Capítulo 4
CARACTERIZACION DE LOS ASPECTOS ESENCIALES DE UN FENOMENO ECONOMICO MEDIANTE TECNICAS ESTADISTICAS DE EXTRACCION DE SEÑALES

Antoni Espasa y José Ramón Cancelo

4.0. Introducción

El análisis de la coyuntura de un fenómeno económico normalmente no se realiza directamente a través de la evolución de sus datos originales, ya que éstos contienen muchas oscilaciones de escaso interés económico que pueden llevar a conclusiones equivocadas. Por contra, dicho análisis se efectúa a través de señales extraídas de los datos, que sirven de estimación de lo que un analista puede considerar como un aspecto esencial del fenómeno en cuestión.

De entre estos aspectos esenciales, quizá el más relevante sea el concepto de evolución subyacente, que se puede estimar identificándolo con los conceptos estadísticos de tendencia o serie ajustada de estacionalidad; esto hace que la forma en que éstas —y otras— señales se extraen sea una cuestión de gran importancia dentro del análisis de la coyuntura económica.

En este capítulo se estudian el tipo de señales que es conveniente utilizar en un análisis de la coyuntura y los principales métodos para estimarlas a partir de la serie temporal observada del agregado.

La exposición de los procedimientos de extracción de señales requiere, forzosamente, el empleo de un instrumental matemático con un nivel de dificultad bastante elevado. En la presentación que se hace en el capítulo se intenta simplificar al máximo la complejidad matemática, rehuyendo en todo momento demostrar los resultados utilizados; pero ello no evita usar una notación aparentemente compleja para poder dar una explicación de lo que cada procedimiento hace, y de cómo se relacionan los resultados de los diferentes procedimientos.

Por ello, y con el fin de facilitar la lectura a aquellos lectores que estén más interesados en las consecuencias de los resultados estadísticos sobre la forma de plantear e interpretar el análisis de la coyuntura, que en el proceso formal de su obtención, el capítulo está adecuadamente dividido en secciones y subsecciones diseñadas para facilitar una lectura resumida.

Las subsecciones con un nivel matemático más elevado están marcadas con dos asteriscos, y se pueden omitir sin pérdida de continuidad. Si se pretende una lectura más superficial, centrada en el planteamiento del problema, una visión general de la forma de proceder para resolverlo, un resumen de los principales resultados, y sus consecuencias desde el punto de vista de la realización del análisis, entonces se pueden omitir también las subsecciones marcadas con un asterisco.

En la sección primera se describen los componentes de una serie temporal y las características de las mismas cuando la serie temporal viene generada por el tipo de modelos ARIMA con Análisis de Intervención considerados en el capítulo 2. En este caso, todos los procedimientos de extracción de señales estiman los componentes de una serie temporal mediante la aplicación de medias móviles simétricas sobre dicha serie. Por ello, en la sección segunda se ofrece una explicación intuitiva del porqué es posible realizar la estimación de los componentes buscados usando este tipo de transformaciones de los datos observados.

De todos los procedimientos posibles que se describen en este libro, los empiricistas son los más sencillos, y a ellos se dedica la sección tercera; si bien la lectura de esta sección se recomienda con carácter general a todos los lectores, el nivel de detalle que se quiera obtener de dicha sección puede ser distinto para cada persona.

Los procedimientos empiricistas tienen el grave inconveniente de que apenas tienen en cuenta las características específicas de la serie a la que se aplican. Por esa razón, en los últimos años ha surgido una línea de investigación basada en relacionar la descomposición en componentes no observables con una representación del modelo teórico de generación de los datos observados: este enfoque ha dado lugar a lo que se conoce en la literatura como métodos de descomposición basados en modelos.

En el epígrafe cuarto se discute el planteamiento general de los métodos basados en modelos de forma reducida: en ellos el modelo ARIMA de la serie agregada juega un papel determinante a la hora de caracterizar las propiedades teóricas de las señales a extraer, así como en el diseño de los filtros que se emplearán. El nivel matemático que se ha de emplear en la discusión de estos temas es necesariamente elevado, y de ahí que esta sección se subdivida en varias subsecciones, algunas de las cuales se pueden omitir en una primera lectura.

Un planteamiento alternativo de la idea básica que subyace en los métodos basados en modelos lleva a formular a priori los modelos teóricos de los componentes, en función de las características deseables que éstos deben mostrar: se tienen así los métodos basados en modelos estructurales. De este tipo de métodos trata el epígrafe quinto, que también se subdivide con el fin de que el lector que sólo esté interesado en una idea general no se vea obligado a entrar en un detalle que sí puede atraer a otro tipo de lectores.

Una de las ventajas de los métodos basados en modelos es que todos los resultados derivados del proceso de descomposición tienen propiedades estocásticas conocidas. Desde el punto de vista de la persona interesada exclusivamente en la interpretación económica de los resultados finales, la principal consecuencia de esto es que se pueden derivar distribuciones de probabilidad para los distintos tipos de errores que surgen en todo proceso de descomposición: el error total, el error en el estimador final y el error de revisión. La sección seis desarrolla estas propiedades, suponiendo que las señales se han estimado usando métodos basados en modelos de forma reducida. Como en las secciones anteriores, lo complejo del tema obliga a proporcionar una orientación que facilite la lectura.

En el epígrafe siete se argumenta la propuesta de utilizar la tendencia como la señal en la que basar el análisis de la coyuntura, en lugar de la serie desestacionalizada como se hace muchas veces en la actualidad: para ello se presentan razones tanto desde el punto de vista económico como estadístico, éstas últimas relacionadas con los desarrollos teóricos de epígrafes precedentes.Las dos últimas secciones tratan de un tipo especial de factores que ocurren de forma más o menos periódica que suelen afectar a las series económicas, y que se modelizan mediante variables artificiales. En el epígrafe octavo se presentan estos factores y se estudia la forma de extender el modelo ARIMA con Análisis de Intervención con el fin de captar estos efectos.

Los procedimientos de descomposición de las secciones tres a seis, e incluso la explicación intuitiva de la sección dos, están basadas en la hipótesis de que la serie a descomponer es estrictamente estocástica, es decir, que su proceso generador de datos se pueda aproximar por un modelo ARIMA. Sin embargo, en la práctica ocurre que:

a) los modelos ARIMA puros no son adecuados en la mayor parte de los casos, ya que la ocurrencia de acontecimientos anómalos que afectan a la serie analizada obliga a ampliarlos con el análisis de intervención;

b) se acaba de decir que existen ciertos factores, los estudiados en el epígrafe ocho, que han de ser modelizados de forma determinista.

En consecuencia, es necesario extender los resultados anteriores al caso en que el proceso generador de la serie observada combine un mecanismo de generación estrictamente estocástico con otro determinista. Esta extensión es la que se realiza en el epígrafe nueve, dedicado a la descomposición de la contribución de los distintos tipos de factores que se modelizan de forma determinista en tendencia, componente estacional y componente irregular.

4.1. La descomposición de una serie temporal económica

4.1.1. Componentes de una serie temporal económica

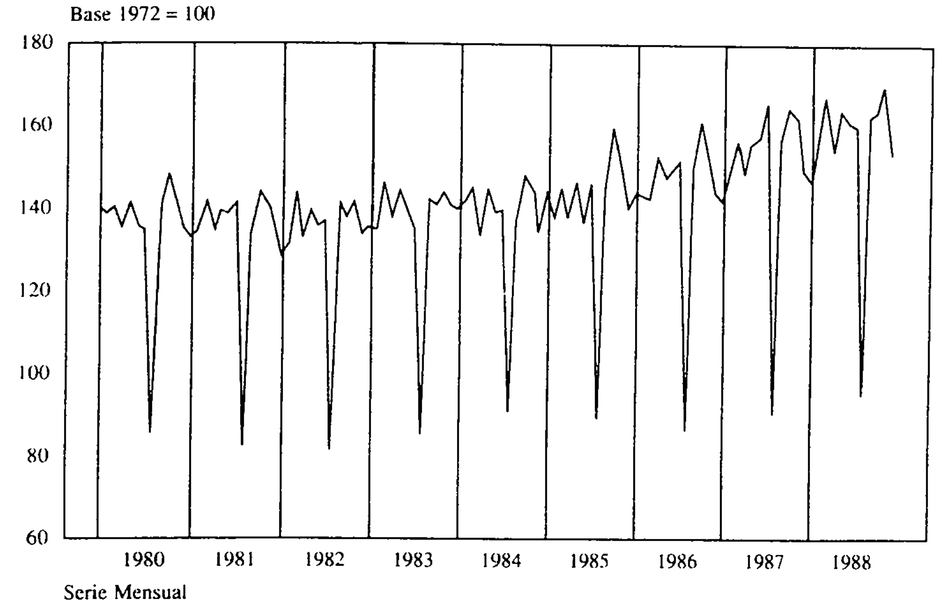
En el primer capítulo se señaló que el análisis de coyuntura tiene que estar basado en los aspectos esenciales del fenómeno económico estudiado. Estos aspectos esenciales no son directamente observables, sino que aparecen combinados con otros aspectos menos importantes para el analista de la coyuntura en los valores observados; por esta razón, es preciso disponer de algún tipo de procedimiento que permita extraerlos de las observaciones. No obstante, con anterioridad a la determinación del procedimiento se ha de definir formalmente lo que se busca, es decir, cuáles son esos aspectos esenciales en los que se centra el interés del analista: de esto trata este primer epígrafe.

La necesidad de basar el análisis de la coyuntura en una línea de evolución firme, con un mínimo de oscilaciones, es clara. Así, por ejemplo, el gráfico 4.1 presenta la evolución del índice de producción industrial (IPI), que muestra diversas oscilaciones: esto hace imprescindible la extracción de una línea firme de evolución del IPI, como paso previo para un análisis de la marcha de la actividad industrial.

El seguimiento de un determinado fenómeno a través de la evolución observada puede generar grandes incertidumbres, y parece aconsejable realizar dicho seguimiento a partir de lo que se conoce como *evolución subyacente* o *nivel subyacente*, y que se puede definir como la evolución firme que hay detrás de la trayectoria observada, una vez que de esta última se eliminan las oscilaciones estacionales y las perturbaciones irregulares o de corto plazo.

Tal y como se ha definido, es evidente que esta evolución subyacente no puede contener más información que la serie observada; al contrario, contiene necesariamente menos, ya que se extrae de ella. Su ventaja respecto a la serie original es haber eliminado de ésta un conjunto de componentes que, muchas veces, no son relevantes desde el punto de vista de las necesidades del analista de la coyuntura. En cualquier caso, incluso cuando todos los componentes tienen un interés económico propio es por motivos diferentes al interés que tiene

GRÁFICO 4.1. *Indice de producción industrial.*



la evolución subyacente como tal, por lo que han de ser considerados por separado.

¿Qué clase de oscilaciones es conveniente eliminar? Pueden ser de muy distinto tipo: por un lado están las oscilaciones irregulares o no sistemáticas, que en general sólo afectan a la serie en el momento en que ocurren y normalmente tienen estructura puramente aleatoria (ruido blanco). Constituyen el llamado *componente irregular*.

La serie puede contener también oscilaciones cuasicíclicas de media cero; de éstas, las más importantes son las que tienen periodicidad anual o submúltiplo del año (trimestrales, mensuales, etc), en cuyo caso se conocen como oscilaciones estacionales o *componente estacional*.

Hay ciertas series que contienen oscilaciones cíclicas debidas a la actividad económica, y cuya periodicidad oscila entre los dos y cinco años. Suelen ser menos frecuentes y menos sistemáticas que las estacionales, pero cuando existen reciben el nombre de *componente cíclico*.

El cuarto componente de una serie temporal es lo que se llama *tendencia* o *componente tendencial*. La tendencia recoge aquella parte de la variable económica que está relacionada, principalmente, con factores de largo plazo. De acuerdo con la exposición anterior, la tendencia es lo que queda de la serie original al eliminar de la misma los componentes cíclicos e irregular.

La agregación de estos cuatro componentes no observables forma la serie observada, de acuerdo con la relación

$$Y_t = T_t + C_t + S_t + I_t, \tag{4.1.1}$$

donde Y_t denota la serie observada, T_t su tendencia, C_t el componente cíclico no estacional, S_t el componente estacional e I_t el componente irregular.

En este capítulo se discutirán una serie de procedimientos para llegar a determinar los valores de los componentes no observables que figuran a la derecha de la igualdad (4.1.1), correspondientes a una serie temporal dada de la variable Y_t.

Cabe destacar que cualquiera de estos componentes o la agregación de algunos de ellos puede ser el objetivo del analista; a este objetivo se le llamará de ahora en adelante la *señal de interés* en un análisis concreto.

En la práctica resulta muy difícil distinguir la tendencia del componente cíclico, sobre todo cuando, por razones que se verán a continuación, ambas se suponen estocásticas[1]. De ahí que, en vez de procurar separarlas, ambas se combinen en un componente que se denomina *tendencia-ciclo*. Para simplificar la terminología y la notación, a esa tendencia-ciclo se le llama simplemente tendencia y se denota por T_t; pero se sobreentiende que, en aquellos casos en que la serie presenta ciclos de actividad, en la tendencia se combinan dichos ciclos y una tendencia en sentido estricto.

Con ello (4.1.1) se simplifica de la siguiente forma

$$Y_t = T_t + S_t + I_t. \tag{4.1.2}$$

En las expresiones (4.1.1) y (4.1.2) se supone que los componentes se combinan de forma aditiva para formar la variable observada; una alternativa es el llamado esquema multiplicativo, dado por

$$Y_t = T_t \cdot S_t \cdot I_t \tag{4.1.3}$$

donde los componentes estacional e irregular son una determinada proporción de la tendencia, y no una cantidad absoluta que se añade a ésta independientemente del valor que dicha tendencia tenga.

En el análisis económico aplicado el esquema multiplicativo es más frecuente que el aditivo. En efecto, la relación (4.1.3) establece

[1] Hay que tener en cuenta, además, que el tamaño de la muestra influye considerablemente en esta separación: lo que parece una evolución tendencial en una muestra de siete años puede ser un efecto cíclico en una muestra de treinta años.

que, por ejemplo, la oscilación estacional de los activos líquidos en manos del público asociada a diciembre es del dos por ciento por encima de la tendencia correspondiente a ese mes, mientras que (4.1.2) implica que diciembre tiene asociada una subida de 100.000 millones por encima del valor de la tendencia. La diferencia consiste en que la oscilación sea proporcional al nivel o una cantidad independiente de éste.

En los casos en que una serie económica sigue el esquema multiplicativo se tiene que su transformación logarítmica sigue un esquema aditivo. Por ello, y sin pérdida de generalidad, a lo largo de este capítulo se centrará la atención en la formulación (4.1.2), entendiendo que está referida a las transformaciones logarítmicas cuando la descomposición es multiplicativa.

En el análisis aplicado es todavía corriente utilizar la descomposición

$$Y_t = Y_t^a + S_t,$$

donde se considera que la variable observada es el resultado de agregar el componente estacional y la llamada *serie ajustada de estacionalidad* o *serie desestacionalizada*, a la que se denotará añadiendo el superíndice «*a*» al símbolo usado para la serie original. Es inmediato que

$$Y_t^a = T_t + I_t,$$

es decir, la serie ajustada es el resultado de combinar en un solo componente la tendencia y el elemento irregular. Sobre el posible interés en el análisis de coyuntura de esta señal, Y^a, se volverá más adelante.

Los componentes son desconocidos y se han de aproximar a partir del dato agregado, única serie que realmente se observa. El problema planteado es un caso particular de lo que se conoce en la literatura estadística como un problema de *extracción de señales*, que pasamos a presentar a continuación.

4.1.2. Extracción de señales y análisis univariante

Es bastante unánime la idea de que el mundo económico es multivariante, por lo que el problema de extracción de señales se debería plantear en ese ámbito. Sin embargo, en la actualidad todavía no se dispone de una metodología suficientemente desarrollada para tratar la cuestión en ese contexto. Ciertos desarrollos teóricos aparecidos a lo largo de la década de los ochenta y vinculados a la teoría de la cointegración constituyen, presumiblemente, la base del tratamiento multivariante, pero hoy por hoy dicho enfoque dista mucho de ser operativo.

Por lo tanto, el problema de extracción de señales se ha de plantear en un contexto univariante. En el tema 2 se vio que, con carácter bastante general, los fenómenos económicos pueden concebirse como generados a partir de un modelo ARIMA con Análisis de Intervención. En consecuencia, parece conveniente que la extracción de señales se realice sobre la base de ese esquema de modelización.

Concretamente, la estimación eficiente de los mencionados aspectos esenciales de un fenómeno económico se ha de basar en el modelo ARIMA que sigue dicho fenómeno. Tal modelo recoge las principales características del fenómeno económico considerado, y no ha de existir contradicción entre el modelo univariante que representa el proceso generador de datos, y la forma en que se extraen de estos datos los valores de las señales no observables.

Una primera implicación, absolutamente determinante, del hecho de que la variable de interés siga un modelo ARIMA —con o sin análisis de intervención— es que los componentes en general serán estocásticos; y, en consecuencia, cualquier forma de aproximarlos de manera determinista —como regresiones sobre la variable tiempo y/o variables artificiales estacionales— será incorrecta.

La extracción de componentes no observables de una serie temporal es una idea muy antigua, que se puede remontar al menos hasta el siglo XVII[2]. Sin embargo hasta la mitad del siglo XX no se dispuso de instrumentos de cálculo potentes y de esquemas teóricos estocásticos generales, lo que limitaba considerablemente el abanico de soluciones disponibles al problema de extracción de señales.

Por estas razones en un primer momento se plantearon esquemas deterministas para aproximar estos componentes: funciones polinomiales del tiempo para la tendencia, sinusoidales para los ciclos, y sinusoidales de período anual o variables artificiales para la estacionalidad. Obviamente pronto se detectó que este tipo de modelización no era satisfactoria, ya que con ella los residuos mostraban una evolución que no coincidía con la que cabía esperar de un componente estrictamente irregular.

De ahí que el siguiente avance consistiera en aproximar las señales mediante medias móviles sobre la serie original[3]. En el contexto de

[2] Véase el capítulo uno de Nerlove et al (1979) para una exposición histórica del análisis de componentes no observables en una serie temporal.

[3] No se debe confundir la expresión medias móviles con el significado que se le va a dar a continuación, con el polinomio sobre el operador de retardos aplicado a las innovaciones de un modelo ARIMA. En el siguiente epígrafe se desarrolla con detalle las similitudes y diferencias entre ambas interpretaciones.

la extracción de señales, se llama *media móvil de tamaño* $2m+1$ a la serie definida como

$$MM(2m+1)_t = \sum_{j=-m}^{m} a_j Y_{t-j} \tag{4.1.4}$$

donde $a_j \neq 0$ para $j=-m, m$, $\sum_{j=-m}^{m} a_j = 1$ y normalmente $a_j = a_{-j}$.

Aproximar un componente mediante una media móvil supone pasar de considerarlo puramente determinista a interpretarlo como estocástico. Si la tendencia se define como $\beta_0 + \beta_1 t$, siendo β_0 y β_1 coeficientes fijos, el valor de tal tendencia en el momento t^* vendrá dado por $\beta_0 + \beta_1 t^*$, con independencia de las condiciones por las que esté atravesando el fenómeno: dicho valor tendencial es predecible sin error desde cualquier momento anterior a t^*. En cambio, en (4.1.4) la serie $MM(2m+1)_t$ es una transformación lineal de las variables aleatorias Y_{t-j}, $j=-m, \ldots, m$, y el carácter aleatorio de estas últimas se traslada a la señal resultante[4].

Actualmente se ve la aplicación de medias móviles del tipo expuesto en (4.1.1), como el procedimiento de estimación óptima de señales en series generadas por un modelo ARIMA. Así, siempre que la serie observada siga un modelo ARIMA, cualquier señal no degenerada de la misma ha de obtenerse empleando una media móvil adecuadamente definida, en función de las características de la señal y de la variable.

Si la serie viene generada por un modelo ARIMA ampliado con un análisis de intervención, se puede considerar dicha serie como formada por dos partes: la puramente estocástica generada en la estructura ARIMA, y la puramente determinista motivada por la(s) anomalía(s) de carácter exógeno y recogida(s) en el análisis de intervención.

En ocasiones, la parte puramente determinista se puede descomponer en distintas señales no observables, y en particular en tendencia y estacionalidad.

Por lo tanto cuando el fenómeno de interés tenga parte determinista y parte estocástica, en el sentido en que se acaban de definir, es preciso proceder a su descomposición por separado, para combinar los resultados al final.

Se hablará así de una tendencia total, que será la suma de la tendencia correspondiente a la parte estocástica, o tendencia estocásti-

[4] Las implicaciones de aproximar una tendencia estocástica de forma determinista se analizan en Chan et al (1977) y Nelson y Kang (1981, 1984).

ca, y de la tendencia correspondiente a la parte determinista, o tendencia determinista. Naturalmente lo mismo se puede decir para los otros componentes.

4.2. El uso de medias móviles para la extracción de señales

El objetivo de esta sección es discutir en términos generales cómo las medias móviles actúan sobre una serie de partida (input) para producir una serie final (output)[5]. Esta transformación, basada en promediar sucesivos valores de una serie de origen, genera una nueva serie en la que se han atenuado determinadas características de la serie original y se han reforzado otras.

Es importante recalcar una vez más que en este capítulo la expresión medias móviles se refiere a una operación empleada para transformar series y constituyen, por tanto, un filtro que se aplica a los datos observados, y no un término utilizado para describir el proceso generador de los mismos como era el caso en el capítulo 2.

Obviamente ambos conceptos tienen grandes similitudes. En ambos casos se trata de una transformación lineal aplicada sobre una sucesión temporal de partida. En el caso del modelo ARIMA el término de «medias móviles» es una combinación lineal de innovaciones pasadas que sirve para recoger la dependencia de cada término de una serie temporal Y_t respecto al pasado. Supóngase, para simplificar, que el modelo ARIMA no tiene parte autorregresiva o que está formulado en términos de la descomposición de Wold (véase sección 2.2): entonces se tiene una explicación de Y_t a partir de una transformación lineal de las innovaciones presente y pasadas. Como se trata de una explicación de los datos Y_t en términos de elementos más simples, las innovaciones a_t, éstas constituyen la serie de partida e Y_t la serie que resulta tras la aplicación de la media móvil.

Además, en este caso tenemos una media móvil unidireccional: del presente al pasado exclusivamente, sin incluir el futuro. Tal unidireccionalidad emana de la restricción de recursividad temporal, por la que se supone que el futuro (una innovación que todavía no se ha producido) no puede afectar al presente Y_t.

En este epígrafe, sin embargo, no se busca una explicación sobre la generación de la serie Y_t, sino que se pretende simplificar alguno de sus aspectos: por ello Y_t es ahora la serie de partida. Por otra parte, en cualquier momento de tiempo (t) alejado de los momentos correspondientes a los extremos de la serie se puede aplicar una transformación lineal que involucre a valores de Y anteriores y posteriores a t. Por tanto ahora la media móvil puede ser bidireccional, calculándose en base a valores pasados, presente y futuros.

[5] Maravall (1989) es una buena referencia para el conjunto de problemas que se abordarán a continuación.

Es más, así conviene que sea: conocida toda la serie temporal $Y_1, Y_2, \ldots, Y_{T-1}, Y_T$ se pretende para cada momento t obtener la transformación lineal que mejor recoja el aspecto que se busca. Para ello lo más eficiente es utilizar toda la información disponible —toda la muestra— y en t aplicar una transformación bidireccional, que incluya a todos los valores observados que puedan aportar información sobre el aspecto concreto que se trata de aproximar.

Para simplificar, sea una serie que sólo presenta tendencia, T_t, y componente irregular, tal que

$$Y_t = T_t + I_t \tag{4.2.1}$$

Promediando esta expresión alrededor del momento de tiempo t

$$\frac{1}{2m+1}\sum_{j=-m}^{m} Y_{t-j} = \frac{1}{2m+1}\sum_{j=-m}^{m} T_{t-j} + \frac{1}{2m+1}\sum_{j=-m}^{m} I_{t-j},$$

se obtiene

$$Y_t^* = T_t^* + I_t^*.$$

Si la tendencia T_t evoluciona suavemente, su promedio T_t^* va a diferir poco de T_t; en cambio, al promediar I_t la nueva perturbación presentará muchas menos oscilaciones, ya que habrá una compensación entre valores positivos y negativos[6].

Por lo tanto Y_t^* será muy parecido a T_t^*, que a su vez será muy similar a T_t, es decir a la tendencia original. Esto es una justificación intuitiva de la operación de promediar sobre la serie original, y a continuación aproximar el componente tendencial no observable mediante dicho promedio. En definitiva, el razonamiento que hay detrás de este proceder es que el promedio afecta poco a la tendencia, pero elimina en gran parte el componente irregular, y el resultado se puede asimilar a aquélla.

Supóngase ahora que la serie agregada sigue un esquema de la forma

$$Y_t = S_t + I_t \tag{4.2.2}$$

[6] Si I_t es ruido blanco normal (0, 1), y usando resultados elementales de estadística teórica, la perturbación original tiene una probabilidad de 95% de tomar valores entre $-1{,}96$ y $1{,}96$; en cambio si $2m+1=9$, I_t^* tiene esa misma probabilidad en el intervalo $(-0{,}65,\ 0{,}65)$.

donde S_t es un componente estacional, en principio estocástico, y I_t el elemento irregular.

Para profundizar en este esquema, supóngase sin pérdida de generalidad que Y_t es una serie mensual, con lo que

$$E\left(\sum_{j=0}^{11} S_{t-j}\right)=0.$$

Si en este caso se toma una media móvil similar a la que se consideró en (4.2.1), el resultado carece de interés; esto se ve claramente calculando un promedio de doce meses consecutivos:

$$\frac{1}{12}\sum_{j=0}^{11} Y_{t-j}=\frac{1}{12}\sum_{j=0}^{11} S_{t-j}+\frac{1}{12}\sum_{j=0}^{11} I_{t-j}.$$

De forma abreviada

$$Y_t^*=S_t^*+I_t^*.$$

Obsérvese que por las razones ya comentadas I_t^* será de un orden de magnitud menor al de la perturbación original I_t. Pero ahora también el sumatorio del componente estacional será próximo a cero, ya que el nuevo componente estacional es el resultado de promediar doce valores consecutivos del antiguo componente, y esta operación tiene una esperanza matemática cero.

En consecuencia, la misma media móvil que en (4.2.1) permitía aproximar la señal de interés T_t, en (4.2.2) no tiene utilidad: como la señal en este caso es radicalmente distinta a la anterior, es preciso utilizar una transformación específicamente diseñada para captar esa nueva señal.

La extracción de la señal en (4.2.2) se puede abordar mediante el empleo de medias móviles estacionales, es decir, promediando observaciones correspondientes al mismo mes de distintos años:

$$\frac{1}{2m+1}\sum_{j=-m}^{m} Y_{t-12j}^l=\frac{1}{2m+1}\sum_{j=-m}^{m} S_{t-12j}^l+\frac{1}{2m+1}\sum_{j=-m}^{m} I_{t-12j}^l,$$

donde el momento t corresponde al mes l del año. A las variables se les añade el superíndice l para recalcar que los promedios se refieren siempre a un mismo tipo de mes.

Por lo que respecta al segundo sumando de la derecha de la igualdad

$$I_t^*=\frac{1}{2m+1}\sum_{j=-m}^{m} I_{t-12j}^l$$

es de un orden de magnitud menor que el de I_t; en el caso en que I es ruido blanco, es indiferente tomar un promedio de $2m+1$ observaciones consecutivas o separadas doce períodos de tiempo entre sí.

Como el promedio se refiere siempre al mismo mes, se tiene que si el proceso S_t evoluciona suavemente,

$$\frac{1}{2m+1}\sum_{j=-m}^{m} S_{t-12j}^l=S_t^l$$

de donde

$$Y_t^*=\frac{1}{2m+1}\sum_{j=-m}^{m} Y_{t-12j}^l\simeq S_t^l$$

y ahora el resultado de la media móvil se puede usar para aproximar el componente estacional subyacente en la observación del momento t.

Los ejemplos anteriores han servido para introducir, de forma intuitiva, la idea de que una media móvil permite amplificar cierto tipo de características de una serie, al tiempo que atenúa otras; en ellos se vio que una media móvil adecuadamente elegida permitía recalcar aquellos aspectos de interés.

Más formalmente, una media móvil es un filtro lineal, es decir, una combinación lineal de valores de una variable temporal referidos a distintos momentos del tiempo, y cuyo resultado es otra variable temporal. Como tal, toda media móvil está plenamente caracterizada por el polinomio temporal que la define, pues

$$\begin{aligned}\sum_{j=-m}^{m} a_j Y_{t-j} &= \cdots+a_2 Y_{t-2}+a_1 Y_{t-1}+a_0 Y_t+a_{-1} Y_{t+1}+a_{-2} Y_{t+2}+\cdots = \\ &=(\cdots+a_2L^2+a_1L+a_0+a_{-1}L^{-1}+a_{-2}L^{-2}+\cdots)\,Y_t= \\ &=(\cdots+a_2L^2+a_1L+a_0+a_{-1}F+a_{-2}F^2+\cdots)\,Y_t= \\ &=a(L,F)\,Y_t\end{aligned} \tag{4.2.3}$$

donde $L^{-j}=F^j$ es el operador de adelantos, tal que$L^{-j}Y_t=F^jY_t=Y_{t+j}$.

Cuando $a_j=a_{-j}$, $\forall j$, la media móvil (4.2.3) es una *media móvil simétrica.*

La utilización de medias móviles para la extracción de señales no observables en procesos estacionarios fue formalizada, en el dominio de las frecuencias, por Wiener (1949) y Kolmogorov (1939, 1941): un tratamiento completo del problema aparece en Whittle (1984). La extensión a procesos no estacionarios pero integrados se encuentra en Bell (1984) y Maravall (1988b).

El uso de medias móviles para extraer señales plantea problemas al principio y al final de la muestra. Para simplificar, supóngase que se trata de aproximar un determinado componente C_t mediante la expresión:

$$C_t = (1/5)[Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2}]. \qquad (4.2.4)$$

Si se dispone de observaciones de Y para el intervalo $t = 1, 2, \ldots, T$, no hay dificultad en calcular los correspondientes valores de C_t para $t = 3, 4, \ldots, T-2$.

Pero aun admitiendo que los valores de C_t correspondientes al principio de la muestra sean de poco interés para el analista, con carácter general uno de los objetivos más habiltuales cuando se calcula una señal es la determinación de sus valores correspondientes al presente (T) y pasado inmediato (T-1), lo que obviamente es imposible a partir de (4.2.4). Esto da lugar a lo que se conoce como el coste informativo asociado al uso de cualquier media móvil del tipo $\sum_{j=-m}^{m} a_j Y_{t-j}$.

En el análisis aplicado la determinación del valor de la señal para los extremos de la muestra se realiza utilizando algún tipo de predicción de los valores que son desconocidos, lo que en definitiva supone truncar el filtro $a(L, F)$ para considerar sólo las observaciones en T y anteriores. Nótese que mientras truncar el filtro sustituyendo los valores futuros desconocidos por cero es una forma muy mala de proceder, el uso de predicciones óptimas, calculadas a partir del correspondiente modelo ARIMA, supone sustituir el verdadero filtro por el filtro truncado más adecuado.

En la literatura sobre medias móviles se utiliza la siguiente terminología:

- *Valor definitivo* o *estimador definitivo*: cuando el filtro utiliza exclusivamente valores verdaderos de la variable observada.
- *Valor condicional* o *estimador condicional*: cuando debido a la falta de información se sustituyen valores desconocidos por predicciones.
- *Valor concurrente* o *estimador concurrente*: es un caso particular de estimador condicional, cuando sólo se dispone de observaciones de la variable agregada hasta el mismo momento para el que se quiere estimar la señal. En los estimadores concurrentes las predicciones constituyen la mitad menos uno de los valores que se utilizan para el cálculo del componente.
- *Revisión* o *actualización*: es el proceso por el cual se sustituyen las predicciones por los valores observados a medida que éstos se van conociendo; esto permite ir mejorando las estimaciones del valor del componente referido a un momento concreto del tiempo, hasta llegar finalmente al valor definitivo.

- *Error de revisión*: es la diferencia entre el valor definitivo y un valor condicional. En el análisis aplicado merece especial atención el error de revisión correspondiente al estimador concurrente.

En la actualidad toda extracción de señales de una serie temporal, y en particular su descomposición en tendencia, estacionalidad e irregular, se basa en: 1) definir la media móvil —el filtro— adecuado para resaltar ese componente (la señal), y 2) aplicar dicho filtro a la serie observada.

Los próximos epígrafes se dedicarán al estudio de los principales métodos que se emplean para definir las medias móviles a utilizar en cada caso en la descomposición de una serie temporal. Butter y Fase (1991) ofrece también una compilación muy exhaustiva de procedimientos de extracción de señales orientados al ajuste estacional.

4.3. Procedimientos empiricistas

4.3.1. Los métodos X-11 y X-11ARIMA: su importancia en la historia del análisis de coyuntura

Un primer grupo de métodos de descomposición de series temporales está formado por los procedimientos empiricistas, que reciben este nombre por haber sido desarrollados a partir del análisis empírico de un gran número de series reales, sin hacer referencia explícita a ningún tipo de modelo teórico de generación de los datos.

De entre los procedimientos empiricistas los más conocidos son el *método X-11* y su derivado *X-11ARIMA*. El primero nace como un método automático de desestacionalización, basado en el uso intensivo del ordenador y capaz de producir resultados para un número grande de series con un mínimo de intervención humana, y supone un cambio radical respecto a lo que había sido hasta el momento de su aparición el modo habitual de proceder.

A mediados de los cincuenta los procesos de desestacionalización que se utilizaban primaban el estudio detallado de cada serie para determinar el tipo de operaciones a realizar, pero éstas tenían que ser sencillas ya que se hacían usando calculadoras mecánicas. Como es lógico, el número de series que se podían desestacionalizar era forzosamente reducido, ya que cada caso requería el análisis específico de un experto. Esta restricción, en un momento en que el análisis de la actividad a corto plazo estaba basado en un número cada vez mayor de indicadores económicos, provocaba un gran número de quejas por parte de los analistas económicos.

La aparición de los ordenadores propició el enfoque contrario: los cálculos ya no eran problema, pero para poder desestacionalizar un número grande de series se hacía necesario reducir el tiempo dedicado a estudiar detalladamente cada caso particular. Esto a su vez requería automatizar gran parte de las decisiones subjetivas que hasta ese momento tomaba el experto, para lo cual el procedimiento de desestacionalización tenía que ser lo suficientemente flexible como para admitir distintas variantes; además, el propio método habría de seleccionar la variante concreta a emplear en cada caso en función de una serie de criterios previamente establecidos.

En 1954 Shiskin, del Bureau of the Census estadounidense, empezó a trabajar en un proceso de desestacionalización basado en las ideas expuestas en el párrafo anterior, un trabajo que culmina en 1965 con el método X-11[7]. El tipo de técnicas utilizadas no era especialmente novedoso[8], pero su gran aportación fue que todo el proceso de desestacionalización estaba automatizado, tanto para series mensuales como trimestrales; esto posibilitaba su aplicación a decenas o incluso cientos de series a un coste accesible.

Una consecuencia de esa automatización, no del todo deseable, fue la de desviar la atención de los usuarios de la técnica estadística utilizada para obtener la señal de interés. El problema no es tanto que el usuario medio no sepa cómo funciona exactamente el método X-11, sino que la mayoría no conoce ni siquiera sus principales rasgos y, por tanto, no puede opinar sobre si el procedimiento puede ser inadecuado para su caso concreto.

El método X-11ARIMA fue desarrollado por E. Dagum de Statistics Canada[9], y es una modificación y perfeccionamiento del anterior. Su principal aportación es añadir un procedimiento automático para construir un modelo ARIMA para la serie a descomponer. Con este modelo se predicen los valores correspondientes a un año en cada extremo de la muestra, y se extraen los componentes utilizando estas predicciones como si fuesen valores reales, lo que permite reducir los errores de revisión.

4.3.2. Descripción de los métodos X-11 y X-11ARIMA

A grandes rasgos, los dos métodos, X-11 y X-11ARIMA, se basan en un proceso iterativo de aplicación de distintos tipos de medias móviles. Este proceso es difícil de resumir, ya que en parte depende de las características de la serie original; así en lo que queda de este epígrafe se supondrá que la serie original es mensual, que no muestra oscilaciones cíclicas de medio plazo y que su componente irregular es poco importante. El procedimiento de cálculo se articula, de forma resumida, en los siguientes ocho pasos[10]:

a) Calcular las diferencias entre la serie original y una media móvil ponderada y centrada de doce términos (es decir, una media móvil 2×12, un promedio de dos términos que a su vez son promedios de doce términos a los que se asigna distinta ponderación), como una primera aproximación de la suma de los componentes estacional e irregular.

b) Calcular una media móvil estacional ponderada 3×3 para cada mes por separado, lo que proporciona una estimación del componente estacional.

c) Ajustar estos componentes estacionales para que sumen cero (aproximadamente) en cualquier período de doce meses, sustrayéndole una media móvil 2×12.

d) Restar a la serie original el componente estacional ajustado, para obtener así una serie preliminar ajustada de estacionalidad.

e) Aplicar una media móvil de Henderson de 9, 13 ó 23 términos a la serie ajustada de estacionalidad, y restar de la serie original esta serie de *tendencia-ciclo* resultante, para obtener una segunda estimación de los componentes estacional e irregular.

f) Calcular una media móvil ponderada 3×5 a cada mes por separado, para llegar a una segunda estimación del componente estacional.

g) Ajustar este componente estacional para que sume cero (aproximadamente) en cualquier período de doce meses, sustrayendo una media móvil centrada de doce términos, obteniendo así el *componente estacional* propiamente dicho.

h) Restar de la serie original estas últimas estimaciones del componente estacional, operación que proporciona la *serie ajustada de estacionalidad*.

Dejando de lado ciertos detalles como el tratamiento de los valores anómalos, la corrección de los totales anuales y la modificación del proceso al principio y al final de la muestra, Wallis (1974) demostró que la serie ajustada de estacionalidad obtenida en el último paso

[7] Shiskin et al. (1967).

[8] Con la excepción del tratamiento del efecto calendario, que incorporaba desarrollos muy recientes para la época como los de Eisenpress (1956), Marris (1960) y Young (1965).

[9] Dagum (1980, 1988).

[10] La formulación de estos pasos está tomada de Wallis (1974); véase así mismo Espasa (1977).

es el resultado de aplicar un filtro de medias móviles simétricas a la serie original, es decir

$$Y_t^a = \sum_{j=-m}^{m} a_j Y_{t-j}, \qquad a_j = a_{-j}$$

donde $m = 82$, 84 ó 89, según el valor de la media móvil de Henderson en la etapa *e*. En el mencionado trabajo de Wallis se representan gráficamente los coeficientes a_j cuando m es igual a 84.

La discusión anterior resume la forma de operar de estos métodos. Sin embargo las series reales presentan una serie de peculiaridades que deben atenderse específicamente.

Entre estas peculiaridades, se puede destacar en primer lugar el que casi todas las series económicas, en un momento u otro, presentan valores anómalos, originados por la aparición de choques exógenos de carácter puntual o permanente no ligados a la evolución normal del fenómeno. Estas anomalías han de ser eliminadas antes de aplicarle a la serie el filtro mencionado, al menos en aquellos casos en que se puedan estimar los efectos de estos acontecimientos especiales. Por esa razón, ha de existir un primer paso donde se procede al tratamiento de las anomalías.

En segundo lugar, las series reales combinan componente tendencial con componente estacional. Si una determinada serie presenta sólo uno de estos componentes, cualquiera que sea, es fácil aproximarlo mediante medias móviles adecuadas. Pero el problema es más complejo cuando se presentan los dos a la vez, ya que para aproximar el componente estacional se deben aplicar medias móviles a la serie desprovista de tendencia; y para haber eliminado el componente tendencial es preciso haberlo calculado previamente, lo que se hace filtrando la serie original desprovista de estacionalidad. Por lo tanto hay que plantear un procedimiento iterativo que en las primeras etapas proporcione aproximaciones brutas de los componentes, para llegar poco a poco a las estimaciones definitivas.

Por estas razones en la práctica los métodos X-11 y X-11ARIMA se estructuran en cuatro grandes fases:

Fase *A*) *Modificaciones a priori de la serie original*: aquí el usuario introduce la corrección de valores atípicos antes mencionada y cualesquiera otros factores deterministas a priori que requiera la serie.

El método incorpora en fases posteriores una corrección automática de la serie, detectando por sí mismo valores anómalos y corrigiéndolos en función de ciertos criterios preestablecidos. Sin embargo cualquier análisis de anomalías de los que hace el programa se puede realizar de forma más eficiente mediante el análisis de intervención (véase la sección 2.6). En consecuencia es preferible, de acuerdo con la recomendación dada al final del primer epígrafe de este capítulo, estimar explícitamente de forma eficiente la incidencia de los factores deterministas, mediante un modelo ARIMA con Análisis de Intervención, e introducir como input en el procedimiento X-11 la serie ya corregida, en vez de dejar al método que la corrija por procedimientos menos eficientes.

Esta serie corregida se puede introducir de dos formas:

1. trasladando las correcciones implícitas en el análisis de intervención a una serie de coeficientes de ajuste a priori, y dejando que el método X-11 corrija la serie original con esos coeficientes; o
2. realizando la corrección con anterioridad, en la línea de lo propuesto en el epígrafe uno, y usando como serie «original observada» directamente la serie corregida, sin facilitar ningún tipo de factores a priori al procedimiento X-11.

En la variante X-11ARIMA, durante esta primera fase también se selecciona de forma automática un modelo ARIMA para la serie original, y se obtienen predicciones para extender la muestra en los extremos. Al igual que sucedía con la modelización de acontecimientos especiales, es más eficiente especificar, estimar y validar el modelo ARIMA con anterioridad, y obtener las predicciones a partir de dicho modelo. Si bien esto requiere la intervención de un experto, los cambios pueden ser considerables si la serie no sigue un modelo sencillo[11].

El resultado finalmente obtenido es la serie que realmente actuará como input del procedimiento de descomposición propiamente dicho; y en las siguientes etapas la expresión «serie original» se refiere a la serie corregida de factores a priori y ampliada con predicciones que se obtiene en esta primera fase.

Fase *B*) *Estimación preliminar de observaciones atípicas y efectos de calendario*: en esta fase tiene lugar la primera iteración de las ocho etapas que se discutieron anteriormente. El objetivo fundamental es lograr una primera aproximación, todavía muy provisional, de los distintos componentes, y obtener estimaciones iniciales de las irregularidades que presenta la serie.

Como es lógico, si se ha corregido previamente la serie usando los resultados de un análisis de intervención, apenas se observarán irregularidades; en cambio, si no se incorpora ningún tipo de corrección a priori, en esta fase se detectan las principales anomalías que experimenta la serie.

[11] Véase Dagum (1988, págs. 1-2 y 12-15) para los modelos disponibles en el método X-11ARIMA/88, que es la versión más reciente.

Es importante recalcar una vez más la necesidad de la corrección previa: para muchas de las anomalías, y en general para todas las realmente importantes, se dispone de información extramuestral que permite su modelización eficiente.

En un análisis de intervención toda la información disponible se maneja a la hora de construir las variables artificiales y definir los filtros dinámicos asociados a ellas; en cambio el procedimiento automático de detección de valores anómalos incorporado al método X-11 no la usa, por lo que en general la serie corregida que con él se obtiene será menos fiable.

En cuanto al efecto calendario, es un tipo especial de estacionalidad determinista que se estudiará en el epígrafe octavo de este capítulo. Por las mismas razones que rigen para los demás factores deterministas, es conveniente no estimarlo de forma automática por este método.

Fase *C*) *Estimación final de observaciones atípicas y efecto calendario*: nueva iteración en la que se repiten los ocho pasos anteriores con el fin de llegar a la estimación final de todo tipo de efectos deterministas. Con ello se puede definir una serie puramente estocástica, corregida de anomalías y de estacionalidad determinista.

Fase *D*) *Estimación final de los distintos componentes:* se aplica por última vez el proceso de ocho pasos, ahora sobre la serie corregida con los resultados de la fase anterior. Al acabar dicho proceso se tienen las estimaciones últimas de tendencia (en el paso *e*), factores estacionales (paso *g*), componente irregular (paso *g*) y serie ajustada de estacionalidad (paso *h*), así como una serie de contrastes y estadísticos útiles para valorar los resultados.

En resumen, el procedimiento X-11, tal y como se aplica en la práctica, se divide en cuatro grandes fases, donde en las tres últimas se repite el proceso en ocho pasos señalado. A su vez estos ocho pasos se pueden agrupar en dos, ya que los cuatro primeros suponen una estimación preliminar de los componentes y los cuatro últimos una estimación final (que sólo es realmente final en la fase *D*).

Cuando se introduce la serie corregida de factores deterministas, tal y como se propone en este libro, las fases *B*, *C* y *D* se parecen mucho, y de ahí que convenga aclarar este punto.

Hay que tener en cuenta que las dos primeras, y especialmente la *B*, añaden al proceso de ocho pasos un tratamiento bastante complejo de las anomalías: así la fase *B* introduce procesos de detección de éstas en los pasos *b*, *e* y *f*. La obtención de los componentes estocásticos realmente no empieza hasta que se considera que la serie está limpia del efecto de dichos factores deterministas, y esto no ocurre hasta el final de la fase *C*.

Por lo tanto, si se dispone de un modelo eficiente que permite cuantificar la influencia de los factores deterministas y obtener predicciones, y si ese modelo se utiliza para corregir previamente la serie y aplicar el método X-11 únicamente a la parte estocástica de la misma, entonces las diferencias entre comenzar por la fase A o hacerlo directamente por la D serán pequeñas.

4.3.3. Los procedimientos empiricistas y los modelos ARIMA

Tal y como se ha dicho al principio de este epígrafe, los métodos X-11 y X-11ARIMA son puramente empiricistas, en el sentido de que su desarrollo se ha producido sin ningún tipo de modelo teórico detrás, basándose exclusivamente en la experiencia de sus autores, adquirida con el contacto con cientos de series distintas a lo largo de muchos años.

Ahora bien, una cosa es que en el desarrollo del método no haya habido ningún, o casi ningún, tipo de consideración teórica, y otra que la forma de proceder del método no tenga una interpretación relacionada con los procesos generadores de los datos económicos.

Por el contrario, si un método puramente empiricista proporciona, para la mayor parte de las series, resultados consistentes con los que cabría esperar de un análisis teórico, parece lógico suponer que existe alguna relación entre las operaciones a las que se ha llegado por ese procedimiento de prueba y error, y las operaciones a las que se llegaría a partir de un desarrollo teórico basado en el proceso generador de los datos.

Más concretamente, cabe preguntar para qué clase de modelo teórico se obtienen resultados para la extracción de señales idénticos a los que obtiene el método X-11, es decir, para qué tipo de modelos la extracción de señales vía X-11 es óptima. Este es un tema que ha sido muy tratado en la literatura[12], y aquí se hará un breve resumen del mismo.

La forma de contestar a esta pregunta es intuitivamente sencilla: las medias móviles que el método X-11 aplica son conocidas[13], y también se pueden calcular las medias móviles óptimas para procesos generadores de datos representados por modelos ARIMA. Se trata entonces de determinar el modelo ARIMA concreto para el cual

[12] Véanse los trabajos pioneros de Cleveland (1972), Cleveland y Tiao (1976), así como el más reciente de Burridge y Wallis (1984). Las referencias Espasa (1984a) y Espasa y Galián (1985) pueden ser también útiles sobre este tema.

[13] La flexibilidad del método X-11 permite realizar ciertas modificaciones que naturalmente van a influir en los valores concretos de los coeficientes de las medias móviles finales. No obstante, las conclusiones del texto son bastante robustas frente a especificaciones alternativas de las opciones del X-11.

ambos filtros de extracción de señales, el del modelo y el del método X-11, coincidan[14].

El modelo ARIMA para el que el método X-11 es óptimo es un modelo que contiene $\Delta\Delta_{12}$, una diferencia regular y una estacional, y cuya estructura estocástica estacionaria es un proceso de medias móviles cuyo orden puede llegar a ser igual a 25. Sin embargo, los únicos coeficientes de este último proceso que toman valores claramente distintos de cero son θ_1 θ_{12} y θ_{13}, correspondientes a los retardos 1, 12 y 13. Además, dados los valores concretos de estos parámetros, no resulta ser un modelo muy distinto de

$$\Delta\Delta_{12} Y_t = (1-\theta_1 L)(1-\theta_{12} L^{12}) a_t = \\ = (1-\theta_1 L - \theta_{12} L^{12} + \theta_1 \theta_{12} L^{13}) a_t .$$

Este resultado explica porqué el método X-11 en general proporciona buenos resultados. Los procesos generadores de muchas series económicas, aunque sean modelos ARIMA más complejos, son razonablemente bien aproximados por este modelo sencillo, propuesto por Box y Jenkins y que en la literatura recibe el nombre de modelo de las líneas aéreas[15].

De ahí que si el modelo de las líneas aéreas por un lado aproxima bien los procesos generadores de muchas series económicas, y por otro es una buena aproximación al modelo para el cual la extracción de señales implícita en el método X-11 es óptima, entonces no es difícil explicar el éxito del procedimiento propuesto por Shiskin. Obsérvese, sin embargo, que el modelo ARIMA que justifica el procedimiento X-11 no es exactamente el modelo de las líneas aéreas, por lo que cuando una serie viene generada por ese modelo se tiene que el método X-11 es una buena aproximación, pero se pueden obtener resultados más correctos aplicando los procedimientos basados en modelos, que serán descritos en las dos secciones siguientes.

En consecuencia es posible adelantar algo sobre los resultados que son de esperar de la aplicación de los métodos X-11 y X-11ARIMA a una serie económica concreta. Cuando el comportamiento de la serie se pueda describir razonablemente bien mediante el modelo de las líneas aéreas, se puede confiar en los resultados que se obtendrán con dichos métodos empiricistas de extracción de señales. Al contrario, cuando el proceso generador de los datos sea muy distinto del modelo ARIMA (0, 1, 1) regular y estacional señalado, cabrá esperar resultados muy poco satisfactorios de la aplicación de los métodos X-11 y X-11ARIMA.

[14] Hay un punto de arbitrariedad en las medias móviles óptimas para la extracción de señales a partir de un modelo ARIMA dado, ya que los parámetros de los filtros a utilizar no están identificados: la hipótesis adicional que se considera para permitir la identificación es la de maximizar la varianza del elemento irregular. En el próximo epígrafe se estudiará con detalle esta cuestión.

[15] El nombre se debe a la serie temporal que Box y Jenkins (1976) utiliza como ejemplo para presentar y discutir las características de este modelo.

Otro punto en que el conocimiento del proceso generador de los datos, o al menos una estimación del mismo, puede contribuir a mejorar la actuación del método X-11 es en la elección de la longitud de determinadas medias móviles. Burridge y Wallis (1984) demuestran que, usando las opciones estándar[16], el método X-11 es cuasi-óptimo para desestacionalizar una serie que sigue el modelo

$$\Delta\Delta_{12} Y_t = (1-0{,}67L)(1-0{,}33L^{12}) a_t . \tag{4.3.1}$$

Si, por ejemplo, la serie a tratar está generada por un modelo con un valor de θ_{12} claramente mayor que 0,33, entonces su componente estacional es más suave que el del modelo de referencia (4.3.1); en este caso es preferible usar una media estacional más amplia, y en consecuencia se elegirá la opción (3 × 9) que permite el procedimiento X-11[17].

El método dispone de un procedimiento automático de selección de la longitud más adecuada para las distintas medias móviles que maneja, pero puede ser preferible imponer directamente dicha longitud en función de los resultados de un modelo univariante cuando éste merezca total confianza.

A modo de conclusión, y terminando un recorrido histórico que comenzó en el subepígrafe 4.3.1, el método X-11 jugó un papel muy destacado en la historia de la descomposición de series temporales económicas. Sin este método el análisis de coyuntura no se hubiera desarrollado a la velocidad en que lo hizo, y muchos resultados de la teoría estadística de extracción de señales todavía serían desconocidos para los economistas.

El hecho de que las medias móviles empleadas, obtenidas mediante tanteos y por el procedimiento de prueba y error, se puedan interpretar como filtros óptimos para el tipo de modelos ARIMA más frecuentemente observados en el análisis aplicado, es una propiedad muy destacable, que sirve para valorar adecuadamente el mérito de sus creadores.

[16] En esencia una media móvil de Henderson de 13 términos para la tendencia y una media móvil ponderada 3 × 5 para el componente estacional. Además Burridge y Wallis distinguen si se va a usar un filtro simétrico o asimétrico, y en este último caso consideran dos tipos distintos de filtros que se diferencian en el número de valores futuros que se suponen conocidos. Toda la discusión del texto se refiere a la utilización del filtro simétrico.

[17] Estamos agradecidos a W. Bell por realizarnos esta sugerencia.

No obstante, es evidente que los métodos X-11 y X-11ARIMA, y como ellos todos los procedimientos empiricistas, empezaron a ser historia desde el momento en que se reconoció su optimalidad para un cierto tipo de modelos. Pues con dicho reconocimiento llegaba la constatación de que la teoría estadística de extracción de señales estaba lo suficientemente desarrollada como para calcular explícitamente el filtro más adecuado en cada caso concreto.

De ahí que desde mediados de los años setenta se produzca un desplazamiento desde un tratamiento general como el que supone el X-11 a filtros específicos; aquí la especifidad se entiende en el sentido de que, en vez de definir filtros de carácter general válidos para un amplio conjunto de series, se diseñan los filtros más adecuados teniendo en cuenta las peculiaridades de la serie a analizar.

Hoy por hoy los métodos X-11 y X-11ARIMA siguen siendo los más utilizados para calcular series desestacionalizadas, componentes estacionales y otras señales, tanto por instituciones públicas como privadas en todo el mundo. Sin embargo, en el futuro otro tipo de procedimientos más específicos tomará su lugar, a medida que se vayan popularizando entre los encargados de elaborar series desestacionalizadas y los usuarios de las mismas. De estos nuevos procedimientos tratan los dos próximos epígrafes.

4.4. Métodos basados en el modelo ARIMA de la serie original (métodos de forma reducida)

4.4.1. Fundamento teórico de los métodos de forma reducida

El segundo gran grupo de métodos de descomposición son los *métodos basados en modelos de forma reducida*. El problema se plantea en los siguientes términos: se observa una serie Y_t que es la suma de tres componentes no observables

$$Y_t = T_t + S_t + I_t, \tag{4.4.1}$$

y se trata de estimar las señales T_t, S_t y I_t a partir de la serie observada aplicando la teoría de extracción de señales, de tal manera que se incorpore explícitamente en el diseño de los filtros a utilizar información sobre los procesos generadores de la variable Y y de sus componentes: de ahí la denominación de métodos basados en modelos.

Para ello se dispone de estimaciones de los parámetros del modelo ARIMA de la serie observada, Y_t, a partir del cual se han de estimar: *a*) los parámetros de los modelos de los componentes, *b*) los filtros adecuados para estimar los componentes, y *c*) las series históricas de los componentes dadas las observaciones del agregado.

El calificativo de «forma reducida» se debe a que el problema es similar al que se plantea en los modelos econométricos estructurales: en éstos se puede determinar de manera única la forma reducida sin restringir y a partir de ella se trata de estimar, incorporando información extramuestral, la estructura concreta, dentro de las muchas posibles, que ha generado dicha forma reducida. Al igual que en los modelos econométricos estructurales, se plantea un problema de identificación, ya que existen infinitas estructuras —en este capítulo diferentes descomposiciones en tendencia, componente estacional y elemento irregular— igualmente compatibles con el modelo ARIMA de la serie observada. Esto requiere introducir un supuesto especial, el llamado requisito canónico, para garantizar la identificación de todos los parámetros implícitos en la expresión (4.4.1).

Si bien el desarrollo matemático que sigue es especialmente complejo, la idea básica de este tipo de métodos es clara: las señales no se obtienen utilizando medias móviles que se han construido a base de analizar cientos de series distintas y elegir los filtros más adecuados en conjunto, como sucede con los procedimientos del epígrafe anterior. Aquí se tienen en cuenta las características específicas de la serie temporal concreta que se pretende descomponer, tal y como se recogen en su modelo ARIMA, y a partir de estas características se diseña la media móvil más adecuada para el problema individual que tenemos planteado.

Las referencias básicas para este enfoque de la descomposición de series económicas son, entre otras, Box, Hillmer y Tiao (1978), Burman (1980), Hillmer y Tiao (1982), Bell y Hillmer (1984), Maravall (1987) y Maravall y Pierce (1987).

4.4.2. Propiedades teóricas de los componentes no observables (*)

Para facilitar la exposición de los fundamentos de este enfoque, se realizará la misma refiriéndola a un ejemplo concreto. Así, sea una serie observada Y_t cuyo proceso generador de datos

$$\phi_Y(L)\, Y_t = \theta_Y(L)\, a_t,$$

resulta ser

$$\Delta\Delta_{12}\, Y_t = (1-\theta_1 L)(1-\theta_{12} L^{12})\, a_t \tag{4.4.2}$$
$$a_t \sim \text{Niid}(0,\ \sigma_a^2)$$

es decir, un modelo de líneas aéreas, al que se le añade la restricción adicional $\theta_{12} \geqslant 0$. Supóngase también que se dispone de una serie temporal de longitud infinita de este proceso, es decir, de un conjunto de observaciones del mismo dado por

$$\ldots, Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}, \ldots$$

Para descomponer esta serie agregada en sus componentes de acuerdo con la ecuación (4.4.1), es preciso definir inequívocamente qué se entiende por tendencia, estacionalidad y componente irregular; como tales, se entienden tres variables temporales caracterizadas por:

1. Cada componente no observable sigue un modelo ARIMA bien definido de la forma

$$\begin{aligned} \phi_T(L)\,T_t &= \theta_T(L)\,b_t \qquad & b_t &\sim \mathrm{Niid}(0, \sigma_b^2) \\ \phi_S(L)\,S_t &= \theta_S(L)\,c_t \qquad & c_t &\sim \mathrm{Niid}(0, \sigma_c^2) \\ \phi_I(L)\,I_t &= \theta_I(L)\,d_t \qquad & d_t &\sim \mathrm{Niid}(0, \sigma_d^2) \end{aligned} \tag{4.4.3}$$

con parámetros desconocidos y donde las innovaciones b_t, c_t y d_t son independientes entre sí.

El imponer que cada componente también sigue un modelo ARIMA es consecuencia inmediata de (4.4.1) y (4.4.2), así como la normalidad de todas las variables implicadas en el análisis.

La indepedencia de las innovaciones, que implica la independencia de los componentes, es un supuesto presente en la literatura desde el mismo momento en que se empezó a interpretar una serie observada como la agregación de componentes no observables. De hecho, la justificación de la descomposición reside precisamente en la necesidad de separar evoluciones muy dispares que se consideran independientes entre sí, pero que sólo se observan en conjunto.

Merece destacarse que, si bien el elemento irregular I_t será normalmente ruido blanco, esto es algo que surge del análisis, pero que no se impone desde el principio: véase Burman (1980, sección 3) o Hillmer y Tiao (1982, sección 1).

2. Los polinomios autorregresivos están relacionados mediante

$$\phi_Y(L) = \phi_T(L) \cdot \phi_S(L) \cdot \phi_I(L) \tag{4.4.4}$$

y no hay raíces comunes entre los polinomios del lado derecho de la ecuación (4.4.4).

Este requisito se impone para garantizar que $\phi_Y(L)$ se puede factorizar de tal forma que todas y cada una de las raíces se puedan asignar a un y sólo un componente.

En el caso particular del modelo (4.4.2), por lo discutido en el capítulo 2 resulta claro que las raíces de $\phi_Y(L) = \Delta\Delta_{12} = \Delta^2 U_{11}(L)$ se han de asignar de manera que $\phi_T(L) = \Delta^2$, $\phi_S(L) = U_{11}(L)$ y $\phi_I(L) = 1$. Para ello se tiene en cuenta el tipo de ciclo que dichas raíces representan, y se asigna cada raíz al componente al que teóricamente corresponde ese ciclo. Cuando la parte autorregresiva sólo contiene raíces unitarias, las reales positivas —que generan ciclos de periodicidad infinita— se adjudican al componente tendencial; la real negativa y las complejas —que generan ciclos estacionales (véase el apéndice al capítulo 2)— al componente estacional; y ninguna al componente irregular.

Precisamente por esta razón no siempre está definida la descomposición por el método basado en modelos de forma reducida: si $\phi_Y(L)$ contiene una raíz cuya asignación no está clara, y por lo tanto resulta arbitraria, diferentes analistas pueden llegar a distintas descomposiciones.

Supóngase, por ejemplo, que $\phi_Y(L)$ contiene un término autorregresivo $(1 - \phi_1 L - \phi_2 L^2)$, que impone a la serie un movimiento cíclico de módulo 0,1 y período tres meses y medio: su asignación podría hacerse a $\phi_S(L)$ o a $\phi_I(L)$. Este ejemplo sirve para ilustrar el hecho de que la extracción de señales es más compleja de lo que los procedimientos empiricistas pueden reflejar y que hay situaciones concretas que pueden necesitar el asesoramiento de un experto.

Por ese motivo se ha elegido presentar el método a partir de un ejemplo concreto, el recogido en (4.4.2), y no a partir de un modelo genérico $\phi_Y(L)\,Y_t = \theta_Y(L)\,a_t$. En todo caso se intentará dejar claro al lector, cuando pueda haber confusión al respecto, qué aspectos son de validez general y cuáles son específicos al ejemplo empleado. En el modelo de las líneas aéreas con $\theta_{12} \geqslant 0$ la descomposición (4.4.1) de acuerdo con (4.4.3) es siempre posible, y de ahí que se haya seleccionado para ilustrar el procedimiento. Al final del epígrafe se tratará el tema desde una perspectiva más general.

3. El orden de $\theta_T(L)$ es menor o igual a 2 (el orden del polinomio $\phi_T(L)$).

4. El orden de $\theta_S(L)$ es menor o igual a 11 (el orden del polinomio $\phi_S(L)$).

Las propiedades tercera y cuarta procuran mantener los modelos de los componentes lo más sencillos posible, de acuerdo con el principio de parquedad en la parametrización. Esto tiene importantes implicaciones para la función de predicción que se deriva para cada uno de los modelos: por ejemplo en el caso del componente estacional, las predicciones hechas con información hasta t siguen un patrón estacional fijo de período s, que naturalmente se actualiza a medida que se cambia el origen de la predicción. Es decir, si los órdenes de los polinomios de las medias móviles no superan el de sus correspondientes polinomios autorregresivos, las funciones de predicción en cada caso tendrán una formulación del tipo (2.8.1) desde el horizonte uno, o, en otras palabras, las soluciones generales de las correspondientes ecuaciones en diferencias finitas serán válidas desde el principio.

5. Se maximiza la varianza de la innovación del componente irregular, σ_d^2 (requisito canónico).

Se puede demostrar que si sólo se imponen las cuatro primeras características la descomposición (4.4.1) no está identificada, ya que existen en (4.4.3) infinitas combinaciones de tendencia, estacional e irregular que las cumplen y que sin embargo generan el mismo agregado Y_t. Para identificar los componentes se introduce esa quinta restricción, que se conoce en la literatura como el requisito canónico.

En otras palabras, esta propiedad del elemento irregular implica concentrar la máxima aleatoriedad posible en él. Visto de otra forma, se trata de que los otros dos componentes, que son las señales sistemáticas, muestren una evolución lo más parecida a una evolución puramente determinista.

Aun siendo arbitrario, éste parece un tipo de restricción bastante razonable, ya que el componente irregular se define intuitivamente como el factor que tiende a concentrar las oscilaciones de muy corto plazo que muestra el agregado.

4.4.3. Determinación de los parámetros de los modelos de los componentes (**)

Para ilustrar el procedimiento seguido para la determinación de los parámetros de los modelos de los componentes es necesario hacer referencia a conceptos, como función generadora de autocovarianzas o espectro, que no son tratados en este libro, por lo que el lector no familiarizado con tales nociones puede, sin pérdida de continuidad, saltar este subepígrafe.

A partir de la hipótesis de independencia de los componentes, es inmediato demostrar que las autocovarianzas de Y y las de los componentes cumplen la propiedad

$$\mathrm{cov}(Y_t, Y_{t-k}) = \mathrm{cov}(T_t, T_{t-k}) + \mathrm{cov}(S_t, S_{t-k}) + \mathrm{cov}(I_t, I_{t-k})$$

para todo t y k. Esto se puede escribir en términos de las funciones generadoras de autocovarianzas como

$$\Gamma_Y(t, k) = \Gamma_T(t, k) + \Gamma_S(t, k) + \Gamma_I(t, k)$$

donde $\Gamma_i(t, k)$ es la función generadora de autocovarianzas de la variable i; nótese que es una función de dos argumentos, ya que se están considerando procesos no estacionarios[18]. Tomando transformadas de Fourier se obtiene

$$h_Y(z) = h_T(z) + h_S(z) + h_I(z), \qquad (4.4.5)$$

donde $z = \cos(w)$ y $h_i(z)$ es el (pseudo)espectro de i. Dada la asignación de la segunda propiedad teórica de los componentes, las limitaciones de la tercera y cuarta y el requisito canónico, se puede calcular a partir de (4.4.5) los parámetros de los polinomios $\theta_i(L)$, $i = T, S, I$, y las varianzas de las innovaciones σ_b^2, σ_c^2 y σ_d^2.

4.4.4. Cálculo de los filtros a emplear y aproximación de las series históricas de los componentes no observables (*)

Una vez determinados los parámetros de los modelos ARIMA de (4.4.3), el siguiente paso es llegar a aproximar los valores de los componentes correspondientes a una serie temporal dada de Y.

Sin pérdida de generalidad, considérese el componente estacional S_t[19]: se trata de diseñar la forma de aproximarlo mediante una variable a definir, S_t^*, es decir, determinar la función g_S

$$S_t^* = g_S(\ldots, Y_{t-1}, Y_t, Y_{t+1}, \ldots),$$

tal que el error cuadrático medio de aproximación sea mínimo: esto es,

$$\min_{g_S(.)} E[S_t - S_t^*]^2 = \min_{g_S(.)} E[S_t - g_S(\ldots, \ldots, Y_{t-1}, Y_t, Y_{t+1}, \ldots)]^2,$$

Es sabido que la aproximación $g_S(.)$ óptima es la esperanza de S_t condicionada al conjunto de información empleado:

$$S_t^* = E(S_t / \ldots, Y_{t-1}, Y_t, Y_{t+1}, \ldots).$$

Bajo el supuesto de normalidad de todas las variables, la esperanza condicionada es una función lineal del conjunto de información,

$$S_t^* = \Sigma \alpha_j Y_{t-j} = \alpha(L, F)\, Y_t \qquad (4.4.6)$$

[18] Con la excepción de I_t, para el cual $\Gamma_I(t, k) = \Gamma_I(k)$.

[19] En todo lo que queda de este capítulo se tenderá a usar el componente estacional con fines de ilustración, pero los procedimientos que se presentan son de aplicación a cualquier otra señal, como tendencia o serie ajustada de estacionalidad.

donde el sumatorio se extiende de $-\infty$ a $+\infty$, y $\alpha(L, F)$ viene dado por[20]

$$\alpha(L, F) = \frac{\sigma_c^2}{\sigma_a^2} \frac{\psi_S(L)\psi_S(F)}{\psi_Y(L)\psi_Y(F)} \tag{4.4.7}$$

siendo $\psi_i(L) = \theta_i(L)/\phi_i(L)$ y $F = L^{-1}$ el operador de adelantos tal que $F^j Y_t = Y_{t+j}$.

Obsérvese que (4.4.7) define precisamente el filtro que se aplica a la serie observada para aproximar el componente estacional, en una operación que se representa en (4.4.6).

El filtro definido en (4.4.7) es un filtro simétrico, con lo que en (4.4.6) el coeficiente de Y_{t+j} coincide con el de Y_{t-j}. Por otra parte, usando (4.4.4) el filtro (4.4.7) se puede reexpresar como

$$\alpha(L, F) = \frac{\sigma_c^2}{\sigma_a^2} \frac{\theta_S(L)\theta_S(F)\phi_T(L)\phi_T(F)\phi_I(L)\phi_I(F)}{\theta_Y(L)\theta_Y(F)}. \tag{4.4.8}$$

Dado que la serie observada sigue un proceso invertible, todas las raíces de los polinomios del denominador están fuera del círculo unidad, y en consecuencia $\alpha(L, F)$ es un filtro convergente. Precisamente por eso se puede aplicar a una serie finita, ya que a partir de un cierto valor k el desconocer los valores Y_{t+j}, $j>k$, no influye en el resultado.

De forma similar se obtendrían los filtros más adecuados para extraer la tendencia y el componente irregular.

El instrumental matemático se ha hecho algo complejo en los últimos desarrollos, pero esto no debe desviar la atención de la idea básica: el resultado final (4.4.8) es una media móvil simétrica sobre la serie original, similar a las que se manejan en el método X-11. Lo que caracteriza a esta forma de proceder es que el filtro (4.4.8) está especialmente diseñado para la serie cuyo proceso generador de los datos viene dado en (4.4.2); y en su construcción se han combinado esta caracterización del proceso generador de datos con un conjunto de propiedades teóricas que definen a los componentes.

Nótese que el verdadero componente S_t y su aproximación S_t^* —y lo mismo se puede decir para cualquier otra señal— no tienen las mismas propiedades estocásticas, o lo que es lo mismo, no están generados por el mismo modelo. Al presentar el método se supuso que el modelo ARIMA de S_t venía dado por (véase 4.4.3):

$$\phi_S(L) S_t = \theta_S(L) c_t \qquad c_t \sim \text{Niid}(0, \sigma_c^2) \tag{4.4.9}$$

[20] Véase, por ejemplo, Maravall (1987).

es decir, $S_t = [\theta_S(L)/\phi_S(L)] c_t$, mientras el modelo del estimador es

$$S_t^* = \alpha(L, F) Y_t = \alpha(L, F)\psi_Y(L) a_t = \frac{\sigma_c^2}{\sigma_a^2} \frac{\theta_S(L)}{\phi_S(L)} \frac{\theta_S(F)\phi_Y(F)}{\phi_S(F)\theta_Y(F)} a_t =$$

$$= \frac{\sigma_c^2}{\sigma_a^2} \frac{\theta_S(L)}{\phi_S(L)} \frac{\theta_S(F)\phi_T(F)\phi_I(F)}{\theta_Y(F)} a_t = \frac{\theta_S(L)}{\phi_S(L)} \theta_S^*(F) a_t$$

con $\theta_S^*(F) = (\sigma_c^2/\sigma_a^2) \cdot (\theta_S(F)\phi_T(F)\phi_I(F)/\theta_Y(F))$; de donde

$$\phi_S(L) S_t^* = \theta_S(L)\theta_S^*(F) a_t \qquad a_t \sim \text{Niid}(0, \sigma_a^2) \tag{4.4.10}$$

es el proceso generador de datos de la aproximación.

Las diferencias entre el modelo de una señal, (4.4.9) para el componente estacional, y el modelo de su correspondiente aproximación, (4.4.10) en este caso, se traducen en:

1. Las aproximaciones vienen generadas por modelos ARIMA con factores que incluyen el operador de adelantos (F), componente $\theta_S^*(F)$ en (4.4.10).
2. Tales polinomios en F representan procesos de medias móviles no invertibles. En el caso concreto del componente estacional $\theta_S^*(F)$ contiene $\phi_T(F)$ que tiene dos raíces unitarias, por lo que $\theta_S^*(F)$ no es invertible.
3. Las transformaciones estacionarias de las aproximaciones de los distintos componentes están correlacionadas entre sí, aunque los verdaderos componentes no lo están.

Estas diferencias entre el proceso generador del verdadero componente y el de su aproximación explican porqué, en una aplicación concreta, es incorrecto juzgar un proceso de descomposición comparando las propiedades muestrales de las aproximaciones obtenidas con las que teóricamente han de presentar las verdaderas señales. Se ha comprobado que en general diferirán significativamente entre sí, sin que ello implique que el procedimiento de extracción de señales utilizado es incorrecto.

Obsérvese que las diferencias recogidas en 1), 2) y 3) son válidas también para los procedimientos empiricistas. Así, por ejemplo, en la mayor parte de los casos el elemento residual obtenido con la aplicación del método X-11 no tendrá estructura de ruido blanco, en el sentido de que mostrará dependencia temporal, sin que ello signifique que tal método es incorrecto en una aplicación concreta.

En ese contexto, la comparación del correlograma muestral del componente irregular resultante se ha de comparar con la función de autocorrelación (teórica) que tiene el estimador. En Espasa (1977,