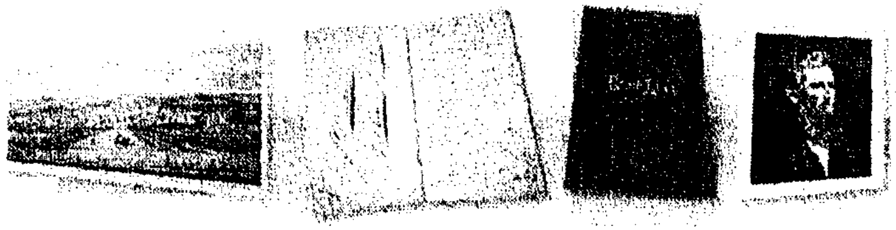


Capítulo 3
MODELOS ECONOMETRICOS

Antoni Espasa y José Ramón Cancelo

3.0. Introducción

Cuando en el análisis de un fenómeno económico se desea plantear un estudio que tenga en cuenta no sólo la información referente al propio fenómeno —problema univariante tratado en el capítulo anterior—, sino también la información correspondiente a otros fenómenos con los que el primero pueda estar relacionado, el analista cuantitativo pasa a desenvolverse en un contexto multivariante, que requiere, al menos en principio, un *marco multiecuacional* en el que se determinen todos los fenómenos considerados. La necesidad de este marco multiecuacional es especialmente imperiosa en el análisis de la economía, pues los fenómenos económicos están, normalmente, interrelacionados. En este libro se denomina *modelo econométrico* a aquel modelo que relaciona una variable económica con otras variables económicas.

El marco multiecuacional conlleva una complicación muy notable en el análisis a realizar, por lo que éste resulta más gravoso y la inversión necesaria para llevarlo a cabo más difícil de justificar, sobre todo si el interés se ciñe exclusivamente a una sola variable, el fenómeno inicialmente considerado. Por ello, en el análisis económico cuantitativo el analista, a menudo, desea utilizar la información correspondiente a otros fenómenos de interés en el marco de un modelo uniecuacional, que explique la generación del fenómeno investigado en función de sí mismo y de los otros, pero que ignore las restantes ecuaciones que explican cómo se determinan los fenómenos adicionales incorporados en el conjunto informativo.

Este modelo uniecuacional pertenece a un modelo más amplio, multiecuacional, y su tratamiento aislado sólo está justificado bajo determinadas condiciones en función de los objetivos del investigador. Este es un punto que a menudo se presta a confusión: los modelos econométricos uniecuacionales no constituyen una tercera categoría de modelos, a añadir a los modelos univariantes y a los modelos econométricos multiecuacionales; al contrario, son una parte de estos últimos, y su estudio no se puede realizar sin haber explicitado claramente cuáles son las restricciones que se han de cumplir para que sus resultados sean correctos. Por ello, *el modelo econométrico uniecuacional no se entiende plenamente si no se discute el contexto general en el que aparece.*

Las siete primeras secciones de este capítulo se dedican a presentar el marco econométrico multiecuacional y a determinar los fines y los supuestos para los cuales es posible aislar una ecuación del sistema n-dimensional. A lo largo de todo el capítulo se supondrá que la hipótesis de que los modelos a considerar son lineales es adecuada.

Cuando el investigador pretende obtener un conocimiento razonablemente completo de un fenómeno económico, incluyendo el tipo de relación en la que se genera, su proyección futura (predicción) y su evolución esperada en distintos escenarios, el marco de análisis adecuado son los modelos econométricos. En tal caso podrá trabajar con un modelo uniecuacional, obviando el planteamiento multiecuacional, si las variables incluidas como explicativas en el conjunto informativo son exógenas. Este término admite distintas acepciones, que se explican en la sección 3.3, pero intuitivamente una variable exógena es aquélla que afecta a la determinación del fenómeno de interés, pero que, en el análisis concreto que se está realizando —determinación de los parámetros de la relación, predicción, simulación, etc.— no viene afectada por aquél. En tal caso, se puede utilizar el modelo uniecuacional.

El lector cuya principal motivación sea el tratamiento de los fenómenos económicos en el marco de modelos econométricos uniecuacionales con sólo variables fuertemente exógenas, puede ir directamente a la sección 3.7.2 y siguientes y dejar para una segunda lectura de este capítulo las secciones iniciales.

Cuando el objetivo del modelo econométrico se centra básicamente en realizar inferencia estadística sobre determinados parámetros, el contexto uniecuacional puede utilizarse aunque las variables no sean exógenas. En tal caso debe emplearse un procedimiento de estimación denominado de variables instrumentales, que requiere algún conocimiento adicional al meramente contenido en la relación uniecuacional. Tal conocimiento es muchísimo menor, en general, que el requerido para la formulación del modelo multiecuacional, por lo que en tal caso se dice que se está haciendo un análisis con información limitada. Con él, la estimación de los parámetros que centran el interés del investigador será consistente, pero menos eficiente que la estimación con información completa que se deriva del modelo multiecuacional. El tratamiento de los modelos econométricos bajo información limitada se hace en la sección 3.7.1.

Los modelos cuantitativos se formulan en función de lo que se pretende obtener con ellos y del grado de precisión o eficiencia que se quiera lograr. Respecto a esto último, mayor eficiencia requiere siempre mayor contenido informativo, y mayor información supone mayor complejidad y mayores costes en la construcción de modelos. En el capítulo anterior se vio que los modelos univariantes, que utilizan un conjunto de información mínimo, no sólo son útiles para fines predictivos, sino que también son instrumentos que describen la naturaleza del largo plazo del fenómeno en cuestión —mediante constantes o polinomios temporales dependientes de las condiciones iniciales— y la forma de acercarse a él. Además, recogen un comportamiento sistemático del fenómeno, de modo que el modelo univariante sirve para determinar períodos de comportamiento anómalo, debidos posiblemente a causas específicas, que, con frecuencia, pueden modelizarse ampliando el modelo ARIMA con un análisis de intervención.

Con los modelos econométricos, si están bien construidos, se pueden lograr todos los objetivos del capítulo anterior con igual o mayor precisión que con modelos univariantes. Así, la formulación de modelos econométricos para fines predictivos exclusivamente sólo tiene sentido cuando la precisión lograda con los modelos univariantes no es suficiente —en general o en determinados momentos— para el problema de decisión en que la predicción está enmarcada.

Además del uso con fines predictivos, los modelos econométricos se pueden utilizar para estimar y contrastar hipótesis sobre los parámetros que definen una relación entre variables. Para ello son imprescindibles, pues los modelos univariantes, por definición, no sirven para tal fin. Los parámetros que definen la relación a largo plazo son de especial interés. Con los modelos econométricos el largo plazo de una variable se define en función de otras variables, y los parámetros de dicha definición representan la ganancia —elasticidad total, en formulaciones logarítmicas— de una variable sobre otra. Ciertamente, el largo plazo en cada una de las nuevas variables vendrá, a su vez, explicado en términos de otras variables o tendrá su correspondiente formulación univariante, con lo que, procediendo de forma sucesiva, se tiene que a partir de la explicación en términos econométricos del largo plazo de una variable, se concluye con una explicación en términos de polinomios temporales con coeficientes estocásticos. En este contexto una constante es un polinomio temporal de orden cero. Esta tendrá que corresponderse con la explicación univariante, si ambos modelos —univariante y econométrico— sobre la variable de interés se han construido correctamente. Esta consistencia entre modelos univariantes y econométricos es la que justifica el que también se utilicen los primeros en el análisis económico.

El llegar a estimar determinados parámetros que definen la relación entre variables constituye, con frecuencia, la principal justificación de los modelos econométricos. El conocimiento de estos parámetros proporciona una utilidad que va mucho más allá de la predicción, pues resume una característica del mundo económico, cuyo conocimiento es importante para luego operar sobre él, por ejemplo, con fines de planificación e incluso de control. Así, estimar cuál es la elasticidad de unas determinadas exportaciones, como por ejemplo la de servicios turísticos españoles respecto a la renta y a los precios relativos de los países clientes y países competidores, sirve para determinar mejor el tipo de acciones más convenientes para lograr un desarrollo sostenido del sector exportador en cuestión. En el capítulo diez se discute con detalle las implicaciones que se derivan del modelo econométrico estimado para los ingresos turísticos en España. El conocimiento de los parámetros de un modelo econométrico sirve pues para utilizar éste con fines de análisis estructural, aspecto que se desarrolla con más detalle en la sección 3.13, en donde se comenta también el uso de modelos econométricos para fines de simulación y control.

La medición de una variable económica ha de hacerse referida a una unidad de tiempo, mes, trimestre, año, etc., lo que supone una medición agregada de las realizaciones de dicha variable en todos los momentos incluidos en la unidad de tiempo considerada. Debido a esta forma de medir, una variable económica puede ser causa y efecto en distintas ecuaciones de un modelo econométrico. Esto da lugar al modelo econométrico simultáneo (SEM), que sólo se puede especificar haciendo uso de resultados a priori de la Teoría Económica, resultados que direccionen la causalidad simultánea y que restrinjan la aparición de ciertas variables en las ecuaciones que determinan a otras. Estos modelos no se pueden construir a partir de la mera información contenida en la muestra temporal sobre un vector de variables; cuando sólo se dispone de información muestral se llega a los modelos de series temporales, modelos VARMA, que se estudian en las secciones primera y segunda.

El modelo SEM se puede resolver, de forma que el vector de variables endógenas en un momento t se exprese en función de su pasado y del presente y pasado de las variables exógenas. Al modelo resultante se le denomina de forma reducida. Este último no es más que un modelo VARMA triangular por bloques, en el que las variables exógenas entran en la determinación del bloque de ecuaciones de las endógenas, pero éstas no entran en la determinación del bloque de ecuaciones de las variables exógenas. Estos temas se desarrollan en las secciones 3.3 y 3.4 y sirven para clarificar la relación entre los modelos multiecuacionales que aparecen en la literatura econométrica y los que se desarrollan en la literatura estadística sobre series temporales. Una estrategia de utilización de ambos tipos de modelos se presenta en la la sección 3.5.

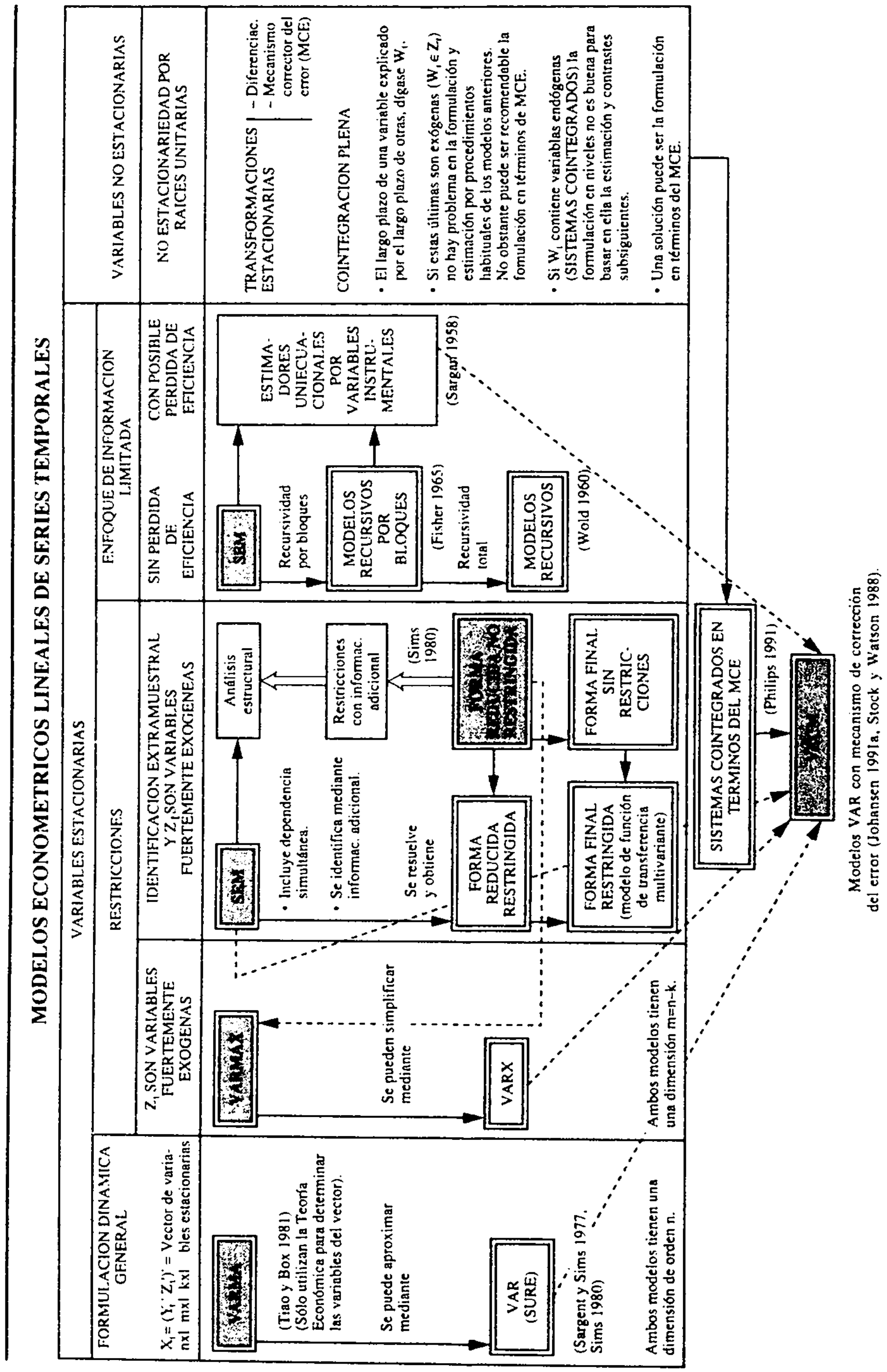
De lo anterior se desprende que el modelo econométrico se puede formular como un modelo SEM o como un modelo de forma reducida. La formulación en unos u otros términos dependerá del objetivo de la investigación. Los primeros permiten un análisis causal que no es posible de forma directa en los segundos. No obstante, tal y como se explica en la sección 3.4, una vez estimada la forma reducida sin información extramuestral se pueden incorporar restricciones procedentes de la Teoría Económica y hacer el tipo de análisis causal que permite el modelo SEM. La conclusión de todo esto es que la variedad de modelos econométricos que se estudian en este capítulo corresponde a distintos objetivos del analista, quien elegirá la formulación adecuada a sus fines. La relación de todos los modelos tratados en este capítulo se recoge en el cuadro 3.0.

La simultaneidad choca con el principio lógico de que una variable no puede ser causa y efecto al mismo tiempo. En la sección 3.6 se estudia la formulación de modelos estructurales recursivos. Cuando eso no es posible, el planteamiento uniecuacional con fines de estimación requiere el empleo de variables instrumentales mencionado anteriormente.

Para fines de predicción sólo se puede operar con modelos uniecuacionales si las variables explicativas son fuertemente exógenas y ese supuesto es el que se realiza en las secciones 3.8 a 3.11, dedicadas a los modelos econométricos uniecuacionales. Las características de estos modelos y el análisis de multiplicadores que de tales modelos se deriva se estudian en las secciones 3.8 y 3.9. Los modelos uniecuacionales dinámicos se suelen formular de forma distinta según sea la metodología cuantitativa que siga el analista. La sección 3.10 se dedica a analizar distintas formulaciones dinámicas del modelo uniecuacional, a relacionarlas entre sí y a discutir las posibles ventajas e inconvenientes de cada una de ellas.

Con el análisis de varias variables económicas puede ocurrir que la no estacionariedad de una venga explicada por las otras. Entonces se dice que están cointegradas y a ese tema, manteniendo la hipótesis de exogeneidad, se dedica la sección 3.11, mientras que en la 3.12 se resumen un conjunto de conclusiones tomadas de Phillips (1991) para el caso de que no se dé la exogeneidad requerida en la sección

CUADRO 3.0. *Modelos econométricos lineales de series temporales*



anterior. El capítulo acaba analizando en la sección 3.13 los distintos usos de un modelo econométrico.

El lector interesado en ampliar los temas que se tratan en este capítulo puede acudir a la amplia literatura existente, entre la que podemos destacar, y sin ánimo de ser exhaustivos, a Amemiya (1985), Berndt (1991), Eatwell et al. (eds., 1990), Engle y Granger (eds, 1991), Goldberger (1991), Granger (ed., 1990), Griliches e Intriligator (eds., 1983, 1984), Hackl y Westlund (eds., 1991), Harvey (1981a), Judge et al (1985), Lutkepohl (1991), Malinvaud (1980), Novales (1988), Peña (1987), Sargan (1988) y Spanos (1986).

3.1. Universos multivariantes, modelos econométricos y modelos de series temporales

En el capítulo 1 se argumentó la necesidad de basar el análisis de la coyuntura en modelos estadístico-econométricos, y en el capítulo 2 se ha visto que un modelo referido a una variable X_{1t} permite una descomposición de la forma[1]

$$X_{1t} = \text{Parte sistemática}_t + \text{Innovación}_t$$

En la construcción de un modelo de comportamiento para X_1 se plantean básicamente dos opciones en función del conjunto de información que se utilice:

a) Emplear únicamente la información referente a X_1; en este caso la parte sistemática es función exclusiva de la propia historia del fenómeno en estudio. Cuando se elige esta opción el resultado es un modelo univariante como los que se han tratado en el capítulo anterior.

b) Emplear, además de la información anterior, la información referida a cualquier otra variable X_j, $j=2, 3, \ldots, n$, con la que X_1 pueda estar relacionada. En tal caso la parte sistemática tendrá una formulación que, en general, vendrá dada como una función del pasado de la variable X_1 y del presente y pasado de cualquier otra variable X_j, $j=2, 3, \ldots, n$; con ello se tiene un modelo multivariante, que, en el caso concreto en que se relacionan variables económicas, se denomina *modelo econométrico*.

Cuando se amplía el conjunto de información univariante con información sobre otras variables, se abandona un universo escalar,

[1] En el capítulo anterior se ha estado denotando a la variable por Y, mientras que en estas primeras secciones del capítulo 3 se usará $X_1, X_2, \ldots, X_n$ para las distintas variables a considerar. A lo largo del capítulo se irá viendo el porqué del cambio.

formado exclusivamente por el fenómeno X_1, para situarse en un universo vectorial, en el que los componentes del vector son el fenómeno X_1 y los restantes fenómenos $X_2, \ldots, X_n$ considerados. Con ello, se tendrá un modelo compuesto por n ecuaciones que explicarán el comportamiento de todas las variables que se consideran en el análisis.

Esto es consecuencia de la clase de relación existente entre las variables económicas, que suelen ser interdependientes. Así, por ejemplo, si se consideran dos variables X_j y X_h puede ocurrir que se cumplan algunas o todas de las siguientes condiciones: a) X_{jt} depende de $X_{h,t-k}$, $k>0$; b) X_{ht} depende de $X_{j,t-k}$, $k>0$; y c) ambas variables X_j y X_k están correlacionadas contemporáneamente. Cuando se cumplen a) y b) a la vez hay realimentación o causalidad bidireccional entre ambas variables, y si además se cumple c) existe también causalidad instantánea (véase Granger 1969). En tales casos, un modelo para X_j que ignorase la bidireccionalidad causal y se limitase a formular una ecuación que recogiese el efecto de X_h sobre X_j, podría conducir a resultados ineficientes e incluso equívocos.

En lo que sigue se va a suponer que todas las variables que se utilizan en el análisis son *estacionarias*, y que están generadas por procesos estocásticos que además de estacionarios son invertibles y no anticipantes[2]. Al igual que en el capítulo anterior, el supuesto de estacionariedad se puede obviar utilizando variables diferenciadas, es decir, introduciendo raíces unitarias en el modelo econométrico. Pero ahora, al considerar n variables, la diferenciación conjunta de todas ellas puede resultar algo excesivo, y derivar, por tanto, en un modelo con un componente de medias móviles no invertible.

Esta diferencia respecto al tratamiento univariante radica en que en economía la trayectoria de largo plazo de una variable no estacionaria puede venir plenamente explicada por otras variables económicas. En tal caso, si un modelo econométrico no relaciona dichas variables en niveles, sino que se plantea directamente sobre las variables diferenciadas, será un modelo mal formulado, pues no recoge una información esencial como es la plena dependencia a largo plazo de la variable analizada respecto a otras variables económicas. Como consecuencia, la estimación del modelo global para las n variables diferenciadas será inconsistente.

Por eso, en la sección once de este capítulo se desarrollará el concepto de *variables cointegradas*, es decir, variables para las que, siendo todas ellas no estacionarias (por ejemplo integradas de orden d), existe al menos una combinación lineal de las mismas que presenta un orden de integración menor que el de las variables que la definen.

[2] Sobre la definición de estos conceptos véase la sección segunda del capítulo 2.

Cuando la combinación lineal es estacionaria, es decir integrada de orden cero, entonces se dice que las variables están *plenamente cointegradas*. Obsérvese que si las variables originales están cointegradas pero no plenamente cointegradas, existe un valor $d^* < d$ tal que aplicando d^* diferenciaciones a todas las variables, las variables resultantes están plenamente cointegradas.

A las combinaciones lineales del párrafo anterior se les denomina *mecanismos de corrección del error*, y con ellos es fácil ver —tal y como se hace en la sección once— que la relación de las variables en las ecuaciones de un modelo econométrico admite una reparametrización en la que, empleando diferenciaciones y mecanismos de corrección del error, todos sus elementos son estacionarios. Así pues, en la formulación de modelos econométricos la no estacionariedad en los datos se aborda mediante la diferenciación y los mecanismos de corrección del error.

En consecuencia, el supuesto de estacionariedad utilizado en estas secciones supone que, en el caso en que las variables originales no sean estacionarias, el modelo está adecuadamente formulado utilizando diferenciaciones y mecanismos de corrección del error. Ambos esquemas son distintas manifestaciones del enfoque de raíces unitarias introducido en el capítulo anterior.

En efecto, supóngase dos variables, X_1 y X_2, que siendo ambas integradas de primer orden su diferencia —una combinación lineal con parámetros uno y menos uno— es estacionaria. Este ejemplo podría darse siendo X_1 y X_2 dos tipos de interés. En tal caso, un modelo econométrico simple podría ser

$$\Delta X_{1t} = \alpha \Delta X_{2t} + \beta(X_1 - X_2)_{t-1} + a_{1t}$$

$$\Delta X_{2t} = a_{2t}$$

$$E(a_{1t}a_{2t'}) = E(a_{1t}a_{1t'}) = E(a_{2t}a_{2t'}) = 0, \qquad t \neq t'$$

Por lo dicho anteriormente, $X_1 - X_2$ es el mecanismo de corrección del error, que formulado directamente sobre las variables originales es estacionario. En el modelo precedente se emplean transformaciones estacionarias de las variables originales. Una de dichas transformaciones es el mecanismo de corrección del error, que explica la no estacionariedad de X_1 en función de la no estacionariedad de X_2, que a su vez descansa en la raíz unitaria que está presente en la segunda ecuación.

Obsérvese que si en este ejemplo el modelo se formulase utilizando solamente diferencias, se estaría omitiendo el término $\beta(X_1 - X_2)_{t-1}$y, por tanto, su estimación sería inconsistente.

En un contexto econométrico es conveniente —siguiendo a Hendry y Richard (1982)— considerar que la información disponible se compone de dos conjuntos: uno, referido a las observaciones de todas las variables contempladas (*información muestral*), y otro referido al conocimiento teórico que, fundamentalmente, consiste en el conocimiento proporcionado por la Teoría Económica (*información extramuestral*). Con ello se pueden distinguir dos grandes categorías de modelos econométricos dinámicos, que se diferencian por la información necesaria para su construcción:

1) *Modelos Multivariantes de Series Temporales o VARMA* (Vector AutoRegresive Moving Average): son una generalización de los modelos ARMA univariantes estudiados en el capítulo anterior. Se caracterizan porque sólo manejan la información contenida en la muestra y no imponen restricciones basadas en la Teoría Económica. Por ello son especialmente recomendables cuando la información de naturaleza teórica es escasa o carece de una aceptación suficientemente generalizada. No obstante, conviene señalar que estos modelos no ignoran completamente la teoría, ya que la elección de las variables incluidas en el modelo, que puede ser una labor compleja, se debe realizar siguiendo los dictámenes de la Teoría Económica.

Por ello, y en relación a los dos conjuntos de información mencionados, es mejor referirse a los modelos VARMA como modelos que se construyen realizando un uso mínimo de la Teoría Económica. Un caso particular de estos modelos lo constituyen los modelos autoregresivos vectoriales (VAR, Vector AutoRegresive), es decir, modelos VARMA sin parte de medias móviles, propuestos por Sims (véase Sargent y Sims (1977), Sims (1980, 1982)).

En el pasado ha sido una práctica habitual formular tanto los modelos VARMA como los VAR sobre las series diferenciadas cuando se utilizaban variables no estacionarias. La literatura sobre el mecanismo de corrección del error ha puesto de manifiesto que la formulación de tales modelos directamente sobre variables diferenciadas, cuando las series originales están cointegradas, es incorrecta e introduce un componente de medias móviles no invertible. No obstante, no existe especial impedimento para incorporar mecanismos de corrección del error en tales modelos: a los modelos que realizan tal incorporación se les denomina modelos Vectoriales con Mecanismo de Corrección del Error (VECM, Vector Error Correction Mechanism)[3].

Sin embargo, la posible existencia de mecanismos de corrección

[3] Sobre la aplicación de tales mecanismos en un modelo VAR véase King et al. (1991).

del error, que no son más que restricciones en la relación a largo plazo de las variables, debe estar en la práctica guiada por la Teoría Económica, por lo que en el caso de variables cointegradas la formulación de un modelo VECM no puede, en general, hacerse de forma completamente independientemente de la teoría. Por esa razón, se puede generalizar la afirmación anterior diciendo que estos modelos sólo utilizan la información teórica para seleccionar las variables e incorporar las relaciones de cointegración.

2) *Modelos Estructurales o SEM* (Simultaneous Equation Models): utilizan intensamente la información teórica disponible pretendiendo una formulación causal en la relación entre variables y, en consecuencia, su uso es tanto más recomendable cuanto más fiable sea dicha información. El planteamiento econométrico multiecuacional tradicionalmente se ha basado en estos modelos, razón por la cual al menos una breve introducción a los mismos figura en casi todos los manuales de econometría; para una crítica de este enfoque véase Sims (1980), trabajo del que se hablará extensamente más adelante.

El tipo de modelo a utilizar depende, como se ha dicho en la introducción, del objetivo del usuario. Así, cuando el interés es exclusivamente la predicción, los modelos VARMA son adecuados, pero cuando el usuario quiere conocer o contrastar el valor de un determinado parámetro que la Teoría Económica señala como «parámetro de interés», como por ejemplo la elasticidad de las importaciones respecto a la producción, entonces los modelos VARMA no son en general adecuados y se requiere un modelo SEM.

En el epígrafe segundo se revisarán las principales características de los modelos VARMA, dejando los modelos SEM para la sección cuarta.

3.2. Modelos VARMA y modelos VAR

Los modelos multivariantes de series temporales fueron introducidos por Quenouille (1957); en Tiao y Box (1981) se estudian desde la perspectiva de la metodología Box-Jenkins, generalizando resultados vistos para el modelo univariante.

Para ilustrar la exposición, considérese el siguiente ejemplo: el objetivo es modelizar el comportamiento de la variable X_1, que se sabe está relacionada con otra variable X_2; se supone para simplificar que la dinámica es de orden uno. En este caso se tiene un modelo VARMA (1,1), cuya expresión general es:

$$\begin{aligned} X_{1t} &= \phi_{11}^1 X_{1,t-1} + \phi_{12}^1 X_{2,t-1} + a_{1t} - \theta_{11}^1 a_{1,t-1} - \theta_{12}^1 a_{2,t-1}, \\ X_{2t} &= \phi_{21}^1 X_{1,t-1} + \phi_{22}^1 X_{2,t-1} + a_{2t} - \theta_{21}^1 a_{1,t-1} - \theta_{22}^1 a_{2,t-1} \end{aligned}$$

donde a_{it} representa la innovación de la variable X_i en el momento t. Ambas innovaciones tienen valor esperado igual a cero y se cumple que

$$E(a_{it}a_{it'})=\begin{cases}0 & \text{si} \quad t\neq t' \\ \sigma_i^2 & \text{si} \quad t=t'\end{cases} \qquad i=1,\, 2$$

y, además,

$$E(a_{1t}a_{2t'})=\begin{cases}0 & \text{si} \quad t\neq t' \\ \sigma_{12} & \text{si} \quad t=t'\end{cases},$$

lo que implica que puede existir correlación contemporánea entre ambas innovaciones, pero no correlación en distintos momentos del tiempo. Para simplificar se suele añadir el supuesto de que las innovaciones siguen una distribución normal.

Operando en el modelo original, se tiene que

$$\begin{aligned} X_{1t}-\phi_{11}^1 X_{1,t-1}-\phi_{12}^1 X_{2,t-1}&=a_{1t}-\theta_{11}^1 a_{1,t-1}-\theta_{12}^1 a_{2,t-1} \\ X_{2t}-\phi_{21}^1 X_{1,t-1}-\phi_{22}^1 X_{2,t-1}&=-\theta_{21}^1 a_{1,t-1}+a_{2t}-\theta_{22}^1 a_{2,t-1}\end{aligned}, \qquad (3.2.1)$$

Esto se puede expresar de forma simplificada definiendo los siguientes polinomios en el operador de retardos L:

$$\begin{aligned} \phi_{11}(L)&=1-\phi_{11}^1 L & \theta_{11}(L)&=1-\theta_{11}^1 L \\ \phi_{12}(L)&=\ \ -\phi_{12}^1 L & \theta_{12}(L)&=\ \ -\theta_{12}^1 L \\ \phi_{21}(L)&=\ \ -\phi_{21}^1 L & \theta_{21}(L)&=\ \ -\theta_{21}^1 L \\ \phi_{22}(L)&=1-\phi_{22}^1 L & \theta_{22}(L)&=1-\theta_{22}^1 L \end{aligned}$$

con lo que (3.2.1) es equivalente a

$$\begin{aligned} \phi_{11}(L)X_{1t}+\phi_{12}(L)X_{2t}&=\theta_{11}(L)a_{1t}+\theta_{12}(L)a_{2t} \\ \phi_{21}(L)X_{1t}+\phi_{22}(L)X_{2t}&=\theta_{21}(L)a_{1t}+\theta_{22}(L)a_{2t}\end{aligned}.$$

En forma matricial

$$\begin{pmatrix}\boldsymbol{\phi_{11}(L)} & \boldsymbol{\phi_{12}(L)} \\ \boldsymbol{\phi_{21}(L)} & \boldsymbol{\phi_{22}(L)}\end{pmatrix}\begin{pmatrix}\mathbf{X_{1t}} \\ \mathbf{X_{2t}}\end{pmatrix}=\begin{pmatrix}\boldsymbol{\theta_{11}(L)} & \boldsymbol{\theta_{12}(L)} \\ \boldsymbol{\theta_{21}(L)} & \boldsymbol{\theta_{22}(L)}\end{pmatrix}\begin{pmatrix}\mathbf{a_{1t}} \\ \mathbf{a_{2t}}\end{pmatrix}. \qquad (3.2.2)$$

Si ahora se definen los vectores

$$\mathbf{X}_t=\begin{pmatrix}X_{1t} \\ X_{2t}\end{pmatrix} \qquad \mathbf{a}_t=\begin{pmatrix}a_{1t} \\ a_{2t}\end{pmatrix}$$

y las matrices polinomiales

$$\Phi_1(L)=\begin{pmatrix}\phi_{11}(L) & \phi_{12}(L) \\ \phi_{21}(L) & \phi_{22}(L)\end{pmatrix} \qquad \Theta_1(L)=\begin{pmatrix}\theta_{11}(L) & \theta_{12}(L) \\ \theta_{21}(L) & \theta_{22}(L)\end{pmatrix}$$

se llega a la expresión sintética del modelo VARMA (1, 1):

$$\begin{aligned} &\boldsymbol{\Phi}_1(L)\mathbf{X}_t=\boldsymbol{\Theta}_1(L)\mathbf{a}_t \\ &E(\mathbf{a}_t\mathbf{a}_t')=\begin{pmatrix}\sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2\end{pmatrix}=\boldsymbol{\Sigma},\end{aligned} \qquad (3.2.3)$$

que es similar a la que se estudió en el capítulo anterior, con la única diferencia de que ahora se modeliza un vector de variables en vez de una variable escalar.

En exacto paralelismo con el caso univariante las condiciones de estacionariedad vienen determinadas por la matriz polinomial autorregresiva, $\Phi_1(L)$. El proceso es estacionario si todas las raíces del polinomio correspondiente a su determinante, $|\Phi_1(z)|$, donde z es el argumento del polinomio, son en valor absoluto mayores que la unidad. Una forma alternativa de expresar la condición de estacionariedad es exigiendo que las raices del polinomio que se obtiene al calcular el determinante de $\Phi_1(z^{-1})$ estén dentro del círculo unitario. Análogamente, la invertibilidad requiere que las raíces del polinomio $|\Theta_1(z^{-1})|$ estéen dentro del círculo unitario.

En la formulación (3.2.2) se compacta la expresión (3.2.1) resumiendo la dinámica y manteniendo explícitas cada una de las variables que entran en el modelo. Una simplificación alternativa se obtiene definiendo las matrices

$$\Phi^{(0)}=\Theta^{(0)}=\begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}=I$$

$$\Phi^{(1)}=\begin{pmatrix}\phi_{11}^1 & \phi_{12}^1 \\ \phi_{21}^1 & \phi_{22}^1\end{pmatrix} \qquad \Theta^{(1)}=\begin{pmatrix}\theta_{11}^1 & \theta_{12}^1 \\ \theta_{21}^1 & \theta_{22}^1\end{pmatrix},$$

de donde (3.2.1) se puede expresar como

$$\mathbf{X}_t - \Phi^{(1)}\mathbf{X}_{t-1} = \mathbf{a}_t - \Theta^{(1)}\mathbf{a}_{t-1}$$

o bien

$$(I - \Phi^{(1)}L)\mathbf{X}_t = (I - \Theta^{(1)}L)\mathbf{a}_t \qquad (3.2.4)$$

En esta expresión se explicita la dinámica del sistema, y la simplificación consiste en que no se especifican cada una de las variables que se consideran, como en (3.2.2), sino que se escriben directamente en forma vectorial.

Naturalmente de (3.2.4) es trivial pasar a la forma (3.2.3), ya que se cumple que

$$\Phi_1(L) = I - \Phi^{(1)}L \qquad \Theta_1(L) = I - \Theta^{(1)}L.$$

Es inmediato generalizar el ejemplo para cualquier número de variables y para estructuras dinámicas de orden finito superior a uno, redefiniendo adecuadamente las matrices y los vectores. Así, en un problema general en donde haya n variables, el orden máximo de desfases de la parte autorregresiva sea p y el de la parte de medias móviles q, se tiene la siguiente expresión general del modelo VARMA (p, q) para n variables:

$$\begin{gathered}\Phi_p(L)\mathbf{X}_t = \Theta_q(L)\mathbf{a}_t \\ E(\mathbf{a}_t\mathbf{a}_t') = \Sigma,\end{gathered} \qquad (3.2.5)$$

que se puede desarrollar como

$$\begin{pmatrix} \phi_{11}(L) & \phi_{12}(L) & \cdots & \phi_{1n}(L) \\ \phi_{21}(L) & \phi_{22}(L) & \cdots & \phi_{2n}(L) \\ \cdots & \cdots & \cdots & \cdots \\ \phi_{n1}(L) & \phi_{n2}(L) & \cdots & \phi_{nn}(L) \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \\ \cdots \\ X_{nt} \end{pmatrix} = \begin{pmatrix} \theta_{11}(L) & \theta_{12}(L) & \cdots & \theta_{1n}(L) \\ \theta_{21}(L) & \theta_{22}(L) & \cdots & \theta_{2n}(L) \\ \cdots & \cdots & \cdots & \cdots \\ \theta_{n1}(L) & \theta_{n2}(L) & \cdots & \theta_{nn}(L) \end{pmatrix} \begin{pmatrix} a_{1t} \\ a_{2t} \\ \cdots \\ a_{nt} \end{pmatrix}$$

en donde los polinomios de las diagonales principales tienen coeficiente unidad en la potencia cero de L, y los demás polinomios coeficientes cero en dicha potencia. Alternativamente el modelo VARMA (p, q) se puede desarrollar como

$$(I - \Phi^{(1)}L - \Phi^{(2)}L^2 - \cdots - \Phi^{(p)}L^p)\mathbf{X}_t = \\ = (I - \Theta^{(1)}L - \Theta^{(2)}L^2 - \cdots - \Theta^{(q)}L^q)\mathbf{a}_t, \qquad (3.2.6)$$

si se quiere poner el acento en la dinámica del modelo en vez de en las variables que lo componen. La matriz de varianzas y covarianzas de las innovaciones es:

$$E(\mathbf{a}_t\mathbf{a}_t') = \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix}. \qquad (3.2.7)$$

En la formulación del modelo VARMA se está sugiriendo que los órdenes p y q son conocidos, y que no existen factores comunes entre la parte autorregresiva y la parte de medias móviles. Además, todas las formulaciones del modelo VARMA se han realizado normalizando el modelo en el retardo cero, es decir, con la restricción

$$\Phi^{(0)} = \Theta^{(0)} = I \qquad (3.2.8)$$

que, junto con la no existencia de factores comunes, garantiza en general[4] la obtención de estimadores únicos para los parámetros del modelo, tal y como se verá más adelante al discutir el problema de la identificación.

En el contexto multivariante la restricción de que no existan factores comunes no hace referencia solamente —como era el caso univariante del capítulo anterior— a factores comunes que una vez eliminados reducen los órdenes $p+q$, como ocurre, por ejemplo, si

$$\begin{gathered}\Phi_p(L) = (I - AL)\Phi_{p-1}(L) \\ \Theta_q(L) = (I - AL)\Theta_{q-1}(L),\end{gathered}$$

sino también a factores unimodulares, $D(L)$.

Una excelente ilustración y discusión de este tema se encuentra en Lütkepohl (1991), sección 7.1. Para los fines de este capítulo baste

[4] Sobre excepciones en las que no se pueden obtener estimadores consistentes en un modelo identificado véase Gabrielson (1978).

decir que los *factores unimodulares* son factores de orden finito que a su vez tienen una inversa de orden finito. Dado que

$$D(L)^{-1} = adj[D(L)]/|D(L)|,$$

estos factores se caracterizan por tener un determinante, $|D(L)|$, que es una constante no nula, es decir, un polinomio en L de orden cero con coeficiente distinto de cero. Los factores unimodulares pueden, aunque no necesariamente, causar problemas, ya que multiplicando $\Phi_p(L)$ y $\Theta_q(L)$ por un factor de este tipo puede ocurrir, según las especificaciones de dichas matrices, que éstas se modifiquen sin alterar sus respectivos órdenes.

Si en el modelo VARMA (p, q) el orden q es cero se tiene un modelo VAR:

$$\Phi_p(L)\mathbf{X}_t = \mathbf{a}_t, \tag{3.2.9}$$

en donde los residuos son ruido blanco. Obsérvese que todo modelo VAR con residuos estacionarios generados por un proceso autorregresivo

$$\begin{aligned} \Phi_p(L)\mathbf{X}_t &= \mathbf{v}_t \\ R_r(L)\,\mathbf{v}_t &= \mathbf{a}_t \end{aligned}$$

se puede reformular como un proceso VAR con residuos ruido blanco, sustituyendo en la primera de las dos ecuaciones anteriores $\mathbf{v}_t$ por $[R_r(L)]^{-1}\mathbf{a}_t$, con lo que

$$R_r(L)\Phi_p(L)\mathbf{X}_t = \mathbf{a}_t,$$

o utilizando $\Phi^*_{p+r}(L) = R_r(L)\Phi_p(L)$, como

$$\Phi^*_{p+r}(L)\mathbf{X}_t = \mathbf{a}_t.$$

Los modelos VAR del tipo (3.2.9) presentan la característica de que se pueden estimar ecuación por ecuación sin pérdida de eficiencia, tal y como se comentará en la sección sexta; esta propiedad no la cumplen los modelos VARMA, en los que el componente de medias móviles obliga a una estimación conjunta. Esta es la razón principal que ha provocado que los modelos VAR sean más utilizados en economía que los modelos VARMA. Un tratamiento a fondo de ambos tipos de modelos se encuentra en Lütkepohl (1991).

El punto de partida del modelo VARMA (p, q) es la función de densidad conjunta de las n variables consideradas en el análisis desde

$t=1$ a $t=T$, y la formulación VARMA es el resultado de un proceso de *condicionar respecto al pasado*. Para ver cómo opera este proceso conviene introducir, siguiendo a Engle et al (1983), la siguiente notación. Sea X_t^1 la matriz de observaciones $t \times n$, correspondiente a una muestra de t observaciones sobre n variables:

$$X_t^1 = (\mathbf{X}_1 \dots \mathbf{X}_t)'$$

donde, siguiendo la notación habitual, $\mathbf{X}_1 = (X_{11} X_{21} \dots X_{n1})'$ representa el vector $n \times 1$ de observaciones de todas las variables para el momento 1 de tiempo, y así sucesivamente. Sea así mismo X_0 la matriz que representa las condiciones iniciales. Finalmente, la información hasta el momento t vendrá dada por la matriz X_t^+, que se define como

$$X_t^+ = \begin{pmatrix} X_0 \\ X_t^1 \end{pmatrix}.$$

Toda función de densidad multivariante se puede escribir como el producto de una función de densidad condicional y la función de densidad marginal de las variables sobre las que se condiciona. Por tanto, si $D(X_T^1/X_0, \delta)$ es la función de densidad de X_T^1 condicionada a una situación inicial X_0, que se supone dada, y al conjunto de parámetros que caracteriza el sistema, se tiene que[5]

$$D(X_T^1/X_0, \delta) = \prod_{t=1}^{T} D(\mathbf{X}_t/X_{t-1}^+, \lambda_t), \tag{3.2.10}$$

siendo λ_t el vector columna de los parámetros asociados a la distribución de $\mathbf{X}_t$ condicionada a X_{t-1}^+. El vector λ^*, que representa el conjunto de parámetros que aparece a la derecha de la igualdad (3.2.10), $\lambda^* = (\lambda_1', \dots, \lambda_T')'$, es función de δ de acuerdo con la representación que se deriva del proceso de condicionamiento secuencial operado.

Si se supone normalidad, tal y como se hace a lo largo de este capítulo,

$$\mathbf{X}_t/X_{t-1}^+ \sim N(\mu_t, \Omega_t), \tag{3.2.11}$$

[5] La expresión (3.2.10) se obtiene observando que

$$D(X_T^1/X_0, \delta) = D(\mathbf{X}_T/X_{T-1}^1, X_0, \lambda_T)\cdot D(X_{T-1}^1/X_0, \delta^1)$$

y repitiendo el proceso de factorización sobre $D(X_{T-1}^1/X_0, \delta^1)$ y las $D(X_{T-j}^1/X_0, \delta^j)$ sucesivas.

entonces μ_t, la esperanza matemática condicional de $\mathbf{X}_t$, es una función lineal del pasado de las n variables consideradas en el análisis[6]. En este caso, el vector λ_t recoge todos los parámetros no redundantes de μ_t y Ω_t.

En el caso particular en que el conjunto de parámetros t es constante en el tiempo, $\lambda_t=\lambda$ para todo t, entonces (3.2.10) se convierte en

$$D(X_T^1/X_0, \delta)=\prod_{t=1}^{T} D(\mathbf{X}_t/X_{t-1}^1, \lambda), \qquad (3.2.12)$$

En las expresiones (3.2.10) o (3.2.12) se está condicionando a una situación inicial a la que no se ha impuesto ningún tipo de restricciones, representada por X_0. Por lo tanto, y al menos teóricamente, la dependencia dinámica no está acotada, en el sentido de que los valores de la variable correspondientes a un pasado muy lejano ($t-k$, con k tendiendo a infinito) pueden estar influyendo en los valores correspondientes a t.

No obstante, en las aplicaciones a variables económicas suele ocurrir que la dependencia dinámica está acotada, de manera que existe un número finito p tal que la dependencia temporal no se remonta más allá del retardo p-ésimo. En ese caso (3.2.10) es equivalente a

$$D(X_T^1/X_0, \delta)=\prod_{t=1}^{T} D(\mathbf{X}_t/X_{t-1}^{t-p}, \lambda_t), \qquad (3.2.13)$$

donde X_{t-1}^{t-p} se define como la matriz de orden $p\times n$ dada por

$$X_{t-1}^{t-p}=(\mathbf{X}_{t-p} \cdots \mathbf{X}_{t-1})',$$

cuando $t-p>0$. Para las p observaciones iniciales de la muestra ($t=1, 2, \ldots, p$), la matriz X_{t-1}^{t-p} se forma seleccionando del conjunto de condiciones iniciales X_0 los valores correspondientes.

También en este caso de dependencia dinámica acotada se puede tener que $\lambda_t=\lambda$ para todo t. Se obtiene así la expresión

$$D(X_T^1/X_0, \delta)=\prod_{t=1}^{T} D(\mathbf{X}_t/X_{t-1}^{t-p}, \lambda), \qquad (3.2.14)$$

[6] La normalidad es una condición suficiente, pero no necesaria, para la linealidad de la esperanza matemática condicionada.

de donde se deriva el modelo VAR (p). Si la dependencia temporal en (3.2.12) está acotada de forma más general en el sentido de que admite una representación racional convergente, entonces de (3.2.14) se obtiene el modelo VARMA (p, q).

Este desarrollo que se acaba de realizar también es válido para los modelos univariantes estudiados en el capítulo anterior, ya que éstos se obtienen como un caso particular tomando $n=1$.

Para la discusión que sigue es importante distinguir —utilizando la terminología econométrica habitual a este respecto— entre estructura y modelo, y para ello se seguirá a Hsiao (1983).

Se denomina *estructura* a una especificación plena de la función de densidad de los datos, en el caso que nos ocupa una especificación plena de $D(\mathbf{X}_t/X_{t-1}^{t-p}, \lambda)$, de tal forma que todos los elementos de λ toman valores concretos. El término *modelo* se define como el conjunto, M, de todas las estructuras en principio posibles. Por lo tanto el concepto de estructura se refiere a un punto del espacio paramétrico Λ^h, y el de modelo a un conjunto de puntos $\Lambda^h\subset R^h$.

Con ello se tiene que dos estructuras, λ_a y λ_b, son observacionalmente equivalentes si

$$D(\mathbf{X}_t/X_{t-1}^{t-p}, \lambda_a)=D(\mathbf{X}_t/X_{t-1}^{t-p}, \lambda_b)$$

para todo X_T^1.

El problema de distinguir entre estructuras consiste pues en distinguir entre puntos en el espacio paramétrico Λ^h en el que se define el modelo; y el problema de identificación consiste en determinar una estructura dado el modelo, M, y las observaciones X_T^1. Así, los parámetros de una estructura λ_0 están globalmente identificados si no existe otra estructura λ_e en Λ^h que sea observacionalmente equivalente.

En determinadas ocasiones puede ocurrir que el conjunto de parámetros de λ_0 no esté identificado globalmente, pero que una parte del mismo λ_0^1 se pueda determinar de forma única. En tal caso se dice que un subconjunto de elementos λ_0^1 de λ_0 está identificado si todas las estructuras observacionalmente equivalentes a λ_0 tienen el mismo valor para λ_0^1.

El efecto que la falta de identificación tiene en la estimación de un modelo es inmediato. Si dos estructuras son observacionalmente equivalentes no es posible distinguirlas si la única información que se posee para ello es una muestra X_T^1, y por tanto a partir de una muestra concreta no se podrá discriminar cuál es la verdadera estructura generadora de los datos.

Esto se ve más claramente si se tiene presente que la función de verosimilitud correspondiente a (3.2.14) viene dada por

$$L(\delta;\, X_T^1,\, X_0) = \prod_{t=1}^{T} L(\lambda;\, \mathbf{X}_t,\, X_{t-1}^{t-p}),$$

y los estimadores de λ se obtienen maximizando dicha expresión.

La falta de identificación ocurre cuando existen dos estructuras distintas, λ_a y λ_b, tales que

$$\prod_{t=1}^{T} L(\lambda_a;\, \mathbf{X}_t,\, X_{t-1}^{t-p}) = \prod_{t=1}^{T} L(\lambda_b;\, \mathbf{X}_t,\, X_{t-1}^{t-p}) = \max_{\lambda} \prod_{t=1}^{T} L(\lambda;\, \mathbf{X}_t,\, X_{t-1}^{t-p})$$

En ese caso no se puede realizar ningún tipo de inferencia, ya que contando únicamente con la información que proporciona la muestra no es posible determinar cuál de las dos estructuras alternativas, λ_a o λ_b, ha generado los datos observados. De ahí que el problema de identificación sea previo al de estimación, ya que, como es fácil demostrar, un parámetro no identificado no se puede estimar de forma consistente.

Por esa razón en la formulación más usual de los modelos VARMA se opta por incorporar como conjunto de restricciones que garantizan la identificación las recogidas en la ecuación (3.2.8). En palabras, lo que esta expresión impone es no permitir que los valores contemporáneos de X_h o de a_h entren en la ecuación de comportamiento de X_j, $j \neq h$, y con ello se garantiza la identificación del modelo.

Es sencillo comprobar como el modelo VARMA no está identificado si no se impone ningún tipo de restricción. Para ello, considérese un caso simplificado en el que q es igual a cero, con lo que se obtiene el modelo VAR (p), que sin la restricción (3.2.8) se puede expresar como:

$$\Phi^{(0)}\mathbf{X}_t - \Phi^{(1)}\mathbf{X}_{t-1} - \cdots - \Phi^{(p)}\mathbf{X}_{t-p} = \mathbf{a}_t \tag{3.2.15}$$

Si en (3.2.15) se resuelve el sistema respecto a $\mathbf{X}_t$, premultiplicando por $[\Phi^{(0)}]^{-1}$ se obtiene

$$\mathbf{X}_t = \pi_1 \mathbf{X}_{t-1} + \cdots + \pi_p \mathbf{X}_{t-p} + \mathbf{v}_t \tag{3.2.16}$$

donde

$$\pi_j = [\Phi^{(0)}]^{-1}\Phi(j) \qquad \mathbf{v}_t = [\Phi^{(0)}]^{-1}\mathbf{a}_t,$$

y $\mathbf{v}_t$ cumple que

$$E(\mathbf{v}_t \mathbf{v}_t') = \Omega = [\Phi^{(0)}]^{-1}\Sigma\{[\Phi^{(0)}]^{-1}\}'. \tag{3.2.17}$$

La función de verosimilitud de la muestra se puede obtener a partir de (3.2.16) y (3.2.17).

Ahora bien, si se premultiplica (3.2.15) por una matriz no singular H se obtiene

$$\bar{\Phi}^{(0)}\mathbf{X}_t - \bar{\Phi}^{(1)}\mathbf{X}_{t-1} - \cdots - \bar{\Phi}^{(p)}\mathbf{X}_{t-p} = \bar{\mathbf{a}}_t. \tag{3.2.18}$$

donde $\bar{\Phi}^{(j)} = H\Phi^{(j)}$ y $\bar{\mathbf{a}}_t = H\mathbf{a}_t$.

Resolviendo (3.2.18) respecto a $\mathbf{X}_t$ se obtiene también el modelo (3.2.16), donde ahora

$$\bar{\pi}_j = [H\Phi^{(0)}]^{-1} H\Phi^{(j)} = [\Phi^{(0)}]^{-1} H^{-1} H\Phi^{(j)} = [\Phi^{(0)}]^{-1}\Phi^{(j)} = \pi_j$$

y de forma similar

$$\bar{\Omega} = [\Phi^{(0)}]^{-1} H^{-1} H\Sigma H'(H')^{-1}\{[\Phi^{(0)}]^{-1}\}' = \Omega.$$

Se ve de forma inmediata que una vez estimadas las matrices π_j y Ω no se puede determinar qué estructura, (3.2.15) o (3.2.18), corresponde a los datos. Así pues, un modelo del tipo VARMA (p, 0) con una matriz $\Phi^{(0)}$ cualquiera —la generalización a modelos con q distinto de cero es inmediata— no se puede obtener de forma única si no se imponen otro tipo de restricciones. El modelo referido no está identificado.

La restricción (3.2.8) es condición suficiente, pero no necesaria, de identificación, pues existen distintas formas de restringir un modelo VARMA para que esté identificado. La que se utiliza en este libro es la más habitual en la literatura (véase por ejemplo Tiao y Box, 1981); para formulaciones alternativas basadas en restringir Ω haciéndola igual a la matriz identidad véase, por ejemplo, Hannan (1970).

En resumen, si sólo se utiliza información de carácter muestral es necesario restringir de alguna manera la interdependencia de las variables, pues en otro caso no será posible obtener estimadores únicos de los parámetros del modelo. No obstante, cuando se considera información teórica sobre las variables estudiadas es posible identificar el modelo sin imponer las restricciones características de los modelos VARMA. Pero esto nos lleva a considerar un tipo distinto de modelo econométrico, el modelo econométrico estructural, que será estudiado en la sección cuarta.

Para terminar esta sección con un ejemplo de los modelos que se verosimilitud correspondiente a (3.2.14) viene dada por

$$L(\delta;\, X_T^1,\, X_0) = \prod_{t=1}^{T} L(\lambda;\, \mathbf{X}_t,\, X_{t-1}^{t-p}),$$

y los estimadores de λ se obtienen maximizando dicha expresión.

La falta de identificación ocurre cuando existen dos estructuras distintas, λ_a y λ_b, tales que

$$\prod_{t=1}^{T} L(\lambda_a;\, \mathbf{X}_t,\, X_{t-1}^{t-p}) = \prod_{t=1}^{T} L(\lambda_b;\, \mathbf{X}_t,\, X_{t-1}^{t-p}) = \max_{\lambda} \prod_{t=1}^{T} L(\lambda;\, \mathbf{X}_t,\, X_{t-1}^{t-p})$$

En ese caso no se puede realizar ningún tipo de inferencia, ya que contando únicamente con la información que proporciona la muestra no es posible determinar cuál de las dos estructuras alternativas, λ_a o λ_b, ha generado los datos observados. De ahí que el problema de identificación sea previo al de estimación, ya que, como es fácil demostrar, un parámetro no identificado no se puede estimar de forma consistente.

Por esa razón en la formulación más usual de los modelos VARMA se opta por incorporar como conjunto de restricciones que garantizan la identificación las recogidas en la ecuación (3.2.8). En palabras, lo que esta expresión impone es no permitir que los valores contemporáneos de X_h o de a_h entren en la ecuación de comportamiento de X_j, $j \neq h$, y con ello se garantiza la identificación del modelo.

Es sencillo comprobar como el modelo VARMA no está identificado si no se impone ningún tipo de restricción. Para ello, considérese un caso simplificado en el que q es igual a cero, con lo que se obtiene el modelo VAR (p), que sin la restricción (3.2.8) se puede expresar como:

$$\Phi^{(0)}\mathbf{X}_t - \Phi^{(1)}\mathbf{X}_{t-1} - \cdots - \Phi^{(p)}\mathbf{X}_{t-p} = \mathbf{a}_t \qquad (3.2.15)$$

Si en (3.2.15) se resuelve el sistema respecto a $\mathbf{X}_t$, premultiplicando por $[\Phi^{(0)}]^{-1}$ se obtiene

$$\mathbf{X}_t = \pi_1\mathbf{X}_{t-1} + \cdots + \pi_p\mathbf{X}_{t-p} + \mathbf{v}_t \qquad (3.2.16)$$

donde

$$\pi_j = [\Phi^{(0)}]^{-1}\Phi(j) \qquad \mathbf{v}_t = [\Phi^{(0)}]^{-1}\mathbf{a}_t,$$

y $\mathbf{v}_t$ cumple que

$$E(\mathbf{v}_t\mathbf{v}_t') = \Omega = [\Phi^{(0)}]^{-1}\Sigma\{[\Phi^{(0)}]^{-1}\}'. \qquad (3.2.17)$$

La función de verosimilitud de la muestra se puede obtener a partir de (3.2.16) y (3.2.17).

Ahora bien, si se premultiplica (3.2.15) por una matriz no singular H se obtiene

$$\bar{\Phi}^{(0)}\mathbf{X}_t - \bar{\Phi}^{(1)}\mathbf{X}_{t-1} - \cdots - \bar{\Phi}^{(p)}\mathbf{X}_{t-p} = \bar{\mathbf{a}}_t. \qquad (3.2.18)$$

donde $\bar{\Phi}^{(j)} = H\Phi^{(j)}$ y $\bar{\mathbf{a}}_t = H\mathbf{a}_t$.

Resolviendo (3.2.18) respecto a $\mathbf{X}_t$ se obtiene también el modelo (3.2.16), donde ahora

$$\bar{\pi}_j = [H\Phi^{(0)}]^{-1}H\Phi^{(j)} = [\Phi^{(0)}]^{-1}H^{-1}H\Phi^{(j)} = [\Phi^{(0)}]^{-1}\Phi^{(j)} = \pi_j$$

y de forma similar

$$\bar{\Omega} = [\Phi^{(0)}]^{-1}H^{-1}H\Sigma H'(H')^{-1}\{[\Phi^{(0)}]^{-1}\}' = \Omega.$$

Se ve de forma inmediata que una vez estimadas las matrices π_j y Ω no se puede determinar qué estructura, (3.2.15) o (3.2.18), corresponde a los datos. Así pues, un modelo del tipo VARMA (p, 0) con una matriz $\Phi^{(0)}$ cualquiera —la generalización a modelos con q distinto de cero es inmediata— no se puede obtener de forma única si no se imponen otro tipo de restricciones. El modelo referido no está identificado.

La restricción (3.2.8) es condición suficiente, pero no necesaria, de identificación, pues existen distintas formas de restringir un modelo VARMA para que esté identificado. La que se utiliza en este libro es la más habitual en la literatura (véase por ejemplo Tiao y Box, 1981); para formulaciones alternativas basadas en restringir Ω haciéndola igual a la matriz identidad véase, por ejemplo, Hannan (1970).

En resumen, si sólo se utiliza información de carácter muestral es necesario restringir de alguna manera la interdependencia de las variables, pues en otro caso no será posible obtener estimadores únicos de los parámetros del modelo. No obstante, cuando se considera información teórica sobre las variables estudiadas es posible identificar el modelo sin imponer las restricciones características de los modelos VARMA. Pero esto nos lleva a considerar un tipo distinto de modelo econométrico, el modelo econométrico estructural, que será estudiado en la sección cuarta.

Para terminar esta sección con un ejemplo de los modelos que se han estudiado en ella, considérese el modelo VAR empleado en Ballabriga y Sebastián (1992)[7] para analizar la relación entre déficit público y tipos de interés en la economía española. El modelo finalmente estimado, usando datos anuales, es

$$\begin{pmatrix} 1-1{,}003L+0{,}000003L^2 & -0{,}001L-0{,}0008L^2 & -0{,}0008L-0{,}0005L^2 \\ 0{,}001L+0{,}0003L^2 & 1-1{,}003L+0{,}000003L^2 & 0{,}0006L+0{,}0004L^2 \\ -0{,}005L-0{,}003L^2 & 0{,}008L+0{,}0005L^2 & 1-1{,}003L-0{,}0001L^2 \end{pmatrix}\begin{pmatrix} r_t \\ dp_t \\ alp_t \end{pmatrix}=$$

$$=\begin{pmatrix} 0{,}0003 & & e_{rt} \\ -0{,}0001 & + & e_{dt} \\ 0{,}0002 & & e_{at} \end{pmatrix}$$

$$\Omega=\begin{pmatrix} 0{,}000166 & & \\ 0{,}000017 & 0{,}000097 & \\ -0{,}000024 & -0{,}000013 & 0{,}000465 \end{pmatrix}$$

donde r es un tipo de interés a largo plazo, dp la capacidad/necesidad de financiación de las administraciones públicas en términos del PIB y alp los activos líquidos en manos del público por unidad de PIB.

A las innovaciones de este modelo se les denota por la letra e, en lugar de a, por razones que quedarán claras en el epígrafe cuarto. Análogamente, para su matriz de varianzas se usa Ω en vez de Σ.

3.3. Exogeneidad: interpretaciones y utilidad en el análisis aplicado

Uno de los puntos de interés en el análisis conjunto de variables económicas es la distinción entre *variables endógenas* y *variables exógenas*. Más adelante se procederá a una discusión formal del concepto de exogeneidad, pero, en términos intuitivos, se dice que una variable es exógena en el análisis que se pretende realizar si éste se puede hacer condicional a dicha variable y, por tanto, sin necesidad de modelizar expresamente la ecuación explicativa de la presunta variable exógena. Por el contrario, a toda variable que no sea exógena se le llama endógena, ya que no es posible realizar el análisis

[7] Estamos agradecidos a estos autores por permitirnos usar los resultados de este trabajo.

condicional sin perder, al menos, eficiencia; en consecuencia, si se desea hacer el máximo uso de la información disponible es necesario considerar su ecuación de comportamiento en el proceso de modelización.

Esto se puede aclarar con un ejemplo: se desea explicar el comportamiento de la variable X_1, que se sabe que está relacionada con otras dos variables, X_2 y X_3. Si X_3 es exógena pero X_2 no lo es, entonces: a) no es preciso construir la ecuación de comportamiento de X_3 para realizar el análisis deseado; b) sí se debe considerar el sistema bivariante que explica conjuntamente a X_1 y X_2, condicionado a X_3, aunque el interés del analista se concentre únicamente en la primera variable.

En el marco de modelos econométricos, el concepto de exogeneidad se desarrolló inicialmente en el seno de la Cowles Comission, véase Koopmans (ed., 1950). Posteriormente diferentes autores han incidido sobre el tema, y en Engle et al. (1983) se realiza un estudio muy completo y preciso del mismo, de modo que este trabajo constituye hoy en día el punto de referencia obligado para la presentación de la exogeneidad. En Ericsson (1992) se realiza una exposición de este concepto enlazándolo con el estudio de variables cointegradas, y Ericsson et al. (1991) contiene un desarrollo muy didáctico de éste y de otros temas econométricos.

El objetivo último del análisis de la exogeneidad es simplificar el modelo econométrico, ya que como se ha dicho el que una variable sea exógena implica que se puede considerar como dada, sin que ello suponga pérdida alguna de información con relación al fin con que se desea utilizar el modelo. Por eso el concepto de exogeneidad depende de la finalidad del modelo, definiéndose la exogeneidad de forma diferente para cada caso: para fines de inferencia estadística (por ejemplo para contrastar si un determinado parámetro del modelo es cero) el concepto relevante es el de *exogeneidad débil*, para fines de predicción el de *exogeneidad fuerte*, y para fines de simulación y control el de *superexogeneidad*.

La notación habitualmente empleada en econometría para referirse a modelos multiecuacionales reserva la letra Y para las variables endógenas, y a las variables exógenas se les representa por la letra Z. A partir de ahora se seguirá esta convención, y en vez de denotar a las n variables por $X_1, X_2, \ldots, X_n$, se usará el vector $X=(Y_1, Y_2, \ldots, Y_m, Z_1, Z_2, \ldots, Z_k)'$, siendo m el número de variables endógenas, k el de exógenas y $m+k=n$. Además, se supondrá que todas las variables que el usuario del modelo desea modelizar expresamente están incluidas en el conjunto formado por $Y_1, Y_2, \ldots, Y_m$, aunque no necesariamente han de ser todas las variables de este conjunto.

De forma similar, se redefinen las matrices y vectores del epígrafe anterior como

$$\mathbf{X}_t=\begin{pmatrix}\mathbf{Y}_t\\ \mathbf{Z}_t\end{pmatrix} \qquad X_t^1=(Y_t^1 \quad Z_t^1) \qquad X_0=(Y_0 \quad Z_0)$$

$$X_t^+=(Y_t^+ \quad Z_t^+)$$

siendo

$$\mathbf{Y}_t=\begin{pmatrix}\mathbf{Y}_{1t}\\ \mathbf{Y}_{2t}\\ \cdots\\ \mathbf{Y}_{mt}\end{pmatrix} \qquad \mathbf{Z}_t=\begin{pmatrix}Z_{1t}\\ Z_{2t}\\ \cdots\\ Z_{kt}\end{pmatrix}$$

$$Y_t^1=(\mathbf{Y}_1\mathbf{Y}_2 \ldots \mathbf{Y}_t)' \qquad Z_t^1=(\mathbf{Z}_1\mathbf{Z}_2 \ldots \mathbf{Z}_t)'$$

$$Y_t^+=\begin{pmatrix}Y_0\\ Y_t^1\end{pmatrix} \qquad Z_t^+=\begin{pmatrix}Z_0\\ Z_t^1\end{pmatrix}$$

y la función de densidad conjunta de la muestra (Y_T^1, Z_T^1) viene dada por

$$D(X_T^1/X_0,\ \delta)=D(Y_T^1,\ Z_T^1/Y_0,\ Z_0,\ \delta).$$

La exogeneidad se corresponde con la operación de *condicionar contemporáneamente* respecto a algunas variables. Se ha visto que condicionando respecto al pasado se llegaba a la expresión

$$D(Y_T^1,\ Z_T^1/Y_0,\ Z_0,\ \delta)=\prod_{t=1}^{T} D(\mathbf{Y}_t,\ \mathbf{Z}_t/Y_{t-1}^+,\ Z_{t-1}^+,\ \lambda),$$

que es en todo equivalente a (3.2.12), y las únicas modificaciones se deben a la nueva notación en términos de Y y de Z. Si se toma ahora la sucesión de funciones de densidad de la derecha de la igualdad, siempre es posible expresar cualquiera de ellas como

$$\begin{aligned} &D(\mathbf{Y}_t,\ \mathbf{Z}_t/Y_{t-1}^+,\ Z_{t-1}^+,\ \lambda)= \\ &=D(\mathbf{Y}_t/\mathbf{Z}_t,\ Y_{t-1}^+,\ Z_{t-1}^+,\ \lambda_1)\cdot D(\mathbf{Z}_t/Y_{t-1}^+,\ Z_{t-1}^+,\ \lambda_2). \end{aligned} \tag{3.3.1}$$

Recuérdese que el analista sólo está interesado en explicar el comportamiento de parte de las variables que maneja, y que todas ellas están incluidas en Y.

El problema que se plantea es bajo qué condiciones es posible utilizar sólo la función de densidad condicional $D(\mathbf{Y}_t/\mathbf{Z}_t,\ Y_{t-1}^+,\ Z_{t-1}^+,\ \lambda_1)$, sin que haya pérdida de información por no considerar la densidad marginal de $\mathbf{Z}_t$.

Para ello, supóngase en primer lugar que se pretende realizar inferencia estadística sobre un conjunto de parámetros, llamados a partir de ahora los parámetros de interés, ψ: éste es el caso cuando, por ejemplo, se trata de estimar el valor de una determinada elasticidad. El concepto de exogeneidad a aplicar es el de *exogeneidad débil*[8], y se dice que las variables contenidas en Z son débilmente exógenas respecto a los parámetros de interés si se cumple:

1) Los conjuntos paramétricos λ_1 y λ_2 en (3.3.1) no tienen elementos comunes y además son de variación libre, es decir, no están sometidos a restricciones cruzadas entre ellos.

2) El conjunto de parámetros de interés ψ es función exclusivamente de λ_1.

Una de las principales aportaciones de Engle et al. (1983) ha consistido en explicitar que existe un planteamiento de la exogeneidad, el reflejado por la exogeneidad débil, que se refiere a parámetros, pudiendo darse el caso de que una variable sea débilmente exógena respecto al parámetro ψ_1 y no lo sea respecto al parámetro ψ_2. Con esto introducen un concepto más preciso que el de *variables predeterminadas* de Koopmans (1950), quien define una variable predeterminada como aquélla cuyo valor en t es independiente de las innovaciones del modelo, $\mathbf{a}_{t+i}$, para todo $i \geqslant 0$.

Sin embargo, si el objetivo último del modelo no es estimar o contrastar ciertos parámetros, sino utilizarlo para hacer predicciones, la exogeneidad débil no es suficiente. Tal y como se observa en (3.3.1) los valores pasados de Y influyen en $\mathbf{Z}_t$; y por lo tanto en el modelo condicional $D(\mathbf{Y}_t/\mathbf{Z}_t,\ Y_{t-1}^+,\ Z_{t-1}^+,\ \lambda_1)$, $\mathbf{Z}_t$ no se puede considerar como dado para fines predictivos. En este caso, se requiere que la función de densidad de $(\mathbf{Y}_t,\ \mathbf{Z}_t)$ condicionada a su pasado se factorice de la siguiente forma:

$$\begin{aligned} &D(\mathbf{Y}_t,\ \mathbf{Z}_t/Y_{t-1}^+,\ Z_{t-1}^+,\ \lambda)= \\ &=D(\mathbf{Y}_t/\mathbf{Z}_t,\ Y_{t-1}^+,\ Z_{t-1}^+,\ \lambda_1)\cdot D(\mathbf{Z}_t/Z_{t-1}^+,\ \lambda_2). \end{aligned} \tag{3.3.2}$$

[8] Cuando los parámetros de interés se refieren a la relación de largo plazo entre dos variables cointegradas la exogeneidad débil es excesivamente restrictiva, y se puede sustituir por el concepto de exogeneidad débil a largo plazo, introducido en Dolado (1992a).

que tiene definiciones distintas y que se da o no en función de una serie de condicionantes: depende de qué tipo de análisis se pretenda, en relación a qué parámetros y según qué supuestos se hagan sobre el cambio de comportamiento en los estudios de simulación y control.

En cuanto a los contrastes de exogeneidad, su presentación detallada excede con mucho el marco de este libro. No obstante, sí es factible discutir el planteamiento general de los mismos, lo que se hace a continuación.

Comenzando con la exogeneidad débil, el problema que se plantea no es tanto contrastar si existe, como la forma de hacerlo sin tener que estimar todo el sistema. Es evidente que si se construye el modelo general representado por la función de densidad conjunta $D(\mathbf{Y}_t, \mathbf{Z}_t/Y^+_{t-1}, Z^+_{t-1}, \lambda)$, resulta sencillo contrastar la exogeneidad débil de Z; pero con la distinción entre variables endógenas y exógenas se trata precisamente de evitar el tener que modelizar todo el sistema. De ahí que en la literatura se hayan propuesto diversos contrastes basados en el multiplicador de Lagrange, que sólo requieren estimar el modelo bajo la hipótesis nula de que $\mathbf{Z}_t$ es débilmente exógena: véase por ejemplo Wu (1973), generalizado posteriormente en Hausman (1978), y Engle (1984).

En cuanto a la exogeneidad fuerte, requiere exogeneidad débil y causalidad en el sentido de Granger. Así, supuesto que Z es débilmente exógena, se trata de verificar si se cumple que

$$D(\mathbf{Z}_t/Y^+_{t-1}, Z^+_{t-1}, \lambda_2)=D(\mathbf{Z}_t/Z^+_{t-1}, \lambda_2) \tag{3.3.3}$$

es decir, si la función de densidad de $\mathbf{Z}_t$ condicionada al pasado del sistema sólo depende del pasado de las variables incluidas en Z, o lo que es lo mismo, que Y no causa en el sentido de Granger a Z. Este tipo de contrastes de causalidad ha sido extensamente tratado en la literatura: véase por ejemplo Granger (1969), Sims (1972) o Geweke et al (1983). Si en ellos se rechaza la hipótesis nula recogida en (3.3.3) se rechaza sin más la hipótesis de exogeneidad fuerte. Sin embargo, para incluir esta hipótesis en un modelo es necesario haber superado además del contraste de causalidad anterior el de exogeneidad débil.

Por último, para la superexogeneidad se completan los contrastes de exogeneidad débil con los de invarianza. Estos últimos están basados en contrastar la constancia de los parámetros de los modelos representados por $D(\mathbf{Y}_t/\mathbf{Z}_t, Y^+_{t-1}, Z^+_{t-1}, \lambda_1)$ —modelo condicional—, y $D(\mathbf{Z}_t/Y^+_{t-1}, Z^+_{t-1}, \lambda_2)$ —modelo marginal—, utilizando para ello métodos de estimación recursiva. Siguiendo a Hendry (1988), si los parámetros del modelo marginal (λ_2) han variado en el tiempo y los

del modelo condicional (λ_1) no lo han hecho, entonces no se rechaza la invarianza[14].

En lo que queda de este libro no entraremos a considerar con detalle el concepto de exogeneidad que se emplea en cada caso, ya que se sobreentenderá que es el necesario para que el desarrollo que se esté haciendo en ese momento sea correcto.

En el primer epígrafe se ha señalado que los modelos VARMA son especialmente útiles para fines predictivos. Ahora bien, si del vector de n variables X sólo se está interesado en predecir el subvector Y, entonces en vez de formular un modelo VARMA para las n variables incluidas en X se puede contrastar si Z es un vector formado exclusivamente por variables fuertemente exógenas[15]. En caso afirmativo, se procede a formular un modelo VARMA para las $n-k$ variables Y, que incluya además a las variables Z como explicativas. Su expresión general es:

$$\mathbf{\Phi}_p(L)\mathbf{Y}_t=\beta_s(L)\mathbf{Z}_t+\Theta_q(L)\mathbf{a}_t \tag{3.3.4}$$

En la literatura de series temporales a estos modelos se les suele denominar modelos VARMAX, en donde la X que aparece en la denominación hace referencia a que el modelo VARMA, respecto a las variables endógenas Y, está ampliado con variables explicativas exógenas[16]. La exogeneidad fuerte permite reducir la dimensión n de un modelo VARMA a la dimensión $n-k$ de un modelo VARMAX.

Los modelos VARMAX parten de una formulación VARMA, por lo que su identificación se basa en las restricciones del tipo (3.2.8) y

---

[14] Este contraste fue completado posteriormente en Engle y Hendry (1989): si el modelo marginal ha cambiado a lo largo del período muestral, estos autores proponen modelizar explícitamente dichos cambios, usando por ejemplo variables artificiales, hasta que se llegue a un modelo válido para toda la muestra. A continuación, se aumenta el modelo condicional con los términos que se añadieron al modelo marginal, y se contrasta si son significativos; en el caso en que no lo sean, no se rechaza la invarianza.

[15] De hecho en ocasiones ésta puede ser una información externa a la muestra, de manera que la exogeneidad no sea el resultado de un contraste estadístico sino del uso de información a priori. Piénsese por ejemplo en el caso en que se desea predecir los ingresos por turismo de nuestro país, y una de las variables explicativas es la renta de los países de origen de los turistas.

[16] Jenkins (1979), capítulo 2 y apéndices A.2 a A.6, proporciona una perspectiva general desde el punto de vista del análisis de series temporales de los distintos tipos de modelos lineales dinámicos, desde el más sencillo —univariante— al más complejo —VARMAX—. A este respecto, es de destacar que no utiliza las expresiones VARMA y VARMAX, sino «modelos estocásticos multivariantes con múltiples outputs» y «modelos multivariantes de función de transferencia con múltiples inputs y múltiples outputs», respectivamente.

en la ausencia de factores comunes; esto es lo que, como se verá en la sección siguiente, los diferencia de los modelos SEM.

3.4. Modelos econométricos de forma estructural y modelos de forma reducida

En las secciones anteriores se ha visto que los modelos multivariantes de series temporales concentran toda la información sobre las relaciones contemporáneas entre las variables endógenas en la matriz de varianzas de las innovaciones.

Pero este tipo de modelización puede ser insatisfactorio para su aplicación en el análisis de ciertos fenómenos económicos, ya que en ocasiones la teoría establece las relaciones causales entre las variables endógenas y, en tales casos, la determinación de los efectos contemporáneos puede ser uno de los principales objetivos del estudio.

Se ha visto anteriormente que en un modelo VARMA o VARMAX la relación contemporánea entre Y_1 y Y_2 se recoge habitualmente en la correlación entre las innovaciones de ambas variables. Sin embargo, si se dispone de información teórica sobre el problema que se está analizando, se puede interpretar esa relación contemporánea en términos de una influencia simultánea, es decir, apareciendo en la parte sistemática de Y_1 e Y_2. Con ello la dependencia contemporánea endógena se introduce también en las ecuaciones de comportamiento, eligiendo para ello, de entre todas las formas posibles de hacerlo, aquélla que señala la teoría.

En ese sentido la información extramuestral permite cambiar las restricciones de normalización del modelo VARMA o VARMAX, por otro conjunto de restricciones que viene especificado por la teoría. Con estas nuevas restricciones el sistema continúa estando identificado, en el sentido utilizado en la sección 3.2.

Si esto es así, es posible replantear el modelo VARMAX (3.3.4) de la sección anterior de la siguiente forma

$$\begin{aligned}\Phi_0^* \mathbf{Y}_t - \Phi_1^* \mathbf{Y}_{t-1} - \cdots - \Phi_p^* \mathbf{Y}_{t-p} = \\ = C_s^*(L)\mathbf{Z}_t + \Theta_0^* \mathbf{a}_t - \Theta_1^* \mathbf{a}_{t-1} - \cdots - \Theta_q^* \mathbf{a}_{t-q},\end{aligned} \tag{3.4.1}$$

$$E(\mathbf{a}_t \mathbf{a}_t') = \Sigma^*,$$

donde los asteriscos indican que la identificación se logra por el uso de restricciones basadas en información extramuestral, que normalmente consisten en restricciones cero sobre los coeficientes de las matrices Φ_j^*, C_s^* y Θ_j^* . Es importante recalcar que las restricciones que propone la teoría se imponen a priori, sin ningún tipo de contrastación previa con los datos, y todo el análisis está condicionado a su validez.

A los modelos cuya identificación se basa en restricciones extramuestrales impuestas por una teoría se les conoce como *Modelos de Ecuaciones Simultáneas* (de ahora en adelante SEM, Simultaneous Equations Model). El término de ecuaciones simultáneas se debe a que Y_{jt} puede depender de $Y_{ht} (h \neq j)$ y, a su vez, Y_{ht} depender de Y_{jt}: supóngase, por ejemplo, en un modelo anual que Y_j es el índice de precios al consumo y Y_h el salario nominal por hora de trabajo.

Los SEM son los modelos econométricos por excelencia, ya que combinan la modelización estadística con la información de la Teoría Económica, siendo ésta última la base de las restricciones de identificación. Precisamente por ello estos modelos no se consideran en un contexto puramente estadístico como es el análisis de series temporales, ya que la formulación y validez de los modelos estructurales dependen de información no contenida en los datos.

Se tiene así que, a diferencia de las restricciones que se imponen en la identificación de los modelos VARMA o VARMAX, las restricciones de un modelo SEM no son neutrales, pues implican direccionar la información contenida en los datos. Por ejemplo, un modelo SEM puede imponer que la única interpretación posible de la correlación contemporánea entre Y_1 y Y_2, es en términos de causalidad unidireccional de Y_2 a Y_1. Esto es algo que nunca se hace con un modelo VARMA, que se basa exclusivamente en la información muestral: en el modelo VARMA, con la convención (3.2.8) utilizada en este libro, la correlación contemporánea entre variables endógenas se recoge exclusivamente en la matriz de varianzas y covarianzas de las innovaciones, y no se interpreta nunca en términos de causalidad unidireccional o bidireccional.

Profundizando en este punto, en un modelo estructural una variable endógena puede venir afectada por el valor contemporáneo de otra variable endógena, pudiendo al mismo tiempo existir correlación contemporánea entre ambas innovaciones. Por el contrario, en un VARMA tal cosa no es posible por definición. Esto es debido a que si sólo se dispone de la información muestral, es imposible determinar qué hay detrás de una correlación, mientras que con información a priori se puede imponer desde fuera de la muestra la interpretación de esa correlación.

Evidentemente la pretendida superioridad de los modelos SEM es relativa, ya que con esa reformulación se deben incluir suficientes restricciones adicionales, que no imponía el modelo VARMA, a lo largo de las diversas matrices de parámetros para que el modelo esté identificado.

De ahí que cuando las restricciones de un modelo SEM son correctas, la mejora en la modelización puede ser considerable en términos de aumento en la eficiencia de los estimadores y de capacidad de explicar la realidad. Pero si esa información a priori es incorrecta, las estimaciones basadas en estos modelos serán generalmente inconsistentes, y los correspondientes contrastes falsos.

Uno de los principales problemas que el analista encuentra a la hora de construir un SEM es el de que la información procedente de la Teoría Económica suele ser insuficiente para restringir en la forma precisa los modelos SEM. De ahí que en la estimación de modelos de este tipo gran parte de las restricciones utilizadas sean subjetivas, y con frecuencia determinadas características impuestas a los modelos resultan erróneas.

La notación habitualmente empleada en la literatura econométrica para referirse a los modelos simultáneos difiere de la utilizada en los modelos VARMA, y por lo tanto de (3.4.1). El modelo SEM se expresa normalmente como

$$B(L)\mathbf{Y}_t + C(L)\mathbf{Z}_t = \mathbf{u}_t, \tag{3.4.2.a}$$

en donde

$$\begin{aligned} \phi(L)\mathbf{u}_t &= \theta(L)\mathbf{a}_t \\ E(\mathbf{a}_t\mathbf{a}_t') &= \Sigma. \end{aligned} \tag{3.4.2.b}$$

Las matrices polinomiales de (3.4.2) se definen como:

$$\begin{aligned} B(L) &= B_0 + B_1 L + \cdots + B_r L^r \\ C(L) &= C_0 + C_1 L + \cdots + C_s L^s \\ \phi(L) &= I - \phi_1 L - \cdots - \phi_p L^p \\ \theta(L) &= I - \theta_1 L - \cdots - \theta_q L^q. \end{aligned}$$

El elemento residual $\mathbf{u}_t$, tal y como se especifica en (3.4.2.a), viene a su vez generado por un proceso VARMA (p, q) y, por tanto muestra en general dependencia temporal. En consecuencia, el vector de innovaciones del sistema es $\mathbf{a}_t$.

Obsérvese que si el modelo SEM es un modelo global, en el sentido de que el vector de variables $\mathbf{X}_t$ incluye todas las variables relacionadas entre sí, los vectores $\mathbf{a}_t$ serán *auténticas innovaciones*, en el sentido de que son impredecibles en función de todo el conjunto de información existente hasta $t-1$. En efecto, ciertamente lo son en relación al conjunto de información utilizado en el modelo, pero dado

que la información no incluida no influye en la determinación de las variables del modelo, tampoco puede servir para predecir $\mathbf{a}_t$.

Para simplificar la exposición de ahora en adelante se supondrá, salvo mención expresa en contrario, $\phi(L=\theta(L)=I$. Con ello, el vector de innovaciones del sistema es directamente $\mathbf{u}_t$. Además, en la formulación (3.4.2) se incorporan suficientes restricciones para que el modelo esté identificado.

Un modelo econométrico estructural se puede expresar de forma más compacta como

$$A(L)\mathbf{X}_t = \mathbf{u}_t$$

siendo $A(L) = (B(L)\ C(L))$, $\mathbf{X}_t = (\mathbf{Y}_t', \mathbf{Z}_t')'$.

A partir de (3.4.2), y resolviendo el sistema para $\mathbf{Y}_t$, se puede obtener la *Forma Reducida* de un modelo econométrico: en ella se pone de manifiesto la dependencia de cada variable endógena respecto a las variables exógenas y a las variables retardadas. Para obtener la forma reducida se reexpresa el modelo SEM (3.4.2) como:

$$B_0\mathbf{Y}_t + B_1\mathbf{Y}_{t-1} + \cdots + B_r\mathbf{Y}_{t-r} + C(L)\mathbf{Z}_t = \mathbf{u}_t,$$

y premultiplicando esta expresión por B_0^{-1} y despejando $\mathbf{Y}_t$ se obtiene el modelo de forma reducida como:

$$\begin{aligned} \mathbf{Y}_t &= -B_0^{-1}B_1\mathbf{Y}_{t-1} - \cdots - B_0^{-1}B_r\mathbf{Y}_{t-r} - B_0^{-1}C(L)\mathbf{Z}_t + B_0^{-1}\mathbf{u}_t = \\ &= B_1^*\mathbf{Y}_{t-1} + \cdots + B_r^*\mathbf{Y}_{t-r} + C^*(L)\mathbf{Z}_t + \mathbf{u}_t^*, \end{aligned} \tag{3.4.3}$$

$$E(\mathbf{u}_t^*\mathbf{u}_t^{*\prime}) = B_0^{-1}\Sigma(B_0^{-1})'.$$

Es importante destacar que, calculada de esta manera, la forma reducida incorpora las restricciones a priori contenidas en la forma estructural, y sus parámetros son función de los parámetros de esa forma estructural: por eso a (3.4.3) se le conoce como *forma reducida restringida*.

Esto no significa que la única forma de estimar un modelo de forma reducida sea estimando el SEM (3.4.2) y luego resolviendo el sistema para obtener (3.4.3). En efecto, un modelo del tipo

$$\begin{aligned} \mathbf{Y}_t &= P_1\mathbf{Y}_{t-1} + P_2\mathbf{Y}_{t-2} + \cdots + P_r\mathbf{Y}_{t-r} + \pi^*(L)\mathbf{Z}_t + \mathbf{v}_t \\ E(\mathbf{v}_t\mathbf{v}_t') &= \Omega, \end{aligned} \tag{3.4.4}$$

está identificado, y no requiere ningún tipo de restricción a priori para ser estimado. De ahí que si se estima directamente (3.4.4) sin usar ningún tipo de información a priori, se obtiene la llamada *forma reducida general o sin restricciones*.

Por otra parte, (3.4.4) es un modelo VARMAX en donde los residuos son ruido blanco como consecuencia del supuesto simplificador previo sobre el elemento residual. Si se levanta este supuesto, entonces en general $\mathbf{v}_t$ sigue un proceso VARMA (p, q) y se tendría un modelo VARMAX general como el recogido en (3.3.4). Se tiene pues que *los modelos VARMAX no son más que modelos econométricos de forma reducida sin restricciones.*

Otro tipo de formulación que se deriva de un modelo estructural consiste en lo que se conoce en la literatura como *forma final*, que hace depender las variables endógenas exclusivamente de las variables exógenas. Despejando $\mathbf{Y}_t$ de (3.4.2):

$$\mathbf{Y}_t = -[B(L)]^{-1}C(L)\mathbf{Z}_t + [B(L)]^{-1}\mathbf{u}_t = Q(L)\mathbf{Z}_t + \mathbf{n}_t \qquad (3.4.5)$$

donde ahora el vector de errores $\mathbf{n}_t$ presenta dependencia en el tiempo, aun manteniendo el supuesto simplificador de que $\mathbf{u}_t$ es ruido blanco.

Implícita en esta formulación está la expresión de cada ecuación de la forma final como una función de transferencia. En efecto, llamando $adj[B(L)]$ a la matriz adjunta de $B(L)$ y $|B(L)|$ a su determinante, se tiene que $[B(L)]^{-1} = adj[B(L)]/|B(L)|$, y (3.4.5) equivale a

$$\mathbf{Y}_t = -\frac{adj[B(L)]\cdot C(L)}{|B(L)|}\mathbf{Z}_t + \frac{adj[B(L)]}{|B(L)|}\mathbf{u}_t$$

siendo $-adj[B(L)]\cdot C(L)/|B(L)|$ la función de transferencia de $\mathbf{Z}_t$ sobre $\mathbf{Y}_t$. En la forma final anterior todas las funciones de transferencia tienen el mismo denominador, con lo que tal forma final se puede expresar también como:

$$|B(L)|\mathbf{Y}_t = -adj[B(L)]\cdot C(L)\mathbf{Z}_t + adj[B(L)]\mathbf{u}_t. \qquad (3.4.6)$$

En (3.4.6) cabe destacar que todas las variables endógenas tienen asociado el mismo polinomio autorregresivo. Esta es una implicación que será considerada con detalle en el siguiente epígrafe.

Una vez introducida la forma reducida resulta más sencillo entender el papel de la información a priori a la hora de identificar un modelo estructural. Sea un ejemplo sencillo, donde hay dos variables endógenas con desfases $r=1$, una exógena, $C(L)=C_0$ y $\mathbf{u}_t$ es ruido blanco. El modelo queda de la forma:

$$B_0\begin{pmatrix} Y_{1t} \\ Y_{2t}\end{pmatrix} + B_1\begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1}\end{pmatrix} + C_0 Z_t = \begin{pmatrix} u_{1t} \\ u_{2t}\end{pmatrix} \qquad (3.4.7)$$

$$E(\mathbf{u}_t\mathbf{u}_t') = \Sigma.$$

Supóngase por un momento que esta forma estructural se formula sin restricciones, con lo que tendrá un total de trece parámetros (cuatro la matriz B_0, cuatro B_1, dos C_0 y tres Σ). Sin embargo, la expresión que realmente es posible estimar es

$$\begin{pmatrix} Y_{1t} \\ Y_{2t}\end{pmatrix} = P_1\begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1}\end{pmatrix} + \pi Z_t + \begin{pmatrix} v_{1t} \\ v_{2t}\end{pmatrix} \qquad (3.4.8)$$

$$E(\mathbf{v}_t\mathbf{v}_t') = \Omega,$$

ya que con (3.4.8), en donde se despeja el vector $(Y_{1t}\ Y_{2t})'$, y la hipótesis de normalidad se obtiene una especificación de la función de densidad de ese vector. En (3.4.8) se incluye un total de nueve parámetros, teniendo en cuenta los tres correspondientes a Ω. En principio los estimadores de los parámetros de la forma estructural habrían de obtener de las relaciones

$$\begin{aligned} -B_0^{-1}B_1 &= P_1 \\ -B_0^{-1}C_0 &= \pi \\ B_0^{-1}\Sigma(B_0^{-1})' &= \Omega, \end{aligned} \qquad (3.4.9)$$

pero es evidente que hay más incógnitas que ecuaciones, o lo que es lo mismo, existen infinitas combinaciones de B_0, B_1, C_0 y Σ que tienen asociados los mismos valores de P_1, π y Ω.

El papel de las restricciones a priori es restringir la forma estructural de tal manera que, dados unos valores para los parámetros de la forma reducida, sea posible determinar los valores de los parámetros de la forma estructural.

Si se imponen exactamente las restricciones necesarias para que el modelo estructural esté identificado, se dice que el modelo está *exactamente identificado*[17]: en este caso las formas reducida restringida y sin restringir coinciden, y la incorporación de información extramuestral no ha supuesto una ganancia de eficiencia en la estimación, aunque sí en la interpretación de los resultados usando la Teoría Económica. En este caso los parámetros estructurales se pueden estimar a partir de los parámetros estimados de la forma reducida no restringida utilizando (3.4.9) —método de mínimos cuadrados indirectos.

Pero si la información extramuestral permite imponer más restricciones que las estrictamente necesarias para identificar los parámetros estructurales, entonces en (3.4.9) «sobran ecuaciones». Así, si se

[17] En términos intuitivos, en (3.4.9) hay «tantas ecuaciones como incógnitas».

imponen restricciones cero en las matrices B_0, B_1 y C_0 de tal forma que sólo haya que estimar cinco parámetros entre las tres, entonces la forma reducida restringida contiene cinco parámetros independientes, mientras la forma reducida sin restringir tiene seis; todo ello aparte de los tres de la matriz de varianzas y covarianzas en ambos modelos. En este caso, en principio[18], se tiene un modelo SEM *sobreidentificado*, que no se puede estimar a partir de la forma reducida no restringida. La estimación se ha de hacer incorporando las restricciones extramuestrales contenidas en el modelo SEM, véase por ejemplo Durbin (1963); con ello la información a priori no sólo permite estimar los parámetros del modelo SEM, sino que, a partir de ellos, proporciona una estimación más eficiente de la forma reducida.

En el contexto del modelo SEM estático de m ecuaciones,

$$B_0\mathbf{Y}_t = C_0\mathbf{Z}_t + \mathbf{a}_t$$
$$E(\mathbf{a}_t\mathbf{a}_t') = \Sigma,$$

se derivan condiciones de orden y rango para la identificación (véase, por ejemplo, Fisher, 1966).

Así, una condición necesaria (*condición de orden*) para que los parámetros de una de las m ecuaciones estén identificados es que en esa ecuación el número de variables exógenas que no aparecen en la misma sea mayor o igual que el número de variables endógenas que aparecen como variables explicativas. Una condición necesaria y suficiente (*condición de rango*) para la identificación de una ecuación de este modelo es que el rango de una determinada submatriz, construida a partir de B_0, C_0 y Σ, sea $m-1$.

El tema de la identificación en modelos estructurales ha sido extensamente tratado en la literatura, y profundizar en el mismo está fuera de los objetivos de este libro, pero conviene añadir que la identificación se puede también conseguir imponiendo restricciones en la matriz de varianzas y covarianzas de las innovaciones (Fisher, 1966), con restricciones cruzadas entre ecuaciones tal y como se hace en la literatura sobre expectativas racionales (Wallis, 1980), etc. Para un estudio más detallado de la identificación en modelos estructurales véase Fisher (1966), Hannan (1971), Schmidt (1976), Hsiao (1983) o Judge et al. (1985).

Desde los trabajos de la Comisión Cowles en los años cincuenta,

[18] Se puede demostrar que el que haya más parámetros en la forma reducida que en la estructural no es condición suficiente para que esta última esté identificada. Precisamente por eso se ha de cumplir la llamada condición de rango para que el modelo esté identificado.

el análisis estructural[19], entendido de forma restringida como el análisis conducente a medir la influencia de cada una de las variables incluidas en el sistema (tanto endógenas como exógenas) en cualquier ecuación del mismo, se ha realizado a partir del modelo SEM. Si tras el análisis estructural se desea realizar ejercicios de predicción y/o calcular distintos multiplicadores —véase los epígrafes nueve y trece— de las variables exógenas sobre las endógenas, se calcula la correspondiente forma reducida restringida.

Sin embargo, la crítica que Sims (1980) hace al modelo SEM, en el sentido de que sus restricciones pueden ser arbitrarias, ha hecho que el uso directo de modelos de forma reducida sin restringir haya adquirido mayor difusión. De hecho, como señalan Clements y Mizon (1991), la crítica de Sims se encuentra ya en Liu (1960), pero en dicho momento los resultados obtenidos con los modelos SEM eran muy esperanzadores y ha sido posteriormente, al deteriorarse la capacidad predictiva de los modelos SEM, cuando el trabajo de Sims (1980), en el que se propone un tipo alternativo de modelos, ha causado impacto en las aplicaciones econométricas. Tal crítica ha conducido a buscar la realización del análisis estructural a partir de la forma reducida sin restringir, o, en otras palabras, a obtener un modelo estructural a partir de una forma reducida en la que no se haya considerado ningún tipo de conocimiento a priori.

Si se admite que los residuos $\mathbf{u}_t$ de la forma estructural pueden venir generados por un proceso VARMA tal y como se especifica en (3.4.2), entonces los residuos $\mathbf{v}_t$ de la forma reducida sin restringir (3.4.4) no son ruido blanco, sino que a su vez vienen generados por el proceso VARMA

$$R(L)\mathbf{v}_t = S(L)\mathbf{e}_t.$$

Al proceso ruido blanco $\mathbf{e}_t$, que está en la base de este proceso VARMA, se le denomina el vector de innovaciones de la forma reducida general.

El análisis estructural a partir de la forma reducida sin restringir se basa en que el vector de innovaciones $\mathbf{e}_t$ está relacionado con el correspondiente vector de innovaciones del modelo estructural (3.4.2), $\mathbf{a}_t$, mediante la expresión[20]

$$B_0^{-1}\mathbf{a}_t = \mathbf{e}_t. \tag{3.4.10}$$

[19] En el epígrafe trece se profundiza en el empleo de un modelo econométrico para realizar un análisis estructural.

[20] Esto se comprueba de forma inmediata en el caso en que los residuos de la forma estructural $\mathbf{u}_t$ son ruido blanco, ya que entonces (3.4.10) simplemente equivale a $B_0^{-1}\mathbf{u}_t = \mathbf{v}_t$.

Análogamente, la matriz de varianzas y covarianzas de $\mathbf{e}_t$, que se denominará Ω, es una matriz no restringida que se relaciona con la matriz de varianzas y covarianzas de $\mathbf{a}_t$, que se denotaba por Σ, del siguiente modo:

$$B_0^{-1}\Sigma(B_0^{-1})'=\Omega$$

Cuando se tienen estimadores de los parámetros de un modelo de forma reducida sin restringir, se dispone de estimadores de las innovaciones $\mathbf{e}_t$ y de la matriz Ω. Con ello, la forma de realizar el análisis estructural consistirá en buscar una transformación $B_0\mathbf{e}_t$ que genere las innovaciones del modelo estructural $\mathbf{a}_t$, con una matriz de varianzas y covarianzas Σ. En este contexto el problema de la identificación estructural tratado anteriormente consiste en establecer condiciones bajo las cuales a partir de $\mathbf{e}_t$ y Ω se puede obtener $\mathbf{a}_t$ y Σ. En este caso la identificación implicará imponer también restricciones en la matriz Σ.

Al igual que el resto de modelos que se han estado considerando, los modelos VAR son un caso particular de la representación de Wold —véase la sección 2.2 para el caso escalar— para un vector de variables. Así King et al. (1991) plantean el problema de identificación de los parámetros estructurales a partir de la forma reducida de un modo más general que el aparecido en la literatura sobre modelos VAR, y lo hacen de la siguiente manera. En la descomposición de Wold el vector de variables económicas viene determinado por

$$\psi(L)\mathbf{e}_t,$$

en donde $\psi(L)$ es un polinomio matricial de orden teóricamente infinito y $\mathbf{e}_t$ el vector de innovaciones. En $\psi(L)$ la matriz ψ_0 es la unidad, por lo que tal descomposición es un modelo de forma reducida, en el que Ω es la matriz de varianzas y covarianzas de $\mathbf{e}_t$. El correspondiente modelo estructural vendría dado por

$$\Gamma(L)\mathbf{a}_t,$$

donde $\Gamma_0 \neq I$ y la matriz de varianzas y covarianzas de a_t es Σ.

El problema de identificación consiste en obtener condiciones mediante las que se puedan deducir las innovaciones estructurales y el polinomio matricial $\Gamma(L)$ a partir de $\mathbf{e}_t$ y $\psi(L)$. Este planteamiento es especialmente útil cuando las variables están cointegradas, ya que entonces se puede identificar el modelo estructural utilizando también restricciones sobre la matriz de multiplicadores a largo plazo

$$\Gamma(1)=\sum_{j=0}^{\infty}\Gamma_j.$$

El desarrollo de este tema va más allá de los objetivos de este libro, por lo que se remite al lector interesado a la referencia citada anteriormente.

Volviendo al planteamiento de los modelos VAR, existen varias formas de determinar $\mathbf{a}_t$ y Σ conocidos $\mathbf{e}_t$ y Ω. La propuesta de Sims (1980) consiste en suponer que Σ es diagonal y B_0 triangular inferior. Otros autores —Blanchard y Watson (1986), King et al. (1991)— han realizado modificaciones sobre el trabajo original de Sims.

Este punto se puede ilustrar acudiendo nuevamente al trabajo de Ballabriga y Sebastián (1992), ya mencionado al final del epígrafe segundo. Allí se presentaba un modelo VAR con una matriz de varianzas y covarianzas de las innovaciones $\mathbf{e}_t$ dada por

$$\Omega=\begin{pmatrix} 0{,}000166 & & \\ 0{,}000017 & 0{,}000097 & \\ -0{,}000024 & -0{,}000013 & 0{,}000465 \end{pmatrix}$$

donde la primera variable era la innovación en la ecuación del tipo de interés (e_{rt}), la segunda la innovación en la ecuación del déficit público (e_{dt}) y la tercera la innovación en la ecuación de los activos líquidos en manos del público (e_{at}).

A partir de aquí los autores proponen distintas estructuras compatibles con el modelo VAR que han estimado. Todas estas estructuras alternativas tienen en común la diagonalidad de Σ y la triangularidad de B_0, siguiendo la propuesta original de Sims (1980).

Una primera posible estructura es la caracterizada por la matriz[21]

$$B_0=\begin{pmatrix} -0{,}172 & 1 & 0 \\ 1 & 0 & 0 \\ 0{,}133 & 0{,}11 & 1 \end{pmatrix}$$

es decir,

$$\begin{aligned} e_{dt}&=a_{dt}\\ e_{rt}&=0{,}172e_{dt}+a_{rt}\\ e_{at}&=-0{,}133e_{rt}-0{,}11e_{dt}+a_{at}\end{aligned}$$

siendo a_{rt}, a_{dt} y a_{at}, las innovaciones de la forma estructural.

[21] Es inmediato comprobar como B_0 es triangular si se reescribe el sistema para que la primera ecuación corresponda a la deuda, la segunda al tipo de interés y la tercera a los activos líquidos en manos del público.

De acuerdo con esta estructura, la innovacion dominante corresponde al déficit público, que afecta al resto de innovaciones del sistema pero no viene afectada por ellas; y la innovacion del tipo de interés afecta a la innovacion de la liquidez, medida por *alp*, pero no a la inversa.

Ahora bien, ésta no es la única estructura compatible con los datos, o equivalentemente con la matriz Ω estimada. Existen varias, así sería igualmente válida la representada por

$$B_0 = \begin{pmatrix} 1 & -0{,}165 & 0{,}047 \\ 0 & 1 & 0{,}027 \\ 0 & 0 & 1 \end{pmatrix}$$

es decir,

$$\begin{aligned} e_{at} &= a_{at} \\ e_{dt} &= -0{,}027 e_{at} + a_{dt} \\ e_{rt} &= 0{,}165 e_{dt} - 0{,}047 e_{at} + a_{rt} \end{aligned}$$

y ahora la perturbación dominante corresponde a la liquidez, siendo la segunda en importancia la del déficit público.

Este ejemplo ilustra cómo dos explicaciones claramente distintas de la realidad son igualmente compatibles con los datos, y corresponde al analista seleccionar, ya en la etapa de interpretación de los resultados finales de una estimación en donde ninguna de ellas se ha impuesto a priori, cuál es la más plausible.

Con la propuesta de Sims no sólo se puede hacer el análisis estructural mencionado, sino que la ortogonalización de las innovaciones estructurales que supone la diagonalización de Σ, junto con la forma que se da a la matriz B_0, permiten realizar el análisis dinámico de cómo la innovación correspondiente a una variable endógena concreta se propaga a través del modelo a todas las variables endógenas. En el modelo SEM, en donde las innovaciones pueden estar correlacionadas y la matriz B_0 no es necesariamente triangular, este análisis dinámico sobre las innovaciones no se puede hacer sin un ulterior desarrollo, lo que implicaría cambiar la especificación inicial del modelo SEM.

La propuesta de Sims para realizar análisis dinámico de las innovaciones consiste en establecer una estructura recursiva en la dependencia contemporánea de las innovaciones. La discusión sobre modelos simultáneos y modelos recursivos se tratará con detalle en la sección 3.6.

Al procedimiento de Sims se le señala la ventaja de partir de una estimación no restringida de la forma reducida, y sobre ella explicitar con toda claridad la manipulación —en principio basada, al menos en parte, sobre los dictámenes de la Teoría Económica— necesaria para obtener las innovaciones estructurales y la matriz B_0 que establece relaciones de causalidad entre las variables endógenas.

Este procedimiento suele realizarse a partir de un modelo VAR para la forma reducida, con lo que se considera que todas las variables son endógenas, y por lo tanto tampoco se incorpora ninguna idea a priori sobre variables exógenas. Como contrapartida se tiene que el número de variables que se considera suele ser pequeño, por ejemplo seis en King et al. (1991). Además, conviene advertir que los modelos VAR no son apropiados —véase Maravall (1992)— para modelizar series ajustadas de estacionalidad o tendencias. La razón de ello quedará clara en el capítulo siguiente, en donde se explica que dichas series estimadas están generadas a partir de procesos que contienen una media móvil no invertible y, por tanto, no admiten una formulación autorregresiva convergente.

No obstante, la Teoría Económica puede sugerir restricciones en otras matrices de parámetros del modelo estructural que el procedimiento de Sims, para evitar arbitrariedades en la implementación de tales restricciones, prefiere ignorar. De hecho el modelo SEM supone un marco en el que es posible incorporar al máximo los resultados de la teoría.

Así por ejemplo, ocurre muchas veces que se puede especificar una o algunas de las ecuaciones del modelo SEM que representa el problema que se quiere estudiar, pero resulta imposible especificar el modelo completo. En tales situaciones, como se verá más adelante, es posible la estimación aislada de la parte del modelo global que se consigue especificar, usando procedimientos de información limitada o de variables instrumentales. En tales casos, la aportación de la teoría es muy importante para determinar y escoger los instrumentos con los que resolver el problema de identificación y estimación. En otras ocasiones la teoría puede ser también de ayuda en la formulación de la dinámica del modelo, tanto en la parte referida a las variables como a los residuos.

Como conclusión se puede decir que ambos modos de realizar análisis estructural, especificando y estimando un modelo SEM o a partir de la estimación de la forma reducida general, son teóricamente aceptables, con ventajas e inconvenientes en ambos casos en la aplicación a datos reales.

Procediendo a partir del modelo SEM se ha visto que se tiene una forma reducida restringida que se puede comparar con la forma reducida no restringida, lo que permite contrastar si las restricciones impuestas en el modelo SEM son compatibles con los datos. Para aclarar la relación entre forma estructural, forma reducida restringida y forma reducida sin restringir, así como la manera de realizar este contraste, considérese el siguiente modelo, extraído de Espasa (1975).

El modelo seleccionado es un modelo trimestral que describe la relación entre precios y salarios en el Reino Unido durante el período 1950-1970. Consta de cuatro ecuaciones, y las variables endógenas son:

— Indice de salarios pactados por convenio (W_t);
— Indice de ganancias salariales totales por empleado (E_t);
— Indice de precios al consumo (P_t);
— Indice de paro (U_t).

Las variables exógenas son:

— Número de huelgas en la industria excluida minería del carbón (SP_t);
— Cotizaciones sociales a cargo de la empresa (CE_t);
— Indice de precios de importaciones (PM_t);
— Producto Interior Bruto (PIB_t).

Todas las variables están expresadas en logaritmos. El modelo además contiene variables artificiales, que para simplificar no se desarrollan en este ejemplo.

La forma estructural que se estima viene dada por[22]

— Ecuación de Salarios Pactados:

$$(1-0{,}775L+0{,}672L^2-0{,}402L^3+0{,}201L^4)W_t+(-0{,}395L-0{,}145L^3)E_t+$$
$$+(-0{,}065L-0{,}091L^3)P_t+(0{,}032-0{,}032L)U_t-0{,}01L\ SP_t=a_{1t}$$

— Ecuación de Ganancias Salariales Totales:

$$(-0{,}539-0{,}150L+0{,}395L^2-0{,}395L^3+0{,}337L^4-0{,}337L^5)W_t+$$
$$+(1-0{,}377L)E_t+(-0{,}078L+0{,}144L^5)P_t+(-0{,}010L-0{,}009L^5)SP_t=a_{2t}$$

— Ecuación de Precios:

$$(-0{,}340-0{,}007L-0{,}240L^3)E_t+(1-0{,}362L+0{,}24L^3-0{,}207L^4)P_t-$$
$$-0{,}830L\ CE_t-0{,}084L\ PM_t=a_{3t}$$

— Ecuación de Paro:

$$(1{,}792-1{,}792L)W_t+(1-L+0{,}339L^2)U_t+4{,}11\ PIB_t=a_{4t}$$

[22] Véase la tabla 20, página 128 del trabajo citado; se ha cambiado la parametrización original para hacerla más coherente con los desarrollos teóricos de este epígrafe.

Obsérvese que la matriz de relaciones contemporáneas entre las variables endógenas, B_0, es igual a

$$B_0=\begin{pmatrix} 1 & 0 & 0 & 0{,}032 \\ -0{,}539 & 1 & 0 & 0 \\ 0 & -0{,}340 & 1 & 0 \\ 1{,}792 & 0 & 0 & 1 \end{pmatrix} \tag{3.4.11}$$

El sistema se puede expresar de forma matricial como

$$B_0\mathbf{Y}_t+B^*(L)\mathbf{Y}_{t-1}+C(L)\mathbf{Z}_t=\mathbf{a}_t \tag{3.4.12}$$

Despejando de (3.4.12), se tiene que

$$\mathbf{Y}_t=-B_0^{-1}B^*(L)\mathbf{Y}_{t-1}-B_0^{-1}C(L)\mathbf{Z}_t+B_0^{-1}\mathbf{a}_t. \tag{3.4.13}$$

Eso da lugar a la forma reducida restringida, cuyos coeficientes se detallan en el cuadro 3.1.

Aparentemente la forma reducida restringida tiene muchos más coeficientes que la forma estructural. Sin embargo el número de parámetros independientes es el mismo, ya que los coeficientes de la forma reducida restringida se obtienen a partir de los de la forma estructural mediante la transformación (3.4.13).

Evidentemente se podría haber estimado directamente un modelo de la forma

$$\mathbf{Y}_t=D(L)\mathbf{Y}_{t-1}+E(L)\mathbf{Z}_t+\mathbf{v}_t \tag{3.4.14}$$

sin imponer a priori las restricciones

$$D(L)=-B_0^{-1}B^*(L) \qquad E(L)=-B_0^{-1}C(L).$$

Si se hace eso se obtiene la forma reducida sin restringir, y si las restricciones impuestas al modelo estructural (3.4.12) son correctas, el ajuste obtenido con (3.4.13) ha de ser similar al obtenido en (3.4.14). En tal situación ocurrirá que los coeficientes de (3.4.13) serán, globalmente considerados, parecidos a los de (3.4.14). La forma reducida sin restringir estimada también figura en el cuadro 3.1: se comprueba que sus estimaciones son similares a las de la forma reducida restringida.

Con las estimaciones anteriores se puede contrastar la forma reducida restringida frente a la que no incluye restricciones. El estadístico formado a partir del cociente de los máximos de las

CUADRO 3.1. *Formas reducidas restringida y sin restringir*

		W_{t-1}	W_{t-2}	W_{t-3}	W_{t-4}	W_{t-5}	E_{t-1}	E_{t-3}	P_{t-1}	P_{t-3}	P_{t-4}	U_{t-1}	U_{t-2}	SP_{t-1}	SP_{t-5}	CE_{t-1}	PM_{t-1}	PIB_t		
W	R	0,761	−0,714	0,428	−0,214	0	0,420	0,150	0,069	0,100	0	0	0,012	0,012	0	0	0	0,140	R	W
	SR	0,748	−0,714	0,388	−0,133	−0,110	0,439	0,194	0,108	0,170	−0,062	0	0,012	0,014	−0,006	−0,017	−0,028	0,212	SR	
E	R	0,561	−0,781	0,625	−0,452	0,337	0,526	0,175	0,115	0,050	−0,144	0	0,006	0,016	0,009	0	0	0,076	R	E
	SR	0,663	−0,790	0,610	−0,373	0,253	0,603	0,083	0,076	0,090	−0,203	0	0,006	0,020	0,008	−0,237	−0,027	0,207	SR	
P	R	0,191	−0,266	0,213	−0,154	0,115	0,212	0,268	0,401	−0,222	0,158	0	0,002	0,006	0,003	0,830	0,084	0,026	R	P
	SR	0,312	−0,251	0,247	−0,186	0,195	0,071	0,326	0,235	−0,180	0,124	0	−0,012	0,005	0	0,811	0,107	−0,122	SR	
U	R	0,430	0,380	−0,760	0,380	0	−0,750	−0,280	−0,130	−0,170	0	1	−0,360	−0,021	0	0	0	−4,360	R	U
	SR	0,520	0,250	−0,270	−2,520	1,430	−0,970	2,060	−0,560	−0,640	−0,810	1	−0,350	−0,094	−0,004	4,940	0,730	−3,910	SR	

correspondientes funciones de verosimilitud toma el valor de 61,45 y, asintóticamente, tiene una distribución χ^2 con cuarenta y cuatro grados de libertad: el valor que toma el estadístico está alrededor del valor crítico con un nivel de significación del cinco por cien. Sin embargo, este contraste, al ser asintótico, no penaliza por la diferencia de grados de libertad entre los dos modelos; si se tiene esto en cuenta, la simplificación de la forma reducida restringida no se rechaza, y en consecuencia las restricciones implícitas en el modelo SEM parecen ser compatibles con los datos.

Como *conclusión* de lo tratado en esta sección, tenemos que:

1) Los modelos econométricos multiecuacionales se pueden clasificar en modelos estructurales (SEM) o modelos de forma reducida. La formulación lineal general de estos últimos es análoga a la de los modelos VARMA, y en su especificación la Teoria Económica prácticamente sólo se utiliza para determinar las n variables que deben incluirse en el vector.

2) Si en esta forma reducida k de las n variables son exógenas, entonces la representación del sistema económico se puede simplificar, ya que basta formular un modelo de forma reducida VARMAX con $m=n-k$ ecuaciones. La reducción de un modelo VARMA de n ecuaciones a un modelo VARMAX de $n-k$ ecuaciones puede ser el resultado de contrastar la exogeneidad de algunas variables, aunque en general es consecuencia de utilizar resultados teóricos que indican que la causalidad va de las k variables exógenas a las $n-k$ variables endógenas.

3) La presencia de un componente de medias móviles en los residuos obliga a que la estimación del modelo no pueda hacerse ecuación por ecuación, sino que tenga que realizarse conjuntamente. Si el modelo VARMA no tiene parte de medias móviles, entonces queda reducido a un modelo VAR, que, en general, sí puede estimarse eficientemente ecuación por ecuación.

4) Aunque el proceso de estimación de la forma reducida no se haya realizado tomando el sistema en conjunto, para fines de predicción es imprescindible la consideración conjunta del sistema siempre que exista causalidad multidireccional.

5) El modelo SEM incorpora, respecto a los modelos de forma reducida, una direccionalidad contemporánea entre sus variables endógenas, que es su principal característica diferenciadora.

6) La construcción de un modelo SEM, bien sea directamente, bien indirectamente por el procedimiento de Sims de transformar las innovaciones de la forma reducida sin restringir en innovaciones estructurales, es ineludible cuando se quiere estimar y contrastar

hipótesis sobre los parámetros que recogen los efectos contemporáneos entre variables endógenas.

7) Si además de lo señalado en el punto anterior se pretende un análisis dinámico de cómo las innovaciones correspondientes a una variable endógena se propagan en el modelo, es necesario —tal y como se hace en Sims (1980)— que la matriz de varianzas y covarianzas de las innovaciones estructurales sea diagonal.

8) La predicción con un modelo SEM se ha de realizar a partir de su correspondiente forma reducida, que incorpora las restricciones del SEM y por eso se le conoce como forma reducida restringida. Cuando el interés del analista es meramente predictivo la posible ventaja del modelo SEM radica exclusivamente en proporcionar una forma reducida restringida, que si es correcta generará predicciones más eficientes que una forma reducida no restringida.

Así pues, la construcción de un tipo u otro de modelo depende de la utilidad que el usuario quiera darle. Para fines predictivos los modelos de forma reducida y, por tanto, los modelos VARMA(X) y VAR(X) desarrollados en la literatura de series temporales son adecuados. Por contra, cuando el objetivo principal que se pretende con un modelo es contrastar determinadas hipótesis sobre parámetros que recogen causalidad contemporánea, es forzoso considerar un modelo SEM, bien estimándolo directamente a partir de una formulación inicial con suficientes restricciones de identificación, o bien deduciéndolo a partir de la forma reducida estimada e introduciendo a posteriori las condiciones de identificación que se consideren convenientes.

3.5. La estrategia SEMTSA en la construcción de modelos econométricos

En el epígrafe anterior se adelantó que en la práctica gran parte de la información en la que se basan los modelos estructurales es de una validez discutible: normalmente no se dispone de un número suficiente de restricciones sólidamente avaladas por la teoría, y la identificación termina apoyándose en supuestos subjetivos, lo que provoca fuertes discrepancias sobre la utilidad de los modelos estructurales.

Por otra parte, el procedimiento de series temporales ha consistido en construir aquellos modelos que los datos admiten, ignorando supuestas restricciones teóricas que no pueden ser directamente contrastadas. La elaboración de tales modelos se dirige principalmente a obtener la estructura dinámica necesaria para explicar los datos; pero

al no hacer uso de la información extramuestral que proporciona la Teoría Económica, terminan siendo, al menos teóricamente, modelos menos eficientes.

Vistas estas dos metodologías de modelización, un requisito básico de coherencia exige que las restricciones que un modelo estructural impone sean consistentes con los resultados que se obtienen mediante modelos VARMA. La información extramuestral que incorporan los modelos estructurales debe permitir aumentar la eficiencia de la estimación, y en ese sentido los modelos SEM y VARMA diferirán; pero puesto que tanto un SEM bien especificado como un VARMA son formulaciones adecuadas de una misma realidad, deben coincidir en lo básico.

Esto sugiere la integración de los enfoques estructural y de series temporales en la construcción de modelos econométricos, en una síntesis que se conoce en la literatura como la *estrategia SEMTSA* (Structural Econometric Modelling Time Series Analysis)[23]. Este enfoque se basa en[24]:

1) Los modelos estructurales lineales y dinámicos son un caso particular de los procesos estocásticos VARMA.

2) El supuesto de que una variable es exógena impone fuertes restricciones en los parámetros de un proceso VARMA.

3) Las estructuras dinámicas de un modelo SEM suelen estar muy restringidas en virtud de los supuestos incorporados al modelo.

4) Si las variables exógenas siguen a su vez un proceso VARMA, entonces cada una de las ecuaciones de la forma final del modelo estructural se puede representar como un modelo ARMA univariante.

De acuerdo con el enfoque SEMTSA de modelización econométrica, en una primera etapa se construye un modelo estructural inicial. Una vez estimado este modelo estructural, se derivan las ecuaciones de la forma final, que como se ha dicho anteriormente están fuertemente restringidas respecto a las que se obtendrían sin usar más información que la muestral.

Por lo tanto la segunda etapa consiste en estimar las funciones de transferencia basadas en los datos, y comprobar si realmente ambos modelos explicativos coinciden en lo esencial.

Estas comparaciones no sólo permiten comprobar hasta qué punto los supuestos incorporados al SEM son compatibles con los datos, sino que también proporcionan una visión alternativa de las características dinámicas de las variables que se analizan.

[23] Este enfoque se propone en Zellner y Palm (1974); Wallis (1977) y Zellner (1979) discuten con profundidad sus implicaciones.

[24] Véase Zellner (1979), sección 3.

Si ambas especificaciones resultan coherentes, entonces las restricciones teóricas no han sido contradichas por los datos, y el SEM parece adecuado para reflejar la realidad observada. Por el contrario, si se produce una contradicción entre las implicaciones del SEM y los modelos de series temporales, es necesario reformular aquél, para lo cual la información que proporcionan estos últimos es de gran utilidad.

Esta contrastación del modelo supone generalizar el contraste realizado en el ejemplo con el que acababa la sección anterior. Allí se comparaba la forma reducida (restringida) implícita en el modelo SEM con una forma reducida supuestamente no restringida, que se definía con las mismas variables exógenas y con los mismos desfases de variables endógenas y exógenas que aparecían en el modelo estructural. Con ello se contrastaba puramente las restricciones cero que imponía el modelo SEM, bajo el supuesto de que el modelo incluía todas las variables relevantes con sus correspondientes retardos. El enfoque SEMTSA va más allá, ya que también se contrasta la formulación dinámica que el modelo SEM incorpora.

Es importante destacar que el hecho de que la forma reducida de un modelo SEM no entre en contradicción con los modelos elaborados a partir de los datos, no es garantía de que sea el modelo estructural correcto. Ello es debido a que pueden existir distintos modelos SEM con formas reducidas restringidas compatibles con un mismo modelo de series temporales, tal y como se vio en la sección anterior con el análisis estructural que realizan Ballabriga y Sebastián (1992).

Desde el punto de vista histórico, la estrategia SEMTSA, y especialmente el trabajo pionero de Zellner y Palm en 1974, ha supuesto la integración de dos enfoques de modelización que hasta ese momento se tenían como alternativos: los econometras tradicionales rechazaban las técnicas estadísticas de análisis de series temporales porque no tenían en cuenta la información derivada de la Teoría Económica, mientras los analistas de series temporales criticaban a los econometras el que no se preocupasen por considerar una formulación dinámica general en sus modelos.

En la actualidad, por el contrario, los enfoques estructurales y de series temporales se consideran instrumentos complementarios en la elaboración de modelos econométricos.

En esta línea se ha desarrollando la metodología propulsada por Sargan y que desarrollada por Hendry, Richard, Mizon, etc., es conocida como «metodología econométrica de la LSE» (London School of Economics)[25]. En ella se proporciona un marco general de

[25] Véase Sargan (1964, 1980), Hendry (1979, 1980, 1983), Davidson et al (1978), Gilbert (1989), Hendry y Wallis (1984), Hendry y Richard (1982, 1983), Mizon (1984, 1989), Mizon y Richard (1986), Spanos (1986, 1988), Ericson et al. (1991), etc.

referencia para los modelos econométricos, definiendo explícitamente la Teoría Económica como un conjunto de restricciones que se imponen al proceso generador de los datos.

3.6. Modelos recursivos

En los epígrafes anteriores se ha discutido extensamente que en un modelo simultáneo general, además de posibles relaciones desfasadas temporalmente entre sus variables, existe relación contemporánea entre todas o parte de las variables endógenas, recogidas en la matriz B_0. Por ello se dice que el modelo es simultáneo, o directamente simultáneo. En este sentido, los modelos de forma reducida o los modelos VARMA/VARMAX son sólo indirectamente simultáneos, ya que la relación contemporánea entre variables endógenas no aparece en la parte funcional, sino en la matriz de varianzas y covarianzas de la innovaciones.

A estos *modelos indirectamente simultáneos* y con un elemento residual ruido blanco se les denomina en Zellner (1962) Ecuaciones Aparentemente No Relacionadas (SURE, Seemingly Unrelated REgressions). En dicho artículo se demuestra que, salvo que se cumpla al menos una de las dos siguientes condiciones:

a) que todas las ecuaciones tengan exactamente las mismas variables explicativas,

b) que la matriz de varianzas y covarianzas de los residuos —que vienen generados por un proceso ruido blanco— sea diagonal,

la estimación individual de las ecuaciones lleva a pérdidas de eficiencia frente a una estimación conjunta. Obsérvese que un modelo VAR corresponde al primero de los casos mencionados, por lo que estos modelos pueden estimarse eficientemente utilizando para cada ecuación estimadores mínimo cuadráticos ordinarios.

Un tipo especial de modelo estructural es aquél en que se puede dar dependencia contemporánea, pero ésta es de forma tal que no hay realimentación. Esto implica que las ecuaciones del modelo se puede ordenar de forma tal que, parafraseando a Fisher (1965) página 609, una variable endógena de orden superior nunca influye en una variable endógena de orden inferior, ni contemporáneamente ni desfasada en el tiempo. A estos modelos se les denomina *modelos recursivos*.

En la literatura econométrica los modelos simultáneos fueron tratados inicialmente por Haavelmo (1943) y con posterioridad en el seno de la comisión Cowles se abordaron los problemas de identificación y estimación de dichos modelos; sus resultados se publicaron en las famosas monografías diez (Koopmans, ed., 1950) y catorce (Hood y Koopmans, eds., 1953) de dicha comisión.

Poco después de aparecer la citada monografía catorce de la comisión Cowles surgen, como un enfoque alternativo a los modelos estructurales simultáneos, los modelos recursivos, propuestos por Wold (1954, 1960 y 1964). La tesis de Wold consiste en que en un modelo la causalidad debe estar encadenada, ya que en el mundo real una variable no puede ser al mismo tiempo causa y efecto de otra.

Sin embargo, y a pesar de los argumentos de tipo lógico en favor de los modelos recursivos, en Economía se da el problema de que en la inmensa mayoría de los casos las variables que entran en un sistema no se pueden observar a intervalos de tiempo tan pequeños como para que la recursividad sea un concepto operativo. Así, los modelos macroeconométricos basados en datos de las contabilidades nacionales, que suelen construirse con periodicidad anual o como mucho trimestral, no pueden formularse en términos recursivos. Sí es cierto que cuanto más se desagrega en el tiempo un modelo, menor es el problema de simultaneidad y más admisible resulta la hipótesis de recursividad; en contrapartida la complejidad dinámica se incrementa considerablemente.

Las restricciones que el modelo recursivo impone sobre el modelo SEM de la sección 3.4 y sus implicaciones sobre la estimación se discuten en Espasa (1973)[26]. Dichas restricciones pueden describirse de la siguiente forma: sea el modelo SEM,

$$B(L)\mathbf{Y}_t + C(L)\mathbf{Z}_t = \mathbf{u}_t,$$

en donde $\mathbf{u}_t$ viene generado por un proceso estacionario, no necesariamente ruido blanco, tal que

$$\begin{aligned} &E(\mathbf{u}_t\mathbf{u}_t') = \Sigma \\ &E(\mathbf{u}_t\mathbf{u}_{t-j}') = V_j, \qquad j>0. \end{aligned} \tag{3.6.1}$$

Las condiciones de recursividad son:

R.1) $B(L) = B_0 + B_1 L + \cdots + B_r L^r$ es una matriz triangular inferior, lo que implica que lo son B_0 y todas las B_j, $j=1, \ldots, r$.

R.2) Σ es una matriz diagonal, y

R.3) V_j es una matriz triangular inferior para cualquier $j>0$ y $B_1, B_2, \ldots, B_r$, o alternativamente todas las matrices $V_j (j>0)$, tienen ceros en todos los elementos de la diagonal principal. Esta restricción obviamente se cumple siempre que $\mathbf{u}$ sea ruido blanco.

Un ejemplo de modelo recursivo con residuos ruido blanco lo constituye el sistema:

$$\begin{pmatrix} b_{11}(L) & 0 & 0 \\ b_{21}(L) & b_{22}(L) & 0 \\ b_{31}(L) & b_{32}(L) & b_{33}(L) \end{pmatrix} \begin{pmatrix} Y_{1t} \\ Y_{2t} \\ Y_{3t} \end{pmatrix} + \begin{pmatrix} C_1(L) \\ C_2(L) \\ C_3(L) \end{pmatrix} \mathbf{Z}_t = \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix} \tag{3.6.2}$$

$$E(\mathbf{u}_t\mathbf{u}_t') = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2).$$

El rasgo distintivo de este modelo es que las variables endógenas pueden ordenarse de manera que la formulación (3.6.2) es posible, con lo que la causalidad existente es siempre de las variables Y_i a las variables Y_j con $j>i$. Esto se aprecia más claramente si se desarrolla la expresión matricial (3.6.2), ya que se obtiene

$$\begin{aligned} &b_{11}(L)Y_{1t} + C_1(L)\mathbf{Z}_t = u_{1t} \\ &b_{21}(L)Y_{1t} + b_{22}(L)Y_{2t} + C_2(L)\mathbf{Z}_t = u_{2t} \\ &b_{31}(L)Y_{1t} + b_{32}(L)Y_{2t} + b_{33}(L)Y_{3t} + C_3(L)\mathbf{Z}_t = u_{3t} \end{aligned}$$

y se comprueba que la parte sistemática de Y_1 sólo depende de su propio pasado y de las variables Z; la de Y_2 de su pasado, de Y_1 y de Z; y la de Y_3 del presente y pasado de todas las variables del sistema. Por lo tanto Y_1 afecta a Y_2, pero no viene afectada por ésta, y lo mismo se podría decir para Y_1 y Y_3 y para Y_2 y Y_3.

Hay dos aspectos que merecen una consideración especial:

1) Es necesario no sólo que $B(L)$ sea triangular sino también que Σ sea diagonal; de no ser así la correlación contemporánea entre las diferentes innovaciones rompe la recursividad.

2) Normalmente se interpreta recursividad como ausencia de simultaneidad: esto se debe a que las presentaciones tradicionales se hacen en un marco estático. Pero la recursividad es más restrictiva que la no simultaneidad: en el modelo

$$\begin{pmatrix} 1 & b_{12}L \\ b_{21}L & 1 \end{pmatrix} \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} + C(L)\mathbf{Z}_t = \mathbf{u}_t$$

$$\mathbf{u}_t \sim \text{Niid}\ (0, \operatorname{diag}(\sigma_1^2, \sigma_2^2))$$

[26] En la transcripción de este artículo se produjeron bastantes erratas, por lo que el lector interesado debe consultar la fe de erratas a dicho artículo publicada en Cuadernos de Economía, 1973, núm. 2.

no hay ningún tipo de relación contemporánea entre Y_1 y Y_2, y sin embargo no es un modelo recursivo, ya que el valor de Y_1 en t afecta al valor de Y_2 en $t+1$, que a su vez entra en la determinación de Y_1 en $t+2$.

En definitiva, un modelo estructural dinámico recursivo se caracteriza por el hecho de que siempre es posible ordenarlo de tal forma que la variable endógena de la primera ecuación, Y_1, dependa sólo de su pasado, de las variables exógenas y de los valores presente y pasados de su propia innovación, y además su innovación no está relacionada con ninguna otra innovación, presente o pasada, que aparezca en el modelo. Esto implica que la variable Y_1 puede estudiarse de manera independiente del resto de variables endógenas, es decir, puede ser modelizada de forma consistente y eficiente especificando y estimando únicamente la ecuación que a ella se refiere, e ignorando las demás ecuaciones.

La segunda ecuación del modelo, referida a la variable Y_2, ha de reflejar que ésta depende de su propio pasado, del presente y pasado de Y_1, de las variables exógenas, de los valores presente y pasados de su propia innovación y de las innovaciones pasadas de Y_1. A esto hay que añadir que su innovación a_2 ha de ser independiente de cualquier otra innovación del modelo. Si esto se cumple, entonces para construir un modelo para Y_2 basta con considerar únicamente su ecuación; a Y_1 se le trata como si fuese una variable exógena similar a las contenidas en Z —y de hecho lo es.

En general, en la ecuación j-ésima correspondiente a la variable Y_j han de entrar únicamente:

— los valores pasados de Y_j;
— los valores presentes y pasados de $Y_1, Y_2, \ldots, Y_{j-1}$;
— las variables exógenas Z;
— los valores presente y pasados de a_j;
— los valores pasados de $a_1, a_2, \ldots, a_{j-1}$;

y además

— la innovación a_j ha de estar incorrelacionada con cualquier otra innovación del modelo.

Si estos requisitos se cumplen, el modelo para Y_{jt} se obtiene considerando exclusivamente su ecuación de comportamiento, ya que en ella $Y_1, Y_2, \ldots, Y_{j-1}$ son variables exógenas.

La principal consecuencia de la recursividad es que en un modelo recursivo la estimación conjunta de todo el sistema proporciona exactamente el mismo resultado que si se estima cada ecuación por separado. Esta propiedad es consecuencia del hecho de que en un sistema recursivo la causalidad siempre es unidireccional: por lo tanto

si el interés se centra en una variable Y_j, no es necesario estimar las ecuaciones correspondientes a $Y_1, Y_2, \ldots, Y_{j-1}$, ya que éstas son variables fuertemente exógenas en la determinación de Y_j; y tampoco es necesario incluir las ecuaciones de $Y_{j+1}, \ldots, Y_n$, pues estas variables no intervienen en la determinación de Y_j.

A priori resulta difícil saber si las matrices de coeficientes y la matriz de varianzas de las innovaciones cumplen las restricciones necesarias para que el modelo sea recursivo. En la práctica la recursividad suele ser un resultado que se obtiene a posteriori: a raíz de estimar un modelo multiecuacional, se observa que los resultados apuntan hacia un modelo recursivo, en vez del modelo con causalidad multidireccional que se había planteado en un principio.

Aunque hasta ahora se ha desarrollado la recursividad en el marco de modelos SEM, esta propiedad no se da exclusivamente en los modelos estructurales, ya que también los modelos multivariantes de series temporales pueden presentarla. Sea un modelo VARMA dado por

$$\Phi(L)\,\mathbf{Y}_t = \Theta(L)\,\mathbf{a}_t \qquad (3.6.3)$$
$$E(\mathbf{a}_t\mathbf{a}_t') = \Sigma$$

Este modelo es recursivo si se cumplen las dos siguientes restricciones:

1) $\Phi(L)$ y $\Theta(L)$ son matrices triangulares;
2) Σ es diagonal.

En este caso (3.6.3) se puede expresar como

$$\begin{aligned}
&\phi_{11}(L)Y_{1t} = \theta_{11}(L)a_{1t}\\
&\phi_{21}(L)Y_{1t} + \phi_{22}(L)Y_{2t} = \theta_{21}(L)a_{1t} + \theta_{22}(L)a_{2t}\\
&\phi_{n1}(L)Y_{1t} + \phi_{n2}(L)Y_{2t} + \cdots + \phi_{nn}(L)Y_{nt} = \theta_{n1}(L)a_{1t} +\\
&\qquad + \theta_{n2}(L)a_{2t} + \cdots + \theta_{nn}(L)a_{nt}
\end{aligned}$$

$$\Sigma = \text{diag}\ (\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$$

es decir, como una sucesión encadenada de modelos uniecuacionales que se estiman eficientemente considerando cada caso de forma individual.

Como ya se ha señalado los sistemas económicos tal y como se observan no pueden, en general, representarse como modelos recursivos. Sin embargo, sí resulta más verosímil que se dé un supuesto alternativo menos restrictivo, propuesto en Fisher (1965), y que se ha denominado *recursividad por bloques*. En un modelo con tal propiedad sus matrices $B_j (j=0, \ldots, r)$, Σ y $V_j (j>0)$ se pueden estructurar por bloques de ecuaciones, de modo que las restricciones de triangularidad, diagonalidad y/o elementos cero en las diagonales principales exigidas en las restricciones R.1 a R.3 se cumplen por bloques.

En otras palabras, un modelo recursivo por bloques es aquél en que cada bloque de ecuaciones individualmente considerado constituye un modelo simultáneo, pero por grupos de ecuaciones la estructura es recursiva: esto implica que para fines de estimación un bloque cualquiera puede tratarse aisladamente sin pérdida de eficiencia. Cuando se da este tipo de recursividad la dimensión de la simultaneidad se reduce enormemente, lo que simplifica bastante la labor del analista.

3.7. El planteamiento uniecuacional de la modelización econométrica

3.7.1. Planteamiento del tema. Procedimientos de Información Limitada y Variables Instrumentales

El análisis de los modelos recursivos ha vuelto a traer a un primer plano la modelización de los fenómenos económicos basada en considerar solamente la ecuación referida a la variable de interés. A lo largo de las secciones anteriores se han encontrado dos situaciones extremas.

1) Una en la que se da causalidad multidireccional entre las variables económicas, por lo que se requiere la modelización conjunta de todo el sistema. Ciertamente esto es siempre necesario si el modelo tiene un componente de medias móviles sobre todo el vector de innovaciones, o si se persiguen fines predictivos.

Cuando lo que se pretende es solamente estimar los parámetros del modelo y éste no tiene parte de medias móviles, conviene distinguir los modelos de forma reducida, por ejemplo VARX, en los que la multidireccionalidad contemporánea se recoge exclusivamente en la matriz de varianzas y covarianzas de las innovaciones, y los modelos estructurales, en los que tal interrelación aparece también en la parte funcional del modelo.

En el primer caso si se estima una ecuación ignorando el resto se obtienen unos resultados que, excepto en los modelos en los que todas las ecuaciones tienen exactamente las mismas variables explicativas —como ocurre en los VAR— son consistentes pero ineficientes; en el segundo, los resultados son también inconsistentes.

2) En el epígrafe 3.6 se estudió una clase especial de modelos, los modelos recursivos, en los que, debido a sus características

particulares, la estimación uniecuacional proporciona exactamente los mismos resultados a los que conduciría el estudio del sistema en conjunto.

Es de destacar que el analista no tiene ningún tipo de control sobre la recursividad del sistema: se encuentra con que las variables que analiza, tal y como se observan, tienen la característica de recursividad, y por lo tanto basta, para fines de estimación, con que centre su esfuerzo en la ecuación que le interesa, o se encuentra con que el sistema de ecuaciones para esas variables no es recursivo y se ve obligado a tratar dicho sistema en conjunto[27].

Desgraciadamente, en la práctica es corriente encontrarse con alguno(s) de los siguientes problemas:

1) Existe incertidumbre sobre la dimensión multivariante del estudio.
2) No se dispone de información teórica suficiente y fiable para todas las ecuaciones del modelo.
3) Algunas ecuaciones no se pueden formular, especialmente por dificultades a la hora de disponer de datos sobre todas las variables explicativas.
4) El coste del análisis multiecuacional no se justifica en función de sus beneficios esperados.

De ahí que al analista aplicado le sea de especial interés poder llegar a una situación intermedia entre las dos que se han considerado hasta ahora. Esto le lleva a preguntarse si existe algún tipo de planteamiento parcial, consistente en fijarse en un número reducido de ecuaciones ignorando el resto, que proporcione resultados aceptables; y en caso afirmativo, cómo se puede medir el coste asociado a dicho planteamiento en relación al enfoque multiecuacional general. Este planteamiento, que en la literatura se denomina de *información limitada*, tiene su versión extrema cuando el investigador sólo se fija en una ecuación ignorando las demás, dando origen al *planteamiento uniecuacional*.

El objetivo último de este enfoque es que el analista pueda escoger entre realizar un estudio complejo que optimice la información disponible, o un análisis más sencillo, donde la sencillez tendrá un coste determinado que será aceptable o no en función del tipo de resultados requeridos.

Los estimadores que se manejan en el análisis estadístico-econométrico de series temporales se valoran en función de dos características, la *consistencia* y la *eficiencia*. El tratamiento multiecuacional del

[27] Y como se discutió en el epígrafe anterior la característica de recursividad normalmente se pone de manifiesto a posteriori, una vez que se ha estimado el sistema.

sistema garantiza la obtención de estimadores consistentes y eficientes; en cambio, si el sistema no es recursivo, el planteamiento uniecuacional dará lugar, salvo las excepciones señaladas, a estimadores ineficientes e incluso, según el tipo de modelos, inconsistentes.

El problema de inconsistencia que aparece con los modelos estructurales no se puede, en el caso de modelos estructurales sobreidentificados, resolver a partir de estimadores consistentes de la forma reducida no restringida, pues para tales modelos de dicha forma reducida no se pueden obtener los parámetros estructurales. Por ello hay que emplear una técnica de estimación especial, la llamada *estimación por variables instrumentales*, que fue introducida por Reiersøl (1941, 1945), aplicada y desarrollada por Geary (1949) y Durbin (1954), y formalizada en su enfoque actual por Sargan (1958). Para ilustrar su funcionamiento, sea el siguiente sistema estático de dos ecuaciones:

$$Y_{1t}=\alpha_0+\alpha_1 Z_{1t}+\alpha_2 Y_{2t}+a_{1t} \tag{3.7.1}$$
$$Y_{2t}=\beta_0+\beta_1 Y_{1t}+\beta_2 Z_{2t}+\beta_3 Z_{3t}+a_{2t} \tag{3.7.2}$$

en donde si $\mathbf{a}_t=(a_{1t}a_{2t})'$,

$$E(\mathbf{a}_t\mathbf{a}_t')=\begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}.$$

Si la variable de interés es Y_{1t} y se estima únicamente la primera ecuación mediante mínimos cuadrados ordinarios, todos los estimadores serán sesgados e inconsistentes, ya que se incumple el supuesto clásico de incorrelación entre regresores y perturbación; a este sesgo se le denomina sesgo de simultaneidad y no desaparece asintóticamente.

Para evitar este problema, el método de estimación recomendado consiste en seleccionar un conjunto de variables, llamadas instrumentos, caracterizadas por: a) estar lo más correlacionadas posible con Y_2, y b) estar incorrelacionadas con la perturbación a_1.

La ecuación referida a Y_{1t}, (3.7.1), se puede expresar de forma matricial como

$$Y_1=X_1\alpha+a_1,$$

donde $\alpha=(\alpha_0\ \alpha_1\ \alpha_2)'$ y X_1 es la matriz $T\times 3$ de las T observaciones de las variables explicativas, incluida la variable artificial correspondiente al término constante. En este contexto el estimador mínimo cuadrático ordinario, $\hat{\alpha}=(X_1'X_1)^{-1}X_1'Y_1$, es inconsistente.

Si en X_1 sustituimos el vector de observaciones de la variable Y_2 por el vector de observaciones de una variable instrumental, y denotamos W_1 a la matriz resultante, el estimador por variables instrumentales se define como

$$\tilde{\alpha}=(W_1'X_1)^{-1}W_1'\,Y_1, \tag{3.7.4}$$

que estima consistentemente el vector original de parámetros dadas las propiedades a) y b) de la variable que actúa de instrumento.

Es fácil demostrar que la eficiencia de $\tilde{\alpha}$ aumenta con la correlación, en valor absoluto, entre Y_2 y su variable instrumental. Por eso, cuando se tiene más de un instrumento Sargan (1958) propone emplear lo que se denomina el Estimador Generalizado de Variables Instrumentales (GIVE, Generalized Instrumental Variable Estimator). Tal estimador consiste en utilizar como variable instrumental una combinación lineal de todos los instrumentos que sea óptima, en el sentido de que maximice la correlación entre la variable instrumental resultante e Y_2.

Las combinaciones lineales de instrumentos se pueden denominar óptimas respecto a dos conjuntos de información diferentes. Uno, de información limitada, que da lugar a los Estimadores de Variables Instrumentales de Información Limitada (LIVE, Limited Information Instrumental Variable Estimator). La información limitada implica que del modelo SEM sólo se conoce la ecuación bajo estudio y el conjunto de variables que entran en el sistema.

Otro, de información completa, que da lugar a los Estimadores de Variables Instrumentales de Información Completa (FIVE, Full Information Instrumental Variable Estimator). La información completa implica que se conoce todo el modelo SEM, por lo que no son de interés cuando la pretensión es realizar un tratamiento uniecuacional. Para una exposición detallada de los estimadores LIVE y FIVE véase Harvey (1981a), capítulo noveno.

Volviendo al ejemplo anterior se puede ilustrar el problema de la estimación calculando el estimador mínimo cuadrático bietápico, que es el estimador LIVE más conocido. En ese sistema los instrumentos disponibles para Y_2 son todas las variables predeterminadas que entran en (3.7.1) y (3.7.2): la constante, Z_1, Z_2 y Z_3; ello lleva a estimar la relación

$$Y_{2t}=\delta_0+\delta_1 Z_{1t}+\delta_2 Z_{2t}+\delta_3 Z_{3t}+e_t \tag{3.7.5}$$

de donde se calculan los valores ajustados $\hat{Y}_{2t}$.

A continuación se estima, también por mínimos cuadrados ordinarios, la primera ecuación del sistema, (3.7.1), sustituyendo Y_{2t} por $\hat{Y}_{2t}$.

Los estimadores $\tilde{\alpha}_0$, $\tilde{\alpha}_1$ y $\tilde{\alpha}_2$ resultantes son estimadores por variables instrumentales en los que $\hat{Y}_2$ actúa como instrumento de Y_2. Obsérvese que no plantean problemas de sesgo, ya que ahora todas las variables explicativas de Y_1 están incorrelacionadas con la perturbación: como por hipótesis la correlación de Z_{1t}, Z_{2t} o Z_{3t} con a_{1t} es cero, cualquier combinación lineal de las mismas —y en particular $\hat{\delta}_0+\hat{\delta}_1 Z_{1t}+\hat{\delta}_2 Z_{2t}+\hat{\delta}_3 Z_{3t}=\hat{Y}_{2t}$— también estará incorrelacionada con la innovación de Y_1.

Para ver que este sistema de estimación bietápico coincide con la aplicación de la ecuación (3.7.4) obsérvese que la regresión (3.7.5) se puede escribir de forma vectorial como

$$Y_2 = Z\delta + e, \tag{3.7.6}$$

en donde $\delta' = (\delta_0\ \delta_1\ \delta_2\ \delta_3)'$ y Z es la matriz de observaciones correspondiente al término constante y a Z_1, Z_2 y Z_3; es decir,

$$Z = (1 \quad Z_1 \quad Z_2 \quad Z_3).$$

Para lo que sigue es conveniente recoger en Z^* las variables de la matriz Z que aparecen en (3.7.1): la constante y Z_1. Así,

$$Z = (Z^* \quad Z_2 \quad Z_3),$$

donde

$$Z^* = (1 \quad Z_1).$$

Con ello el valor de $\hat{Y}_2$ que se obtiene en la primera etapa es

$$\hat{Y}_2 = Z(Z'Z)^{-1}Z'Y_2.$$

En la segunda etapa se sustituye Y_2 por $\hat{Y}_2$ en (3.7.1) con lo que se tiene

$$Y_1 = (Z^* \mid \hat{Y}_2)\alpha + a_1^* \tag{3.7.7}$$

donde α es el vector de parámetros que aparecen en (3.7.1). Aplicando mínimos cuadrados ordinarios a (3.7.7) se tiene:

$$\begin{aligned}\tilde{\alpha} &= [(Z^* \mid Z(Z'Z)^{-1}Z'Y_2)'\ (Z^* \mid Z(Z'Z)^{-1}Z'Y_2)]^{-1} \times \\ &\times [(Z^* \mid Z(Z'Z)^{-1}Z'Y_2)'\ Y_1]\end{aligned} \tag{3.7.8}$$

Teniendo en cuenta que

$$Z^* = Z\begin{pmatrix} I \\ 0 \end{pmatrix},$$

se puede escribir

$$Z^* = Z(Z'Z)^{-1}Z'Z\begin{pmatrix} I \\ 0 \end{pmatrix} = Z(Z'Z)^{-1}Z'Z^* \tag{3.7.9}$$

y

$$(Z^* \mid Z(Z'Z)^{-1}Z'Y_2) = Z(Z'Z)^{-1}Z'(Z^* \mid Y_2) \tag{3.7.10}$$

Si ahora se utiliza (3.7.10) en (3.7.8) y se cancelan términos $Z'Z$ y $(Z'Z)^{-1}$, se tiene que

$$\tilde{\alpha} = \left[\begin{pmatrix} Z^{*\prime} \\ Y_2' \end{pmatrix} Z(Z'Z)^{-1}Z'(Z^* \mid Y_2)\right]^{-1} \left[\begin{pmatrix} Z^{*\prime} \\ Y_2' \end{pmatrix} Z(Z'Z)^{-1}Z'Y_1\right] \tag{3.7.11}$$

Alternativamente se puede obtener $\tilde{\alpha}$ aplicando (3.7.4) directamente. Para ello hay que observar que las variables instrumentales en W_1 son tal que

$$W_1 = (Z^* \mid \hat{Y}_2) = Z(Z'Z)^{-1}Z'(Z^* \mid Y_2) \tag{3.7.12}$$

y

$$X_1 = (Z^* \mid Y_2)$$

con lo que la aplicación de (3.7.4) concluye también en (3.7.11).

Este ejemplo sirve para demostrar que teniendo varios instrumentos recogidos en la matriz de observaciones Z la combinación lineal óptima de los mismos, desde un enfoque de información limitada, es la recogida en (3.7.12), que consiste en dejar a las variables Z^* que aparezcan en (3.7.1) como instrumentos de sí mismas y sustituir Y_2 por su ajuste —$\hat{Y}_2$— en una regresión sobre todos los instrumentos disponibles.

El ejemplo anterior sirve para señalar también que cuantas más restricciones se hayan empleado en la identificación de una ecuación del modelo SEM, más variables instrumentales se disponen para su estimación. Así, en la ecuación (3.7.1) la condición de orden mencionada anteriormente se cumple, ya que habiendo una variable endógena como explicativa, hay dos variables exógenas que no entran en