

Modelos vectoriales autoregresivos (VAR)

Alfonso Novales
Universidad Complutense

Septiembre 2011

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 las distintas ecuaciones. 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.

Así, en un modelo vectorial autoregresivo de primer orden, VAR(1), las variables explicativas de cada ecuación son: una constante, más un retardo de cada una de las variables del modelo. Si el modelo pretende explicar el comportamiento temporal de 3 variables, habría 3 variables explicativas, más constante, en cada ecuación, para un total de 12 coeficientes a estimar. Si el modelo fuera de segundo orden, VAR(2), habría 7 coeficientes a estimar en cada una de las 3 ecuaciones que componen el modelo VAR. Como puede verse, todas las variables son tratadas simétricamente, siendo explicadas por el pasado de todas ellas.

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 un 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.

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.

Por el contrario, en un modelo VAR todas las variables se tratan de igual modo: el modelo tienen 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.

2. El modelo VAR(1)

En el caso más simple, con sólo dos variables y un retardo, el modelo VAR₂(1) es,

$$\begin{aligned} y_{1t} &= \beta_{10} + \beta_{11}y_{1t-1} + \beta_{12}y_{2t-1} + u_{1t} \\ y_{2t} &= \beta_{20} + \beta_{21}y_{1t-1} + \beta_{22}y_{2t-1} + u_{2t} \end{aligned} \quad (2.1)$$

o, en forma matricial,

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \quad (2.2)$$

donde los términos de error satisfacen,

$$\begin{aligned}
E(u_{1t}) &= E(u_{2t}) = 0, \forall t \\
E(u_{1t}u_{1s}) &= E(u_{2t}u_{2s}) = E(u_{1t}u_{2s}) = 0, \forall t \neq s \\
Var \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \Sigma, \forall t
\end{aligned} \tag{2.3}$$

En el modelo VAR anterior, valores negativos de β_{12} y β_{21} tienden a inducir correlación negativa entre y_{1t} e y_{2t} si bien no la garantizan.

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} , influye sobre y_{1t} , a través de de la correlación entre las innovaciones de ambas variables. En general, una sorpresa en y_{2t} vendrá acompañada de un valor no nulo de la innovación u_{1t} , salvo en el caso excepcional en que $\sigma_{u_1u_2} = 0$. Estos efectos se propagan en el tiempo debido a la presencia de los valores retardados como variables explicativas.

En general, un modelo VAR se especifica,

$$Y_t = A_0 + \sum_{s=1}^K A_s Y_{t-s} + u_t \tag{2.4}$$

donde Y_t es un vector columna $nx1$, K es el orden del modelo VAR, o número de retardos de cada variable en cada ecuación, y u_t es un vector $nx1$ de innovaciones, es decir, procesos sin autocorrelación, con $Var(u_t) = \Sigma$, constante. El elemento (i, j) en la matriz A_s , $1 \leq s \leq K$ mide el *efecto directo* o *efecto parcial* de un cambio en Y_j en el instante t sobre la variable explicativa al cabo de s períodos, $Y_{i,t+s}$. 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 . Con esta notación el modelo VAR(1) se escribiría: $Y_t = A_0 + A_1 y_{t-1} + u_t$.

3. Un modelo estructural

Es útil interpretar el modelo VAR como *forma reducida* de un modelo estructural,

$$\begin{aligned}
y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + \varepsilon_{1t} \\
y_{2t} &= \alpha_{20} + \alpha_{21}y_{1t} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{2t-1} + \varepsilon_{2t}
\end{aligned} \tag{3.1}$$

donde y_{1t}, y_{2t} son variables estacionarias, y $\varepsilon_{1t}, \varepsilon_{2t}$ son *innovaciones*, procesos ruido blanco con esperanza cero y varianzas $\sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2$. Este es un modelo de

ecuaciones simultáneas con la única peculiaridad de que sus dos variables son endógenas. Un *shock* inesperado en 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 como variables explicativas. Es natural pensar que los términos de error del modelo estructural estén mutuamente incorrelacionados, puesto que la correlación contemporánea entre y_{1t} e y_{2t} ya está capturada por la presencia de sus valores contemporáneos como variables explicativas en ambas ecuaciones. Por tanto, suponemos que $Cov(\varepsilon_{1t}, \varepsilon_{2t}) = \sigma_{\varepsilon_1, \varepsilon_2} = 0$.

Si dicha covarianza no fuese cero, el sistema podría transformarse en otro sistema, observacionalmente equivalente, con tal propiedad (ver Apéndice).

De forma resumida, la representación matricial del modelo estructural anterior puede escribirse,

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

con,

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

y si suponemos que la matriz B tiene inversa, lo cual requiere que $\alpha_{11}\alpha_{21} \neq 1$, tenemos,

$$y_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 y_{t-1} + B^{-1}\varepsilon_t = A_0 + A_1 y_{t-1} + u_t$$

donde,

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

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = B^{-1}\varepsilon_t = B^{-1} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \frac{1}{1 - \alpha_{11}\alpha_{21}} \begin{pmatrix} \varepsilon_{1t} + \alpha_{11}\varepsilon_{2t} \\ \varepsilon_{2t} + \alpha_{21}\varepsilon_{1t} \end{pmatrix} \quad (3.2)$$

$$A_1 = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} = \frac{1}{1 - \alpha_{11}\alpha_{21}} \begin{pmatrix} \alpha_{12} + \alpha_{11}\alpha_{22} & \alpha_{13} + \alpha_{11}\alpha_{23} \\ \alpha_{22} + \alpha_{21}\alpha_{12} & \alpha_{23} + \alpha_{21}\alpha_{13} \end{pmatrix} \quad (3.3)$$

$$A_0 = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} = \frac{1}{1 - \alpha_{11}\alpha_{21}} \begin{pmatrix} \alpha_{10} + \alpha_{11}\alpha_{20} \\ \alpha_{20} + \alpha_{21}\alpha_{10} \end{pmatrix} \quad (3.4)$$

con lo que habremos pasado a la forma reducida, o modelo VAR.

Como puede verse, si los términos de error del modelo estructural eran ruido blanco, también los términos de error del modelo VAR tendrán estructura ruido blanco. Sin embargo, las innovaciones del VAR estarán correlacionadas entre sí, puesto que,

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

de modo que, incluso si los términos de error del modelo estructural están incorrelacionados, $\sigma_{\varepsilon_1\varepsilon_2} = 0$, las perturbaciones del modelo VAR tendrán correlación no nula, a no ser que $\alpha_{11} = \alpha_{21} = 0$.

Es 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 (??) y (3.3), más las 3 relaciones entre los elementos de las respectivas matrices de covarianzas,

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

4. Identificación en un modelo VAR

Definition 4.1. *Recuperación de los parámetros del modelo estructural a partir de estimaciones de los parámetros del modelo en forma reducida.*

La estimación de un modelo VAR(1) bivalente 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 del vector u_t en (2.3). Sin embargo, el modelo estructural consta de 10 parámetros: las dos constantes, los 6 coeficientes, y los 3 parámetros de la matriz de covarianzas del vector ε_t , por lo que no es posible recuperar los parámetros del modelo estructural.

Recordemos que los dos términos de perturbación del modelo en forma estructural tienen correlación cero.

En el ejercicio 1 se prueba que el modelo estructural recursivo bivalente de orden 1,

$$\begin{aligned} y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + \varepsilon_{1t} \\ y_{2t} &= \alpha_{20} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{2t-1} + \varepsilon_{2t} \end{aligned} \quad (4.1)$$

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.

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 las relaciones,

$$\hat{u}_{2t} = \frac{\hat{\varepsilon}_{2t}}{1 - \alpha_{11}\alpha_{21}}; \quad \hat{u}_{1t} = \frac{\hat{\varepsilon}_{1t} + \alpha_{11}\hat{\varepsilon}_{2t}}{1 - \alpha_{11}\alpha_{21}}$$

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* respecto de y_{1t} , puesto que $\text{Cov}(\varepsilon_1, \varepsilon_2) = 0$. Examinando los modelos anteriores, es fácil ver que las dos restricciones que hemos impuesto, $\alpha_{21} = \alpha_{22} = 0$ hacen que en el modelo VAR, $\beta_{21} = 0$, restricción que puede contrastarse utilizando el estadístico tipo t habitual de dicho coeficiente, sin más que las dificultades habituales en el uso de este estadístico.

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 todo un continuo de 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.

El problema de obtener las innovaciones estructurales a partir de las las estimaciones de los residuos del modelo VAR equivale a la posibilidad de disponer de valores numéricos para los elementos de la matriz B , puesto que $\varepsilon_t = Bu_t$. 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; recuérdese que sus covarianzas son nulas. Así, tenemos 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, $(k^2 - k)/2$ restricciones adicionales, si queremos tener alguna posibilidad de identificar el modelo. En el caso de un modelo VAR(1) con 2 variables, hemos de imponer $(2^2 - 2)/2 = 1$ restricciones 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.

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} , por lo que el modelo estaría, en principio, exactamente identificado. Esta estructura

recursiva es consistente con una estructura de covarianzas que se conoce como de tipo Cholesky, dado que la matriz que transforma el vector ε en el vector u es triangular inferior. La recursividad mediante una matriz B trinagular inferior o superior, como en este caso, proporciona siempre el número exacto de restricciones que se precisan para identificar un modelo VAR, que es de $k^2 - k$.

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.

Otro tipo de restricciones consistiría en imponer un determinado valor numérico para una respuesta. Por ejemplo, podemos pensar que la innovación ε_{2t} tiene un efecto unitario sobre y_{1t} . Esto equivaldría a suponer $\alpha_{11} = -1$ en la matriz:

$$\varepsilon_t = Bu_t = \begin{pmatrix} 1 & -\alpha_{11} \\ -\alpha_{21} & 1 \end{pmatrix} u_t$$

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 covarianza entre ambos. 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. En el caso del modelo de 2 variables, esta condición de simetría de efectos conduce asimismo a una condición de igualdad de varianzas para las innovaciones estructurales, lo que no ocurre en modelos con más de 2 variables.

4.1. Identificación y respuestas del sistema

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) bivalente, (2.2), en el que queremos calcular cómo reacciona cada variable ante una innovación en una de ellas, lo que denominamos como *funciones de*

respuesta al impulso. Sería poco adecuado, sin embargo, calcular 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, hemos de transformar primero el modelo estimado en otro modelo en que los términos de error, siendo innovaciones, estén incorrelacionados entre sí. Para ello, podríamos seguir una estrategia similar a la discutida más arriba, proyectando por mínimos cuadrados una de los dos innovaciones, u_{1t} , por ejemplo, sobre u_{2t} ,

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

cuyo residuo \hat{a}_t , definido por $\hat{a}_t = u_{1t} - \hat{\rho}u_{2t}$, estaría incorrelacionado, por construcción, con u_{2t} .

Premultiplicando el modelo (2.2) por la matriz $\begin{pmatrix} 1 & -\hat{\rho} \\ 0 & 1 \end{pmatrix}$, tendríamos,

$$\begin{aligned} y_{1t} &= (\beta_{10} - \hat{\rho}\beta_{20}) + \hat{\rho}y_{2t} + (\beta_{11} - \hat{\rho}\beta_{21})y_{1t-1} + (\beta_{12} - \hat{\rho}\beta_{22})y_{2t-1} + \hat{a}_t \\ y_{2t} &= \beta_{20} + \beta_{21}y_{1t-1} + \beta_{22}y_{2t-1} + u_{2t} \\ Cov(\hat{a}_t, u_{2t}) &= 0, \end{aligned}$$

un modelo en el que la variable y_2 tiene efectos contemporáneos sobre y_1 . Este es el modelo estructural exactamente identificado (4.1) que antes consideramos.

En este modelo, tiene sentido preguntarse por las respuestas de ambas variables a una perturbación en \hat{a}_t o en u_{2t} , puesto que ambos están incorrelacionados, por construcción. En respuesta a un impulso en u_{2t} , ambas variables reaccionarán en el mismo instante, y también en períodos siguientes, hasta que dichas respuestas decaigan a cero. En cambio, en respuesta a una perturbación en \hat{a}_t , y_1 responderá en el mismo período y períodos siguientes, mientras que y_2 sólo responderá en períodos siguientes al de la perturbación.

4.1.1. Generalizando el orden del VAR

Como es sabido, dada una matriz simétrica, definida positiva, como Σ , existe una única matriz triangular inferior A , 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 admite una descomposición,

$$\Sigma = ADA'$$

Si consideramos la transformación lineal del vector de error precisamente con esta matriz, $\varepsilon_t = A^{-1}u_t$, tenemos,

$$Var(\varepsilon_t) = E(\varepsilon_t \varepsilon_t') = E(A^{-1}u_t u_t' (A^{-1})') = E(A^{-1}\Sigma (A^{-1})') = D,$$

que es una matriz diagonal, por lo que, a diferencia de los componentes del vector u , los elementos del vector ε están incorrelacionados entre sí. Deshaciendo la transformación, tenemos,

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ \dots \\ u_{kt} \end{pmatrix} = A\varepsilon_t = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{12} & 1 & 0 & \dots & 0 \\ a_{13} & a_{23} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{1k} & a_{2k} & a_{3k} & \dots & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \dots \\ \varepsilon_{kt} \end{pmatrix}$$

por lo que,

$$\varepsilon_{kt} = u_{kt} - a_{1k}\varepsilon_{1t} - a_{2k}\varepsilon_{2t} - \dots - a_{k-1,k}\varepsilon_{k-1,t} \quad (4.2)$$

Si los coeficientes $a_{1k}, a_{2k}, \dots, a_{k-1,k}$ se obtienen mediante una estimación de mínimos cuadrados ordinarios de la ecuación (4.2), que tiene a u_{kt} como variable dependiente, y a $\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{k-1,t}$ como variables explicativas,

$$\varepsilon_{kt} = u_{kt} - \hat{a}_{1k}\varepsilon_{1t} - \hat{a}_{2k}\varepsilon_{2t} - \dots - \hat{a}_{k-1,k}\varepsilon_{k-1,t} \quad (4.3)$$

entonces tendremos, por construcción, $E(\varepsilon_{kt} \cdot \varepsilon_{1t}) = E(\varepsilon_{kt} \cdot \varepsilon_{2t}) = \dots = E(\varepsilon_{kt} \cdot \varepsilon_{k-1,t}) = 0$. 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, llamémosle ε_{it} , tendremos un componente de u_{it} que, por construcción, estará incorrelacionado con $u_{1t}, u_{2t}, \dots, u_{i-1,t}$. Nótese que los espacios generados por las variables $u_{1t}, u_{2t}, \dots, u_{i-1,t}$ y por las variables $\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{i-1,t}$ son los mismos, es decir, que ambos conjuntos de variables contienen la misma información. La única diferencia entre ambos es que las variables $u_{1t}, u_{2t}, \dots, u_{i-1,t}$ tiene correlaciones no nulas, mientras que las variables $\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{i-1,t}$ están incorrelacionadas entre sí.

5. Condiciones de estabilidad

Si resolvemos recursivamente el modelo VAR(1) tenemos,

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

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. Recordemos que en este modelo: $A = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}$.

Cuando se cumplen las condiciones de estabilidad, tomando límites, tenemos,

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

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

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

Además,

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

En el caso bivalente, $\mu_1 = E(u_{1t}), \mu_2 = E(u_{2t})$, con

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

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

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

Un modelo VAR estable define momentos incondicionales para cada una de las variables del vector Y_t . En ese caso, hay que distinguir entre la distribución y los momentos incondicionales y condicionales del vector Y_t

6. 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 VAR(1), como (2.2), 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,

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

que es un proceso ARMA(2,1).

7. Estimación de un modelo VAR

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 variable explicativas, el menor número de retardos que permita eliminar la autocorrelación residual en todas las ecuaciones. Existen contrastes del tipo de razón de verosimilitud sobre el número de retardos a incluir en el modelo.

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.

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

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

siendo $\hat{\Sigma}$ la estimación de la matriz de covarianzas del vector de innovaciones u ,

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

una matriz simétrica, definida positiva, por construcción.

8. Contratación de hipótesis

8.1. 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 . Si, por ejemplo, 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. Por eso ya comentamos con anterioridad que debe incluirse en cada ecuación 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, por el número de variables al cuadrado.

Sin embargo, no puede olvidarse que la elección del número de retardos debe tener muy en cuenta la eliminación de autocorrelación residual 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 y Schwartz,

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

siendo $n = k(d + pk)$ el número de parámetros estimados en el modelo VAR. d es el número de variables exógenas, p el orden del VAR, y k el número de variables. En ocasiones, se ignora el término constante, y los criterios anteriores se aproximan por,

$$\begin{aligned}
AIC &= T \ln |\Sigma| + 2n \\
SBC &= T \ln |\Sigma| + n \ln(T)
\end{aligned}$$

siendo N el número de parámetros que se estima, y Σ la matriz de covarianzas de los residuos. 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.

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.

8.2. Contrastes de causalidad

Un contraste especialmente interesante es el conocido 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; 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

9. Representación MA de un modelo VAR

Para simplificar la notación, sin perder ningún elemento relevante del modelo, en esta sección ignoramos la presencia de constantes en las ecuaciones del modelo VAR. Un modo de justificar tal ausencia es pensar que las variables del modelo están en diferencias con respecto a sus respectivas medias muestrales.

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

$$Y_t = \sum_{s=0}^{\infty} B_s u_{t-s}$$

a la que se llega tras sucesivas sustituciones de Y_{t-s} en (2.4). La representación MA puede obtenerse asimismo en función de las innovaciones estructurales. 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 $Y_{j,t+s}$.

Si volvemos al modelo de dos variables de orden 1, tenemos,

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

que, como vimos, puede escribirse,

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

y, en términos de las innovaciones del modelo estructural, incorrelacionadas entre sí,

$$\begin{aligned} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \frac{1}{1 - \alpha_{11}\alpha_{21}} \sum_{s=0}^{\infty} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}^s \begin{pmatrix} 1 & \alpha_{11} \\ \alpha_{21} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-s} \\ \varepsilon_{2t-s} \end{pmatrix} \\ &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \sum_{s=0}^{\infty} \begin{pmatrix} \phi_{11}(s) & \phi_{12}(s) \\ \phi_{21}(s) & \phi_{22}(s) \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-s} \\ \varepsilon_{2t-s} \end{pmatrix} = \mu + \sum_{s=0}^{\infty} \Phi(s)\varepsilon_{t-s} \end{aligned} \quad (9.1)$$

donde,

$$\begin{pmatrix} \phi_{11}(s) & \phi_{12}(s) \\ \phi_{21}(s) & \phi_{22}(s) \end{pmatrix} = \frac{1}{1 - \alpha_{11}\alpha_{21}} \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}^s \begin{pmatrix} 1 & \alpha_{11} \\ \alpha_{21} & 1 \end{pmatrix} \quad (9.2)$$

Existe un procedimiento recursivo para obtener las matrices de coeficientes de la representación de medias móviles que utiliza la relación que buscamos,

$$\begin{aligned} Y_t &= A_1 Y_{t-1} + \dots + A_p Y_{t-p} + u_t = (I_k - A_1 L - A_2 L - \dots - A_p L^p)^{-1} u_t = \\ &= (\phi_0 + \phi_1 L + \phi_2 L^2 + \dots) u_t \end{aligned}$$

de modo que tenemos,

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

que conduce a,

$$\phi_0 = I_k; \phi_1 - A_1\phi_0 = 0; \phi_2 - A_1\phi_1 - A_2\phi_0 = 0; \dots$$

de donde, finalente, obtenemos:

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

que pueden utilizarse para calcular recursivamente las matrices de coeficientes de la representación de medias móviles.

Si trabajamos con un modelo VAR(1), es facil ver de lo anterior que las matrices ϕ_s son las sucesivas potencias de la matriz A_1 .

10. Funciones de respuesta al impulso

La ecuación (9.1) es la representación de medias móviles del modelo VAR(1) bi-variente. Los coeficientes de la sucesión de matrices $\Phi(s)$ 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 $\varepsilon_{1t}, \varepsilon_{2t}$. Por ejemplo, los coeficientes $\phi_{12}(s)$ reflejan el impacto que en los distintos períodos $s, s \geq 1$, tiene sobre y_1 una perturbación del tipo impulso en ε_2 .

Es decir, consideramos que ε_2 está en su valor de equilibrio, cero, excepto en un período, en que toma un valor igual a 1; como consecuencia, tanto y_1 como y_2 reaccionan, porque ε_{2t} aparece en ambas ecuaciones en (9.1), y dicha respuesta se extiende a varios períodos, hasta que la sucesión $\phi_{12}(s)$ se hace cero. La sucesión de valores numéricos $\{\phi_{12}(s)\}$ se conoce como la *respuesta de y_1 a un impulso en ε_2* . El efecto, *multiplicador o respuesta a largo plazo* es la suma $\sum_{s=0}^{\infty} \phi_{12}(s)$. Esta suma existe si las variables son estacionarias, pues en tal caso ha de cumplirse que $|\sum_{s=0}^{\infty} \phi_{12}(s)| < \infty$.

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 (9.2). En el modelo recursivo que antes vimos, se tiene $\alpha_{21} = 0$. Además, se prueba en el ejercicio 1 que en este modelo el parámetro α_{11} puede recuperarse mediante $\hat{\alpha}_{11} = \sigma_{u_1 u_2} / \sigma_{u_2}^2$. En ese caso, $u_{2t} = \varepsilon_{2t}$ y $u_{1t} = \varepsilon_{1t} + \alpha_{11} \varepsilon_{2t} = \varepsilon_{1t} + \alpha_{11} u_{2t}$.

Las funciones de respuesta al impulso sólo pueden obtenerse bajo restricciones de identificación de este tipo. La que hemos descrito es la más habitual, y equivale a admitir que una de las dos variables afecta a la otra sólo con retraso, si bien permitimos que en la otra dirección haya respuesta contemporánea. Estaremos caracterizando las respuestas del sistema a un impulso en cada una de las innovaciones del modelo estructural o, lo que es lo mismo, en la innovación u_{2t} y en $u_{1t} - \alpha_{11} u_{2t}$. Esta última es la componente de u_{1t} que no está explicada por u_{2t} o, si se prefiere, la componente de u_{1t} que no está correlacionada con u_{2t} ¹.

De hecho, si $\alpha_{21} = 0$, entonces $u_{1t} - \alpha_{11} u_{2t}$ es, precisamente, igual a la perturbación estructural ε_{1t} .

Como hemos visto, las funciones de respuesta al impulso sólo pueden obtenerse después de haber introducido restricciones acerca del retraso con que unas variables inciden sobre otras. Esta elección condiciona bastante, en general, el aspecto de las funciones de respuesta, excepto si las innovaciones del modelo VAR, u_{1t} y u_{2t} están incorrelacionadas, en cuyo caso, coinciden con las innovaciones del modelo estructural.

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. El inconveniente del segundo tipo de representación es que las respuestas de las distintas variables dependen de sus respectivas volatilidades, por lo que la comparación de las respuestas de dos variables diferentes a un determinado impulso no permite decir cuál de las variables *responde más*. Recordando

¹En general, si proyectamos u_{1t} sobre u_{2t} , el coeficiente estimado será igual a $\frac{Cov(u_1, u_2)}{\sqrt{Var(u_2)}}$. Pero $u_{1t} = \frac{\varepsilon_{1t} + \alpha_{11} \varepsilon_{2t}}{\Delta}$ y $u_{2t} = \frac{\varepsilon_{2t} + \alpha_{21} \varepsilon_{1t}}{\Delta}$, por lo que $Cov(u_1, u_2) = \frac{\alpha_{11} \sigma_{\varepsilon_2}^2 + \alpha_{21} \sigma_{\varepsilon_1}^2}{(1 - \alpha_{11} \alpha_{21})^2}$

que la desviación típica es una medida adecuada del tamaño de toda variable aleatoria de esperanza nula, debemos dividir las respuestas de cada variable por su desviación típica antes de representarlas en un mismo gráfico. Tampoco un impulso de tamaño unidad tiene el mismo significado en cada variable, por lo que conviene calcular las respuestas normalizadas a un impulso de tamaño igual a una desviación típica en cada innovación.

Consideremos un VAR(1) sin constante (es decir, las variables tiene esperanza igual a 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 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,

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

y examinar la sucesión definida en (9.2).

Nótese que en este modelo VAR las respuestas al impulso iniciales (en $t = 0$) son todas nulas.

11. Descomposición de la varianza

Si utilizamos la representación MA para obtener predicciones de las variables y_1, y_2 , tenemos,

$$E_t y_{t+n} = E_t \begin{pmatrix} y_{1t+n} \\ y_{2t+n} \end{pmatrix} = \mu + \sum_{s=n}^{\infty} \Phi(s) \varepsilon_{t+n-s}$$

donde Φ es la misma matriz que aparece en (9.1).

por lo que el error de predicción es,

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

cuya varianza es,

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

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,

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

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.

La descomposición de la varianza está sujeta al mismo problema de identificación que vimos antes para las funciones de respuesta al impulso, siendo necesario introducir alguna restricción 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 las restricciones de identificación de la sección anterior, ε_2 explica el 100% de la varianza del error de predicción un período hacia adelante en la variable y_2 . Si, en vez de dicha restricción, excluyéramos y_{2t} de la primera ecuación, entonces ε_1 explicaría el 100% de la varianza del error de predicción un período hacia adelante en la variable y_1 .

11.1. Identificación recursiva: la descomposición de Cholesky

Para eliminar la correlación contemporánea existente entre las innovaciones u_t de distintas ecuaciones, podemos transformar el vector u_t en un vector e_t mediante la transformación definida por la descomposición de Cholesky de la matriz de covarianzas Σ , $\Sigma = \text{Var}(u_t)$. Esta descomposición nos proporciona una matriz triangular inferior G tal que $GG' = \Sigma$. Como consecuencia, $G^{-1}\Sigma G'^{-1} = I$, y el sistema VAR puede escribirse,

$$Y_t = \sum_{s=0}^{\infty} A_s u_{t-s} = \sum_{s=0}^{\infty} (A_s G) (G^{-1} u_{t-s}) = \sum_{s=0}^{\infty} \tilde{A}_s e_{t-s} \quad (11.1)$$

con $\tilde{A}_s = A_s G$, $e_{t-s} = G^{-1} u_{t-s}$, $\text{Var}(e_{t-s}) = G^{-1} \text{Var}(u_{t-s}) G^{-1'} = I$.

El efecto de e_{it} sobre $Y_{j,t+s}$ viene medido por el elemento (j, i) de la matriz \tilde{A}_s . La sucesión de dichos elementos, para $1 \leq s \leq \infty$ proporciona la respuesta dinámica de la variable Y_j a una innovación en la variable Y_i . esto se conoce como función de respuesta de Y_j a un impulso sorpresa en Y_i . Como e_{it} es el error de predicción un período hacia adelante en Y_{it} , la representación MA ortogonalizada nos permite computar el error de predicción de Y_{it} , m -períodos hacia adelante, en el instante $t - m + 1$, a través del elemento i -ésimo en el vector $\sum_{s=0}^{m-1} \tilde{A}_s e_{t-s}$. Su varianza, el elemento i -ésimo en la diagonal de $\sum_{s=0}^{m-1} \tilde{A}_s \tilde{A}_s'$, puede escribirse, $\sum_{j=1}^K \sum_{s=0}^{m-1} \tilde{a}_s(i, j) \tilde{a}_s(j, i)$, siendo $\tilde{a}_s(i, j)$ el elemento (i, j) genérico de la matriz element \tilde{A}_s . Al aumentar m , a partir de $m = 1$, esta descomposición de la varianza del error de predicción de Y_{it+m} entre las k variables del vector Y_t se conoce como descomposición de la varianza de Y_{it} . Proporciona una estimación de la relevancia de cada variable del sistema para explicar los errores de predicción de las fluctuaciones futuras en Y_{it} .

12. 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. Note 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.

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

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 ε_{1t} y ε_{2t} .

13. Apéndice

13.1. Transformando un VAR con covarianza no nula en otro con tal propiedad

Supongamos que en el sistema,

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

se tiene: $Cov(u_{1t}, u_{2t}) = \sigma_{12} \neq 0$. Si estimamos la proyección: $u_{2t} = \beta u_{1t} + \varepsilon_{2t}$ por mínimos cuadrados, tendremos: $\beta = \frac{\sigma_{12}}{\sigma_1^2}$, con $Cov(\varepsilon_{2t}, u_{1t}) = 0$. Premultiplicando el sistema por la matriz $\begin{pmatrix} 1 & 0 \\ -\beta & 1 \end{pmatrix}$, se tiene:

$$\begin{aligned}y_{1t} &= \beta_{10}y_{2t} + u_{1t} \\y_{2t} &= \frac{\beta + \beta_{20}}{1 + \beta\beta_{10}}y_{1t} + \frac{1}{1 + \beta\beta_{10}}\varepsilon_{2t}\end{aligned}$$

cuyos dos términos de error tiene covarianza nula, como queríamos.

13.2. 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, ε_{2t} , por ejemplo, sobre ε_{1t} ,

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

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

Si representamos el modelo estructural en forma matricial,

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

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

$$\begin{aligned}
y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + \varepsilon_{1t} \\
(1 + \hat{\rho}\alpha_{11})y_{2t} &= (\alpha_{20} - \hat{\rho}\alpha_{10}) + (\hat{\rho} + \alpha_{21})y_{1t} + (\alpha_{22} - \hat{\rho}\alpha_{12})y_{1t-1} + (\alpha_{23} - \hat{\rho}\alpha_{13})y_{2t-1} + \hat{a}_t
\end{aligned} \tag{13.1}$$

un modelo VAR en el que, una vez despejáramos y_{2t} en la segunda ecuación, sería indistinguible del modelo (3.1) con $Cov(\varepsilon_{1t}, \hat{a}_t) = 0$. *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.*

13.3. Errata en Enders, página 299,

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

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

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