre t y $t + s$, esto es:

$$y_{t+s} = \Delta y_{t+s} + \Delta y_{t+s-1} + \cdots + \Delta y_{t+1} + y_t$$

por lo tanto

$$E_t(y_{t+s}) = \nu + \nu + \cdots + \nu + y_t = s\nu + y_t$$

Un paseo aleatorio con deriva se espera que crezca a una tasa constante ν por periodo a partir del valor que haya tomado la serie en y_t .

Por lo tanto, el parámetro ν en ambos procesos juega papeles similares ya que con cualquiera de las dos especificaciones la predicción es una función lineal en el horizonte de predicción s con pendiente ν . La **diferencia está en el intercepto**. En el caso del proceso **trend-stationary** es **constante** e igual a ω independientemente del valor que tome y_t mientras que en el caso del **proceso con raíz unitaria cambia con cada nuevo valor de y en el origen de predicción**.

¹Sigue el capítulo 15 del Hamilton

■ **Varianza del error de predicción.**

1. En el ejemplo del proceso **trend-stationary** se tiene que el error de predicción es

$$y_{t+s} - E_t(y_{t+s}) = u_{t+s}$$

por lo que el **error cuadrático medio de la predicción** es

$$E(y_{t+s} - E_t(y_{t+s}))^2 = E(u_{t+s})^2 = \sigma_u^2 = 1$$

constante e independiente del horizonte de predicción.

2. En el **paseo aleatorio**, como ejemplo particular de proceso *difference stationary* se tiene que el error de predicción es,

$$y_{t+s} - E_t(y_{t+s}) = u_{t+s} + u_{t+s-1} + \cdots + u_{t+1}$$

por lo que el error cuadrático medio de la predicción es

$$E(y_{t+s} - E_t(y_{t+s}))^2 = E\left(\sum_{j=1}^s u_{t+j}\right)^2 = s$$

esto es, el **error cuadrático medio crece linealmente con el horizonte de predicción**²

- Efecto de las innovaciones o shocks sobre el nivel de la serie. Para los ejemplos que estamos considerando (que son algo restrictivos pero ilustrativos)
 1. En el proceso **trend-stationary** el efecto sobre y_{t+s} de un aumento en una unidad en u_t permaneciendo constantes el resto de innovaciones es cero.
 2. En el caso del **paseo aleatorio**, con o sin deriva, este efecto es igual al efecto sobre y_t de un aumento en u_t . que es igual a uno. Una innovación o shock en t , u_t , tiene un efecto permanente en el nivel de la serie³.

²Dado que el error cuadrático medio crece linealmente con el horizonte de predicción, **la desviación típica del error de predicción crece con \sqrt{s}** . Por otro lado, si la deriva ν es positiva, entonces la predicción crece linealmente con s . De esta forma, un intervalo del 95 % de confianza para y_{t+s} crece más despacio que el nivel de la serie, lo que implica que **si se observan los datos que genera un proceso con raíz unitaria y con deriva positiva por un periodo suficientemente largo, estos exhibirán una tendencia creciente**. En este sentido, la tendencia determinista que introduce el término de deriva domina asintóticamente la creciente variabilidad generada por el componente de raíz unitaria o tendencia estocástica.

³Notar que:

$$y_{t+s} = \Delta y_{t+s} + \Delta y_{t+s-1} + \cdots + \Delta y_{t+1} + y_t$$

4.2. Algunos contrastes de raíces unitarias

Aquellos procesos univariantes que presentan d raíces unitarias en su parte autorregresiva se conocen con el nombre de *procesos integrados de orden d* , $I(d)$.

Un proceso $I(0)$ es aquel proceso que es estacionario en niveles e $I(1)$ si es estacionario después de tomar primeras diferencias. En general, un proceso $I(d)$ es tal que $(1 - L)^d y_t$ tiene una representación ARMA estacionaria e invertible.

En esta sección vamos a ver algunos contrastes de raíces unitarias a nivel univariante. En particular, nos centraremos en contrastes de procesos $I(1)$ frente a procesos $I(0)$ quizás alrededor de una tendencia determinista.

4.2.1. Contrastes de Dickey-Fuller⁴

Se supone en todos los casos $u_t \sim iid(0, \sigma_u^2)$ y se conoce la condición inicial y_0

- **Caso 1:** Modelo bajo la hipótesis nula es un paseo aleatorio sin deriva y la regresión MCO es:

$$y_t = \alpha y_{t-1} + u_t \Leftrightarrow \Delta y_t = \gamma y_{t-1} + u_t \quad t = 1, \dots, T$$

- **Caso 2:** Modelo bajo la nula es un paseo aleatorio sin deriva y la regresión MCO es:

$$y_t = \nu + \alpha y_{t-1} + u_t \Leftrightarrow \Delta y_t = \nu + \gamma y_{t-1} + u_t \quad t = 1, \dots, T$$

- **Caso 3:** Modelo bajo la nula es un paseo aleatorio con deriva y la regresión MCO es:

$$y_t = \nu + \alpha y_{t-1} + u_t \Leftrightarrow \Delta y_t = \nu + \gamma y_{t-1} + u_t \quad t = 1, \dots, T$$

- **Caso 4:** Modelo bajo la hipótesis nula es un paseo aleatorio con o sin deriva y la regresión es:

$$y_t = \nu + \delta t + \alpha y_{t-1} + u_t \Leftrightarrow \Delta y_t = \nu + \delta t + \gamma y_{t-1} + u_t \quad t = 1, \dots, T$$

⁴Dickey y Fuller (1979, JASA 74; 1981, Econometrica 49).

Hipótesis:

$H_0 : \alpha = 1$, versus $H_1 : |\alpha| < 1$ o $H_0 : \gamma = 0$, versus $H_1 : \gamma < 0$

Estadísticos:

- $T(\hat{\alpha} - 1)$ o equivalentemente $T\hat{\gamma}$
- Estadístico t usual de la regresión MCO

$$(\hat{\alpha} - 1)/\hat{\sigma}_{\hat{\alpha}} \text{ o equivalentemente } \hat{\gamma}/\hat{\sigma}_{\hat{\gamma}}$$

1. Para los casos 1, 2 y 4 los estadísticos no tienen las distribuciones usuales bajo la hipótesis nula como en el caso de regresores fijos o estacionarios porque estos no lo son. Por ejemplo el estadístico t no tiene distribución t-student ni distribución asintótica $N(0,1)$ bajo H_0 sino que dependen de movimientos Brownianos (ver Hamilton (1994, pp. 487-489)).
2. Sus distribuciones no son simétricas sino que son asimétricas hacia la izquierda. Han sido tabulados valores críticos para diferentes tamaños muestrales mediante simulación suponiendo cierta la hipótesis nula y normalidad de u_t .
3. Estas distribuciones no son invariantes a incluir en la regresión un término constante y/o una tendencia lineal determinista. Por eso en cada caso los valores críticos son distintos.
 - Caso 1: Tablas B.5 y B.6 case 1 del Hamilton(1994), pp. 762-763.
 - Caso 2: Tablas B.5 y B.6 case 2 del Hamilton (1994).
 - Caso 4: Tablas B.5 y B.6 case 4 del Hamilton (1994)
4. A diferencia del resto de casos, en el caso 3 el estadístico t bajo la hipótesis nula tiene distribución asintótica $N(0,1)$.

En los casos 2 y 3 podemos estar interesados en contrastar la hipótesis nula conjunta de que $\nu = 0, \alpha = 1$ mediante el estadístico F usual en la regresión anterior. Así como en el caso 3, este estadístico multiplicado por el número de restricciones a contrastar bajo la nula se distribuye asintóticamente con distribución ji-cuadrado con dos grados de libertad, esto no es así en el caso 2. Por otro lado, contrastar la hipótesis nula conjunta $\delta = 0, \alpha = 1$ en el caso 4, requiere utilizar otros valores críticos distintos⁵.

⁵Los valores críticos han sido tabulados mediante simulación para diferentes tamaños muestrales (Caso 2: Tabla B.7 case 2 y Caso 4: Tabla B.7 case 4 del Hamilton (1994)).

Potencia

La potencia de estos contrastes de raíz unitaria disminuye a medida que se incluyen más términos deterministas en la regresión. Esto se debe a que, al añadir en la regresión un término constante y una tendencia determinista, los valores críticos a considerar aumentan en valor absoluto, lo que hace más difícil rechazar la hipótesis nula incluso cuando debiera de ser rechazada.

¿Cuál es el caso correcto a utilizar? La respuesta depende del interés en contrastar la existencia de una raíz unitaria. Si se tiene en mente una hipótesis nula específica sobre el proceso que ha generado los datos, esto nos puede guiar en la elección del contraste más adecuado. Si no, un principio general sería ajustar la especificación que parece la mejor descripción de la serie observada tanto bajo la hipótesis nula como alternativa. Este principio sugiere por tanto el caso 4 para series que exhiben una tendencia muy evidente y el caso 2 para aquellas donde no lo es tanto.

Controlando por autocorrelación: Dickey-Fuller aumentado (ADF):

Hasta ahora se ha supuesto que $u_t \sim iid(0, \sigma_u^2)$. ¿Cómo generalizar los contrastes para tener en cuenta un proceso autorregresivo de mayor orden o componentes de medias móviles?

Una posibilidad son los contrastes de DF aumentado (ADF). Incluir en la regresión más retardos de y_t , (o de Δy_t). Por ejemplo, para el primer caso de antes, ahora la regresión a considerar sería:

$$y_t = \alpha_1 y_{t-1} + \sum_{j=2}^p \alpha_j y_{t-j} + u_t$$

o equivalentemente

$$\Delta y_t = \gamma y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + u_t$$

donde $\gamma = (\alpha_1 + \alpha_2 + \dots + \alpha_p) - 1$ y $\gamma_j = -(\alpha_{j+1} + \alpha_{j+2} + \dots + \alpha_p)$ para $j = 1, 2, \dots, p-1$. La hipótesis nula de una raíz unitaria es

$$H_0 : \gamma = 0 \text{ frente a } H_1 : \gamma < 0$$

Ejemplo:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + u_t$$

El polinomio $(1 - \alpha_1 z - \alpha_2 z^2)$ se puede expresar como

$$(1 - \alpha_1 z - \alpha_2 z^2) = (1 - \lambda_1 z)(1 - \lambda_2 z)$$

donde $\alpha_1 = (\lambda_1 + \lambda_2)$ y $\alpha_2 = -\lambda_1 \lambda_2$. El polinomio tiene una raíz unitaria si $1/\lambda_1 = 1$ o $1/\lambda_2 = 1$. Esto implica en ambos casos que $\alpha_1 + \alpha_2 = 1$. Podemos reparametrizar el modelo de la siguiente forma:

$$y_t - y_{t-1} = \alpha_1 y_{t-1} + \alpha_2 y_{t-1} - y_{t-1} - \alpha_2 y_{t-1} + \alpha_2 y_{t-2} + u_t$$

reorganizando términos

$$y_t - y_{t-1} = (\alpha_1 + \alpha_2 - 1)y_{t-1} - \alpha_2(y_{t-1} - y_{t-2}) + u_t$$

$$\Delta y_t = \gamma y_{t-1} + \gamma_1 \Delta y_{t-1} + u_t$$

donde $\gamma = \alpha_1 + \alpha_2 - 1$ y $\gamma_1 = -\alpha_2$.

El contraste de $H_0 : \gamma = 0$ frente a $H_1 : \gamma < 0$ se llevaría acabo con el estadístico t usual y los valores críticos a considerar serían los mismos que en el caso 1 anterior. También podríamos tener en la regresión intercepto y tendencia determinista como anteriormente.

1. Los coeficientes de regresión que acompañan a Δy_{t-j} tienen distribución asintótica Normal a la tasa usual \sqrt{T} por lo que se pueden utilizar los estadísticos t y F de la forma habitual. No ocurre lo mismo para los estimadores de los parámetros que acompañan a y_{t-1} y los componentes deterministas.

¿Cómo determinar el número de retardos a considerar en la regresión aumentada si el orden del AR no es conocido?

2. Un procedimiento propuesto en la literatura es comenzar con un orden arbitrariamente grande \tilde{p} e ir contrastando la significatividad de γ_j comenzando por $j = \tilde{p} - 1$. Si no se rechaza $H_0 : \gamma_{\tilde{p}-1} = 0$ proceder a contrastar la significatividad conjunta de γ_j para $j = \tilde{p} - 1, \tilde{p} - 2$ y así sucesivamente hasta llegar a rechazar la hipótesis nula $H_0 : \gamma_{\tilde{p}-1} = 0, \gamma_{\tilde{p}-2} = 0, \dots, \gamma_{\tilde{p}-l} = 0$ para algún l . Esa última regresión sería la que consideraríamos.
3. Si el proceso presenta componentes de medias móviles, Said y Dickey (1984, Biometrika 71) sugieren que estos componentes se pueden aproximar bien por un proceso autorregresivo de un orden suficientemente grande tal que es función del tamaño muestral utilizado. La elección de ese número de retardos es algo que puede afectar al tamaño y la potencia del contraste. Una propuesta que parece adecuada es elegir el número de retardos igual a la parte entera de $(12(T/100))^{1/4}$.

4.2.2. Contrastes de Phillips y Perron⁶

Una propuesta alternativa a la de ADF es la que consideran Phillips y Perron. Estos autores consideran una corrección no paramétrica de los estadísticos $T(\hat{\alpha} - 1)$ y $(\hat{\alpha} - 1)/\hat{\sigma}_{\hat{\alpha}}$ de Dickey-Fuller en cada caso 1, 2 y 4, para tener en cuenta que el término de error u_t esté autocorrelacionado y que no sea homocedástico, sin especificar la forma concreta de esta dependencia o heterocedasticidad. Las distribuciones asintóticas de estos estadísticos coinciden con las obtenidas con DF, por lo que los contrastes a una cola se pueden llevar a cabo utilizando las mismas tablas.

Estos estadísticos corregidos serían Z_{α} y Z_t :

$$Z_{\alpha} = T(\hat{\alpha} - 1) - (1/2) \left\{ T^2 \frac{\hat{\sigma}_{\hat{\alpha}}^2}{\hat{\sigma}_u^2} \right\} (\hat{\lambda}^2 - \hat{\gamma}_0)$$

$$Z_t = \frac{(\hat{\alpha} - 1)}{\hat{\sigma}_{\hat{\alpha}}} \left(\frac{\hat{\gamma}_0}{\hat{\lambda}^2} \right)^{1/2} - (1/2)(\hat{\lambda}^2 - \hat{\gamma}_0) \left(\frac{1}{\hat{\lambda}} \right) \left\{ T \frac{\hat{\sigma}_{\hat{\alpha}}}{\hat{\sigma}_u} \right\}$$

donde

$$\begin{aligned} \hat{\gamma}_j &= T^{-1} \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j} \\ \hat{u}_t &= \text{residuos de la regresión estimada en cada caso 1, 2 y 4} \\ \hat{\lambda}^2 &= \hat{\gamma}_0 + 2 \sum_1^q \left(1 - \frac{j}{q+1} \right) \hat{\gamma}_j \\ \hat{\sigma}_u^2 &= (T - c)^{-1} \sum_{t=1}^T \hat{u}_t^2 \\ \hat{\sigma}_{\hat{\alpha}} &= \text{desviación típica estimada de } \hat{\alpha} \text{ en la regresión} \\ c &= \text{número de parámetros estimados en la regresión.} \end{aligned}$$

4.2.3. Otros

La literatura sobre contrastes de raíces unitarias en procesos univariantes ha crecido exponencialmente pero estos serán los que más se han utilizado aunque han sido criticados por su falta de potencia especialmente cuando la raíz es cercana a la unidad. Una posible razón en el caso de los contrastes de DF es que son fáciles de computar.

Han sido propuestos otros contrastes, como por ejemplo el contraste de KPSS de Kwiatkowsky, Phillips, Schmidt y Shin (1992, *J. of Econometrics* 54), que consideran como hipótesis nula la de estacionariedad en lugar de raíz unitaria. También se ha analizado la influencia en estos contrastes de la existencia de cambios estructurales, por ejemplo ver Perron (1989, *Journal of Economic Dynamics and Control*), Perron (1989, *Econometrica*), Perron (1990, *Journal of Business and Economic Statistics*) y Stock (1993, *Hanbook of Econometrics* 4).

⁶Phillips (1987, *Econometrica* 55), Phillips y Perron (1988, *Biometrica*, 75)

Otra rama importante de artículos se centra en contrastes de raíces unitarias en series con estacionalidad (Hylleberg (1990) *Modelling Seasonality*, Oxford University Press) y cada vez están tomando más importancia los procesos de larga memoria o procesos de integración fraccional (volumen especial *Journal of Econometrics*, 110 (2002)).

4.3. Regresión espuria y Cointegración

La existencia de tendencias en los datos puede originar correlaciones que impliquen el aceptar la existencia de relaciones entre dos o más variables en una regresión cuando no existen. Este es un problema que fue descubierto por Granger y Newbold (1974, *J. of Econometrics* 2) y más tarde explicado teóricamente por Phillips (1986, *J. of Econometrics* 33).

Regresión espuria

Considerar dos paseos aleatorios totalmente independientes:

$$y_{1t} = y_{1,t-1} + \epsilon_{1t}$$

$$y_{2t} = y_{2,t-1} + \epsilon_{2t}$$

donde ϵ_{1t} y ϵ_{2t} son ruidos blancos independientes. Por lo tanto ambas series son $I(1)$ y cada una presenta una tendencia estocástica independiente de la otra.

Si se realizara la regresión

$$y_{1t} = \beta_0 + \beta_1 y_{2t} + u_t$$

se debería aceptar la hipótesis nula de que $\beta_1 = 0$. Lo que ocurre es que los estadísticos usuales t y F de la regresión MCO no tienen las distribuciones usuales y hay una tendencia a rechazar la hipótesis nula aumentando a medida que crece el tamaño muestral. Este tipo de regresiones normalmente se caracterizan por un R^2 cercano a uno y un DW cercano a cero. Ver ejemplo de M. Jeréz y S. Sotoca (UCM) y la presentación de J. Gonzalo (UC3M).

Cointegración

El concepto de cointegración fué originariamente introducido por Granger (1981, *J. of Econometrics* 16) y Engel y Granger (1987, *Econometrica* 55) como una forma de considerar relaciones de equilibrio a largo plazo entre variables integradas en principio del mismo orden.

Definición de Engle y Granger (1987): Los componentes del vector y_t se dice que están cointegrados de orden d, b , denotado como $y_t \sim CI(d, b)$ si todos los componentes del vector y_t son $I(d)$ y existe un vector β tal que $\beta'y_t \sim I(d - b)$, $d \geq b > 0$. •

Lütkepohl (1991) utiliza una definición de cointegración algo diferente para no excluir componentes del vector y_t con un orden de integración menor que d .

Definición Lütkepohl (1991): Un proceso y_t de dimensión k es $I(d)$ si $\Delta^d y_t$ es estable pero $\Delta^{d-1} y_t$ no lo es. Entonces el proceso y_t está cointegrado si existe una combinación lineal $\beta'y_t$ que es integrado de un orden menor que d . •

En lo que sigue nos vamos a centrar en el caso específico donde todas las variables individualmente son $I(1)$ o $I(0)$.

Consideremos el caso de un vector y_t bivariante. Si los componentes del vector, y_{1t} e y_{2t} son $I(1)$, en general cualquier combinación de estas dos series seguirá siendo $I(1)$.

Si existe una combinación lineal de ellas que es estacionaria o $I(0)$ se dirá que estas series están cointegradas. Por ejemplo si existe un valor b tal que $y_{1t} - by_{2t}$ es $I(0)$ el vector de cointegración será $\beta = (1, -b)'$. En el corto plazo, las variables se desviarán de esa relación de equilibrio pero esas desviaciones o errores $z_t = y_{1t} - by_{2t}$ serán estacionarios, esto es, oscilarán alrededor de esa relación de equilibrio a largo plazo. Por lo tanto se podrá percibir un movimiento tendencial común. De hecho, podemos ver con un ejemplo sencillo cómo asociado a la cointegración está la existencia de un **modelo de factor de tendencia común**. Consideremos el siguiente modelo:

$$\begin{aligned} y_{1t} &= b\omega_t + \eta_t \\ y_{2t} &= \omega_t + \varepsilon_t \end{aligned}$$

donde ω_t es $I(1)$, η_t , ε_t son $I(0)$. Por lo tanto y_{1t} e y_{2t} son ambas $I(1)$ porque son la combinación de una $I(1)$ con una $I(0)$. El origen de esa tendencia es común a ambas. La combinación $y_{1t} - by_{2t}$ anula esa tendencia convirtiéndose en $I(0)$

$$y_{1t} - by_{2t} = b\omega_t + \eta_t - b(\omega_t + \varepsilon_t) = \eta_t - b\varepsilon_t$$

Por lo tanto y_{1t} e y_{2t} están cointegradas con vector de cointegración $(1, -b)'$. La cointegración surge porque la fuente $I(1)$ de ambas series viene del factor común ω_t o tendencia estocástica común.

Comentarios:

1. Si el vector y_t es de dimensión k entonces al menos pueden existir r relaciones de cointegración tal que $0 \leq r < k$. El número de tendencias comunes será $k - r$.
2. Si una serie es $I(1)$ y la otra $I(0)$ entonces el vector de cointegración sería uno trivial que recogería el hecho de que una serie es estacionaria. Por ejemplo si y_{1t} es $I(1)$ e y_{2t} es $I(0)$ la combinación $\beta' y_t$ es $I(0)$ si $\beta' = (0, b)$ donde b es cualquier valor distinto de cero. El vector de cointegración $\beta' = (0, 1)$ es uno trivial. En el caso de que ambas series fueran $I(0)$ en niveles entonces $r = 2$ pero ambos vectores de cointegración serían triviales.
3. Si tenemos más de dos variables en el vector puede existir más de una relación de cointegración. Si nos centramos en series que son $I(1)$ y permitimos que algunas series en el vector sean $I(0)$ y otras $I(1)$ entonces tenemos que tener en cuenta que habrá tantos vectores de cointegración triviales como series $I(0)$ haya en el vector.
4. Aunque pueden existir r vectores de cointegración linealmente independientes en general no están identificados. Esto es, si β es un vector de cointegración entonces para toda constante $c \neq 0$ $c\beta$ también es un vector de cointegración ya que si $\beta' y_t$ es $I(0)$ también lo es $c\beta' y_t$.

Ejemplos:**PPP (Purchasing Power Parity)**

La teoría de la paridad del poder adquisitivo establece que el tipo de cambio real es igual a uno.

$$S_t P_t^* / P_t = 1 \quad \Leftrightarrow \quad P_t = S_t P_t^*$$

$$p_t - s_t - p_t^* = 0$$

donde $p_t = \ln P_t$ y $p_t^* = \ln P_t^*$ son los niveles de precios (en logaritmos) de cada país, digamos nacional y extranjero respectivamente, y $s_t = \ln S_t$ es el tipo de cambio nominal (en logaritmos) entre las dos monedas. Esto es, una vez considerado el tipo de cambio, dos bienes iguales se deberían de vender al mismo precio en cada país.

Esta relación puede que no se satisfaga exactamente en todo momento debido a costes de transporte entre países, errores en la medición

de los precios etc. Pero podemos contrastar esta hipótesis como una relación de equilibrio en el largo plazo, permitiendo desviaciones en el corto plazo estacionarias esto es que

$$z_t = p_t - s_t^* - p_t^*$$

sea $I(0)$. Siendo el vector de variables $y_t = (p_t, s_t, p_t^*)$, el vector de cointegración será $(1, -1, -1)'$.

Modelos de valor presente descontado

Campbell y Shiller (1987, *J. of Political Economy*, 95)

$$y_{1t} = \theta(1 - \delta) \sum_{i=0}^{\infty} \delta^i E_t y_{2,t+i} + c$$

donde y_{1t} e y_{2t} pueden ser por ejemplo el tipo de interés a largo y a corto plazo, el precio de un activo y los dividendos o el Consumo y la renta laboral respectivamente, donde c es una constante, δ es el factor de descuento y θ es el coeficiente de proporcionalidad.

El modelo de valor presente descontado implica que si las dos series y_{1t} e y_{2t} son $I(1)$, entonces tienen que estar cointegradas siendo el *spread* $S_t = y_{1t} - \theta y_{2t}$ estacionario o $I(0)$. El vector de cointegración sería $(1, -\theta)'$ y el spread recogería las desviaciones en el corto plazo de estas variables alrededor de esta relación de equilibrio a largo plazo.

4.4. Modelo de corrección del error

Consideremos el siguiente modelo VAR(p) en niveles para un vector y_t de dimensión $(k \times 1)$

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t$$

o

$$A(L)y_t = \nu + u_t$$

donde $A(L) = I_k - A_1 L - \dots - A_p L^p$. El proceso es estable si

$$\det(I_k - A_1 z - \dots - A_p z^p) = 0$$

tiene todas sus raíces fuera del círculo unidad.

En ese caso y_t es $I(0)$ y en particular para $z = 1$ $\det(A(1)) \neq 0$ siendo el rango de $A(1) = I_k - A_1 - \dots - A_p$ completo e igual al número de variables en el vector.

Si el polinomio tiene una raíz en el círculo unidad y el resto de raíces caen fuera, entonces y_t es $I(1)$ y pueden existir r relaciones de cointegración $0 \leq r < k$. En ese caso $\det(A(1)) = 0$ siendo la matriz $A(1)$ de rango reducido.

El **VAR(p)** en niveles se puede **reparametrizar** en términos del llamado **VECM** (Vector Error Correction Model) o Modelo de Corrección del Error de la siguiente⁷ forma:

$$\Delta y_t = \nu + \Pi y_{t-1} + D_1 \Delta y_{t-1} + \cdots + D_{p-1} \Delta y_{t-p+1} + u_t$$

o

$$D(L) \Delta y_t = \nu + \Pi y_{t-1} + u_t$$

donde

$$\Pi = -(I_k - A_1 - \cdots - A_p) = -A(1)$$

$$D(L) = I_k - \sum_{j=1}^{p-1} D_j L^j$$

$$D_j = - \sum_{i=j+1}^p A_i, \quad j = 1, \dots, p-1$$

Notar que los parámetros del VAR en niveles se pueden recuperar de los parámetros Π y $D_j, j = 1, \dots, p-1$ del VECM de la siguiente forma:

$$A_1 = D_1 + \Pi + I_k$$

$$A_j = D_j - D_{j-1}, \quad j = 2, \dots, p.$$

La transformación utilizada para obtener el VECM del VAR en niveles es análoga a la utilizada para derivar la regresión del contraste ADF de raíz unitaria o Dickey-Fuller aumentado en el caso univariante.

Por **ejemplo**, considerar el modelo VAR(2)

$$y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + u_t$$

Restando y_{t-1} de ambos lados se obtiene

$$\Delta y_t = \nu - y_{t-1} + A_1 y_{t-1} + A_2 y_{t-2} + u_t.$$

⁷S. Johansen (1991) considera otra reparametrización del VAR(p) en términos de un VECM

$$\Delta y_t = \nu + \sum_{i=1}^{p-1} \Gamma_i L^i \Delta y_t + \Pi y_{t-p} + u_t$$

donde $\Pi = -A(1) = -(I_k - A_1 - \cdots - A_p)$ y $\Gamma_i = -I_k + A_1 + A_2 + \cdots + A_i$ para $i = 1, \dots, p-1$.

Si ahora sumamos y restamos $A_2 y_{t-1}$ a la derecha de la ecuación obtenemos

$$\begin{aligned}\Delta y_t &= \nu - y_{t-1} + A_1 y_{t-1} + A_2 y_{t-1} - A_2 y_{t-1} + A_2 y_{t-2} + u_t \\ &= \nu + (A_1 + A_2 - I_k) y_{t-1} - A_2 (y_{t-1} - y_{t-2}) + u_t \\ &= \nu + \Pi y_{t-1} + D_1 \Delta y_{t-1} + u_t\end{aligned}$$

donde $\Pi = -(I_k - A_1 - A_2)$ y $D_1 = -A_2$.

Consideremos los siguientes casos en función del rango de la matriz $\Pi = -A(1)$:

1. El rango de Π , r es igual a k , $r = k$. En este caso el vector y_t es $I(0)$ en niveles. Por lo tanto $|A(1)| \neq 0$ y Π es de rango completo.
2. El rango de Π , r es igual a cero, $r = 0$. En este caso $\Pi = 0$, la matriz nula. El vector y_t es $I(1)$ y no existe ninguna relación de cointegración. El VECM se convierte en un VAR(p-1) en primeras diferencias que sería la especificación estable adecuada para Δy_t en este caso.
3. El rango de Π es $0 < r < k$. En este caso el vector y_t es $I(1)$ y existen r vectores linealmente independientes de cointegración. La matriz Π es de rango reducido pero no es la matriz nula. Se puede factorizar como el producto

$$\Pi = \alpha \beta' = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1r} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{k1} & \alpha_{k2} & \cdots & \alpha_{kr} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1k} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{r1} & \beta_{r2} & \cdots & \beta_{rk} \end{pmatrix}$$

donde α y β son matrices ($k \times r$) con rango igual a r . Las filas de β' forman la base para los r vectores de cointegración y los elementos de α distribuyen el impacto o el ajuste en el corto plazo de cada variable en el sistema a las desviaciones o errores de desequilibrio $\beta' y_{t-1} \sim I(0)$.

La existencia de cointegración por tanto impone una serie de restricciones no lineales sobre los coeficientes del VAR(p), esto es que $\Pi = \alpha \beta'$. La especificación adecuada si existe cointegración no es el VAR(p-1) en primeras diferencias porque impondría la restricción falsa de que Π es la matriz nula cuando si hay cointegración no lo es aunque sí de rango reducido.

El VECM en este caso es

$$\Delta y_t = \nu + \alpha \beta' y_{t-1} + D_1 \Delta y_{t-1} + \cdots + D_{p-1} \Delta y_{t-p+1} + u_t$$

Todas las variables que aparecen en el VECM son estacionarias.

Ejemplo:

VAR(1) bivalente en niveles:

$$y_t = A_1 y_{t-1} + u_t$$

En forma de VECM:

$$\Delta y_t = \Pi y_{t-1} + u_t$$

donde $\Pi = A_1 - I_2$. Si y_t está cointegrado existe un vector $\beta = (\beta_1, \beta_2)'$ tal que $\beta' y_t = \beta_1 y_{1t} + \beta_2 y_{2t}$ es I(0). Utilizando la normalización $\beta_1 = 1$ y $\beta_2 = -b$ la relación de cointegración sería $y_{1t} - b y_{2t}$. Esta normalización sugiere la relación de equilibrio a largo plazo estocástica

$$y_{1t} = b y_{2t} + z_t$$

donde z_t es I(0) y representa las desviaciones de la relación de equilibrio a largo plazo $y_{1t} = b y_{2t}$. Esta variable aleatoria entraría a formar parte de la dinámica de corto plazo ponderada por los coeficientes de ajuste en el vector $\alpha = (\alpha_1, \alpha_2)'$ de forma que al menos una de las dos variables se tendría que ajustar a estos desequilibrios.

La matriz Π tiene rango reducido e igual a uno y por tanto $\Pi = \alpha \beta'$ tal que el VECM es:

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 & -b \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

Escribiendo el VECM ecuación por ecuación:

$$\Delta y_{1t} = \alpha_1 (y_{1,t-1} - b y_{2,t-1}) + u_{1t}$$

$$\Delta y_{2t} = \alpha_2 (y_{1,t-1} - b y_{2,t-1}) + u_{2t}$$

Los parámetros α_1 y α_2 se conocen con el nombre de coeficientes de ajuste ya que capturan los ajustes de cada una de las variables, si estas están fuera del equilibrio a largo plazo, tal que tiendan a volver hacia esa relación estable.

Comentarios:

- Las matrices α y β no son únicas, ya que para cualquier matriz no singular ($r \times r$) C tenemos que

$$\alpha \beta' = \alpha C C^{-1} \beta' = (\alpha C) (\beta C^{-1})' = \alpha^* \beta^{*'}.$$

Por lo tanto, α y β no están identificados a no ser que se impongan ciertas restricciones.

- Si existe cointegración entonces α no puede ser la matriz nula. Por lo tanto al menos una de las variables del sistema se ha de ajustar en el corto plazo a las desviaciones de las relaciones de largo plazo. Pueden existir ciertas variables que no se ajusten a esas relaciones. En ese caso estas variables se dicen que son débilmente exógenas respecto a la matriz de vectores de cointegración.
- Si y_t está cointegrado con rango r , una forma simple de obtener la presentación de las $k - r$ tendencias comunes es considerar la matriz α_{\perp} de orden $k \times (k - r)$ tal que $\alpha'_{\perp}\alpha = 0$, y utilizar $q_t = \alpha_{\perp}y_t$. Dado que α_{\perp} anula a $\alpha\beta'$, esta serie $(k - r)$ -variante tiene la representación

$$\Delta q_t = \alpha'_{\perp}\nu + D_1\Delta q_{t-1} + \cdots + D_{p-1}\Delta q_{t-p+1} + \alpha'_{\perp}u_t$$

que es un VAR en primeras diferencias de q_t por lo que este vector tiene $k - r$ raíces unitarias.

4.5. Contrastes de Cointegración y Estimación.

4.5.1. Contrastes basados en residuos

Consideremos que, de existir, sólo hubiera una relación de cointegración en el vector y_t . Previamente se ha contrastado y aceptado que las series en y_t son todas individualmente $I(1)$.

- **Vector de cointegración conocido.**

Por ejemplo en el caso de la PPP $\beta' = (1, -1, -1)$ por lo que se quiere contrastar si la serie $z_t = \beta'y_t = p_t - s_t^* - p_t^*$ es o no estacionaria.

Se pueden aplicar los contrastes de raíz unitaria a la serie z_t como los de DF (ADF) o PP. La hipótesis nula sería que esa relación concreta no es de cointegración, esto es $H_0 : z_t$ es $I(1)$ versus la alternativa de que sí lo sea, $H_1 : z_t$ es $I(0)$. En este caso los estadísticos se calcularían de la misma forma y los valores críticos a utilizar serían los mismos que en los contrastes de raíz unitaria.

Podemos no rechazar H_0 pero esto no implica que no pueda existir un vector de cointegración alternativo a $\beta' = (1, -1, -1)$.

- **Vector de cointegración no conocido.**

En ese caso los contrastes de cointegración se basarían en los residuos de la regresión de cointegración (Residual based test):

1. Estimar por MCO la posible relación de cointegración una vez elegida una normalización, esto es la variable dependiente en la regresión. Por ejemplo, si se elige la primera variable y se considera la regresión

$$y_{1t} = b_2 y_{2t} + \dots + b_k y_{kt} + z_t$$

entonces esto implica que el supuesto vector de cointegración es $\beta = (1, -b_2, \dots, -b_k)'$. Esto supone que la variable y_{1t} entra en el vector de cointegración con un coeficiente distinto de cero.

2. Obtener los residuos de la regresión por MCO

$$\hat{z}_t = y_{1t} - \hat{b}_2 y_{2t} - \dots - \hat{b}_k y_{kt}$$

y aplicar los contrastes de raíz unitaria a la serie de residuos \hat{z}_t tales como DF (ADF) o PP.

Aunque los estadísticos se construyen de la misma forma, utilizando la serie de residuos \hat{z}_t , los valores críticos a considerar no son los mismos que los utilizados en los contrastes de raíz unitaria porque las distribuciones asintóticas cambian.

Phillips y Ouliaris (1990, *Econometrica* 58) derivan las distribuciones asintóticas para los estadísticos de Dickey-Fuller y Phillips-Perron y tabulan valores críticos para contrastar la hipótesis nula de no cointegración o de raíz unitaria en los residuos de la regresión de cointegración (ver tablas B.8 y B.9 del Hamilton (1994), pp. 765-766).

Comentarios:

- Si las series están cointegradas entonces el estimador MCO de los coeficientes de la regresión de cointegración es consistente. De hecho converge en probabilidad al verdadero valor a una tasa T , mayor que en el caso de regresores estacionarios \sqrt{T} . Este hecho se denomina *superconsistencia* (Stock (1987), *Econometrica* 55). Pero para muestras finitas es sesgado debido a que la regresión de cointegración considerada es estática por lo que el término de error capturará la dinámica de corto plazo omitida función de Δy_{t-j} y correlacionada con los regresores. Este sesgo puede influir negativamente en la potencia de los contrastes de cointegración basados en los residuos de esta regresión estática.
- Si no sabemos cómo normalizar podemos elegir erróneamente la variable dependiente, por ejemplo si la variable elegida no entra

en la relación de cointegración. Además el estimador MCO no es invariante a cuál sea la variable dependiente en la regresión. Esto puede influir en los resultados de la estimación del vector de cointegración y en la conclusión del contraste de cointegración.

- La distribución asintótica del estimador MCO del vector de cointegración en la regresión estática en general no es la estandar como en el caso de regresores estacionarios, por lo que no se puede hacer inferencia sobre el vector de cointegración con los estadísticos t y F usuales. En general las dificultades surgen por la correlación que puede existir entre el error de la ecuación de cointegración z_t y los regresores como antes hemos comentado. Distintos autores consideran paliar este problema introduciendo dinámica en la regresión de cointegración o corrigiendo el estimador MCO de la regresión estática, por ejemplo el estimador Fully-Modified OLS de Phillips y Hansen (1990) (ver Hamilton (1994) pp. 608-617).
- Si el número de variables en el vector y_t es mayor que dos puede existir más de una relación de cointegración. En ese caso, ¿qué vector de cointegración estaría estimando la regresión de cointegración? Wooldridge (1991) demuestra que la regresión elige aquella relación cuyos residuos están incorrelados con cualquier otra combinación de los regresores que sea $I(1)$.
- Otros contrastes de cointegración que no presentan estos problemas son por ejemplo el método de Johansen (1988,1991) basado en la estimación del VECM por Máxima Verosimilitud con Información completa y en determinar el rango de la matriz de largo plazo Π o el de Stock y Watson (1988) que utiliza la representación de medias móviles, contrastando el número de tendencias comunes.

4.5.2. Método en dos etapas de Engle y Granger

Engle y Granger (1987, *Econometrica*) proponen un estimador de los parámetros del modelo de corrección de error en dos etapas en el caso de que exista solamente una relación de cointegración $r = 1$.

$$\Delta y_t = \nu + \alpha \beta' y_{t-1} + D_1 \Delta y_{t-1} + \dots + D_{p-1} \Delta y_{t-p+1} + u_t$$

1. En una primera etapa, una vez elegida la normalización (por ejemplo y_{1t} como variable dependiente) estimar el vector de cointegración $\beta' = (1, -b_2, \dots, -b_k)$ por MCO en la regresión estática

$$y_{1t} = b_2 y_{2t} + \dots + b_k y_{kt} + z_t$$

2. Estimar cada una de las ecuaciones del modelo de corrección del error separadamente por MCO sustituyendo previamente el error de desequilibrio z_t por los residuos de la regresión de cointegración \hat{z}_t de la primera etapa.

$$\Delta y_{jt} = \nu_j + \alpha_j \hat{z}_{t-1} + \sum_{s=1}^{p-1} \sum_{i=1}^k d_{ji,s} \Delta y_{i,t-s} + \varepsilon_{jt}$$

para cada una de las ecuaciones $j = 1, \dots, k$, donde $\hat{z}_t = y_{1t} - \hat{b}_2 y_{2t} - \dots - \hat{b}_k y_{kt}$. De esta forma se obtienen los estimadores de los parámetros de corto plazo α y D_1, \dots, D_{p-1} .

Dado que el estimador del vector de cointegración β en la primera etapa es superconsistente, Lütkepohl demuestra que los estimadores de D_1, \dots, D_{p-1} y $\Pi = \alpha\beta'$ son consistentes y asintóticamente equivalentes a los estimadores por Máxima Verosimilitud bajo Normalidad imponiendo la restricción de que el rango de Π es igual a uno ($r = 1$ y $\Pi = \alpha\beta'$). En el modelo de corrección del error estimado en la segunda etapa, todas las variables son I(0). En consecuencia, la inferencia sobre los parámetros de corto plazo se puede llevar a cabo utilizando los contrastes t y F usuales ya que serán válidos asintóticamente.

4.5.3. Método de Johansen

Estimación por Máxima Verosimilitud Información Completa

En esta sección resumimos brevemente la estimación por máxima verosimilitud de un modelo VAR(p) cointegrado⁸. El método de Johansen se basa en la estimación por Máxima Verosimilitud del VECM⁹

$$\Delta y_t = \nu d_t + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-p} + u_t$$

donde $t = 1, \dots, T$, conocidas p condiciones iniciales y se supone que $u_t \sim N(0, \Sigma_u)$. En d_t permitimos la existencia de un posible término constante, y otros términos deterministas como tendencia t o variables ficticias que puedan recoger por ejemplo estacionalidad.

Si el rango de Π es $0 < r < k$ entonces existen r relaciones de cointegración y $\Pi = \alpha\beta'$ distinta de la matriz nula. El espacio de cointegración vendría dado por la matriz $(k \times r)$, β y la matriz $(k \times r)$, α

⁸Una derivación más detallada se encuentra en el apéndice de estas notas.

⁹Equivalentemente se puede considerar la parametrización alternativa

$$\Delta y_t = \nu d_t + \Pi y_{t-1} + D_1 \Delta y_{t-1} + \dots + D_{p-1} \Delta y_{t-p+1} + u_t$$

recoge los coeficientes de ajuste de cada variable en el corto plazo a las desviaciones de las r relaciones de equilibrio a largo plazo.

El esquema a seguir para obtener los estimadores MV de los parámetros sujeto a que $\Pi = \alpha\beta'$ es el siguiente:

- Realizar las siguientes regresiones multivariantes auxiliares¹⁰ utilizando las p condiciones iniciales y $t = 1, \dots, T$ observaciones

$$\Delta y_t = \gamma_0 d_t + \Upsilon_1 \Delta y_{t-1} + \Upsilon_2 \Delta y_{t-2} + \Upsilon_{p-1} \Delta y_{t-p+1} + u_{0t}$$

$$y_{t-p} = \gamma_p d_t + \Xi_1 \Delta y_{t-1} + \Xi_2 \Delta y_{t-2} + \Xi_{p-1} \Delta y_{t-p+1} + u_{pt}$$

- Sean R_{0t} y R_{pt} los vectores de residuos de las regresiones anteriores. Definimos las matrices de covarianzas muestrales siguientes:

$$S_{00} = T^{-1} \sum_{t=1}^T R_{0t} R_{0t}' \quad S_{0p} = T^{-1} \sum_{t=1}^T R_{0t} R_{pt}' \quad S_{pp} = T^{-1} \sum_{t=1}^T R_{pt} R_{pt}'$$

- Obtener los valores y vectores propios de $S_{p0} S_{00}^{-1} S_{0p}$ con respecto a S_{pp} . Esto equivale a resolver el problema de obtención de valores propios

$$\left| \lambda S_{pp} - S_{p0} S_{00}^{-1} S_{0p} \right| = 0$$

para los r mayores valores propios de los k existentes $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq \dots \geq \lambda_k$.

El conjunto completo de vectores propios se obtienen de resolver

$$(\lambda_i S_{pp} - S_{p0} S_{00}^{-1} S_{0p}) v_i = 0 \quad i = 1, 2, \dots, k$$

sujeto a la normalización $V' S_{pp} V = I$ donde $V = (v_1, v_2, \dots, v_k)$.

Las r columnas de $\tilde{\beta}_{MV}$ corresponden a los vectores propios (v_1, v_2, \dots, v_r) asociados a $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$.

- La estimación MV del resto de parámetros del VECM se obtiene de la regresión lineal multivariante:

$$\Delta y_t = \nu d_t + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \alpha \tilde{\beta}_{MV} y_{t-p} + u_t$$

donde $t = 1, \dots, T$, conocidas p condiciones iniciales.

¹⁰En el caso de la parametrización en función de y_{t-1} en lugar de y_{t-p} las regresiones serían:

$$\Delta y_t = \gamma_0 d_t + \Upsilon_1 \Delta y_{t-1} + \Upsilon_2 \Delta y_{t-2} + \Upsilon_{p-1} \Delta y_{t-p+1} + u_{0t}$$

$$y_{t-1} = \gamma_1 d_t + \Xi_1 \Delta y_{t-1} + \Xi_2 \Delta y_{t-2} + \Xi_{p-1} \Delta y_{t-p+1} + u_{1t}$$

El logaritmo de la función de verosimilitud evaluada en el máximo utilizando r siendo el rango de cointegración igual a r es:

$$L_r = \text{constante} - \frac{T}{2} \ln \left| \tilde{\beta}'_{MV} (S_{pp} - S_{p0} S_{00}^{-1} S_{0p}) \tilde{\beta}_{MV} \right|$$

o equivalentemente, siendo $\Lambda_r = \tilde{\beta}'_{MV} S_{p0} S_{00}^{-1} S_{0p} \tilde{\beta}_{MV}$

$$L_r = \text{constante} - \frac{T}{2} \ln |I_r - \Lambda_r|$$

$$L_r = \text{constante} - \frac{T}{2} \sum_{j=1}^r \ln(1 - \lambda_j)$$

Este valor se utiliza en los contrastes de razón de verosimilitudes para contrastar la hipótesis nula de que el rango de cointegración sea r .

Finalmente, las estimaciones de los complementos ortogonales de α y β se pueden obtener utilizando:

$$\hat{\alpha}_\perp = S_{00}^{-1} S_{pp} [v_{r+1}, \dots, v_k] \quad \hat{\beta}_\perp = S_{pp} [v_{r+1}, \dots, v_k]$$

4.5.4. Contrastes sobre el rango de cointegración

El objetivo es contrastar el rango de Π . Sea $H(r)$ la hipótesis nula de que el rango de Π es r . Por ejemplo, bajo $H(0)$, el rango de Π es igual a cero, por lo que $\Pi = 0$ y no hay cointegración. Las hipótesis de interés son

$$H(0) \subset H(1) \subset \dots \subset H(r) \subset \dots \subset H(k)$$

Por lo tanto, se comienza por $H(0)$ y se sigue mientras se rechace la hipótesis nula. En caso de no rechazarse se pararía el proceso siendo el número de vectores de cointegración los de esa hipótesis nula.

Contraste de la traza: Considerar las hipótesis

$$H(r) : \text{Rango } \Pi = r \quad \text{frente a} \quad H_a : \text{Rango } \Pi > r$$

Johansen (1988) propone un contraste de la razón de verosimilitudes. Bajo la hipótesis nula el logaritmo de la función de verosimilitud es

$$L_r = c - \frac{T}{2} \sum_{j=1}^r \ln(1 - \lambda_j)$$

Bajo la hipótesis alternativa se consideran los k vectores propios como estimador de β esto es V y por lo tanto el logaritmo neperiano de la función de verosimilitud sin restringir es:

$$L_a = c - \frac{T}{2} \ln |I_k - \Lambda| = L_a = c - \frac{T}{2} \sum_{j=1}^k \ln(1 - \lambda_j)$$

El estadístico de razón de verosimilitudes es el llamado por Johansen el estadístico de la traza (Trace statistic):

$$\begin{aligned} LR_{tr}(r) &= -2(L_r - L_a) \\ &= T \left(\sum_{j=1}^r \ln(1 - \lambda_j) - \sum_{j=1}^k \ln(1 - \lambda_j) \right) \\ &= -T \sum_{j=r+1}^k \ln(1 - \lambda_j) \end{aligned}$$

Si la hipótesis nula fuera cierta, entonces λ_j debería ser pequeño para $j > r$ y por lo tanto $LR_{tr}(r)$ debería tomar valores pequeños.

Su distribución asintótica no es la convencional Chi-cuadrado sino función de movimientos Brownianos. Esta distribución depende del número de variables en el vector y_t y de si se incluyen intercepto y componentes deterministas en el VECM. Los valores críticos están obtenidos via simulación y tabulados por Johansen y Juselius y por Osterwald-Lenum .

Contraste del mayor valor propio

Este es un proceso secuencial que también propone Johansen (1988) para determinar el número de vectores de cointegración. En este caso

$$H(r) : \text{Rango } \Pi = r \quad \text{frente a} \quad H(r+1) : \text{Rango } \Pi = r + 1$$

También es un contraste de razón de verosimilitudes.

$$\begin{aligned} LR_{\lambda-max}(r) &= -2(L_r - L_{r+1}) \\ &= T \left(\sum_{j=1}^r \ln(1 - \lambda_j) - \sum_{j=1}^{r+1} \ln(1 - \lambda_j) \right) \\ &= -T \ln(1 - \lambda_{r+1}) \end{aligned}$$

Lo mismo ocurre con la distribución asintótica de este estadístico. En general las conclusiones de estos dos contrastes no tienen por qué coincidir ya que las alternativas son distintas. La del estadístico de la traza es más general que la del mayor valor propio.

En este marco de contrastes de razón de verosimilitudes se pueden llevar a cabo otros tipos de contrastes. Por ejemplo sobre ciertas restricciones en el espacio de cointegración: si alguno de los vectores de cointegración es trivial o si alguna variable no entra en ninguno de los vectores de cointegración etc. Johansen ha derivado distintos contrastes de este tipo¹¹ y en muchos de los casos las distribuciones asintóticas sí son la usual Chi-cuadrado.

¹¹Podeis ver como ejemplo de aplicación de esta metodología Johansen y Juselius (1990) Oxford Bulletin of Economics and Statistics, Johansen y Juselius (1992) Journal of Econometrics, Gardeazabal y Regúlez (1992) Lecture Notes in Economics and Mathematical Systems, vol. 385, Springer Verlag, entre otros.

Capítulo 5

Apéndice¹

5.1. Valores y Vectores propios-Raíces características

Los valores propios o valores característicos o raíces características de una matriz cuadrada ($m \times m$) A son las raíces del polinomio en λ dados por

$$\det(A - \lambda I_m) \quad \text{o} \quad \det(\lambda I_m - A)$$

El determinante se conoce a veces con el nombre de *determinante característico* y el polinomio es el *polinomio característico* de A . Dado que las raíces de un polinomio son números complejos en general los valores propios son también complejos. Un número λ_i es un valor propio de A , si las columnas de $(A - \lambda_i I_m)$ son linealmente dependientes. En consecuencia, existe un vector ($m \times 1$) $v_i \neq 0$ tal que

$$(A - \lambda_i I_m)v_i = 0 \quad \text{o} \quad Av_i = \lambda_i v_i.$$

Un vector con esta propiedad es un vector característico o vector propio correspondiente al valor propio λ_i . Cualquier escalar distinto de cero por v_i es también un vector propio correspondiente a λ_i .

Ejemplo:

Consideremos la matriz $\begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}$.

Sus valores propios son las raíces de

$$|A - \lambda I_2| = \det \begin{bmatrix} 1 - \lambda & 0 \\ 1 & 3 - \lambda \end{bmatrix} = (1 - \lambda)(3 - \lambda)$$

¹Varias de las secciones se basan en los apéndices del libro de Lütkepohl.

Por lo tanto $\lambda_1 = 1$ y $\lambda_2 = 3$ son los valores propios de la matriz A anterior. Los correspondientes vectores propios se obtienen resolviendo

$$\begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix} = 3 \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix}$$

Por lo tanto

$$\begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

son los vectores propios correspondientes a λ_1 y λ_2 respectivamente, o los mismos multiplicados por un escalar no nulo.

Reglas:

1. Si A es simétrica, entonces todos sus valores propios son números reales.
2. Los valores propios de una matriz diagonal son sus elementos de la diagonal.
3. Los valores propios de una matriz triangular son los elementos de la diagonal principal.
4. Una matriz $(m \times m)$ tiene como máximo m valores propios.
5. Sea $\lambda_1, \dots, \lambda_m$ los valores propios de la matriz A , entonces $|A| = \lambda_1 \cdots \lambda_m$.
6. Sean λ_i y λ_j dos valores propios de A distintos con correspondientes vectores propios v_i y v_j . Entonces v_i y v_j son linealmente independientes.
7. Todos los valores propios de la matriz $(m \times m)$ A tienen módulo¹ menor que la unidad si y solo si $\det(I_m - Az) \neq 0$ para $|z| \leq 1$, esto es, el polinomio $\det(I_m - Az)$ no tiene raíces ni dentro ni en el círculo unidad.

Sea A una matriz real $(m \times m)$ con valores propios $\lambda_1, \dots, \lambda_m$ los cuales tienen todos módulo menor que la unidad, esto es, $|\lambda_i| < 1$ para $i = 1, \dots, m$. Entonces se satisface lo siguiente:

- i) $A^j \rightarrow 0$ cuando $j \rightarrow \infty$
- ii) $\sum_{j=0}^{\infty} A^j = (I_m - A)^{-1}$ existe.

¹Sea z un número complejo, $z = z_1 + iz_2$. El módulo se define como $|z| = \sqrt{z_1^2 + z_2^2}$. Si $z_2 = 0$ entonces z es un número real y el módulo es el valor absoluto de z lo que justifica la notación

- iii) La sucesión A^j , $j = 0, 1, 2, \dots$, es absolutamente sumable, esto es $\sum_{j=0}^{\infty} |a_{kl,j}|$ es finita para todo $k, l = 1, \dots, m$, donde $a_{kl,j}$ es un elemento típico de A^j .

5.2. Descomposición de Choleski

Si A es una matriz definida positiva ($m \times m$) existe una matriz triangular por arriba y una matriz triangular por abajo P con la diagonal principal positiva tal que

$$P^{-1}A(P')^{-1} = I_m \quad \text{o} \quad A = PP'$$

La descomposición $A = PP'$, donde P es triangular por abajo con la diagonal principal positiva se conoce con el nombre de *descomposición de Choleski*. Existen programas de ordenador que determinan la matriz P de una matriz definida positiva A . Si se quiere obtener una matriz Q triangular por arriba y el programa de ordenador nos da una matriz triangular por abajo se puede proceder como sigue: Se define una matriz ($m \times m$)

$$G = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & \ddots & \vdots & \\ 1 & & 0 & 0 \end{bmatrix}$$

con unos en la diagonal que parte desde arriba de la esquina de la derecha hacia abajo a la esquina de la izquierda y ceros en el resto de elementos. Notar que $G = G'$ y $G^{-1} = G$. Supongamos que queremos obtener una descomposición de la matriz A . Entonces descomponer primeramente la matriz $B = GAG$ como $B = PP'$ donde P es la matriz triangular por abajo. Entonces

$$A = GBG = GPGGP'G = QQ',$$

donde $Q = GPG$ es triangular por arriba.

5.3. Sumas infinitas de variables aleatorias

La representación MA de un proceso VAR es a menudo una suma infinita de vectores aleatorios. Al igual que en el estudio de sumas infinitas de números reales debemos especificar con precisión qué queremos decir con tal suma infinita. El concepto de convergencia absoluta es básico en lo que sigue.

Una sucesión $\{a_i\}$, $i = 0, \pm 1, \pm 2, \dots$, doblemente infinita de números reales es *absolutamente sumable* si $\lim \sum_{i=-n}^n |a_i|$ cuando $n \rightarrow \infty$ existe y es finito. El límite se suele denotar por $\sum_{i=-\infty}^{\infty} |a_i|$.

El siguiente teorema provee una justificación para trabajar con sumas infinitas de variables aleatorias.

Proposición(*Existencia de sumas infinitas de variables aleatorias*)

Sea $\{a_i\}$ una sucesión absolutamente sumable de números reales y $\{z_t\}$ $t = 0, \pm 1, \pm 2, \dots$, una sucesión de variables aleatorias que satisfacen $E(z_t^2) \leq c$, $t = 0, \pm 1, \pm 2, \dots$, para alguna constante finita c .

Entonces existe una sucesión $\{y_t\}$, $t = 0, \pm 1, \pm 2, \dots$, de variables aleatorias tal que $\sum_{i=-n}^n a_i z_{t-i}$ converge en media cuadrática² a y_t cuando $n \rightarrow \infty$ y por lo tanto también converge en probabilidad.

Las variables aleatorias y_t están determinadas de forma única excepto por un conjunto de probabilidad nula.

Estos conceptos se pueden generalizar a procesos vectoriales.

Una sucesión de matrices $(k \times k)$, $\{A_i\}$, $i = 0, \pm 1, \pm 2, \dots$, es *absolutamente sumable* si cada sucesión $\{a_{mn,i}\}$, $m, n = 1, \dots, k$, es absolutamente sumable.

Proposición(*Existencia de sumas infinitas de vectores de variables aleatorias*)

Sea $\{A_i\}$ una sucesión absolutamente sumable de matrices reales $(k \times k)$ y $\{z_t\}$ $t = 0, \pm 1, \pm 2, \dots$, una sucesión de variables aleatorias k -dimensionales que satisfacen $E(z_t' z_t) \leq c$, $t = 0, \pm 1, \pm 2, \dots$, para alguna constante finita c . Entonces existe una sucesión de variables aleatorias k -dimensionales $\{y_t\}$, $t = 0, \pm 1, \pm 2, \dots$, tal que $\sum_{i=-n}^n A_i z_{t-i}$ converge en media cuadrática³ a y_t cuando $n \rightarrow \infty$ y por lo tanto también converge en probabilidad. Las variables aleatorias y_t están determinadas de forma única excepto por un conjunto de probabilidad nula.

Con respecto a los momentos de sumas infinitas tenemos el siguiente resultado

Proposición(*Momentos de sumas infinitas de vectores de variables aleatorias*)

Supongamos que $\{z_t\}$ satisface las condiciones de la proposición anterior, $\{A_i\}$ y $\{B_i\}$ son sucesiones absolutamente sumables de matrices

²Una sucesión de variables aleatorias $\{x_T\}$ converge en media cuadrática a x si $\lim E(x_T - x)^2 = 0$ cuando $T \rightarrow \infty$

³Una sucesión de variables aleatorias $\{x_T\}$ k -dimensionales converge en media cuadrática a la variable aleatoria k -dimensional x si $\lim E(x_T - x)'(x_T - x) = 0$ cuando $T \rightarrow \infty$

$(k \times k)$ e

$$y_t = \sum_{i=-\infty}^{\infty} A_i z_{t-i}, \quad x_t = \sum_{i=-\infty}^{\infty} B_i z_{t-i}$$

Entonces,

$$E(y_t) = \lim_{n \rightarrow \infty} \sum_{i=-n}^n A_i E(z_{t-i})$$

$$E(y_t x_t') = \lim_{n \rightarrow \infty} \sum_{i=-n}^n \sum_{j=-n}^n A_i E(z_{t-i} z_{t-j}') B_j',$$

donde el límite de una sucesión de matrices es la matriz de los límites de las sucesiones de los elementos individuales.

5.4. Método de Johansen

El método de Johansen se basa en la estimación por Máxima Verosimilitud del VECM

$$\Delta y_t = \nu + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-p} + u_t$$

donde $\Gamma_i = -I_k + A_1 + A_2 + \dots + A_i$ para $i = 1, \dots, p-1$ y $\Pi = -(I_k - A_1 - \dots - A_p)$ suponiendo $u_t \sim N(0, \Sigma_u)$.

Si el rango de Π , r , es igual a k entonces y_t es estacionario I(0). Si el rango de Π , r es cero, entonces Π es la matriz nula e y_t es I(1) pero no habría ninguna relación de cointegración. Si el rango de Π es $0 < r < k$ entonces existen r relaciones de cointegración y $\Pi = \alpha \beta'$ distinta de la matriz nula. El espacio de cointegración vendría dado por la matriz $(k \times r)$, β y la matriz $(k \times r)$, α recoge los coeficientes de ajuste de cada variable en el corto plazo a las desviaciones de las r relaciones de equilibrio a largo plazo.

El logaritmo de la función de verosimilitud de (y_1, y_2, \dots, y_T) condicionando a los valores iniciales $(y_{1-p}, y_{2-p}, \dots, y_0)$ es:

$$L(\Gamma, \Pi, \Sigma_u) = -\frac{Tk}{2} \ln(2\pi) - \frac{T}{2} \ln |\Sigma_u| - \frac{1}{2} \sum_{t=1}^T u_t' \Sigma_u^{-1} u_t$$

donde $u_t = \Delta y_t - \nu - \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} - \Pi y_{t-p}$ y $\Gamma = (\nu, \Gamma_1, \dots, \Gamma_{p-1})$.

El esquema a seguir para obtener los estimadores MV de los parámetros sujeto a que $\Pi = \alpha \beta'$ es primeramente obtener la función de verosimilitud concentrada que dependa solamente de la matriz Π de la siguiente forma:

- Obtener $\tilde{\Sigma}_u$ que maximiza L dados Π y Γ . Este es:

$$\tilde{\Sigma}_u = T^{-1} \sum_{t=1}^T u_t u_t'$$

donde $u_t = \Delta y_t - \nu - \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} - \Pi y_{t-p}$.

- Obtener $\tilde{\Gamma} = (\tilde{\nu}, \tilde{\Gamma}_1, \dots, \tilde{\Gamma}_{p-1})$ que maximiza L dados Π y Σ_u .

Si definimos $Z_{0t} = \Delta y_t$, $Z_{pt} = y_{t-p}$ y

$$Z_{1t} = \begin{pmatrix} 1 \\ \Delta y_{t-1} \\ \Delta y_{t-2} \\ \vdots \\ \Delta y_{t-p+1} \end{pmatrix}$$

Podemos escribir, dado un valor de Π ,

$$(Z_{0t} - \Pi Z_{pt}) = \Gamma Z_{1t} + u_t \quad t = 1, \dots, T$$

donde $u_t \sim N(0, \Sigma_u)$. El estimador MV de Γ dado Π y Σ_u coincide con el estimador por MCO en el sistema (que coincide con el de MCO ecuación por ecuación) ya que es un sistema de ecuaciones aparentemente no relacionadas con idénticos regresores en cada ecuación.

$$\tilde{\Gamma}' = \left(\sum_{t=1}^T Z_{1t} Z_{1t}' \right)^{-1} \left(\sum_{t=1}^T Z_{1t} (Z_{0t} - \Pi Z_{pt})' \right)$$

La i -ésima columna de $\tilde{\Gamma}'$ es el vector de coeficientes de la regresión de $(\Delta y_{it} - \Pi \Delta y_{i,t-p})$ sobre $Z_{1t}' = (1, \Delta y'_{i,t-1}, \dots, \Delta y'_{i,t-p+1})$.

Podemos escribir

$$\tilde{\Gamma}' = \left(\sum_{t=1}^T Z_{1t} Z_{1t}' \right)^{-1} \left(\sum_{t=1}^T Z_{1t} Z_{0t}' \right) - \Pi' \left(\sum_{t=1}^T Z_{1t} Z_{1t}' \right)^{-1} \left(\sum_{t=1}^T Z_{1t} Z_{pt}' \right)$$

Sustituyendo $\tilde{\Gamma}$ en

$$(Z_{0t} - \Pi Z_{pt}) = \Gamma Z_{1t} + u_t$$

y reorganizando tenemos

$$R_{0t} = \Pi R_{pt} + u_t \quad t = 1, \dots, T$$

donde

$$R_{0t} = Z_{0t} - \left(\sum_{t=1}^T Z_{1t} Z'_{0t} \right) \left(\sum_{t=1}^T Z_{1t} Z'_{1t} \right)^{-1} Z_{1t} \quad t = 1, \dots, T$$

Cada uno de los elementos del vector R_{0t} es el residuo en t de la regresión de cada uno de los elementos de Δy_t sobre $(1, \Delta y'_{t-1}, \dots, \Delta y'_{t-p+1})$, $t = 1, \dots, T$.

$$R_{pt} = Z_{pt} - \left(\sum_{t=1}^T Z_{1t} Z'_{pt} \right) \left(\sum_{t=1}^T Z_{1t} Z'_{1t} \right)^{-1} Z_{1t} \quad t = 1, \dots, T$$

Cada uno de los elementos del vector R_{pt} es el residuo en t obtenido de la regresión de cada uno de los elementos de y_{t-p} sobre $(1, \Delta y'_{t-1}, \dots, \Delta y'_{t-p+1})$, $t = 1, \dots, T$.

- Sustituyendo en el logaritmo de la función de verosimilitud $\tilde{\Sigma}_u$ y $\tilde{\Gamma}$ obtenidos anteriormente dada la matriz Π obtenemos la función de verosimilitud concentrada que solamente depende de Π

$$L_c(\Pi) = \text{const.} - \frac{T}{2} \ln \left| \frac{1}{T} \sum_{t=1}^T (R_{0t} - \Pi R_{pt})(R_{0t} - \Pi R_{pt})' \right|$$

Donde hemos utilizado el resultado de que

$$\sum_{t=1}^T u'_t \tilde{\Sigma}_u^{-1} u_t = T \text{tr}(I_k) = Tk$$

El siguiente paso es maximizar $L_c(\Pi)$ sujeto a que se satisface $\Pi = \alpha\beta'$ dado que es de rango reducido r . Si denotamos por

$$S_{ij} = T^{-1} \sum_{t=1}^T R_{it} R'_{jt} \quad i, j = 0, p$$

podemos escribir la función de verosimilitud concentrada solamente función de α y β

$$L_c(\alpha, \beta) = \text{const.} - \frac{T}{2} \ln |S_{00} - \alpha\beta' S_{p0} - S_{0p}\beta\alpha' + \alpha\beta' S_{pp}\beta\alpha'|$$

Finalmente obtenemos el estimador MV de β una vez hemos obtenido el estimador MV de α dado β . Este último se obtiene de maximizar $L_c(\alpha, \beta)$ con respecto de α tomando como dado β ,

$$\tilde{\alpha} = S_{0p}\beta(\beta' S_{pp}\beta)^{-1}$$

Concentrando de nuevo la función de verosimilitud ahora en $\tilde{\alpha}$ obtenemos

$$L_{cc}(\beta) = \text{const.} - \frac{T}{2} \ln \left| S_{00} - S_{0p}\beta (\beta' S_{pp}\beta)^{-1} \beta' S_{p0} \right|$$

o equivalentemente ⁴

$$L_{cc}(\beta) = \text{const.} - \frac{T}{2} \ln \left(|\beta' S_{pp}\beta|^{-1} |S_{00}| \left| \beta' (S_{pp} - S_{p0}S_{00}^{-1}S_{0p})\beta \right| \right)$$

Maximizar esta función con respecto a β corresponde a minimizar con respecto de β el cociente

$$\frac{\left| \beta' (S_{pp} - S_{p0}S_{00}^{-1}S_{0p})\beta \right|}{|\beta' S_{pp}\beta|}$$

Imponiendo la normalización $\beta' S_{pp}\beta = I$ el estimador por Máxima verosimilitud de β requiere minimizar

$$\left| \beta' (S_{pp} - S_{p0}S_{00}^{-1}S_{0p})\beta \right| \text{ sujeto a } \beta' S_{pp}\beta = I$$

Este problema de minimización equivale a resolver el conjunto de ecuaciones de la forma

$$(\lambda S_{pp} - S_{p0}S_{00}^{-1}S_{0p})\beta = 0$$

donde λ viene dado de resolver

$$\left| \lambda S_{pp} - S_{p0}S_{00}^{-1}S_{0p} \right| = 0$$

para los r mayores valores propios de los k existentes $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq \dots \geq \lambda_k$.

Las columnas de $\tilde{\beta}_{MV}$ corresponden a los vectores propios (v_1, v_2, \dots, v_r) asociados a $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$. El conjunto completo de vectores propios se obtienen de resolver

$$(\lambda_i S_{pp} - S_{p0}S_{00}^{-1}S_{0p})v_i = 0 \quad i = 1, 2, \dots, k$$

sujeto a la normalización $V' S_{pp} V = I$ donde $V = (v_1, v_2, \dots, v_k)$.

⁴Dado que se tiene la siguiente igualdad

$$|\beta' S_{pp}\beta| \left| S_{00} - S_{0p}\beta (\beta' S_{pp}\beta)^{-1} \beta' S_{p0} \right| = |S_{00}| \left| \beta' S_{pp}\beta - \beta' S_{p0}S_{00}^{-1}S_{0p}\beta \right|$$

se tiene que

$$\left| S_{00} - S_{0p}\beta (\beta' S_{pp}\beta)^{-1} \beta' S_{p0} \right| = |\beta' S_{pp}\beta|^{-1} |S_{00}| \left| \beta' (S_{pp} - S_{p0}S_{00}^{-1}S_{0p})\beta \right|$$

Una vez obtenido $\tilde{\beta}_{MV}$ se pueden obtener los estimadores MV del resto de parámetros:

$$\begin{aligned}\tilde{\alpha}_{MV} &= S_{0p}\tilde{\beta}_{MV} \\ \tilde{\Pi}_{MV} &= \tilde{\alpha}_{MV}\tilde{\beta}'_{MV} \\ \tilde{\Gamma}'_{MV} &= \left(\sum_{t=1}^T Z_{1t}Z'_{1t}\right)^{-1} \left(\sum_{t=1}^T Z_{1t}(Z_{0t} - \tilde{\Pi}_{MV}Z_{pt})'\right) \\ \tilde{\Sigma}_u &= T^{-1} \sum_{t=1}^T \tilde{u}_t\tilde{u}'_t\end{aligned}$$

donde $\tilde{u}_t = R_{0t} - \tilde{\Pi}_{MV}R_{pt}$.

Para resolver el problema de obtener los valores propios λ_i , $i = 1, \dots, k$ tal que

$$\left|\lambda S_{pp} - S_{p0}S_{00}^{-1}S_{0p}\right| = 0$$

se pueden utilizar métodos convencionales de obtención de valores propios de una matriz. Dado que S_{pp} es simétrica y definida positiva esta se puede expresar como $S_{pp} = GG'$ donde G es una matriz no singular. Por lo tanto $G^{-1}S_{pp}(G')^{-1} = I$ y

$$\left|G^{-1}\left|\lambda S_{pp} - S_{p0}S_{00}^{-1}S_{0p}\right|(G')^{-1}\right| = 0$$

equivalentemente

$$\begin{aligned}\left|\lambda G^{-1}S_{pp}(G')^{-1} - G^{-1}S_{p0}S_{00}^{-1}S_{0p}(G')^{-1}\right| &= 0 \\ \left|\lambda I - G^{-1}S_{p0}S_{00}^{-1}S_{0p}(G')^{-1}\right| &= 0\end{aligned}$$

Por ejemplo RATS tiene instrucciones para descomponer S_{pp} en GG' y para obtener los valores propios de una matriz⁵. Además de

$$(\lambda_i S_{pp} - S_{p0}S_{00}^{-1}S_{0p})v_i = 0 \quad i = 1, 2, \dots, k$$

tenemos que

$$S_{p0}S_{00}^{-1}S_{0p}V = S_{pp}V\Lambda$$

donde Λ es la matriz diagonal con los valores propios en su diagonal principal. Así dado que $V'S_{pp}V = I$ tenemos que

$$V'S_{p0}S_{00}^{-1}S_{0p}V = \Lambda$$

y

$$\tilde{\beta}'_{MV}S_{p0}S_{00}^{-1}S_{0p}\tilde{\beta}_{MV} = \Lambda_r$$

es decir la matriz de vectores propios V asociados a los valores propios $\lambda_1, \dots, \lambda_k$ diagonaliza la matriz $S_{p0}S_{00}^{-1}S_{0p}$. La matriz Λ está ordenada tal que los primeros r elementos de la diagonal principal son los r mayores valores propios (Λ_r) y los restantes $k - r$ serían los menores. Estos valores propios juegan un papel muy importante en la derivación de los contrastes de razón de verosimilitud sobre el rango de cointegración r .

⁵Aunque los valores propios son los mismos los vectores propios cambian. Denotando por e_i a estos últimos podemos obtener $v_i = (G')^{-1}e_i$.