dicha ecuación. Así pues, la sobreidentificación en este ejemplo indica que para sustituir Y_2 en (3.7.1) tenemos más de un instrumento: Z_2 y Z_3[28]. Es decir, las restricciones que identifican el modelo garantizan la existencia de instrumentos suficientes en la estimación, y cuantas más restricciones, más instrumentos y, por lo tanto, una estimación más eficiente. Así, los resultados de la Teoría Económica que sirven para identificar el modelo pueden verse también como la aportación de la teoría para obtener los máximos instrumentos posibles.

El método de variables instrumentales, en sus múltiples variantes, ha recibido mucha atención en la literatura, ya que su aplicación tampoco está libre de dificultades. No obstante, para los propósitos de este libro su característica principal es la de permitir un tratamiento uniecuacional de variables que pertenecen a un sistema; el lector interesado en profundizar en su estudio puede acudir a la monografía de Bowden y Turkington (1984). En Hendry (1976) se deriva un sistema de ecuaciones normales del que se obtienen prácticamente todos los tipos de estimadores conocidos para el modelo SEM con residuos ruido blanco o autorregresivos de primer orden. En Espasa (1975) esto se generaliza, mediante métodos espectrales, para sistemas con cualquier tipo de residuos con la única restricción de que sean estacionarios.

La elección de los instrumentos con los que obtener las variables instrumentales es una labor ardua ya que se suele disponer de poca información al respecto. Los instrumentos han de cumplir las propiedades a) y b) señaladas anteriormente, y en este sentido el analista encontrará en la Teoría Económica una orientación importante[29].

Tal y como se acaba de comentar, esta elección está plenamente vinculada a las restricciones con las que se ha especificado el modelo SEM. Así, ciñéndonos al caso de restricciones cero en los parámetros de las matrices estructurales en un modelo estático, el número de instrumentos realmente disponibles para estimar los parámetros correspondientes a las variables endógenas que entran como explicativas en una ecuación dada, equivale exactamente al número de restricciones cero que afectan a las variables exógenas en esa ecuación. Obsérvese, por tanto, que el tener suficiente restricciones para que la ecuación esté identificada equivale a tener suficientes instrumentos

[28] La constante y Z_1 no son realmente instrumentos, ya que estas variables aparecen como regresores en (3.7.1). En efecto, el problema planteado en la estimación de (3.7.1) se podría abordar, de acuerdo con el procedimiento de regresión por etapas, como un modelo entre Y_1^* y Y_2^* donde las variables con asterisco son los residuos de las regresiones de Y_1 y Y_2 sobre la constante y Z_1. En tal caso se ve con claridad que los instrumentos realmente disponibles son Z_2 y Z_3, que no aparecen en (3.7.1).

[29] Un excelente trabajo sobre la problemática y posibles soluciones de la selección de instrumentos es Fisher (1965).

para aplicar el procedimiento de variables instrumentales; además, la sobreidentificación implica más instrumentos que los estrictamente necesarios, lo que normalmente conducirá a obtener una variable instrumental más correlacionada con la variable endógena que instrumentaliza y, por tanto, a proporcionar un estimador más eficiente.

En cuanto a la eficiencia de los estimadores LIVE, no hay ninguna manera de compensar la pérdida de información que, normalmente, resulta del tratamiento uniecuacional en modelos que no son recursivos. El no tener en cuenta las interrelaciones entre todas las variables endógenas del modelo forzosamente conlleva una visión limitada y parcial del sistema económico, lo que se traduce, en general, en una pérdida de precisión de los estimadores resultantes.

En ese sentido el coste de un planteamiento uniecuacional se puede medir como la pérdida de precisión, en términos de varianza de la innovación, que se produce al pasar de un tratamiento multiecuacional al uniecuacional.

Al analista le corresponde así la decisión de intercambiar eficiencia por simplificación del análisis: si prefiere la eficiencia, necesariamente deberá plantearse un modelo multivariante; si por el contrario espera que la ganancia en eficiencia no compense la mayor complejidad del modelo multivariante, puede decidirse por la solución uniecuacional.

3.7.2. Conclusión

Como conclusión a lo tratado hasta este momento se puede decir que en la formulación de un modelo econométrico con n variables es importante explicitar las variables exógenas, pues ello permite reducir la dimensión multiecuacional. Si de las n variables que se consideraron al principio k son exógenas, entonces basta con plantear el subsistema formado por las $n-k$ ecuaciones correspondientes a las variables endógenas.

Este procedimiento de condicionar respecto a las variables exógenas se realiza principalmente a partir de los dictámenes de la Teoría Económica, que establece cuáles son los parámetros de interés. Por tanto, este condicionamiento no es algo plenamente basado en los datos, por lo que autores como Sims, véase Sims (1980), lo ponen en entredicho y sugieren operar a través de modelos VAR sin establecer a priori si hay variables exógenas.

Cuando el análisis se realiza con fines predictivos la reducción de la dimensionalidad, pasando de n a $n-k$ ecuaciones, sólo se logra si las variables exógenas son fuertemente exógenas. Bien entendido que aun en este caso habrá que estimar el sistema de k ecuaciones respecto a las variables exógenas, para generar predicciones sobre las mismas que entrarán como inputs en la predicción realizada con el sistema de $n-k$ ecuaciones correspondientes a las variables endógenas.

La ventaja habrá consistido en operar con dos sistemas de menor dimensión, en vez de con un gran sistema de dimensión n. En la práctica, la predicción de las variables exógenas se realiza mediante modelos univariantes, con lo que su sistema de k ecuaciones conjuntas se sustituye por k modelos univariantes.

La predicción obliga a operar con todo el sistema de $n-k$ ecuaciones, ya que aunque sólo se esté interesado en la variable Y_h, si ésta depende de otras variables endógenas es necesario utilizar todo el sistema en el que se determinan las variables endógenas ($n-k$ ecuaciones) para obtener una predicción óptima de Y_h.

Sin embargo, si el objetivo es estimar los parámetros de un número reducido de ecuaciones, que con frecuencia puede ser una única ecuación, cabe adoptar un enfoque de información limitada, que incluso puede llegar a ser uniecuacional. Este enfoque comportará, en general, una pérdida de eficiencia que es absolutamente inevitable; y además planteará problemas de consistencia, que tendrán que ser resueltos mediante la aplicación de procedimientos de variables instrumentales. No obstante, las pérdidas de eficiencia pueden muy bien ser aceptables dado que gracias al enfoque de información limitada habrá sido posible estimar una relación económica de interés inmersa en un sistema multiecuacional que puede ser muy complejo de formular.

En resumen, las características de exogeneidad que un sistema multiecuacional concreto posee constituyen una propiedad específica del mismo, que permite siempre operar con el subsistema de las $n-k$ ecuaciones correspondientes a las variables endógenas. En tanto en cuanto tales características de exogeneidad son correctas no generan distorsiones en las conclusiones obtenidas a partir del sistema reducido de $n-k$ ecuaciones. Adicionalmente el analista puede plantearse aislar su ecuación (o ecuaciones) de interés del sistema reducido. Las posibilidades y costes de realizar tal aislamiento depende de la finalidad del análisis: predicción o estimación. Pagando como coste una pérdida de eficiencia el análisis uniecuacional es, en general, posible para fines de estimación de los parámetros. Para fines predictivos o de simulación el aislamiento de una ecuación sólo es posible si sus variables explicativas son fuertemente exógenas y superexógenas, respectivamente.

En este libro la predicción es uno de los objetivos, por lo que el estudio de modelos uniecuacionales que se realiza en las secciones siguientes supondrá que las variables explicativas son fuertemente exógenas.

Cuando esto no se cumpla la predicción tendrá que realizarse con

modelos multiecuacionales con el consiguiente aumento de complejidad que ello comporta. El lector interesado en el tema puede remitirse al libro de Lütkepohl (1991). En cualquier caso las secciones que siguen estudian con detalle la relación dinámica entre variables económicas en un marco sencillo —uniecuacional— que puede ser muy útil para entender posteriormente las relaciones dinámicas en un modelo de mayor dimensión. También se aborda con cierto detalle el problema de cointegración en su mayor nivel de sencillez.

Habiendo insistido en los posibles costes que entraña el análisis uniecuacional, conviene señalar una ventaja importante del mismo. En efecto, el enfoque multiecuacional tiene las propiedades de eficiencia comentadas, siempre que el analista sea capaz de formular adecuadamente todo el sistema; en caso contrario, los errores cometidos en la especificación de una ecuación se trasladan, en general, a las restantes. Por tanto, cuando sobre la ecuación de la variable de interés Y_h se tiene bastante seguridad sobre su formulación, pero en la formulación de las restantes ecuaciones que completan el sistema de variables endógenas que intervienen en la explicación de Y_h se tienen grandes incertidumbres o incluso ignorancia sobre cómo proceder, el enfoque uniecuacional es más seguro, pues para su estimación se requiere simplemente ser capaz de formular un conjunto válido de instrumentos. En tales casos el enfoque uniecuacional tiene una solidez y firmeza, frente a errores de especificación en las otras ecuaciones, que el enfoque multiecuacional carece y hace que aquél sea preferible a éste.

3.8. El modelo lineal general uniecuacional: su formulación; variables explicativas y elemento residual

Una vez conocidas las ventajas e inconvenientes de la modelización uniecuacional de los fenómenos económicos, el resto del capítulo se va a centrar en este tratamiento uniecuacional, bajo el supuesto de que las variables explicativas son fuertemente exógenas. El acento se va a poner en la explotación de la información sobre la relación dinámica entre las variables, tal y como la refleja el modelo econométrico uniecuacional.

Para hacer la exposición más general se abandonará el supuesto de estacionariedad, y se permitirá que las variables que entran en el análisis sigan procesos integrados. Tal y como se demuestra en Phillips (1991), véase también Dolado (1992a, b), la hipótesis de exogeneidad introducida garantiza que los estimadores uniecuacionales usuales tienen propiedades asintóticas óptimas y por lo tanto los contrastes de hipótesis pueden realizarse por los procedimientos habituales. Además, se usará la notación común en modelos uniecuacionales, y desde ahora se denotará por Y_t a la variable a explicar y por X_{jt}, $j=1, 2, \ldots, m$, al conjunto formado por las m variables explicativas.

En ocasiones, y siguiendo la terminología habitual en las funciones de transferencia, a las variables explicativas X_j se les llamará los *inputs* de la función, y a Y se le denominará el *output* de la función de transferencia.

La forma general de representar el modelo lineal uniecuacional es en términos de retardos racionales distribuidos o funciones de transferencia[30]:

$$Y_t = \sum_{j=1}^{m} \frac{w_{0,j} + w_{1,j}L + \cdots + w_{s,j}L^{s,j}}{1 - \delta_{1,j}L - \cdots - \delta_{r,j}L^{r,j}} X_{jt} + \frac{\theta_q(L)}{\phi_{p+q}(L)} a_t =$$

$$= \sum_{j=1}^{m} v_{\infty,j}(L) X_{jt} + \frac{\theta_q(L)}{\phi_{p+q}(L)} a_t, \quad (3.8.1)$$

donde

$$\frac{w_{0,j} + w_{1,j}L + \cdots + w_{s,j}L^{s,j}}{1 - \delta_{1,j}L - \cdots - \delta_{r,j}L^{r,j}} = \frac{w_{s,j}(L)}{\delta_{r,j}(L)}$$

es un filtro racional con el que se aproxima el efecto dinámico de la variable explicativa X_j sobre Y, y en el que los órdenes de los polinomios pueden ser distintos según las variables X_j; y

$$\frac{\theta_q(L)}{\phi_{p+d}(L)}$$

es un filtro que aproxima la estructura dinámica general de la perturbación o término residual.

Esta especificación permite reflejar cualquier clase de dependencia dinámica que se pueda dar entre variables económicas, que en algunos casos puede llegar a ser bastante compleja. Así, la función de respuesta de Y a cambios de tipo impulso en X_j[31],

$$v_{\infty,j}(L) = v_{0,j} + v_{1,j}L + v_{2,j}L^2 + \cdots,$$

[30] En la sección 3.10 se verán esquemas alternativos para representar la forma general del modelo lineal uniecuacional.

[31] Esta función será objeto de un estudio detallado en el próximo epígrafe.

puede ser de orden infinito, aunque convergente, en cuyo caso siempre se podrá aproximar mediante un cociente de polinomios, $w_{s,j}(L)/\delta_{r,j}(L)$, de órdenes finitos, en donde las raíces de $\delta_{r,j}(L)$ están fuera del círculo unitario.

Esto implica que una variación transitoria en la variable explicativa no tiene un efecto permanente en el nivel de la variable dependiente, ya que este efecto tiende a cero a medida que nos alejamos del momento en que ocurrió la variación transitoria de X_j. Dado que tal efecto sobre Y tiende a anularse, siempre se puede señalar un momento en el tiempo, $t^* + m_j$, a partir del cual el efecto de una variación transitoria de X_j ocurrida en t^* sobre Y es despreciable, y por lo tanto puede considerarse prácticamente nulo.

Por ejemplo, supóngase que una variable explicativa presenta la siguiente evolución:

$$X_{jt} = \begin{cases} X_{jt}^e & \text{si } t < t^* \\ X_{jt}^e + 1 & \text{si } t = t^* \\ X_{jt}^e & \text{si } t > t^* \end{cases} \quad (3.8.2)$$

es decir, toma siempre el valor X_{jt}^e, que puede ser considerado de equilibrio, excepto en el momento t^* en que pasa a valer una unidad más.

En este caso el efecto del cambio transitorio en X_j sobre la variable endógena puede llegar a ser como el recogido en el gráfico 3.1 (para simplificar la representación se realiza, sin pérdida de generalidad, para $X_{jt}^e = 0$). En él, después de un desfase de b períodos —fase cero de la función de respuesta— se produce la respuesta de Y. Con frecuencia, en este segundo período de la función de respuesta se puede distinguir otras dos fases: fase I, en la que la respuesta, en valor absoluto (ya que el punto A del gráfico puede estar en la zona de valores negativos), va aumentando, y fase II, en la que la respuesta en valor absoluto va disminuyendo hasta anularse.

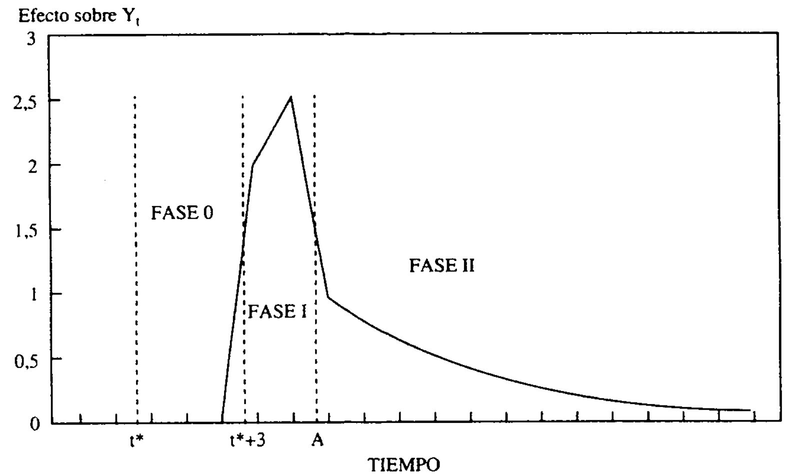
Este esquema puede mostrar distintas variantes: ambas fases, I y II, pueden mostrar su correspondiente evolución creciente o decreciente de forma más compleja de lo que aparece en el gráfico, por ejemplo de forma oscilante. También es posible que no exista un punto A y que las fases I y II se confundan en una única en la que la respuesta oscila sobre cero y que, posteriormente, de modo repentino o paulatino se anula.

Como se aprecia en el ejemplo anterior, lo que caracteriza a la relación entre variables económicas temporales es la estructura dinámica de respuesta, que puede llegar a ser bastante compleja, y que en consecuencia demanda una formulación muy general como la

GRÁFICO 3.1. *Ejemplo de posible respuesta a una variación impulso en una variable explicativa.*

Función de respuesta asociada al filtro

$$2L^3+2{,}5L^4+\frac{L^5}{1-0{,}8L}$$



recogida en (3.8.1). El tipo de relaciones que hay detrás del modelo dinámico general uniecuacional se puede representar esquemáticamente a partir del cuadro 3.2.

Al filtro general dinámico por el que se pasa del valor de la variable explicativa X_{jt} al valor X_{jt}^* con que dicha variable contribuye en la variable a explicar, se le conoce con el nombre de *función de transferencia de X_j sobre Y.*

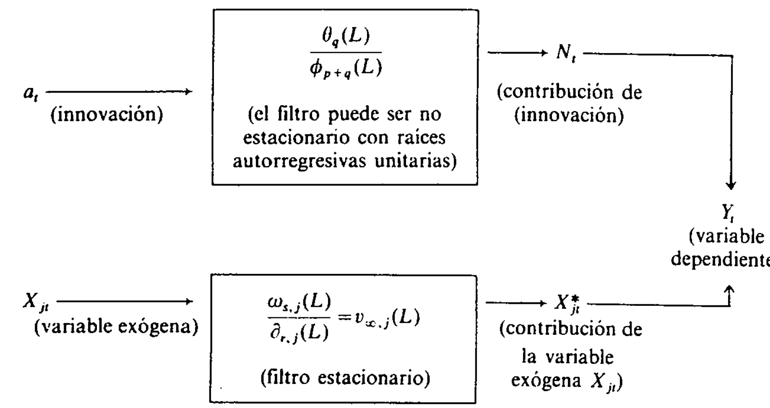
Al conjunto de las funciones de transferencia $w_j(L)/\delta_j(L)$, $j=1, \ldots, m$, se les conoce como la *estructura dinámica de las variables explicativas*[32].

Se ha mencionado anteriormente la contribución de la variable explicativa X_{jt}: por *contribución* se entiende la influencia de la variable explicativa en el valor de la variable endógena:

$$X_{jt}^*=\frac{w_j(L)}{\delta_j(L)}X_{jt}=v_j(L)X_{jt}. \tag{3.8.3}$$

[32] Los subíndices de $w_j(L)$, $\delta_j(L)$ y $v_j(L)$ no se refieren al orden de los polinomios sino sólo a la variable explicativa a la que están asociados; se suprime la información sobre el orden de $w(L)$, $\delta(L)$ y $v(L)$ para no complicar en demasía la notación.

CUADRO 3.2. *El modelo lineal dinámico uniecuacional*



Por otra parte, obsérvese en el cuadro 3.2 que la contribución de las variables explicativas no agota lo que hay de predecible en la variable endógena, sino que puede quedar un elemento residual, denotado por N_t, que siga un proceso ARIMA; es decir

$$N_t=\frac{\theta_q(L)}{\phi_{p+d}(L)}a_t=\hat{N}_t+a_t \tag{3.8.4}$$

en el que el componente $\hat{N}_t$ es perfectamente predecible con información hasta $t-1$.

Por analogía con las funciones de transferencia de las variables explicativas, al filtro por el que se pasa de la innovación a_t a la perturbación N_t se le conoce como *estructura dinámica residual.*

De acuerdo con (3.8.3) y (3.8.4), el modelo lineal general uniecuacional (3.8.1) se puede expresar como

$$Y_t=\sum_{j=1}^{m}X_{jt}^*+N_t. \tag{3.8.5}$$

Es interesante comparar la formulación alternativa (3.8.5) con la descomposición del valor observado de Y_t en parte sistemática e innovación; es evidente que la parte sistemática vendrá dada por

$$PS_t=\sum_{j=1}^{m}X_{jt}^*+\hat{N}_t$$

La innovación a_t se obtiene como la diferencia entre N_t y $\hat{N}_t$.

De esto se desprende que en un modelo uniecuacional la parte sistemática de la variable a explicar está formada por dos tipos de componentes:

— la contribución de las variables explicativas, o lo que es lo mismo, aquella parte del valor de Y en el momento t que se debe a los valores que han tomado, en el presente o en el pasado, las variables explicativas; y
— lo que hay de predecible en el elemento residual o perturbación N_t y que puede ser anticipado con la información que se dispone hasta el periodo inmediatamente anterior.

En lo que queda de este epígrafe y en el siguiente se analizarán estos componentes de la parte sistemática de Y_t, tal y como se obtienen utilizando un modelo uniecuacional[33]. En el resto de este epígrafe se prestará especial atención al papel de la perturbación, y en el siguiente se analizará la relación entre las variables explicativas y la variable a explicar, tal y como se refleja en sus correspondientes estructuras dinámicas.

La perturbación N_t comprende principalmente el efecto de variables explicativas omitidas que, por la hipótesis de exogeneidad, son ortogonales a las variables X_j incluidas. En todo análisis económico existe un conjunto muy amplio de variables de menor importancia que es imposible, o al menos muy difícil, cuantificar, y que se suelen relegar al término residual. Esto no significa que éste tenga que seguir un proceso ruido blanco: al contrario, sus propiedades estocásticas están inducidas por las propiedades de aquellas variables explicativas que se han omitido, y cualquier clase de restricción a priori que se haga sobre N_t puede resultar incorrecta en un problema específico.

El elemento N_t se puede ver como un conjunto de variables que se desconocen, por lo que se le formula, a su vez, un modelo univariante sobre las innovaciones del modelo. Así puede existir una diferencia importante entre el filtro asociado a la innovación y cualquier función de transferencia de las variables explicativas. Antes se comentó que cualquier $v_j(L)$ tenía que ser necesariamente estable, o lo que es lo mismo, una variación transitoria en la variable explicativa a la que estaba asociado tenía que originar una variación en la variable dependiente que con el tiempo tiende a cero. Por lo tanto si la contribución de X_j resulta no estacionaria, esto se debe a que la propia X_j es no estacionaria, de tal forma que la no estacionariedad de Y por tal contribución es consecuencia directa del tipo de evolución que presenta la variable explicativa.

[33] Una vez más es necesario insistir que cuando se habla de la parte sistemática de una variable es conveniente recalcar el tipo de modelo que se está considerando, ya que para una misma variable la parte sistemática es distinta según se considere un modelo univariante o un modelo econométrico.

Por el contrario, la restricción de estacionariedad no se impone al filtro asociado a la innovación: ésta es por definición ruido blanco, y por tanto estacionaria, por lo que si se impusiese sobre $\theta_q(L)/\phi_{p+d}(L)$ una restricción similar a la que pesa sobre las funciones de transferencia de los inputs $v_j(L)$, se estaría obligando a que las variables omitidas fuesen estrictamente estacionarias. Ahora bien, dado que en algún contexto parece razonable que la no estacionariedad de Y venga explicada al menos en parte por la no estacionariedad de las variables omitidas, resulta conveniente, en general, no imponer esta restricción de estacionariedad sobre la estructura dinámica residual.

¿En qué situaciones cabe esperar la presencia de elementos no estacionarios en el filtro de las innovaciones? En el epígrafe 3.10 se tratará este tema con más detalle, pero se pueden avanzar algunas ideas generales.

La inclusión del factor Δ^d, con un valor para d distinto de cero, en la parte autorregresiva de la estructura dinámica residual implica que la evolución no estacionaria de carácter tendencial de la variable dependiente no se explica plenamente por la evolución no estacionaria de las variables explicativas.

En la modelización de relaciones macroeconómicas sería de esperar que Δ^d apareciese poco frecuentemente, o al menos que el valor de d fuese menor al número de diferenciaciones del modelo univariante de Y_t. Los modelos que ligan variables macroeconómicas normalmente tratan de captar las relaciones tendenciales o de largo plazo entre las variables, y la presencia de raíces unitarias en el filtro de la perturbación indica que las variables explicativas que se han incluido no explican en su totalidad el componente permanente latente en la evolución de Y_t.

No obstante, en ocasiones el nivel de una variable macroeconómica Y_t puede depender de variables de población, institucionales o de otras variables macroeconómicas no estacionarias difíciles de medir y que, en consecuencia, no pueden incluirse en el modelo, y cuyo efecto sobre la variable dependiente se refleja en una perturbación de carácter no estacionario. Un ejemplo de variable difícil de medir es el progreso técnico: hay cierta evidencia de que esta variable contiene una raíz unitaria, lo que justifica la presencia de perturbaciones no estacionarias en aquellos casos en que el progreso técnico pueda ejercer algún tipo de influencia.

Por contra, en aplicaciones a problemas microeconómicos el nivel de Y_t puede depender, entre otros factores, de la evolución de la situación económica general, y si no se incluyen en el modelo variables para aproximarla es probable que aparezca el operador Δ^d.

Otro tipo de raíces unitarias que pueden surgir en el elemento residual son las del operador $U_{s-1}(L)$, siendo s el período estacional, que son las responsables de las oscilaciones de carácter estacional (véase la sección 2.4). Este operador puede aparecer tanto en problemas macro como microeconómicos, y su presencia indica que la estacionalidad de carácter no estacionario de la variable dependiente no viene plenamente explicada por las variables incluidas en la ecuación: en tanto en cuanto muchas veces la estacionalidad de las variables económicas viene también inducida por factores institucionales, parece razonable la presencia de $U_{s-1}(L)$ en modelos econométricos.

No obstante, la presencia de variables económicas explicativas pueden cubrir una parte importante de la oscilación estacional de la variable dependiente, de modo que la parte restante sea muy estable, en cuyo caso un esquema de variables artificiales estacionales, en vez del operador $U_{s-1}(L)$, puede ser más adecuado.

Puesto que en la práctica Δ también resulta ser relativamente común, $U_{s-1}(L)$ y Δ se combinan en Δ_s, y no es raro encontrar modelos econométricos que requieran el operador de diferencia estacional en la estructura dinámica residual.

Admitiendo que el filtro de la perturbación contenga diferencias de los dos tipos, regular y estacional, normalmente se podrá factorizar en

$$\frac{\theta_q(L)}{\phi_{p+q}(L)}=\frac{\theta_{q1}(L)\cdot\Theta_{q2}(L^s)}{\phi_{p1}(L)\cdot\Phi_{p2}(L^s)\cdot\Delta^d\cdot\Delta_s^D}$$

y por lo tanto si en la expresión general

$$\begin{gathered} Y_t=\sum_{j=1}^{m}\frac{w_j(L)}{\delta_j(L)}X_{jt}+N_t \\ N_t=\frac{\theta_{q1}(L)\cdot\Theta_{q2}(L^s)}{\phi_{p1}(L)\cdot\Phi_{p2}(L^s)\cdot\Delta^d\cdot\Delta_s^D}a_t \end{gathered} \tag{3.8.6}$$

se define

$$n_t=\frac{\theta_{q1}(L)\cdot\Theta_{q2}(L^s)}{\phi_{p1}(L)\cdot\Phi_{p2}(L^s)}a_t=\Delta^d\Delta_s^D N_t$$

entonces el modelo lineal dinámico uniecuacional (3.8.1) se puede expresar como

$$\Delta^d\Delta_s^D Y_t=\sum_{j=1}^{m}\frac{w_j(L)}{\delta_j(L)}\Delta^d\Delta_s^D X_{jt}+n_t \tag{3.8.7}$$

Ambas expresiones, (3.8.6) y (3.8.7), son equivalentes: mientras una utiliza variables originales y una perturbación no estacionaria, la otra emplea variables diferenciadas y un residuo estacionario. Esta relación es importante tenerla presente, ya que implica que en toda función de transferencia en la que las variables originales aparecen diferenciadas, las variables explicativas no consiguen captar en su totalidad la evolución de carácter no estacionario de la variable dependiente; y eso motiva que la perturbación tenga una contribución no estacionaria en la explicación del nivel de Y_t.

3.9. La contribución de las variables explicativas

Anteriormente se ha definido la contribución de la variable explicativa X_{jt}, denotada por X_{jt}^*, como

$$X_{jt}^*=\frac{w_j(L)}{\delta_j(L)}X_{jt}=v_j(L)X_{jt}=(v_{0,j}+v_{1,j}L+v_{2,j}L^2+\cdots)X_{jt}.$$

Como se ha indicado anteriormente, los filtros $v_j(L)$ cumplen la propiedad

$$v_{i,j}\xrightarrow[i\to\infty]{}0 \qquad j=1,2,\ldots,m$$

y para un k suficientemente grande

$$\frac{w_j(L)}{\delta_j(L)}\simeq v_{0,j}+v_{1,j}L+\cdots+v_{k,j}L^{k,j}$$

Al conjunto de coeficientes $v_{0,j}$, $v_{1,j}$, $v_{2,j}$, ..., se le conoce como la *función de respuesta a un impulso* de la variable explicativa X_j, y caracteriza plenamente la relación dinámica existente entre X_j y Y.

Para ahondar en esta última afirmación volvamos al ejemplo recogido en (3.8.2) y supóngase, para simplificar, que existe una sola variable explicativa, X_t. Si en el intervalo de tiempo $t<t^*$ el sistema está en equilibrio, se tiene que

$$N_t = N^e = 0$$
$$X_t = X^e$$
$$Y_t = v(L)X_t + N_t = v(L)X^e + 0 = Y^e;$$

la perturbación permanece siempre igual a cero y supóngase ahora que en $t=t^*$, X_t experimenta una variación impulso, pasando a valer X^e+1; y con posterioridad a t^* vuelve a su valor de equilibrio, $X_t = X^e$, $t>t^*$. En este caso, el efecto de la perturbación impulso de X sobre Y vendrá dado por

$$\begin{aligned} Y_{t^*-1} &= v(L)X_{t^*-1} = v_0 X_{t^*-1} + v_1 X_{t^*-2} + v_2 X_{t^*-3} + \cdots = \\ &= v_0 X^e + v_1 X^e + v_2 X^e + \cdots = v(L)X^e = Y^e \end{aligned}$$

$$\begin{aligned} Y_{t^*} &= v(L)X_{t^*} = v_0 X_{t^*} + v_1 X_{t^*-1} + v_2 X_{t^*-2} + \cdots = \\ &= v_0(X^e+1) + v_1 X^e + v_2 X^e + \cdots = Y^e + v_0 \end{aligned}$$

$$\begin{aligned} Y_{t^*+1} &= v(L)X_{t^*+1} = v_0 X_{t^*+1} + v_1 X_{t^*} + v_2 X_{t^*-1} + \cdots = \\ &= v_0 X^e + v_1(X^e+1) + v_2 X^e + \cdots = Y^e + v_1 \end{aligned}$$

$$\begin{aligned} Y_{t^*+2} &= v(L)X_{t^*+2} = v_0 X_{t^*+2} + v_1 X_{t^*+1} + v_2 X_{t^*} + \cdots = \\ &= v_0 X^e + v_1 X^e + v_2(X^e+1) + \cdots = Y^e + v_2 \end{aligned}$$

y en general

$$Y_{t^*+k} = v(L)X_{t^*+k} = Y^e + v_k,$$

hasta que para un valor suficientemente grande se cumpla que $v_k \simeq 0$, con lo que

$$Y_{t^*+k} \simeq Y^e$$

y se restablece el equilibrio inicial que había sido perturbado por la variación transitoria de la variable explicativa.

Resulta así que la perturbación de tipo impulso experimentada por la variable explicativa alteró el equilibrio de la variable endógena durante un período de tiempo finito; además, la desviación entre el valor real de Y y su teórico valor de equilibrio, i unidades de tiempo después del cambio en X, es precisamente igual a v_i. En consecuencia, se pueden interpretar los coeficientes v_i como el impacto que ex-

perimenta la variable dependiente debido a una perturbación impulso registrada en el input i períodos atrás.

Si todas las variables están expresadas en logaritmos los coeficientes v_i son las elasticidades a diferentes plazos: v_0 es la *elasticidad contemporánea*, y en general v_i es la *elasticidad de Y respecto a X tras i períodos de desfase*.

Cuando exista un desfase de b períodos entre el momento en que la variable explicativa experimenta el impulso y el instante en que el output empieza a reaccionar, los b primeros coeficientes de la función serán iguales a cero,

$$v_0 = v_1 = \cdots = v_{b-1} = 0$$

y es preferible destacar esta situación explícitamente, expresando la contribución del input como[34]

$$X_t^* = \frac{w_s(L)}{\delta_r(L)} L^b X_t = \frac{w_s(L)}{\delta_r(L)} X_{t-b}.$$

Es importante estudiar con detalle los valores, forma, posibles desfases, etc., de esta función de respuesta a un impulso, ya que por regla general su caracterización constituye uno de los principales objetivos del proceso de modelización econométrica de fenómenos económicos.

Conocida la expresión racional de la función de transferencia de X, la función $v(L)$ se calcula de forma sencilla igualando polinomios

$$\frac{w_s(L)}{\delta_r(L)} L^b = v(L)$$

y despejando

$$w_s(L) \cdot L^b = \delta_r(L) \cdot v(L).$$

Si ahora se particulariza esta igualdad en la variable auxiliar z se tiene que

$$w_s(z) \cdot z^b = \delta_r(z) \cdot v(z),$$

[34] Los subíndices que se asignan a los polinomios finitos vuelven a representar su orden, ya que como sólo hay una variable explicativa no es preciso hacer explícito a cuál se refieren.

y puesto que dos polinomios son iguales si y sólo si los coeficientes de sus distintas potencias lo son, los coeficientes de la función de respuesta a un impulso vienen dados por

$$v_i = \begin{cases} 0 & i<b \\ w_0 & i=b \\ \delta_1 v_{i-1} + \delta_2 v_{i-2} + \cdots + \delta_r v_{i-r} + w_{i-b} & b+1 \leqslant i \leqslant b+s \\ \delta_1 v_{i-1} + \delta_2 v_{i-2} + \cdots + \delta_r v_{i-r} & i>b+s \end{cases}$$

De la última línea de la expresión anterior se comprueba que a partir de $i=b+s+1$ los coeficientes v_i vienen determinados por una ecuación en diferencias finitas, en la que $v_{b+s}, v_{b+s-1}, \ldots, v_{b+s-r+1}$ constituyen las condiciones iniciales. En consecuencia para $i \geqslant b+s-r+1$ la sucesión de coeficientes presentará una cierta estructura en su evolución, ya que los v_i vienen dados por funciones exponenciales y sinusoidales decrecientes.

Es importante recalcar esta idea, ya que la función de respuesta caracteriza la ley de comportamiento de Y frente a X, lo que puede ser de interés a la hora de estudiar el fenómeno. En general, y dependiendo de los valores de b, s y r, la función de respuesta a un impulso $v(L)$ se caracteriza por:

a) b primeros valores cero: $v_0, v_1, \ldots, v_{b-1}$;
b) $s-r+1$ valores siguientes (si $s>r$) que no siguen un patrón determinado: $v_b, v_{b+1}, \ldots, v_{b+s-r}$; y
c) el resto de valores que siguen el esquema que proporciona una ecuación en diferencias finitas de orden r, inducida por el denominador de la función de transferencia.

En resumen, el numerador $w_s(L)$ alarga sin estructura el efecto de un impulso, mientras el denominador $\delta_r(L)$ lo alarga imponiendo un patrón de comportamiento determinado a los coeficientes v_i, que será exponencial u oscilante en función de las raíces de $\delta_r(L)$. Se tiene así el mismo resultado que ya se obtuvo en el capítulo 2 al tratar el análisis de intervención; de hecho, este último se puede interpretar como la contribución de un tipo especial de variables explicativas, las variables artificiales. Por lo tanto todo lo que se dijo allí es de aplicación aquí, y viceversa.

Al mismo tiempo, una detenida consideración de la estructura de respuesta implícita en la función de transferencia constituye una herramienta útil a la hora de validar los resultados de la estimación.

Esta última afirmación queda mejor ilustrada con un ejemplo. Así, si en la estimación la función de respuesta a un impulso en un problema concreto se llega a una expresión de la forma

$$\frac{w_0}{1+0{,}5L} = w_0 - 0{,}5 w_0 L + 0{,}5^2 w_0 L^2 - 0{,}5^3 w_0 L^3 + \cdots,$$

este resultado puede ser evidencia de que se ha cometido algún tipo de error de especificación durante el proceso de modelización, por ejemplo la omisión de una variable explicativa con evolución oscilante. Tal omisión puede provocar que en la estimación aparezca la raíz negativa de $\delta_r(L)$, que induce una respuesta oscilante de la variable dependiente ante cambios en la correspondiente variable explicativa, lo que en ocasiones puede ser algo extraño entre variables económicas.

En situaciones como la de este ejemplo es muy probable que la variable dependiente presente bastantes oscilaciones, determinadas por una variable omitida, de manera que el proceso de búsqueda de un modelo que se ajuste a los datos intentará aproximarlas introduciendo un filtro que genere oscilaciones. De ahí que si se estima una determinada función de transferencia con un denominador que tiene raíces complejas o negativas, el analista debe asegurarse de que el correspondiente esquema de la función de respuesta tiene sentido teórico, antes de dar por válido el modelo.

Otra manera de analizar la función de transferencia consiste en la llamada *Función de Respuesta a un Escalón*. Sea una situación nuevamente caracterizada por:

a) $N_t = N^e = 0$, para todo t
b) $X_t = X^e$, $t<t^*$
c) $Y_t = v(L)X_t + N_t = v(L)X^e + 0 = Y^e$, $t<t^*$

Si en t^* la variable X_t experimenta un cambio permanente en su valor de equilibrio previo, de tal forma que $X_t = X^e + 1$, para todo t a partir de t^* la respuesta dinámica de Y vendrá dada por

$$\begin{aligned} Y_{t^*} &= v(L)X_{t^*} = v_0 X_{t^*} + v_1 X_{t^*-1} + v_2 X_{t^*-2} + \cdots = \\ &= v_0(X^e+1) + v_1 X^e + v_2 X^e + \cdots = Y^e + v_0 = Y^e + V_0 \end{aligned}$$

$$\begin{aligned} Y_{t^*+1} &= v(L)X_{t^*+1} = v_0 X_{t^*+1} + v_1 X_{t^*} + v_2 X_{t^*-1} + \cdots = \\ &= v_0(X^e+1) + v_1(X^e+1) + v_2 X^e + \cdots = Y^e + v_0 + v_1 = Y^e + V_1 \end{aligned}$$

$$\begin{aligned} Y_{t^*+2} &= v(L)X_{t^*+2} = v_0 X_{t^*+2} + v_1 X_{t^*+1} + v_2 X_{t^*} + \cdots = \\ &= v_0(X^e+1) + v_1(X^e+1) + v_2(X^e+1) + \cdots = \\ &= Y^e + v_0 + v_1 + v_2 = Y^e + V_2 \end{aligned}$$

y en general

$$Y_{t^*+k} = Y^e + v_0 + v_1 + \cdots + v_k = Y^e + V_k.$$

Los coeficientes V_0, V_1, V_2, ..., definen la función de respuesta a un escalón, dada por

$$V(L) = V_0 + V_1 L + V_2 L^2 + \cdots$$

El coeficiente V_i mide el cambio en la variable dependiente originado por una perturbación de carácter permanente en la variable explicativa que ha comenzado i períodos atrás.

Normalmente a cada uno de estos coeficientes se les conoce con el nombre de *multiplicadores*, y si las variables están expresadas en logaritmos estos multiplicadores se pueden interpretar como *elasticidades acumuladas*.

Es fácil comprobar la relación existente entre las funciones de respuesta a un impulso y a un escalón; del desarrollo anterior es inmediato que:

$$Y_{t^*+k} = Y_e + v(L) S(t^*)_{t^*+k} \qquad (3.9.1)$$

siendo $S(t^*)_{t^*+k}$ el valor en t^*+k de una variable artificial de tipo escalón que empieza a tomar el valor uno en t^*. Este mismo resultado se puede expresar también de la siguiente forma:

$$Y_{t^*+k} = Y_e + V(L) D(t^*)_{t^*+k} \qquad (3.9.2)$$

donde $D(t^*)_{t^*+k}$ es una variable impulso definida de forma similar. Igualando ambas expresiones, y teniendo en cuenta que

$$D(t^*)_{t^*+k} = \Delta S(t^*)_{t^*+k},$$

se tiene que

$$v(L) S(t^*)_{t^*+k} = V(L) D(t^*)_{t^*+k},$$

con lo que

$$v(L) = \Delta V(L)$$

y la diferencia entre V_i y V_{i-1} es precisamente igual a v_i.

Alternativamente

$$V(L) = v(L)/\Delta, \qquad (3.9.3)$$

de donde se ve que $V(L)$ es la suma o integración de $v(L)$.

La función de respuesta a un escalón admite dos interpretaciones. Una la que se viene dando, en la que los coeficientes V_i recogen la perturbación puntual, es decir, en t^*+i, en la variable Y por una perturbación escalón en la variable explicativa i períodos antes. Esta interpretación aparece claramente en (3.9.2). Sin embargo, no es necesario que la variable explicativa sufra una perturbación constante y permanente para poder hablar de función de respuesta a un escalón. En efecto, de acuerdo con (3.9.3) los coeficientes V_i pueden interpretarse como la acumulación de perturbaciones puntuales v_0, v_1, ..., v_i, sufridas por la variable endógena debido a una perturbación impulso en la variable explicativa i períodos antes.

La restricción de estacionariedad de la función de transferencia, es decir

$$v_i \xrightarrow[i \to \infty]{} 0$$

implica que

$$V_i \xrightarrow[i \to \infty]{} \text{constante}$$

y por lo tanto el cambio permanente que registra la variable explicativa tiene un efecto permanente y finito en la variable dependiente.

Si en la función de transferencia $w_s(L)/\delta_r(L)$ el polinomio autorregresivo es de orden cero, es decir, un filtro MA puro, entonces

$$V_i = V_s \qquad i > s$$

y a partir de s períodos ya se ha absorbido totalmente el cambio en X: en ese caso a V_s se le llama *ganancia*, *multiplicador a largo plazo o total*, o en el caso de que las variables estén en logaritmos, *elasticidad a largo plazo o total*.

Si por contra se tiene un polinomio de orden mayor que cero en el denominador de la función de transferencia, v_i no será exactamente igual a cero para valores finitos de i; pero como a partir de un determinado retardo k los coeficientes de la función de respuesta a un impulso serán aproximadamente cero, se tendrá que

$$V_i \simeq V_k \qquad i > k$$

y se puede tomar V_k como una aproximación al multiplicador (elasticidad) total o a largo plazo.

El valor exacto de dicho multiplicador viene dado por

$$\sum_{i=0}^{\infty} v_i = \frac{w_s(1)}{\delta_r(1)} = \frac{w_0 + w_1 + \cdots + w_s}{1 - \delta_1 - \cdots - \delta_r} = g$$

también llamado *ganancia del filtro*.

En aquellos casos en que los coeficientes v_i sean todos positivos, y por lo tanto también lo sean los V_i, es de interés calcular las correspondientes funciones normalizadas respecto a la ganancia total.

Se define la *función normalizada de respuesta a un escalón* a

$$V^*(L) = V(L)/g$$

cuyos coeficientes V_i^* vienen dados por V_i/g, y su interpretación es la de multiplicadores o elasticidades acumuladas medidos en tanto por uno del multiplicador a largo plazo. Si por ejemplo $V_{12}^* = 0{,}7$, doce unidades de tiempo después del momento del cambio en la variable explicativa, la variable dependiente habrá experimentado ya el 70 % del efecto total. Como es evidente, este tipo de información es de mucho interés a la hora de caracterizar la relación existente entre variables económicas.

Análogamente se puede definir la *función normalizada de respuesta a un impulso* como

$$v^*(L) = v(L)/g$$

cuyo coeficiente v_i^* es igual a v_i/g. Si estos coeficientes no son negativos, se verifica que

$$v_i^* > 0, \qquad \sum_{i=0}^{\infty} v_i^* = 1$$

y en conjunto se pueden interpretar como probabilidades de una variable auxiliar z cuya función de cuantía sea

$$P(z = i) = v_i^* \qquad i = 0,\ 1,\ 2,\ \ldots$$

y a partir de aquí se puede definir el *retardo medio*, denotado por $\bar{m}$, como

$$\bar{m} = \sum_{i=0}^{\infty} i \cdot v_i^* = \sum_{i=0}^{\infty} \frac{i \cdot v_i}{\left(\sum_{l=0}^{\infty} v_l\right)}.$$

Teniendo en cuenta que la derivada de $v(L)$ respecto a L viene dada por

$$v'(L) = v_1 + 2v_2 L + 3v_3 L^3 + \cdots$$

donde ′ denota derivada respecto a L, una expresión alternativa del retardo medio es

$$\bar{m} = \left.\frac{v'(L)}{v(L)}\right|_{L=1} = \frac{v_1 + 2v_2 + 3v_3 + \cdots}{v_0 + v_1 + v_2 + \cdots}$$

Esto sugiere una forma de obtener el retardo medio directamente a partir de los polinomios originales de la función de transferencia, $w_s(L)$ y $\delta_r(L)$; como

$$v'(L) = \frac{\delta_r(L) w_s'(L) - \delta_r'(L) w_s(L)}{[\delta_r(L)]^2},$$

entonces

$$\frac{v'(L)}{v(L)} = \frac{[\delta_r(L) w_s'(L) - \delta_r'(L) w_s(L)]/[\delta_r(L)]^2}{w_s(L)/\delta_r(L)} =$$

$$= \frac{w_s'(L)}{w_s(L)} - \frac{\delta_r'(L)}{\delta_r(L)}$$

y

$$\bar{m} = \left(\frac{w_s'(L)}{w_s(L)} - \frac{\delta_r'(L)}{\delta_r(L)}\right)_{L=1}$$

Con frecuencia, también resulta útil calcular el *retardo mediana*, que se define como el número de períodos de tiempo que tarda el sistema en absorber el cincuenta por cien de la ganancia.

Para ilustrar los conceptos teóricos que se han desarrollado en este epígrafe y en el anterior considérese el modelo del efectivo en manos del público de Espasa (1980):

$$\Delta_4 \ln E_t = \frac{0{,}18}{1 - 0{,}84L} \Delta_4 \ln P_t + \frac{0{,}20}{1 - 0{,}60L} \Delta_4 \ln Y_t +$$

$$+ \frac{1}{(1 - 0{,}99L + 0{,}40L^2)\,(1 + 0{,}65L^4)}\, a_t$$

siendo:

— E_t: serie trimestral de efectivo; sigue un proceso no estacionario, y su modelo ARIMA contiene una diferencia regular y una estacional[35];
— P_t: índice del coste de vida, equivalente al actual índice de precios al consumo;
— Y_t: una aproximación al PIB trimestralizado sin variación estacional[36].

Lo primero que se observa es que las variables explicativas no captan en su totalidad la evolución no estacionaria del efectivo: la permanencia de la diferencia estacional indica que la perturbación del modelo deambula con estacionalidad.

No es extraño que las variables explicativas no hayan conseguido captar la estacionalidad; ésta podría venir explicada por el verdadero PIB, pero nunca por una aproximación de éste desprovista del componente estacional.

En cuanto a la presencia de una raíz unitaria autorregresiva en los residuos puede deberse a cambios en los hábitos de pago, que se haya pasado a utilizar masivamente las tarjetas de crédito y las domiciliaciones de pagos, y a oscilaciones en la economía sumergida, que no viene recogida en el PIB y que utiliza mayor cantidad de efectivo.

El modelo de la perturbación viene dado por

$$\Delta_4 N_t = \frac{1}{(1-0{,}99L+0{,}40L^2)\ (1+0{,}65L^4)}\, a_t$$

El polinomio $(1-0{,}99L+0{,}40L^2)$ refleja un ciclo de período nueve trimestres y factor de amortiguamiento 0,63; esto está indicando que el efectivo presenta un componente de este tipo que no viene explicado por precios o producción, y que termina reflejándose en la perturbación. Lo mismo puede decirse de $(1+0{,}65L^4)$, una oscilación en el efectivo que no es explicada por los precios o el PIB.

La función de respuesta a un impulso del efectivo respecto a los precios es igual a

$$0{,}18/(1-0{,}84L)=0{,}18(1+0{,}84L+0{,}84^2L^2+0{,}84^3L^3+\cdots)=$$
$$=0{,}18+0{,}151L+0{,}127L^2+0{,}107L^3+\cdots$$

[35] Nótese que como la serie es trimestral el período estacional es igual a cuatro.

[36] Véase Rodríguez (1978) para un mayor detalle en la elaboración de esta variable.

y respecto al PIB

$$0{,}20/(1-0{,}60L)=0{,}20(1+0{,}60L+0{,}60^2L^2+0{,}60^3L^3+\cdots)=$$
$$=0{,}20+0{,}12L+0{,}072L^2+0{,}043L^3+\cdots$$

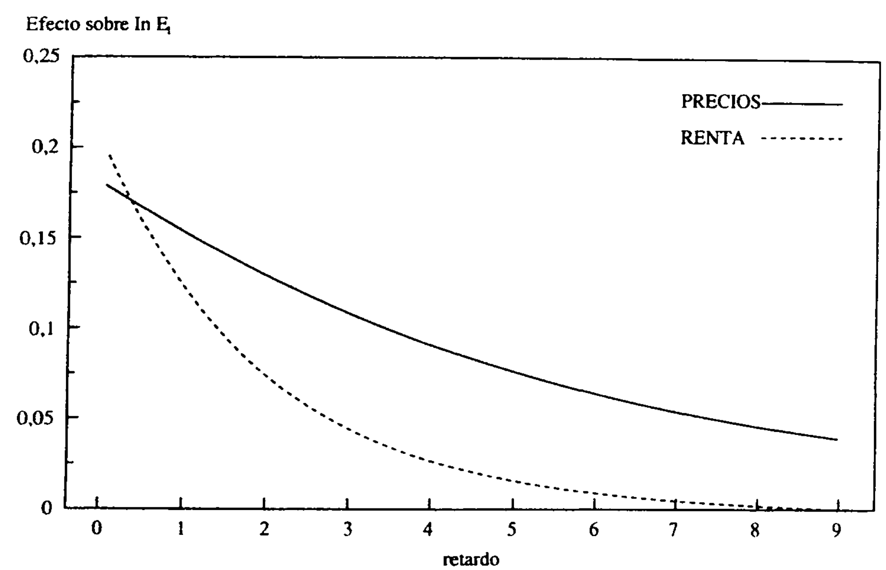
Ambas funciones se representan en el gráfico 3.2.

Aplicando las demás fórmulas se obtendrían el resto de funciones y estadísticos. Los resultados se resumen en el cuadro 3.3; sin necesidad de entrar en un análisis detallado de los mismos, se observa claramente como: 1) la elasticidad a largo plazo respecto a precios es mayor que respecto a PIB, y 2) el ajuste desencadenado por una variación en una variable explicativa es mucho más rápido para PIB que para precios.

3.10. Formulaciones alternativas del modelo lineal uniecuacional

En los dos epígrafes anteriores se ha desarrollado la idea de que un modelo uniecuacional para la variable Y descompone, para cada momento t, el valor de esa variable en:

GRÁFICO 3.2. *Funciones de respuesta a un impulso para el modelo del efectivo.*



CUADRO 3.3. *Funciones de respuesta del modelo para el efectivo*

		PRECIOS				PIB			
		F.R.I.	*F.R.I.N.*	*F.R.E.*	*F.R.E.N.*	*F.R.I.*	*F.R.I.N.*	*F.R.E.*	*F.R.E.N.*
Retardos	0	0,180	0,16	0,180	0,16	0,200	0,40	0,200	0,40
	1	0,151	0,13	0,331	0,29	0,120	0,24	0,320	0,64
	2	0,127	0,11	0,458	0,41	0,072	0,14	0,392	0,78
	3	0,107	0,10	0,565	0,50	0,043	0,09	0,435	0,87
	4	0,090	0,08	0,655	0,58	0,026	0,05	0,461	0,92
	5	0,075	0,07	0,730	0,65	0,016	0,03	0,477	0,95
	6	0,063	0,06	0,793	0,70	0,009	0,01	0,486	0,97
	7	0,053	0,05	0,846	0,75	0,006	0,01	0,492	0,98
	8	0,045	0,04	0,891	0,79	0,003	0,01	0,495	0,99
Elasticidad a largo plazo		1,125				0,5			
Retardo medio		5,25				1,5			
Retardo mediana		3				1			

Nota: F.R.I.: Función de respuesta a un impulso.
F.R.I.N.: Id. normalizada.
F.R.E.: Función de respuesta a un escalón.
F.R.E.N.: Id. normalizada.

— la contribución de las variables explicativas, o cómo los valores en distintos momentos del tiempo de éstas repercuten en el valor presente de Y; y
— la contribución del elemento residual, que básicamente refleja el efecto de variables explicativas omitidas sobre la variable que se pretende modelizar.

En el caso de que sólo se haya incluido una variable explicativa en el modelo, se tiene que éste toma la forma

$$Y_t = \frac{w_s(L)}{\delta_r(L)} X_t + \frac{\theta_q(L)}{\phi_p(L)} a_t \tag{3.10.1}$$

donde el primer sumando del lado derecho de (3.10.1) refleja la contribución de la variable explicativa y el segundo la del elemento residual. En lo que sigue supondremos en general que sólo se incluye una variable explicativa, y que ésta es fuertemente exógena.

Quizás la forma más clara de entender el modelo uniecuacional es la representada en (3.10.1), pero otras formulaciones, totalmente equivalentes a (3.10.1), son también factibles[37]. Puesto que distintas formulaciones recalcan diferentes aspectos de la relación entre X e Y, es recomendable que el analista de la coyuntura las conozca, y el objetivo de este epígrafe es presentarlas.

Para simplificar, supóngase de momento que en (3.10.1) las posibles raíces unitarias del polinomio autorregresivo residual han sido eliminadas tomando diferencias hasta obtener las variables X e Y.

Definiendo

$$\alpha^*(L) = \delta_r(L) \cdot \phi_p(L)$$
$$\beta^*(L) = w_s(L) \cdot \phi_p(L)$$
$$\mu(L) = \theta_q(L) \cdot \delta_r(L)$$

entonces

$$\alpha^*(L) Y_t = \beta^*(L) X_t + \mu(L) a_t \tag{3.10.2}$$

representa el mismo modelo que (3.10.1). A la formulación en términos de tales polinomios se le llama *formulación en ecuación de diferencias finitas estocástica*, mientras que (3.10.1) es la *formulación*

[37] Harvey (1981a) dedica dos capítulos al modelo lineal uniecuacional dinámico, uno desde la perspectiva de la función de transferencia y otro desde la que se adopta en este epígrafe.

en función de transferencia, que también se conoce como *modelo de retardos racionales distribuidos*.

En (3.10.2) el residuo sigue un proceso de medias móviles, que bajo el supuesto habitual de invertibilidad se podrá aproximar, véase el capítulo 2, mediante un proceso autorregresivo, con lo que

$$\mu(L)a_t \simeq a_t/r(L).$$

Con ello, si se multiplica (3.10.2) por $r(L)$ se obtiene

$$\alpha(L)Y_t = \beta(L)X_t + a_t \tag{3.10.3}$$

en donde

$$\alpha(L) = \alpha^*(L)\cdot r(L)$$
$$\beta(L) = \beta^*(L)\cdot r(L)$$

Si el modelo tiene m variables explicativas toma la forma

$$\alpha(L)Y_t = \sum_{j=1}^{m} \beta_j(L)X_{jt} + a_t \tag{3.10.4}$$

En (3.10.3) o (3.10.4) se tiene una ecuación en diferencias finitas estocásticas con un residuo que es ruido blanco. En este modelo, los estimadores por mínimos cuadrados ordinarios son lineales en las observaciones. Sin embargo, en (3.10.1) o en su generalización (3.8.1) con m variables explicativas, los estimadores mínimo cuadráticos son no lineales. Esta complejidad en la estimación ha llevado con frecuencia a rehuir modelos del tipo (3.8.1)-(3.10.1) en favor de (3.10.4)-(3.10.3). Sin embargo, a lo largo de esta sección veremos cómo la formulación en ecuación en diferencias presenta otro tipo de problemas.

La formulación (3.10.3) indica que con polinomios $\alpha(L)$ y $\beta(L)$ bastante generales sobre Y y X se puede expresar el modelo con unos residuos ruido blanco; esta formulación constituye una *generalización dinámica del modelo de regresión*.

La formulación en términos de ecuaciones en diferencias finitas estocástica es la empleada en la denominada metodología econométrica de la London School of Economics (LSE) —véase Gilbert (1986), Hendry y Richard (1982), Spanos (1988), etc.—, en la que se empieza especificando el modelo con polinomios dinámicos generales para a continuación, y a través de una una sucesión de contrastes, llegar a un modelo final más simple. A dicha estrategia se le denomina

«de lo general a lo particular» y se basa en el principio de verificación secuencial de hipótesis desarrollado en Anderson (1971).

La estrategia de lo general a lo particular tiene el riesgo de acabar con un modelo más general que lo que la realidad requiere, a cambio de minimizar el riesgo de concluir con un modelo excesivamente restrictivo. Lo positivo de esta estrategia es que si el modelo final es demasiado general la estimación será ineficiente, pero si es demasiado restrictivo la estimación será inconsistente. Además, el procedimiento secuencial de Anderson tiene la propiedad de ser uniformemente más potente dentro de la clase de procedimientos que fijan la probabilidad de no rechazar una hipótesis menos restrictiva que la verdadera.

Volviendo a la comparación entre (3.10.1) y (3.10.3), ambos modelos son idénticos: sin embargo si entre $\delta_r(L)$ y $\phi_p(L)$ no hay raíces comunes, la formulación en términos de ecuación de diferencias estocástica es innecesariamente compleja.

Si por el contrario existen raíces comunes entre $\delta_r(L)$ y $\phi(L)$, de tal forma que

$$\delta_r(L) = \alpha^+(L)\cdot\delta^+(L)$$
$$\phi_p(L) = \alpha^+(L)\cdot\phi^-(L)$$

se tiene

$$\alpha^+(L)Y_t = \frac{w_s(L)}{\delta^+(L)}X_t + \frac{\theta_q(L)}{\phi^+(L)}a_t,$$

que vuelve a ser una función de transferencia, pero no sobre Y_t sino sobre $\alpha^+(L)Y_t = Y_t^+$. No obstante, si en la expresión anterior el polinomio racional de X_t se aproxima por un polinomio simple, el de a_t por un esquema autorregresivo —$r(L)$— y se multiplica toda la ecuación por $r(L)$, se vuelve a obtener una formulación en términos de ecuación de diferencias estocásticas.

Estas diferentes presentaciones conectan con lo que en Teoría Económica se estudia con los nombres de mecanismo de ajuste parcial y mecanismo de expectativas adaptativas.

El *mecanismo de ajuste parcial* se puede formular en este contexto de la siguiente manera: se tiene un valor deseado dado por

$$Y_t^d = \frac{w_s(L)}{\delta^+(L)}X_t$$

y una relación entre realizaciones y deseos:

$$Y_t - Y_{t-1} = \gamma(Y_t^d - Y_{t-1}) + \frac{\theta_q(L)}{\phi^+(L)} a_t$$

de donde

$$Y_t - (1-\gamma)Y_{t-1} = \frac{\gamma w_s(L)}{\delta^+(L)} X_t + \frac{\theta_q(L)}{\phi^+(L)} a_t$$

que es una función de transferencia para $[1-(1-\gamma)L]Y_t$. Nótese que la ganancia de ese filtro es $1-(1-\gamma)=\gamma$.

A su vez, el *mecanismo de expectativas adaptativas* parte de la relación

$$Y_t = \beta X^e_{t+1} + \frac{\theta_q(L)}{\phi^+(L)} a_t$$

con un mecanismo de generación del valor esperado de X

$$X^e_{t+1} - X^e_t = \gamma(X_t - X^e_t)$$

o bien

$$X^e_{t+1} = [\gamma/1-(1-\gamma)L]X_t$$

por lo que el modelo final es

$$Y_t = \frac{\gamma\beta}{1-(1-\gamma)L} X_t + \frac{\theta_q(L)}{\phi^+(L)} a_t$$

es decir, de la forma de una función de transferencia.

Cuando el modelo es conocido todas las expresiones en términos de función de transferencia y de ecuación en diferencias finitas son equivalentes. Pero cuando el modelo no se conoce y el problema planteado es cómo llegar a una expresión dinámica que aproxime el correspondiente esquema teórico desconocido, parece más conveniente establecer una estrategia de búsqueda, usando la formulación de función de transferencia o la de ecuación en diferencias estocástica, sobre la variable original. La formulación en términos de función de transferencia sobre $\alpha^+(L)Y_t$ es menos operativa en este proceso de búsqueda de especificación dinámica.

Cuando el modelo es desconocido el investigador tiene que decidir con qué tipo de formulación —(3.10.1) o (3.10.3)— debe comenzar su tarea de construir un modelo econométrico. Ciertamente lo importante no es el punto de partida sino la validación del resultado final; no obstante, según los casos el comenzar por una u otra especificación puede tener sus ventajas.

El modelo (3.10.1) en términos de función de transferencia es especialmente útil cuando *se dispone de información a priori* sobre las funciones de respuesta de la variable Y a cambios en cada una de las variables explicativas. Este tipo de información será normalmente sobre si la función de respuesta: a) se da desde el principio —impacto contemporáneo o instantáneo— o si actúa con algún retraso; b) es larga o corta; c) si a partir de un determinado retardo muestra un patrón de comportamiento, o si por el contrario las respuestas puntuales son relativamente libres; d) si el posible patrón de comportamiento es uniforme u oscilante; etc.

Así:

a) Si la variable dependiente responde con retraso a los cambios en la correspondiente variable explicativa, se emplearán formulaciones del tipo

$$\frac{w_s(L)}{\delta_r(L)} L^b,$$

dando a b un valor adecuado a partir de la información disponible. Si se cree que la respuesta se produce desde el principio, se asignará a b el valor cero.

b) Una función de respuesta larga se puede lograr con un polinomio $w_s(L)$ con un valor de s relativamente largo, y por lo tanto con una formulación pródiga en parámetros. Alternativamente se pueden utilizar esquemas autorregresivos de orden r muy pequeño, normalmente uno o dos, que requieren pocos parámetros. En general, y teniendo siempre en cuenta los condicionamientos de los apartados posteriores c) y d), una respuesta larga tenderá a implementarse mediante esquemas autorregresivos, como se hacía en el ejemplo anterior del efectivo, precios y producto interior bruto.

c) Una función de respuesta del tipo $w_s(L)$ implica que la variable a explicar se ve perturbada durante solamente $s+1$ períodos de tiempo —t^* a t^*+s—, y en cada momento el efecto puntual viene recogido por un parámetro w_i específico. Por el contrario, para respuestas largas con efectos puntuales que muestran un determinado patrón de comportamiento, se requiere el operador autorregresivo $\delta_r(L)$, que con pocos parámetros produce una respuesta convergente, aunque teóricamente infinita. En este último caso los (infinitos) parámetros de la función de respuesta $v(L)$ muestran una relación funcional entre ellos, ya que son generados a partir de los r parámetros autorregresivos operando sobre el polinomio $w_r(L)$: la interacción entre los polinomios $w_s(L)$ y $\delta_r(L)$ y cómo ésta afecta a los coeficientes de $v(L)$ se ha discutido con detalle en el epígrafe anterior.

d) Las características de la relación funcional entre los coeficientes de la función de respuesta dependen de las raíces del polinomio autorregresivo $\delta_r(L)$, tal y como se desarrolló también en el epígrafe anterior. Resumiendo lo que allí se dijo para el caso de que r valga la unidad, si la raíz es positiva se tendrá una respuesta de tipo exponencial convergente; y si es negativa una respuesta oscilatoria donde a un efecto positivo le sigue uno negativo, y viceversa. Una estructura de tipo cíclico requerirá que $\delta_r(L)$ contenga un par de raíces complejas conjugadas, por lo que en tal caso r debe tomar al menos el valor dos. Respecto a las funciones de respuesta oscilante —sean cíclicas o simplemente con alternancia de signo— ya se ha advertido al lector que antes de aceptarlas conviene asegurarse de que no contradicen la información a priori que pueda existir sobre la función en cuestión.

El esquema autorregresivo de primer orden con raíz positiva puede ser muy útil en la práctica, tanto operando sobre un polinomio $w_s(L)$ de orden cero, véase el cuadro 3.3, como sobre polinomios $w_s(L)$ de orden mayor que cero. En este último caso la estructura funcional de los coeficientes de la función de respuesta sólo se dará a partir del retardo m, siendo $m = \text{máx}(0, (b+s-r+1))$.

En cualquier caso, cuando se dispone de información a priori sobre la función de respuesta conviene formular el modelo asignando a s y r valores lo más pequeños posibles, dentro del campo de valores capaces de captar las características a priori de la función. Con ello se cumplirá por un lado con lo que la información teórica requiere y los datos permiten, y por otro con el principio de parquedad en la parametrización. Ser parco a costa de omitir características básicas de la relación dinámica no sólo no es una virtud, sino un error.

La formulación del modelo para el término residual se puede hacer a partir del modelo univariante de la variable dependiente eliminado de él aquellos factores —raíces unitarias, factores que generan oscilaciones periódicas, etc.— que reflejen comportamientos que se esperan pasen a ser recogidos por las variables explicativas. La modelización de la demanda de efectivo, presentada al final del epígrafe anterior, ilustra claramente este punto.

En la formulación del modelo de función de transferencia con información a priori es pues importante tener un buen conocimiento de los modelos univariantes de todas las variables que entran en el mismo: de la variable a explicar, para saber qué es lo que hay que modelizar; de las variables explicativas, para conocer el potencial de

éstas en explicar el comportamiento de la variable dependiente y, en consecuencia, poder aproximar la estructura que debe tener el componente residual. Esto supone realizar un análisis bastante completo de los datos disponibles; este análisis es algo básico en la modelización econométrica, y algunos programas, como el PC-GIVE[38], incorporan en una fase previa a la modelización econométrica en sentido estricto el tipo de análisis previo de datos que se considera conveniente.

Con todo lo anterior el modelo econométrico estará *equilibrado*[39], es decir, todo lo contenido en la variable a explicar —parte izquierda de la ecuación— tendrá su contrapartida en la parte derecha, pudiendo ésta descomponerse en una parte sistemática y una innovación (ruido blanco). Si el modelo no está equilibrado las pretendidas innovaciones —$\hat{a}_t$— que se estiman no serán tales.

En efecto, al estimar un modelo para Y_t se tiene que en el modelo se han fijado (véase cuadro 3.2) las variables explicativas, la estructura dinámica de las variables explicativas y la estructura dinámica residual, y sólo quedan por fijar las presuntas innovaciones —$\hat{a}_t$— durante el proceso de estimación.

Si el modelo no está equilibrado, para que ambas partes del modelo sean iguales o bien se terminará generando unas innovaciones estimadas $\hat{a}_t$ que violan la hipótesis de ruido blanco, o bien se concluirá con funciones de respuesta estimadas que incumplen las condiciones impuestas en el epígrafe octavo.

Así por ejemplo, si las presuntas innovaciones estimadas no son ruido blanco, pero al menos constituyen una serie temporal estacionaria, es necesario incorporar al modelo más estructura dinámica de carácter estacionario: funciones de respuesta más amplias en las variables explicativas incluidas, inclusión de variables explicativas omitidas y/o estructuras más complejas para el término residual.

Pero incluso puede ocurrir que la serie temporal formada por las innovaciones estimadas no sea estacionaria, lo que indicará que la regresión puede ser espúrea, que se han omitido variables explicativas capaces de explicar el comportamiento no estacionario residual, o que debe formularse sobre variables diferenciadas.

Especificando el modelo en términos de función de transferencia tal y como se ha indicado, se puede aplicar una secuencia de contrastes de hipótesis sobre el modelo general de partida para analizar si es posible llegar a un modelo más simple capaz de explicar los datos utilizados. Por lo tanto, la estrategia de lo general a lo

[38] Sobre PC-GIVE véase Hendry (1989, 1991) o Ericsson et al. (1991).

[39] Sobre modelos equilibrados (balanced models) véase Granger (1990), páginas 12-13.

particular no sólo es posible, sino muy recomendable, en modelos econométricos formulados como funciones de transferencia.

Cuando *no existe información a priori* sobre las características de las funciones de transferencia que relacionan a la variable dependiente con cada una de las explicativas, la formulación del modelo econométrico en términos de función de transferencia suele ser muy arbitraria, y constituir un mal punto de partida. En tal caso, todavía es posible una formulación dinámica del modelo adecuada a los datos, que es la recogida en (3.10.3), y que consiste en aplicar polinomios bastante generales a cada una de las variables que entran en el modelo, para asegurar así que el residuo es ruido blanco. En los próximos párrafos supondremos que hay más de una variable explicativa, con el fin de que el lector pueda apreciar más claramente el razonamiento.

La consecuencia de buscar una especificación del modelo del tipo (3.10.3)-(3.10.4) es que la estructura dinámica que se obtenga puede no tener ahora interpretación económica: los polinomios $\hat{\alpha}(L)$ y $\hat{\beta}(L)$ a los que conduzcan los datos tenderán a ser más simples que los correspondientes polinomios teóricos, ya que éstos normalmente contienen factores comunes. Por ello, determinar funciones de respuesta a partir de (3.10.4) será, en general, muy difícil.

Esto se debe a que la modelización de la función de respuesta no se plantea a partir de una especificación inicial basada en la Teoría Económica, sino que se impone una estructura dinámica a ciegas, sin información extramuestral, con el solo propósito llegar a un modelo en el que los residuos sean ruido blanco. Si bien es cierto que esta manera de proceder garantiza que la forma dinámica introducida pueda ser aceptable para la muestra empleada, es bastante probable también que simplifique en exceso las estructuras polinomiales teóricas.

En el modelo verdadero y, por tanto, desconocido, las expresiones

$$\beta_j(L)/\alpha(L)$$

nos dan la función de respuesta de Y respecto a X_j y, obviamente, son iguales a

$$w_j(L)/\delta_j(L),$$

que es la expresión que se obtiene directamente de (3.8.1). Las funciones de respuesta coinciden porque, tal y como se ha puesto de manifiesto al principio de esta sección, los cocientes $\beta_j(L)/\alpha(L)$ se simplifican y conducen a los correspondientes cocientes $w_j(L)/\delta_j(L)$.

En la búsqueda de especificación y posterior estimación, sin embargo, no habrá sido en general posible obtener polinomios $\hat{\alpha}(L)$ y $\hat{\beta}_j(L)$ con la multiplicidad de factores comunes que hay en los verdaderos $\alpha(L)$ y $\beta_j(L)$, debido a los problemas de multicolinealidad que esa multiplicidad ocasiona; en consecuencia, es posible que los cocientes

$$\hat{\beta}_j(L)/\hat{\alpha}(L)$$

no recojan la correspondiente función de respuesta.

La validez de la aproximación $\hat{\beta}(L)/\hat{\alpha}(L)$ a la función de respuesta teórica dependerá de los polinomios $\delta_j(L)$ y $\phi(L)$ en (3.8.1) y, especialmente, de que $\phi(L)$ tenga o no raíces negativas o complejas.

En efecto, se ha indicado anteriormente que los polinomios $\delta_j(L)$ deben, en general, imprimir una estructura suave a la función de respuesta, y que para ello puede valer con que tales polinomios sean de orden uno con raíz positiva. En consecuencia las diferencias entre los distintos $\delta_j(L)$ estarán fundamentalmente en la magnitud de la raíz: cuanto mayor sea ésta, más corta será la respuesta y viceversa. Por tanto, si en el modelo teórico (3.8.1) el residuo es ruido blanco, la aproximación

$$\hat{\alpha}(L) \simeq \delta_1(L) \simeq \cdots \simeq \delta_m(L)$$

puede ser bastante aceptable, teniendo en cuenta que los polinomios $\hat{\beta}_j(L)$ pueden compensar tal aproximación alterando la magnitud de sus parámetros respecto a los teóricos.

Si el residuo, denotado por n_t, en vez de ser ruido blanco es autorregresivo de primer orden, $n_t = [\phi(L)]^{-1} a_t$, también con raíz positiva, la aproximación anterior puede extenderse a $\hat{\alpha}(L) \simeq \phi(L)$, y continuar siendo muy aceptable la obtención de funciones de respuesta a partir del modelo (3.10.4) estimado. De hecho en este caso (3.10.4) es preferible a efectos de estimación.

Sin embargo, si el término residual estacionario en (3.8.1) tiene una estructura temporal muy oscilante, reflejando la existencia de oscilaciones en la variable dependiente que se deben a variables omitidas, su modelización requerirá esquemas autorregresivos de primer o segundo orden con raíces negativas o complejas. En tal caso el polinomio $\hat{\alpha}(L)$ estimado puede venir dominado por la estructura residual, y tener raíces que no sean reales y positivas.

En dicha situación al estimar un modelo del tipo (3.10.4), en el que el término residual teórico ha sido blanqueado, se está forzando a que tal oscilación se recoja en el polinomio $\hat{\alpha}(L)$ mediante raíces negativas o complejas, que generarán funciones de respuesta oscilantes. Para las variables en las que la función teórica de respuesta sea corta, tal oscilación temporal no será importante. Pero en variables con función teórica de respuesta larga, la distorsión del polinomio $\hat{\alpha}(L)$ puede ser importante y provocar la aparición de polinomios $\hat{\beta}_j(L)$ con parámetros que generen un mayor impacto inicial en la función de respuesta.

En conclusión, se puede decir que cuando no se tenga información para formular el modelo en forma de función de transferencia, la formulación en términos de ecuación en diferencias finitas estocástica permitirá una explicación adecuada de los datos. Ahora bien, la posibilidad de recuperar funciones de respuesta a partir de tal estimación dependerá de la naturaleza del elemento residual teórico (variables omitidas), y de la similitud de las estructuras autorregresivas de las distintas funciones teóricas de respuesta correspondientes a cada una de las variables explicativas.

Cuando no existe información para especificar la dinámica es conveniente que la formulación dinámica inicial proporcione un residuo ruido blanco. En otras palabras, se ha de llegar, en el caso de una sola variable explicativa, a un modelo del tipo (3.10.3), y no simplemente a un modelo como (3.10.2), en el que los residuos son meramente estacionarios. Esto último, como punto de partida, es peligroso, pues esa dinámica residual puede no ser tal, sino una forma restringida —y por tanto incorrecta en general— de poner mayor dinámica en la relación entre X e Y[40].

Como conclusión de la discusión que antecede se puede decir que cuando existe información a priori sobre las funciones de respuesta es conveniente construir los modelos econométricos a partir de la formulación en términos de funciones de transferencia. Si la información a priori no existe, el modelo econométrico todavía se puede formular *blanqueando el término residual*, es decir, empleando una estructura dinámica que genere residuos que sean ruido blanco. En ambos casos se estará partiendo de una especificación dinámica inicial general, sobre la que apoyar el procedimiento secuencial de contrastación de hipótesis de Anderson (1971), y se llegará, normalmente, a un modelo final más simple. Si se han empleado funciones de transferencia este modelo final podrá interpretarse en términos de función de respuesta; en caso contrario puede ocurrir que tal interpretación no sea posible.

Un ejemplo de modelo dinámico planteado como una ecuación

[40] Una exposición simple de este problema basada en los trabajos de Sargan, Hendry y Mizon se encuentra en Espasa (1978a). En este trabajo se señala también que la estructura dinámica residual, cuando es plenamente válida en el proceso de construcción del modelo, supone una forma eficiente de recoger factores comunes (COMFAC) en los polinomios $\alpha(L)$ y $\beta(L)$; sobre esto último véase Hendry y Mizon (1978).

en diferencias finitas estocástica se encuentra en la ecuación explicativa de la inversión de Andrés et al (1990). En este trabajo todo el proceso de búsqueda de especificación dinámica, estimación y validación se realiza adoptando desde el principio una formulación del tipo (3.10.4).

El modelo finalmente estimado es (página 153):

$$\begin{aligned}(1-0{,}46L+0{,}25L^2)I_t = & -2{,}73+(2{,}72-1{,}85L^2)Y_{t-1}+\\ & +(2{,}29-2{,}29L)CU_t+(-1{,}54+0{,}43L)\ (C/P)_{t-1}+\\ & +(-0{,}74+1{,}48L-0{,}74L^2)\pi_t+a_t \end{aligned}\qquad (3.10.5)$$

donde I representa el logaritmo de la inversión productiva privada, Y el logaritmo del PIB, CU el logaritmo del grado de utilización de la capacidad productiva, (C/P) el coste de uso del capital y π la tasa de inflación del deflactor del PIB.

El polinomio asociado a la variable endógena tiene raíces complejas que recogen un movimiento cíclico de período 5,74 años y factor de amortiguamiento 0,50. Nótese que este ciclo va a estar presente en todas las relaciones dinámicas de la inversión con las variables explicativas. Así por ejemplo, el polinomio temporal asociado a la renta viene dado por el desarrollo del cociente de polinomios

$$(2{,}72-1{,}85L^2)L/(1-0{,}46L+0{,}25L^2) \qquad (3.10.6)$$

cuya representación figura en el gráfico 3.3.

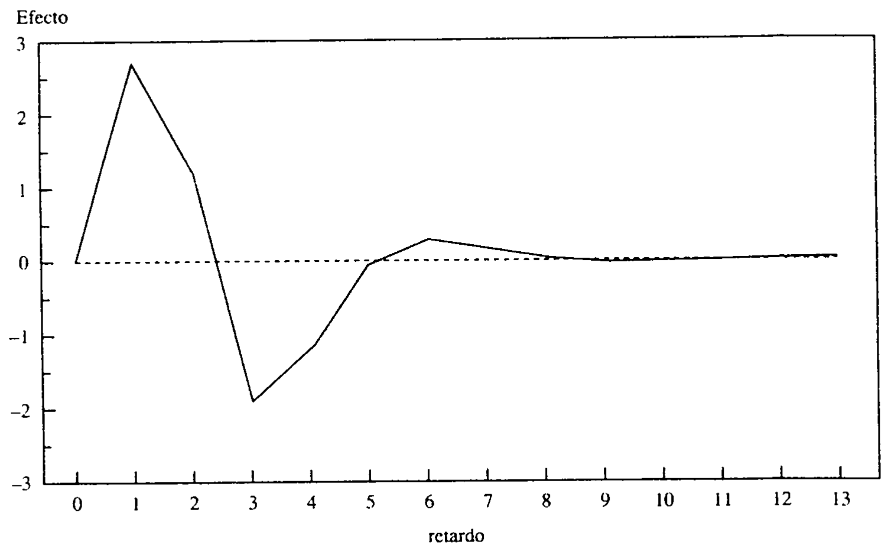
En el ejemplo que nos ocupa las variables explicativas no son exógenas y, por tanto, el concepto de función de respuesta, con sus correspondientes multiplicadores, carece de sentido. En cualquier caso, (3.10.6) recoge la relación dinámica entre inversión y PIB en la ecuación determinante de la inversión, y todavía cabe preguntarse si (3.10.6) está recogiendo adecuadamente la correspondiente relación dinámica teórica o si, por el contrario, la estimación que aparece en (3.10.6) se debe a: a) no haber especificado un componente residual autorregresivo; o b) a que sea(n) alguna(s) de las otras variables explicativas la(s) que requiere(n) una dinámica oscilante y, por tanto, la impone(n) al polinomio autorregresivo de la variable dependiente.

A la pregunta anterior se puede intentar responder volviendo a estimar el modelo. Supongamos que se ha estimado y validado el siguiente modelo:

$$\alpha(L)Y_t=\beta_1(L)X_{1t}+\beta_2(L)X_{2t}+a_t, \qquad (3.10.7)$$

y se presenta la duda de si $\alpha(L)$ es común a todas las funciones dinámicas, la de las dos variables explicativas y la residual, o sólo

GRÁFICO 3.3. *Función de respuesta de la inversión ante un cambio de tipo impulso en la renta.*



corresponde a alguna de ellas. En concreto, supóngase que se sospecha que no debe entrar en la relación dinámica con el primer regresor. En tal caso se podría estimar el modelo

$$Y_t = \beta_1^*(L) X_{1t} + \frac{\beta_2(L)}{\alpha(L)} X_{2t} + \frac{a_t}{\alpha(L)},$$

con la restricción de que los denominadores polinomiales de X_2 y de la innovación sean iguales.

Para concluir este epígrafe conviene hacer otra puntualización sobre las relaciones dinámicas. En el caso de variables exógenas a los polinomios $w_j(L)/\delta_j(L)$ se les ha interpretado como una función de respuesta, pero esto es correcto si la variable explicativa X_j es la variable teórica correspondiente. Si no es exactamente así, el polinomio $w_j(L)/\delta_j(L)$ se puede interpretar como la transformación que hay que operar sobre X_j para aproximar mejor a la variable teórica. En concreto, si la verdadera variable explicativa, X_j^*, tiene un comportamiento muy suave, pero la variable X_j disponible tiene, por ejemplo debido a la forma de medirla, una evolución con una oscilación que no está presente en X_j^*, el polinomio $w_j(L)/\delta_j(L)$ —o parte de él— se podría interpretar como el filtro necesario para pasar de X_j a X_j^*.

3.11. Cointegración y mecanismo de corrección del error

Durante la década de los setenta se produjo un cierto debate entre determinados económetras y los analistas de series temporales, centrado en la especificación de modelos referidos a variables no estacionarias.

La postura de estos últimos era la siguiente: si X_t y Y_t son series integradas de orden uno, es decir, X_t, Y_t no son estacionarias pero ΔX_t y ΔY_t sí lo son, entonces en una relación del tipo

$$Y_t = v(L) X_t + N_t$$

puede ocurrir que N_t sea integrada de orden uno.

Esto se manifestaba en el llamado problema de las regresiones espúreas: Granger y Newbold (1974), incidiendo en un tema que ya había apuntado previamente Yule (1926), mostraron como era posible obtener regresiones con un coeficiente de determinación R^2 muy alto relacionando variables independientes, sencillamente por el hecho de que ambas fueran no estacionarias. En ese caso el estadístico Durbin-Watson era muy bajo, indicando la alta correlación muestral positiva en los residuos el carácter no estacionario de éstos[41].

En la literatura econométrica de la época[42] era corriente encontrar ecuaciones con un R^2 muy alto y un estadístico *dw* pequeño, lo que parecía señalar que muchas de las relaciones que se presentaban eran casos similares al tratado por Granger y Newbold en el trabajo anteriormente mencionado. Esto además coincide en el tiempo con la verificación de que modelos univariantes sencillos tenían una capacidad predictiva superior a la de muchos macromodelos econométricos pese a manejar mucha menos información[43], lo que daba pie para pensar que estos macromodelos estaban incorrectamente especificados.

Como consecuencia de todo ello en la segunda mitad de los años setenta aparece una corriente que propone relacionar exclusivamente variables estacionarias. Tales autores propugnaban que en aquellas aplicaciones en que se manejasen variables no estacionarias, se tomasen previamente las diferencias necesarias para obtener variables estacionarias[44].

Esta postura fue contestada desde la perspectiva econométrica en una serie de trabajos que desarrollan las ideas expuestas origi-

[41] Sargan (1979) y Sargan y Bhargava (1983) desarrollan el contraste de la estacionariedad de los residuos a partir del estadístico Durbin-Watson.

[42] Véase por ejemplo Hickman (1972).

[43] Véase Nelson (1972).

[44] Pierce (1977) sirve como ejemplo al tipo de conclusiones que se pueden llegar con dicha forma de proceder.

nalmente en Sargan (1964). En efecto, para que la propuesta de diferenciar previamente las variables resultase recomendable con carácter general, tendría que ocurrir que cuando la relación se plantease en niveles se obtuviera siempre un residuo integrado del mismo orden que las variables consideradas. No obstante, en muchas ocasiones no sucede así, pues se tiene que al relacionar variables integradas de orden uno se originan residuos inequívocamente estacionarios; y esto entra en contradicción clara con la propuesta de aquellos analistas de series temporales que defendían que sólo se relacionaran variables estacionarias. Es más, en tal situación formular el modelo sobre variables diferenciadas sería incorrecto, pues generaría residuos con una media móvil no invertible.

La aportación de Sargan (1964) pasó a ser más conocida con el trabajo de Davidson et al. (1978), en el que se criticaba el enfoque de series temporales explicando más detalladamente que la diferenciación previa eliminaba la información sobre el comportamiento de largo plazo o tendencial de las variables, con lo que se perdía una parte importante de la información contenida en los datos.

Finalmente Granger (1981) introduce el concepto de *cointegración* y con él se establecen las bases para realizar un tratamiento matemático-estadístico profundo del tema, y que ha servido para aproximar posturas. La cointegración vuelve a traer a primera línea muchas de las ideas que expuestas en Sargan (1964), y que ya habían formado el núcleo de las críticas contenidas en Davidson et al. (1978).

Sean X, Y dos variables integradas de orden uno: se dice que están cointegradas cuando existe una combinación lineal, de la forma a_1X+a_2Y, que es estacionaria. En general, si X, Y son integradas de orden d, $I(d)$, entonces presentarán *cointegración de ordenes d y b*, $CI(d,b)$, si existe una combinación lineal de ambas que sea integrada de orden $d-b$, $0<b\leqslant d$.

Como ya se dijo en el capítulo anterior, a lo largo de todo el libro sólo se consideran procesos $I(d)$ con d un número natural, lo que excluye los modelos con diferencias fraccionales. Por lo tanto todo el análisis de la cointegración también se limita al caso en que tanto d como b son números naturales.

Si dos variables están cointegradas, su comportamiento a largo plazo está relacionado. Así, si X e Y son dos tipos de interés a distintos plazos que se generan en mercados eficientes, entonces lo más probable es que por separado sean variables $I(1)$, con todo lo que ello acarrea: un nivel medio cambiante, incertidumbre sobre el futuro no acotada, etc. En este caso, la Teoría Económica sugiere, y varios estudios empíricos parecen confirmarlo, que el diferencial $X-Y$, una relación lineal en la que $a_1=-a_2=1$, sigue una evolución estrictamente estacionaria. Por lo tanto, ambas variables guardan una relación similar a largo plazo o, en otras palabras, existe una relación de equilibrio entre los dos tipos de interés que se mantiene en el largo plazo, de modo que sobre el diferencial entre ambos hay una incertidumbre futura acotada, sin importar la lejanía en que se sitúe en el futuro.

La *cointegración* $CI(1, 1)$ es la que ha sido más tratada en la literatura especializada, y sobre ella se basará el desarrollo de este tema, de modo que cuando se hable de cointegración y no se exprese lo contrario se supondrá que se trata de variables $CI(1, 1)$. Al final de esta sección se harán una serie de comentarios para la extensión de la cointegración a series económicas que sean $I(2)$.

Para que dos variables estén cointegradas con orden $CI(1, 1)$, ha de existir una descomposición de ambas series del tipo:

$$Y_t=aW_t+Y_{1t}$$
$$X_t=W_t+X_{1t},$$

en donde Y_{1t} y X_{1t} son variables $I(0)$ y W_t es $I(1)$. Es decir, la no estacionariedad de X_t y Y_t viene generada por un factor común W_t.

A partir de aquí es inmediato comprobar que si definimos Y_t-aX_t, una combinación lineal de la forma $a_1X_t+a_2Y_t$ con $a_1=-a$ y $a_2=1$, se tiene que

$$Y_t-aX_t=aW_t+Y_{1t}-a(W_t+X_{1t})=Y_{1t}-aX_{1t}=n_t$$

que por definición es estacionaria. Por lo tanto

$$Y_t=aX_t+n_t$$

y el largo plazo de Y viene plenamente explicado por el largo plazo de X.

Esta simplificación se generaliza, y si las variables son $CI(d, d)$ ocurre que el largo plazo de una variable viene plenamente explicado por el largo plazo de la otra. A la cointegración $CI(d, d)$ se le puede denominar *cointegración plena*, para distinguirla de la cointegración $CI(d, b)$ con $b<d$, que se caracteriza porque en la relación

$$Y_t=aX_t+N_t,$$

N es $I(d-b)$ y por tanto no estacionario. En este caso el largo plazo de Y_t no viene plenamente explicado por X_t, ya que el elemento N_t contribuye también a dicho largo plazo. No obstante, X_t tiene una contribución decisiva —aunque no plena— en el largo plazo de Y_t, ya que X_t es de orden de integración d superior al orden $d-b$ de N_t.

La cointegración se observa con muchas variables económicas: por ejemplo, en el caso español Andrés et al. (1990) observan cointegración entre la inversión, el PIB y el coste de uso de capital con datos anuales. Dolado y Escrivá (1992) encuentran cointegración entre el nivel de precios, el producto interior bruto, la masa monetaria, la inflación y los tipos de interés.

Desde el punto de vista estrictamente económico, el estudio de la existencia o no de cointegración entre las variables del análisis es uno de los resultados básicos del proceso de modelización, ya que lo que se analiza es hasta qué grado se ha conseguido captar el comportamiento a largo plazo de la variable a explicar mediante las variables explicativas que se han considerado en el proceso de modelización.

Conviene recalcar aquí que si se están considerando dos variables X e Y, el que no se detecte ningún tipo de cointegración entre ambas no indica necesariamente que las dos variables no estén relacionadas a largo plazo mediante una relación de equilibrio. Si en dicho equilibrio a largo plazo interviene una tercera variable Z que no se ha tenido en cuenta, entonces es evidente que la perturbación está reflejando esa omisión y por lo tanto no será estacionaria.

Precisamente por eso el que al final se tenga un residuo no estacionario se puede tomar como indicación de que se ha omitido una variable relevante.

Una característica importante de los sistemas formados por variables cointegradas es que admiten la llamada representación en términos del *mecanismo de corrección del error* (MCE)[45]: sean dos variables X, Y que son $CI(1, 1)$, y supóngase sin pérdida de generalidad que la relación de cointegración o de largo plazo es $Y_t - gX_t$. Si su relación dinámica viene recogida por el modelo en términos de *ecuación en diferencias finitas* (3.10.3), que para mayor comodidad en la exposición se vuelve a recoger aquí:

$$\alpha(L)\, Y_t = \beta(L) X_t + a_t \tag{3.11.1}$$

se tiene que dicha relación a largo plazo es

$$Y_t = \frac{\beta(1)}{\alpha(1)} X_t.$$

[45] El mecanismo de corrección del error fue propuesto originalmente en Phillips (1957) y Sargan (1964). Este resultado se deduce del teorema de representación de Granger: véase Engle y Granger (1987). Véase también Escribano (1990) y demás artículos recogidos en el volumen monográfico de Cuadernos Económicos de ICE (número 44), y Dolado (1990, 1992a,b).

Llamando β a la ganancia —$\beta(1)$— del polinomio $\beta(L)$, y α a la ganancia —$\alpha(1)$— del polinomio $\alpha(L)$, tenemos que

$$g = \beta/\alpha$$

es la ganancia a largo plazo de X sobre Y. En lo sucesivo, dado un polinomio $a(L)$ llamaremos a a su ganancia, $a(1)$.

Para formular (3.11.1) en términos del mecanismo de corrección del error obsérvese que cualquier polinomio $a(L)$ se puede escribir como

$$\mathbf{a(L) = a + a^*(L)\,(1 - L)} \tag{3.11.2}$$

Para ello sólo hay que arreglar los términos a la derecha de (3.11.2) para que se cumpla la igualdad.

En (3.11.2) el polinomio $a^*(L)\,(1-L)$ tiene ganancia nula; $a^*(L)$, que se aplica sobre una serie diferenciada, recibe el nombre de *polinomio dinámico transitorio*.

Como ejemplo de lo anterior, sea

$$a(L) = 1 - a_1 L - a_2 L^2,$$

su ganancia es

$$a = 1 - a_1 - a_2,$$

y la formulación (3.11.2) toma la forma:

$$a(L) = a + (1 - L)\,[(a_1 + a_2) + a_2 L].$$

Si en (3.11.1) se formula $\alpha(L)$ como

$$\alpha(L) = 1 - L\tilde{\alpha}(L)$$

y se escribe $\tilde{\alpha}(L)$ utilizando (3.11.2) se obtiene

$$\alpha(L) = 1 - L\tilde{\alpha} - L\tilde{\alpha}^*(L)\,(1 - L) \tag{3.11.3}$$

Utilizando (3.11.3) y escribiendo $\beta(L)$ de acuerdo con (3.11.2), el modelo (3.11.1) se puede escribir como

$$Y_t = \tilde{\alpha}\, Y_{t-1} + \tilde{\alpha}^*(L)\Delta Y_{t-1} + \beta X_t + \beta^*(L)\Delta X_t + a_t. \tag{3.11.4}$$

Restando a ambos lados de (3.11.4) Y_{t-1}, y sumando y restando βX_{t-1} en el término de la derecha de (3.11.4) y reagrupando términos, se obtiene

$$\Delta Y_t = \bar{\alpha}^*(L)\Delta Y_{t-1} + \beta\Delta X_t + \beta^*(L)\Delta X_t - $$
$$-(1-\bar{\alpha})\left(Y_{t-1} - \frac{\beta}{1-\bar{\alpha}} X_{t-1}\right) + a_t \qquad (3.11.5)$$

Es inmediato comprobar que en (3.11.5)

$$\beta/(1-\bar{\alpha}) = \beta/\alpha = g,$$

con lo que denominando

$$\beta^{**}(L) = \beta^*(L) + \beta$$

se tiene que

$$(A) \quad \Delta Y_t = \bar{\alpha}^*(L)\Delta Y_{t-1} + \beta^{**}(L)\Delta X_t - \alpha[Y_{t-1} - gX_{t-1}] + a_t \qquad (3.11.6)$$

que es la formulación del modelo (3.11.1) en términos de mecanismo de corrección del error.

En (3.11.6) todas las variables que entran en el modelo son estacionarias. En este modelo se explica el incremento de Y_t en función de una estructura dinámica sobre incrementos pasados y presente de X e incrementos pasados de Y, más la desviación de Y_{t-1} respecto a la ganancia total g aplicada a X_{t-1}. El modelo (3.11.6) aparentemente recoge una relación de X e Y en diferencias, pero realmente no es así, ya que en (3.11.6) se puede despejar Y_t y obtener el modelo original (3.11.1): que (3.11.6) recoge una relación entre niveles viene explícitamente señalado por la presencia de las variables en niveles Y_{t-1} y X_{t-1}. El modelo (3.11.6) es una formulación estacionaria de una relación de variables no estacionarias, haciendo uso de la restricción a largo plazo que las liga a ambas por el hecho de estar cointegradas.

Este modelo pone de manifiesto que cuando hay más de una variable y hay cointegración la transformación estacionaria no se logra diferenciando, sino que requiere la inclusión del mecanismo de corrección del error junto con términos en diferencias.

La plena cointegración conlleva que los residuos del modelo (3.11.1) sean estacionarios[46], lo que implica a su vez que una formulación del modelo (3.11.1) sobre variables diferenciadas conduciría a un modelo con una media móvil no invertible en los residuos. Habiendo plena cointegración no se puede formular un modelo invertible usando sólo variables diferenciadas: la presencia de cointegración exige que la formulación estacionaria incluya variables en niveles, y la formulación en términos de mecanismo de corrección del error es uno de los modos más útiles de llevarla a cabo.

[46] Obsérvese que con variables $I(1)$ la cointegración, si existe, siempre es plena.

En conclusión, se puede decir:

1) La ecuación (3.11.6) expresa una relación en niveles, ya que aparecen ambas variables, X e Y, sin diferenciar.

2) El término $Y_{t-1} - gX_{t-1}$ representa la desviación entre el valor observado de la variable dependiente en $t-1$ y el valor que resulta de aplicar la ganancia total al valor de la variable explicativa en $t-1$; esta desviación está necesariamente afectada por un coeficiente negativo, de tal forma que, por ejemplo, cuando la discrepancia en $t-1$ sea positiva, su contribución al incremento de Y en el momento t será negativa.

3) La restante dinámica de la relación está parametrizada en términos de ΔY y ΔX, que siendo X e Y variables $I(1)$, sólo tienen efectos transitorios sobre el nivel de Y_t. Por eso se dice que los polinomios $\bar{\alpha}^*(L)$ y $\beta^{**}(L)$ junto con el mecanismo de corrección del error recogen una dinámica transitoria.

Obsérvese que cuando las variables X e Y son cointegradas de orden $CI(1, 1)$, al menos uno de los modelos uniecuacionales, de X o de Y, tomará la forma (3.11.6), tal y como se demuestra en el teorema de representación de Granger (Engle y Granger, 1987).

Considérese ahora el caso en que X e Y vienen relacionadas por el siguiente *modelo de función de transferencia*:

$$Y_t = \frac{w(L)}{\delta(L)} X_t + n_t \qquad (3.11.7)$$

en donde si X e Y están cointegradas $CI(1, 1)$, el residuo n_t es estacionario y se puede suponer que viene generado por el siguiente proceso ARMA,

$$n_t = \frac{\theta(L)}{\phi(L)} a_t.$$

Operando en (3.11.7) se tiene que

$$\delta(L)Y_t = w(L)X_t + n_t^*, \qquad (3.11.8)$$

donde n_t^* es igual a $\delta(L)n_t$ y por tanto también es estacionario.

Procediendo con los polinomios $w(L)$ y $\delta(L)$ de forma similar a como se ha hecho previamente con los polinomios $\beta(L)$ y $\alpha(L)$, se obtiene para (3.11.8) la siguiente formulación en términos de mecanismo de corrección del error:

$$\text{(A)}\quad \Delta Y_t = \bar{\delta}^*(L)\Delta Y_{t-1} + w^{**}(L)\Delta X_t - \delta[Y_{t-1} - gX_{t-1}] + n_t^*, \qquad (3.11.9)$$

en donde

$$g = \frac{w(1)}{\delta(1)} = \frac{w}{1-\bar{\delta}} = \frac{w}{\delta}. \qquad (3.11.10)$$

Así mismo, formulando la función de transferencia como

$$\frac{w(L)}{\delta(L)} = g + \frac{w^+(L)}{\delta(L)}(1-L), \qquad (3.11.11)$$

donde g viene definida por (3.11.10), se tiene que

$$Y_t = \left(g + \frac{w^+(L)}{\delta(L)}(1-L)\right) X_t + n_t, \qquad (3.11.12)$$

con lo que restando Y_{t-1} a ambos lados de (3.11.12), y sumando y restando gX_{t-1} en el término de la derecha de (3.11.12), se obtiene

$$\text{(B)}\quad \Delta Y_t = \frac{w^{++}(L)}{\delta(L)}\Delta X_t - (Y_{t-1} - gX_{t-1}) + n_t, \qquad (3.11.13)$$

donde

$$w^{++}(L) = w^*(L) + g\delta(L). \qquad (3.11.14)$$

Es importante recalcar aquí que (A) y (B) son dos formas alternativas de representar el mecanismo de corrección del error. Más adelante se discutirá este punto en detalle, pero la diferencia básica entre ambas es la asignación a distintas fuentes, mecanismo de corrección del error o los polinomios sobre ΔX_t y ΔY_{t-1}, de la dinámica transitoria.

Para ilustrar todos los desarrollos que se han hecho hasta este momento, considérese el siguiente ejemplo donde $w(L)$ y $\delta(L)$ son de orden uno:

$$Y_t = \frac{w_0 + w_1 L}{1 - \delta_1 L} X_t + n_t, \qquad (3.11.15)$$

Las formulaciones alternativas del mecanismo de corrección del error, representadas en (3.11.9) y (3.11.13), correspondientes a (3.11.15) son:

$$\text{(A)}\quad \Delta Y_t = w_0 \Delta X_t - (1-\delta_1)(Y_{t-1} - gX_{t-1}) + n_t^* \qquad (3.11.16)$$

$$\text{(B)}\quad \Delta Y_t = \left(\frac{w_0 - g\delta_1 L}{1 - \delta_1 L}\right)\Delta X_t - (Y_{t-1} - gX_{t-1}) + n_t \qquad (3.11.17)$$

En (3.11.16) se tiene que el residuo n_t^* viene dado por

$$n_t^* = n_t - \delta_1 n_{t-1} = n_t - \delta_1\left(Y_{t-1} - \frac{w(L)}{\delta(L)} X_{t-1}\right) \qquad (3.11.18)$$

con lo que, pasando

$$-\delta_1\left(Y_{t-1} - \frac{w(L)}{\delta(L)} X_{t-1}\right)$$

del término n_t^* a los otros componentes que aparecen en el término de la derecha de (3.11.16), se llega a (3.11.17).

Volviendo a la discusión en términos generales, el mecanismo de corrección del error se puede formular de más de una manera también para el modelo (3.11.1). Hasta el momento se ha expresado dicho mecanismo para este tipo de modelos en lo que hemos llamado la forma (A) —véase (3.11.6)—; pero siguiendo un desarrollo similar al que nos permitió pasar de (3.11.9) a (3.11.13), sería igualmente factible expresar el mecanismo de corrección del error correspondiente a (3.11.1) como

$$\text{(B)}\quad \Delta Y_t = \frac{\beta^{++}(L)}{\alpha(L)}\Delta X_t - (Y_{t-1} - gX_{t-1}) + \frac{a_t}{\alpha(L)}, \qquad (3.11.19)$$

que es el equivalente de (3.11.13) para los modelos formulados como ecuación en diferencias finitas estocástica, y en donde

$$\beta^{++}(L) = \beta^{**}(L) - gL\alpha^*(L) = \beta^+(L) + g\alpha(L)$$
$$\alpha(L) = 1 - L\tilde{\alpha}(L).$$

en donde $\beta^+(L)$ se ha obtenido aplicando (3.11.11) a $\beta(L)/\alpha(L)$.

Así, por ejemplo, para el modelo

$$(1 - \alpha_1 L) Y_t = (\beta_0 + \beta_1 L) X_t + a_t \qquad (3.11.20)$$

las dos expresiones mencionadas del mecanismo de corrección del error son:

$$\text{(A)} \qquad \Delta Y_t = \beta_0 \Delta X_t - (1-\alpha_1)(Y_{t-1} - gX_{t-1}) + a_t \qquad (3.11.21)$$

$$\text{(B)} \qquad \Delta Y_t = \frac{\beta_0 - g\alpha_1 L}{1-\alpha_1 L} \Delta X_t - (Y_{t-1} - gX_{t-1}) + \frac{a_t}{1-\alpha_1 L} \qquad (3.11.22)$$

El resumen 3.1 recoge de manera sucinta las diferentes formulaciones consideradas.

El desarrollo anterior ha permitido llegar a *dos formas de expresar el mecanismo de corrección del error*, que para distinguirlas denominaremos (A) y (B):

(A) Formulación con un residuo que corresponde al residuo del modelo puesto en términos de ecuaciones en diferencias finitas, y un mecanismo de corrección del error cuyo coeficiente es la ganancia del polinomio del denominador de la función de transferencia en niveles cambiada de signo[47].

En (3.11.1) tal residuo es ruido blanco, y la formulación (3.11.6) en términos de mecanismo de corrección del error tiene un residuo que es ruido blanco. En (3.11.8) se ha partido de un modelo en que el residuo —n_t^*— no es ruido blanco y su correspondiente mecanismo de corrección del error (3.11.9) tiene como residuo n_t^*. En tales casos el mecanismo de corrección del error tiene un coeficiente en valor absoluto menor que la unidad. Ejemplos de esta formulación se encuentran en (3.11.21) y (3.11.16).

En la formulación (A) es importante partir de un modelo en el que los polinomios Y_t y X_t no tengan raíces comunes. Si las hay convendrá pasarlas al término residual, en cuyo caso no será ruido blanco. Para mantener los polinomios de Y_t y X_t como primos es por lo que en (3.11.8) se utiliza el residuo n_t^*. Si tales polinomios son primos la formulación (A) se realiza explicitando dos parámetros de especial interés: la ganancia g en la relación a largo plazo y la ganancia del polinomio en el denominador de la función de transferencia —α en el caso del modelo (3.11.1) y δ en el modelo (3.11.8)—. Esta última ganancia es importante porque, tal y como se ha visto en la sección 3.9, es la que determina la rapidez o lentitud con que la variable Y acaba incorporando un cambio en X.

(B) Formulación con un residuo que recoge las desviaciones de Y_t sobre la función de transferencia respecto a X_t y un mecanismo de

[47] Como se explica más adelante es importante que este polinomio no tenga factores comunes con el del numerador.

Resumen 3.1
FORMULACIONES DEL MECANISMO DE CORRECCION DEL ERROR

1. Modelo de ecuación en diferencias finitas estocástica

$$\alpha(L)\, Y_t = \beta(L) X_t + a_t \qquad (3.11.1)$$

$$Y_t = \frac{\beta(L)}{\alpha(L)} X_t + \frac{a_t}{\alpha(L)}$$

En (3.11.1) a_t es ruido blanco, pero las desviaciones de Y_t sobre la función de transferencia de X_t son $a_t/\alpha(L)$ y en consecuencia no son ruido blanco.

Formulación (A)

$$\Delta Y_t = \bar{\alpha}^*(L)\Delta Y_{t-1} + \beta^{**}(L)\Delta X_t - (1-\bar{\alpha})[Y_{t-1} - gX_{t-1}] + a_t \qquad (3.11.6)$$

$$\alpha(L) = 1 - L\bar{\alpha}(L) = 1 - L\bar{\alpha} - L\bar{\alpha}^*(L)\,(1-L)$$

$$\beta(L) = \beta + \beta^*(L)\,(1-L)$$

$$\beta^{**}(L) = \beta^*(L) + \beta$$

donde para cualquier polinomio $a(L)$, $a = a(1)$. Además, g es el multiplicador a largo plazo de Y sobre X. En este caso, $g = \beta(1)/\alpha(1) = \beta/\alpha$.

Formulación (B)

$$\Delta Y_t = \frac{\beta^{++}(L)}{\alpha(L)} \Delta X_t + (Y_{t-1} - gX_{t-1}) + \frac{a_t}{\alpha(L)} \qquad (3.11.19)$$

$\beta^{++}(L)$ se obtiene de la siguiente forma

$$\beta^{++}(L) = \beta^+(L) + g\alpha(L) = \beta^{**}(L) - gL\alpha^*(L)$$

$$\frac{\beta(L)}{\alpha(L)} = g + \frac{\beta^+(L)}{\alpha(L)}(1-L)$$

2. Modelo de función de transferencia

$$Y_t = \frac{w(L)}{\delta(L)} X_t + n_t \qquad (3.11.7)$$

$$\delta(L) Y_t = w(L) X_t + \delta(L) n_t \qquad (3.11.8)$$

Formulación (A)

$$\Delta Y_t = \bar{\delta}^*(L) \Delta Y_{t-1} + w^{**}(L) \Delta X_t - (1-\bar{\delta})[Y_{t-1} - g X_{t-1}] + n_t^* \qquad (3.11.9)$$

donde

$$\delta(L) = 1 - L\bar{\delta}(L)$$

$$\delta(L) = 1 - L\bar{\delta} - L\bar{\delta}^*(L)\,(1-L)$$

$$w^{**}(L) = w^*(L) + w$$

$$g = w(1)/\delta(1)$$

$$n_t^* = \delta(L) n_t$$

Formulación (B)

$$\Delta Y_t = \frac{w^{++}(L)}{\delta(L)} \Delta X_t - (Y_{t-1} - g X_{t-1}) + n_t \qquad (3.11.13)$$

$$\frac{w(L)}{\delta(L)} = g + \frac{w^{+}(L)}{\delta(L)} (1-L)$$

$$w^{++}(L) = w^{+}(L) + g\delta(L)$$

$$w^{**}(L) = w^{++}(L) + gL\delta^*(L)$$

Nota: Si el orden de $w(L)$ es s y el de $\delta(L)$ es r y $s \leqslant r$, entonces en $w^{++}(L)$, que en tal caso es de orden r, hay restricciones en los coeficientes.

corrección del error con coeficiente igual a menos uno. La expresión (3.11.19) recoge esta formulación para el modelo en términos de (3.11.1) y la expresión (3.11.13) para el modelo en la forma (3.11.7). Ejemplos concretos de esta formulación se en cuentran en (3.11.22) y (3.11.17). En la formulación (B) la ganancia del denominador de la función de transferencia no se explicita, pues dicho polinomio no se pierde en (B), sino que continúa apareciendo en la función de transferencia sobre ΔX_t.

La formulación (A) del mecanismo de corrección del error es preferible cuando hay que estimar el modelo, por los argumentos mencionados previamente sobre la mayor sencillez de estimar un modelo en forma de ecuación en diferencias finitas estocástica, respecto a la estimación de la forma en función de transferencia. Además, si $\beta(L)$ o $w(L)$ son de orden s y sus correspondientes polinomios $\alpha(L)$ o $\delta(L)$ son de orden r y $s \leqslant r$, entonces el número de parámetros de la formulación (B) es $2r$ y por tanto mayor que $s+r$, con lo que en $\beta^{++}(L)$ o $w^{++}(L)$ existen restricciones que hay que tenerlas en cuenta en la estimación.

Las formulaciones del mecanismo de corrección del error sirven para poner de manifiesto que en la evolución de Y existe una trayectoria de largo plazo, la que se obtiene despejando Y en el mecanismo de corrección del error, y una dinámica transitoria sin efecto a largo plazo, recogida por los polinomios sobre ΔY_{t-1}, ΔX_t y el mecanismo de corrección del error.

Al mecanismo de corrección del error se le puede asignar un coeficiente igual a -1, o un coeficiente también negativo pero menor en valor absoluto, dependiendo de qué formulación se considere. Según sea este coeficiente así serán los efectos sobre ΔY_t de la dinámica transitoria debida a los polinomios sobre ΔY_{t-1} y ΔX_t, ya que el efecto global ha de ser el mismo en ambos casos. Por ello no es fácil darle al coeficiente del mecanismo de corrección del error una interpretación económica, ya que depende de la división concreta de los efectos transitorios por la que el analista haya optado. En ausencia de información extramuestral que imponga una especificación concreta de la dinámica transitoria, cualquier interpretación está condicionada a la formulación de la misma que el analista subjetivamente haya considerado preferible presentar.

Lo anterior se ve claramente con un ejemplo. Sea el modelo (3.11.20) con sus formulaciones (A) y (B) del mecanismo de corrección del error recogidos en (3.11.21) y (3.11.22). Supóngase que X e Y son *variables* $I(1)$ y que hasta $t-2$ el sistema está en equilibrio,

$$X = X^e$$
$$Y = gX^e = Y^e$$

y en $t-1$ la variable X sufre una perturbación unitaria transitoria, de modo que en t vuelve al valor X^e. Con ello, a partir de (3.11.20), observando que los residuos son cero, se tiene

$$\begin{aligned} Y_{t-1} &= Y^e + \beta_0 \\ Y_t &= Y^e + \alpha_1\beta_0 + \beta_1 \\ Y_{t+1} &= Y^e + \alpha_1^2\beta_0 + \alpha_1\beta_1 \\ &\cdots\cdots\cdots\cdots \\ Y_{t+k} &= Y^e + \alpha_1^{k+1}\beta_0 + \alpha_1^k\beta_1 . \end{aligned} \tag{3.11.23}$$

Como $|\alpha_1|<1$, Y_{t+k} tiende a Y^e cuando k tiende a infinito.

En (3.11.23) se observa que la desviación de Y_{t+k} sobre su valor de equilibrio, Y^e, debido a la desviación sufrida por X en $t-1$ es $\alpha_1^k(\alpha_1\beta_0+\beta_1)$, con lo que

$$\begin{aligned} \Delta Y_{t+k} &= (1-\alpha_1)\alpha_1^{k-1}(\alpha_1\beta_0+\beta_1) = \\ &= (1-\alpha_1)\alpha_1^k\beta_0 - (1-\alpha_1)\alpha_1^{k-1}\beta_1 \end{aligned} \tag{3.11.24}$$

En el cuadro 3.4 se recoge cómo explican las formulaciones (A) y (B) los valores de ΔY_t. En él se observa que en la formulación (A) muy pronto toda la explicación descansa exclusivamente en el mecanismo de corrección del error. Sin embargo, la formulación (B) considera que hay una relación dinámica a corto plazo entre X e Y además de un efecto unitario del mecanismo de corrección del error. En ambos casos, la esperanza matemática

$$E(\Delta Y_t / I_{t-1}, X_t)$$

donde I_{t-1} recoge toda la información hasta t, es, obviamente, la misma y la posible cuestión es cómo contribuyen a ella el mecanismo de corrección del error y la restante dinámica transitoria del sistema.

Las formulaciones (A) y (B) indican que esa división en cómo explicar ΔY_{t+k} no es única y que, por tanto, la forma elegida puede ser arbitraria. De hecho, en el ejemplo el equilibrio se volverá a recobrar cuando de nuevo ΔY sea cero, pero los valores de ΔY son exactamente los mismos en (A) que en (B), por lo que ambos modelos se acercan igual —como no podía ser de otra forma— al equilibrio.

Por tanto, el cómo se formule el mecanismo de corrección del error no parece tener ningún significado económico especial. El *acercamiento al equilibrio*, como ya se ha indicado y se discutió con detalle en la sección 3.9, *depende básicamente del denominador de la función de transferencia en niveles.*

La formulación (A), correspondiente a un modelo en el que los

CUADRO 3.4. *Explicación de la evolución de Y en $(1-\alpha_1 L)Y_t = (\beta_0+\beta_1 L)X_t + a_t$ ante un cambio transitorio en X ocurrido en $t-1$*

Valor de ΔY	Explicación del valor de ΔY según los distintos esquemas del MCE			
	ESQUEMA (A) Dinámica transitoria debida a		ESQUEMA (B) Dinámica transitoria debida a	
	$\beta_0\Delta X_t$	$-MCE$	$\frac{\beta_0 - g\alpha_1 L}{1-\alpha_1 L}\Delta X_t$	$-MCE$
$\Delta Y_{t-1} = \beta_0$	β_0	(a)	β_0	(a)
$\Delta Y_t = -(1-\alpha_1)\beta_0 + \beta_1$	$-\beta_0$	$\alpha_1\beta_0+\beta_1$	$-(1-\alpha_1)\beta_0 - g\alpha_1$	$-\beta_0+g$
$\Delta Y_{t+1} = -(1-\alpha_1)\alpha_1\beta_0 - (1-\alpha_1)\beta_1$	—	$-(1-\alpha_1)\alpha_1\beta_0-(1-\alpha_1)\beta_1$	$-(1-\alpha_1)\alpha_1\beta_0+(\beta_0+\beta_1)\alpha_1$	$-\alpha_1\beta_0-\beta_1$
$\Delta Y_{t+2} = -(1-\alpha_1)\alpha_1^2\beta_0 - (1-\alpha_1)\alpha_1\beta_1$	—	$-(1-\alpha_1)\alpha_1^2\beta_0-(1-\alpha_1)\alpha_1\beta_1$	$-(1-\alpha_1)\alpha_1^2\beta_0+(\beta_0+\beta_1)\alpha_1^2$	$-\alpha_1^2\beta_0-\alpha_1\beta_1$
....	...	....	....	....
$\Delta Y_{t+k} = -(1-\alpha_1)\alpha_1^k\beta_0 - (1-\alpha_1)\alpha_1^{k-1}\beta_1$	—	$-(1-\alpha_1)\alpha_1^k\beta_0-(1-\alpha_1)\alpha_1^{k-1}\beta_1$	$-(1-\alpha_1)\alpha_1^k\beta_0+(\beta_0+\beta_1)\alpha_1^k$	$-\alpha_1^k\beta_0-\alpha_1^{k-1}\beta_1$

(a) El mecanismo de corrección del error no actúa desde el principio.

polinomios sobre Y y X no tienen factores comunes, recoge en el coeficiente del mecanismo de corrección del error la ganancia, con signo cambiado, de dicho polinomio. La formulación (B) mantiene dicho polinomio en la función de transferencia que incluye sobre ΔX_t. Pero en ambos casos hay que tener presente que el equilibrio se alcanza en variables $CI(1,1)$ cuando ΔY colapsa a cero si las variables son $I(1,0)$, o aun valor fijo, que no se altera por las perturbaciones que sufre el sistema, si las variables son $I(1,1)$.

El mecanismo de corrección del error tiene la ventaja de incorporar más información al modelo explicitando la relación de largo plazo entre las variables, pero no incorpora información alguna sobre la dinámica transitoria. Esta información tiene la ventaja de poder restringir fácilmente el modelo para que g tenga un valor dado por la Teoría Económica, por ejemplo la unidad. De hecho la formulación del mecanismo de corrección del error en Sargan (1964) se realiza principalmente para imponer una elasticidad unitaria ($g=1$) a largo plazo entre precios y salarios, en un modelo de determinación de los salarios nominales.

El ejemplo recogido en el cuadro 3.4 pone también de manifiesto que sin haberse modificado el largo plazo de X —ya que a partir de t vuelve al valor de equilibrio— y, por tanto, sin haberse modificado el largo plazo de Y, el mecanismo de corrección del error puede estar operando durante bastante tiempo.

Supóngase en el ejemplo anterior que X viene generada por un sendero aleatorio. En $t-1$ se estimaría que el largo plazo de X se había modificado de forma permanente, ya que su esperanza condicional a largo plazo pasaba a ser X^e+1. Sin embargo, el mecanismo de corrección del error no actúa en ese momento en ninguna de las dos formulaciones expuestas anteriormente. Por el contrario, en t el largo plazo de X vuelve a ser el de siempre en el ejercicio utilizado, X^e, y es a partir de ese momento cuando se desencadena el efecto del mecanismo de corrección del error.

Lo anterior pone de manifiesto que gX_{t-1} no es en sentido estricto el valor a largo plazo de Y, pues X_{t-1} no es, en general, el valor a largo plazo de X[48]. Esto sugiere que, al menos en el caso en que X es fuertemente exógena, *el mecanismo de corrección del error podría formularse utilizando en cada momento t el valor de largo plazo de X*, estimado como la esperanza matemática a largo plazo de X condicional a la información hasta t a partir, por ejemplo, del modelo ARIMA que genera a X. Esto supondría incorporar más información en el modelo y, por tanto, tal formulación podría ser más eficiente.

[48] X_{t-1} es el valor de largo plazo de X tal y como se estima en $t-1$ en el caso en que X venga generada por un sendero aleatorio.

La presencia de variables no estacionarias plantea problemas a la hora de estimar un modelo sobre las variables en niveles, ya que aunque los coeficientes de la relación a largo plazo se estiman de forma superconsistente (Stock, 1987), la inferencia sobre los mismos es compleja debido a que los estimadores no siguen las distribuciones asintóticas habituales. Estudiar la problemática de la estimación en tales casos supera con mucho los objetivos de este capítulo, si bien en la sección siguiente se realizan una serie de comentarios al respecto.

Aquí, baste señalar que cuando las variables explicativas son fuertemente exógenas entonces el modelo, tanto formulado en niveles como en mecanismos de corrección del error, se puede estimar por procedimientos uniecuacionales de minimizar la suma de cuadrados. Además los estimadores tienen las propiedades óptimas y se pueden aplicar los procedimientos habituales de contrastación de hipótesis.

Hasta ahora sólo se ha hablado de cointegración plena $CI(1,1)$. Cuando las *variables son* $I(2)$ la cointegración puede ser de dos tipos, $CI(2,1)$ y $CI(2,2)$. Pero aun ciñéndonos a la cointegración plena $CI(2,2)$, los problemas son más complejos que los mencionados hasta ahora, ya que la cointegración plena puede requerir que en la relación de cointegración aparezca una misma variable en niveles y en primeras diferencias. Cuando en dicha relación de largo plazo aparecen también las primeras diferencias, se dice que hay *cointegración polinomial*. Este es, por ejemplo, el caso de cointegración encontrado en el trabajo mencionado anteriormente de Dolado y Escrivá (1992), en donde en la relación de cointegración entran los precios y sus primeras diferencias (inflación).

El análisis de la cointegración en el caso $I(2)$ está poco desarrollado, véase Johansen (1991b) y Dolado (1992a,b). No obstante, pueden darse también relaciones de cointegración entre los niveles de las variables solamente —cointegración sin cointegración polinomial—, en cuyo caso los modelos de esta sección serían sin duda adecuados, y siendo las variables fuertemente exógenas no se presentarían problemas especiales de estimación y contrastación. No obstante, las formulaciones (A) y (B) no constituirán auténticas formulaciones en términos de mecanismo de corrección del error, pues si bien el factor $(Y\text{-}gX)_{t-1}$ es el mecanismo de corrección del error, la variable dependiente, ΔY_t, y ΔX_t no son estacionarias.

Cuando esto ocurra no habrá que preocuparse si las variables con crecimiento sostenido son $I(1,1)$ o $I(2,0)$. Los resultados sobre modelos en niveles con variables exógenas se pueden justificar adecuadamente tanto si realmente las variables son $I(1,1)$ como si son $I(2,0)$. Esto es importante porque tal y como se deriva del trabajo de Maravall (1992) sobre extracción de señales, se tiene que «desde el punto de vista de la estimación la tendencia $I(2)$ será indistinguible de un modelo $I(1)$ con deriva».

Siendo cierto lo dicho en los dos párrafos anteriores, hay que advertir que aunque las variables sean exógenas si son $I(2)$ la formulación del modelo debe ser muy cuidadosa, ya que se puede requerir, como en el caso indicado de Dolado y Escrivá (1992), el empleo de niveles y primeras diferencias. Detectar tales relaciones no es fácil, y en tal caso el basar el análisis econométrico en los resultados de la Teoria Económica resulta esencial.

3.12. El planteamiento de información limitada en sistemas cointegrados

Cuando un sistema de ecuaciones incorpora variables endógenas no estacionarias que están cointegradas[49], los problemas que aparecen respecto a sistemas estacionarios son varios. Phillips (1991) realiza un excelente tratamiento de esta cuestión. Los principales resultados de dicho trabajo se pueden resumir así.

1) Es importante que la estimación de sistemas cointegrados se realice incorporando la información sobre las raíces unitarias presentes. Con ello los estimadores tiene propiedades asintóticas óptimas, y los contrastes de hipótesis pueden realizarse mediante los procedimientos habituales.

2) Por el contrario, si las raíces unitarias se estiman como unos parámetros más del sistema, se obtiene que las distribuciones asintóticas necesarias para la inferencia no son distribuciones de tipo estándar, con lo que es necesario tabular tales distribuciones. En dichas tabulaciones hay que tener en cuenta el efecto de los parámetros molestos (nuisance parameters), que además tienen que ser estimados. En tales casos no se dispone de una teoría asintótica óptima aplicable.

3) En la estimación e inferencia de sistemas cointegrados lo importante no es la forma precisa de especificar el modelo, por ejemplo en la forma de mecanismo de corrección del error, sino la incorporación de la información sobre raíces unitarias, que también se puede realizar especificando las variables dependientes de cada ecuación en niveles y las explicativas en niveles y diferencias.

4) La formulación de modelos de información limitada, incluso modelos uniecuacionales, es posible y relativamente fácil en sistemas en los que sólo aparecen las relaciones a largo plazo, es decir, que no incorporan una estructura dinámica transitoria. En dichos casos los modelos de información limitada se formulan incluyendo las variables explicativas en niveles y en diferencias presentes y pasadas. Con tales modelos se pueden obtener estimaciones óptimas de las relaciones a largo plazo o relaciones de cointegración.

5) Si estos sistemas incluyen una formulación dinámica transitoria además de las relaciones de largo plazo, la reformulación necesaria para la construcción de modelos de información limitada óptimos puede ser muy compleja, así como requerir también diferenciaciones futuras de las variables explicativas. En tales casos Phillips recomienda la estimación semiparamétrica.

6) Para la estimación óptima de las relaciones de cointegración sólo se requieren estimadores consistentes de los parámetros que recogen la estructura dinámica transitoria, incluso en el caso en que estos parámetros y los de las relaciones de cointegración no tengan campos de variabilidad independientes entre sí.

7) Con fines de estimación e inferencia el enfoque uniecuacional sólo es válido si las variables explicativas son fuertemente exógenas.

8) Si se aplica un enfoque uniecuacional cuando no se cumple la propiedad de exogeneidad anterior, se tiene que los estimadores resultantes incorporan los inconvenientes señalados en el punto segundo. Por ello en sistemas cointegrados el tratamiento simultáneo es más apremiante que en el contexto clásico de variables estacionarias.

Respecto a los modelos VAR, la presencia de variables no estacionarias y cointegradas tiene las siguientes implicaciones:

a) El problema de no estacionariedad no se puede solucionar formulando un modelo VAR sobre variables diferenciadas. Esa formulación es incorrecta ya que omite los términos —mecanismos de corrección del error— que recogen las relaciones de largo plazo entre los niveles de las variables.

b) Si se formula el modelo VAR sobre las variables en niveles, el modelo no incorpora la información sobre las raíces unitarias presentes entre las variables, sino que dichas raíces se estiman. En tal caso los estimadores obtenidos tienen los problemas señalados más arriba en el punto segundo.

c) La solución consiste en formular el modelo VAR sobre series diferenciadas incluyendo los términos correspondientes a los mecanismos de corrección del error. Al modelo resultante se le denomina en ocasiones VECM (Vector Error Correction Mechanism).

Como comentario final se puede señalar que la teoría de la cointegración ha proporcionado una terminologia que se ha aceptado y

[49] Si no hay ninguna relación de cointegración entre ellas, diferenciando las variables se obtiene un sistema estacionario.

sobre todo asumido tanto en econometría como en macroeconomía, lo que ha servido para establecer una comunicación mucho más fluida entre ambas disciplinas. Además, la cointegración está sirviendo para unificar distintas metodologías econométricas. En Granger (ed., 1990) se recogen una serie de trabajos conducentes a ilustrar el estado actual de la econometría de series temporales, y en ellos se presentan las principales metodologías existentes en la actualidad. Pues bien, todas estas metodologías: (a) de modelos uniecuacionales dinámicos con mecanismo de corrección del error (LSE), (b) de modelos VAR, o (c) de ecuaciones simultáneas, están mostrando una fuerte convergencia con el desarrollo de la teoría de la cointegración.

En efecto, la teoría de la cointegración ha puesto de manifiesto que la formulación de modelos simultáneos puede exigir la estimación de raíces unitarias, lo que conduce a problemas en la estimación y sobre todo en la contrastación a partir de tales estimaciones. Por eso, la formulación en términos de mecanismos de corrección del error es mucho más interesante, ya que con ellos los problemas anteriores desaparecen.

Sin embargo, la formulación que se emplea en la metodología de la LSE de modelos de información limitada puede ser, en presencia de variables cointegradas, extraordinariamente más compleja, con lo que la formulación del sistema multiecuacional aparece más importante. Este sistema puede ser del tipo VECM mencionado anteriormente, que supone una ampliación de los VAR tradicionales mediante la incorporación de mecanismos de corrección del error.

Todos estos puntos aparecen tratados con profundidad en el excelente trabajo de Phillips (1991).

3.13. Uso de modelos econométricos en el análisis aplicado

En este epígrafe final se abordarán una serie de cuestiones relacionadas con la explotación de la información contenida en un modelo econométrico, y su incidencia en la evaluación de la situación que atraviesa el fenómeno de interés. Se ha reservado el estudio de estos temas para el final del capítulo debido a que el potencial de un modelo econométrico sólo se revela plenamente una vez que se conocen sus fundamentos.

Las distintas utilizaciones de un modelo econométrico, sea multiecuacional o uniecuacional, en el análisis aplicado se pueden agrupar en tres grandes apartados: *análisis estructural*, *predicción* y *evaluación de trayectorias alternativas*.

3.13.1. Análisis Estructural

Uno de los fines que muchas veces se persigue en la construcción de un modelo econométrico es caracterizar la relación teórica que liga a las distintas variables que lo componen. De ahí que, incluso en aquellos enfoques más reluctantes al uso de información impuesta a priori, se hayan desarrollado procedimientos para aproximar o proponer una o varias estructuras causales posibles en los datos[50].

De manera más formal, el análisis estructural se puede definir como «el uso de un modelo econométrico estimado para medir las interrelaciones subyacentes del sistema que se considera»[51]. Esta es una definición amplia, ya que basa el análisis estructural en:

1) La estimación de los parámetros de la forma estructural, que en definitiva no es más que buscar la contrapartida empírica de aquellos parámetros que caracterizan las relaciones teóricas en que se ha fundado el modelo. El objetivo en este caso es contrastar una determinada propuesta o principio teórico mediante la contrastación estadística sobre un parámetro o conjunto de parámetros del modelo econométrico estructural, ya que éste ha sido especificado de forma que si sobre los parámetros mencionados no se rechaza una determinada hipótesis, eso implica que la supuesta teoría es compatible con los datos, al tiempo que de dicho contraste estadístico puede derivarse también que otras teorías alternativas quedan rechazadas.

2) La estimación de los parámetros de la forma reducida restringida correspondientes a las variables exógenas, que muestran el efecto del cambio en una de estas variables sobre las variables endógenas. De entre estos parámetros destacan los multiplicadores impacto, definidos en la sección nueve para el modelo uniecuacional.

3) La estimación de los parámetros de la forma final, para recoger los multiplicadores acumulados y totales de las variables endógenas frente a cambios en las exógenas.

En resumen, el objetivo del análisis estructural es claro: poner de manifiesto la interdependencia de los fenómenos económicos reales, tal y como se aproxima por un modelo econométrico que combina la información a priori con la información muestral.

[50] Véase en la sección cuarta la discusión sobre la caracterización de la estructura en modelos VAR.

[51] Intriligator (1983), pag. 208.

3.13.2. Predicción

Una vez se dispone de una caracterización del sistema económico, tal y como la resume el modelo econométrico, resulta posible emplearla para predecir la evolución futura de las variables endógenas.

La principal característica de la predicción econométrica es que no sólo tiene en cuenta la información sobre el pasado de la variable a predecir, sino también los valores que toman cualesquiera variables que puedan estar relacionadas con ella. Por lo tanto, y al menos a priori, cabe esperar que la calidad de la predicción econométrica sea superior a la que se deriva de los modelos univariantes, ya que se basa en una caracterización del sistema económico que genera la variable a predecir y no en una simple extrapolación del pasado.

Como conclusión, teóricamente la predicción hecha con un modelo econométrico es preferible a la qué se hace con un modelo univariante, ya que, como contiene más información, es más eficiente, o lo que es lo mismo, la incertidumbre asociada es menor.

No obstante, y ya situados ante un problema práctico de predicción, no se debe olvidar que la construcción de modelos econométricos puede ser, y en general lo es, compleja. Con frecuencia las variables teóricas que el modelo requiere no se observan —recuérdese el ejemplo del efectivo en la sección novena—, y en esas situaciones no se tienen garantías de que el modelo especificado sea la mejor aproximación posible de la realidad.

Esto evidentemente afectará a cualquier uso que se haga del modelo, y por lo tanto a la predicción. En tales casos si las características de las variables omitidas cambian en el período postmuestral respecto al muestral, las predicciones con dicho modelo estarán sesgadas.

Existe además un problema de índole práctico: necesariamente un modelo econométrico ha de ser más eficiente que un modelo univariante, ya que incorpora más información. Ahora bien, esto realmente sólo se traduce en una ganancia para la predicción cuando se conoce el valor exacto de las variables explicativas. Pero en muchas aplicaciones no sucederá así, y habrá que predecir previamente los valores de las variables explicativas para llegar a la predicción de la variable de interés.

En tal caso la predicción econométrica pasa a depender de la calidad de las predicciones de los inputs, y entonces ya no tiene porqué ser preferible a la predicción univariante, pues grandes errores de predicción de las variables explicativas pueden acumularse e inducir grandes errores de predicción en la variable dependiente.

Por ejemplo, supóngase que se trata de explicar la evolución de índice de precios al consumo en España, y para ello se construye un modelo econométrico uniecuacional donde se proponen como variables explicativas una medida de la liquidez, un conjunto de variables fiscales y un índice de precios de importaciones; no entramos en la definición detallada de estas últimas por ser accesorio para la discusión que sigue.

Está claro que un modelo con tal riqueza de información va a predecir, en principio, mejor que un modelo que se limita a incorporar la historia pasada del IPC. Ahora bien, puede suceder que las variables explicativas mencionadas no se puedan predecir con márgenes de error reducidos; y sus errores en la predicción pueden llevar, si no se compensan entre sí, a errores importantes en la predicción de los precios al consumo, incluso mayores que los que cometería un modelo univariante.

La necesidad de hacer predicciones ha llevado a que surgieran una categoría especial de modelos econométricos, los *Modelos con Indicador*.

Los modelos con indicador están basados, fundamentalmente, en la idea de correlación, y en la precedencia temporal del indicador sobre la variable endógena.

Un ejemplo claro se tiene con el IPC y la relación que muestra su evolución con la de los precios industriales (IPRI)[52]: se comprueba que la trayectoria del IPC es parecida a la de los precios industriales con unos pocos meses de retraso.

Esta es una regularidad basada en el hecho de que los efectos en los precios al consumo de las variables determinantes de los mismos tenderán a reflejarse previamente en los precios industriales. Esta regularidad resulta útil para predecir el IPC, ya que el IPRI actúa como un indicador adelantado: ciertas variables que sí tienen un efecto causal sobre el IPC también lo tienen sobre el IPRI, y necesariamente ejercen su influencia sobre éste antes que sobre el IPC.

Por esa razón a los modelos con indicador a veces se les llama *Modelos en Forma Semirreducida*: en vez de plantear un modelo econométrico sobre las variables causales, se plantea sobre una variable intermedia que actúa de puente, y que reacciona a los cambios en las variables causales antes que la variable de interés.

Son modelos más eficientes que los univariantes, ya que incluyen información sobre variables «explicativas». Al mismo tiempo son más sencillos que los econométricos, ya que la complejidad

[52] Esto no es del todo cierto, ya que habría que desagregar el IPC por categorías del consumo y determinar exactamente cuáles son los subíndices que muestran esa relación: el trabajo de Matea que figura en el capítulo nueve desarrolla extensamente este punto. Sin embargo la simplificación que se hace aquí facilita la exposición de la idea principal.

implícita en la formulación de un modelo general con varias variables explicativas se reduce al trabajar con un número pequeño de indicadores. Como contrapartida son menos eficientes, pues los mecanismos de transmisión de los factores explicativos no se explicitan, lo que sí ocurre en un modelo con verdaderas variables explicativas.

A la hora de predecir un modelo con indicador tiene una gran ventaja: por regla general un indicador sólo se considera tal si es un indicador adelantado, es decir, si presenta la misma evolución que la variable a predecir antes de que lo haga ésta. Un modelo típico es:

$$Y_t = \beta X_{t-4} + N_t$$

Si se conoce la historia de ambas variables, Y y X, hasta el momento t inclusive, con este modelo se pueden predecir los valores de Y correspondientes a $t+1$, $t+2$, $t+3$ *y* $t+4$ sin necesidad de usar predicciones del indicador, gracias al desfase de cuatro períodos entre variable a explicar e indicador.

En cambio con modelos econométricos algunas de las variables explicativas tendrán una relación contemporánea con la variable dependiente, y será preciso predecir los correspondientes valores de tales variables. Esto supone, como hemos señalado anteriormente, que las predicciones econométricas pueden perder eficiencia, por lo que en una aplicación concreta no necesariamente se han de obtener predicciones mejores con un modelo econométrico que con un modelo con indicador.

3.13.3. Evaluación de Trayectorias Alternativas: Simulación y Control

Por último, el modelo también se puede emplear para aproximar la evolución previsible de la variable dependiente bajo distintos supuestos sobre la evolución de la(s) variable(s) que incide(n) en ella. A esto se le conoce como *evaluación de trayectorias alternativas*, y es una generalización del uso de los modelos macroeconométricos para evaluar distintas medidas de política económica.

Un primer tipo de evaluación lo forman los ejercicios de simulación, donde el analista fija, de acuerdo con criterios ajenos al modelo en sí, la trayectoria de una o varias variables exógenas. Nótese que no necesariamente el usuario ha de poder fijar los valores de las variables exógenas en las cifras que él desee, sino que en la práctica se tendrá una combinación de variables que el analista sí puede controlar con variables que son totalmente inaccesibles a cualquier intento de influir en ellas.

En el ejemplo del modelo econométrico que liga el IPC con variables de liquidez, fiscales y un índice de precios de importaciones, supongamos que el usuario es la autoridad económica; en ese caso, puede determinar —dentro de ciertos límites— los valores de las variables monetarias y fiscales, y en consecuencia dicho modelo se puede emplear para simular la evolución futura de los precios bajo distintas combinaciones de políticas monetaria y fiscal. En cuanto a la otra variable, el índice de precios de importaciones, aunque no se pueda actuar directamente sobre ella se puede usar el modelo para simular escenarios, que se basan en distintos supuestos acerca de cómo se va a comportar en el futuro.

Para la economía española existe un modelo reciente especialmente diseñado para simular distintos tipos de políticas, el modelo MOISEES (Molinas et al., 1990).

Dentro de este apartado de evaluación, el modelo econométrico también se puede utilizar para determinar los valores óptimos de las variables explicativas que el analista controla con vistas a alcanzar unos valores deseados de la variable dependiente. Esta es la aplicación del modelo con fines de *control*.

En ese caso el usuario tiene una función objetivo donde se recoge una trayectoria de valores deseados para un conjunto de variables, así como los costes asociados a las desviaciones de esos valores deseados. El modelo econométrico representa la estructura del sistema, es decir, las relaciones existentes entre la(s) variable(s) dependiente(s) y la(s) variable(s) que el usuario maneja y con la(s) que trata de alcanzar los valores deseados. Se plantea un problema de optimización dinámica, que se resuelve aplicando la teoría del control óptimo.

La relación con la simulación es evidente: en la simulación se parte de cualquier combinación de las variables explicativas que al usuario le interese considerar, y se trata de determinar su efecto sobre la variable de interés. Por el contrario, en el control óptimo el punto de partida es una evolución deseada de la variable de interés, y lo que se calcula son los valores que han de tomar las variables controladas para que la evolución real de la variable modelizada esté lo más cerca posible de esa serie de valores deseados.

Una exposición teórica del control de sistemas económicos dinámicos se encuentra en Chow (1975); una aplicación relativa a la política monetaria y fiscal en el caso español se encuentra en Cancelo (1988a).

3.13.4. La Complementareidad de los Distintos Tipos de Modelos en el Análisis de un Fenómeno Económico

Para terminar este repaso general a la modelización de variables económicas resta contestar a una pregunta; en este capítulo y en el anterior se han estudiado modelos univariantes, VARMA, estructurales, uniecuacionales, con indicador, etc. La pregunta que surge después de todo ello es casi obligada: ¿para qué tantos tipos de modelos? ¿Hasta qué punto estos modelos resultan redundantes en el análisis aplicado? ¿Existe algún modelo que incluya a los demás, o por el contrario todos contienen una información diferencial y propia que los hace imprescindibles?

Considérese primero el caso en que se dispone de distintos modelos econométricos. Este problema ha sido muy tratado en la literatura y la propuesta consiste en analizar si es posible escoger un modelo que comprenda (encompass) a los demás, en el sentido de que sea capaz de explicar las características relevantes de los otros modelos y de apuntar los motivos por los que éstos no explican características de los datos que con el modelo que se pretende escoger quedan aclaradas. Este principio de comprensión (encompassing principle) de un modelo sobre los demás constituye una fase muy importante en el progreso de la econometría aplicada. Sobre este tema el lector interesado puede consultar Mizon y Richard (1986), Hendry y Richard (1989), Mizon (1989), etc.

Otro aspecto distinto de multiplicidad de modelos pero quizás muy patente en problemas de predicción, es el que se refiere a situaciones en que se dispone de modelos univariantes, econométricos (uniecuacionales) y modelos con indicador sobre una determinada variable. Si se utilizan dichos modelos para predecir se encontrará que las predicciones diferirán entre sí, y de esas diferencias se puede extraer una información útil.

Las predicciones univariantes no incorporan los últimos acontecimientos del sistema hasta que éstos empiezan a afectar al pasado de la serie. En la medida en que estos acontecimientos ya hayan incidido en los indicadores, su efecto sí se reflejará en las predicciones con indicador. Por esa razón la comparación de esas predicciones con las univariantes sirve para evaluar el efecto de esos últimos acontecimientos.

A su vez las predicciones econométricas incorporan la información sobre la evolución de lo que supuestamente son las verdaderas causas que afectan a la serie. Estas predicciones econométricas se explotan en mayor medida al compararlas con las otras dos, pues de ese modo se puede determinar con precisión el efecto de la evolución más reciente de las variables causales.

En definitiva las diferencias entre las predicciones de los distintos modelos, lejos de ser algo negativo, pueden servir para comprender mejor qué está pasando y cuáles son los determinantes últimos en la evolución de la variable de interés.

Por otra parte, con información muy desagregada en el tiempo, por ejemplo, mensual o semanal, la construcción de modelos econométricos puede no ser posible porque algunas variables determinantes no se observen al nivel de desagregación temporal de la variable endógena. Esto obligará a agregar en el tiempo la información y construir modelos econométricos, por ejemplo, trimestrales o anuales. Con tales modelos no se puede explotar la información desagregada que ha llegado de la variable endógena. Sin embargo, si en ese caso además de un modelo econométrico se dispone de un modelo univariante, mensual o semanal, sobre la variable endógena, a medida que llegan datos mensuales o semanales sobre ella se podrá realizar una predicción univariante, con el fin de evaluar si las innovaciones mensuales o semanales incorporadas implican cambios importantes respecto la predicción trimestral o anual del modelo econométrico.

Los trabajos de Delrieu (capítulo 8) y Matea (capítulo 9) incluidos en este mismo libro ilustran claramente la importancia de disponer de distintos modelos sobre un mismo fenómeno económico, así como la información diferencial que se extrae de ellos en las aplicaciones concretas (comercio exterior e inflación, respectivamente) que realizan.