

Modelos vectoriales autoregresivos (VAR)

Alfonso Novales
Universidad Complutense

Noviembre 2017
Preliminary version
No citar sin permiso del autor
©Copyright A.Novales 2016

Contents

1	Introducción	2
2	Un modelo estructural dinámico	3
3	Modelos VAR	4
3.1	El modelo VAR(1)	4
3.2	Modelos VAR de orden superior	7
3.3	VAR y modelos univariantes	8
3.4	Estimación y contrastación de hipótesis	9
3.4.1	Contrastes de especificación	10
3.4.2	Contrastes de causalidad	11
3.5	Identificación en un modelo VAR	12
3.5.1	Algunos ejemplos	14
3.5.2	Identificación ortogonal: el enfoque Cholesky	16
3.6	Representación MA de un modelo VAR	22
3.7	Funciones de respuesta al impulso	24
3.8	Descomposición de la varianza	27
3.9	VAR Generalizado: un análisis sin condiciones de identificación	29
3.9.1	Funciones de respuesta al impulso	29
3.9.2	Descomposición de la varianza del error de predicción a h periodos	32
3.9.3	Spillovers entre variables: medidas de contagio	34
3.10	Apéndice: Suma de respuestas a un impulso	37
4	Cointegración	37
4.1	Contraste de cointegración	39
4.2	El contraste de cointegración de Johansen	43
4.3	El modelo de corrección del error	47

4.4	Cointegración por umbrales	51
4.5	Aplicaciones	52
4.5.1	Index Benchmark tracking	52
4.5.2	Pairs trading	54
5	Ejercicios	56
6	Técnicas de cointegración en el análisis de “Asset allocation”	56
7	Apéndices	58
7.1	Las innovaciones de un modelo estructural deben estar incorrelacionadas entre sí.	58
7.2	Errata en Enders, página 299,	58

1 Introducción

Utilizamos un modelo del tipo vector autoregresivo (VAR) cuando queremos caracterizar las interacciones simultáneas entre un grupo de variable. Un VAR es un modelo de ecuaciones simultáneas formado por un sistema de ecuaciones de forma reducida sin restringir. Que sean ecuaciones de forma reducida quiere decir que los valores contemporáneos de las variables del modelo no aparecen como variables explicativas en ninguna de las ecuaciones. Por el contrario, el conjunto de variables explicativas de cada ecuación está constituido por un bloque de retardos de cada una de las variables del modelo. Que sean ecuaciones no restringidas significa que aparece en cada una de ellas el mismo grupo de variables explicativas.

Pueden incluirse también como variables explicativas algunas variables de naturaleza determinista, como una posible tendencia temporal, variables ficticias estacionales, o una variable ficticia de tipo impulso o escalón, que sirve para llevar a cabo una análisis de intervención en el sistema. Por último, podría incluirse como explicativa una variable, incluso en valor contemporáneo, que pueda considerarse *exógena* respecto a las variables que integran el modelo VAR.

Se dice que la variable z_t es predeterminada en una ecuación si $E(z_{t-s}u_t) = 0, \forall s \geq 0$, siendo u_t el término de error de dicha ecuación. En una ecuación cuyo término de error no tiene autocorrelación, los retardos de la variable endógena, si aparecen como variables explicativas, serían variables explicativas predeterminadas. La variable z_t sería exógena en dicha ecuación si: $E(z_{t-s}u_t) = 0, \forall s \neq 0$. Nótese que ninguna de ellas son propiedades que una variable tenga con carácter absoluto, como parte de su naturaleza, sino que lo es en relación con una determinada ecuación. De hecho, una variable z puede ser exógena en una ecuación y no serlo en otra ecuación diferente. También puede decirse que una variable z es exógena (o predeterminada) en determinada ecuación si satisface las condiciones anteriores. En realidad, dado que se trata de propiedades de una variable explicativa respecto del componentes no explicado de una variable que se considera endógena, se podría decir que la variable z es exógena (o predeter-

minada) respecto de y . Y , por supuesto, podría no ser exógena con respecto a otras variables.

El modelo VAR es muy útil cuando existe evidencia de simultaneidad entre un grupo de variables, y que sus relaciones se transmiten a lo largo de un determinado número de períodos. Al no imponer ninguna restricción sobre la versión estructural del modelo, no se incurre en los errores de especificación que dichas restricciones pudieran causar al ejercicio empírico. De hecho, la principal motivación detrás de los modelos VAR es la dificultad en identificar variables como exógenas, como es preciso hacer para identificar un modelo de ecuaciones simultáneas.

2 Un modelo estructural dinámico

Partimos de un *modelo estructural* dinámico,

$$\begin{aligned} y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + \gamma'_1 z_t + \varepsilon_{1t} \\ y_{2t} &= \alpha_{20} + \alpha_{21}y_{1t} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{2t-1} + \gamma'_2 z_t + \varepsilon_{2t} \end{aligned} \quad (1)$$

donde y_{1t}, y_{2t} son variables estacionarias, y $\varepsilon_{1t}, \varepsilon_{2t}$ son procesos ruido blanco con esperanza cero, varianzas $\sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2$ y covarianza σ_{12} . Este es un modelo de ecuaciones simultáneas con dos variables endógenas, y_{1t}, y_{2t} , y un vector z_t de variables exógenas. Un *shock* inesperado sobre y_{2t} , en la forma de un valor no nulo de la innovación estructural ε_{2t} , afecta directamente a y_{2t} , pero también influye sobre y_{1t} a través de la presencia de y_{2t} como variable explicativa en la primera ecuación. Además, este efecto se propaga en el tiempo, debido a la presencia de los valores retardados de ambas variables como variables explicativas. Por simplicidad, suponemos de momento que únicamente aparece un primer retardo de ambas variables como explicativas en cada ecuación, aunque la identificación del número de retardos forma parte del análisis de especificación del modelo, como veremos más adelante.

El modelo estructural puede incorporar asimismo un vector de variables explicativas exógenas z_t en cada ecuación, que pueden aparecer asimismo con retardos. Un ejemplo de este tipo de variables serían una tendencia determinista, o variables ficticias estacionales. También podrían ser variables que se determinan claramente fuera de la influencia de y_{1t} e y_{2t} , de modo que pueda justificarse que $E(z_{t-s}\varepsilon_{1t}) = E(z_{t-s}\varepsilon_{2t}) = 0 \forall s$. Por ejemplo, el precio del barril de petróleo Brent, determinado en mercados internacionales con poca influencia de España, siendo y_{1t} e y_{2t} variables de ámbito nacional (por ejemplo, las rentabilidades del Ibex35 y del futuro sobre Ibex35).

De forma resumida, la representación matricial del *modelo estructural* (nótese que los coeficientes de y_t no son la matriz identidad) de primer orden puede escribirse,

$$By_t = \Gamma_0 + \Gamma_1 y_{t-1} + Gz_t + \varepsilon_t$$

En el caso del modelo de dos ecuaciones, las matrices de esta representación son:

$$B = \begin{pmatrix} 1 & -\alpha_{11} \\ -\alpha_{21} & 1 \end{pmatrix}; \Gamma_0 = \begin{pmatrix} \alpha_{10} \\ \alpha_{20} \end{pmatrix}; \Gamma_1 = \begin{pmatrix} \alpha_{12} & \alpha_{13} \\ \alpha_{22} & \alpha_{23} \end{pmatrix}; G = \begin{pmatrix} \gamma'_1 \\ \gamma'_2 \end{pmatrix};$$

El modelo estructural VAR presenta dos dificultades para su estimación: a) la simultaneidad, al aparecer cada una de las dos variables como variable explicativa en la ecuación de la otra, lo que genera inconsistencia del estimador MCO, b) si los términos de error tuviesen autocorrelación, las estimaciones MCO serían inconsistentes, al tratarse de un modelo dinámico. En cualquiera de estos casos se incumpliría la condición de ausencia de correlación entre cada una de las variables explicativas y el término de error de una misma ecuación $E(x_{1t}\varepsilon_{1t}) = 0, E(x_{2t}\varepsilon_{2t}) = 0$, lo que garantizaría la consistencia del estimador de mínimos cuadrados. Pero en este modelo: $x_{1t} = (y_{2t}, y_{1t-1}, y_{2t-1}, z_t)$, y $x_{2t} = (y_{1t}, y_{1t-1}, y_{2t-1}, z_t)$, por lo que las condiciones necesarias para la consistencia se incumplen. Por tanto, habría que esperar que dicho estimador fuese en este caso inconsistente.

La primera dificultad podría resolverse estimando por variables instrumentales, siempre que contemos con instrumentos adecuados, lo cual no es sencillo de justificar. Por eso, para evitarla, transformamos el modelo. Pero la segunda dificultad persistirá, y se debe resolver tratando de ampliar la estructura dinámica del modelo hasta lograr que los términos de error carezcan de autocorrelación.

Supongamos que la matriz B tiene inversa, lo cual requiere que $\det(B) \neq 0$. Tendríamos entonces:

$$y_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 y_{t-1} + B^{-1}Gz_t + B^{-1}\varepsilon_t = A_0 + A_1 y_{t-1} + Mz_t + u_t \quad (2)$$

con $A_0 = B^{-1}\Gamma_0, A_1 = B^{-1}\Gamma_1, M = B^{-1}G, u_t = B^{-1}\varepsilon_t$. De este modo habremos pasado a la forma reducida, o *modelo vectorial autoregresivo (VAR)*, en la cual los coeficientes de los valores actuales de y_t son la matriz identidad:

$$\begin{aligned} y_{1t} &= \beta_{10} + \beta_{11}y_{1t-1} + \beta_{12}y_{2t-1} + m_{11}z_t + u_{1t} \\ y_{2t} &= \beta_{20} + \beta_{21}y_{1t-1} + \beta_{22}y_{2t-1} + m_{21}z_t + u_{2t} \end{aligned} \quad (3)$$

3 Modelos VAR

3.1 El modelo VAR(1)

El modelo vectorial autoregresivo (VAR) tiene un orden, que es el número de retardos con que las variables entran en cada ecuación. Así, los modelos que hasta ahora hemos ido describiendo son distintas variantes del modelo VAR(1).

En el caso de dos variables (modelo $VAR_2(1)$) la condición suficiente para poder realizar la transformación mencionada sería $\alpha_{11}\alpha_{21} \neq 1$. De no cumplirse,

ambas ecuaciones serían indistinguibles una de otra. La relación entre parámetros de la forma reducida y de la forma estructural es entonces:

$$B^{-1} = \frac{1}{1 - \alpha_{11}\alpha_{21}} \begin{pmatrix} 1 & \alpha_{11} \\ \alpha_{21} & 1 \end{pmatrix}$$

$$A_1 = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} = \frac{1}{1 - \alpha_{11}\alpha_{21}} \begin{pmatrix} \alpha_{12} + \alpha_{11}\alpha_{22} & \alpha_{13} + \alpha_{11}\alpha_{23} \\ \alpha_{22} + \alpha_{21}\alpha_{12} & \alpha_{23} + \alpha_{13}\alpha_{21} \end{pmatrix} \quad (4)$$

$$A_0 = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} = \frac{1}{1 - \alpha_{11}\alpha_{21}} \begin{pmatrix} \alpha_{10} + \alpha_{11}\alpha_{20} \\ \alpha_{20} + \alpha_{21}\alpha_{10} \end{pmatrix}; M = \frac{1}{1 - \alpha_{11}\alpha_{21}} \begin{pmatrix} \gamma'_1 + \alpha_{11}\gamma'_2 \\ \alpha_{21}\gamma'_1 + \gamma'_2 \end{pmatrix} \quad (5)$$

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = B^{-1}\varepsilon_t = B^{-1} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \frac{1}{1 - \alpha_{11}\alpha_{21}} \begin{pmatrix} \varepsilon_{1t} + \alpha_{11}\varepsilon_{2t} \\ \alpha_{21}\varepsilon_{1t} + \varepsilon_{2t} \end{pmatrix} \quad (6)$$

Aunque esta es la versión más general del modelo VAR, es habitual hacer supuestos simplificadores.¹ Para simplificar la notación supondremos que no hay variables exógenas o deterministas en el modelo VAR: $G = 0_k$. Sin embargo, todos los resultados que vamos a desarrollar se extienden sin dificultad cuando tales variables están presentes en el modelo. Asimismo, debe notarse que hemos introducido desde el comienzo el supuesto normalizador de que la matriz B tenga unos en su diagonal principal.

Suponemos asimismo que la matriz de covarianzas de las innovaciones ε_t del modelo VAR, Σ_ε , es diagonal, es decir que las innovaciones asociadas a distintas variables tienen covarianza nula, puesto que la correlación entre y_{1t} e y_{2t} ya está recogida por la presencia de cada una de estas variables en la ecuación de la otra variable en el modelo estructural. De este modo, la posible correlación entre las dos variables del modelo (pensemos en las rentabilidades del Ibex35 y de su futuro) están explicadas por el efecto contemporáneo de cada una de ellas sobre la otra [ver Apéndice 1].

Estos supuestos simplificadores anteriores nos llevan a un *modelo estructural* dinámico de la forma,

$$\begin{aligned} y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + \varepsilon_{1t} \\ y_{2t} &= \alpha_{20} + \alpha_{21}y_{1t} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{2t-1} + \varepsilon_{2t} \end{aligned} \quad (7)$$

y a su correspondiente forma reducida o modelo VAR₂(1) :

$$\begin{aligned} y_{1t} &= \beta_{10} + \beta_{11}y_{1t-1} + \beta_{12}y_{2t-1} + u_{1t} \\ y_{2t} &= \beta_{20} + \beta_{21}y_{1t-1} + \beta_{22}y_{2t-1} + u_{2t} \end{aligned} \quad (8)$$

o, en forma matricial,

¹En realidad, ya hemos introducido uno de tales supuestos al hacer que la matriz B tenga unos en su diagonal principal.

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (9)$$

o:

$$y_t = A_0 + A_1 y_{t-1} + u_t \quad (10)$$

cuyos términos de error satisfacen:

$$E(u_{1t}) = E(u_{2t}) = 0, \forall t \quad (11)$$

$$E(u_{1t}u_{1s}) = E(u_{2t}u_{2s}) = E(u_{1t}u_{2s}) = 0, \forall t \neq s$$

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = B^{-1} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \frac{1}{1 - \alpha_{11}\alpha_{21}} \begin{pmatrix} \varepsilon_{1t} + \alpha_{11}\varepsilon_{2t} \\ \varepsilon_{2t} + \alpha_{21}\varepsilon_{1t} \end{pmatrix} \quad (12)$$

$$\Sigma_u = Var \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \quad (13)$$

$$= \frac{1}{(1 - \alpha_{11}\alpha_{21})^2} \begin{pmatrix} \sigma_{\varepsilon_1}^2 + \alpha_{11}^2 \sigma_{\varepsilon_2}^2 & \alpha_{21} \sigma_{\varepsilon_1}^2 + \alpha_{11} \sigma_{\varepsilon_2}^2 \\ \alpha_{21} \sigma_{\varepsilon_1}^2 + \alpha_{11} \sigma_{\varepsilon_2}^2 & \alpha_{21}^2 \sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 \end{pmatrix} \quad (14)$$

Es importante observar que las innovaciones del modelo VAR estarán correlacionadas entre sí, $\sigma_{u_1 u_2} \neq 0$, incluso si las innovaciones del modelo estructural están incorrelacionadas, $\sigma_{\varepsilon_1 \varepsilon_2} = 0$, como hemos supuesto en la expresión anterior. La única excepción requeriría $\alpha_{11} = \alpha_{21} = 0$, el caso en que no hay efectos contemporáneos de ninguna variable sobre la otra. Únicamente en este caso límite tendríamos $\sigma_{u_1 u_2} = 0$.

En este modelo VAR, valores negativos de β_{12} y β_{21} tienden a inducir correlación negativa entre y_{1t} e y_{2t} , si bien no la garantizan, y valores positivos de β_{12} y β_{21} tienden a generar correlación positiva. Un *shock* inesperado en y_{2t} , en la forma de un valor no nulo de la innovación u_{2t} , además de afectar a y_{2t} , también influye sobre y_1 en períodos futuros, debido a la presencia del retardo y_{2t-1} como variable explicativa en la ecuación de y_{1t} . Por otra parte, dada la correlación existente entre ambos términos de error, un valor no nulo de u_{2t} vendrá habitualmente acompañado de un valor positivo o negativo (según sean los signos de u_{2t} y de σ_{12}) de u_{1t} , por lo que la reacción de y_{1t} vendrá acompañada también de una reacción de y_{2t} .

Dada la relación existente entre los vectores ε_t y u_t , si los términos de error del modelo estructural eran ruido blanco, los términos de error del modelo VAR también tendrán estructura de ruido blanco: $E(u_{1t} \cdot u_{1t-s}) = 0 \forall s \neq 0$. Es asimismo importante examinar las relaciones entre los parámetros de ambos modelos, que son, en el caso del modelo VAR(1), las 6 relaciones entre los parámetros β y los parámetros α que aparecen en (4) y (5), más las 3 relaciones entre los elementos de las respectivas matrices de covarianzas,

$$\begin{aligned}\sigma_{u_1}^2 &= \frac{1}{(1 - \alpha_{11}\alpha_{21})^2} (\sigma_{\varepsilon_1}^2 + \alpha_{11}^2\sigma_{\varepsilon_2}^2) \\ \sigma_{u_2}^2 &= \frac{1}{(1 - \alpha_{11}\alpha_{21})^2} (\sigma_{\varepsilon_2}^2 + \alpha_{21}^2\sigma_{\varepsilon_1}^2) \\ \sigma_{u_1u_2} &= \frac{1}{(1 - \alpha_{11}\alpha_{21})^2} (\alpha_{21}\sigma_{\varepsilon_1}^2 + \alpha_{11}\sigma_{\varepsilon_2}^2)\end{aligned}$$

3.2 Modelos VAR de orden superior

Los modelos anteriores se dicen que son de orden 1 porque en ellos las variables explicativas aparecen unicamente con un retardo. En general, un modelo VAR de orden n , con variables endógenas, se especifica,

$$Y_t = A_0 + \sum_{s=1}^n A_s Y_{t-s} + GW_t + u_t \quad (15)$$

donde Y_t es un vector columna $k \times 1$, n es el orden del modelo VAR, o número de retardos de cada variable en cada ecuación, y u_t es un vector $k \times 1$ de innovaciones, es decir, procesos sin autocorrelación, con $Var(u_t) = \Sigma$, constante. W_t es un vector de variables exógenas.

A lo largo de la exposición consideraremos generalmente, unicamente por simplicidad, que no hay variables exógenas en el modelo, $G = 0_k$.

El elemento (i, j) en la matriz A_s , $1 \leq s \leq n$ mide el *efecto directo o parcial* de un cambio unitario en Y_j en el instante t sobre los valores numéricos de Y_i al cabo de s períodos, es decir, sobre el vector $Y_{i,t+s}$. La columna j de la matriz A_s mide el efecto que un cambio unitario en Y_j en el instante t tiene sobre el vector Y_{t+s} . El elemento i -ésimo en u_t es el componente de Y_{it} que no puede ser previsto utilizando el pasado de las variables que integran el vector Y_t .

Algunas consideraciones sobre la estimación de un modelo VAR:

- Como puede verse, en un modelo VAR todas las variables son tratadas simétricamente, siendo explicadas por el pasado de todas ellas. El modelo tiene tantas ecuaciones como variables, y los valores retardados de todas las ecuaciones aparecen como variables explicativas en todas las ecuaciones.
- Una vez estimado el modelo, puede procederse a excluir algunas variables explicativas, en función de su significación estadística, pero hay razones para no hacerlo. Por un lado, si se mantiene el mismo conjunto de variables explicativas en todas las ecuaciones, entonces la estimación por mínimos cuadrados ordinarios ecuación por ecuación es eficiente, por lo que el proceso de estimación del modelo es verdaderamente sencillo. Por otro, la presencia de bloques de retardos como variables explicativas hace que la colinealidad entre variables explicativas sea importante, lo que hace perder precisión en la estimación del modelo y reduce los valores numéricos de los

estadísticos tipo t de Student. Por tanto, no es buena estrategia proceder en varias etapas, excluyendo del modelo las variables cuyos coeficientes resultan estadísticamente no significativos, por cuanto que esto puede ser consecuencia de la colinealidad inherente al modelo, y no tanto de la falta de contenido informativo de las variables.

- En el modelo VAR pueden estimarse con bastante precisión los elementos globales del modelo, como el R^2 , la desviación típica residual, y los mismos residuos, o el efecto global de una variable sobre otra, lo que se resume en los contrastes de causalidad que veremos más adelante. Sin embargo, no cabe hacer interpretaciones de coeficientes individuales en distintos retardos, ni llevar a cabo contrastes de hipótesis sobre coeficientes individuales.
- En un modelo vectorial autoregresivo de primer orden, $\text{VAR}_2(1)$, las variables explicativas de cada ecuación son: una constante, más un retardo de cada una de las variables del modelo, más 3 parámetros en la matriz de covarianzas, con un total de 9 parámetros para explicar el movimiento conjunto de 2 variables. En el modelo $\text{VAR}_3(1)$, que explica el comportamiento temporal de 3 variables, habría 3 variables explicativas retardadas, más una constante, en cada ecuación, para un total de 12 coeficientes a estimar, 4 en cada ecuación, más los 6 elementos de la matriz de covarianzas de las innovaciones (18 parámetros en total). Si trabajamos con 4 variables y establecemos un orden 3 para el VAR, tendremos 12 variables explicativas, más el término constante, en cada ecuación, con un total de 52 coeficientes en el sistema de ecuaciones, más 10 parámetros en la matriz de varianzas-covarianzas de las innovaciones. El número de parámetros a estimar crece muy rápidamente con el número de retardos. Si pasamos de 3 a 4 retardos, tendríamos 68 coeficientes más los 10 parámetros de la matriz de covarianzas. Como se ve, el número de parámetros a estimar en un modelo VAR aumenta muy rápidamente con el orden del mismo o con el número de ecuaciones.

3.3 VAR y modelos univariantes

Es útil asimismo pensar en términos de cuáles son los modelos univariantes que se deducen de una representación VAR, en línea con el trabajo de Zellner y Palm (19xx). En este sentido, si partimos de un $\text{VAR}_2(1)$, como (9), escrito en función del operador de retardos,

$$\begin{aligned} y_{1t} &= \beta_{10} + \beta_{11}Ly_{1t} + \beta_{12}Ly_{2t} + u_{1t} \\ y_{2t} &= \beta_{20} + \beta_{21}Ly_{1t} + \beta_{22}Ly_{2t} + u_{2t} \end{aligned}$$

tenemos,

$$y_{2t} = \frac{\beta_{20} + \beta_{21}Ly_{1t} + u_{2t}}{1 - \beta_{22}L}$$

con lo que,

$$(1 - \beta_{11}L) y_{1t} = \beta_{10} + \beta_{12}L \frac{\beta_{20} + \beta_{21}Ly_{1t} + u_{2t}}{1 - \beta_{22}L} + u_{1t}$$

y, finalmente,

$$[(1 - \beta_{11}L)(1 - \beta_{22}L) - \beta_{12}\beta_{21}L^2] y_{1t} = [(1 - \beta_{22})\beta_{10} + \beta_{12}\beta_{20}] + [(1 - \beta_{22}L)u_{1t} + \beta_{12}u_{2t-1}]$$

que es un proceso ARMA(2,1) con término de error (no innovación): $\xi_{1t} = (1 - \beta_{22}L)u_{1t} + \beta_{12}u_{2t-1}$.

Las dos variables tienen la misma ecuación característica, e iguales propiedades dinámicas, por tanto. Nótese que el hecho de que el modelo VAR sea estable garantiza que los modelos univariantes con estructura AR(2) sean asimismo estables, pues ambas raíces quedan fuera del círculo unidad.

3.4 Estimación y contrastación de hipótesis

Como ya hemos mencionado, en ausencia de restricciones, la estimación por mínimos cuadrados, ecuación por ecuación, de un modelo VAR produce *estimadores eficientes* a pesar de que ignora la información contenida en la matriz de covarianzas de las innovaciones. Junto con el hecho de que la colinealidad entre las variables explicativas no permite ser muy estricto en la interpretación de los estadísticos t , sugiere que es preferible mantener todas las variables explicativas iniciales en el modelo.

El estimador es *consistente* siempre que los términos de error sean innovaciones, es decir, procesos ruido blanco, pues en tal caso, estarán incorrelacionados con las variables explicativas, por la misma razón que en un modelo univariante. Por tanto, la ausencia de autocorrelación en los términos de error de todas las ecuaciones es muy importante. Tomando ambos hechos conjuntamente, es fácil concluir que debe incluirse en cada ecuación, como variables explicativas, el menor número de retardos que permita eliminar la autocorrelación residual en todas las ecuaciones.

Un modelo VAR no se estima para hacer inferencia acerca de coeficientes de variables individuales. Precisamente la baja precisión en su estimación, desaconseja cualquier análisis de coeficientes individuales. Tiene mucho sentido, por el contrario, el análisis conjunto de los coeficientes asociados a un bloque de retardos en una determinada ecuación.

Existen contrastes del tipo de razón de verosimilitud sobre el número de retardos a incluir en el modelo.

Bajo la hipótesis de Normalidad del vector de innovaciones, el logaritmo de la función de verosimilitud es,

$$L = \prod_{t=p}^T \frac{1}{2\pi^{k/2} |\Sigma|} e^{-\frac{1}{2} \hat{u}_t' \Sigma^{-1} \hat{u}_t}$$

$$l = \ln L = -\frac{Tk}{2} \ln(2\pi) - \frac{T}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^T \hat{u}_t' \Sigma^{-1} \hat{u}_t$$

y, tras estimar $\hat{\Sigma}$ mediante la matriz de covarianzas del vector de residuos \hat{u} ,

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T \hat{u}_t \hat{u}_t' \Rightarrow \hat{u}_t' \hat{\Sigma}^{-1} \hat{u}_t = k$$

llegamos a:

$$l = -\frac{Tk}{2} (1 + \ln 2\pi) - \frac{T}{2} \ln |\Sigma|$$

3.4.1 Contrastes de especificación

Uno de los contratos más habituales en un modelo VAR es el relativo al número de retardos que deben incluirse como variables explicativas. Hay que tener en cuenta que en cada ecuación entra un bloque de retardos de todas las variables del vector y . Ya hemos visto que el número de parámetros a estimar crece rápidamente con el número de variables del modelo. Para evitar una excesiva sobreparametrización del modelo VAR, una estrategia razonable consiste en incluir en cada ecuación del modelo el *menor* número de retardos que permita eliminar la autocorrelación del término de error de todas ellas.

Existe un contraste formal de significación de un conjunto de retardos, que utiliza un estadístico de razón de verosimilitudes,

$$\lambda = (T - k)(\ln |\Sigma_R| - \ln |\Sigma_{SR}|)$$

donde $|\Sigma_R|$, $|\Sigma_{SR}|$ denotan los determinantes de las matrices de covarianzas de los modelos restringido y sin restringir, respectivamente. Si queremos contrastar si un cuarto retardo es significativo, deberíamos estimar el modelo con 3 y con 4 retardos, y construir el estadístico anterior, que tiene una distribución chi-cuadrado con un número de grados de libertad igual al número de restricciones que se contrastan. Al pasar del modelo con 3 retardos al modelo con 4 retardos, hay que añadir un retardo más de cada variable en cada ecuación, por lo que el número de restricciones es igual al incremento en el número de retardos, multiplicado por el número de variables, elevado al cuadrado.

Sin embargo, no puede olvidarse que la elección del número de retardos debe tener muy en cuenta la eliminación de autocorrelación en los residuos. Los estadísticos anteriores no examinan este importante aspecto y, por tanto, no deben utilizarse por sí solos. En consecuencia, una buena estrategia es comenzar de un número reducido de retardos, y examinar las funciones de autocorrelación de los residuos, junto con estadísticos del tipo Ljung-Box o Box-Pierce para contrastar

la posible existencia de autocorrelación, lo que requeriría aumentar el número de retardos y con ello, el número de parámetros a estimar. Lamentablemente, sin embargo, es muy poco probable que pueda eliminarse la autocorrelación residual con menos de 4 retardos cuando se trabaja con datos trimestrales, o con menos de 12 retardos, cuando se trabaja con datos mensuales.

Una estrategia distinta para encontrar el orden del modelo VAR consiste en examinar los denominados *criterios de Información*, que son determinadas correcciones sobre el valor muestral de la función logaritmo de Verosimilitud. Los más conocidos son los de Akaike (AIC) y Schwartz (SBC o BIC),

$$\begin{aligned} AIC &= -2\frac{l}{T} + 2\frac{p}{T} \\ SBC &= -2\frac{l}{T} + p\frac{\ln(T)}{T} \\ Hannan - Quinn &= -2\frac{l}{T} + 2\frac{k\ln(\ln(T))}{T} \end{aligned}$$

siendo $l = -\frac{Tk}{2}(1 + \ln 2\pi) - \frac{T}{2} \ln |\hat{\Sigma}|$, y $p = k(d+nk)$ el número de parámetros estimados en el modelo VAR, siendo d es el número de variables exógenas, n el orden del VAR, k el número de variables endógenas.

En ocasiones, se ignora el término constante de la función de verosimilitud, y los criterios anteriores se aproximan por:

$$\begin{aligned} AIC &= T \cdot \ln(|\Sigma|) + 2p \\ SBC &= T \cdot \ln(|\Sigma|) + p \ln(T) \\ Hannan - Quinn &= 2 \ln(|\Sigma|) + 2k \ln(\ln(T)) \end{aligned}$$

siendo N el número de parámetros que se estima, y Σ la matriz de covarianzas de los residuos. Nótese que, en ocasiones, los criterios se dan por observación (como en el primer grupo que hemos presentado), y en otras se dan de modo agregado para toda la muestra (como en el segundo grupo). Estos estadísticos se calculan para una sucesión de modelos con distinto número de retardos y se comparan, seleccionando aquél modelo que produce un *menor* valor del estadístico. Pero no cabe esperar que los tres criterios conduzcan a la misma conclusión.

Un estadístico de razón de verosimilitudes como el antes descrito puede utilizarse para contrastar cualquier tipo de hipótesis, y no sólo la significación de grupos de variables, siempre que el modelo restringido esté anidado dentro del modelo sin restringir.

3.4.2 Contrastes de causalidad

Un contraste especialmente interesante es el conoce como de causalidad en el sentido de Granger: supongamos que estamos explicando el comportamiento

de una variable y utilizando su propio pasado. Se dice que *una variable z no causa a la variable y* si al añadir el pasado de z a la ecuación anterior no añade capacidad explicativa. El contraste consiste en analizar la significación estadística del bloque de retardos de z en la ecuación mencionada, y la hipótesis nula es que la variable z *no causa*, en el sentido de Granger, a la variable y .

En realidad, la propuesta inicial de Granger hacía referencia a que la predicción de y basada en el pasado de las dos variables y y z , sea estrictamente mejor (es decir, con menos error) que la predicción de y basada exclusivamente en su propio pasado. Así, se diría que la *variable z no causa a la variable y* si se tiene,

$$E(y_t / y_{t-1}, y_{t-2}, \dots; z_{t-1}, z_{t-2}, \dots) = E(y_t / y_{t-1}, y_{t-2}, \dots)$$

Sin embargo, esta propiedad no suele analizarse utilizando predicciones. Se contrasta exclusivamente la significación del bloque de retardos de z en la ecuación de y , y se supone que si dicho bloque de variables es significativo, contribuirá a mejorar la predicción de la variable y . Esta manera de proceder se basa en que, analíticamente, es evidente que la presencia del bloque de retardos de z en la ecuación de y hace que la esperanza de y condicional en el pasado de las dos variables, y y z , sea distinta de la esperanza de y condicional en su propio pasado exclusivamente, si bien esta propiedad teórica no siempre se manifiesta en resultados prácticos, y es bien sabido que un buen ajuste no necesariamente conduce a una buena predicción.

El contraste puede llevarse a cabo utilizando el estadístico F habitual en el contraste de significación de un bloque de variables, o mediante el estadístico de razón de verosimilitudes anterior. Con más de dos variables, existen muchos posibles contrastes de causalidad y en algunos casos, el estadístico de razón de verosimilitudes puede resultar más útil que el estadístico F , al permitir contrastar la exclusión de algún bloque de retardos en varias ecuaciones *simultáneamente*. Asimismo, el contraste de causalidad o, lo que es lo mismo, el contraste de significación de un bloque de retardos puede llevarse a cabo mediante un estadístico de razón de verosimilitudes, en el que el modelo restringido excluye un grupo de retardos de una ecuación.

3.5 Identificación en un modelo VAR

Consideremos el caso sencillo de un modelo VAR₂(1) bivalente (8). Su estimación proporciona valores numéricos para 9 parámetros: las dos constantes más los cuatro coeficientes en las variables retardadas, más los 3 parámetros de la matriz de covarianzas (13) del vector u_t . Sin embargo, el modelo estructural consta de 10 parámetros: las dos constantes, los 6 coeficientes, y los 2 parámetros de la matriz de covarianzas del vector ε_t (sus varianzas), por lo que no es posible recuperar valores numéricos para todos los parámetros del modelo estructural a partir de la estimación de la forma reducida. A este proceso de recuperación de parámetros se le conoce como *identificación* del modelo VAR.

Si se examina la relación que existe entre un modelo VAR y el modelo estructural del cual procede, se aprecia que el problema de identificación se reduce a

obtener las innovaciones estructurales a partir de los residuos del modelo VAR. Por tanto, identificar el modelo es encontrar valores numéricos para los elementos de la matriz B que define la transformación: $\varepsilon_t = Bu_t$. Esta matriz tiene unos en la diagonal principal, pero no es simétrica, por lo que tiene $k^2 - k$ parámetros por determinar. Además, debemos encontrar las k varianzas de las innovaciones estructurales ε_t ; es habitual suponer que sus covarianzas son nulas. Así, tenemos $k^2 - k + k = k^2$ parámetros del modelo estructural, que querríamos recuperar a partir de los $(k^2 + k)/2$ elementos de $Var(u_t)$. Necesitamos, por tanto, $(k^2 - k)/2 = k(k - 1)/2$ restricciones, si queremos tener alguna posibilidad de identificar el modelo. En el caso de un modelo VAR(1) con 2 variables, hemos de imponer $(2^2 - 2)/2 = 1$ restricción para identificar el sistema exactamente, como hemos constatado en los ejemplos anteriores. En un modelo con 3 variables necesitaríamos imponer $(3^2 - 3)/2 = 3$ restricciones. El número de restricciones necesarias para identificar el modelo es independiente del orden de retardos del modelo VAR.

Este análisis se refiere a la versión simplificada del modelo VAR que estamos considerando. Bajo el supuesto de que Σ_ε es diagonal ya estimamos k parámetros pero si la matriz de covarianzas Σ_ε no fuese diagonal, tendríamos que estimar otros $k(k - 1)/2$ parámetros adicionales a estimar. Esto se debe a que la matriz de covarianzas Σ_ε tendría entonces $k(k + 1)/2$ elementos distintos.² La inclusión de variables exógenas no empeora esta situación, pues estimaríamos tantos coeficientes para ellas en el VAR como parámetros debemos recuperar en la forma estructural.

La denominación de "VAR estructural" que utilizamos en esta sección no debe confundirse con "un modelo estructural" que consideramos en la primera sección. Son denominaciones comunmente aceptadas. Denominamos modelo "estructural" a todo modelo cuyas ecuaciones tienen una interpretación en cuanto a que reflejan relaciones entre variables motivadas por la teoría económica. Es un modelo en el que podemos hablar de una "función de consumo", una función o ecuación de inversión", una "función de demanda de dinero", etc.. El modelo de la primera sección era "estructural" pero no era un modelo VAR. Se denomina VAR estructural a un modelo VAR al que se ha añadido una determinada estrategia de identificación, supuestamente basada en conceptos de teoría económica o financiera, o a características del mercado en estudio.

Una vez que hubiésemos escogido una determinada especificación de la matriz B , es decir, un determinado conjunto de condiciones de identificación, podremos pasar del modelo VAR estimado:

$$y_t = A_0 + A_1 y_{t-1} + M z_t + u_t \quad (16)$$

al modelo estructural que supuestamente subyace al VAR, mediante:

$$By_t = (BA_0) + (BA_1) y_{t-1} + (BM) z_t + Bu_t = \Gamma_0 + \Gamma_1 y_{t-1} + G z_t + \varepsilon_t$$

²Luego estaríamos añadiendo un número de parámetros igual a la diferencia $k(k+1)/2 - k = k(k - 1)/2$.

debiendo quedar claro que hay más de una forma de hacer esta transformación, por lo que hay más de un modelo estructural compatible con un determinado modelo VAR estimado. De hecho, existen tantas maneras de identificar el modelo estructural como conjuntos de $k(k-1)/2$ restricciones podamos introducir sobre dicho modelo.

3.5.1 Algunos ejemplos

Que haya tantos parámetros estimados como parámetros queremos recuperar (estimar) en la forma estructural es una condición necesaria, aunque no suficiente, para lograr la identificación del modelo estructural. En el ejercicio 1 se prueba que el modelo estructural recursivo bivalente de orden 1,

$$\begin{aligned} y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + \varepsilon_{1t} \\ y_{2t} &= \alpha_{20} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{2t-1} + \varepsilon_{2t} \end{aligned} \quad (17)$$

está exactamente identificado, es decir, que sus parámetros pueden recuperarse de forma única a partir de las estimaciones del modelo VAR asociado. Este es un modelo interesante, en el que se consigue identificar todos los parámetros del modelo estructural a partir de las estimaciones de la forma reducida (modelo VAR), introduciendo la hipótesis de que la variable y_{1t} afecta a la variable y_{2t} únicamente con un retardo, mientras que la dirección de influencia de y_{2t} hacia y_{1t} se manifiesta ya dentro del mismo período.

En este modelo se tiene:

$$\begin{aligned} \beta_{11} &= \alpha_{12} + \alpha_{11}\alpha_{22}; \beta_{12} = \alpha_{13} + \alpha_{11}\alpha_{23}; \\ \beta_{21} &= \alpha_{22}; \beta_{22} = \alpha_{23}; \beta_{10} = \alpha_{10} + \alpha_{11}\alpha_{20}; \beta_{20} = \alpha_{20}; \beta_{22} = \alpha_{23}; \\ \sigma_{u_1}^2 &= \sigma_{\varepsilon_1}^2 + \alpha_{11}^2\sigma_{\varepsilon_2}^2; \sigma_{u_2}^2 = \sigma_{\varepsilon_2}^2; \sigma_{u_1u_2} = \alpha_{11}\sigma_{\varepsilon_2}^2; \end{aligned}$$

que puede resolverse, con solución única, para encontrar los coeficientes α junto con $\sigma_{\varepsilon_1}^2$ y $\sigma_{\varepsilon_2}^2$.

En este modelo no sólo se pueden recuperar estimaciones de todos los parámetros que aparecen en el modelo estructural. También las series temporales de los residuos del modelo estructural pueden recuperarse a partir de los residuos obtenidos en la estimación del modelo VAR, mediante,

$$\begin{aligned} \hat{\varepsilon}_{2t} &= \hat{u}_{2t}; \\ \hat{\varepsilon}_{1t} + \alpha_{11}\hat{\varepsilon}_{2t} &= \hat{u}_{1t} \Rightarrow \hat{\varepsilon}_{1t} = \hat{u}_{1t} - \alpha_{11}\hat{u}_{2t} \end{aligned}$$

Un modelo más restringido,

$$\begin{aligned} y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + \varepsilon_{1t} \\ y_{2t} &= \alpha_{20} + \alpha_{23}y_{2t-1} + \varepsilon_{2t} \end{aligned}$$

implicaría que la variable y_{1t} no afecta ni de forma contemporánea, ni retardada, a la variable y_{2t} , por lo que ésta puede considerarse *exógena*³ respecto de y_{1t} . Estas dos restricciones, $\alpha_{21} = \alpha_{22} = 0$, hacen que en el modelo VAR, $\beta_{21} = 0$, restricción que debería resultar evidente al estimar dicho coeficiente. Al haber introducido una restricción más, el modelo estructural está ahora *sobreidentificado*, es decir, hay más de una manera de recuperar valores numéricos para los parámetros de dicho modelo a partir de las estimaciones numéricas del modelo VAR.

Más dificultades plantea el modelo,

$$\begin{aligned} y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \varepsilon_{1t} \\ y_{2t} &= \alpha_{20} + \alpha_{21}y_{1t} + \alpha_{23}y_{2t-1} + \varepsilon_{2t} \end{aligned}$$

que está asimismo *sobreidentificado*, habiendo varias maneras de recuperar las estimaciones de los parámetros del modelo estructural. Sin embargo, en este caso no hay ninguna restricción contrastable sencilla que nos permita discutir esta representación. En este caso, las restricciones del modelo estructural introducen restricciones no lineales entre los parámetros del modelo VAR. Una posible estrategia consiste en estimar el modelo VAR sujeto a las restricciones no lineales generadas por las condiciones de sobreidentificación.

Si imponemos condiciones de *recursividad* en un modelo con 3 variables, tenemos,

$$\begin{aligned} u_{1t} &= \varepsilon_{1t} \\ u_{2t} &= c_{21}\varepsilon_{1t} + \varepsilon_{2t} \\ u_{3t} &= c_{31}\varepsilon_{1t} + c_{32}\varepsilon_{2t} + \varepsilon_{3t} \end{aligned}$$

que implica imponer 3 restricciones sobre los elementos de la matriz B^{-1} : $c_{12} = c_{13} = c_{23} = 0$. Con estas restricciones el modelo estaría exactamente identificado. La recursividad, que consiste en que la matriz B tiene estructura triangular inferior, siempre impone $(k^2 - k)/2$ restricciones. Luego veremos que este tipo de condiciones de identificación se conoce como identificación de Choleski.

Hay conjuntos alternativos de restricciones, como,

$$\begin{aligned} u_{1t} &= \varepsilon_{1t} + c_{13}\varepsilon_{3t} \\ u_{2t} &= c_{21}\varepsilon_{1t} + \varepsilon_{2t} \\ u_{3t} &= c_{32}\varepsilon_{2t} + \varepsilon_{3t} \end{aligned}$$

que también lograría la identificación exacta del modelo. Aquí estamos imponiendo: $c_{12} = c_{23} = c_{31} = 0$. Esta estructura podría obedecer al tipo de

³Una variable y_2 es exógena en la ecuación de otra variable y_{1t} , con innovación u_{1t} , si $E(y_{2t}u_{1s}) = 0 \forall t, s$. La variable y_2 es predeterminada en la ecuación de la variable y_{1t} si: $E(y_{2t}u_{1s}) = 0 \forall t \leq s$.

relaciones entre las tres variables sugerido por la teoría económica. La representación inversa es, en este caso:

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} = \frac{1}{1 + c_{21}c_{13}c_{32}} \begin{pmatrix} 1 & c_{13}c_{32} & -c_{13} \\ -c_{21} & 1 & c_{13}c_{21} \\ c_{21}c_{32} & -c_{32} & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}$$

Otro tipo de restricciones consistiría en imponer un determinado valor numérico para una respuesta. Por ejemplo, en un modelo VAR₂(1) podemos pensar que y_{2t} tiene un efecto unitario sobre y_{1t} , es decir, que $\alpha_{11} = 1$. En tal caso:

$$\varepsilon_t = Bu_t = \begin{pmatrix} 1 & -1 \\ -\alpha_{21} & 1 \end{pmatrix} u_t$$

Una posibilidad diferente consistiría en identificar el modelo estructural imponiendo restricciones sobre la matriz de covarianzas, ya sea imponiendo un valor numérico para la varianza de ε_{1t} , la varianza de ε_{2t} , o la relación entre ambas. Este tipo de restricciones conduce generalmente a soluciones múltiples, aunque en número finito, por lo que el modelo estructural está en tal caso, sobreidentificado.

Por último, puede conseguirse la identificación imponiendo restricciones razonables entre los valores numéricos de los parámetros estructurales. Por ejemplo, puede imponerse una *condición de simetría*, $\alpha_{11} = \alpha_{21}$, o cualquier otra que resulte adecuada en la aplicación que se analiza.

VAR estructural (omitir)

$$\begin{aligned} u_t^p &= \varepsilon_t^{as} \\ u_t^y &= B_1 u_t^p + B_2 u_t^r + B_3 u_t^m + \varepsilon_t^{is} \\ u_t^r &= A_4 u_t^m + \varepsilon_t^{ms} \\ u_t^m &= A_5 (u_t^y + u_t^p) + A_6 u_t^r + \varepsilon_t^{md} \end{aligned}$$

3.5.2 Identificación ortogonal: el enfoque Cholesky

Otra manera de entender los problemas de identificación es la siguiente: supongamos que, sin considerar el posible modelo estructural, hemos estimado un modelo VAR(1) bivariante (9) en el que queremos calcular cómo reacciona cada variable ante una innovación en una de ellas, lo que denominamos como *función de respuesta al impulso*. Sería poco adecuado, sin embargo, estimar las respuestas a un impulso en una de las innovaciones, u_1 , por ejemplo, sin que u_2 experimente ningún impulso, pues ambas innovaciones están correlacionadas entre sí. Por tanto, parece lógico transformar primero el modelo estimado en otro modelo en que las innovaciones del modelo estén incorrelacionadas entre sí.

Para ello, nos basamos en la descomposición LDL de una matriz simétrica, definida positiva. Dada una matriz simétrica, definida positiva, como es la matriz de covarianzas Σ_u , existe una única matriz triangular inferior C , con unos

en su diagonal principal, y una única matriz diagonal D , con elementos positivos a lo largo de su diagonal principal, tal que Σ_u admite la descomposición:

$$\Sigma_u = CDC'$$

La descomposición LDL está ligada a la descomposición de Cholesky, que garantiza que existe una matriz triangular inferior P , que es única en el caso en que la matriz que se quiere descomponer sea positiva definida, tal que: $\Sigma_u = PP'$.⁴ Nótese que a partir de la descomposición LDL podemos obtener la descomposición de Cholesky con $P = CD^{1/2}$: $\Sigma_u = CDC' = (CD^{1/2})(CD^{1/2})'$. En algunos textos es la descomposición LDL la que se denomina descomposición de Cholesky. El nombre que reciba es irrelevante, pero es crucial entender que la descomposición LDL permite transformar el modelo VAR en otro modelo con innovaciones incorrelacionadas con varianzas no unitarias, mientras que la descomposición de Cholesky permite obtener un modelo transformado con innovaciones incorrelacionadas y con varianzas unitarias. Esto afecta a las expresiones analíticas de los principales estadísticos del VAR, como veremos en las secciones siguientes, por lo que en el trabajo empírico habrá que estar muy atentos al tipo de identificación que se esté utilizando, con objeto de utilizar las matrices de covarianzas adecuadas.

Si consideramos la transformación lineal del vector de los términos de error del modelo VAR mediante la matriz C^{-1} , obtenemos:⁵

$$v_t = C^{-1}u_t,$$

$$\text{Var}(v_t) = E(v_t v_t') = E(C^{-1}u_t u_t' (C^{-1})') = C^{-1} \Sigma_u (C^{-1})' = D$$

por lo que, a diferencia de los componentes del vector u , los elementos del vector v están incorrelacionados entre sí, ya que su matriz de covarianzas es diagonal.

La matriz C y su inversa tienen la forma:

⁴Ejemplo: La descomposición de Cholesky de la matriz $\begin{pmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{pmatrix}$ es:

$$\begin{pmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 6 & 1 & 0 \\ -8 & 5 & 3 \end{pmatrix} \begin{pmatrix} 2 & 6 & -8 \\ 0 & 1 & 5 \\ 0 & 0 & 3 \end{pmatrix}$$

y la descomposición LDL:

$$\begin{pmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -4 & 5 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 9 \end{pmatrix} \begin{pmatrix} 1 & 3 & -4 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{pmatrix}$$

⁵En esta estrategia de identificación, v_t juega el papel de innovación estructural ε_t , y C^{-1} juega el papel de la matriz B de secciones anteriores.

$$C = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ c_{12} & 1 & 0 & \dots & 0 \\ c_{13} & c_{23} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ c_{1k} & c_{2k} & c_{3k} & \dots & 1 \end{pmatrix} \Rightarrow C^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -c_{12} & 1 & 0 & \dots & 0 \\ -c_{13} & -c_{23} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -c_{1k} & -c_{2k} & -c_{3k} & \dots & 1 \end{pmatrix}$$

por lo que:

$$\begin{aligned} v_{1t} &= u_{1t}, \\ v_{2t} &= u_{2t} - \hat{c}_{12}v_{1t}, \\ v_{3t} &= u_{3t} - \hat{c}_{13}v_{1t} - \hat{c}_{23}v_{2t}, \dots \\ &\dots \\ v_{kt} &= u_{kt} - \hat{c}_{1k}v_{1t} - \hat{c}_{2k}v_{2t} - \dots - \hat{c}_{k-1,k}v_{k-1,t} \end{aligned} \quad (18)$$

¿Quiénes son estas nuevas innovaciones v_{it} ? La primera innovación, v_{1t} , es igual a u_{1t} . La segunda innovación, v_{2t} , es el residuo de la regresión de mínimos cuadrados de u_{2t} sobre v_{1t} (o lo que es lo mismo, sobre u_{1t}): $v_{2t} = u_{2t} - c_{12}v_{1t}$. La tercera innovación, v_{3t} , es el residuo de la regresión de mínimos cuadrados de u_{3t} sobre v_{1t} y v_{2t} (o lo que es lo mismo, sobre u_{1t} y u_{2t} , ya que v_{1t} y v_{2t} contienen la misma información que u_{1t} y u_{2t}).⁶

Por construcción, el residuo de mínimos cuadrados en una regresión lineal tiene correlación nula con cada una de las variables explicativas, por lo que: $E(v_{kt} \cdot v_{1t}) = E(v_{kt} \cdot v_{2t}) = \dots = E(v_{kt} \cdot v_{k-1,t}) = 0$. Dicho de otra manera, si estimamos regresiones de cada innovación u_{it} sobre todas las que le preceden dentro del vector u y nos quedamos con el residuo de dicha regresión, vamos generando innovaciones v_{it} , cada una de las cuales será el componente de u_{it} que está incorrelacionado con $v_{1t}, v_{2t}, \dots, v_{i-1,t}$. Pero los valores numéricos que toman las posibles combinaciones lineales que podamos formar con las variables $u_{1t}, u_{2t}, \dots, u_{i-1,t}$ y con las variables $v_{1t}, v_{2t}, \dots, v_{i-1,t}$ son los mismos, es decir, que ambos conjuntos de variables contienen la misma información. La única diferencia entre ambos es que las variables $u_{1t}, u_{2t}, \dots, u_{i-1,t}$ tienen correlaciones no nulas, mientras que las variables $v_{1t}, v_{2t}, \dots, v_{i-1,t}$ están incorrelacionadas entre sí.

Estamos llevando a cabo un proceso de identificación del modelo VAR. Las nuevas innovaciones v_{it} van a ser las innovaciones del modelo estructural que identifiquemos, es decir, hacemos $\varepsilon_{it} = v_{it}$. El resultado será un modelo estructural compatible con el modelo VAR que hemos estimado. Esta transformación de los términos de error, de las u_{it} a las v_{it} , basada en la descomposición LDL de la matriz de covarianzas de los términos de error originales del VAR, permite identificar el modelo, pues la matriz C incorpora precisamente $(k^2 - k)/2$

⁶Los residuos de una regresión lineal estimada por mínimos cuadrados ordinarios no varían si llevamos a cabo una transformación lineal de las variables explicativas.

restricciones. El procedimiento introduce una ordenación de variables, al conceder a los términos de error transformados una relevancia distinta: una vez ordenadas las variables del vector Y_t , el primer error transformado v_{1t} coincide con el antiguo, u_{1t} . El segundo, v_{2t} , es la parte de u_{2t} no explicada por u_{1t} , y así sucesivamente.

La ordenación que se establezca entre las variables condiciona los resultados que posteriormente se obtengan. Cuanto más correlacionadas estén las variables, más dependientes son los resultados de la ordenación de variables; cuanto menores sean las correlaciones entre las variables de Y_t , menos dependerán los resultados de la ordenación establecida. En presencia de correlaciones significativas no cabe sino analizar el modelo con distintas ordenaciones y tratar de interpretar los resultados obtenidos. En algunos casos, la propia naturaleza de las variables hará natural pensar que una variable se mueve en anticipación a la otra, o que es causal-previa a la otra.

En ocasiones, la transformación que pasa de las innovaciones u_{it} a las ε_{jt} se basa en la descomposición de Cholesky:

$$\Sigma_u = CD^{1/2}D^{1/2}C'$$

y consiste en: $w_t = D^{-1/2}C^{-1}u_t = P^{-1}u_t = D^{-1/2}v_t$, de modo que la matriz de covarianzas del vector transformado w_t es en este caso no solo es una matriz diagonal, sino que es igual a la matriz identidad, es decir todas las varianzas son iguales a uno:

$$Var(w) = D^{-1/2}Var(v_t)D^{-1/2} = D^{-1/2}DD^{-1/2} = I_k$$

En lo sucesivo, utilizamos la notación ε_{it} o v_{it} para denotar innovaciones estructurales con varianza no unitaria, y la notación w_{it} para denotar innovaciones con varianza unitaria.

Un shock unitario en w_{it} es igual a un shock en ε_{it} de tamaño igual a σ_{ε_i} , la desviación típica de ε_{it} . Las respuestas del sistema a un shock unitario en ε_{it} son respuestas a shocks unitarios en y_{it} , mientras que las respuestas a w_{it} son respuestas a shocks de magnitud σ_{ε_i} en y_{it} . Lógicamente, las respuestas a shocks en w_{it} son iguales a las respuestas a shocks en ε_{it} , multiplicadas por σ_{ε_i} .

Cuando las variables que se incluyen en el VAR tienen varianzas diferentes, esta transformación facilita la comparación al calcular respuestas a shocks de tamaño comparable en cada variable. Son comparables las respuestas a shocks de tamaño una desviación típica de cada innovación; es decir, las respuestas a un shock unitario en las innovaciones w_{it} .⁷ No son comparables las respuestas a un shock unitario en cada innovación ε_{it} , porque éste puede ser un shock grande para algunas variables y muy reducido para otras, dependiendo de cuáles sean sus varianzas respectivas.

Por ejemplo, en el caso del modelo VAR₂(1), la descomposición LDL será:

⁷Recordemos que la desviación típica es una medida adecuada del tamaño de toda variable aleatoria de esperanza nula.

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} m & 0 \\ 0 & n \end{pmatrix} \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}$$

y es sencillo probar que debe ser: $m = \sigma_1^2$, $n = \sigma_2^2 - \sigma_{12}^2/\sigma_1^2$, $c = \sigma_{12}/\sigma_1^2$. Por tanto, la constante c no es sino el estimador de mínimos cuadrados en la regresión de u_{2t} sobre u_{1t} . La relación entre innovaciones de la forma reducida y de la forma estructural será:

$$\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \equiv \begin{pmatrix} \hat{\varepsilon}_{1t} \\ \hat{\varepsilon}_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

con matriz de covarianzas:

$$\begin{aligned} Var(\varepsilon_t) &= \begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix} \cdot Var \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \cdot \begin{pmatrix} 1 & -c \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & -c\sigma_1^2 + \sigma_{12} \\ -c\sigma_1^2 + \sigma_{12} & c^2\sigma_1^2 + \sigma_2^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2} \end{pmatrix} \end{aligned}$$

Premultiplicando el modelo VAR₂(1):

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

por la matriz $\begin{pmatrix} 1 & 0 \\ -\hat{c} & 1 \end{pmatrix}$, tendríamos,

$$\begin{aligned} y_{1t} &= \beta_{10} + \beta_{11}y_{1t-1} + \beta_{12}y_{2t-1} + \hat{\varepsilon}_{1t} \\ y_{2t} &= (\beta_{20} - \hat{c}\beta_{10}) + \hat{c}y_{1t} + (\beta_{21} - \hat{c}\beta_{11})y_{1t-1} + (\beta_{22} - \hat{c}\beta_{12})y_{2t-1} + \hat{\varepsilon}_{2t} \end{aligned} \quad (19)$$

que es un modelo estructural en el que la variable y_1 tiene efectos contemporáneos sobre y_2 , pero y_2 no tiene efectos contemporáneos sobre y_1 . La estructura recursiva que apreciamos en este modelo, como en (17), también se obtendría en el caso de contar el modelo VAR con k ecuaciones. Es una estructura generada por la transformación del modelo VAR mediante el factor de la descomposición LDL.

Si transformamos el modelo VAR mediante el factor de la descomposición de Cholesky, tenemos:

$$PP' = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a^2 & ab \\ ab & b^2 + c^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

lo que conduce a:

$$a = \sqrt{\sigma_1^2}; \quad b = \frac{\sigma_{12}}{\sqrt{\sigma_1^2}}; \quad c = \sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}}$$

En este caso, premultiplicaríamos el modelo VAR₂(1) por la matriz P^{-1} que es, en este caso: $P^{-1} = \begin{pmatrix} 1/a & 0 \\ -b/ac & 1/c \end{pmatrix}$, obteniendo un modelo con estructura recursiva:

$$\begin{aligned} \frac{y_{1t}}{\sigma_1} &= \frac{1}{\sigma_1}\beta_{10} + \beta_{11}\frac{y_{1t-1}}{\sigma_1} + \beta_{12}\frac{y_{2t-1}}{\sigma_1} + w_{1t}, \quad w_{1t} = \frac{u_{1t}}{\sigma_1} \\ \frac{1}{c}y_{2t} &= \left(\frac{1}{c}\beta_{20} - \frac{b}{ac}\beta_{10}\right) - \frac{b}{ac}y_{1t} + \left(\frac{1}{c}\beta_{21} - \frac{b}{ac}\beta_{11}\right)y_{1t-1} + \left(\frac{1}{c}\beta_{22} - \frac{b}{ac}\beta_{12}\right)y_{2t-1} + w_{2t}, \quad w_{2t} = \frac{u_{2t}}{c} - \frac{b}{ac}w_{1t} \end{aligned}$$

en el que las innovaciones satisfacen:

$$\begin{aligned} Var(w_t) = P^{-1}Var(u_t) &= \begin{pmatrix} 1/\sqrt{\sigma_1^2} & 0 \\ -\sigma_{12}/\left(\sigma_1^2\sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}}\right) & 1/\sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}} \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \\ &= \begin{pmatrix} 1/\sqrt{\sigma_1^2} & -\sigma_{12}/\left(\sigma_1^2\sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}}\right) \\ 0 & 1/\sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Puede comprobarse que llegaríamos al mismo resultado si obtenemos el producto matricial $CD^{1/2}$ que aparece en la descomposición LDL y lo invertimos. En ambos modelos, tiene sentido preguntarse por las respuestas de ambas variables a un impulso en $\hat{\varepsilon}_{1t}$ o en $\hat{\varepsilon}_{2t}$, puesto que ambas están incorrelacionadas, por construcción, como también lo están w_{1t} y w_{2t} en el segundo modelo. La diferencia es que en el segundo modelo las innovaciones tienen varianza unitaria. En respuesta a un impulso en $\hat{\varepsilon}_{1t}$, ambas variables, y_{1t} y y_{2t} , reaccionarán en el mismo instante, y también en períodos siguientes, hasta que dichas respuestas decaigan a cero si las variables del VAR son estacionarias. En cambio, en respuesta a un impulso en $\hat{\varepsilon}_{2t}$, y_2 responderá en el mismo período y períodos siguientes, mientras que y_1 sólo responderá en períodos siguientes al del impulso. Si utilizamos esta transformación, los shocks observados en $\hat{\varepsilon}_1$ son shocks en u_{1t} , mientras que los shocks en $\hat{\varepsilon}_2$ son atribuidos al componente de u_{2t} que no es común con u_{1t} . Si el modelo VAR contase con 3 variables endógenas entonces, en respuesta a un impulso en $\hat{\varepsilon}_{3t}$, tanto y_1 como y_2 reaccionarán solo a partir del período siguiente.

Por tanto, la primera variable debe ser la que se considere más importante en el orden causal de relaciones entre todas ellas, la segunda variable es la segunda en importancia, y así sucesivamente. Como vemos, la estrategia de identificación LDL , al igual que la estrategia de identificación de Cholesky conducen a un modelo VAR estructural recursivo. Estos son modelos estructurales exactamente identificados, puesto que imponen el número exacto de restricciones precisas para la identificación. Estas dos estrategias de identificación apenas utilizan conceptos teóricos, por lo que se distingue entre estas estrategias de identificación de un modelo VAR, que permiten identificar un modelo estructural y que se basan en la descomposición de la matriz de covarianzas, pero

no utilizan conceptos de teoría económica, y los denominados modelos estructurales, que sí los utilizan.

3.6 Representación MA de un modelo VAR

Todo modelo VAR estacionario admite una representación de medias móviles (MA),

$$Y_t = \Phi(L)u_t = \sum_{s=0}^{\infty} \Phi_s u_{t-s}$$

a la que se llega tras sucesivas sustituciones de Y_{t-s} en (15). Esta representación permite resumir las propiedades de las relaciones cruzadas entre las variables que componen el vector Y_t , que queda representado como una combinación lineal de valores actuales y pasados del vector de innovaciones. La simultaneidad vuelve a quedar palpable en el sentido de que cualquier innovación u_{it} afecta a todas las variables $Y_{j,t+s}$.

Por ejemplo, si resolvemos recursivamente el modelo $\text{VAR}_k(1)$ tenemos,

$$\begin{aligned} Y_t &= A_0 + A_1 Y_{t-1} + u_t = A_0 + A_1(A_0 + A_1 Y_{t-2} + u_{t-1}) + u_t = \\ &= (I_k + A_1)A_0 + A_1^2 Y_{t-2} + (A_1 u_{t-1} + u_t) = \\ &= (I_k + A_1 + A_1^2 + \dots + A_1^{n-1})A_0 + \sum_{i=0}^{n-1} A_1^i u_{t-i} + A_1^n Y_{t-n} \end{aligned}$$

Como puede verse, para la estabilidad del sistema es preciso que las sucesivas potencias de la matriz A_1 decaigan hacia cero, pues de lo contrario, el futuro lejano tendría efectos sobre el presente, en contra de la rápida amortiguación temporal de efectos inherente a todo proceso estacionario. Esto requiere que las raíces del polinomio característico de dicha matriz $|I_k - A_1 \lambda| = 0$, caigan fuera del círculo unidad, condición análoga a la que se tiene para un proceso autoregresivo univariante.

Cuando se cumplen las condiciones de estabilidad, tomando límites en la expresión anterior, tenemos la *representación de medias móviles* del VAR,

$$Y_t = \mu + \sum_{i=0}^{\infty} A_1^i u_{t-i} \quad (21)$$

donde $\mu = E(Y)$ es el vector de esperanzas matemáticas, que viene dado por,⁸

$$\mu = (I_k - A_1)^{-1} A_0$$

Por tanto,

⁸Nótese la similitud con la expresión de la suma de una progresión geométrica de razón menor que la unidad, si bien en este caso se trata de una progresión geométrica de matrices.

$$Var(Y_t) = E[(Y_t - \mu)^2] = E\left[\sum_{i=0}^{\infty} A_1^i u_{t-i}\right]^2 = \sum_{i=0}^{\infty} A_1^i Var(u_{t-i}) (A_1^i)' = \sum_{i=0}^{\infty} A_1^i \Sigma_u (A_1^i)'$$

En el modelo $VAR_2(1)$:

$$\begin{aligned} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \\ &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \sum_{s=0}^{\infty} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}^s \begin{pmatrix} u_{1t-s} \\ u_{2t-s} \end{pmatrix} \end{aligned}$$

por lo que en este caso: $\Phi_s = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}^s$, y el término independiente es:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \left[I_2 - \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \right]^{-1} \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} \beta_{10}(1 - \beta_{22}) + \beta_{12}\beta_{20} \\ \beta_{20}(1 - \beta_{11}) + \beta_{21}\beta_{10} \end{pmatrix}$$

siendo $\Delta = (1 - \beta_{11})(1 - \beta_{22}) - \beta_{12}\beta_{21}$.

Además:

$$Var(Y_t) = \sum_{i=0}^{\infty} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}^i \begin{pmatrix} \sigma_{u_1}^2 & \sigma_{u_1 u_2} \\ \sigma_{u_1 u_2} & \sigma_{u_2}^2 \end{pmatrix} \left[\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}^i \right]'$$

En el modelo VAR general, podemos encontrar las matrices de coeficientes de la representación de medias móviles:

$$Y_t = (\phi_0 + \phi_1 L + \phi_2 L^2 + \dots) u_t$$

comparándola con la propia expresión analítica del modelo VAR,

$$Y_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + u_t \Rightarrow (I_k - A_1 L - A_2 L - \dots - A_p L^p) Y_t = u_t$$

de modo que,

$$Y_t = (\phi_0 + \phi_1 L + \phi_2 L^2 + \dots) u_t = (\phi_0 + \phi_1 L + \phi_2 L^2 + \dots) (I_k - A_1 L - A_2 L - \dots - A_p L^p) Y_t$$

por lo que tenemos:

$$\begin{aligned} I_k &= (\phi_0 + \phi_1 L + \phi_2 L^2 + \dots) (I_k - A_1 L - A_2 L - \dots - A_p L^p) = \\ &= \phi_0 + (\phi_1 - A_1 \phi_0) L + (\phi_2 - A_1 \phi_1 - A_2 \phi_0) L^2 + \dots \end{aligned}$$

que conduce a,

$$\begin{aligned}
\phi_0 &= I_k \\
\phi_1 &= A_1 \\
\phi_2 &= A_1\phi_1 + A_2 \\
\phi_3 &= A_1\phi_2 + A_2\phi_1 + A_3 \\
&\dots \\
\phi_s &= A_1\phi_{s-1} + A_2\phi_{s-2} + \dots + A_p\phi_{s-p}, \quad \forall s \geq p
\end{aligned}$$

con $\phi_s = 0_k$ para $s < 0$, de donde podemos obtener recursivamente las matrices ϕ_i de coeficientes de la representación de medias móviles:

$$Y_t = u_t + \phi_1 u_{t-1} + \phi_2 u_{t-2} + \phi_3 u_{t-3} + \dots$$

A efectos de la interpretación del modelo VAR estimado, es aún más importante considerar la representación MA *en términos de las innovaciones del modelo estructural*, ya que dichas innovaciones admiten generalmente una interpretación conceptual. En este sentido, si volvemos al modelo VAR₂(1), tenemos:

$$\begin{aligned}
\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \frac{1}{1 - \alpha_{12}\alpha_{21}} \sum_{s=0}^{\infty} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}^s \begin{pmatrix} 1 & \alpha_{11} \\ \alpha_{21} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-s} \\ \varepsilon_{2t-s} \end{pmatrix} \\
&= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \sum_{s=0}^{\infty} \begin{pmatrix} \tilde{\phi}_{11}(s) & \tilde{\phi}_{12}(s) \\ \tilde{\phi}_{21}(s) & \tilde{\phi}_{22}(s) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-s} \\ \varepsilon_{2t-s} \end{pmatrix} = \mu + \sum_{s=0}^{\infty} \tilde{\Phi}_s \varepsilon_{t-s}
\end{aligned}$$

donde,

$$\tilde{\Phi}(s) = \begin{pmatrix} \tilde{\phi}_{11}(s) & \tilde{\phi}_{12}(s) \\ \tilde{\phi}_{21}(s) & \tilde{\phi}_{22}(s) \end{pmatrix} = \frac{1}{1 - \alpha_{12}\alpha_{21}} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}^s \begin{pmatrix} 1 & \alpha_{11} \\ \alpha_{21} & 1 \end{pmatrix} \quad (23)$$

para lo cual necesitaremos conocer los parámetros α_{ij} , es decir, necesitaremos una determinada estrategia de identificación.

La representación de medias móviles de un modelo VAR tiene dos aplicaciones muy importantes: la estimación de las funciones de respuesta al impulso y la estimación de la descomposición de la varianza del error de predicción, que analizamos en las secciones siguientes.

3.7 Funciones de respuesta al impulso

Las funciones de respuesta al impulso miden la reacción de cada una de las variables a un shock en una de las innovaciones estructurales. En un sistema de interrelaciones, todas las variables reaccionarán a dicho shock; además, tratándose de un modelo dinámico, puede haber reacciones contemporáneas pero también

en todos los períodos siguientes. Por tanto, para cada innovación del modelo hay tantas funciones de respuesta al impulso como variables endógenas; cada una de dichas funciones dependen del tiempo transcurrido desde que se produce el shock. Tenemos por tanto una matriz $k \times k$ de tales funciones de respuesta al impulso.

Las funciones de respuesta al impulso generan una gran cantidad de números, pues se calcula el impacto que, en *cada instante futuro* tendría, *sobre cada variable* del modelo, un impulso en una determinada innovación, y ello puede repetirse para las *innovaciones en cada una de las ecuaciones*. Por eso, suelen representarse en varios gráficos, cada uno de los cuales incluye las respuestas a través del tiempo, de una determinada variable a un impulso en cada una de las innovaciones; de este modo se tiene tantos gráficos como variables en el modelo, cada uno de ellos conteniendo tantas curvas como variables. Alternativamente, pueden construirse gráficos, cada uno de los cuales representa la respuesta temporal de todas las variables del modelo a un impulso en una de las innovaciones. Nuevamente hay tantos gráficos como variables, cada uno de ellos conteniendo tantas curvas como variables.

La manera más natural de estimar las funciones de respuesta a un impulso es a partir de la representación MA del mismo. Consideremos la representación de medias móviles del modelo VAR(1) bivalente (22). Los coeficientes de la sucesión de matrices $\tilde{\Phi}_s = \begin{pmatrix} \tilde{\phi}_{11}(s) & \tilde{\phi}_{12}(s) \\ \tilde{\phi}_{21}(s) & \tilde{\phi}_{22}(s) \end{pmatrix}$ representan el impacto que, a lo largo del tiempo, tienen sobre las dos variables del modelo y_{1t} e y_{2t} una perturbación en las innovaciones estructurales $\varepsilon_{1t}, \varepsilon_{2t}$. Por ejemplo, los coeficientes $\tilde{\phi}_{12}(s)$ reflejan el impacto que en los distintos períodos $s, s \geq 1$, tiene sobre y_1 una perturbación de tipo impulso en ε_2 . Consideramos que ε_2 está en su valor de equilibrio, cero, excepto en un período determinado, en el que toma un valor numérico igual a 1; como consecuencia, tanto y_1 como y_2 reaccionan en dicho período, puesto que ε_{2t} aparece en ambas ecuaciones en (22) y dicha respuesta se extiende a varios períodos, hasta que la sucesión $\tilde{\phi}_{12}(s)$ se hace cero. La sucesión de valores numéricos $\{\tilde{\phi}_{12}(s)\}$ se conoce como la *respuesta de y_1 a un impulso en ε_2* . El *multiplicador o respuesta a largo plazo* es la suma $\sum_{s=0}^{\infty} \tilde{\phi}_{12}(s)$. Esta suma existe si las variables son estacionarias, pues en tal caso ha de cumplirse que $\sum_{s=0}^{\infty} |\tilde{\phi}_{12}(s)| < \infty$.

El problema al que nos enfrentamos al tratar de calcular las funciones de respuesta al impulso es que, si bien contamos con estimaciones numéricas de los parámetros $\beta_{ij}, i, j = 1, 2$, desconocemos los parámetros α_{11} y α_{21} que aparecen en (23). Por tanto, la estimación de las funciones de respuesta al impulso descansa en una determinada estrategia de identificación, es decir, en una determinada elección de matriz B , y dependerán de dichas condiciones de identificación.⁹

Para estimar las funciones de respuesta a un impulso sobre una determinada innovación del modelo estructural, fijamos el valor de dicha innovación

⁹Excepto si las innovaciones del modelo VAR, u_{1t} y u_{2t} , estuviesen incorrelacionadas.

estructural al tamaño deseado, y a cero en los instantes sucesivos, mientras que fijamos el valor numérico de todas las demás innovaciones igual a cero en todos los períodos, el corriente y los periodos futuros. Por ejemplo, en el modelo VAR₂(1) identificado mediante la descomposición LDL: (19),

$$\begin{aligned} y_{1t} &= \beta_{10} + \beta_{11}y_{1t-1} + \beta_{12}y_{2t-1} + \hat{\varepsilon}_{1t} \\ y_{2t} &= (\beta_{20} - \hat{c}\beta_{10}) + \hat{c}y_{1t} + (\beta_{21} - \hat{c}\beta_{11})y_{1t-1} + (\beta_{22} - \hat{c}\beta_{12})y_{2t-1} + \hat{\varepsilon}_{2t} \end{aligned} \quad (24)$$

un shock en ε_{1t} de tamaño σ_{ε_1} genera una respuesta contemporánea de igual tamaño, σ_{ε_1} , en y_1 y una respuesta igual a $c\sigma_{\varepsilon_1}$ en y_2 . En $t+1$, la respuesta de y_1 es igual a $(\beta_{11} + \beta_{12}c)\sigma_{\varepsilon_1}$, mientras que la respuesta de y_2 es $[\hat{c}(\beta_{11} + \beta_{12}c) + (\beta_{21} - \hat{c}\beta_{11}) + (\beta_{22} - \hat{c}\beta_{12})c]\sigma_{\varepsilon_1} = [\beta_{21} + \beta_{22}]\sigma_{\varepsilon_1}$, y así sucesivamente.

En respuesta a un shock en ε_{2t} de tamaño σ_{ε_2} , y_1 no respondería en dicho período, mientras que y_2 mostraría una respuesta de tamaño σ_{ε_2} . En $t+1$, la respuesta de y_1 sería igual a $\beta_{12}\sigma_{\varepsilon_2}$, mientras que la respuesta de y_2 sería igual a $\hat{c}\beta_{12}\sigma_{\varepsilon_2} + (\beta_{22} - \hat{c}\beta_{12})\sigma_{\varepsilon_2} = \beta_{22}\sigma_{\varepsilon_2}$. Las respuestas en periodos posteriores se obtendrían igual que en el caso anterior. La respuesta contemporánea de las variables que preceden en la ordenación de Cholesky a la variable que experimenta el shock son siempre nulas.

El lector puede comprobar que si hubiésemos situado a y_2 primera en la ordenación de Cholesky, hubiésemos recuperado un modelo estructural diferente, y las respuestas a los shocks en cada una de ellas habría sido diferente. De hecho, como puede fácilmente intuirse, en ese caso, la respuesta contemporánea de y_2 a un shock en ε_{1t} sería igual a cero.

En el caso particular de que utilicemos la transformación de Cholesky para identificar el modelo, tendremos perturbaciones estructurales:

$$\begin{pmatrix} \hat{\varepsilon}_{1t} \\ \hat{\varepsilon}_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \text{ con: } c = \sigma_{u_1u_2}/\sigma_{u_1}^2,$$

por lo que:

$$\begin{aligned} Y_t &= u_t + \tilde{\phi}_1 u_{t-1} + \tilde{\phi}_2 u_{t-2} + \tilde{\phi}_3 u_{t-3} + \dots \\ \tilde{\phi}_s &= \phi_s \cdot \begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix} \end{aligned}$$

Exercise 1 Consideremos un VAR(1) en forma reducida sin constante (es decir, las variables tiene esperanza igual a cero y las innovaciones u_{it} tienen covarianza cero),

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

y supongamos que antes del instante t_0 las innovaciones toman un valor igual a cero en todos los períodos, las variables están en sus niveles de equilibrio,

$y_i = y_i^* = 0, i = 1, 2, 3$. En dicho instante, la innovación u_{1t_0} toma un valor unitario, $u_{1t_0} = 1$, y vuelve a ser cero en los períodos siguientes. ¿Cuál es la respuesta del sistema?

En el instante t_0 ,

$$\begin{pmatrix} y_{1t_0} \\ y_{2t_0} \\ y_{3t_0} \end{pmatrix} = \begin{pmatrix} u_{1t_0} \\ u_{2t_0} \\ u_{3t_0} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

por lo que y_{2t_0} e y_{3t_0} estarán en sus niveles de equilibrio, $y_2 = y_2^* = 0$, $y_3 = y_3^* = 0$, mientras que $y_{1t_0} = y_1^* + 1 = 1$.

Posteriormente,

$$\begin{aligned} \begin{pmatrix} y_{1t_0+1} \\ y_{2t_0+1} \\ y_{3t_0+1} \end{pmatrix} &= \begin{pmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{pmatrix} \begin{pmatrix} y_{1t_0} \\ y_{2t_0} \\ y_{3t_0} \end{pmatrix} + \begin{pmatrix} u_{1t_0+1} \\ u_{2t_0+1} \\ u_{3t_0+1} \end{pmatrix} = \\ &= \begin{pmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{pmatrix} \begin{pmatrix} y_1^* + 1 \\ y_2^* \\ y_3^* \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,5 \\ 0,1 \\ 0 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \begin{pmatrix} y_{1t_0+2} \\ y_{2t_0+2} \\ y_{3t_0+2} \end{pmatrix} &= \begin{pmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{pmatrix} \begin{pmatrix} y_{1t_0+1} \\ y_{2t_0+1} \\ y_{3t_0+1} \end{pmatrix} + \begin{pmatrix} u_{1t_0+2} \\ u_{2t_0+2} \\ u_{3t_0+2} \end{pmatrix} = \\ &= \begin{pmatrix} 0,5 & 0 & 0 \\ 0,1 & 0,1 & 0,3 \\ 0 & 0,2 & 0,3 \end{pmatrix} \begin{pmatrix} 0,5 \\ 0,1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0,25 \\ 0,06 \\ 0,02 \end{pmatrix} \end{aligned}$$

que van proporcionando la primera columna de las matrices que obtenemos calculando las sucesivas potencias de la matriz de coeficientes A_1 .

De este modo, tendríamos las respuestas del sistema a sorpresas en las innovaciones del modelo VAR. Si queremos calcular las respuestas a innovaciones estructurales, debemos utilizar la representación,

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \frac{1}{1 - \alpha_{11}\alpha_{21}} \sum_{s=0}^{\infty} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}^s \begin{pmatrix} 1 & \alpha_{11} \\ \alpha_{21} & 1 \end{pmatrix} \begin{pmatrix} v_{1t-s} \\ v_{2t-s} \end{pmatrix}$$

y examinar la sucesión definida en (23).

3.8 Descomposición de la varianza

Consideremos la representación MA de un VAR en función de las innovaciones estructurales para un valor futuro del vector de variables:

$$y_{t+m} = \mu + \sum_{s=0}^{\infty} \tilde{\Phi}_s \varepsilon_{t+m-s}$$

A partir de esta expresión, las predicciones de las variables y_1, y_2 , a horizonte n serían:

$$E_t y_{t+m} = E_t \begin{pmatrix} y_{1t+m} \\ y_{2t+m} \end{pmatrix} = \mu + \sum_{s=m}^{\infty} \tilde{\Phi}_s \varepsilon_{t+m-s}$$

donde $\Phi(s) = (\phi_{ij}(s))$ es la sucesión de matrices que antes introdujimos. Nótese que hemos utilizado: $E_t \varepsilon_{t+1} = E_t \varepsilon_{t+2} = \dots = E_t \varepsilon_{t+m} = 0$

En consecuencia, el error de predicción a horizonte m es,

$$\begin{aligned} e_t(m) &= y_{t+m} - E_t y_{t+m} = \left(\mu + \sum_{s=0}^{\infty} \tilde{\Phi}_s \varepsilon_{t+m-s} \right) - E_t \left(\mu + \sum_{s=0}^{\infty} \tilde{\Phi}_s \varepsilon_{t+m-s} \right) = \sum_{s=0}^{m-1} \tilde{\Phi}_s \varepsilon_{t+m-s} = \\ &= \left(\begin{pmatrix} \tilde{\phi}_{11}(0)\varepsilon_{1t+m} + \dots + \tilde{\phi}_{11}(m-1)\varepsilon_{1t+1} \\ \tilde{\phi}_{21}(0)\varepsilon_{1t+m} + \dots + \tilde{\phi}_{21}(m-1)\varepsilon_{1t+1} \end{pmatrix} + \begin{pmatrix} \tilde{\phi}_{12}(0)\varepsilon_{2t+m} + \dots + \tilde{\phi}_{12}(m-1)\varepsilon_{2t+1} \\ \tilde{\phi}_{22}(0)\varepsilon_{2t+m} + \dots + \tilde{\phi}_{22}(m-1)\varepsilon_{2t+1} \end{pmatrix} \right) \end{aligned}$$

que, como vemos, depende de la realización de las perturbaciones que se produzcan en ambas innovaciones desde $t+1$ hasta $t+m$. El tamaño del error de predicción viene indicado por su varianza. Si hemos utilizado la estrategia de identificación basada en una descomposición ortogonal (ya sea mediante la descomposición LDL o mediante el factor Cholesky) mediante la descomposición LDL, las innovaciones $\varepsilon_{it}, i = 1, 2$ estarán incorrelacionadas, y tendremos :

$$Var \begin{bmatrix} e_{1t}(m) \\ e_{2t}(m) \end{bmatrix} = \begin{pmatrix} \sigma_{\varepsilon_1}^2 \sum_{s=0}^{m-1} \tilde{\phi}_{11}(s)^2 + \sigma_{\varepsilon_2}^2 \sum_{s=0}^{m-1} \tilde{\phi}_{12}(s)^2 \\ \sigma_{\varepsilon_1}^2 \sum_{s=0}^{m-1} \tilde{\phi}_{21}(s)^2 + \sigma_{\varepsilon_2}^2 \sum_{s=0}^{m-1} \tilde{\phi}_{22}(s)^2 \end{pmatrix}$$

que, inevitablemente, aumentan con el horizonte de predicción. La expresión anterior nos permite descomponer la varianza del error de predicción en dos fuentes, según tenga a ε_1 o a ε_2 como causa. Con ello, estamos examinando el inevitable error de predicción en cada variable a un determinado horizonte, y atribuyéndolo a la incertidumbre acerca de la evolución futura en cada una de las variables. Es, por tanto, una manera de hacer inferencia acerca de las relaciones intertemporales entre las variables que componen el vector y . Para ello, se expresan los componentes de cada varianza en términos porcentuales,

$$\left(\frac{\sigma_{\varepsilon_1}^2 \sum_{s=0}^{m-1} \tilde{\phi}_{11}(s)^2}{Var(e_{1t}(m))}; \frac{\sigma_{\varepsilon_2}^2 \sum_{s=0}^{m-1} \tilde{\phi}_{12}(s)^2}{Var(e_{1t}(m))} \right) y \left(\frac{\sigma_{\varepsilon_1}^2 \sum_{s=0}^{m-1} \tilde{\phi}_{21}(s)^2}{Var(e_{2t}(m))}; \frac{\sigma_{\varepsilon_2}^2 \sum_{s=0}^{m-1} \tilde{\phi}_{22}(s)^2}{Var(e_{2t}(m))} \right)$$

Si una variable es prácticamente exógena respecto a las demás, entonces explicará casi el 100% de la varianza de su error de predicción a todos los horizontes posibles. Esto es lo más habitual a horizontes cortos, mientras que a horizontes largos, otras variables pueden ir explicando un cierto porcentaje de la varianza del error de predicción.

Si hubiéramos utilizado la identificación basada en el factor Cholesky, las innovaciones estructurales estimadas tendrían además varianza unitaria, por lo que la descomposición de la varianza vendría dada por:

$$\left(\frac{\sum_{s=0}^{m-1} \tilde{\phi}_{11}(s)^2}{\text{Var}(e_{1t}(m))}, \frac{\sum_{s=0}^{m-1} \tilde{\phi}_{12}(s)^2}{\text{Var}(e_{1t}(m))} \right) y \left(\frac{\sum_{s=0}^{m-1} \tilde{\phi}_{21}(s)^2}{\text{Var}(e_{2t}(m))}, \frac{\sum_{s=0}^{m-1} \tilde{\phi}_{22}(s)^2}{\text{Var}(e_{2t}(m))} \right)$$

La descomposición de la varianza está sujeta al mismo problema de identificación que vimos antes para las funciones de respuesta al impulso, ya que depende de los parámetros $\tilde{\phi}_{ij}(s)$ siendo necesario introducir un conjunto de $k(k-1)/2$ restricciones como las consideradas en la sección anterior. Nuevamente, si la correlación entre las innovaciones del VAR es muy pequeña, la ordenación que se haga de las variables del vector y o, lo que es lo mismo, las restricciones de exclusión de valores contemporáneos que se introduzcan serán irrelevantes. En general, sin embargo, tales restricciones condicionan muy significativamente la descomposición de la varianza resultante. De hecho, con estas restricciones de identificación ε_1 explica el 100% de la varianza del error de predicción un período hacia adelante en la variable y_1 . Si, en vez de dicha restricción, excluyéramos y_{1t} de la segunda ecuación, entonces ε_2 explicaría el 100% de la varianza del error de predicción un período hacia adelante en la variable y_2 .

3.9 VAR Generalizado: un análisis sin condiciones de identificación

En esta sección hacemos un análisis de funciones de respuesta al impulso y de descomposición de la varianza del error de predicción que no descansa en la ordenación que se haga de las variables en el VAR

3.9.1 Funciones de respuesta al impulso

Comenzamos comparando la función de respuesta al impulso ortogonalizada, que antes ya vimos, y que utiliza innovaciones ortogonalizadas, con la función de respuesta al impulso generalizada, introducida por Diebold y Yilmaz (2007), que utiliza las innovaciones originales u_{it} del modelo VAR, no las innovaciones estructurales. Es un procedimiento que evita tener que imponer una determinada ortogonalización de la matriz de covarianzas como hemos hecho en el apartado anterior, en el que hemos utilizado la descomposición LDL o el factor Cholesky de la matriz de covarianzas para identificar el modelo VAR.

Partimos del modelo VAR:¹⁰

$$Y_t = \sum_{i=1}^p A_i Y_{t-i} + u_t, \quad t = 1, 2, \dots, T$$

¹⁰Nada varía sustancialmente si consideramos un modelo VAR aumentado para incluir posibles variables explicativas exógenas o deterministas Z_t :

$$Y_t = \sum_{i=1}^p A_i Y_{t-i} + \Psi Z_t + u_t, \quad t = 1, 2, \dots, T$$

pero el álgebra se hace más complicado.

Bajo las condiciones de regularidad habituales, la representación de medias móviles es:

$$Y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \quad t = 1, 2, \dots, T \quad (25)$$

en la que las matrices Φ_i , de orden $k \times k$, pueden obtenerse recursivamente del modo que antes vimos. Por definición, la función de respuesta al impulso describe la reacción del sistema Y_t a un hipotético vector de shocks: $\delta = (\delta_1, \delta_2, \dots, \delta_k)$ en las innovaciones $(u_{1t}, u_{2t}, \dots, u_{kt})$ del modelo VAR, en relación con la evolución que habría seguido el sistema en ausencia de tales shocks.

Función de respuesta al impulso ortogonalizada La función de respuesta al impulso ortogonalizada utiliza las condiciones de identificación que surgen de aplicar la descomposición de Cholesky de la matriz de covarianzas: $PP' = \Sigma_u$, con P triangular inferior, y $\varepsilon_t = P^{-1}u_t$. Entonces:

$$Y_t = \sum_{i=0}^{\infty} (\Phi_i P)(P^{-1}u_{t-i}) = \sum_{i=0}^{\infty} (\Phi_i P)\varepsilon_{t-i} = \sum_{i=0}^{\infty} \tilde{\Phi}_i \varepsilon_{t-i}, \quad t = 0, 1, 2, \dots, T$$

con $E(\varepsilon_t \varepsilon_t') = I_k$.

La función de respuesta ortogonalizada $\psi_j^o(h)$ del vector de variables Y_{t+h} a un impulso unitario en la innovación estructural ε_j de la ecuación j (que es equivalente a un impulso de una desviación típica en ε_j al haber utilizado el factor Cholesky) es la diferencia entre la trayectoria seguida por el sistema después de un haber experimentado un shock en alguna de sus variables, y la trayectoria que habría seguido en ausencia de dicho shock.

En este modelo, es claro que:

$$\psi_j^o(h) = \tilde{\Phi}_h 1_j = \Phi_h P 1_j, \quad h = 0, 1, 2, \dots$$

donde 1_j es un k -vector columna con un 1 en su coordenada j y ceros en el resto, que selecciona por tanto la columna j de las matrices $\tilde{\Phi}_h$ o P . La matriz $\Phi_n P$ contiene todas las respuestas de las variables del sistema a todos los shocks. El producto $\Phi_n P 1_j$, al ser la columna j de $\Phi_n P$, contiene las respuestas de todas las variables a shocks en ε_j .

Función de respuesta al impulso generalizada (no ortogonalizada) El enfoque sin condiciones de identificación sigue otro razonamiento: si denotamos la historia de la economía hasta el instante $t-1$ por la sucesión no-decreciente de conjuntos de información Ω_{t-1} , tenemos la función generalizada de respuesta al impulso:

$$GI_Y(h, \delta, \Omega_{t-1}) = E(Y_{t+h} | u_t = \delta, \Omega_{t-1}) - E(Y_{t+h} | \Omega_{t-1}) \quad (26)$$

Utilizando (25) tenemos:

$$\begin{aligned} E(Y_{t+h} \mid u_t = \delta, \Omega_{t-1}) &= \Phi_h \delta + \Phi_{h-1} u_{t-1} + \Phi_{h-2} u_{t-2} + \dots, \\ E(Y_{t+h} \mid \Omega_{t-1}) &= \Phi_{h-1} u_{t-1} + \Phi_{h-2} u_{t-2} + \dots \end{aligned}$$

por lo que:

$$GI_Y(h, \delta, \Omega_{t-1}) = \Phi_h \delta$$

que es independiente de Ω_{t-1} pero depende de la composición de los shocks recogidos en δ . La elección de este vector δ condiciona las propiedades de las funciones de respuesta al impulso.

Generalmente, en vez de perturbar todos los elementos de u_t , consideramos un shock en su j -ésimo elemento:

$$GI_Y(h, \delta_j, \Omega_{t-1}) = E(Y_{t+h} \mid u_{jt} = \delta_j, \Omega_{t-1}) - E(Y_{t+h} \mid \Omega_{t-1})$$

para lo que tendremos que integrar los efectos de los restantes shocks utilizando la distribución histórica de los errores. Es decir, tenemos que inferir los valores de los shocks en las demás variables, teniendo en cuenta las correlaciones históricas entre las innovaciones del modelo VAR.¹¹

Suponiendo que u_t sigue una distribución Normal multivariante, se tiene:¹²

$$E(u_t \mid u_{jt} = \delta_j) = (\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{kj})' (\sigma_j^2)^{-1} \delta_j = \Sigma_u 1_j (\sigma_j^2)^{-1} \delta_j$$

por lo que el k -vector de respuestas generalizadas en el instante $t+n$ a un shock de tipo impulso de tamaño δ_j en la ecuación j en el instante t es:

$$\psi_j^g(h) = \Phi_h E(u_t \mid u_{jt} = \delta_j) = \left(\frac{\Phi_h \Sigma_u 1_j}{\sqrt{\sigma_j^2}} \right) \left(\frac{\delta_j}{\sqrt{\sigma_j^2}} \right), \quad h = 0, 1, 2, \dots$$

Si tomamos: $\delta_j = \sigma_j$ obtenemos la función de respuesta generalizada (no ortogonalizada) de Y_{t+h} a un impulso "normalizado" en la ecuación j (de tamaño igual a una desviación típica de la innovación original):

$$\psi_j^g(h) = \sqrt{\sigma_j^2} \Phi_h \Sigma_u 1_j, \quad h = 0, 1, 2, \dots$$

que es un vector de dimensión k .

¹¹De modo similar a como se relacionan los Efectos Parciales y los Efectos Totales de una variable explicativa sobre la variable endógena en una regresión lineal.

¹²Recordemos que si el vector $x = (x_1, x_2)$ sigue una distribución Normal multivariante $N \left(\begin{matrix} \mu_1 \\ \mu_2 \end{matrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$, la distribución del subvector x_1 condicional en $x_2 = a$ es: $(x_1 \mid x_2 = a) \sim N(\mu, \Sigma)$, con: $\mu = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (a - \mu_2)$, $\Sigma = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$. En nuestro caso, x_2 es u_j , x_1 son todos los demás shocks, y tenemos: $\mu_1 = 0, \mu_2 = 0, a = \delta_j, \Sigma_{22} = \sigma_j^2, \Sigma_{12} = (\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{kj})$.

3.9.2 Descomposición de la varianza del error de predicción a h periodos

Como ya vimos, la descomposición de la varianza nos permite dividir la varianza del error de predicción de cada variable en los componentes que son atribuibles a los distintos shocks que puede experimentar el sistema. Así podemos responder a preguntas del tipo: ¿qué porcentaje de la varianza del error de predicción en $y_{1,t}$ es debido a shocks en $y_{2,t}$? La descomposición de la varianza se obtiene a partir de la función de respuesta al impulso, y ambas se obtienen a partir de la representación de medias móviles del proceso.

La función generalizada de respuesta al impulso también puede utilizarse para deducir la descomposición de la varianza del error de predicción, definido como el porcentaje de la varianza del error de predicción h periodos hacia adelante en la variable i que es atribuible a shocks en la variable j .

Comencemos recordando las expresiones del error de predicción:

Descomposición de la varianza bajo condiciones de identificación (ortogonalizada)

Puesto que:

$$Y_t = \sum_{i=0}^{\infty} \tilde{\Phi}_i \varepsilon_{t-i}, \text{ con } E(\varepsilon_t \varepsilon_t') = I_k,$$

el error de predicción de Y_t^i dentro de h periodos es:

$$\begin{aligned} Y_{t+h}^i - E_t Y_{t+h}^i &= 1_i' \tilde{\Phi}_0 \varepsilon_{t+h} + 1_i' \tilde{\Phi}_1 P \varepsilon_{t+h-1} + \dots + 1_i' \tilde{\Phi}_h P \varepsilon_t = \\ &= 1_i' \Phi_0 P \varepsilon_{t+h} + 1_i' \Phi_1 P \varepsilon_{t+h-1} + \dots + 1_i' \Phi_h P \varepsilon_t \end{aligned}$$

Antes vimos que las respuestas de la variable Y_t^i a un impulso unitario (es decir, un impulso de una desviación típica, ya que esta es igual a 1) en la variable j son:

$$1_i' \psi_j^o(h) = 1_i' \Phi_h P 1_j$$

que es un escalar.

Al igual que en modelos univariantes, los errores de predicción, así como las funciones de respuesta al impulso, se obtienen directamente de la representación de medias móviles del proceso. Es decir, el modo en que una sorpresa de una desviación típica en la variable ε_t^j contribuye a un error de predicción en Y_{t+h}^i es: $1_i' \Phi_h P \varepsilon_t^j$, con varianza: $(1_i' \Phi_h P 1_j)^2$.

Volviendo a la expresión del error de predicción, $Y_{t+h}^i - E_t Y_{t+h}^i$, tenemos que el tamaño de cada sumando, medido por su varianza es:

$$Var(1_i' \Phi_0 P \varepsilon_{t+h}) = 1_i' \Phi_0 P Var(\varepsilon_{t+h}) P' \Phi_0 1_i = 1_i' \Phi_0 P P' \Phi_0 1_i = 1_i' \Phi_0 \Sigma_u \Phi_0 1_i$$

y una expresión similar se obtiene para los demás sumandos. Por tanto, la varianza del error de predicción es:

$$Var(Y_{t+h}^i - E_t Y_{t+h}^i) = \sum_{n=0}^h 1_i' \Phi_n \Sigma_u \Phi_n 1_i$$

Por tanto, la descomposición ortogonalizada de la varianza del error de predicción en Y_{t+h}^i es:

$$\left(\frac{\sum_{n=0}^h (1_i' \Phi_n P 1_j)^2}{\sum_{n=0}^h 1_i' \Phi_n \Sigma_u \Phi_n 1_i}, j = 1, 2, \dots, k \right)$$

cuyos componentes suman 1.

El producto $1_i \Phi_n P$ es la fila i de $\Phi_n P$ y contiene las respuestas de la variable Y_i a todos los shocks. El producto $1_i \Phi_n P 1_j$ es el elemento (i, j) de la matriz $\Phi_n P$. El producto $1_i' \Phi_n \Sigma_u \Phi_n 1_i = 1_i' \Phi_n P P' \Phi_n 1_i$ es la varianza del error de predicción de la variable i . Es también la suma de cuadrados de la fila i de $\Phi_n P$ y, por tanto, la suma de los elementos del numerador.

Descomposición de la varianza generalizada

Veamos ahora la descomposición generalizada de la varianza del error de predicción.

Partiendo de:

$$Y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \text{ con } E(u_t u_t') = \Sigma_u,$$

el error de predicción de la variable Y_t^i dentro de h periodos es:

$$Y_{t+h}^i - E_t Y_{t+h}^i = 1_i' \Phi_0 u_{t+h} + 1_i' \Phi_1 u_{t+h-1} + \dots + 1_i' \Phi_h u_t$$

teniendo que:

$$Var(1_i' \Phi_0 u_{t+h}) = 1_i' \Phi_0 Var(u_{t+h}) \Phi_0 1_i = 1_i' \Phi_0 \Sigma_u \Phi_0 1_i$$

y una expresión similar se obtiene para los demás sumandos. Este es el mismo resultado que obtuvimos con las innovaciones ortogonalizadas, por lo que obtenemos la misma expresión para la varianza del error de predicción:

$$Var(Y_{t+h}^i - E_t Y_{t+h}^i) = \sum_{n=0}^h 1_i' \Phi_n \Sigma_u \Phi_n 1_i$$

De hecho, no podría ser diferente, pues incluso el propio error de predicción debe ser el mismo en ambas representaciones. Otra cuestión diferente es el modo en que dicha varianza se distribuye entre las innovaciones de las variables del modelo VAR.

Puesto que la función de respuesta generalizada y normalizada (es decir, con: $\delta_j = \sigma_j$) de Y_{t+h} a un impulso en la ecuación j es:

$$\psi_j^g(h) = \sqrt{\sigma_j^{-2}} \Phi_h \Sigma_u 1_j, h = 0, 1, 2, \dots$$

la respuesta de la variable Y_t^i es: $\sigma_j^{-1} 1'_i \Phi_h \Sigma_u 1_j$, y su varianza es: $(1'_i \Phi_h \Sigma_u 1_j)^2 / \sigma_j^2$, por lo que el efecto relativo de un shock en ε_j sobre Y_{t+n}^i será;

$$\frac{\sum_{n=0}^h \text{Var} \left[(1'_i \Phi_n \Sigma_u 1_j) / \sqrt{\sigma_j^2} \right]}{\sum_{n=0}^h \text{Var}(e_{t+n}^i)} = \frac{\sum_{n=0}^h (1'_i \Phi_n \Sigma_u 1_j)^2}{\sigma_j^2 \sum_{n=0}^h 1'_i \Phi_n \Sigma_u \Phi'_n 1_i}$$

y la descomposición de la varianza del error de predicción h períodos hacia adelante en la variable Y_{t+h}^i se obtiene:

$$\left(\frac{\sum_{n=0}^h (1'_i \Phi_n \Sigma_u 1_j)^2}{\sigma_j^2 \sum_{n=0}^h 1'_i \Phi_n \Sigma_u \Phi'_n 1_i}, j = 1, 2, \dots, k \right)$$

La suma de esta descomposición sólo será igual a 1 si la matriz de covarianzas Σ_u fuese diagonal. En efecto, si Σ_u fuese diagonal, tendríamos que:

$$\begin{aligned} 1'_i \Phi_n &= (\phi_{i1}(n), \phi_{i2}(n), \dots, \phi_{ik}(n)) \\ \Sigma_u 1_j &= \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2) 1_j = (0, 0, \dots, \sigma_j^2, 0, \dots, 0) \Rightarrow \\ \Rightarrow 1'_i \Phi_n \Sigma_u 1_j &= \phi_{ij}(n) \sigma_j^2 \Rightarrow \sum_{j=1}^k (1'_i \Phi_n \Sigma_u 1_j)^2 = \sum_{j=1}^k (\phi_{ij}(n))^2 (\sigma_j^2)^2 \end{aligned}$$

mientras que en el denominador:

$$\begin{aligned} 1'_i \Phi_n \Sigma_u &= (\phi_{i1}(n), \phi_{i2}(n), \dots, \phi_{ik}(n)) \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2) = \\ &= (\phi_{i1}(n) \sigma_1^2, \phi_{i2}(n) \sigma_2^2, \dots, \phi_{ik}(n) \sigma_k^2) \Rightarrow \\ 1'_i \Phi_n \Sigma_u \Phi'_n 1_i &= \phi_{i1}(n)^2 \sigma_1^2 + \phi_{i2}(n)^2 \sigma_2^2 + \dots + \phi_{ik}(n)^2 \sigma_k^2 \end{aligned}$$

por lo que:

$$\sum_{j=1}^k \frac{\sum_{n=0}^h (1'_i \Phi_n \Sigma_u 1_j)^2}{\sigma_j^2 \sum_{n=0}^h 1'_i \Phi_n \Sigma_u \Phi'_n 1_i} = \frac{\sum_{j=1}^k \sigma_j^2 (\phi_{ij}(n))^2}{\sum_{n=0}^h 1'_i \Phi_n \Sigma_u \Phi'_n 1_i} = 1$$

Solo en este caso serán ambas descomposiciones de varianza iguales entre sí.

3.9.3 Spillovers entre variables: medidas de contagio

Por simplicidad, consideremos ahora un modelo VAR₂(1), con vector de variables: $Y_t = (y_{1t}, y_{2t})$. La predicción óptima 1 periodo hacia adelante en este modelo VAR₂(1) es:

$$Y_{t+1,t} = A_1 Y_t,$$

con vector de errores de predicción:

$$e_{t+1,t} = Y_{t+1} - Y_{t+1,t} = \tilde{\Phi}_0 \varepsilon_{t+1} = \begin{pmatrix} \tilde{\phi}_{11}(0) & \tilde{\phi}_{12}(0) \\ \tilde{\phi}_{21}(0) & \tilde{\phi}_{22}(0) \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t+1} \\ \varepsilon_{2,t+1} \end{pmatrix}$$

que tiene matriz de covarianzas:

$$Var(e_{t+1,t}) = E(e_{t+1,t}e'_{t+1,t}) = \tilde{\Phi}_0\tilde{\Phi}'_0$$

ya que bajo la identificación de Cholesky, $Var(\varepsilon_t) = I_2$. Por tanto, en particular, la varianza del error de la predicción 1 periodo hacia adelante de la variable $y_{1,t}$ es $\tilde{\phi}_{11}(0)^2 + \tilde{\phi}_{12}(0)^2$, mientras que la varianza del error de la predicción 1 periodo hacia adelante de la variable $y_{2,t}$ es $\tilde{\phi}_{21}(0)^2 + \tilde{\phi}_{22}(0)^2$.

Definamos el *componente de varianza propia* como la fracción de la varianza del error de predicción 1 periodo hacia adelante que es debida a shocks en la propia variable, y el *componente cruzado de la varianza o spillover* como el porcentaje de dicha varianza que es debida a shocks en la otra variable.

En un VAR con dos variables, hay dos posibles spillovers: shocks en $y_{1,t}$ que afectan a la varianza del error de predicción en $y_{2,t}$, con una contribución de $\tilde{\phi}_{21}(0)^2$, y shocks en $y_{2,t}$ que afectan a la varianza del error de predicción de $y_{1,t}$, con una contribución de $\tilde{\phi}_{12}(0)^2$. Por tanto, el spillover total es $\tilde{\phi}_{12}(0)^2 + \tilde{\phi}_{21}(0)^2$. Podemos convertir el spillover total en un índice fácilmente interpretable expresandolo en términos relativos a la variación total del error de predicción: $\tilde{\phi}_{11}(0)^2 + \tilde{\phi}_{12}(0)^2 + \tilde{\phi}_{21}(0)^2 + \tilde{\phi}_{22}(0)^2 = traza(\tilde{\Phi}_0\tilde{\Phi}'_0)$. Si expresamos este ratio como un porcentaje, tenemos el Índice de Spillover:

$$\text{Índice de Spillover: } S = \frac{\tilde{\phi}_{12}(0)^2 + \tilde{\phi}_{21}(0)^2}{traza(\tilde{\Phi}_0\tilde{\Phi}'_0)}.100$$

que como puede verse, mide la intensidad de la interacción entre ambas variables, en ambas direcciones.

Consideremos ahora un $VAR_k(p)$, con k variables. Para las predicciones un periodo hacia adelante tenemos:

$$S = \frac{\sum_{i,j=1,i \neq j}^k \tilde{\phi}_{ij}(0)^2}{traza(\tilde{\Phi}_0\tilde{\Phi}'_0)}.100$$

y para predicciones h periodos hacia adelante:

$$S = \frac{\sum_{n=0}^h \sum_{i,j=1,i \neq j}^k \tilde{\phi}_{ij}(n)^2}{\sum_{n=0}^h traza(\tilde{\Phi}_n\tilde{\Phi}'_n)}.100$$

que son de nuevo medidas que toman en cuenta las relaciones existentes entre las variables, en todas las direcciones.

Expresiones de este tipo pueden utilizarse para definir efectos contagio o spillover entre variables a partir de las expresiones para la descomposición de la varianza que antes vimos.¹³ En este caso, estaremos tomando en cuenta las relaciones que se producen en una determinada dirección:

¹³Lutkepohl (1991)

$$\theta_{j \rightarrow i}^o(h) = \frac{\sum_{n=0}^h (1'_i \Phi_n P 1_j)^2}{\sum_{n=0}^h 1'_i \Phi_n \Sigma_u \Phi'_n 1_i},$$

$$\theta_{j \rightarrow i}^g(h) = \frac{\sum_{n=0}^h \text{Var} \left[(1'_i \Phi_n \Sigma_u 1_j) / \sqrt{\sigma_j^2} \right]}{\sum_{n=0}^h \text{Var}(e_{t+n}^i)} = \frac{\sum_{n=0}^h (1'_i \Phi_n \Sigma_u 1_j)^2}{\sigma_j^2 \sum_{n=0}^h 1'_i \Phi_n \Sigma_u \Phi'_n 1_i}$$

Para entender la presencia del factor σ_j^2 en la función generalizada de respuesta al impulso, recordemos que un shock de tamaño σ_j en ε_j equivale a un shock unitario en w_j .

Por construcción, $\sum_{j=1}^k \theta_{j \rightarrow i}^o(h) = 1$. Sin embargo, debido a las covarianzas no nulas entre los shocks originales, en general tendremos: $\sum_{j=1}^k \theta_{j \rightarrow i}^g(h) \neq 1$. En efecto, el producto $1'_i \Phi_n \Sigma_u 1_j$ que aparece en el numerador de $\theta_{j \rightarrow i}^g(h)$ es el elemento (i, j) de $\Phi_n \Sigma_u$ y sumando sus cuadrados ya no tendremos el denominador. Por esta razón, conviene normalizar cada componente de la matriz de descomposición de varianza utilizando la suma de una fila:

$$\tilde{\theta}_{j \rightarrow i}^g(h) = \frac{\theta_{j \rightarrow i}^g(h)}{\sum_{j=1}^k \theta_{j \rightarrow i}^g(h)} \cdot 100, \text{ para } i, j = 1, 2, \dots, k$$

Nótese que, por construcción: $\sum_{j=1}^k \tilde{\theta}_{j \rightarrow i}^g(h) = 100$, y $\sum_{i,j=1}^k \tilde{\theta}_{j \rightarrow i}^g(h) = k \times 100$, lo que incluye los efectos a shocks propios. Estos efectos propios son precisamente lo que excluimos en el cálculo del índice de spillover que introducimos a continuación.

También podemos construir un *Total Spillover index*, que mide la contribución de los spillovers entre variables a la varianza del error de predicción de todo el sistema de k variables:

$$TSG(h) = \frac{\sum_{i,j=1, i \neq j}^k \tilde{\theta}_{j \rightarrow i}^g(h)}{k}, \text{ para } i, j = 1, 2, \dots, k$$

El Total Spillover indica, en promedio, el porcentaje de la varianza del error de predicción en todas las variables que es debido a spillovers, es decir, al contagio de los shocks on otras variables. Es de nuevo, una medida que analiza todas las direcciones de las relaciones entre las variables. Antes vimos un caso particular, para dos variables, de esta medida.

Los índices *net directional spillover* mide el spillover neto transmitido desde la variable i a las demás:

$$NDS_{i \rightarrow all}^g(h) = \sum_{j=1, i \neq j}^k \tilde{\theta}_{i \rightarrow j}^g(h) - \sum_{j=1, i \neq j}^k \tilde{\theta}_{j \rightarrow i}^g(h), \text{ para } i, j = 1, 2, \dots, k$$

Este índice mide la diferencia entre la transmisión del shock en i hacia las demás variables, y los efectos sobre la variable i de shocks en las demás variables.

Valores positivos del índice NDS indican que la variable i es un transmisor de efectos spillover, mientras que un valor numérico negativo de NDS indica que es un receptor de efectos spillover.

Finalmente, los índices *net pairwise spillover* entre las variables i y j :

$$NPS_{i \rightarrow j}^g(n) = \tilde{\theta}_{i \rightarrow j}^g(n) - \tilde{\theta}_{j \rightarrow i}^g(n),$$

mide la diferencia entre la transmisión de shocks de i hacia j , menos la transmisión de efectos spillover de j a i . Nuevamente, valores positivos o negativos indicarán que la variable i es, respecto de la variable j , un emisor o un receptor de efectos spillover.

3.10 Apéndice: Suma de respuestas a un impulso

4 Cointegración

Una variable y_t se dice que es integrada de orden d , lo cual representamos como $y_t \sim I(d)$ si es preciso diferenciar serie temporal d veces para conseguir una transformación que sea estacionaria. Una variable estacionaria es integrada de orden 0. Dicho de otro modo una variable es integrada de orden d si el polinomio característico asociado a su representación autoregresiva tiene exactamente d raíces unitarias. Si una serie temporal es $I(d)$, entonces puede escribirse:

$$y_t = (1 - L)^d z_t$$

con $z_t \sim I(0)$, estacionario.

Consideremos dos variables integradas: $x_t \sim I(d)$, $z_t \sim I(e)$ con $e > d$. Toda variable que sea combinación lineal de ambas variables es, en general, $I(e)$, puesto que habría que diferenciar dicha combinación e veces para conseguir una variable estacionaria. Lo sorprendente es que si las dos variables fuesen integradas del mismo orden, d , podría haber situaciones en que una combinación lineal fuese integrada de orden inferior a d . Se dice entonces que ambas variables están cointegradas de orden $d - b$.

El caso quizá más interesante es: $x_t \sim I(1)$, $z_t \sim I(1)$. en general, si $y_t = ax_t + bz_t$, entonces $y_t \sim I(1)$. Pero es posible en algunos casos que $y_t \sim I(0)$ sea estacionaria. En tal caso, se dice que x_t y z_t están cointegradas. Por ejemplo, supongamos que el IBEX35 y el Dow Jones son variables $I(1)$ ¿Existirá alguna combinación lineal de ambos índices que sea estacionario? Si esto sucediera, interpretaríamos que la raíz unitaria que contienen ambas variables es la misma y, por tanto, una cartera que tomase posiciones en ambos índices de acuerdo con los coeficientes de dicha combinación lineal tendría un precio de mercado estacionario, mostrando reversion a la media (o a su tendencia, si tiene comportamiento alcista). Se diría entonces que ambos mercados mantienen una relación de equilibrio a largo plazo. Esta denominación pretende recoger el hecho de que cualquier distanciamiento *permanente* respecto de su nivel de equilibrio de uno de dichos activos x_t , debe ir acompañado de otro distanciamiento, asimismo *permanente*, en el otro activo z_t , de tal manera que su combinación

lineal $y_t = ax_t + bz_t$ permanezca proxima a su nivel de equilibrio. Pero no impide en absoluto que se produzcan desviaciones *transitorias* en ambos activos que separen a y_t *transitoriamente* de su nivel de equilibrio. Entre otras cosas, esto sugeriria que podriamos tomar posiciones en ambos mercados (muy posiblemente de signo opuesto) para aprovechar desviaciones *transitorias* en ambos mercados respecto de la relacion de equilibrio que mantendrian a largo plazo. Este es el fundamento estadístico de la negociacion por pares, que ha suscitado mucho interés en los últimos años.

Se puede tener cointegración entre varias variables, incluso con diferente orden de integración. Entonces, están cointegradas si existe una combinación lineal de las variables de mayor orden de integración que están cointegradas con las variables con menor orden de integración del grupo de variables. Por ejemplo, si $x_t \sim I(2)$, $z_t \sim I(2)$, $w_t \sim I(1)$, las tres variables están cointegradas si existen coeficientes a, b tales que $v_t = ax_t + bz_t \sim I(1)$ y $cv_t + w_t \sim I(0)$.

Si x_t y z_t están cointegradas, entonces x_t y z_{t-i} también estan cointegradas.

Si $x_t \sim I(1)$ y $z_t \sim I(1)$ están cointegradas, entonces también están cointegradas con cualquier variable estacionaria.

Si x_t y z_t están cointegradas, entonces o bien x_t causa en el sentido de Granger a z_t , o z_t causa en el sentido de Granger a x_t , o suceden ambas cosas a la vez. Esto se debe a la existencia del *modelo de corrección del error*.

El estimador MCO de una relación de cointegración es superconsistente, es decir, converge al verdadero valor de la parámetro al aumentar el tamaño de la muestra mas rapidamente que lo hace el estimador MCO de un modelo de regresión entre variables estacionarias. Además, su consistencia no requiere la ausencia de correlación entre variables explicativas de la regresión de cointegración y el término de error. De hecho, en la regresión de cointegracion puede tomarse como variable dependiente cualquiera de las variables del conjunto cuya cointegración se esta contrastando. Sin embargo, dicho estimador es sesgado en muestras finitas, y su distribución limite no es estándar. El tamaño del sesgo está relacionado con 1-R2 de la regresión de cointegración, de modo que si el R2 es alto, dicho sesgo será pequeño.

Engle y Granger (1987) demostraron que el estimador en dos etapas del modelo de corrección del error es asimismo consistente.

Consideremos un vector z de variables de naturaleza $I(1)$. Se dice que dichas variables están *cointegradas* si existe una combinación lineal de las mismas, definida por un vector α , tal que $\alpha'z$ es una variable aleatoria $I(0)$, es decir, *estacionaria*. Más generalmente, se dice que un vector z de variables cuyo máximo orden de integración es q están *cointegradas* si existe una combinación lineal de las mismas, definida por un vector α tal que $\alpha'z$ es una variable aleatoria $I(p)$, con $p < q$. El vector α se denomina *vector de cointegración*. Una relación de cointegración entre variables $I(1)$ puede involucrar asimismo alguna variable $I(0)$, estacionaria, pero no variables $I(2)$.

4.1 Contraste de cointegración

Si partimos de variables y_t, x_t de naturaleza $I(1)$, sus primeras diferencias, $\Delta y_t, \Delta x_t$ son estacionarias. Contrastaríamos entonces la cointegración de y_t, x_t estimando una regresión,

$$y_t = \beta_0 + \beta_1 x_t + v_t, \quad t = 1, 2, \dots, T \quad (27)$$

y contrastando la estacionariedad de los residuos, como propusieron Engle y Granger (1987). Sin embargo, los niveles críticos para el contraste de esta hipótesis no son los mismos que para el contraste de raíces unitarias en una variable, pues ahora, el contraste se lleva a cabo después de haber estimado el modelo de regresión (27).

Esto no es irrelevante: el procedimiento de mínimos cuadrados busca los valores del espacio paramétrico (β_0 y β_1 en la regresión anterior) que minimizan la varianza del residuo resultante, y éste tiene una varianza infinita para los valores de β_1 que no hacen que las variables estén cointegradas. Por tanto, si y_t, x_t están cointegradas, el método de MCO tenderá a seleccionar el valor de β_1 que genera residuos estacionarios, es decir, la constante de cointegración. Aunque esto es lo que pretendemos, ello significa que hay una cierta tendencia a concluir con más frecuencia de la que debiéramos que las variables están cointegradas. En consecuencia, los valores críticos para el contraste de raíz unitaria de los residuos de (27) deben ser más elevados en valor absoluto que los utilizados para el contraste de raíz unitaria habitual.

La contraste de cointegración β_1 , si existe, es única. En efecto supongamos que existe otra constante γ , con $\gamma \neq \beta$, tal que $y_t - \gamma x_t$ es estacionario. Pero $y_t - \gamma x_t = y_t - \beta x_t + (\beta - \gamma)x_t$, siendo $y_t - \beta x_t$ y $y_t - \gamma x_t$ estacionarios por hipótesis. Sin embargo, el último sumando no es estacionario, por lo que el supuesto de partida no es válido, y γ no puede ser una constante de cointegración.

Si los residuos de esta regresión resultan ser estacionarios, decimos que las variables y_t, x_t están cointegradas, siendo (27) o $\alpha'z = y_t - \beta_0 - \beta_1 x_t$ la *relación de cointegración* entre ambas, en la que ya habríamos normalizado una de las coordenadas del vector α para que sea igual a 1. Tal normalización es siempre posible, ya que si una variable es estacionaria, dividirla o multiplicarla por un número no va a cambiar dicha condición. La relación de cointegración se interpreta como la relación de largo plazo entre ellas, alrededor de la cual experimentan ambas variables desviaciones a corto plazo que *revierten* posteriormente. Es decir, si en un determinado período, y_t está por encima del valor numérico de $\beta_0 + \beta_1 x_t$ para ese mismo período, generalmente y_t crecerá por encima de $\beta_0 + \beta_1 \Delta x_t$, de manera que y_{t+1} tenderá a acercarse a $\beta_0 + \beta_1 x_{t+1}$. En el caso de dos variables y_t, x_t , decimos que β_1 es la constante de cointegración entre ambas.

El contraste de cointegración descrito, mediante la contrastación de la existencia de una raíz unitaria en los residuos de una regresión, se conoce como contraste de Engle y Granger. Cuando el vector z_t consta de más de 2 variables pueden existir varias relaciones de cointegración. Esto es lo que sucede, por

ejemplo, al considerar un vector de tipos de interés a distinto vencimiento, dentro de un mismo mercado, ya sea el mercado secundario de deuda pública, un mercado de swap en una determinada divisa, etc.. En este caso, el procedimiento de Engle-Granger para estimar vectores de cointegración es problemático, pues estimaremos solamente una combinación lineal de las posibles relaciones de cointegración existentes entre las variables que componen el vector. De hecho, la estimación resultante dependerá de la normalización de coeficientes utilizada en (27), a diferencia de lo que ocurre en el caso de dos variables. En tal caso hay que utilizar el procedimiento de Johansen, que explicamos en la sección siguiente.

Si generamos ambas series temporales a partir de procesos independientes, este contraste nos sugerirá en una mayoría de simulaciones que y_t, x_t no están cointegradas, lo que aparecerá en la forma de residuos de naturaleza $I(1)$ en (27). En tal caso, habríamos de estimar un modelo en diferencias de ambas variables,

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + v_t, \quad t = 1, 2, \dots, T$$

que arrojará un coeficiente β_1 no significativo y un coeficiente de determinación muy reducido.

Al contrastar cointegración, estamos tratando de detectar la posible existencia de relaciones de largo plazo entre las variables del modelo. En ese sentido, la naturaleza del contraste sugiere el uso de una datos no necesariamente frecuentes, y una muestra temporal suficientemente amplia. De lo contrario, predominará en la muestra, en términos relativos, la información acerca de las fluctuaciones de corto plazo en las variables, frente a la de su evolución tendencial, que es lo que tratamos de detectar. Por tanto, una elección inapropiada de la muestra, ya sea por una frecuencia alta de observación de los datos, o por el uso de un período muestral no muy amplio, sesgará el resultado del contraste hacia la no detección de relaciones de cointegración.

Efectuar un análisis de cointegración significa relacionar los niveles de variables como oferta monetaria y precios, o los precios de dos activos financieros, es decir, variables no estacionarias, y no sus tasas de variación. Por el contrario, basar la caracterización de la relación entre variables como las citadas utilizando coeficientes de correlación estándar es delicado, pues puede conducir a la detección de regresiones espúreas. El concepto de cointegración generaliza el concepto de correlación en la dirección adecuada. La existencia de una tendencia estocástica común generaría una relación sostenible a largo plazo entre ambas variables, lo que hará que sus diferenciales reviertan a través del tiempo, es decir, que sean *mean-reverting*. No tiene sentido analizar las relaciones entre los niveles de variables $I(1)$ si no están cointegradas.¹⁴

Cointegración en un modelo estimado por mínimos cuadrados

¹⁴El concepto de cointegración, como existencia de una tendencia o de una raíz unitaria común a varias variables, se extiende a otros conceptos. Así, se puede hablar de una estructura ARCH común a varias variables, si todas ellas tienen estructura ARCH, pero existe una combinación lineal de todas ellas que no tiene tal estructura. *Common Features*.

Al estimar la relación anterior por mínimos cuadrados hay que tener en cuenta que las propiedades de dicho estimador son válidas únicamente en el caso de variables estacionarias. Cuando las variables están cointegradas, el uso de mínimos cuadrados en la estimación de la regresión está justificado estadísticamente, pero la distribución de probabilidad del estimador MCO no es la habitual. Por tanto, aunque el programa de estimación que utilizemos nos proporcionará las desviaciones típicas estimadas y los ratios tipo- t de cada coeficiente, estos no son válidos en este caso, y no deben utilizarse, por ejemplo, para contrastar hipótesis sobre los coeficientes de la relación de cointegración.

Hay muchos casos en que el modelo teórico sugiere que las variables y_t, x_t deben estar relacionadas con un determinado valor numérico del coeficiente, por ejemplo, $\beta^0 = 1$, por lo que el investigador estará interesado en contrastar dicha hipótesis. Esto puede hacerse por un procedimiento indirecto, sustituyendo el valor teórico de β , $\beta = \beta^0$ en la relación entre ambas variables. Ello significa que construimos la variable auxiliar $w_t = y_t - \beta^0 x_t$, y contrastamos la estacionariedad de esta variable. Por ejemplo, la Hipótesis de las Expectativas en la formación de la Estructura Temporal de los Tipos de Interés implica que el diferencial de tipos a largo, R_t , y corto plazo, r_t , debe ser estacionario, es decir, en la relación: $R_t = \alpha + \beta r_t + u_t$, los residuos \hat{u}_t son estacionarios, y no se rechaza la hipótesis: $H_0 : \beta = 1$. Esto equivale a contrastar la estacionariedad del spread $R_t - r_t$.

Cuando se procede de este modo, es importante repetir el contraste para valores de β^0 en un entorno de β^0 , con el objeto de analizar la precisión con que hemos identificado la constante de cointegración.

Correlación y cointegración no son sinónimos

El problema de correlación espúrea surge entre variables no estacionarias, con independencia de que estén o no cointegradas, luego puede haber alta correlación (de hecho, muy elevada) sin cointegración. Alternativamente, el hecho de que exista una relación de largo plazo entre variables no estacionarias no impide que éstas experimenten desviaciones respecto de la misma que, si son de apreciable magnitud, reducirán la correlación existente entre dichas variables. Un ejemplo sería la evolución temporal de la cotización de un valor en Bolsa, analizada conjuntamente con un índice que lo incluya, ya sea el índice de mercado, un índice de los valores más capitalizados, o un índice sectorial; dado que todo índice es un promedio ponderado de las cotizaciones de los valores en él incluidos, cabría esperar que ambas series temporales estuvieran relacionadas. Sin embargo, las fluctuaciones que ambos experimentan a corto plazo pueden ser suficientes para que su coeficiente de correlación sea reducido. Un ejemplo típico de variables posiblemente relacionadas pero habitualmente no cointegradas lo constituye algunos tipos de cambio.

Por esta misma razón, cointegración tampoco es sinónimo de alta capacidad explicativa. Podemos encontrar, por ejemplo, un conjunto reducido de acciones que están cointegradas con el Ibex35, pero al constituir la cartera formada por dichas acciones utilizando los coeficientes de la relación de cointegración podríamos descubrir que la capacidad explicativa de dicha cartera sobre el Ibex35 fuese muy reducida. La varianza de los residuos de dicha regresión, conocida

como *tracking error* de la cartera que intenta replicar el Ibex35, sería muy alta.

A partir de dos variables no estacionarias, pero cointegradas, es sencillo construir dos variables no cointegradas, sin más que añadir en cada período a una de ellas, la suma hasta dicho periodo de un ruido blanco. Si la varianza de este componente no es muy grande, mantendremos una correlación análoga a la inicial, que podía ser elevada. Sin embargo, por construcción, las dos variables no están cointegradas. Por tanto, hay que tener sumo cuidado en interpretar los resultados obtenidos en contrastes de cointegración.

Generalized Purchasing Power Parity

La PPP afirma que el tipo de cambio nominal debe estar determinado por el diferencial de tipos de interés de los dos países (*Paridad cubierta* de tipos de cambio). Dado que este supuesto es claramente refutado por los datos, una versión más flexible del modelo permite que los *tipos de cambio reales* estuviesen determinados no solo por el diferencial de tipos de interés, sino también por el diferencial de crecimiento de la renta, el diferencial de crecimiento de la oferta monetaria, etc.. Tampoco esta versión del modelo parece dar buenos resultados. Una interpretación es que los determinantes de los tipos de cambio, no siendo estacionarios, inducen la no estacionariedad de los propios tipos de cambio. Sin embargo, si los determinantes de distintos países, no siendo estacionarios entre sí, mantienen relaciones de largo plazo entre ellos, entonces los tipos de cambio deberían estar cointegrados. [Ejercicio]

Variables cointegradas: ejemplos

Ejemplo 1: Un ejemplo de variables cointegradas es,

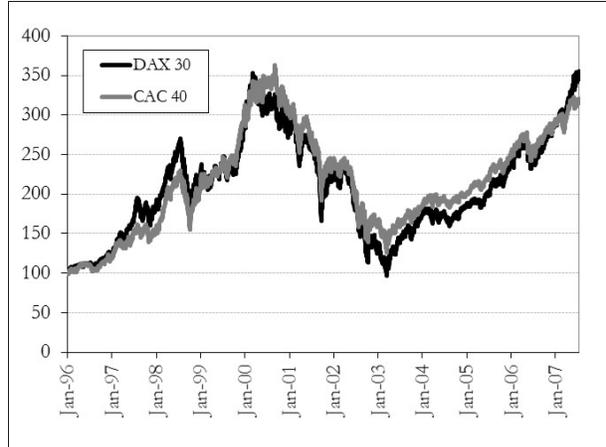
$$\begin{aligned}x_t &= \alpha_x + \beta_x w_t + \varepsilon_{x_t} \\y_t &= \alpha_y + \beta_y w_t + \varepsilon_{y_t} \\w_t &= w_{t-1} + \varepsilon_t\end{aligned}$$

donde w_t es la tendencia (estocástica) común a x_t e y_t , siendo ε_{x_t} , ε_{y_t} variables aleatorias $N(0, \sigma_x^2)$, $N(0, \sigma_y^2)$, sin autocorrelación. Las variables x_t e y_t están cointegradas, puesto que

$$y_t - \frac{\beta_y}{\beta_x} x_t = \left(\alpha_y - \frac{\beta_y}{\beta_x} \alpha_x \right) + \left(\varepsilon_{y_t} - \frac{\beta_y}{\beta_x} \varepsilon_{x_t} \right)$$

que es una variable estacionaria. El vector $(1, -\beta_y/\beta_x)$ se denomina vector de cointegración, mientras que se dice que $-\beta_y/\beta_x$ es la constante de cointegración, y la combinación lineal $\xi_t = y_t - \frac{\beta_y}{\beta_x} x_t$, que es estacionaria, es la cuantía en la que se incumple la relación de equilibrio a largo plazo en cada período t .

Ejemplo 2: En Example II.5.11 (C. Alexander) se proporcionan datos diarios de los índices CAC y DAX, desde el inicio de 1996 a mitad de julio de 2007. El gráfico presenta dichos datos, cuando estos se representan como índices, con base 100 en el primer dato.



El contraste de existencia de raíz unitaria, en GRETL, incluyendo constante y tendencia, arroja para el logaritmo del índice CAC un estadístico para el contraste de Dickey-Fuller ampliado de -1,695, con p-value de 0,753, incluyendo 7 retardos en la regresión de Dickey-Fuller ampliada. Para el DAX tenemos un estadístico de contraste de -1,489, con p-value 0,833, incluyendo asimismo 7 retardos en la regresión de Dickey-Fuller. Ambos índices tienen una raíz unitaria.

En primeras diferencias, es decir, trabajando con rentabilidades diarias, para el CAC tenemos un estadístico de -38,612, con p-value 0,0, incluyendo un retardo en la regresión de Dickey-Fuller. Para el DAX tenemos un estadístico de -26,78, con p-value de 0,0 incluyendo 3 retardos en la regresión. Podemos concluir que las rentabilidades son estacionarias.

¿Están cointegrados ambos índices? Continuamos trabajando con sus logaritmos. Estimamos la regresión:

$$\ln CAC_t = 0,195 + 0,959 \cdot \ln DAX_t + u_t, R^2 = 0,912$$

(0,005)

Para los residuos de esta regresión, incluyendo 5 retardos en la regresión de *DF*, obtenemos un estadístico de contraste de -1,713, con p-value de 0,672, lo que nos lleva a no rechazar la hipótesis nula de existencia de raíz unitaria. En consecuencia, los residuos no son estacionarios y los índices no están cointegrados.

4.2 El contraste de cointegración de Johansen

Si consideramos un vector autoregresivo $VAR_k(p)$,

$$y_t = \mu_t + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + Bx_t + \epsilon_t$$

donde y_t es un vector de k variables no estacionarias, $I(1)$, x_t es un vector de variables deterministas, y ϵ_t es un vector de innovaciones. Denotemos $\Phi(B) =$

$I_k - A_1B - A_2B^2 - \dots - A_pB^p$. Si todas las raíces de la ecuación característica: $|\Phi(B)| = 0$ están fuera del círculo unidad, entonces y_t es estacionario, en el sentido de no contener raíces unitarias. Si la matriz de multiplicadores a largo plazo satisface: $|\Phi(1)| = |I_k - A_1 - A_2 - \dots - A_p| = 0$, entonces y_t es no estacionario, conteniendo al menos una raíz unitaria. Una raíz unitaria es un factor del tipo $(1-B)$ en la descomposición factorial del polinomio autoregresivo $\Phi(B)$. Suponemos por el momento que y_t tiene como mucho una raíz unitaria. En ese caso, $(1-B)y_{it}$ sería estacionario.

El proceso $VAR(p)$ puede escribirse en la forma de *Modelo Vectorial de Corrección del Error Vectorial (VECM)* que justificaremos más adelante:

$$\Delta y_t = \mu_t + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + Bx_t + \epsilon_t$$

con

$$\Pi = \sum_{i=1}^p A_i - I_k = -\Phi(1), \quad \Gamma_i = - \sum_{j=i+1}^p A_j, \quad i = 1, 2, \dots, p-1$$

Por simplicidad, ignoramos la posible presencia de variables exógenas en el modelo. El término Πy_{t-1} se conoce como término de corrección del error, y juega un papel central en el análisis de cointegración.

Yendo en dirección contraria, la representación $VAR(p)$ puede recuperarse a partir de la estimación del MCE mediante:

$$\begin{aligned} A_1 &= I_k + \Pi - \Gamma_1 \\ A_i &= \Gamma_i - \Gamma_{i-1}, \quad i = 2, \dots, p, \quad \text{con } \Gamma_p = 0_{k \times k} \end{aligned}$$

Si y_t contiene raíces unitarias, entonces $|\Phi(1)| = 0$ y la matriz $\Pi = -\Phi(1)$ será singular. Cabe considerar 3 casos:

1. $Rango(\Pi) = 0$. Esto implica $\Pi = 0_{k \times k}$ y las variables que componen el vector y_t no están cointegradas. El modelo ECM se reduce a un modelo VAR en primeras diferencias:

$$\Delta y_t = \mu_t + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + \epsilon_t$$

1. $Rango(\Pi) = k$. Esto implica que $|\Phi(1)| \neq 0$ e y_t no contiene raíces unitarias, es decir, y_t es $I(0)$. El modelo ECM no es informativo, y analizamos directamente un modelo VAR en y_t :

$$y_t = \mu_t + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \dots + \Gamma_p y_{t-p} + \epsilon_t$$

1. $0 < Rango(\Pi) = m < k$. Entonces, podemos escribir Π como: $\Pi = \alpha\beta'$, donde α y β son matrices $k \times m$ con $Rango(\alpha) = Rango(\beta) = m$. El modelo ECM se convierte en:

$$\Delta y_t = \mu_t + \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + \epsilon_t$$

En este caso, las variables que componen el vector y_t están cointegradas, existiendo m vectores de cointegración linealmente independientes, que vienen dados por las combinaciones lineales: $w = \beta' y_t$. La ecuación característica $|\Phi(1)| = 0$ tiene $k-m$ raíces unitarias, que generan $k-m$ tendencias estocásticas comunes a las variables que componen y_t .

Por tanto, un enfoque para contrastar cointegración es analizar el rango de la matriz Π de multiplicadores a largo plazo. Este es el enfoque de Johansen. De hecho, el enfoque de Johansen, que consiste en estimar el modelo VAR por máxima verosimilitud y analizar el rango de la estimación de la matriz Π , sugiere dos contrastes: el contraste de la traza y el contraste del máximo autovalor. Los contrastes se ejecutan secuencialmente.

En el caso del contraste de la traza, la hipótesis nula es del tipo:

$$H_0 : \text{Rango}(\Pi) \leq m \text{ versus } H_1 : \text{Rango}(\Pi) > m$$

o, lo que es lo mismo: H_0 : Las variables que componen el vector y_t tienen un máximo de m relaciones de cointegración, frente a la hipótesis alternativa: H_1 : Las variables que componen el vector y_t tienen más de m relaciones de cointegración.

El estadístico de contraste es:

$$LK_{tr}(m) = -(T-p) \sum_{i=m+1}^k \ln(1 - \hat{\lambda}_i)$$

siendo $\hat{\lambda}_i$ los autovalores generalizados estimados para una determinada matriz que surge en el proceso de estimación por Máxima Verosimilitud [el lector interesado puede consultar detalles en Hamilton (1994) *Time Series Analysis*, capítulo 19]. Si $\text{Rango}(\Pi) = m$, entonces los $\hat{\lambda}_i$ serán pequeños para $i > m$ y por tanto, el estadístico de contraste será pequeño. Este es el contraste de la traza, aunque en realidad, es una secuencia de contrastes. La distribución de probabilidad del estadístico es función de movimientos Brownianos, y depende del número de variables en el modelo y del componente determinista que se incluya.

El primer contraste se efectúa para $m = 0$: H_0 : El vector y_t no tiene ninguna relación de cointegración, frente a la hipótesis alternativa: H_1 : El vector y_t tiene al menos 1 relación de cointegración. Si rechazamos H_0 , contrastaríamos con $m = 1$: H_0 : El vector y_t tiene como máximo una relación de cointegración, frente a la hipótesis alternativa: H_1 : El vector y_t tiene al menos 2 relaciones de cointegración, y así sucesivamente.

El contraste del máximo autovalor especifica:

$$H_0 : \text{Rango}(\Pi) = m \text{ versus } H_1 : \text{Rango}(\Pi) = m + 1$$

y el estadístico de contraste es:

$$LK_{\max}(m) = -(T - p) \ln(1 - \hat{\lambda}_{m+1})$$

Lamentablemente, la distribución de probabilidad del estadístico de contraste de cointegración depende del componente determinista del vector y_t , al igual que sucedía con el contraste de raíz unitaria en el caso univariante. Las distintas especificaciones pueden considerar una constante y/o una tendencia lineal en la relación de cointegración, como cuando decimos que $y_t - \beta x_t - \alpha$ es estacionario, siendo y_t y x_t variables $I(1)$, o que $y_t - \beta x_t - \gamma t - \alpha$ es estacionario. Los distintos paquetes estadísticos solicitan del usuario que especifique los componentes deterministas, para posteriormente proporcionarle los valores críticos correspondientes a la especificación que se haya propuesto para el componente determinista.

En algunas ocasiones se leen trabajos en los que el investigador analiza la posible cointegración bajo todas las especificaciones posibles del componente determinista, para luego tomar una decisión en función de cuántos de dichos contrastes hayan dado una respuesta u otra. Incluso algunos programas estadísticos permiten al usuario dar esa opción. Sin embargo, tal enfoque debe no es una buena práctica estadística. El usuario debe analizar previamente sus series para tener una cierta idea acerca de la estructura del componente determinista. Si las series tienen tendencia, entonces deberá incluir una tendencia en la relación de cointegración y una constante en el modelo VECM. Si las series tienen constante pero no tendencia, entonces deberá incluir una constante en la relación de cointegración, y nada en el modelo VECM. Si las variables no tienen constante ni tendencia, entonces no deberá incluir componente determinista ni en la relación de cointegración ni en el modelo VECM.

Las matrices α, β no son únicas. Si tomamos una matriz no singular $k \times k, C$, y formamos: $\alpha^* = \alpha C, \beta^* = \beta C^{-1}$, tenemos:

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

por lo que la identificación de α y β requiere imponer algunas restricciones.
Aplicación práctica del contraste de Johansen

1. Buscar, posiblemente mediante contraste, la especificación apropiada para el orden del VAR. El número de retardos puede condicionar los resultados. Hay que asegurarse de que los residuos están libres de autocorrelación.
2. Estimar el modelo y determinar el rango de Π ¹⁵
3. Analizar los vectores de cointegración normalizados β y los coeficientes α de velocidad de ajuste al desequilibrio.
4. Contrastar restricciones, si se desea.

¹⁵ Si bien, como hemos dicho, el análisis de autovalores se realiza para otra matriz diferente de Π

Contraste de restricciones

Estimar los modelos restringido y sin restringir, y utilizar el estadístico:

$$-T \sum_{i=m+1}^k \left[\ln(1 - \hat{\lambda}_i^*) - \ln(1 - \hat{\lambda}_i) \right]$$

donde $\hat{\lambda}_i^*$ denotan las raíces características de la matriz Π sin restricciones, y $\hat{\lambda}_i$ las raíces características obtenidas en el modelo estimado después de imponer las restricciones.

4.3 El modelo de corrección del error

Teorema de representación de Engle y Granger: el caso de 2 variables

Este teorema afirma que si dos variables y_t, x_t de naturaleza $I(1)$, están cointegradas, sus relaciones dinámicas están caracterizadas por el *modelo vectorial de corrección de error (VECM)*:

$$\begin{aligned} \Delta y_t &= \alpha_y + \sum_{i=1}^m \delta_{1i}^y \Delta x_{t-i} + \sum_{i=1}^n \delta_{2i}^y \Delta y_{t-i} + \gamma_y z_{t-1} + \varepsilon_{y_t} \\ \Delta x_t &= \alpha_x + \sum_{i=1}^p \delta_{1i}^x \Delta x_{t-i} + \sum_{i=1}^q \delta_{2i}^x \Delta y_{t-i} + \gamma_x z_{t-1} + \varepsilon_{x_t} \end{aligned} \quad (28)$$

donde z_{t-1} denota la desviación del período anterior respecto de la relación de equilibrio a largo plazo $z_{t-1} = y_{t-1} - \beta x_{t-1}$, siendo β el coeficiente de cointegración entre y_t y x_t , y Δ es el operador de primeras diferencias. En el modelo de corrección de error todas las variables son estacionarias, $I(0)$, por lo que las propiedades habituales del estimador MCO en dicho contexto, son válidas. Los términos $\gamma_y z_{t-1}$ y $\gamma_x z_{t-1}$ se denominan *términos de corrección de error*, y han de aparecer en las ecuaciones anteriores con un determinado signo, que depende del modo en que se haya definido el desequilibrio z_{t-1} . Con nuestra definición, cabría esperar que $\gamma_y < 0, \gamma_x > 0$; un valor negativo de γ_y indicará que períodos en que y_t es alto, es decir, superior a βx_t , tenderán a venir seguidos de crecimientos relativamente reducidos de dicha variable. Un valor positivo de γ_x indica que siguiendo a períodos en que y_t es alto, x_t tenderá a experimentar un crecimiento mayor; la conjunción de ambos efectos hace que y_{t+1} tienda a aproximarse a βx_{t+1} . Lo dual ocurrirá tras períodos en que y_t haya sido bajo, es decir, inferior a βx_t . Si hubiéramos normalizado la relación de cointegración de otro modo, habríamos definido el término de desequilibrio como $z_{t-1} = \beta y_{t-1} - x_{t-1}$, y los signos de los coeficientes γ_y, γ_x en (28) deberían ser entonces los contrarios a los antes descritos.

No es preciso, sin embargo, que se produzcan estos signos: la aproximación entre ambas variables puede conseguirse asimismo si ambas aumentan o disminuyen simultáneamente, pero x_t experimenta la mayor variación. Por tanto, si ambos coeficientes tienen igual signo, γ_x debe ser significativamente mayor que

γ_y en valor absoluto. De hecho, podría ocurrir también que sólo unos de los dos coeficientes resulte estadísticamente significativo, lo que podría interpretarse en el sentido de que la variable asociada soporta todo el peso del ajuste hacia la relación de equilibrio a largo plazo.

La cointegración entre variables no lleva añadida ninguna interpretación concreta en términos de causalidad entre dichas variables. De hecho, como la relación de cointegración puede normalizarse de distintas maneras, puede presentarse una apariencia de causalidad en cualquiera de las dos direcciones. El modelo de corrección de error muestra que, en presencia de cointegración, existe importante causalidad entre ambas variables, en principio, con carácter bidireccional. Sólo si algunos de los coeficientes del modelo VECM resultan ser estadísticamente no significativos, podría hablarse de causalidad unidireccional. Si dos variables están cointegradas, al menos una de ellas causa a la otra; sin embargo, ello podría también reflejar el efecto común de una tercera variable, no considerada en el modelo. La ausencia de causalidad en un sistema cointegrado implica que una de las variables no reacciona a variaciones en la otra. Esto significa que los retardos de la segunda no aparecen en la ecuación de la primera ni en la forma de diferencias, ni tampoco a través del término de corrección del error.

Por ejemplo, al trabajar con datos de precios de contado y del futuro sobre un determinado activo financiero, es habitual hallar un mayor número de retardos del precio del futuro en la ecuación del contado, que viceversa, lo que sugiere que los mercados de derivados (en este caso, de futuros), incorporan la nueva información más rápidamente que los mercados de contado, por lo que los últimos parecen responder a fluctuaciones en los primeros. En este tipo de ejemplos, en ocasiones el término de corrección de error resulta no significativo en la ecuación de precios del mercado de contado.

Cuando el vector z_t incorpora más de dos variables, y existe más de una relación de cointegración entre ellas, el modelo de corrección de error adopta una expresión similar a la antes propuesta. La diferencia estriba en que aparecen retardos de todas las variables, en diferencias, en todas las ecuaciones, y aparecen tantos términos de corrección de error como relaciones de cointegración en cada una de las ecuaciones. Dichos términos serán los valores retardados de dichas relaciones de cointegración; la normalización escogida afecta únicamente a la interpretación de los valores numéricos estimados.

La búsqueda de variables cointegradas abundan en la literatura financiera, donde trata de caracterizarse las posibles relaciones de equilibrio a largo plazo entre precios de activos. Así, se han analizado las posibles relaciones de cointegración entre tipos de cambio, entre tipos de interés dentro de una misma estructura temporal, entre mercados de contado y futuro, entre commodities, valoración de divisas. También se ha utilizado este tipo de análisis para discutir el grado de integración entre mercados de valores o de deuda, si bien parece existir más evidencia favorable en el primer tipo de mercados. Este análisis tiene asimismo implicaciones para la gestión financiera: en principio, debería ser posible encontrar una cesta reducida de valores cointegrada con el índice, lo que podría utilizarse en la gestión pasiva de carteras. Lo mismo debería ocurrir

con un pequeño conjunto de índices sectoriales, etc..

Ejemplo 1: Antes hicimos un análisis de cointegración de logaritmos de los datos diarios de los índices CAC y DAX, desde el inicio de 1996 a mitad de julio de 2007 siguiendo el enfoque de Engle y Granger, concluyendo que ambas variables no están cointegradas en dicha muestra temporal. La aplicación del contraste de Johansen sugiere nuevamente que no existe ninguna relación de cointegración entre ambos índices.

Cuando las variables no están cointegradas, como en este caso, el modelo a estimar es un VAR en primeras diferencias. Para estimar un VAR en primeras diferencias usando GRETL, el criterio de Akaike sugiere un orden 8, el criterio bayesiano BIC sugiere orden 1 y el criterio Hannan-Quinn sugiere orden 4. Un orden 8 parece excesivo con datos diarios, por lo que tomamos el orden 4 como compromiso. Aunque hay coeficientes individualmente significativos en ambas ecuaciones, los R2 son muy reducidos y las desviaciones típicas residuales prácticamente coinciden con las desviaciones típicas de la variable dependiente, como cabe esperar en series temporales largas, de datos frecuentes volátiles, como son las rentabilidades. No hay evidencia de autocorrelación residual en ninguna de las dos ecuaciones.

Los contrastes de causalidad de Granger arrojan un valor numérico del estadístico de 16,421 cuando contrastamos la presencia de la rentabilidad del DAX en la ecuación del CAC, y de 33,95 cuando contrastamos la presencia de la rentabilidad CAC en la ecuación del DAX. Ambos tienen p-valores reducidos, por lo que concluimos acerca de la influencia mutua entre ambos mercados. Cabe observar que si hubiésemos optado por un orden inferior, como el sugerido por el criterio de Akaike, $n = 1$, hubiésemos concluido que el bloque de retardos del DAX no añade información a los propios retardos del CAC para explicar el comportamiento en rentabilidad de este último índice. Sin embargo, los contrastes de autocorrelación muestran claros indicios de presencia de la misma en los residuos de ambas ecuaciones, al contrario de lo que sucede con $n = 4$, por lo que desecharmos la evidencia empírica arrojada por el modelo más sencillo, con $n = 1$.

Ejemplo 2: Tomamos datos del Ibex contado y del Futuro sobre Ibex, desde el inicio de 1993 a 9/9/2014.

Los criterios de Akaike, BIC y Hannan-Quinn coinciden en sugerir un orden del VAR en niveles de $n = 6$. El estadístico de la traza de Johansen para el contraste de no existencia de relaciones de cointegración es 93,878, con p-value 0,0, y el estadístico del máximo autovalor es de 87,658, asimismo con p-value 0,0. Por tanto, se rechaza dicha hipótesis y pasamos a pensar que existe al menos una relación de cointegración. Para el contraste de H_0 : existe un máximo de una relación de cointegración, ambos estadísticos toman el valor 6,2206, con p-value de 0,0126. Dado que trabajamos con una muestra larga, tiene sentido ser estrictos en el contraste y tomar como nivel de significación 99%, con lo cual no rechazamos dicha hipótesis nula. Otra razón para no hacerlo es que de hacerlo, concluiríamos que existen dos relaciones de cointegración, es decir, que ambas variables: Ibex y Futuro sobre Ibex, son estacionarias, lo cual evidentemente no es cierto.

La estimación de máxima verosimilitud por el método de Johansen conduce a a una relación de cointegración:

$$\log(Ibex_t) = 0,052 + 1,0059 \cdot \ln(FIbex_t) + u_t \quad (29)$$

(0,0013)

El término de corrección del error (el residuo en la relación anterior) entra en la ecuación de $\Delta \log(Ibex_t)$ con coeficiente 0,0014(0,0034) y en la ecuación de $\Delta \log(FIbex_t)$ con coeficiente 0,0689(0,0359). En dichas ecuaciones aparecen como variables explicativas, además de una constante y del término de corrección de error, 4 retardos de la primera diferencia de cada una de las dos variables. Por tanto, podemos interpretar que es el Futuro quien reacciona ante desviaciones de los precios de futuro y contado con respecto de su relación de equilibrio a largo plazo, definida por (29).

El signo positivo es el esperado: valores positivos de u_t indican precios de contado "relativamente altos" dado el precio del futuro y la relación (29), que se ajustan en periodos sucesivos mediante elevaciones del precio del futuro, dado el efecto positivo que u_t tiene en su ecuación. Valores negativos de u_t indicarían precios del futuro relativamente altos, que se corregirían con descensos en el precio del futuro, ya que u_t entra con signo positivo en su ecuación. Lo relevante es que es el futuro quien ajusta estas desviaciones, lo cual es un resultado habitual al analizar muchos mercados de futuros. Es un efecto que se conoce como "*price discovery*", que se refiere al hecho de que el mercado de futuros reacciona antes que el mercado de contado ante cualquier noticia, debido a que mientras que es sencillo negociar el futuro sobre Ibex, es complicado negociar una cartera de contado de Ibex.

En consonancia con este resultado, el estadístico F para el contraste de causalidad de Granger desde el futuro hacia el contado es de 37,75, con p-value 0,0, siendo de 6,61 con p-value de 0,16 cuando se contrasta la causalidad del contado al futuro.

Un ejemplo similar puede verse en el ExII.5.11 del libro de Alexander considera la cointegración entre el índice Hang Seng y el contrato de futuros sobre dicho índice.

Ejemplo 3: Tomamos datos diarios de tipos de interés cupón cero de UK, con vencimientos 1 mes, 3 meses, 12 meses y 60 meses, extraídos del archivo PCA_short_Spot.xls, de C.Alexander, para el periodo: 4/1/1997 a 31/12/2007. Las cuatro series contienen una raíz unitaria (es decir, son integradas de orden 1, $I(1)$). Los estadísticos de la traza y del máximo autovalor de Johansen para el contraste de cointegración, utilizando 5 retardos son: 149,93 y 103,32, con p-values de 0,0, para la existencia de un máximo de 0 relaciones de cointegración; son 46,61 y 33,10 con p-values asimismo nulos cuando se contrasta la hipótesis nula de un máximo de 1 relación de cointegración, y son de 13,51 y 10,30, con p-values de 0,097 y 0,19 cuando se contrasta la hipótesis nula de un máximo de 2 relaciones de cointegración. Por tanto, esta hipótesis no se rechaza, y pasamos a creer que existen 2 relaciones de cointegración.

Las dos relaciones de cointegración se estiman:

$$\begin{aligned}
r_{1,t} - 1,479r_{12,t} + 0,855r_{60,t} &= u_{2t} \\
(0,072) & \quad (0,105) \\
r_{3,t} - 1,438r_{12,t} + 0,753r_{60,t} &= u_{2t} \\
(0,058) & \quad (0,085)
\end{aligned}$$

donde puede verse que la diferencia de ambas implica, muy aproximadamente, que $r_{1,t} - r_{3,t}$ es estacionario. La segunda relación de cointegración sería una combinación más compleja entre los tipos de interés.

Los coeficientes de respuesta a los dos términos de corrección del error son ambos significativos en las ecuaciones de las ecuaciones de $\Delta r_{1,t}$ y $\Delta r_{3,t}$ y no son significativos en las ecuaciones de $\Delta r_{12,t}$ y $\Delta r_{60,t}$, indicando que son los tipos a corto plazo los que se ajustan para corregir desviaciones respecto de las dos relaciones de equilibrio a largo plazo. Son además, numéricamente más elevados en la ecuación de $\Delta r_{1,t}$ que en la ecuación de $\Delta r_{3,t}$, sugiriendo que el tipo a 1 mes es quien soporta la mayor parte del ajuste.

4.4 Cointegración por umbrales

Es habitual que los precios de contado y de contratos de futuros de un activo que se negocia con suficiente liquidez estén cointegrados. En el caso de la renta variable, si $f_{t,l}$ denota el logaritmo del precio del futuro con vencimiento en l ($l > t$) y s_t es el log-precio del contado, una versión del modelo de *cost-of-carry* específica que:

$$f_{t,l} - s_t = (r_{t,l} - q_{t,l})(l - t) + z_t$$

donde $r_{t,l}$ es el tipo de interés en t con vencimiento en l y $q_{t,l}$ es la tasa de dividendo durante el periodo (t, l) . El tiempo a vencimiento es $l - t$. El proceso z_t debe carecer de raíces unitarias, pues de lo contrario, existirían oportunidades persistentes de arbitraje. Consistirían en tomar una posición corta en el índice y una posición comprada en el futuro cuando los logaritmos de los precios se separan en más del *cost-of-carry*, siendo más alto el precio del contado. La estrategia contraria se tomaría si es el precio del futuro el que es más alto. Para que el arbitraje resulte económicamente interesante, z_t debe exceder de la suma de costes de transacción y de cualquier otro factor que acarree costes.

Es habitual que la serie de un índice como S&P500 tenga una raíz unitaria, al igual que el precio de sus contratos de futuros, pero que resulten cointegrados una vez que se ajusta por tipos de interés y por dividendos. Tras el ajuste, es frecuente que $[1, -1]$ sea el vector de cointegración, como se recoge en la ecuación anterior, y la serie cointegrada es z_t . Por tanto, debemos especificar el VECM para las series de rentabilidades: $r_t = (\Delta f_t, \Delta s_t)'$, donde Δ denota primeras diferencias.

Las consideraciones de costes que hemos hecho conducen a una especificación:

$$\begin{aligned}
r_t &= \mu_{1t} + \Gamma_1 r_{t-1} + \Gamma_2 r_{t-2} + \dots + \Gamma_p r_{t-p} + \beta_1 z_{t-1} + \epsilon_{1t}, \text{ si } z_{t-1} \leq \gamma_1 \\
r_t &= \mu_{2t} + \Upsilon_1 r_{t-1} + \Upsilon_2 r_{t-2} + \dots + \Upsilon_p r_{t-p} + \beta_2 z_{t-1} + \epsilon_{2t}, \text{ si } \gamma_1 < z_{t-1} \leq \gamma_2 \\
r_t &= \mu_{3t} + \Lambda_1 r_{t-1} + \Lambda_2 r_{t-2} + \dots + \Lambda_p r_{t-p} + \beta_3 z_{t-1} + \epsilon_{3t}, \text{ si } z_{t-1} \geq \gamma_2
\end{aligned}$$

por lo que tenemos un modelo con 3 regímenes. El arbitraje solo será posible si z_t es suficientemente grande, es decir, en los regímenes 1 y 3, en cada uno de ellos tomando una posición opuesta a la del otro régimen. El segundo régimen corresponde al funcionamiento normal del mercado, y esperaríamos que $\beta_2 = 0$, pues las dos series de precios se comportarían como un camino aleatorio. [ver ejemplo con datos en Sección 8.7.2 en Tsay]

4.5 Aplicaciones

4.5.1 Index Benchmark tracking

Case Study II.5.4.7 Alexander y Dimitriou (2005), "Indexing and statistical arbitrage: Tracking error or cointegration", *Journal of Portfolio Management*, 31(2), 50-63.

La construcción de una cartera que pueda replicar muy aproximadamente el comportamiento de un índice ofrece muchas posibilidades en la gestión de carteras. Ha sido habitual tratar de caracterizar la cartera réplica que más se aproxima al índice siguiendo procedimientos de mínimos cuadrados. Para ello, se estima una regresión de la rentabilidad del índice (si el objetivo es replicar el índice) sobre las rentabilidades de un conjunto de acciones individuales, y se utilizan los coeficientes de dicha regresión como pesos para configurar la cartera. Los signos positivos y negativos se corresponderían con posiciones largas y cortas en cada activo. El residuo es el tracking error, y su desviación típica es el tamaño de dicho error.

Pero si la cartera resultante no está cointegrada con el índice (y no hay garantía de que lo esté), su precio podría experimentar desviaciones duraderas respecto del índice. Esto se evita utilizando procedimientos de cointegración, como se hace en este Case Study para el Dow Jones IA, para conseguir un *tracking error* estacionario.

Siguiendo el método de Engle-Granger para analizar la cointegración entre un índice y un conjunto de n activos, estimaríamos el modelo:

$$\ln(I_t) = \alpha + \sum_{i=1}^n \beta_i \ln(P_{it}) + \varepsilon_t$$

Si los residuos son estacionarios, habríamos encontrado una cartera cointegrada con el índice I_t , combinando los n activos con ponderaciones:

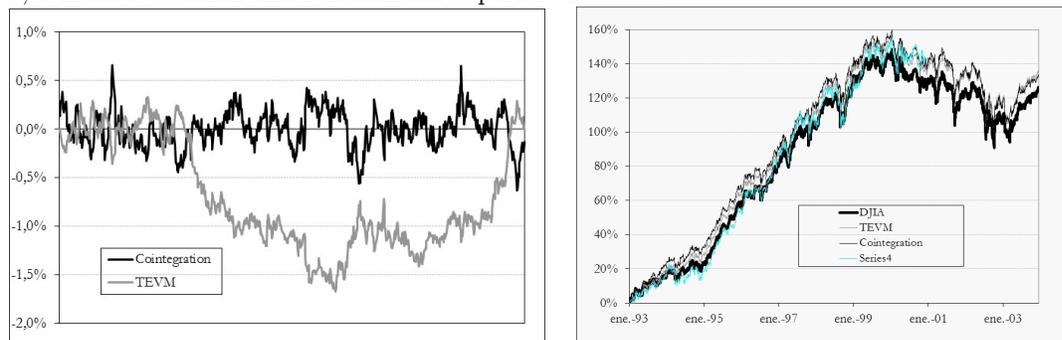
$$\beta^* = \left(\sum_{i=1}^n \beta_i \right)^{-1} (\beta_1, \beta_2, \dots, \beta_n)'$$

Nótese que así resolvemos el problema de *asset allocation*, es decir, de distribución del importe nominal entre los n activos, pero no el problema de selección de dichos activos. Generalmente, cabría probar con distintos conjuntos de activos para encontrar una cartera que combine una evidencia clara de cointegración con el índice, a la vez que una menor varianza del error de ajuste. Generalmente, este análisis requerirá tomar posiciones cortas y largas en algunos de los activos.

Un procedimiento de gestión pasiva de la cartera del fondo que comparase ambos procedimientos durante el periodo muestral tomaría una forma del siguiente tipo:

1. estimar el modelo con 5 años de datos diarios para calcular los pesos de la cartera en ambos métodos
2. mantener la cartera durante 2 semanas; al término de ellas, anotar las rentabilidades y desplazar la muestra dos semanas para volver a estimar los pesos
3. repetir hasta agotar los datos.

En este tipo de ejercicios, el analista debe decidir sobre cuál es el número de activos que va a utilizar en la cartera réplica, qué submuestra va a utilizar en la primera estimación de las ponderaciones, y con qué frecuencia realizará el rebalanceo de la cartera. El gráfico superior muestra los residuos que se obtiene cuando se sigue un enfoque de mínimos cuadrados (Tracking error de mínima varianza, TEVM), o un enfoque de cointegración. Como puede verse, las desviaciones del modelo respecto del índice son estacionarias en el segundo caso, mientras que las desviaciones que surgen al estimar las ponderaciones por mínimos cuadrados pueden tener carácter permanente. A pesar de ello, el gráfico inferior muestra que ambos enfoques hacen un buen papel en la réplica del índice, e incluso obtienen una rentabilidad superior al mismo.

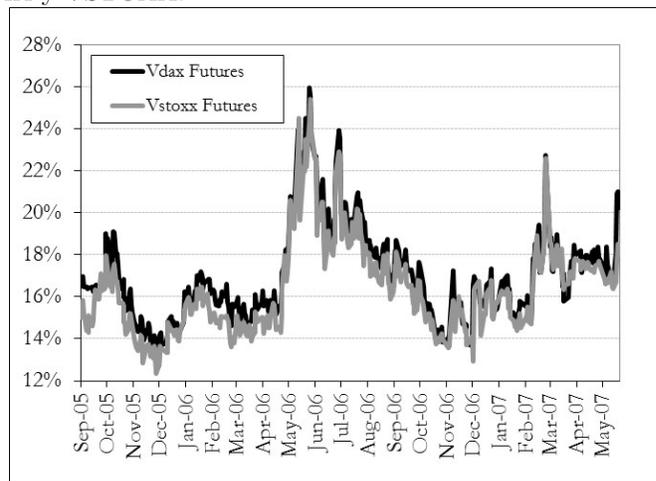


Alexander y Dimitriu (2005) "Indexing, cointegration and equity market regimes", *International Journal of Finance and Economics*, 10, 213-231, explican la preferencia por el método de cointegración. Sin embargo, hay que tener

presente la posibilidad de encontrar más de una cartera cointegrada con el índice que se pretende replicar. Actualmente, se propone seguir estrategias que toman una posición larga en la cartera de cointegración, y una posición corta en el futuro sobre el índice.

4.5.2 Pairs trading

[Case Study II.5 Pairs] Una estrategia que está popularizándose se basa en que si los precios de dos activos están cointegrados, su spread será estacionario. Esto es lo que sucede con dos índices de volatilidad sobre contratos de futuros: VDAX y VSTOXX.

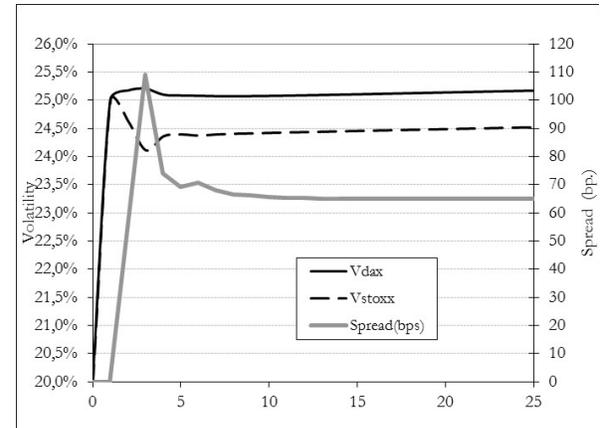
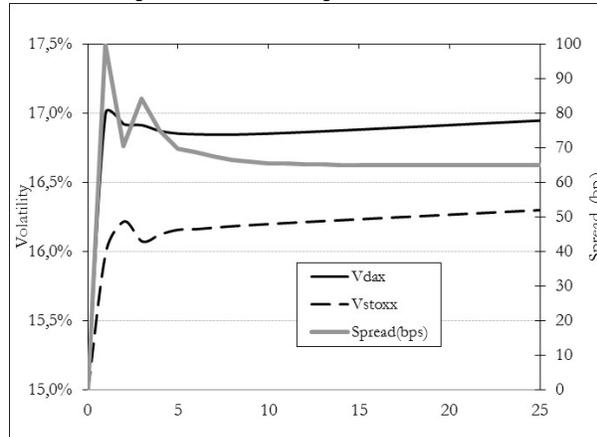


Teniendo cuidado en utilizar el spread en las mismas unidades que los índices de volatilidad (por ejemplo, como porcentajes; en todo caso, el spread se calcula por diferencia de ambos índices, por lo que generalmente no habrá problemas). Tras comprobar la cointegración de ambos índices de volatilidad, en el Case Study se estima un VECM que, posteriormente, se re-escribe en formato regresión, dejando un índice a la izquierda de cada ecuación, no sus primeras diferencias. Esta representación nos permite calcular las respuestas a shocks en ambos índices de volatilidad, así como la evolución que seguiría su spread en respuesta a ambos shocks.

El primer gráfico considera una situación en la que ambos índices de volatilidad comienzan de niveles estables de 15% y experimentan un shock de 1% en VDAX y de 2% en VSTOXX. En ausencia de otros shocks, los índices de volatilidad alcanzan un nuevo nivel de casi 16,94% para el futuro sobre VDAX y de 16,30% para el futuro sobre VSTOXX. Siendo estacionario, el spread revierte a su media de largo plazo, que estimamos mediante su media muestral 0,00649. Este análisis permitiría diseñar estrategias tomando posiciones contrarias en ambos activos. El segundo gráfico considera una situación en la que los índices están en 18% dos días antes de sufrir un shock excepcionalmente elevado, se sitúan en 19% un día antes, y en 20% el día que sufren el shock, pasando a un

25% al día siguiente. Eso hace que se establezcan en niveles más elevados que en la simulación anterior. Sin embargo, el spread, siendo estacionario, se estabiliza en torno al mismo nivel de 65 puntos básicos.

La posibilidad de predecir la evolución de dos índices cointegrados tras sufrir un determinado shock y, con ello, predecir la evolución de su diferencial, puede permitir tomar posiciones en el par de activos.



(ver asimismo "Pairs Trading: an empirical study", A. Fiz, tesina del Master en Banca y Finanzas Cuantitativas, Julio 2014)

5 Ejercicios

- Considere el modelo estructural recursivo,

$$\begin{aligned}y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + \varepsilon_{1t} \\y_{2t} &= \alpha_{20} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{2t-1} + \varepsilon_{2t}\end{aligned}$$

donde y_{1t} afecta a y_{2t} sólo con cierto retraso. Observe que este modelo permite identificar el término de error ε_{2t} a partir de las observaciones de la variable y_{2t} . Pruebe que este modelo está exactamente identificado, en el sentido de que todos sus coeficientes, así como las varianzas de los dos términos de error pueden recuperarse a partir de la estimación del modelo VAR(1) en estas dos variables.

$$\begin{aligned}\beta_{10} &= \alpha_{10} + \alpha_{11}\alpha_{20}; \beta_{11} = \alpha_{12} + \alpha_{11}\alpha_{22}; \beta_{12} = \alpha_{13} + \alpha_{11}\alpha_{23}; \\ \beta_{20} &= \alpha_{20}; \beta_{21} = \alpha_{22}; \beta_{22} = \alpha_{23}; \\ \sigma_{u_1}^2 &= \sigma_{v_1}^2 + \alpha_{11}^2\sigma_{v_2}^2; \sigma_{u_2}^2 = \sigma_{v_2}^2; \sigma_{u_1, u_2} = \alpha_{11}\sigma_{v_2}^2;\end{aligned}$$

sistema que puede resolverse para obtener los 9 parámetros del modelo estructural recursivo.

Muestre que en este modelo, no sólo se pueden recuperar estimaciones de todos los parámetros que aparecen en el modelo estructural, sino también las series temporales de los términos de error v_{1t} y v_{2t} .

6 Técnicas de cointegración en el análisis de “Asset allocation”

Válidas para decisiones de inversión a largo plazo

Si las decisiones se toman en base a tendencias de largo plazo, no requieren una actualización muy frecuente.

El análisis de cointegración trata de maximizar la estacionariedad y, con ello, minimizar la varianza del “tracking error”. En contraste con el análisis de frontera eficiente, en el que nada asegura que los “tracking errors” sean “mean-reverting”.

Los modelos de benchmarking o de index tracking utilizan generalmente regresiones con logaritmos de precios. La variable dependiente puede ser el log de un índice más un pequeño incremento que equivalga a un $x\%$ anual. Las variables explicativas son los logs de los precios de los activos que pueden incluirse en la cartera que sigue al índice.

El problema tiene dos partes: a) seleccionar los activos, b) optimizar las ponderaciones de la cartera. El primero es difícil. Opciones: “fuerza bruta”: estimar muchos modelos con distintas combinaciones de activos y ver cuál ajusta

mejor. Basarse en las preferencias de riesgo del inversor, o en las limitaciones de inversión que se nos impongan. Segundo problema: Técnicas de regresión, de análisis de series temporales multivariante, o de cointegración. Si se construye una cartera mediante regresión o cointegración, las ponderaciones se normalizan de modo que sumen 1, para ser ponderaciones de cartera.

En global asset management: primero, elegir los países, y después, comprar o vender futuros sobre los índices de los países, o repetir el problema dentro de cada país, para obtener carteras de seguimiento de cada índice.

En modelos de un sólo país: Primero, seleccionar los sectores; después, seleccionar los activos dentro de cada sector.

Constrained allocations:

Ejemplos:

- Seleccionar una cartera que invierte al menos un 50% en España.
- Fijar como rentabilidad benchmark la del SP100 más un 5% anual
- Construir un fondo corto-largo en 12 países prefijados, con un índice mundial como benchmark

Utilizar como variable dependiente, el índice que se quiere replicar, menos w -veces el precio del activo en el que se nos restringe a tomar una posición del $w\%$. Si se nos exige una posición en el activo “no superior al $w\%$ ” el problema es más difícil. Primero, resolver el problema sin restringir, a ver si hay suerte: ¿satisface la solución la restricción?

Supongamos que no podemos tomar posiciones cortas. Si alguna ponderación es negativa, se fija en cero, y se vuelve a resolver, iterando de este modo. Confiar en alcanzar una solución.

Selección de parámetros:

El modelo básico de “index tracking” mediante cointegración se define en función de ciertos parámetros:

- Una rentabilidad “alfa” por encima de l índice
- El intervalo de datos diarios que se utiliza en la estimación “training period”
- La relación de activos en la cartera
- Las posibles restricciones

Los parámetros se selecciona en base al resultado de pruebas dentro y fuera de la muestra.

Dentro de la muestra:

- Estadísticos ADF
- Standard error of regresión
- Turnover

“Testing period”:

- Tracking error variance: varianza de los errors de réplica diarios RMSE
- Diferencial de rentabilidades entre la cartera y el índice
- Information ratio: (Mean daily tracking error)/(desviación típica del daily tracking error) a lo largo del “testing period”
- La decisión puede basarse sobre un “alfa” o sobre el número de activos que queremos incluir en la cartera.

7 Apéndices

7.1 Las innovaciones de un modelo estructural deben estar incorrelacionadas entre sí.

De hecho, si dicha covarianza no fuese nula, podríamos transformar el modelo del siguiente modo: proyectaríamos uno de los dos errores, v_{2t} , por ejemplo, sobre v_{1t} ,

$$v_{2t} = \rho v_{1t} + a_t$$

obteniendo que el residuo \hat{a}_t , definido por $\hat{a}_t = v_{2t} - \hat{\rho}v_{1t}$, estaría incorrelacionado, por construcción, con v_{1t} .

Si representamos el modelo estructural en forma matricial,

$$\begin{pmatrix} 1 & -\alpha_{11} \\ -\alpha_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{10} \\ \alpha_{20} \end{pmatrix} + \begin{pmatrix} \alpha_{12} & \alpha_{13} \\ \alpha_{22} & \alpha_{23} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

y premultiplicamos por la matriz $\begin{pmatrix} 1 & 0 \\ -\hat{\rho} & 1 \end{pmatrix}$, lo que equivale a sustituir la segunda ecuación por la diferencia entre ésta y el resultado de multiplicar la primera ecuación por $\hat{\rho}$, tendríamos,

$$\begin{aligned} y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + v_{1t} \\ (1 + \hat{\rho}\alpha_{11})y_{2t} &= (\alpha_{20} - \hat{\rho}\alpha_{10}) + (\hat{\rho} + \alpha_{21})y_{1t} + (\alpha_{22} - \hat{\rho}\alpha_{12})y_{1t-1} + (\alpha_{23} - \hat{\rho}\alpha_{13})y_{2t-1} + \hat{a}_t \end{aligned} \quad (30)$$

un modelo VAR en el que, una vez despejáramos y_{2t} en la segunda ecuación, sería indistinguible del modelo (7) con $Cov(v_{1t}, \hat{a}_t) = 0$. Siempre debemos estar considerando esta última representación con errores ortogonalizados, por lo que la condición de ausencia de correlación entre los errores de las distintas ecuaciones en el modelo VAR estructural debe satisfacerse siempre.

7.2 Errata en Enders, página 299,

$$Var(Y_t) = E[(Y_t - \mu)^2] = E\left[\sum_{i=0}^{\infty} A_1^i u_{t-i}\right]^2 = \sum_{i=0}^{\infty} A_1^{2i} (Var(u_{t-i})) = (I_k - A_1^2)^{-1} \Sigma$$

$$Var(Y_t) = (I_2 - A_1^2)^{-1} \Sigma = \frac{1}{M} \begin{pmatrix} \beta_{21}\beta_{12} + \beta_{22}^2 & -(\beta_{11} + \beta_{22})\beta_{12} \\ -(\beta_{11} + \beta_{22})\beta_{21} & \beta_{21}\beta_{12} + \beta_{11}^2 \end{pmatrix}$$

$$\text{con } M = [1 - (\beta_{21}\beta_{12} + \beta_{11}^2)] [1 - (\beta_{21}\beta_{12} + \beta_{22}^2)] - (\beta_{11} + \beta_{22})^2 \beta_{12}\beta_{21}.$$