

**UNIVERSIDAD DE LA LAGUNA**

**Modelos estructurales y estacionalidad en  
series temporales económicas de alta frecuencia**

**Autor: Martín Rodríguez, Gloria**

**Directores: Carlos Murillo Fort  
y Ginés Guirao Pérez**

**Departamento de Economía de las Instituciones,  
Estadística Económica y Econometría**

# Índice

<b>Introducción</b>	<b>1</b>
<b>1 Modelos estructurales de series temporales</b>	<b>7</b>
1.1 Formulación de los modelos estructurales de series temporales	9
1.1.1 Modelos de series temporales univariantes	10
1.1.2 Modelos de series temporales multivariantes y cointegración	22
1.2 Representación en el espacio de los estados y filtro de Kalman	27
1.2.1 Representación en el espacio de los estados	30
1.2.2 Filtro de Kalman y algoritmos de suavizado	34
1.2.3 Inicialización y estimación máximo verosímil	48
1.2.4 Observaciones anómalas y filtro de Kalman	52
1.3 Modelos estructurales y filtro de Kalman	54
1.4 Modelos estructurales versus modelos ARIMA	64
<b>2 Estacionalidad y funciones <i>splines</i></b>	<b>69</b>
2.1 Funciones <i>splines</i>	78
2.1.1 Procedimientos que no imponen el paso por los nodos	82
2.1.2 Procedimientos que imponen el paso por los nodos	96
2.2 Funciones <i>splines</i> cúbicas periódicas y estacionalidad	101
2.2.1 Procedimientos que no imponen el paso por los nodos	102
2.2.2 Procedimientos que imponen el paso por los nodos	105
2.3 Aplicaciones	109
<b>3 La exportación semanal de tomate canario</b>	<b>113</b>
3.1 Descripción de la serie semanal de exportaciones	115
3.2 Modelo estructural con variables cualitativas estacionales	117
3.3 Modelo estructural con funciones <i>splines</i>	125
3.4 Conclusiones	137

<b>4</b>	<b>Número diario de visitas a un servicio de urgencias del Barcelonès Nord</b>	<b>139</b>
4.1	Estacionalidad de las visitas diarias registradas	143
4.2	Especificación de la función <i>spline</i> para la fluctuación anual	147
4.3	Estimación del modelo estructural	155
4.4	Conclusiones	166
<b>5</b>	<b>Demanda horaria de energía eléctrica en España</b>	<b>169</b>
5.1	Descripción de la demanda horaria de energía eléctrica	173
5.2	Especificación y estimación del modelo estructural	179
5.2.1	Extracción de la fluctuación de periodo anual	180
5.2.2	Extracción de la fluctuación de periodo semanal	191
5.2.3	Extracción de la fluctuación de periodo diario	195
5.2.4	Estimación del modelo estructural	198
5.3	Conclusiones	212
	<b>Conclusiones y extensiones</b>	<b>215</b>
	<b>Anexos</b>	<b>223</b>
1.1	Ejemplos de la especificación del modelo de tendencias comunes	225
1.2	Propiedades estadísticas de un modelo en el espacio de los estados	227
1.3	Distribución del estado condicionada a las observaciones	233
2.1	Ilustración de los procedimientos de estimación simultánea del número y localización de los nodos	237
3.1	Evolución por campaña de las exportaciones semanales de tomate canario a Europa	245
3.2	Resultados para la transformación logarítmica de la serie de exportaciones semanales	247
4.1	Evolución semanal del número diario de visitas registradas en el servicio de urgencias del Barcelonès Nord	253
4.2	Evolución anual del número diario de visitas registradas en el servicio de urgencias del Barcelonès Nord	257
4.3	Resultados del modelo estructural para la transformación logarítmica de la serie de número diario de visitas	259

5.1	Evolución por hora de la semana de las medias móviles de periodo 24 sobre la serie de demanda horaria de energía eléctrica filtrada de componente tendencial determinístico y de fluctuación estacional de periodo anual	263
5.2	Evolución por hora del día de la serie de demanda horaria de energía eléctrica filtrada de tendencia lineal determinística y de las fluctuaciones estacionales de periodo anual y semanal	267
5.3	Evolución por hora del día de la fluctuación estacional diaria en un modelo con nivel local y fluctuaciones estacionales diaria estocástica y anual y semanal determinísticas ( <i>splines</i> )	269
5.4	Evolución por hora de la semana de la fluctuación estacional semanal estimada en un modelo con nivel local y fluctuación semanal estocástica ( <i>dummies</i> ) para la serie de demanda horaria de energía eléctrica filtrada de las estimaciones de la fluctuación estacional estocástica de periodo diario y del resto de componentes de naturaleza determinística	271
5.5	Resultados del modelo estructural para la transformación logarítmica de la serie de demanda horaria de energía eléctrica	275

<b>Bibliografía</b>	279
---------------------	-----



# Introducción

El análisis coyuntural de fenómenos económicos a partir de su observación directa puede conducir a conclusiones equivocadas, ya que los datos que se toman como expresión de la magnitud analizada suelen estar afectados por múltiples factores que introducen distorsiones e impiden percibir con claridad los patrones evolutivos que se pretende examinar. De ahí, la conveniencia de transformar los datos de manera que pueda aislarse el componente ruidoso y extraer señales informativas del comportamiento de la serie en el corto y largo plazo. En otras palabras, es necesario descomponer la magnitud observada en términos de variaciones no directamente observables tales como el nivel subyacente o tendencia, el ciclo o las fluctuaciones estacionales.

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. Sin embargo, hasta la mitad del siglo XX no se dispuso de instrumentos de cálculo potentes y de esquemas teóricos generales para el manejo de procesos estocásticos, lo que limitaba considerablemente el abanico de soluciones al problema de la extracción de señales. En el ámbito del análisis clásico de series temporales se asignaba a cada uno de los componentes básicos de una serie temporal un comportamiento rígido que correspondía a la evolución dictada por una función determinística del tiempo y cuya agregación daba como resultado el valor de la serie, al margen de un componente ruidoso que recogía la evolución no sistemática, imposible de captar con la formulación utilizada para el resto de componentes. Posteriormente, en el análisis moderno de series temporales, cuyo paradigma son los conocidos modelos ARIMA, se produce un sustancial cambio de enfoque. Aunque se admite que tanto la tendencia como las fluctuaciones periódicas pueden seguir pautas más flexibles y, de hecho, poseer carácter estocástico, la dependencia teórica del supuesto de estacionariedad conduce a evitar formulaciones explícitas de estas inestabilidades que, por el contrario, tratan de ser eliminadas mediante *adecuadas* transformaciones de la serie que convierten el componente ruidoso

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

en el núcleo del análisis. Por supuesto, este planteamiento oculta precisamente los aspectos más sustanciales de la evolución de dicha serie.

El enfoque de los modelos estructurales de series temporales aporta, en cambio, una especificación que permite estimar la evolución de cada uno de los componentes no observables, para los cuales se introducen formulaciones estocásticas. De este modo, no sólo se perciben las señales de interés desde el punto de vista del análisis económico de una serie temporal, sino que también se describen los cambios que, como reacción al entorno, hayan podido experimentar tales señales.

El planteamiento básico de los modelos estructurales consiste pues en asumir que la evolución de una serie temporal es resultado de la agregación de componentes estocásticos; pero, obviamente, existen infinitas formas de efectuar esta descomposición, sobre todo si, como de hecho se argumenta, los componentes no se ajustan a ningún patrón rígido. En este sentido, el filtro de Kalman es la herramienta estadística que, en virtud de procedimientos recursivos a partir de valores iniciales y de acuerdo con especificaciones explícitas de cada uno de los componentes, permite que este planteamiento no se quede sólo en una idea atractiva, sino que, gracias también a los avances computacionales, puedan obtenerse aproximaciones sensatas de los componentes no observables de buena parte de las series temporales económicas.

Estos progresos computacionales en la segunda mitad del siglo pasado significaron una notable reducción de los tiempos de cálculo que, a su vez, alentó los desarrollos teóricos. Y estos dos *ingredientes* hicieron posible aplicar el enfoque estructural al estudio de series más largas y con mayor frecuencia de observación, que comenzaban a ser publicadas y cuyo uso más extendido en la literatura estadístico-econométrica es, sin duda, una medida del progreso en la Economía aplicada. Es evidente que una agregación temporal excesiva puede impedir que se aprecien determinados comportamientos o respuestas dinámicas. En este sentido, y salvo que el objetivo del análisis no vaya más allá de la descripción de tendencias generales en el largo plazo, la elección de la frecuencia con la que se observan los datos es un elemento determinante del éxito del estudio y, por tanto, quizás es mejor *pecar por exceso que por defecto*. Ahora bien, sin entrar en el ámbito de las denominadas series de muy alta frecuencia, como por ejemplo las series temporales financieras *tick-by-tick* —que quedan fuera de la perspectiva analítica de esta investigación—, el estudio de series semanales, diarias u horarias introduce elementos de complejidad cuyo tratamiento puede abordarse también mediante el enfoque de los modelos estructurales.

En primer lugar, la elevada frecuencia de los datos suele venir acompañada de problemas de heterogeneidad causados, por ejemplo, porque el periodo estacional anual no es siempre de la misma longitud, como ocurre cuando se trabaja con datos semanales o también con frecuencias diarias u horarias y combinación de años bisiestos y no bisiestos. Por otra parte, en series largas resulta más difícil asumir que el patrón estacional es estable. Además, en este tipo de series suelen coexistir fluctuaciones de distinto periodo cuya estimación simultánea no es siempre factible con la formulación determinística o estocástica que cada una de estas fluctuaciones requiera. En este sentido, un modelo estructural admite la combinación de formulaciones convencionales con otras especificaciones alternativas más parsimoniosas, funciones *splines*, que suponen, en general, un ahorro computacional significativo y pueden, de hecho, ser la única vía posible de estimación.

Por ejemplo, en una serie horaria la variación de periodo anual, aunque se asuma la estabilidad de ésta a lo largo de los años, difícilmente podrá ser estimada a través de un *sencillo* modelo de variables cualitativas. Ahora bien, si el patrón estacional de periodo anual se ajusta a una curva relativamente suave, dicho patrón puede ser descrito adecuadamente dividiendo el periodo estacional en segmentos apropiados y especificando en cada uno un determinado polinomio expresado como función de la estación a la que pertenezca la observación. La imposición de condiciones de continuidad que vinculan los polinomios anteriores permite aproximar con precisión este tipo de comportamientos, que, sin embargo, no tienen por qué responder a la evolución generada por una única función polinomial de grado elevado definida para todo el periodo estacional, incluso aunque el grado sea suficiente para obtener un buen ajuste.

Pues bien, el objetivo fundamental de este trabajo es diseñar procedimientos adaptados a las particularidades de series económicas de alta frecuencia que muestren la utilidad de estas funciones *splines* para recoger comportamientos estacionales en el seno metodológico de los modelos estructurales. Para alcanzar este objetivo, el texto que sigue se estructura como se indica a continuación.

En el capítulo primero, se pone el énfasis en describir la racionalidad subyacente en los modelos estructurales, indicar la forma en la que podrían ser empleados y señalar aquellos aspectos en que su metodología difiere de enfoques alternativos. En primer lugar se introduce la formulación básica de los modelos estructurales de series temporales. A continuación, se expone la representación más general en el espacio de

## **Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia**

los estados y se explica cómo se puede usar el filtro de Kalman para estimar el valor de las denominadas variables de estado. Se muestra, entonces, que los modelos estructurales constituyen un caso particular de los modelos en el espacio de los estados en el que los componentes típicos de la serie temporal son las variables de estado inobservables y la propia serie es la magnitud observable relacionada con el vector de variables de estado. Por último, se indica el modo de obtener una estimación de dichos componentes en cada uno de los instantes del tiempo a través del filtro de Kalman.

Dado que la esencia de estos modelos consiste en recoger la inestabilidad de los comportamientos a lo largo del tiempo y que tal variabilidad puede apreciarse con más claridad en series largas y ser más acentuada conforme aumenta la frecuencia de observación, este tipo de series constituye un campo sugerente de aplicación de la metodología propia del enfoque estructural. Aunque muchas series temporales pueden ser descritas mediante modelos que recogen las variaciones estacionales a través de variables cualitativas o términos trigonométricos, cuando las observaciones se registran con alta frecuencia —por ejemplo, semanal, diaria o por hora—, estas formulaciones del componente estacional no se caracterizan por su parsimonia.

En un escenario marcado por la transformación operada en las publicaciones estadísticas —en las que las series de alta frecuencia han dejado de ser la excepción— y por los progresos en el campo de la computación —que han permitido el tratamiento de gran cantidad de información con mayor eficiencia en términos de tiempo de cálculo—, la búsqueda de nuevas técnicas que amplíen el conjunto de herramientas estadísticas a disposición del analista para la descripción de series de este tipo es una tarea atractiva. En este sentido, en el capítulo segundo se exponen con detalle las características de las funciones *splines* en el ámbito del análisis de series temporales y, más concretamente, se describen los distintos procedimientos a través de los cuales estas funciones pueden, en el contexto de los modelos estructurales, utilizarse como herramienta adecuada para recoger fluctuaciones periódicas de naturaleza determinística y también estocástica. Aunque las funciones *splines* no constituyen una herramienta nueva, muchos de los desarrollos teóricos a partir de los cuales estas funciones aproximan el comportamiento de otras en determinados espacios de tiempo, necesitan ser adaptados al caso particular en el que la magnitud aproximada es alguna de las fluctuaciones periódicas presentes en una serie temporal, tales como el ciclo o las variaciones estacionales. En este trabajo se presenta, en concreto, la adaptación de aquellos procedimientos que implican la

optimización de alguna función relativa al error cometido en el ajuste sin necesidad de imponer que la función *spline* pase necesariamente por determinados puntos.

Una vez expuesto el marco metodológico, en la segunda parte de este trabajo, capítulos tercero a quinto, se valora el funcionamiento de estas técnicas cuando se aplican a series temporales económicas que contienen fluctuaciones periódicas de alta frecuencia. En concreto, el uso de funciones *splines* en cada uno de los casos de estudio ha requerido la introducción de modificaciones metodológicas cuyo valor añadido es que aportan soluciones a los distintos problemas de heterogeneidad en la variación estacional que caracterizan a las tres series analizadas.

Aunque, desde un punto de vista sectorial, una de las series pertenece al ámbito agrario —exportaciones semanales de tomate canario—, otra al de servicios —visitas diarias a un centro de urgencias de la ciudad de Barcelona— y la tercera podría ubicarse en el industrial —demanda horaria de energía eléctrica en España—, éste no ha sido el criterio para su elección, que está justificada más bien por la naturaleza específica de los datos disponibles en cada caso. Precisamente por esta razón, cada uno de estos tres capítulos se inicia con una breve descripción de las características fundamentales del contexto socioeconómico al que hace referencia cada serie, haciendo especial hincapié en los rasgos esenciales de su patrón estacional.

En el caso de la serie semanal, el componente estacional puede ser estimado a través de formulaciones convencionales, de modo que la estimación mediante funciones *splines* permite calibrar el buen funcionamiento de éstas a la hora de aproximar esta fluctuación. En cambio, en las otras dos series, la coexistencia de dos o más fluctuaciones periódicas y la excesiva longitud del periodo asociado a alguna de ellas impide la estimación directa de las variaciones estacionales y, en consecuencia, las funciones *splines* se convierten en una alternativa necesaria.

Dada la diversidad de los ámbitos de aplicación escogidos, al final de cada uno de estos tres capítulos se ha incluido un epígrafe que contiene las principales conclusiones que, desde una perspectiva económica, se desprenden del análisis realizado. Mientras que el apartado con el que termina este trabajo se reserva para enumerar las conclusiones metodológicas de la investigación emprendida y las líneas abiertas que ésta ha sugerido.

En las páginas finales se incluyen los anexos correspondientes a cada uno de los capítulos. En el caso de los dos primeros capítulos, estos anexos tratan de aportar nociones sobre determinados aspectos que resultaba difícil incorporar en el texto sin

## **Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia**

romper el hilo conductor del mismo. Sin embargo, son útiles para aclarar dudas y facilitar la exposición de las ideas sin detenerse en explicaciones más o menos complejas de ciertos detalles. Los anexos restantes, en los que se ha tratado de proporcionar una presentación simplificada, se ocupan de mostrar algunos de los resultados en los que se apoyan los comentarios realizados en los capítulos tercero al quinto.

# 1. Modelos estructurales de series temporales

El análisis de series temporales se asienta en la teoría de procesos estocásticos, originalmente construida sobre la hipótesis de estacionariedad. De ahí que, en el enfoque tradicional, es decir, en la conocida metodología propuesta por Box y Jenkins (1976), se opte por eliminar la no estacionariedad mediante la aplicación del operador diferencia. Entonces, el comportamiento en el tiempo del proceso estocástico generador de una serie temporal observada se recoge a través de los denominados modelos *ARIMA*.

Sin embargo, pocas series temporales económicas y sociales son estacionarias. De hecho, la mayor disponibilidad de datos en los últimos años y, sobre todo, la posibilidad de construir series más largas y con mayor frecuencia de observación, han puesto de manifiesto las dificultades para mantener el supuesto de un patrón de comportamiento fijo a lo largo del tiempo. Así, parece más acertado asumir que las propiedades estadísticas de buena parte de las series socioeconómicas presentan un carácter evolutivo.

En presencia de comportamientos no estacionarios, la aplicación del operador diferencia y el análisis posterior de la transformación estacionaria, no es siempre la opción adecuada. En este sentido, los modelos estructurales de series temporales constituyen —a pesar de la equivalencia existente, en términos predictivos, entre las formulaciones más simples y las representaciones *ARIMA*— una alternativa apropiada a los modelos convencionales; puesto que, lejos de eliminar la fuente de no estacionariedad, la incorporan en el modelo admitiendo la posibilidad de que cada uno de los componentes inobservables típicos de una serie temporal posea una naturaleza estocástica.

Por ejemplo, si una serie temporal se puede descomponer como suma de tendencia, estacionalidad y un componente irregular, su comportamiento se podría recoger mediante la formulación de un modelo de regresión cuyas variables explicativas son una tendencia determinística y un conjunto de variables cualitativas estacionales. Si estos componentes no son estables, dicha formulación sería inadecuada y sería preciso que los coeficientes de regresión cambiaran en el tiempo. Esta flexibilidad es la que pretenden recoger los modelos

## **Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia**

estructurales que, en tal sentido, « ... no son más que modelos de regresión en los que las variables explicativas son funciones del tiempo y los parámetros varían en el tiempo» (Harvey, 1989:10).

En general, los componentes que describen la evolución de una serie temporal — tendencia, estacionalidad, ciclo y componente irregular— se han modelado tradicionalmente de forma determinista; sin embargo, cuando las series son suficientemente largas, se suele apreciar la inestabilidad de tales componentes e, incluso, puede ser razonable considerar que evolucionan aleatoriamente en el tiempo<sup>1</sup>. Este es el supuesto de partida de los modelos estructurales. Este capítulo pone el énfasis en describir la racionalidad subyacente en estos modelos, indicar la forma en la que podrían ser empleados y señalar aquellos aspectos en que su metodología difiere de enfoques alternativos<sup>2</sup>.

Un modelo estructural de series temporales es un modelo estadístico definido con objeto de representar las características más sobresalientes de una serie temporal mediante su descomposición en componentes que, aún siendo no observables, tienen interés en sí mismos y una interpretación directa. Concretamente, esta formulación permite describir el comportamiento de dicha serie, es decir, es posible indicar qué componentes y en qué grado son responsables de sus principales movimientos o, simplemente, es factible, por ejemplo, extraer el componente estacional y obtener una serie ajustada estacionalmente sobre la que se observan con mayor facilidad los principales movimientos causados por los restantes componentes. También, a partir de un modelo estructural se pueden evaluar los efectos que determinada medida de política económica tiene sobre el comportamiento a largo, medio y corto plazo de la serie. Finalmente, si el modelo es adecuado para describir las características más notables de una serie temporal, puede constituir el elemento sobre el que apoyar la predicción de observaciones futuras.

La formulación de los modelos estructurales se puede encuadrar en el marco más general de modelos en el espacio de los estados, lo que permite abordar su estimación a

---

<sup>1</sup> «Una tendencia global puede ser representada por una única función determinística del tiempo para toda la muestra. Una tendencia local puede cambiar de dirección a lo largo de la muestra y constituye la dirección más reciente que se desea extrapolar hacia el futuro» (Harvey, 1989:24). Particularmente en series económicas, una tendencia determinística constituye la excepción más que la regla y, por tanto, no hay justificación para imponerla a un conjunto de datos (Harvey, 1989:290). En Harvey y otros (1986) y Slade (1989) se introduce una tendencia estocástica para aproximar el progreso tecnológico.

través del filtro de Kalman. Estos modelos en el espacio de los estados resultan útiles cuando se desea estimar el valor de una magnitud —estado— no directamente observable que evoluciona en el tiempo y que, sin embargo, depende de su valor en el instante anterior, siendo posible, además, establecer una relación contemporánea entre una magnitud observable y el estado no observable.

En el epígrafe siguiente, se introduce la formulación básica de los modelos estructurales de series temporales. A continuación, se expone la representación más general en el espacio de los estados y se explica cómo se puede usar el filtro de Kalman para estimar el valor de las denominadas variables de estado. Se muestra, entonces, que los modelos estructurales constituyen un caso particular de los modelos en el espacio de los estados en el que los componentes típicos de la serie temporal son las variables de estado inobservables y la propia serie es la magnitud observable relacionada con el vector de variables de estado<sup>3</sup>. Por último, se indica el modo de obtener una estimación de dichos componentes en cada uno de los instantes del tiempo a través del filtro de Kalman. También se incluyen anexos en los que se detallan aspectos formales tratados en el texto principal.

### 1.1. Formulación de los modelos estructurales de series temporales

En este apartado, se exponen distintos modelos que se diferencian no sólo en los componentes considerados, sino también en la naturaleza determinística o estocástica asignada a cada uno de ellos. Se ha comenzado por las especificaciones más sencillas en el caso univariante; en concreto, aquéllas que sólo introducen componente de nivel, pero no componente cíclico ni estacional. Dada la orientación metodológica específica de este trabajo se concede especial atención a las diversas formulaciones del componente estacional. El epígrafe concluye con la especificación de modelos estructurales

---

<sup>2</sup> El desarrollo de modelos estructurales de series temporales se inició a finales de los 70, destacando las contribuciones tempranas de Pagan (1975), Engle (1978) y Kitagawa y Gersch (1984). El cuerpo metodológico de este enfoque aparece recogido en Harvey (1989).

<sup>3</sup> Como afirma Aoki (1990:3), una serie temporal se puede entender como un sistema que transforma información contenida en señales exógenas pasadas y presentes en observaciones futuras. Los estados son resúmenes de la información contenida en la señal exógena que es transmitida por la dinámica generadora de la serie temporal.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

multivariantes, en los que se introduce la noción de cointegración, que aparece como resultado de la presencia de tendencias comunes o compartidas por dos o más series<sup>4</sup>.

### 1.1.1. Modelos de series temporales univariantes

Un modelo univariante de series temporales se puede definir como

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t,$$

donde  $\mu_t$  representa la tendencia,  $\psi_t$  el ciclo,  $\tilde{a}_t$  el componente estacional y  $\hat{a}_t$  el componente irregular, que recoge los movimientos no sistemáticos<sup>5</sup>. La peculiaridad de los modelos estructurales radica en su flexibilidad para recoger los cambios en el comportamiento de la serie mediante la consideración de sus diferentes componentes como procesos estocásticos dirigidos por perturbaciones aleatorias<sup>6</sup>. Como se podrá comprobar en las distintas especificaciones de este apartado, un modelo estructural no necesariamente debe ser definido en términos de todos los componentes inobservables<sup>7</sup>. A continuación se exponen algunas de las formulaciones posibles para cada uno de los componentes o combinaciones de ellos.

La formulación más simple para el componente irregular es un proceso ruido blanco, es decir, una secuencia de variables aleatorias no correlacionadas serialmente, con media constante e igual a cero y varianza constante e igual a  $\sigma_\varepsilon^2$ . También se pueden plantear modelos más generales en los que el proceso estocástico  $\varepsilon_t$  sea un proceso autorregresivo estacionario de orden  $p$ .

La tendencia —es decir, el componente que indica la dirección en que se mueve la serie en el largo plazo<sup>8</sup>— no se especifica como una función determinista del tiempo, sino

---

<sup>4</sup> Los modelos multivariante no serán utilizados explícitamente en esta investigación, pero ha parecido oportuno incluirlos para dar coherencia a la descripción de los modelos estructurales y, sobre todo, para que pueda entenderse una metodología que sí será necesaria para abordar algunas de las líneas abiertas que se proponen.

<sup>5</sup> Aunque un modelo multiplicativo podría ser más adecuado en algunos casos, éste podría ser tratado bajo el enfoque aditivo aplicando la transformación logarítmica.

<sup>6</sup> De este modo, los comportamientos no sistemáticos de la serie no tienen que recaer exclusivamente en el componente irregular, sino que pueden ser recogidos a través de los términos de perturbación de los otros componentes inobservables de naturaleza estocástica.

<sup>7</sup> Concretamente, los componentes considerados dependerán, entre otras cosas, de la periodicidad y la longitud de la muestra.

<sup>8</sup> Parafraseando a Harvey, la tendencia «... es aquella parte de la serie que, cuando se extrapola, proporciona la indicación más clara del movimiento futuro en el largo plazo de la serie. La definición ... es consistente con la idea de indicar una dirección general» (Harvey, 2000:1).

que, por el contrario, se asume que el componente tendencial es un proceso estocástico cuyo valor en un instante dado del tiempo (nivel estocástico) es el resultado de añadir al valor del componente en el instante anterior un determinado incremento de carácter aleatorio (pendiente estocástica) y un término de perturbación aleatoria. Por supuesto, cabe la posibilidad de que el nivel o la pendiente no posean un carácter estocástico y, más aún, la pendiente puede estar o no presente.

El ejemplo más sencillo lo constituye una serie cuyas observaciones oscilan alrededor de un nivel medio que permanece constante en el tiempo. Si dichas oscilaciones son estacionarias —en el sentido de que, aunque algunos valores se alejen en el corto plazo del nivel medio, la serie siempre tiende a dicho nivel— y se supone, además, que no están correlacionadas entre sí, una formulación capaz de recoger dicho comportamiento dinámico es  $y_t = \mu_0 + \varepsilon_t$ , para  $t = 1, \dots, T$ , donde  $\hat{a}_t$  es un proceso ruido blanco gaussiano con varianza  $\sigma_\varepsilon^2$ . Se trata de un modelo con componente de nivel,  $\mu_0$ , determinista.

Esta formulación se puede flexibilizar para permitir que el nivel de la serie cambie en el tiempo, dando lugar al *modelo de nivel local*, en el que el nivel en cada momento del tiempo es el nivel del periodo anterior más un elemento aleatorio. En este modelo, definido como

$$y_t = \mu_t + \varepsilon_t,$$

con

$$\mu_t = \mu_{t-1} + \eta_t,$$

donde  $\eta_t \sim NID(0, \sigma_\eta^2)$ , existe un término de perturbación aleatoria alrededor de un nivel subyacente que oscila sin una dirección particular<sup>9</sup>. De ahí que también se conozca por el nombre de paseo aleatorio con ruido. Si la varianza del nivel es cero, se obtiene el modelo de nivel determinista antes especificado. Si la varianza del componente irregular es cero, pero la varianza del nivel es distinta de cero, la serie sólo posee componente de nivel cuyo proceso generador es un camino aleatorio<sup>10</sup>.

---

<sup>9</sup> La expresión  $NID(0, \sigma^2)$  denota que las variables aleatorias son serialmente independientes y normalmente distribuidas con media nula y varianza  $\sigma^2$ .

<sup>10</sup> La inclusión de un componente de camino aleatorio significa que el proceso es no estacionario y, además, que es equivalente a un proceso ARIMA(0,1,1), donde el parámetro MA sólo cubre la mitad del espacio

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

Si a los elementos descritos hasta el momento para el componente de nivel se añade una pendiente, es posible considerar que el incremento medio por unidad de tiempo sea constante, pero también se puede recurrir a formulaciones más flexibles. Una de ellas es el *modelo de tendencia lineal local*<sup>11</sup>, definido como

$$y_t = \mu_t + \varepsilon_t,$$

con

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$

$$\beta_t = \beta_{t-1} + \zeta_t,$$

donde  $\zeta_t \sim NID(0, \mathbf{s}_z^2)$ <sup>12</sup>. El término de perturbación  $\zeta_t$  otorga a la pendiente un carácter estocástico, mientras que, dados el nivel y la pendiente del periodo anterior, la naturaleza estocástica del nivel en el instante actual se deriva de la presencia del término de perturbación  $\eta_t$ <sup>13</sup>. Así pues, si la varianza de  $\zeta_t$  es nula, de forma que la pendiente es constante, el modelo resultante —conocido como *modelo de nivel local con deriva*— considera un nivel estocástico y una pendiente fija, es decir,

$$\mu_t = \mu_{t-1} + \beta + \eta_t$$
<sup>14</sup>.

Si la pendiente no es sólo constante sino nula, el modelo anterior se transforma en el que se ha denominado modelo de nivel local. Finalmente, es posible mantener el carácter estocástico de la pendiente y, al mismo tiempo, suponer que, dados los valores del nivel y la pendiente en el instante  $t-1$ , el valor del nivel en el instante  $t$  es no estocástico, es decir,

$$\mu_t = \mu_{t-1} + \beta_{t-1},$$

$$\beta_t = \beta_{t-1} + \zeta_t$$
<sup>15</sup>.

---

usual del parámetro, es decir,  $-1 \leq \theta \leq 0$ ; además, si  $\sigma_\eta^2 = 0$ , se obtiene un proceso ARIMA en el que el parámetro MA toma el valor  $-1$ .

<sup>11</sup> Aproximación local de una tendencia lineal.

<sup>12</sup> En este tipo de modelos, la tendencia lineal local puede ser generalizada a polinomios de órdenes más altos y, aunque dichas tendencias no son lineales en el sentido de que dan lugar a funciones de predicción, por ejemplo, cuadráticas o cúbicas, pueden ser formuladas como modelos lineales de series temporales (véase Harvey, 1989:295, para el caso de un modelo de tendencia cuadrática). El modelo de tendencia lineal local es equivalente a un proceso ARIMA (0,2,2); además, será un proceso no invertible cuando  $\sigma_\zeta^2 = 0$  y, por el contrario, si  $\sigma_\zeta^2 > 0$  el proceso es invertible independientemente de cual sea el valor de  $\sigma_\eta^2$ .

<sup>13</sup> Además, se supone que estos dos términos de perturbación aleatoria son mutua y serialmente independientes.

<sup>14</sup> Si se supone, además, que  $\sigma_\eta^2 = 0$ , se obtiene una tendencia lineal determinística.

Por otra parte, el modelo de tendencia lineal local se puede modificar introduciendo un factor de amortiguación en el componente de pendiente de la tendencia, es decir,

$$\beta_t = \rho\beta_{t-1} + \zeta_t,$$

donde  $0 \leq \rho \leq 1$ . Se obtiene entonces el *modelo de tendencia lineal local amortiguada*<sup>16</sup>.

Por supuesto, a la hora de elegir una de estas especificaciones, es importante, *a priori*, examinar el comportamiento de la serie a lo largo de toda la muestra y, *a posteriori*, comprobar cuál produce un mejor ajuste; aunque tampoco se pueden ignorar las implicaciones de cada especificación y los objetivos que se persiguen con el modelo.

En muchas series económicas es importante distinguir entre una tendencia de largo plazo y los movimientos asociados a un componente cíclico, que admite también tanto una especificación determinística como estocástica<sup>17</sup>. Un ciclo determinista describe un comportamiento perfectamente periódico con una frecuencia dada, que puede ser recogido mediante una función de senos y cosenos del tipo siguiente

$$\Psi_t = \alpha \cos(\lambda t) + \beta \sin(\lambda t),$$

donde  $\lambda$  es la frecuencia angular del ciclo, en radianes<sup>18</sup>. Si se admite que los parámetros  $\alpha$  y  $\beta$  evolucionan aleatoriamente en el tiempo, el ciclo adquiere naturaleza estocástica. Pero antes de incorporar el carácter estocástico, conviene garantizar la continuidad del componente cíclico expresándolo de la forma recursiva siguiente<sup>19</sup>:

---

<sup>15</sup> Asumir que  $\sigma_\eta^2 = 0$  puede significar predicciones más pobres que las del modelo de tendencia lineal local; sin embargo, se trata de una formulación que puede ser más apetecible para los gestores de la política económica (véase Harvey, 1989:286). El modelo de nivel fijo y pendiente estocástica se denomina también *modelo de tendencia alisada (smoothed)*, dado que el componente tendencial es un camino aleatorio integrado cuya estimación fluctúa suavemente, sin incrementos o decrementos abruptos. Estas tendencias alisadas constituyen especificaciones equivalentes a las denominadas funciones *splines*, diseñadas, precisamente, con objeto de aproximar una serie de datos a través de una función relativamente suave. En concreto, el modelo anterior es equivalente a una *spline* cúbica (Wecker y Ansley, 1983, y Durbin y Koopman, 2001:61-63). Estas funciones se exponen con detalle en el capítulo siguiente y, como se ha comentado, constituyen de hecho el objetivo metodológico básico de este trabajo.

<sup>16</sup> Estos modelos pueden ser utilizados para recoger tendencias globales no lineales como, por ejemplo, las presentes en aquellos casos en los que existe un nivel de saturación. Esta solución, aunque es viable en algunos momentos del tiempo, no capta ciertas características de comportamiento de manera tan adecuada como otras formulaciones alternativas, tales como curvas de crecimiento (Harvey, 1989:296-298).

<sup>17</sup> Componentes cíclicos de este tipo resultan útiles para modelar ciclos económicos (Harvey, 1985, y Harvey y Jaeger, 1993).

<sup>18</sup> Es decir, el parámetro  $\lambda$  representa el número de veces que el ciclo se repite en un periodo de tiempo de longitud  $2\pi$ . De este modo, el cociente  $\lambda/2\pi$  expresa el número de veces que se repite el ciclo por unidad de tiempo.

<sup>19</sup> Véase Harvey (1989:39).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$\begin{pmatrix} \Psi_t \\ \Psi_t^* \end{pmatrix} = \begin{pmatrix} \cos \lambda & \text{sen} \lambda \\ -\text{sen} \lambda & \cos \lambda \end{pmatrix} \begin{pmatrix} \Psi_{t-1} \\ \Psi_{t-1}^* \end{pmatrix}, \quad t = 1, \dots, T$$

con  $\Psi_0 = \alpha$  y  $\Psi_0^* = \beta$ <sup>20</sup>. Los nuevos parámetros son  $\Psi_{t-1}$ , el valor del ciclo en  $t-1$ , y  $\Psi_{t-1}^*$ , que aparece por construcción en la formación de  $\Psi_t$ . Entonces, el ciclo es estocástico al introducir dos términos de perturbación ruido blanco gaussianos  $k_t$  y  $k_t^*$  de forma que

$$\begin{pmatrix} \Psi_t \\ \Psi_t^* \end{pmatrix} = \begin{pmatrix} \cos \lambda & \text{sen} \lambda \\ -\text{sen} \lambda & \cos \lambda \end{pmatrix} \begin{pmatrix} \Psi_{t-1} \\ \Psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} k_t \\ k_t^* \end{pmatrix}$$

Por razones de identificabilidad y parsimonia, se supone que  $k_t$  y  $k_t^*$  están mutuamente incorreladas y poseen la misma varianza  $\sigma_k^2$ . Por otra parte, para recoger los comportamientos pseudocíclicos que caracterizan a muchas series temporales económicas, conviene incluir un factor de amortiguación (*damping*),  $\rho$ , que aporta mayor flexibilidad al ciclo estocástico, es decir,

$$\begin{pmatrix} \Psi_t \\ \Psi_t^* \end{pmatrix} = \rho \begin{pmatrix} \cos \lambda & \text{sen} \lambda \\ -\text{sen} \lambda & \cos \lambda \end{pmatrix} \begin{pmatrix} \Psi_{t-1} \\ \Psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} k_t \\ k_t^* \end{pmatrix}$$

donde  $0 \leq \rho \leq 1$ <sup>21</sup>. Si  $\lambda = 0$  o  $\lambda = \pi$ , el ciclo estocástico se convierte en un proceso autorregresivo de primer orden<sup>22</sup>. Si  $I = 0$  y  $r = 1$ , el componente cíclico se transforma en el modelo de nivel local especificado para la tendencia<sup>23</sup>.

Considerando el ciclo como una parte intrínseca de la tendencia más que como un componente separado de ésta, puede resultar adecuado acudir a especificaciones que combinen estos dos componentes. Las formulaciones de este tipo más habituales son: *modelo cíclico*,

$$y_t = \mu + \Psi_t + \varepsilon_t,$$

<sup>20</sup> Véase Harvey (1993:182, 227).

<sup>21</sup> Adviértase que la incorporación de un factor de amortiguación menor que la unidad hace que el ciclo presente una representación equivalente a la de un proceso VAR(1) y que, en términos predictivos, la fluctuación cíclica tienda a desaparecer. Harvey y Streibel (1996) proponen un contraste de la hipótesis nula  $\rho = 1$  frente a la alternativa de que este parámetro sea menor que la unidad.

<sup>22</sup> Aunque  $\rho$  es no negativo, el parámetro en la formulación autorregresiva de primer orden equivalente puede ser tanto positivo ( $\lambda = 0$ ) como negativo ( $\lambda = \pi$ ).

<sup>23</sup> Nótese que la varianza del ciclo,  $\sigma_k^2$ , está relacionada con la varianza del término de perturbación del ciclo,  $\sigma_\varepsilon^2$ , a través de la expresión  $\sigma_k^2 = (1 - \rho^2)\sigma_\varepsilon^2$ , cuando  $0 \leq \rho \leq 1$ .

## Modelos estructurales de series temporales

con  $0 \leq \rho \leq 1$ <sup>24</sup>; *modelo tendencia-ciclo*,

$$y_t = \mu_t + \psi_t + \varepsilon_t,$$

donde  $\mu_t$  recoge una tendencia lineal local<sup>25</sup>; *modelo de tendencia cíclica*<sup>26</sup>,

$$y_t = \mu_t + \varepsilon_t,$$

donde

$$\mu_t = \mu_{t-1} + \psi_{t-1} + \beta_{t-1} + \eta_t,$$

$$\beta_t = \beta_{t-1} + \zeta_t$$
<sup>27</sup>.

Otras especificaciones consisten en añadir un componente AR(2) a  $y_t = \mu_t + \varepsilon_t$  o  $\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$ <sup>28</sup>.

Si las observaciones de la serie en cuestión se registran con periodicidad inferior a la anual, suelen aparecer efectos estacionales y, por ello, conviene introducir este componente en el modelo. Si las fluctuaciones estacionales son estables, el componente estacional se puede considerar determinista o fijo. Si el número de estaciones es  $s$ , y dado que la suma de los efectos estacionales a lo largo del año, o, en general, del periodo en el que se completa la fluctuación estacional, debe ser nula, el componente estacional en el instante  $t$  se puede especificar como:

$$\gamma_t = \sum_{j=1}^{s-1} z_{j,t} \gamma_j$$

---

<sup>24</sup> Es equivalente a un proceso ARMA(2,2) con constante. El ciclo en sí mismo es un proceso ARMA(2,1). La parte MA está sometida a restricciones, pero las más importantes se establecen con respecto a los parámetros AR. Si  $0 < \lambda < \pi$ , las raíces del polinomio AR son un par de raíces complejas conjugadas con módulo  $\rho^{-1}$  y fase  $\lambda$ . Si, además,  $0 \leq \rho < 1$ , dichas raíces están fuera del círculo de radio unidad y el proceso es estacionario. Dado que las raíces de un polinomio AR(2) pueden ser tanto reales como complejas, la formulación de un modelo cíclico restringe la región admisible de valores de los coeficientes AR a aquella capaz de generar un comportamiento pseudocíclico (Harvey, 1993:193).

<sup>25</sup> Esta representación es equivalente a un proceso ARIMA(2,2,4) (Harvey, 1989:511).

<sup>26</sup> En la práctica resulta muy difícil distinguir el componente de tendencia del cíclico, sobre todo cuando ambos se suponen estocásticos. De ahí que, en lugar de separarlos, ambos se combinen en un componente, denominado *tendencia-ciclo*. Hay que tener en cuenta, además, que el tamaño de la muestra influye considerablemente en esta separación porque lo que parece una evolución tendencial en una muestra pequeña puede resultar un efecto cíclico en otra mayor.

<sup>27</sup> Si en este modelo se supone que  $\lambda = 0$  y se elimina el término  $\beta_{t-1}$ , se puede asumir que  $\psi_t$  recoge el componente tendencial y, de hecho, se obtiene la formulación del modelo de tendencia lineal local amortiguada. Si se supone, además, que  $\rho = 1$  y  $\sigma_k^2 = 0$ , se obtiene una tendencia lineal determinística.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

donde  $z_{j,t}$ ,  $j=1, \dots, s-1$ , es igual a 1 si  $t$  pertenece a la estación  $j$ ,  $-1$  si  $t$  pertenece a la estación  $s$  y 0 en otro caso.

Formalmente, se tiene que

$$\gamma_t = \begin{cases} \gamma_j, & \text{si } t \text{ pertenece a la estación } j, \\ -\sum_{j=1}^{s-1} \gamma_j, & \text{si } t \text{ pertenece a la estación } s \end{cases} \quad j = 1, \dots, s-1$$

En cambio, si el patrón estacional evoluciona a lo largo del tiempo, parece más adecuado incorporar un componente estacional estocástico, que admite varias formulaciones posibles.

Una forma de conseguir que el patrón estacional evolucione en el tiempo es permitir que la suma de los efectos estacionales a lo largo del año no sea estrictamente nula, sino igual a un término de perturbación aleatoria. Es decir, los valores del componente estacional en los instantes del tiempo  $t$ ,  $t-1$ , ...,  $t-s+1$  están ligados por la relación

$$\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t, \text{ o bien, } \gamma_t = \sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t, \text{ siendo } \omega_t \text{ un término de perturbación aleatoria tal}$$

que  $\omega_t \sim NID(0, \sigma_\omega^2)$ . Otra posibilidad es suponer que el efecto de cada estación evolucione de acuerdo a un camino aleatorio (Harrison y Stevens, 1976:217-218). Es decir, si  $t$  pertenece a la estación  $j$ ,  $\gamma_t = \gamma_{j,t}$ , donde

$$\mathbf{g}_{j,t} = \mathbf{g}_{j,t-1} + \mathbf{w}_{j,t}, \quad j = 1, \dots, s$$

y  $\omega_{j,t}$  sigue siendo un término de perturbación aleatoria tal que  $\omega_{j,t} \sim NID(0, \sigma_\omega^2)$ <sup>29</sup>.

Una alternativa a ésta es la formulación trigonométrica<sup>30</sup>, en la que

<sup>28</sup> Nótese que si la serie temporal presenta un comportamiento cíclico de frecuencia  $\lambda$ , dicho comportamiento puede recogerse mediante un polinomio autorregresivo de la forma  $(1-2\cos(\lambda)L+L^2)$ , siendo  $L$  el operador de retardos.

<sup>29</sup> Esta especificación se conoce también como modelo de *dummies* estacionales estocásticas. En esta formulación, la restricción  $\sum_{j=1}^s \mathbf{g}_{j,t} = 0$ , en cualquier instante del tiempo  $t$ , se incorpora estableciendo una

estructura de correlación apropiada entre los términos  $\omega_{j,t}$ ,  $j = 1, \dots, s$ , de modo que  $\sum_{j=1}^s \omega_{j,t} = 0$ .

<sup>30</sup> El modelo de un patrón estacional fijo a partir de un conjunto de términos trigonométricos fue inicialmente introducido por Hannan y otros (1970). Y una de las primeras aplicaciones económicas en el contexto de los modelos ARIMA correspondió a Pierce y otros (1984).

## Modelos estructurales de series temporales

$$\gamma_t = \sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{j,t},$$

donde  $\lfloor s/2 \rfloor$  denota la parte entera de  $s/2$  y cada término  $\tilde{a}_{j,t}$  es generado por

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} = \begin{pmatrix} \cos \lambda_j & \text{sen} \lambda_j \\ -\text{sen} \lambda_j & \cos \lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix},$$

para  $j = 1, \dots, \lfloor s/2 \rfloor$  y  $t = 1, \dots, T$ ;  $\lambda_j = 2\pi j/s$  es la frecuencia, en radianes, y  $\omega_{j,t}$  y  $\omega_{j,t}^*$  son dos términos de perturbación aleatoria mutuamente incorrelados distribuidos  $NID(0, \sigma_\omega^2)$  en ambos casos<sup>31</sup>. Si  $s$  es par,  $\gamma_{s/2,t} = -\gamma_{s/2,t-1} + \omega_{s/2,t}$ . Nótese que  $\gamma_{j,t}^*$  interviene por razones de construcción con el mismo fin que en el componente cíclico<sup>32</sup>.

Nótese también que en cualquiera de las dos especificaciones el número de parámetros que intervienen en el componente estacional es el mismo y, si  $\sigma_\omega^2 = 0$ , el patrón estacional es determinístico<sup>33</sup>.

Debido a que el patrón estacional suele cambiar de forma relativamente suave a lo largo del tiempo, podría ser razonable eliminar algunos de los términos trigonométricos correspondientes a las frecuencias más altas<sup>34</sup>. Aunque, cuando dichos términos son responsables de una parte importante de la variación estacional, ésta no sería una forma adecuada de proceder. Así ocurre cuando se dispone de observaciones semanales en las que existe un importante efecto intramensual, que se refleja en ciclos de periodos dos y cuatro semanas<sup>35</sup>; en este caso, el componente estacional podría formularse usando los términos

<sup>31</sup> Asignando diferentes varianzas a los términos de perturbación correspondientes a términos trigonométricos asociados a diferentes frecuencias, se consigue que las distintas fluctuaciones cíclicas posean diferente grado de variabilidad. Sin embargo, desde un punto de vista práctico es deseable que tengan la misma varianza porque se pierde poco en términos de bondad de ajuste y se gana mucho en tiempo de computación. Por otro lado, la especificación del componente estacional se podría modificar reemplazando el término de perturbación ruido blanco por un proceso ARMA.

<sup>32</sup> Esta formulación puede ser más efectiva para modelar comportamientos estacionales en series económicas debido a que en ella se asume que el patrón estacional evoluciona en el tiempo de acuerdo con un proceso MA(s-2), lo que implica oscilaciones más suaves que las generadas por un término de perturbación ruido blanco, que caracteriza la formulación a partir de variables cualitativas.

<sup>33</sup> Se recomienda esta opción cuando el número de años para los que se disponga de información sea pequeño y no haya un número suficiente de observaciones que permitan la consideración de un patrón estacional cambiante.

<sup>34</sup> Anderson (1971) y Abraham y Box (1978).

<sup>35</sup> Pierce y otros (1984).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

trigonométricos asociados a las frecuencias bajas así como a las frecuencias  $\pi/2$  y  $\pi$ , es decir,

$$y_t = \mu_t + \gamma_t^{bf} + \gamma_t^{im} + \varepsilon_t,$$

donde  $\gamma_t^{bf}$  contiene, por ejemplo, las variaciones correspondientes a las frecuencias  $2\pi j/52$ ,  $j = 1, \dots, 6$ , y  $\gamma_t^{im}$  las asociadas a las frecuencias  $\pi/2$  y  $\pi$ <sup>36</sup>.

También en el caso de datos semanales, el número de semanas por año varía entre 52 y 53 y la misma semana no recoge siempre el mismo periodo del año. Además, determinadas fiestas como la Semana Santa tampoco caen siempre en el mismo periodo del año. Una forma de abordar este problema consiste en considerar que el componente estacional es el resultado de combinar dos componentes, es decir,

$$\gamma_t = \gamma_t^p + \gamma_t^f,$$

donde  $\gamma_t^p$  representa el componente periódico, que es una función del número de días del año que han transcurrido hasta que se registra la observación<sup>37</sup>, y  $\gamma_t^f$  el componente calendario, que recoge los efectos asociados a determinadas fiestas móviles o incluso fiestas con fecha fija en el calendario que no corresponden siempre al mismo estadio estacional de una determinada fluctuación periódica<sup>38</sup>.

Si se dispone de observaciones diarias se puede incorporar al modelo un componente  $\theta_t$  para recoger el comportamiento atribuible al día de la semana al que pertenece la observación registrada en el momento  $t$ . Si se denota por  $w$  el número de tipos diferentes de días en una semana y por  $k_j$  el número de días del tipo  $j$ -ésimo en la semana, para  $j = 1, \dots, w$ , el efecto asociado con el  $j$ -ésimo tipo de día es  $\theta_{j,t}$  donde

$$\theta_{j,t} = \theta_{j,t-1} + \chi_{j,t}, \quad j = 1, \dots, w-1,$$

---

<sup>36</sup> Una formulación simplificada alternativa del componente estacional se puede construir a partir de funciones polinomiales de grado tres definidas por tramos (*splines* cúbicas), como se expondrá en el capítulo siguiente.

<sup>37</sup> Las especificaciones más habituales utilizadas para recoger este componente periódico son las conocidas funciones *splines* (véase capítulo segundo).

<sup>38</sup> Este planteamiento requiere trabajar sobre una base diaria, de modo que se conozca el número de días del año que han transcurrido desde el comienzo del año hasta que se registra cada observación (Harvey, Koopman y Riani, 1997).

## Modelos estructurales de series temporales

siendo  $\chi_{j,t}$  un término de perturbación de media cero, varianza  $\sigma_\chi^2(1-k_j^2/K)$ ,  $j=1, \dots, w-1$ , con  $K = \sum_{j=1}^w k_j^2$ , y covarianzas entre los términos de perturbación  $\chi_{j,t}$  y  $\chi_{h,t}$  igual a  $-\sigma_\chi^2 k_j k_h / K$ ,  $h, j=1, \dots, w-1$ ,  $j \neq h$ . Aunque cada uno de los  $w$  efectos diarios,  $\theta_{j,t}$ , evoluciona en el tiempo, el valor del componente diario en un instante dado del tiempo,  $\theta_t$ , viene dado por el valor del efecto asociado al día de la semana al que pertenece la observación registrada en el instante  $t$ , es decir,  $\theta_t = \theta_{j,t}$ ,  $t=1, \dots, T$ , donde  $j$  es el tipo de día al que pertenece la observación. Como el modelo se formula para los primeros  $w-1$  efectos diarios, el efecto para el tipo de día  $w$  se obtiene a partir de la condición de que, en cualquier momento del tiempo, la suma de los efectos diarios correspondientes a los diferentes días de una semana es igual a cero, es decir,  $\theta_{w,t} = -k_w^{-1} \sum_{j=1}^{w-1} k_j \theta_{j,t}$ <sup>39</sup>. El ámbito de aplicación de este último modelo no está limitado a los efectos diarios, sino que puede ser interpretado como una generalización del modelo formulado para el componente estacional en la que se admite que no todas las estaciones sean igualmente ponderadas<sup>40</sup>.

Por último, si aún siendo la periodicidad de las observaciones inferior a la anual, el análisis revela la ausencia de variaciones estacionales o se da el caso de que los datos han sido ajustados estacionalmente<sup>41</sup>, se puede plantear un modelo estructural sin componente estacional.

Combinando la tendencia bajo la formulación del modelo de tendencia lineal local y el componente estacional bajo cualquiera de las dos formulaciones estocásticas básicas, es decir, sin considerar posibles efectos intramensuales ni de calendario, se obtiene el *modelo*

<sup>39</sup>  $\theta_{w,t}$  tiene exactamente las mismas propiedades estadísticas que  $\theta_{j,t}$ ,  $j=1, \dots, w-1$ .

<sup>40</sup> Nótese que, a partir de la formulación anterior, el modelo convencional del componente estacional se obtiene asumiendo que  $w=s$  y  $k_j=1$  para  $j=1, \dots, s$ .

<sup>41</sup> Nótese que los datos ajustados estacionalmente no siempre tienen las propiedades deseadas, particularmente si el patrón estacional cambia de una forma que no es la considerada por los métodos de ajuste estándar. En general, se recomienda no emplear datos que hayan sido transformados con objeto de eliminar algunos de sus componentes; de ahí, la ventaja que proporciona la formulación estructural, que permite estimar conjuntamente los diferentes elementos de una serie temporal.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

estructural básico (MEB)<sup>42</sup>. Si se elige la formulación trigonométrica de la estacionalidad, el MEB queda especificado, por tanto, del modo siguiente:

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim NID(0, \sigma_\eta^2)$$

$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim NID(0, \sigma_\zeta^2)$$

$$\tilde{a}_t = \sum_{j=1}^{\lfloor s/2 \rfloor} \tilde{a}_{j,t},$$

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} = \begin{pmatrix} \cos \lambda_j & \text{sen } \lambda_j \\ -\text{sen } \lambda_j & \cos \lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix}, \quad j = 1, \dots, \lfloor s/2 \rfloor, \quad \lambda_j = 2\pi j / s,$$

$$\omega_{j,t} \sim NID(0, \sigma_\omega^2), \quad \omega_{j,t}^* \sim NID(0, \sigma_\omega^2), \quad \text{Cov}\{\omega_{j,t}, \omega_{j,t}^*\} = 0.$$

La especificación del modelo se completa suponiendo que no existe correlación entre los términos de perturbación de los distintos componentes debido a razones de identificabilidad<sup>43</sup>. Y, dado que las propiedades estadísticas del modelo dependen del conjunto completo de perturbaciones, la función de autocovarianzas de la serie observada es igual a la suma de las autocovarianzas de cada uno de los componentes.

Por último, como una extensión natural, el modelo puede incluir valores retardados de la variable dependiente así como variables explicativas exógenas, es decir,

$$y_t = \mu_t + \gamma_t + \psi_t + \sum_{\delta=1}^p \varphi_\delta y_{t-\delta} + \sum_{i=1}^k \sum_{\tau=0}^{q_k} \delta_{i\tau} x_{i,t-\tau} + \tilde{a}_t,$$

<sup>42</sup> Es equivalente a un proceso  $\Delta \Delta_s y_t \sim MA(s+1)$  o también  $y_t = \zeta_t / \Delta^2 + S^{-1}(L)\omega_t + \varepsilon_t$ . Dado que los operadores  $\Delta^2$  y  $S(L)$  no poseen raíces comunes, los cambios en el patrón estacional no se confunden con los cambios en la tendencia (Harvey, 1989:56). Un modelo estructural básico con variables cualitativas estacionales constituye una formulación equivalente en términos predictivos a la propuesta por Box y Jenkins (1976) para analizar la serie del *número de pasajeros*.

<sup>43</sup> La independencia de las innovaciones, que implica la independencia entre los respectivos componentes, es un supuesto establecido en la literatura desde el mismo momento en que se empezó a interpretar una serie observada como la agregación de componentes no observables. 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 (véase Espasa y Cancelo, 1993:280). Ahora bien, como señalan Harvey y Koopman (1992) y Koopman (1992:111), las estimaciones de dichas innovaciones pueden presentar correlaciones elevadas. De forma general, se puede asumir que existen estructuras de correlación entre los términos de perturbación y, de hecho, la relación entre los distintos componentes puede ser un supuesto más adecuado en series temporales económicas.

donde  $x_{i,t}$  son variables exógenas<sup>44</sup> y  $\delta_\tau$  y  $\tilde{a}_{i,\tau}$  son parámetros desconocidos. En el caso de que el modelo tenga incorporados retardos de la variable dependiente, por ejemplo

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \mu_t + \varepsilon_t,$$

la tendencia puede ser redefinida como

$$\mu_t^+ = \Phi_p^{-1}(L)\mu_t,$$

de forma que

$$y_t = \mu_t^+ + \Phi_p^{-1}(L)\varepsilon_t,$$

siendo  $\Phi_p(L) = (1 - \phi_1 L - \dots - \phi_p L^p)$ . En este caso, tanto el componente tendencial como el irregular están sometidos a efectos autorregresivos similares.

Las variables exógenas del modelo anterior pueden ser, en ocasiones, los regresores necesarios para la especificación de una función *spline* que recoja el componente estacional de una serie, como se expondrá en el capítulo siguiente. Otro caso de particular interés es aquél en el que las variables explicativas son variables de intervención que permiten recoger efectos de calendario<sup>45</sup>, observaciones anómalas y cambios estructurales<sup>46</sup>. En el caso más sencillo, una observación anómala que tiene lugar en un momento determinado del tiempo se recoge mediante una variable impulso. Sin embargo, un cambio estructural caracterizado por un aumento o disminución permanente en el componente de nivel de la serie a partir de un determinado momento del tiempo puede ser capturado añadiendo una variable de intervención tipo impulso en la ecuación del nivel, es decir,

$$\mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{I}_{t_0} I_{t_0} + \mathbf{h}_t, \quad \mathbf{h}_t \sim NID(0, \sigma_\eta^2),$$

donde  $I_{t_0}$  es una variable de intervención que toma el valor uno si la observación corresponde al instante del tiempo  $t_0$  y cero en cualquier otro caso<sup>47</sup>. Si, por el contrario, se produce un cambio permanente en la pendiente del componente tendencial a partir de un

<sup>44</sup> Se dice que  $x_{i,t}$  es exógena si no proporciona información sobre los componentes no observables en los instantes del tiempo  $t+s$ , para  $s=0,1,2,\dots$ , más allá de la contenida en  $y_{t-1}, \dots, y_1$  (véase Hamilton, 1994:373).

<sup>45</sup> Algunos patrones de comportamiento económico están muy relacionados con vacaciones o fiestas cuya fecha puede, por ejemplo, localizarse en meses distintos según el año. Estos efectos de calendario pueden ser recogidos mediante variables de intervención. Véase Liu (1980) y Bell y Hillmer (1983).

<sup>46</sup> Véase Harvey (1989:397-399), Harvey y Koopman (1992) y Harvey, Koopman y Penser (1997).

<sup>47</sup> Alternativamente, puede utilizarse una variable de intervención tipo escalón especificada junto al resto de componentes de la serie.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

determinado momento, dicha alteración puede ser recogida mediante una variable de intervención que toma los valores 1, 2, 3, ... a partir del momento del tiempo en el que es observado el cambio o, alternativamente, añadiendo una variable de intervención tipo impulso en la ecuación de la pendiente<sup>48</sup>.

### 1.1.2. Modelos de series temporales multivariantes y cointegración

Cuando se dispone de más de una serie temporal, cada una de ellas puede ser analizada de forma separada pero, debido a que las series pueden estar sometidas a similares influencias, se pueden generar estructuras de correlación que hacen interesante su consideración conjunta. De ahí, la necesidad de recurrir a los modelos estructurales de series temporales multivariantes, en los que la citada correlación se introduce a través de los elementos no nulos fuera de la diagonal de la matriz de covarianzas de los términos de perturbación de las ecuaciones<sup>49</sup>.

En este sentido, el enfoque multivariante significa una mayor aproximación a la realidad, en tanto que se pueden capturar las interacciones dinámicas entre las series; pero, en cualquier caso, los cimientos sobre los que se construye dicho planteamiento descansan en los establecidos para el modelo univariante. Este último se puede generalizar fácilmente al caso multivariante, en el que se analiza un vector  $N \times 1$  de variables observables  $y_t = (y_{1,t}, \dots, y_{N,t})'$ . A modo de ejemplo, la formulación multivariante del modelo de nivel local es  $y_{i,t} = \mu_{i,t} + \varepsilon_{i,t}$ , con  $\mu_{i,t} = \mu_{i,t-1} + \eta_{i,t}$ , para  $i = 1, \dots, N$  y  $t = 1, \dots, T$ , es decir,

$$\begin{pmatrix} y_{1,t} \\ \vdots \\ y_{N,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t} \\ \vdots \\ \mu_{N,t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{N,t} \end{pmatrix}$$

con

$$\begin{pmatrix} \mu_{1,t} \\ \vdots \\ \mu_{N,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t-1} \\ \vdots \\ \mu_{N,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \vdots \\ \eta_{N,t} \end{pmatrix}$$

---

<sup>48</sup> Aunque las variables de intervención tipo impulso o escalón resultan apropiadas para recoger cambios bruscos en el nivel o en la pendiente, también se pueden emplear para recoger cambios inesperados en el componente estacional. Si los comportamientos anómalos afectan específicamente al componente estacional, puede ser adecuado introducir un tipo particular de variables de intervención tales como el denominado *cambio de nivel estacional* (SLS) propuesto por Kaiser y Maravall (1999).

donde  $\mathbf{e}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})' \sim NID(0, \Sigma_\varepsilon)$  y  $\mathbf{h}_t = (\eta_{1,t}, \dots, \eta_{N,t})' \sim NID(0, \Sigma_\eta)$ . Cada una de las variables del vector  $\mathbf{y}_t$  tiene su propio nivel estocástico y la matriz  $\mathbf{S}_\eta$  mide las covarianzas entre los diferentes niveles. Este modelo se denomina *modelo SUTSE* (*seemingly unrelated time series equations*) o modelo de ecuaciones de series temporales aparentemente no relacionadas, dado que las series sólo están conectadas a través de las correlaciones contemporáneas que se establezcan en las matrices  $\mathbf{S}_\varepsilon$  y  $\mathbf{S}_\eta$ .

Por supuesto, estos modelos multivariantes pueden incorporar especificaciones más complejas del componente tendencial, así como componentes cíclico y estacional. En cuanto al componente cíclico, y en el caso particular de series económicas, parece razonable asumir que los movimientos de este tipo en las diferentes series estén relacionados con un ciclo económico común, de modo que posean también un periodo común. Consideraciones similares se pueden efectuar en relación al componente estacional.

Bajo ciertos supuestos, el modelo estructural multivariante de nivel local admite la representación de un modelo de componentes de error de la forma

$$y_{i,t} = \mu_{i,t} + \varepsilon_t + \varepsilon_{i,t}^*,$$

$$\mu_{i,t} = \mu_{i,t-1} + \eta_t + \eta_{i,t}^*,$$

siendo  $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$ ,  $\varepsilon_{i,t}^* \sim NID(0, \sigma_{\varepsilon^*}^2)$ ,  $\eta_t \sim NID(0, \sigma_\eta^2)$ ,  $\eta_{i,t}^* \sim NID(0, \sigma_{\eta^*}^2)$ , para todo  $i = 1, \dots, N$  y  $t = 1, \dots, T$ <sup>50</sup> y donde  $\varepsilon_t$  y  $\eta_t$  recogen los efectos comunes, mientras que  $\varepsilon_{i,t}^*$  y  $\eta_{i,t}^*$  recogen los efectos específicos. Un caso especialmente interesante es aquél en el que  $\sigma_\eta^2 > 0$ , pero  $\sigma_{\eta^*}^2 = 0$ ; entonces, las  $N$  series tienen componentes tendenciales que se mueven de forma paralela y están cointegradas.

En general, los modelos multivariantes deben admitir la posibilidad de que existan tendencias comunes para algunas de las variables que componen el vector  $\mathbf{y}_t$ <sup>51</sup>. En este caso, existe cointegración entre las series, que puede ser incorporada en un modelo del tipo

<sup>49</sup> Véase Harvey y Koopman (1996).

<sup>50</sup> Nótese que  $\mathbf{S}_\varepsilon = \sigma_{\varepsilon^*}^2 \mathbf{I} + \sigma_\varepsilon^2 \mathbf{ii}'$  y  $\mathbf{S}_\eta = \sigma_{\eta^*}^2 \mathbf{I} + \sigma_\eta^2 \mathbf{ii}'$ , donde  $\mathbf{i}$  es el vector unitario de dimensión  $N$  e  $\mathbf{I}$  es la matriz identidad.

<sup>51</sup> Aquellas situaciones en las que las series tienen la misma fuente de crecimiento constituyen un ejemplo en el que tiene sentido la formulación de tendencias comunes.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$y_t = \mathbf{Q}\mathbf{m}_t^* + \mathbf{m}_0 + \mathbf{e}_t,$$

$$\mathbf{m}_t^* = \mathbf{m}_{t-1}^* + \mathbf{h}_t^*,$$

con  $\mathbf{e}_t \sim NID(0, \Sigma_\varepsilon)$  y  $\mathbf{h}_t^* \sim NID(0, \Sigma_{\eta^*})$ , para  $t = 1, \dots, T$ , siendo  $\mathbf{S}_{\eta^*}$  una matriz diagonal  $k \times k$ ,  $\mathbf{m}_0$  un vector  $N \times 1$  cuyos  $k$  primeros elementos son nulos y  $\mathbf{Q}$  una matriz  $N \times k$  fija tal que  $\Theta_{ij} = 0$ ,  $\forall j > i$ , y  $\Theta_{ii} = 1$ ,  $\forall i$ . Si hay  $k$  tendencias comunes, habrá  $N-k$  relaciones de cointegración. Los correspondientes  $N-k$  vectores de cointegración definen las  $N-k$  filas de la matriz  $\mathbf{A}_{(N-k) \times N}$ , tal que  $\mathbf{A}\mathbf{Q} = \mathbf{0}$ , es decir,  $\mathbf{A}y_t = \mathbf{A}\mathbf{m}_0 + \mathbf{A}\mathbf{e}_t$ , de modo que  $\mathbf{A}y_t$  es un proceso estacionario de dimensión  $(N-k) \times 1$ . Si  $k = N$ , hay  $N$  tendencias y, por tanto, no hay cointegración<sup>52</sup>. Como  $\mathbf{S}_{\eta^*}$  es diagonal, las tendencias comunes (los elementos de  $\mathbf{m}_t^*$ ) son independientes. Para un modelo con componente tendencial de nivel local con deriva, la especificación del modelo de tendencias comunes es

$$y_t = \mathbf{Q}\mathbf{m}_t^* + \mathbf{m}_0 + \mathbf{e}_t,$$

$$\mathbf{m}_t^* = \mathbf{m}_{t-1}^* + \mathbf{b} + \mathbf{h}_t^*,$$

para  $t = 1, \dots, T$ ,  $\mathbf{e}_t \sim NID(0, \Sigma_\varepsilon)$  y  $\mathbf{h}_t^* \sim NID(0, \Sigma_{\eta^*})$ , siendo  $\mathbf{b}$  un vector  $k \times 1$  de pendientes. A modo de ejemplo, el modelo de nivel local bivalente se puede representar como

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix},$$

$$\begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix},$$

para  $t = 1, \dots, T$ , donde

$$\mathbf{S}_\varepsilon = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & \sigma_{\varepsilon_{12}}^2 \\ \sigma_{\varepsilon_{21}}^2 & \sigma_{\varepsilon_2}^2 \end{pmatrix}$$

y

$$\mathbf{S}_\eta = \begin{pmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_{12}}^2 \\ \sigma_{\eta_{21}}^2 & \sigma_{\eta_2}^2 \end{pmatrix}.$$

<sup>52</sup> Una forma de determinar  $k$  se basa en estimar el modelo original como si no hubieran factores comunes, aplicar un análisis de componentes principales sobre la matriz  $\mathbf{S}_\eta$  estimada y tomar el valor de  $k$  que recoja

Alternativamente, su representación como un modelo de componentes de error es

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \end{pmatrix} + \varepsilon_t \mathbf{i}_2 + \begin{pmatrix} \varepsilon_{1,t}^* \\ \varepsilon_{2,t}^* \end{pmatrix},$$

$$\begin{pmatrix} \mu_{1,t} \\ \mu_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{pmatrix} + \eta_t \mathbf{i}_2 + \begin{pmatrix} \eta_{1,t}^* \\ \eta_{2,t}^* \end{pmatrix},$$

para  $t = 1, \dots, T$ , siendo  $\mathbf{i}_2 = (1, 1)'$ ,  $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$ ,  $\varepsilon_{i,t}^* \sim NID(0, \sigma_{\varepsilon^*}^2)$  para  $i = 1, 2$ ,  $\eta_t \sim NID(0, \sigma_\eta^2)$  y  $\eta_{i,t}^* \sim NID(0, \sigma_{\eta^*}^2)$  para  $i = 1, 2$ . Si  $\sigma_{\eta^*}^2 = 0$ , se obtiene que  $\mu_{1,t} = \mu_{1,t-1} + \eta_t$  y  $\mu_{2,t} = \mu_{2,t-1} + \eta_t$  de forma que los componentes tendenciales de ambas series siguen sendas paralelas. Resulta que los procesos  $y_{1,t}$  e  $y_{2,t}$  son no estacionarios, puesto que no se mueven alrededor de un nivel constante, pero existe una tendencia común compartida por ambas series. Es decir,  $y_{1,t}$  e  $y_{2,t}$  son I(1), pero existe una relación lineal estacionaria y, por tanto, las series están cointegradas. Si  $\mu_{1,0} = \mu_{2,0}$ , dicha relación es  $y_{1,t} - y_{2,t}$ .

En general, la presencia de cointegración en el caso bivalente se puede recoger mediante un modelo de tendencias comunes especificado como

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 1 \\ \theta_{2,1} \end{pmatrix} \mu_t^* + \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix},$$

donde

$$\mu_t^* = \mu_{t-1}^* + \eta_t^*.$$

Es decir, se tiene la siguiente representación para las series consideradas

$$y_{1,t} = \mu_t^* + \varepsilon_{1,t},$$

$$y_{2,t} = \theta_{2,1} \mu_t^* + \mu_2 + \varepsilon_{2,t},$$

donde

$$\mu_t^* = \mu_{t-1}^* + \beta + \eta_t^*.$$

Por tanto,

$$y_{1,t} - \frac{1}{\theta_{2,1}} y_{2,t} = -\frac{\mu_2}{\theta_{2,1}} + \varepsilon_{1,t} - \frac{1}{\theta_{2,1}} \varepsilon_{2,t}$$

---

una parte importante de la variabilidad total. Un procedimiento más formal puede llevarse a cabo en la línea

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

es una combinación lineal estacionaria que implica cointegración. El vector de cointegración  $(1 \quad -1/\theta_{2,1})$  es tal que  $(1 \quad -1/\theta_{2,1}) \begin{pmatrix} 1 \\ \theta_{2,1} \end{pmatrix} = 0$ , eliminándose, de este modo, el elemento causante de la no estacionariedad<sup>53</sup>.

Las tendencias comunes especificadas implican cointegración en la frecuencia cero, pero todos estos conceptos pueden ser extendidos a otras frecuencias, admitiendo entonces, la existencia de componentes cíclicos o estacionales comunes, que pueden ser interpretados en el sentido de que los cambios en los patrones cíclicos o estacionales de las diferentes series proceden de una fuente común.

Las propiedades de un modelo multivariante  $y_t$ , en concreto, del modelo de nivel local multivariante dependen de las estructuras de las matrices de varianzas-covarianzas  $\mathbf{S}_\varepsilon$  y  $\mathbf{S}_\eta$ . Bajo la restricción de homogeneidad, que establece que  $\mathbf{S}_\eta = q\mathbf{S}_\varepsilon$ <sup>54</sup>, todas las combinaciones lineales de los elementos de  $y_t$  tienen las mismas propiedades estocásticas<sup>55</sup>.

Cuando en un modelo multivariante todas menos una de las variables componentes del vector  $y_t$  son exógenas, no es necesario plantear un sistema con  $N$  ecuaciones sino que se puede formular un modelo de regresión dinámico uniecuacional. En el contexto de los modelos estructurales, se puede obtener un modelo uniecuacional con variables exógenas del tipo

$$y_t = \mathbf{m}_t + X_t \mathbf{d} + \mathbf{e}_t,$$

con

$$\mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{h}_t,$$

$$X_t = (x_{1,t}, \dots, x_{N-1,t})$$

y

sugerida por Stock y Watson (1988).

<sup>53</sup> Otros ejemplos de la especificación del modelo de tendencias comunes se pueden encontrar en el anexo 1.1.

<sup>54</sup> Véase Harvey (1989:435-436).

<sup>55</sup> Es decir, para un modelo de nivel local,  $y_t$  es homogéneo si el proceso estacionario  $\Delta \mathbf{w}' y_t$  posee la misma potencia espectral y la misma función de autocorrelación para todo vector  $\mathbf{w}$  de parámetros fijos (no estocásticos). En ese caso, el modelo estructural para cualquier combinación  $\mathbf{w}' y_t$  verifica la relación  $\mathbf{S}_h = q\mathbf{S}_\varepsilon$  y las funciones de autocorrelación y de densidad espectral dependen sólo de  $q$ .

$$\mathbf{d} = (\delta_1, \dots, \delta_{N-1})'$$

Si la varianza del nivel es igual a cero, se obtiene un modelo de regresión clásico. Cuando la varianza del componente irregular es cero pero la varianza del componente de nivel es distinta de cero, se obtiene una regresión en primeras diferencias. Finalmente, cuando ambas varianzas son distintas de cero, se obtiene un modelo de función de transferencia. Casos particulares de inclusión de variables explicativas en modelos estructurales multivariantes son el análisis de intervención y el modelo del mecanismo de corrección del error<sup>56</sup>.

### 1.2. Representación en el espacio de los estados y filtro de Kalman

El tratamiento estadístico de los modelos estructurales de series temporales está basado en su representación en el espacio de los estados o *SSF* (*state space form*), una herramienta de enorme potencial que permite el tratamiento de un amplio rango de modelos de series temporales<sup>57</sup>. Los modelos en el espacio de los estados surgieron en el campo de la ingeniería. El artículo de Kalman (1960) significó un importante salto cualitativo, ya que mostró que una extensa clase de problemas podía representarse mediante un modelo lineal simple en el espacio de los estados, cuya naturaleza markoviana permite que los cálculos necesarios para su aplicación práctica se efectúen de forma recursiva. Estas ideas se desarrollaron posteriormente en el ámbito de investigación de los ingenieros<sup>58</sup>; mientras que las contribuciones de estadísticos y econométricos fueron esporádicas hasta finales de los 70<sup>59</sup>, aunque en los últimos años el interés mostrado por estos últimos ha crecido notablemente<sup>60</sup>.

---

<sup>56</sup> Véase Harvey y Scott (1994).

<sup>57</sup> Concretamente, el espacio de los estados es una representación útil en el caso de magnitudes que puedan expresarse como suma de procesos estocásticos o que son observadas con errores de medida (véase Hamilton, 1994:376). Así, los modelos en el espacio de los estados pueden entenderse como una representación unificada de modelos tales como los modelos ARIMA (Box y Jenkins, 1976), modelos ARMAX (Hannan y Deistler, 1988), modelos estructurales (Harvey, 1989), modelos tipo intervención (Box y Tiao, 1975) y modelos de regresión con perturbaciones ARMA.

<sup>58</sup> Cuando se publica el famoso artículo de Kalman, la metodología que en él se describe se estaba utilizando en la industria aeroespacial para estimar la posición de los planetas (Jones, 1993:77).

<sup>59</sup> Algunos buenos ejemplos de análisis de series temporales en los que la representación en el espacio de los estados se aplica de forma satisfactoria son Harvey y Phillips (1979), Engle y Watson (1981), Burrige y Wallis (1985) y Burmeister y otros (1986).

<sup>60</sup> Véase Durbin y Koopman (2001:52). Algunas aportaciones recientes de autores españoles son las realizadas por Vargas (1999), Vargas (2000) y Vargas y Mínguez (2000). Resulta curioso que, aunque la

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

El estado o situación de un sistema en un momento dado del tiempo se describe a través de un conjunto de variables que forman el llamado vector de estado. El espacio de los estados es, por tanto, el espacio donde los sucesivos vectores de estado describen la evolución del sistema como función del tiempo. En un modelo estructural de series temporales los elementos del vector de estado son los componentes no observables de la serie.

Una vez que el modelo ha sido formulado en el espacio de los estados, el filtro de Kalman proporciona el medio de estimar el estado inobservable a partir de alguna magnitud observable relacionada con éste, de forma que la estimación se actualice cada vez que se dispone de nueva información. Se pueden emplear varios algoritmos de alisado, que constituyen extensiones de la formulación original diseñada por Kalman para obtener una estimación óptima del estado en cualquier punto de la muestra.

Supóngase un sistema descrito por un vector de estado  $m$ -dimensional  $\mathbf{a}_t$ , en los instantes del tiempo  $t = 0, 1, 2, \dots, T$ , como por ejemplo una determinada economía nacional a lo largo de diferentes años. Las variables de estado que describen la economía nacional no son observables<sup>61</sup> o, en general, no se observan en un estado puro sino que están contaminadas por un componente ruidoso, de modo que es posible que lo realmente observado sea  $y_t$ , siendo

$$y_t = \mathbf{a}_t + \mathbf{e}_t,$$

donde  $\mathbf{e}_t$  es el componente ruidoso<sup>62</sup>; o, de forma más general, se observa un vector  $N \times 1$  de variables,  $y_t$ , relacionadas con el vector de estado a través de la expresión

---

formulación en el espacio de los estados surgió al margen del campo específico del análisis de series temporales, los nuevos avances en los modelos estructurales de series temporales y sus aplicaciones han contribuido a la representación en el espacio de los estados (Harvey y Shephard, 1993).

<sup>61</sup> Los modelos de componentes no observables se han mostrado apropiados para el tratamiento de problemas económicos en los que una variable, que se supone que juega un papel importante en este contexto, no es directamente observable; y, de hecho, han sido extensamente utilizados en la investigación económica en áreas tales como el estudio de la reacción de los agentes a cambios (transitorios y permanentes) en el nivel de precios (Lucas, 1976), la estabilidad de algunos de los *grandes* ratios macroeconómicos (Pagan, 1975), el nivel natural de la oferta laboral (Bull y Frydman, 1983), el progreso técnico (Slade, 1989, y Harvey y Marshall, 1991) o la evolución de varias macromagnitudes de las economías de algunos países europeos y de EEUU (Aoki, 1990).

<sup>62</sup> Como afirma Maybeck, «... los sistemas dinámicos están dirigidos no sólo por *inputs* sometidos a nuestro propio control, sino también por perturbaciones que no podemos controlar ni modelar determinísticamente» (Maybeck, 1979:2). Un ejemplo económico de esta situación fue analizado por Moody y Wu (1998), que utiliza la formulación en el espacio de los estados para descomponer los precios observados en los

$$y_t = Z_t \mathbf{a}_t + \mathbf{e}_t,$$

donde  $Z_t$  es alguna transformación lineal. Pues bien, si se desea estimar el valor del estado, magnitud que evoluciona en el tiempo y que no se observa directamente sino contaminada por dicho componente ruidoso, se puede recurrir a los denominados filtros adaptativos, que poseen características de autoajuste, es decir, permiten corregir la estimación cuando se recibe nueva información. Uno de estos filtros fue desarrollado por Kalman (1960) para el caso de sistemas dinámicos en tiempo discreto<sup>63</sup>.

El propósito del análisis en el espacio de los estados es inferir las propiedades relevantes de los estados  $\mathbf{a}_1, \dots, \mathbf{a}_T$  a partir de los valores observados  $y_1, \dots, y_T$  de la magnitud relacionada con el estado y del conocimiento de las propiedades estadísticas del ruido. Para ello, es preciso establecer, además, una relación entre los estados en los momentos  $t$  y  $t-1$ , es decir, entre  $\mathbf{a}_t$  y  $\mathbf{a}_{t-1}$ . En la propuesta de Kalman se asume que

$$\mathbf{a}_t = \mathbf{S}_t \mathbf{a}_{t-1} + \mathbf{u}_t,$$

donde  $\mathbf{S}_t$  es alguna transformación lineal que describe la evolución en el tiempo del sistema,  $\mathbf{u}_t$  es un término de ruido y se asume que  $\mathbf{e}_t$  y  $\mathbf{u}_t$  son vectores de variables aleatorias gaussianas, de media nula y serialmente incorreladas. Kalman concluye que la mejor estimación para el estado  $\mathbf{a}_t$  puede ser obtenida recursivamente a partir de la mejor estimación de  $\mathbf{a}_{t-1}$  y su matriz de covarianzas. A continuación, se hace una exposición más detallada de los aspectos fundamentales de la representación en el espacio de los estados y del filtro de Kalman<sup>64</sup>.

---

*verdaderos* precios no observables y un ruido derivado del error de medida. Otro ejemplo similar es la aplicación que realizan Bolland y otros (1998).

<sup>63</sup> Kalman y Bucy (1961) desarrollaron el filtro para el caso de sistemas dinámicos en tiempo continuo. El filtro de Kalman se puede aplicar tanto para magnitudes estacionarias como para no estacionarias. Hasta su formulación por Kalman y Bucy en 1960-1961, se aplicaba otro filtro adaptativo, desarrollado por Wiener en los años 40, que sólo es válido para magnitudes estacionarias.

<sup>64</sup> Véase, entre otros, Kalman (1960), Kalman y Bucy (1961), Harrison y Stevens (1976), Anderson y Moore (1979), Wegman (1982), Meinhold y Singpurwalla (1983), Harvey (1989), Steyn (1996) y Durbin y Koopman (2001).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

### 1.2.1. Representación en el espacio de los estados

Sea un sistema descrito por un vector de estado  $m \times 1$ ,  $\mathbf{a}_t$ , en los instantes del tiempo  $t = 1, \dots, T$ <sup>65</sup> y sea  $\mathbf{y}_t$  un vector  $N \times 1$  de variables observables. La formulación básica del modelo en el espacio de los estados consta de dos ecuaciones. Bajo el supuesto de que el estado sigue un proceso de Markov de primer orden, la *ecuación de transición* establece una relación entre el estado actual y el estado en el instante del tiempo anterior<sup>66</sup>, es decir,

$$\mathbf{a}_t = \mathbf{S}_t \mathbf{a}_{t-1} + \mathbf{c}_t + \mathbf{R}_t \mathbf{u}_t, \quad t = 1, \dots, T,$$

donde  $\mathbf{S}_t$  es una matriz  $m \times m$ ,  $\mathbf{c}_t$  es un vector  $m \times 1$ ,  $\mathbf{R}_t$  es una matriz  $r \times m$  ( $r \leq m$ ) y  $\mathbf{u}_t$  es un vector  $r \times 1$  de perturbaciones no correlacionadas serialmente con media cero y matriz de covarianzas  $\mathbf{Q}_t$ <sup>67</sup>. Dado que el estado es inobservable, es preciso establecer una relación entre el vector  $N \times 1$  de variables observables y el vector de estado  $m \times 1$ , que se define a través de la *ecuación de medida*

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{d}_t + \mathbf{e}_t, \quad t = 1, \dots, T,$$

donde  $\mathbf{Z}_t$  es una matriz  $N \times m$ ,  $\mathbf{d}_t$  es un vector  $N \times 1$  y  $\mathbf{e}_t$  es un vector  $N \times 1$  de perturbaciones no

---

<sup>65</sup> Aunque hay muchas formas equivalentes de expresar un sistema en la representación *SSF*, cada una de ellas proporciona los mismos valores de la función de verosimilitud y, por tanto, las mismas predicciones del sistema, pero las soluciones no son siempre igualmente fáciles de obtener. Así, resulta deseable definir  $\mathbf{a}_t$  de forma que contenga toda la información relevante del sistema en el momento  $t$  con el menor número de elementos. De modo que, si las demás condiciones se mantienen, se prefiere aquella especificación que posea una realización mínima, es decir, aquella que minimice la longitud del vector de estado.

<sup>66</sup> La ecuación de transición resume el comportamiento dinámico del estado del sistema. Aunque normalmente se asume que los estados varían de acuerdo a un proceso de camino aleatorio, también se podrían especificar otras formas. Sin embargo, la opción de camino aleatorio suele ser la alternativa a la que se recurre en primer lugar, dado que es el proceso más simple y, desde el punto de vista práctico, funciona correctamente en muchos modelos de series económicas (Bohara y Sauer, 1992 y Kim, 1993).

<sup>67</sup> En muchas aplicaciones  $\mathbf{R}_t = \mathbf{I}_m$ , con lo que la ecuación de transición puede escribirse como  $\mathbf{a}_t = \mathbf{S}_t \mathbf{a}_{t-1} + \mathbf{c}_t + \mathbf{u}_t$ , donde  $\mathbf{u}_t$  es un vector de dimensión  $m \times 1$ . Pero en otras conviene definir  $\mathbf{R}_t$  como una matriz  $r \times m$ , con  $r < m$ . Por ejemplo, Durbin y Koopman (2001:38) asumen que  $\mathbf{R}_t$  es un subconjunto de columnas de  $\mathbf{I}_m$ , de modo que se trata de una matriz que selecciona aquellas filas del conjunto de ecuaciones de estado con términos de perturbación distintos de cero. En ocasiones, la ecuación de transición se expresa como  $\mathbf{a}_{t+1} = \mathbf{S}_t \mathbf{a}_t + \mathbf{c}_t + \mathbf{u}_t$ ,  $t = 1, \dots, T$  (Jazwinski, 1970, y Anderson y Moore, 1979), y, aunque desde el punto de vista práctico existen pocas diferencias, es necesario tener cuidado en ciertos casos como, por ejemplo, aquéllos en los que existe correlación entre las perturbaciones de la ecuación de medida y de transición.

Nótese también que si  $m = 1$ ,  $\mathbf{a}_t$  es una variable de estado unidimensional,  $\mathbf{S}_t$  es un coeficiente para cada instante del tiempo y  $\mathbf{u}_t$  es un término de perturbación con elementos serialmente incorrelados con media cero y varianza  $q_t$ .

## Modelos estructurales de series temporales

correlacionadas serialmente<sup>68</sup> con media nula y matriz de covarianzas  $H_t$ <sup>69</sup>. La especificación de la formulación en el espacio de los estados requiere dos supuestos adicionales.

1) Respecto a las condiciones iniciales, el vector de estado inicial,  $\mathbf{a}_0$ , tiene media  $\mathbf{a}_0$  y matriz de covarianzas  $\mathbf{P}_0$ , es decir,

$$E[\mathbf{a}_0] = \mathbf{a}_0$$

y

$$Var[\mathbf{a}_0] = \mathbf{P}_0.$$

2) Los términos de perturbación de la ecuación de medida y de la ecuación de transición están incorrelados entre sí y con el vector de estado inicial en cualquier momento del tiempo, es decir,

$$E[\mathbf{e}_t \mathbf{u}'_s] = \mathbf{0}, \forall s, t = 1, \dots, T,$$

$$E[\mathbf{e}_t \mathbf{a}'_0] = \mathbf{0}, t = 1, \dots, T,$$

y

$$E[\mathbf{u}_t \mathbf{a}'_0] = \mathbf{0}, t = 1, \dots, T$$
<sup>70</sup>.

Las matrices  $\mathbf{Z}_t$ ,  $\mathbf{d}_t$ ,  $\mathbf{H}_t$ ,  $\mathbf{S}_t$ ,  $\mathbf{c}_t$ ,  $\mathbf{R}_t$  y  $\mathbf{Q}_t$  se denominan matrices del sistema y, a menos que se especifique lo contrario, se consideran no estocásticas, aunque podrían cambiar a lo largo del tiempo de forma predeterminada. De esta manera, el sistema es lineal y para cualquier momento del tiempo, el vector de variables observables,  $\mathbf{y}_t$ , se puede expresar como una combinación lineal de valores presentes y pasados de perturbaciones,  $\mathbf{u}_t$  y  $\mathbf{e}$ , y del vector de estado inicial,  $\mathbf{a}_0$ . Si las matrices del sistema no cambian con el tiempo, el modelo se denomina homogéneo o invariante respecto al tiempo<sup>71</sup> y los modelos no

<sup>68</sup> Nótese que la incorrelación serial de los elementos del vector de perturbaciones implica la ausencia de correlación serial para cada uno de los elementos del vector así como para dos elementos cualesquiera de éste.

<sup>69</sup> Si  $N = 1$ ,  $\mathbf{y}_t$  es una variable unidimensional,  $\mathbf{Z}_t$  es un vector  $1 \times m$  de coeficientes para cada instante del tiempo y  $\mathbf{e}_t$  es un término de perturbación con media cero y varianza  $h_t$  y, además, los elementos de este proceso estocástico están serialmente incorrelados.

<sup>70</sup> Este supuesto se puede relajar dando lugar a modificaciones en varios de los algoritmos.

<sup>71</sup> En un modelo homogéneo respecto al tiempo la ecuación de transición es un proceso  $VAR(1)$ , pero este hecho no constituye un supuesto restrictivo debido a que pueden considerarse procesos de órdenes superiores. Por otro lado, un modelo no tiene que ser autorregresivo para que pueda ser representado en el espacio de los estados sino que, por el contrario, cualquier proceso estocástico admite dicha representación.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

estacionarios constituyen casos particulares que, siendo invariantes respecto al tiempo, se transforman en estacionarios al aplicarles el operador diferencia.

Las matrices del sistema  $\mathbf{Z}_t$ ,  $\mathbf{H}_t$ ,  $\mathbf{S}_t$ ,  $\mathbf{R}_t$  y  $\mathbf{Q}_t$  pueden depender de un conjunto de parámetros desconocidos que determinan las propiedades estocásticas del modelo. Se denominan hiperparámetros para distinguirlos de los que intervienen en los vectores  $\mathbf{d}_t$  y  $\mathbf{c}_t$ , que sólo afectarán al valor esperado del estado y las observaciones de forma determinística<sup>72</sup>.

Si se adopta una representación simplificada en el espacio de los estados de la forma:

$$\begin{aligned}\mathbf{a}_t &= \mathbf{S}\mathbf{a}_{t-1} + \mathbf{u}_t, \\ \mathbf{y}_t &= \mathbf{Z}_t\mathbf{a}_t + \mathbf{e}_t,\end{aligned}$$

donde  $\mathbf{a}_t$  es un vector de estado  $m \times 1$ ,  $\mathbf{S}$  es una matriz de transición  $m \times m$ ,  $\mathbf{u}_t$  es el vector  $m \times 1$  de perturbaciones de la ecuación de transición tal que  $\mathbf{u}_t \sim i.i.d.N_m(\mathbf{0}, \mathbf{Q})$ ,  $\mathbf{y}_t$  es un vector  $N \times 1$  de variables observables,  $\mathbf{Z}_t$  es una matriz de medida  $N \times m$  y  $\mathbf{e}_t$  es el vector  $N \times 1$  de perturbaciones de la ecuación de medida tal que  $\mathbf{e}_t \sim i.i.d.N_N(\mathbf{0}, \mathbf{H})$ , entonces, por sustituciones recursivas, se tiene que

$$\mathbf{a}_t = \mathbf{S}^t \mathbf{a}_0 + \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{u}_{t-i}$$

y, por tanto,

$$\mathbf{y}_t = \mathbf{Z}_t \left[ \mathbf{S}^t \mathbf{a}_0 + \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{u}_{t-i} \right] + \mathbf{e}_t.$$

A partir de estas igualdades, las propiedades estadísticas del modelo son<sup>73</sup>:

- 1)  $E[\mathbf{a}_t] = \mathbf{S}^t E[\mathbf{a}_0] = \mathbf{S}^t \mathbf{a}_0$ ,
- 2)  $E[\mathbf{y}_t] = \mathbf{Z}_t \mathbf{S}^t E[\mathbf{a}_0] = \mathbf{Z}_t \mathbf{S}^t \mathbf{a}_0$ ,
- 3)  $E[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'] = \left[ \mathbf{S}^t \mathbf{P}_0 (\mathbf{S}^{t-k})' + \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \mathbf{Q} (\mathbf{S}^i)' \right] = \mathbf{P}_{t,t-k}$ ,

<sup>72</sup> Esta conclusión no se sostiene si  $\mathbf{d}_t$  es función de valores retardados de  $\mathbf{y}_t$ . Por otra parte, como señalan Durbin y Koopman (2001:30), los parámetros de modelos en el espacio de los estados se suelen denominar hiperparámetros con objeto de distinguirlos de los elementos de los vectores de estado, que pueden ser considerados como parámetros aleatorios.

<sup>73</sup> En el anexo 1.2 se pueden encontrar las demostraciones de estas propiedades.

con  $\mathbf{P}_{t,t} \equiv \mathbf{P}_t$ ,

$$4) E\left[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_{t-k} - E[\mathbf{y}_{t-k}])'\right] = \left[ \mathbf{Z}_t \mathbf{S}^t \mathbf{P}_0 (\mathbf{Z}_{t-k} \mathbf{S}^{t-k})' + \mathbf{Z}_t \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \mathbf{Q} (\mathbf{Z}_{t-k} \mathbf{S}^i)' \right],$$

si  $k \neq 0$ ,

$$5) E\left[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_t - E[\mathbf{y}_t])'\right] = \left[ \mathbf{Z}_t \mathbf{S}^t \mathbf{P}_0 (\mathbf{Z}_t \mathbf{S}^t)' + \mathbf{Z}_t \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{Q} (\mathbf{Z}_t \mathbf{S}^i)' + \mathbf{H} \right].$$

Nótese que la estacionariedad en media del estado se produce si  $\mathbf{a}_0 = 0$ , o bien, si  $\mathbf{S}$  posee algún valor propio unitario tal que  $\mathbf{S}'\mathbf{a}_0$  es constante. Por otra parte, se puede demostrar que los momentos de segundo orden son invariantes respecto a  $t$  si y sólo si todos los autovalores de  $\mathbf{S}$  caen dentro del círculo de radio unidad (Steyn, 1996:10).

Finalmente, la función de verosimilitud de los  $T$  vectores de observaciones se puede obtener del modo siguiente. Sean los vectores de vectores  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$ ,  $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_T)'$  y  $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_T)'$ , y sea la matriz  $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_T)$ ; entonces,

$$\mathbf{y} = \mathbf{Z}\mathbf{a} + \mathbf{e},$$

donde

$$\mathbf{e} \sim N_{NT}(\mathbf{0}, (\mathbf{I}_T \otimes \mathbf{H})).$$

Por otro lado, se tiene que

$$\mathbf{a} = \mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{S}_2 \mathbf{u},$$

donde  $\mathbf{S}_1 = \text{diag}(\mathbf{S}, \mathbf{S}^2, \dots, \mathbf{S}^T)$ ,  $\mathbf{i}_T$  es un vector unitario  $T \times 1$ ,

$$\mathbf{S}_2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{S} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}^{T-1} & \mathbf{S}^{T-2} & \dots & \mathbf{I} \end{bmatrix}$$

y

$$\mathbf{u} \sim N_{mT}(\mathbf{0}, (\mathbf{I}_T \otimes \mathbf{Q})).$$

Por tanto,

$$E[\mathbf{y}] = \mathbf{Z} \left[ (\mathbf{S}\mathbf{a}_0)', \dots, (\mathbf{S}^T \mathbf{a}_0)' \right]' \equiv \mathbf{m},$$

mientras que, dado que

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$\mathbf{y} = \mathbf{Z}[\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{S}_2 \mathbf{u}] + \mathbf{e},$$

se obtiene que

$$E[(\mathbf{y} - \mathbf{m})(\mathbf{y} - \mathbf{m})'] = [\mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \mathbf{i}_T' \otimes \mathbf{P}_0)\mathbf{S}_1' \mathbf{Z}' + \mathbf{Z}\mathbf{S}_2(\mathbf{I}_T \otimes \mathbf{Q})\mathbf{S}_2' \mathbf{Z}' + (\mathbf{I}_T \otimes \mathbf{H})] \equiv \mathbf{S}.$$

Como  $\mathbf{y}$  es normal multivariante de dimensión  $NT$ , con media  $\mathbf{m}$  y varianza  $\mathbf{S}$ , la función de verosimilitud y el logaritmo natural de ésta son

$$L(\mathbf{y}) = \frac{\sqrt{|\mathbf{S}|^{-1}}}{(2\pi)^{NT/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{m})\mathbf{S}^{-1}(\mathbf{y} - \mathbf{m})'\right\}$$

y

$$\ln L(\mathbf{y}) = -\frac{1}{2} \ln |\mathbf{S}| - \frac{NT}{2} \ln(2\pi) - \frac{1}{2}(\mathbf{y} - \mathbf{m})\mathbf{S}^{-1}(\mathbf{y} - \mathbf{m})',$$

respectivamente.

Para ilustrar los desarrollos del apartado siguiente, se tomará como punto de partida esta última representación simplificada del modelo en el espacio de los estados.

### 1.2.2. Filtro de Kalman y algoritmos de alisado

Una vez que un modelo ha sido formulado en el espacio de los estados, su aplicación a series temporales reales exige calcular cantidades tales como las medias y varianzas de las variables  $\mathbf{a}_t / \mathbf{y}_1, \dots, \mathbf{y}_{t-1}$ ,  $\mathbf{a}_t / \mathbf{y}_1, \dots, \mathbf{y}_t$  y  $\mathbf{a}_t / \mathbf{y}_1, \dots, \mathbf{y}_T$ , y ajustar el modelo a los datos obteniendo estimaciones máximo verosímiles de los parámetros desconocidos, tales como las varianzas de los términos de perturbación de las diferentes ecuaciones. En principio, como señalan Durbin y Koopman (2001:10-11), esta tarea puede ser abordada usando los métodos de inferencia clásica con el apoyo de los resultados de la teoría de la probabilidad respecto a la normal multivariante. Sin embargo, como consecuencia de la correlación serial entre las observaciones, la computación se complica notablemente cuando aumenta el número de éstas. En este sentido, las técnicas de filtrado y alisado constituyen algoritmos de computación que proporcionan los mismos resultados de forma más eficiente.

En concreto, el algoritmo conocido como filtro de Kalman permite obtener información sobre el estado  $\mathbf{a}_t$  a partir de la magnitud observada  $\mathbf{y}_t$ , relacionada con éste

desde dos perspectivas diferentes, en función del nivel de información que se pretende alcanzar. La primera tiene por objeto encontrar un estimador insesgado del estado que minimice el error cuadrático medio (Harvey, 1989:104 y ss.). En la segunda, el objetivo se formula en términos más amplios, ya que se pretende obtener la distribución del estado  $\mathbf{a}_t$  condicionada al conjunto de información disponible  $I_t : \{\mathbf{y}_1, \dots, \mathbf{y}_t\}$ , es decir,  $f\left(\frac{\mathbf{a}_t}{I_t}\right)$  (Steyn, 1996:14 y ss.). Los resultados de ambas aproximaciones no son esencialmente diferentes, dado que el estimador insesgado de error cuadrático medio mínimo viene dado, precisamente, por la media de la distribución condicionada obtenida bajo el supuesto de normalidad<sup>74</sup>.

Para obtener la distribución del estado  $\mathbf{a}_t$  condicionada al conjunto de observaciones  $I_t : \{\mathbf{y}_1, \dots, \mathbf{y}_t\}$ , la representación simplificada del modelo en el espacio de los estados expuesta al final del apartado anterior, se puede expresar como:

$$\begin{aligned} \text{Ecuación de transición:} \quad & f\left(\frac{\mathbf{a}_t}{\mathbf{a}_{t-1}}\right) \sim N_m(\mathbf{S}\mathbf{a}_{t-1}, \mathbf{Q}) \\ \text{Ecuación de medida:} \quad & f\left(\frac{\mathbf{y}_t}{\mathbf{a}_t}\right) \sim N_N(\mathbf{Z}_t\mathbf{a}_t, \mathbf{H}) \end{aligned}$$

Dado que los términos de perturbación de las ecuaciones de medida y de transición están serialmente incorrelados e incorrelados entre sí, es decir,

$$\begin{aligned} E[\mathbf{e}_t \mathbf{e}_s'] &= \mathbf{0}, \quad t \neq s; \\ E[\mathbf{u}_t \mathbf{u}_s'] &= \mathbf{0}, \quad t \neq s; \\ E[\mathbf{e}_t \mathbf{u}_s'] &= \mathbf{0}, \quad \forall s, \quad t = 1, \dots, T, \end{aligned}$$

entonces,

$$f\left(\frac{\mathbf{a}_t}{\mathbf{a}_{t-1}, \mathbf{y}_1, \dots, \mathbf{y}_{t-1}}\right) = f\left(\frac{\mathbf{a}_t}{\mathbf{a}_{t-1}}\right),$$

es decir, conocido  $\mathbf{a}_{t-1}$ , la historia pasada de  $\mathbf{y}_t$ ,  $\{\mathbf{y}_1, \dots, \mathbf{y}_{t-1}\}$ , no aporta información adicional sobre  $\mathbf{a}_t$ , y

<sup>74</sup> Téngase en cuenta que en el segundo planteamiento formulado se está adoptando un enfoque bayesiano. Y, una vez obtenida la distribución de  $\mathbf{a}_t$  condicionada a  $I_t$ , la estimación óptima podría ser tanto la media, la moda como la mediana, pero, bajo el supuesto de normalidad, las tres posibilidades coinciden.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$f\left(\mathbf{y}_t / \mathbf{a}_t, \mathbf{y}_1, \dots, \mathbf{y}_{t-1}\right) = f\left(\mathbf{y}_t / \mathbf{a}_t\right),$$

es decir, conocido  $\mathbf{a}_t$ , la historia pasada de  $\mathbf{y}_t$ ,  $\{\mathbf{y}_1, \dots, \mathbf{y}_{t-1}\}$ , no aporta información adicional sobre  $\mathbf{y}_t$ .

El objetivo del filtro es obtener  $f\left(\mathbf{a}_t / I_t\right)$ , pero, para ello, conviene obtener primero  $f\left(\mathbf{a}_t / \mathbf{y}_t\right)$ , lo que exige conocer  $f\left(\mathbf{a}_{t-1} / \mathbf{a}_t\right)$ , que se puede expresar como

$$f\left(\mathbf{a}_{t-1} / \mathbf{a}_t\right) = \frac{f\left(\mathbf{a}_{t-1}, \mathbf{a}_t\right)}{f\left(\mathbf{a}_t\right)} = \frac{f\left(\mathbf{a}_t / \mathbf{a}_{t-1}\right) f\left(\mathbf{a}_{t-1}\right)}{f\left(\mathbf{a}_t\right)}.$$

Sabiendo que

$$\begin{aligned} f\left(\mathbf{a}_t / \mathbf{a}_{t-1}\right) &\sim N_m\left(\mathbf{S}\mathbf{a}_{t-1}, \mathbf{Q}\right) \\ f\left(\mathbf{a}_{t-1}\right) &\sim N_m\left(\mathbf{S}^{t-1}\mathbf{a}_0, \mathbf{P}_{t-1}\right) \end{aligned}$$

y

$$f\left(\mathbf{a}_t\right) \sim N_m\left(\mathbf{S}^t\mathbf{a}_0, \mathbf{P}_t\right),$$

sustituyendo y reagrupando términos, se obtiene que

$$f\left(\mathbf{a}_{t-1} / \mathbf{a}_t\right) \sim N_m\left(\left[\mathbf{P}_{t-1}^{-1} + \mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}\right]^{-1}\left[\mathbf{S}'\mathbf{Q}^{-1}\mathbf{a}_t + \mathbf{P}_{t-1}^{-1}\mathbf{S}^{t-1}\mathbf{a}_0\right], \left[\mathbf{P}_{t-1}^{-1} + \mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}\right]^{-1}\right)$$

(Steyn, 1996:15)<sup>75</sup>.

Además, se sabe que

$$\begin{bmatrix} \mathbf{a}_t \\ \mathbf{y}_t \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{S}^t\mathbf{a}_0 \\ \mathbf{Z}_t\mathbf{S}^t\mathbf{a}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{P}_t & \mathbf{P}_t\mathbf{Z}_t' \\ \mathbf{Z}_t\mathbf{P}_t & \mathbf{Z}_t\mathbf{P}_t\mathbf{Z}_t' + \mathbf{H} \end{bmatrix}\right).$$

<sup>75</sup> Si  $\mathbf{a}_t$  no es estacionario, se puede asumir que los elementos de  $\mathbf{P}_t$  son suficientemente grandes como para despreciar  $\mathbf{P}_t^{-1}$  y, entonces,  $f\left(\mathbf{a}_{t-1} / \mathbf{a}_t\right) \sim N_m\left(\left[\mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}\right]^{-1}\left[\mathbf{S}'\mathbf{Q}^{-1}\mathbf{a}_t\right], \left[\mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}\right]^{-1}\right)$ ; y si, además,  $\mathbf{S}$  es invertible,  $f\left(\mathbf{a}_{t-1} / \mathbf{a}_t\right) \sim N_m\left(\mathbf{S}^{-1}\left[\mathbf{S}'\mathbf{Q}^{-1}\right]^{-1}\left[\mathbf{S}'\mathbf{Q}^{-1}\right]\mathbf{a}_t, \mathbf{S}^{-1}\mathbf{Q}\left[\mathbf{S}^{-1}\right]'\right)$ , es decir,  $f\left(\mathbf{a}_{t-1} / \mathbf{a}_t\right) \sim N_m\left(\mathbf{S}^{-1}\mathbf{a}_t, \mathbf{S}^{-1}\mathbf{Q}\left[\mathbf{S}^{-1}\right]'\right)$ .

Nótese que

$$\begin{aligned} Cov[\mathbf{a}_t, \mathbf{y}_t] &= E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{y}_t - E[\mathbf{y}_t])'\right] \\ &= E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{Z}_t(\mathbf{a}_t - E[\mathbf{a}_t]) + \mathbf{e}_t)'\right] \\ &= E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_t - E[\mathbf{a}_t])' \mathbf{Z}_t'\right] + E[(\mathbf{a}_t - E[\mathbf{a}_t])\mathbf{e}_t'] = \mathbf{P}_t \mathbf{Z}_t' \end{aligned}$$

teniendo en cuenta que  $E[(\mathbf{a}_t - E[\mathbf{a}_t])\mathbf{e}_t'] = \mathbf{0}$ , dado que  $\mathbf{a}_t - E[\mathbf{a}_t]$  es, por sustitución recursiva, función de  $\mathbf{a}_0$  y retardos de  $\mathbf{u}_t$ .

Entonces, y dado que tanto las perturbaciones como el vector de estado inicial están distribuidos normalmente, la distribución de  $\mathbf{a}_t / \mathbf{y}_t$  es también normal, de forma que queda totalmente caracterizada a partir del vector de medias y la matriz de covarianzas. En concreto, y dadas las propiedades de una distribución normal multivariante<sup>76</sup>,  $f\left(\mathbf{a}_t / \mathbf{y}_t\right)$

sigue una distribución normal  $m$ -variante con media

$$E\left[\mathbf{a}_t / \mathbf{y}_t\right] = \mathbf{S}' \mathbf{a}_0 + \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H})^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{S}' \mathbf{a}_0)$$

y varianza

$$Var\left[\mathbf{a}_t / \mathbf{y}_t\right] = [\mathbf{P}_t^{-1} + \mathbf{Z}_t' \mathbf{H}^{-1} \mathbf{Z}_t]^{-1}.$$

De forma más general, el objetivo del filtro de Kalman consiste en encontrar  $f\left(\mathbf{a}_t / \mathbf{I}_k\right)$ , donde  $\mathbf{I}_k : \{\mathbf{y}_1, \dots, \mathbf{y}_k\}$ . Dependiendo del conjunto de información,  $\mathbf{I}_k$ , empleado en la obtención de la distribución condicionada del estado se pueden distinguir tres casos. Si  $k < t$ , se trata de un problema de predicción; si  $k = t$ , se habla de un problema de filtrado; y si  $k = T$  (o, en general, si  $k > t$ ), se habla de un problema de alisado. En otras palabras, la solución al problema de predicción ofrecerá estimaciones que son extrapolaciones hacia el futuro de los componentes del vector de estado; por otro lado, el filtrado permitirá actualizar las estimaciones del estado no observable a medida que se disponga de una nueva observación de la magnitud observable relacionada con éste; y,

<sup>76</sup> Véase anexo 1.3.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

finalmente, el alisado ofrece la mejor estimación del estado en cualquier instante del tiempo con toda la información de la muestra. A continuación se obtienen las distribuciones del estado  $\mathbf{a}$ , condicionadas a  $I_k$  en los diferentes casos.

### - Predicción un periodo hacia delante y filtrado <sup>77</sup>

La predicción un periodo hacia delante consiste en obtener  $f\left(\frac{\mathbf{a}_t}{I_{t-1}}\right)$ . Dicha función de densidad se puede expresar como

$$\begin{aligned} f\left(\frac{\mathbf{a}_t}{I_{t-1}}\right) &= \int_{-\infty}^{\infty} f\left(\frac{\mathbf{a}_t, \mathbf{a}_{t-1}}{I_{t-1}}\right) d\mathbf{a}_{t-1} \\ &= \int_{-\infty}^{\infty} f\left(\frac{\mathbf{a}_t}{\mathbf{a}_{t-1}, I_{t-1}}\right) f\left(\frac{\mathbf{a}_{t-1}}{I_{t-1}}\right) d\mathbf{a}_{t-1} \\ &= \int_{-\infty}^{\infty} f\left(\frac{\mathbf{a}_t}{\mathbf{a}_{t-1}}\right) f\left(\frac{\mathbf{a}_{t-1}}{I_{t-1}}\right) d\mathbf{a}_{t-1} \end{aligned}$$

Por tanto, la función de densidad del problema de predicción un periodo hacia delante puede obtenerse si se conoce  $f\left(\frac{\mathbf{a}_t}{\mathbf{a}_{t-1}}\right)$ , que viene dada por la ecuación de transición del modelo, y  $f\left(\frac{\mathbf{a}_{t-1}}{I_{t-1}}\right)$ , que es la solución al problema del filtro en el periodo anterior.

El problema de filtrado consiste en obtener  $f\left(\frac{\mathbf{a}_t}{I_t}\right)$ . Si dicha función de densidad se expresa como

$$\begin{aligned} f\left(\frac{\mathbf{a}_t}{I_t}\right) &= f\left(\frac{\mathbf{a}_t}{\mathbf{y}_t, I_{t-1}}\right) = \frac{f\left(\frac{\mathbf{a}_t, \mathbf{y}_t}{I_{t-1}}\right)}{f\left(\frac{\mathbf{y}_t}{I_{t-1}}\right)} \\ &= \frac{f\left(\frac{\mathbf{y}_t}{\mathbf{a}_t, I_{t-1}}\right) f\left(\frac{\mathbf{a}_t}{I_{t-1}}\right)}{f\left(\frac{\mathbf{y}_t}{I_{t-1}}\right)} = \frac{f\left(\frac{\mathbf{y}_t}{\mathbf{a}_t}\right) f\left(\frac{\mathbf{a}_t}{I_{t-1}}\right)}{f\left(\frac{\mathbf{y}_t}{I_{t-1}}\right)} \end{aligned}$$

<sup>77</sup> La predicción y el filtrado se consideran simultáneamente porque los procesos recursivos de ambas derivaciones están relacionados.

la solución al problema requiere conocer  $f\left(\mathbf{y}_t/\mathbf{a}_t\right)$ , que viene dada por la ecuación de medida del modelo,  $f\left(\mathbf{a}_t/I_{t-1}\right)$ , que se obtiene en la fase de predicción, y  $f\left(\mathbf{y}_t/I_{t-1}\right)$ , que es un factor de normalización que se obtiene como se indica más adelante.

Para derivar las distribuciones del problema de predicción y de filtrado, se definen

$$E\left[\mathbf{a}_t/I_k\right] = \mathbf{a}_{t/k},$$

con  $\mathbf{a}_{t/t} \equiv \mathbf{a}_t$ , y

$$E\left[(\mathbf{a}_t - \mathbf{a}_{t/k})(\mathbf{a}_t - \mathbf{a}_{t/k})' | I_k\right] = \mathbf{P}_{t/k},$$

con  $\mathbf{P}_{t/t} \equiv \mathbf{P}_t$ , de forma que, conocidos  $\mathbf{a}_{t-1}$  y  $\mathbf{P}_{t-1}$ , se tiene que

$$f\left(\mathbf{a}_{t-1}/I_{t-1}\right) \sim N_m(\mathbf{a}_{t-1}, \mathbf{P}_{t-1}).$$

A partir de la ecuación de transición del modelo, se sabe que

$$f\left(\mathbf{a}_t/\mathbf{a}_{t-1}\right) \sim N_m(\mathbf{S}\mathbf{a}_{t-1}, \mathbf{Q}).$$

Además,

$$f\left(\mathbf{a}_t/I_{t-1}\right) = \int_{-\infty}^{\infty} f\left(\mathbf{a}_t/\mathbf{a}_{t-1}\right) f\left(\mathbf{a}_{t-1}/I_{t-1}\right) d\mathbf{a}_{t-1} \sim N_m(\mathbf{S}\mathbf{a}_{t-1}, \mathbf{S}\mathbf{P}_{t-1}\mathbf{S}' + \mathbf{Q}),$$

es decir,

$$f\left(\mathbf{a}_t/I_{t-1}\right) \sim N_m(\mathbf{S}\mathbf{a}_{t-1}, \mathbf{S}\mathbf{P}_{t-1}\mathbf{S}' + \mathbf{Q})^{78}.$$

Con la notación antes señalada, y dada la normalidad de la distribución, este resultado se puede expresar indicando que la media es  $\mathbf{a}_{t/t-1} = \mathbf{S}\mathbf{a}_{t-1}$  y la varianza  $\mathbf{P}_{t/t-1} = \mathbf{S}\mathbf{P}_{t-1}\mathbf{S}' + \mathbf{Q}$ . Nótese que, dada la ecuación de transición, se tiene que

$$E\left[\mathbf{a}_t/I_{t-1}\right] = \mathbf{S} E\left[\mathbf{a}_{t-1}/I_{t-1}\right] = \mathbf{S}\mathbf{a}_{t-1},$$

---

<sup>78</sup> Si  $f(x) \sim N(\mathbf{m}, \mathbf{S})$ ,  $f(y/x) \sim N(\mathbf{A}x, \mathbf{W})$  y  $\mathbf{A}$  es invertible, entonces se tiene que  $\int_{-\infty}^{\infty} f(y/x)f(x)dx \sim N(\mathbf{A}\mathbf{m}, \mathbf{A}\mathbf{S}\mathbf{A}' + \mathbf{W})$ .

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

y, dado que  $Cov[\mathbf{a}_{t-1}, \mathbf{u}_t] = \mathbf{0}$ , resulta que

$$Var\left[\frac{\mathbf{a}_t}{\mathbf{I}_{t-1}}\right] = Var\left[\frac{\mathbf{S}\mathbf{a}_{t-1}}{\mathbf{I}_{t-1}}\right] + Var[\mathbf{u}_t] = \mathbf{S}\mathbf{P}_{t-1}\mathbf{S}' + \mathbf{Q}.$$

Respecto al problema de filtrado, se necesita obtener la función de densidad condicionada  $f\left(\frac{\mathbf{y}_t}{\mathbf{I}_{t-1}}\right)$ . Teniendo en cuenta que

$$f\left(\frac{\mathbf{y}_t}{\mathbf{I}_{t-1}}\right) = \int_{-\infty}^{\infty} f\left(\frac{\mathbf{y}_t}{\mathbf{a}_t}, \mathbf{I}_{t-1}\right) f\left(\frac{\mathbf{a}_t}{\mathbf{I}_{t-1}}\right) d\mathbf{a}_t = \int_{-\infty}^{\infty} f\left(\frac{\mathbf{y}_t}{\mathbf{a}_t}\right) f\left(\frac{\mathbf{a}_t}{\mathbf{I}_{t-1}}\right) d\mathbf{a}_t,$$

se obtiene que

$$f\left(\frac{\mathbf{y}_t}{\mathbf{I}_{t-1}}\right) \sim N_N(\mathbf{Z}_t\mathbf{S}\mathbf{a}_{t-1}, \mathbf{Z}_t\mathbf{P}_{t/t-1}\mathbf{Z}_t' + \mathbf{H}),$$

es decir,

$$f\left(\frac{\mathbf{y}_t}{\mathbf{I}_{t-1}}\right) \sim N_N(\mathbf{Z}_t\mathbf{S}\mathbf{a}_{t-1}, \mathbf{F}_t),$$

donde  $\mathbf{F}_t = \mathbf{Z}_t\mathbf{P}_{t/t-1}\mathbf{Z}_t' + \mathbf{H}$  <sup>79</sup>.

Antes de obtener la solución al problema del filtrado, conviene expresar esta función de densidad a partir de los denominados errores de predicción, definidos como  $\mathbf{e}_t = \mathbf{y}_t - E[\mathbf{y}_t | \mathbf{I}_{t-1}] = \mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_{t/t-1}$ , cuya media es nula y cuya varianza es, precisamente, la varianza de  $\frac{\mathbf{y}_t}{\mathbf{I}_{t-1}}$ , es decir,

$$E\left[\left(\mathbf{y}_t - E\left[\frac{\mathbf{y}_t}{\mathbf{I}_{t-1}}\right]\right)\left(\mathbf{y}_t - E\left[\frac{\mathbf{y}_t}{\mathbf{I}_{t-1}}\right]\right)' / \mathbf{I}_{t-1}\right] = \mathbf{F}_t.$$

Finalmente, se puede obtener la función de densidad del problema del filtro

$$f\left(\frac{\mathbf{a}_t}{\mathbf{I}_t}\right) = \frac{f\left(\frac{\mathbf{y}_t}{\mathbf{a}_t}\right) f\left(\frac{\mathbf{a}_t}{\mathbf{I}_{t-1}}\right)}{f\left(\frac{\mathbf{y}_t}{\mathbf{I}_{t-1}}\right)}$$

y, de hecho, se demuestra que

---

<sup>79</sup> Véase la nota anterior.

$$f\left(\mathbf{a}_t / \mathbf{I}_t\right) \sim N_m\left(\mathbf{a}_{t/t-1} + \mathbf{K}_t \mathbf{e}_t, \mathbf{P}_{t/t-1} - \mathbf{K}_t \mathbf{F}_t \mathbf{K}_t'\right),$$

donde  $\mathbf{K}_t = \mathbf{P}_{t/t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1}$  es la llamada ganancia de Kalman.

Con la notación antes señalada, y dada la normalidad de la distribución, este resultado se puede expresar indicando que la media es  $\mathbf{a}_t = \mathbf{a}_{t/t-1} + \mathbf{K}_t \mathbf{e}_t$  y la varianza

$\mathbf{P}_t = \mathbf{P}_{t/t-1} - \mathbf{K}_t \mathbf{F}_t \mathbf{K}_t'$ . Nótese que la distribución de  $\mathbf{a}_t / \mathbf{I}_t$  es también la distribución de

$\mathbf{a}_t / \mathbf{I}_{t-1}, \mathbf{e}_t$ , puesto que, dados  $\mathbf{I}_{t-1}$  y  $\mathbf{e}_t$ ,  $\mathbf{I}_t$  queda determinado. Por tanto, dadas las propiedades de la normal multivariante<sup>80</sup>, resulta que

$$E\left[\mathbf{a}_t / \mathbf{I}_t\right] = E\left[\mathbf{a}_t / \mathbf{I}_{t-1}, \mathbf{e}_t\right] = E\left[\mathbf{a}_t / \mathbf{I}_{t-1}\right] + \text{Cov}\left[\mathbf{a}_t, \mathbf{e}_t / \mathbf{I}_{t-1}\right] \left[\text{Var}[\mathbf{e}_t]\right]^{-1} \mathbf{e}_t$$

y

$$\text{Var}\left[\mathbf{a}_t / \mathbf{I}_t\right] = \text{Var}\left[\mathbf{a}_t / \mathbf{I}_{t-1}, \mathbf{e}_t\right] = \text{Var}\left[\mathbf{a}_t / \mathbf{I}_{t-1}\right] + \text{Cov}\left[\mathbf{a}_t, \mathbf{e}_t / \mathbf{I}_{t-1}\right] \left[\text{Var}[\mathbf{e}_t]\right]^{-1} \text{Cov}\left[\mathbf{a}_t, \mathbf{e}_t / \mathbf{I}_{t-1}\right]'$$

Y dado que

$$\text{Cov}\left[\mathbf{a}_{t-1}, \mathbf{e}_t / \mathbf{I}_{t-1}\right] = \text{Cov}\left[\mathbf{a}_{t-1}, \mathbf{y}_t / \mathbf{I}_{t-1}\right] = \mathbf{P}_{t/t-1} \mathbf{Z}_t',$$

se tiene que

$$E\left[\mathbf{a}_t / \mathbf{I}_t\right] = \mathbf{a}_{t/t-1} + \mathbf{P}_{t/t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{e}_t$$

y

$$\text{Var}\left[\mathbf{a}_t / \mathbf{I}_t\right] = \mathbf{P}_{t/t-1} - \mathbf{P}_{t/t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}_{t/t-1}.$$

Resumiendo, el problema de predicción un periodo hacia delante<sup>81</sup> y el problema de filtrado se resuelven recursivamente, bajo el supuesto de normalidad, a través de las ecuaciones de predicción del estado —que permiten proyectar hacia el futuro el estado actual y estimar la varianza de la predicción— y las ecuaciones de actualización —que

<sup>80</sup> Véase Durbin y Koopman (2001:37).

<sup>81</sup> La predicción  $m$  periodos hacia delante es relativamente trivial una vez solucionado el problema del filtro. Véase, por ejemplo, Steyn (1996:19).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

permiten incorporar una nueva observación a la estimación del estado y obtener una estimación mejorada del estado<sup>82</sup>.

Ecuaciones de predicción	Predicción del estado	$\mathbf{a}_{t/t-1} = \mathbf{S}\mathbf{a}_{t-1}$
	Varianza de la predicción del estado	$\mathbf{P}_{t/t-1} = \mathbf{S}\mathbf{P}_{t-1}\mathbf{S}' + \mathbf{Q}$
Ecuaciones de actualización	Estimación filtrada del estado	$\mathbf{a}_t = \mathbf{a}_{t/t-1} + \mathbf{K}_t\mathbf{e}_t$
	Varianza de la estimación filtrada del estado	$\mathbf{P}_t = \mathbf{P}_{t/t-1} - \mathbf{K}_t\mathbf{F}_t\mathbf{K}_t'$

Es decir, el algoritmo se ejecuta en dos fases. En una primera fase se hace una predicción del estado en  $t$  a partir de la estimación filtrada del estado en  $t-1$  mediante la ecuación  $\mathbf{a}_{t/t-1} = \mathbf{S}\mathbf{a}_{t-1}$  y la varianza del error de predicción viene dada por  $\mathbf{P}_{t/t-1} = \mathbf{S}\mathbf{P}_{t-1}\mathbf{S}' + \mathbf{Q}$ . En la segunda fase, se actualiza la predicción incorporando la observación  $\mathbf{y}_t$  y se obtiene la estimación filtrada del estado, es decir,  $\mathbf{a}_t = \mathbf{a}_{t/t-1} + \mathbf{K}_t\mathbf{e}_t$ , donde  $\mathbf{e}_t = \mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_{t/t-1}$ , que es una combinación lineal de la predicción del estado y de la diferencia observada entre la medida registrada en el instante  $t$ ,  $\mathbf{y}_t$ , y la predicción realizada para dicha medida con la estimación  $\mathbf{a}_{t/t-1}$ . El factor que pondera dicha diferencia,  $\mathbf{K}_t$ , es la ya definida ganancia de Kalman.

### - Alisado.

El problema de alisado de intervalo fijo<sup>83</sup> consiste en obtener la función de densidad condicionada  $f\left(\frac{\mathbf{a}_t}{I_T}\right)$ , siendo  $T$  el último instante del tiempo en el que se observa la serie. Esta función de densidad se obtiene también recursivamente. Para ello, es preciso

<sup>82</sup> Como señalan Welch y Bishop (2001:4-5), las primeras pueden interpretarse como ecuaciones de predicción y las segundas de corrección. Cuando las secuencias  $\{\mathbf{P}_{t/t-1}\}_{t=1, \dots, T}$  y  $\{\mathbf{K}_t\}_{t=1, \dots, T}$  convergen a matrices fijas, el sistema alcanza un estado estacionario (*steady state*).

<sup>83</sup> Para obtener las distribuciones alisadas se puede recurrir a diferentes algoritmos (Anderson y Moore, 1979:cap.7 y Harvey, 1989:149-155): a) alisado de punto fijo, que obtiene  $\mathbf{a}_{\tau/T}$  para determinados instantes del tiempo  $\tau$ ; b) alisado de retardo fijo, que obtiene  $\mathbf{a}_{t-j/t}$ ,  $j = 1, \dots, M$ , siendo  $M$  el máximo retardo especificado (Anderson y Moore 1979:cap.7); c) alisado de intervalo fijo, que obtiene  $\mathbf{a}_{t/T}$ ,  $t = 1, \dots, T$ . Este último algoritmo es el más utilizado (veáanse Anderson y Moore, 1979:cap.7, Ansley y Kohn, 1982, Kohn y Ansley, 1983, y Koopman, 1996).

relacionar  $f\left(\mathbf{a}_t/I_T\right)$  con  $f\left(\mathbf{a}_{t+1}/I_T\right)$  y, en este sentido, la función de densidad conjunta viene dada por

$$f\left(\mathbf{a}_t, \mathbf{a}_{t+1}/I_T\right) = f\left(\mathbf{a}_{t+1}/I_T\right) f\left(\mathbf{a}_t/\mathbf{a}_{t+1}, I_T\right) = f\left(\mathbf{a}_{t+1}/I_T\right) f\left(\mathbf{a}_t/\mathbf{a}_{t+1}, I_t\right).$$

Nótese que, conocido  $\mathbf{a}_{t+1}$ , la información  $\{\mathbf{y}_{t+1}, \dots, \mathbf{y}_T\}$  es irrelevante para la distribución de  $\mathbf{a}_t$ . Además, a partir de la expresión de  $f\left(\mathbf{a}_t, \mathbf{a}_{t+1}/I_T\right)$  en  $T = t$ , se tiene que

$$f\left(\mathbf{a}_t, \mathbf{a}_{t+1}/I_t\right) = f\left(\mathbf{a}_{t+1}/I_t\right) f\left(\mathbf{a}_t/\mathbf{a}_{t+1}, I_t\right)$$

y, entonces,

$$f\left(\mathbf{a}_t/\mathbf{a}_{t+1}, I_t\right) = \frac{f\left(\mathbf{a}_t, \mathbf{a}_{t+1}/I_t\right)}{f\left(\mathbf{a}_{t+1}/I_t\right)}.$$

Por tanto,  $f\left(\mathbf{a}_t, \mathbf{a}_{t+1}/I_T\right)$  se puede expresar como

$$\begin{aligned} f\left(\mathbf{a}_t, \mathbf{a}_{t+1}/I_T\right) &= f\left(\mathbf{a}_{t+1}/I_T\right) \frac{f\left(\mathbf{a}_t, \mathbf{a}_{t+1}/I_t\right)}{f\left(\mathbf{a}_{t+1}/I_t\right)} = f\left(\mathbf{a}_{t+1}/I_T\right) \frac{f\left(\mathbf{a}_{t+1}/\mathbf{a}_t, I_t\right) f\left(\mathbf{a}_t/I_t\right)}{f\left(\mathbf{a}_{t+1}/I_t\right)} \\ &= f\left(\mathbf{a}_{t+1}/I_T\right) \frac{f\left(\mathbf{a}_{t+1}/\mathbf{a}_t\right) f\left(\mathbf{a}_t/I_t\right)}{f\left(\mathbf{a}_{t+1}/I_t\right)}. \end{aligned}$$

Integrando la función de densidad conjunta con respecto a  $\mathbf{a}_{t+1}$  se obtiene la función de densidad deseada. Es decir,

$$f\left(\mathbf{a}_t/I_T\right) = \int_{-\infty}^{\infty} f\left(\mathbf{a}_t, \mathbf{a}_{t+1}/I_T\right) d\mathbf{a}_{t+1} = f\left(\mathbf{a}_t/I_T\right) \int_{-\infty}^{\infty} f\left(\mathbf{a}_{t+1}/I_T\right) \frac{f\left(\mathbf{a}_{t+1}/\mathbf{a}_t\right)}{f\left(\mathbf{a}_{t+1}/I_t\right)} d\mathbf{a}_{t+1}.$$

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

A partir de esta expresión, se obtiene  $f\left(\mathbf{a}_t/I_T\right)$ , una vez determinado  $f\left(\mathbf{a}_{t+1}/I_T\right)$ . Dado que  $f\left(\mathbf{a}_T/I_T\right)$  es la última función de densidad obtenida en el problema del filtro, la función de densidad  $f\left(\mathbf{a}_{T-1}/I_T\right)$  se puede obtener a partir de la anterior. De este modo, se obtienen recursivamente las funciones de densidad alisadas  $f\left(\mathbf{a}_t/I_T\right)$  para cada uno de los instantes del tiempo  $t = 1, \dots, T$ .

Si se asumen funciones de densidad normales,  $\mathbf{a}_{t/T} = E\left[\mathbf{a}_t/I_T\right]$  es también la moda de la distribución de  $\mathbf{a}_t/I_T$ , de modo que las  $\mathbf{a}_{1/T}, \dots, \mathbf{a}_{T/T}$  pueden obtenerse como los valores de  $\mathbf{a}_1, \dots, \mathbf{a}_T$  que maximizan  $\log f\left(\mathbf{a}_1, \dots, \mathbf{a}_T | I_T\right)$ , es decir,  $\log f\left(\mathbf{a}_1, \dots, \mathbf{a}_T / \mathbf{y}_1, \dots, \mathbf{y}_T\right)$  o, en forma compacta,  $\log f\left(\mathbf{a} / \mathbf{y}\right)$ . Teniendo en cuenta que

$$\frac{\partial \log f\left(\mathbf{a} / \mathbf{y}\right)}{\partial \mathbf{a}} = \frac{\partial [\log f(\mathbf{a}, \mathbf{y}) - \log f(\mathbf{y})]}{\partial \mathbf{a}} = \frac{\partial \log f(\mathbf{a}, \mathbf{y})}{\partial \mathbf{a}},$$

las estimaciones alisadas del estado pueden encontrarse como solución a las ecuaciones

$$\frac{\partial \log f(\mathbf{a}_1, \dots, \mathbf{a}_T, \mathbf{y}_1, \dots, \mathbf{y}_T)}{\partial \mathbf{a}_t} = \mathbf{0}, \quad t = 1, \dots, T^{84}.$$

Nótese que, dadas las propiedades de la normal multivariante<sup>85</sup>, y teniendo en cuenta que  $I_T$  está determinado si se conoce  $I_t, \mathbf{e}_{t+1}, \dots, \mathbf{e}_T$ , así como el hecho de que  $\mathbf{e}_{t+1}, \dots, \mathbf{e}_T$  son independientes de  $I_t$  y, a su vez, independientes entre sí con medias nulas, resulta que

$$\mathbf{a}_{t/T} \equiv E\left[\mathbf{a}_t/I_T\right] = E\left[\mathbf{a}_t/I_t, \mathbf{e}_{t+1}, \dots, \mathbf{e}_T\right] = \mathbf{a}_t + \sum_{j=t+1}^T \text{Cov}\left[\mathbf{a}_t, \mathbf{e}_j / I_t\right] \mathbf{F}_j^{-1} \mathbf{e}_j$$

y

<sup>84</sup> Véase Harvey y Shephard (1993).

$$\mathbf{P}_{t/T} \equiv \text{Var}\left[\mathbf{a}_t / \mathbf{I}_T\right] = \text{Var}\left[\mathbf{a}_t / \mathbf{I}_t, \mathbf{e}_{t+1}, \dots, \mathbf{e}_T\right] = \mathbf{P}_t - \sum_{j=t+1}^T \text{Cov}\left[\mathbf{a}_t, \mathbf{e}_j / \mathbf{I}_t\right] \mathbf{F}_j^{-1} \text{Cov}\left[\mathbf{a}_t, \mathbf{e}_j / \mathbf{I}_t\right]'$$

Estas expresiones pueden calcularse recursivamente a través de diferentes algoritmos<sup>86</sup>. Nótese que  $\mathbf{P}_{t/T} \leq \mathbf{P}_t$ , dado que se está utilizando más información, y la diferencia entre  $\mathbf{P}_t$  y  $\mathbf{P}_{t/T}$  mide la ganancia obtenida con el alisado.

Anderson y Moore (1979) propusieron su ya clásico algoritmo en el que la media y la varianza de  $\mathbf{a}_t / \mathbf{I}_T$  se derivan recursivamente a través de las ecuaciones

$$\mathbf{a}_{t/T} = \mathbf{a}_t + \mathbf{P}_t^* [\mathbf{a}_{t+1/T} - \mathbf{S}\mathbf{a}_t]$$

y

$$\mathbf{P}_{t/T} = \mathbf{P}_t + \mathbf{P}_t^* (\mathbf{P}_{t+1/T} - \mathbf{P}_{t+1/t}) \mathbf{P}_t^*$$

donde

$$\mathbf{P}_t^* = \mathbf{P}_t \mathbf{S}' \mathbf{P}_{t+1/t}^{-1}$$

para  $t = T-1, \dots, 1$ , respectivamente (Steyn, 1996:20)<sup>87</sup>. Estas expresiones pueden obtenerse

si se calcula primero  $E\left[\mathbf{a}_t / \mathbf{a}_{t+1}, \mathbf{I}_t\right]$ . Pues bien, se tiene que

$$E\left[\mathbf{a}_t / \mathbf{a}_{t+1}, \mathbf{I}_t\right] = E\left[\mathbf{a}_t / \mathbf{I}_t\right] + \text{Cov}\left[\mathbf{a}_t, \mathbf{a}_{t+1} / \mathbf{I}_t\right] \text{Var}\left[\mathbf{a}_{t+1} / \mathbf{I}_t\right]^{-1} \left[\mathbf{a}_{t+1} - E\left[\mathbf{a}_{t+1} / \mathbf{I}_t\right]\right]$$

Y, dado que

$$\begin{aligned} \text{Cov}\left[\mathbf{a}_t, \mathbf{a}_{t+1} / \mathbf{I}_t\right] &= E\left[\left(\mathbf{a}_t - E\left[\mathbf{a}_t / \mathbf{I}_t\right]\right)\left(\mathbf{a}_{t+1} - E\left[\mathbf{a}_{t+1} / \mathbf{I}_t\right]\right)'\right] \\ &= E\left[\left(\mathbf{a}_t - \mathbf{a}_t\right)\left(\mathbf{a}_t - \mathbf{a}_t\right)' \mathbf{S}' / \mathbf{I}_t\right] = \mathbf{P}_t \mathbf{S}' \end{aligned}$$

resulta que

<sup>85</sup> Véase Durbin y Koopman (2001:37).

<sup>86</sup> Véase Durbin y Koopman (2001:70-73).

<sup>87</sup> El algoritmo propuesto por Anderson y Moore (1979) es poco eficiente en términos de tiempo de cálculo y memoria. Un procedimiento más eficiente ha sido propuesto por Koopman (1993).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$\begin{aligned} E\left[\frac{\mathbf{a}_t}{\mathbf{a}_{t+1}}, I_t\right] &= \mathbf{a}_t + \mathbf{P}_t \mathbf{S}' \mathbf{P}_{t+1}^{-1} [\mathbf{a}_{t+1} - \mathbf{a}_{t+1/t}] \\ &= \mathbf{a}_t + \mathbf{P}_t \mathbf{S}' \mathbf{P}_{t+1}^{-1} [\mathbf{a}_{t+1} - \mathbf{S} \mathbf{a}_t] = \mathbf{a}_t + \mathbf{P}_t^* [\mathbf{a}_{t+1} - \mathbf{S} \mathbf{a}_t] \end{aligned}$$

Si ahora se advierte que  $E\left[\frac{\mathbf{a}_t}{\mathbf{a}_{t+1}}, I_T\right] = E\left[\frac{\mathbf{a}_t}{\mathbf{a}_{t+1}}, I_t\right]$ <sup>88</sup>, entonces, por la ley de las proyecciones iteradas<sup>89</sup>, se tiene que

$$E\left[\frac{\mathbf{a}_t}{I_T}\right] = E\left[E\left[\frac{\mathbf{a}_t}{\mathbf{a}_{t+1}}, I_T\right] / I_T\right] = \mathbf{a}_t + \mathbf{P}_t^* \left[E\left[\frac{\mathbf{a}_{t+1}}{I_T}\right] - \mathbf{S} \mathbf{a}_t\right] = \mathbf{a}_t + \mathbf{P}_t^* [\mathbf{a}_{t+1/T} - \mathbf{S} \mathbf{a}_t].$$

Para obtener

$$\mathbf{P}_{t/T} \equiv E\left[(\mathbf{a}_t - \mathbf{a}_{t/T})(\mathbf{a}_t - \mathbf{a}_{t/T})' / I_T\right],$$

basta con considerar que

$$\mathbf{a}_t - \mathbf{a}_{t/T} = \mathbf{a}_t - \mathbf{a}_t - \mathbf{P}_t^* [\mathbf{a}_{t+1/T} - \mathbf{a}_{t+1/t}].$$

Entonces,

$$\mathbf{a}_t - \mathbf{a}_{t/T} = \mathbf{a}_t - \mathbf{a}_t - \mathbf{P}_t^* \mathbf{a}_{t+1/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/t},$$

o bien,

$$\mathbf{a}_t - \mathbf{a}_{t/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/T} = \mathbf{a}_t - \mathbf{a}_t + \mathbf{P}_t^* \mathbf{a}_{t+1/t},$$

de modo que

$$(\mathbf{a}_t - \mathbf{a}_{t/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/T})(\mathbf{a}_t - \mathbf{a}_{t/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/T})' = (\mathbf{a}_t - \mathbf{a}_t + \mathbf{P}_t^* \mathbf{a}_{t+1/t})(\mathbf{a}_t - \mathbf{a}_t + \mathbf{P}_t^* \mathbf{a}_{t+1/t})',$$

y tomando esperanzas, resulta que

$$\begin{aligned} &E\left[(\mathbf{a}_t - \mathbf{a}_{t/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/T})(\mathbf{a}_t - \mathbf{a}_{t/T} + \mathbf{P}_t^* \mathbf{a}_{t+1/T})'\right] \\ &= E\left[(\mathbf{a}_t - \mathbf{a}_{t/T})(\mathbf{a}_t - \mathbf{a}_{t/T})'\right] + \mathbf{P}_t^* E[\mathbf{a}_{t+1/T} \mathbf{a}_{t+1/T}'] \mathbf{P}_t^{*'} \end{aligned}$$

y

$$E\left[(\mathbf{a}_t - \mathbf{a}_t + \mathbf{P}_t^* \mathbf{a}_{t+1/t})(\mathbf{a}_t - \mathbf{a}_t + \mathbf{P}_t^* \mathbf{a}_{t+1/t})'\right] = E\left[(\mathbf{a}_t - \mathbf{a}_t)(\mathbf{a}_t - \mathbf{a}_t)'\right] + \mathbf{P}_t^* E[\mathbf{a}_{t+1/t} \mathbf{a}_{t+1/t}'] \mathbf{P}_t^{*'} \quad ^{90}$$

<sup>88</sup> Nótese que, conocido  $\alpha_{t+1}$ , la información  $\{y_{t+1}, \dots, y_T\}$  es irrelevante para la distribución de  $\alpha_t$ .

<sup>89</sup> Véase Hamilton (1994:396).

Se obtiene que

$$\mathbf{P}_{t/T} + \mathbf{P}_t^* E[\mathbf{a}_{t+1/T} \mathbf{a}'_{t+1/T}] \mathbf{P}_t^{*\prime} = \mathbf{P}_t + \mathbf{P}_t^* E[\mathbf{a}_{t+1/t} \mathbf{a}'_{t+1/t}] \mathbf{P}_t^{*\prime},$$

de modo que

$$\begin{aligned} \mathbf{P}_{t/T} &= \mathbf{P}_t - \mathbf{P}_t^* E[\mathbf{a}_{t+1/T} \mathbf{a}'_{t+1/T}] \mathbf{P}_t^{*\prime} + \mathbf{P}_t^* E[\mathbf{a}_{t+1/t} \mathbf{a}'_{t+1/t}] \mathbf{P}_t^{*\prime} \\ &= \mathbf{P}_t + \mathbf{P}_t^* [-E[\mathbf{a}_{t+1/T} \mathbf{a}'_{t+1/T}] + E[\mathbf{a}_{t+1/t} \mathbf{a}'_{t+1/t}]] \mathbf{P}_t^{*\prime} \end{aligned}$$

y sumando y restando  $E[\mathbf{a}_{t+1} \mathbf{a}'_{t+1}]$ , se obtiene que

$$\mathbf{P}_{t/T} = \mathbf{P}_t + \mathbf{P}_t^* [\mathbf{P}_{t+1/T} - \mathbf{P}_{t+1/t}] \mathbf{P}_t^{*\prime}.$$

Bajo el supuesto de gaussianidad, se puede demostrar que las estimaciones alisadas del estado coinciden con los valores de una función *spline*<sup>91</sup>. Así, si deseamos aproximar una función alisada  $\mu_t$  a una serie  $y_1, \dots, y_T$  de valores observados en instantes equidistantes en el tiempo, una aproximación estándar consiste en elegir  $\mu_t$  tal que minimice la expresión

$$\sum_{t=1}^T (y_t - \mu_t)^2 + \lambda \sum_{t=1}^T (\Delta^2 \mu_t)^2,$$

para un valor dado de  $\lambda$  ( $\lambda > 0$ )<sup>92</sup>. La función  $\mu_t$  se conoce como función *spline*<sup>93</sup>. Este problema puede formularse en el marco de los modelos en el espacio de los estados mediante las ecuaciones

$$y_t = \mu_t + \varepsilon_t$$

y

$$\Delta^2 \mu_t = \xi_t,$$

donde  $Var[\varepsilon_t] = \sigma^2$  y  $Var[\xi_t] = \lambda^{-1} \sigma^2$ ,  $\lambda > 0$ . Bajo el supuesto de normalidad, el logaritmo natural de la función de densidad de  $\mathbf{a}_1, \dots, \mathbf{a}_T$ ,  $y_1, \dots, y_T$  es, al margen de las constantes irrelevantes, igual a

<sup>90</sup> Nótese que  $E[(\mathbf{a}_t - \mathbf{a}_{t/T}) \mathbf{a}'_{t+1/T}] = E[\mathbf{a}_t \mathbf{a}'_{t+1/T}] - E[\mathbf{a}_{t/T} \mathbf{a}'_{t+1/T}] = \mathbf{0}$  y, del mismo modo,  $E[(\mathbf{a}_t - \mathbf{a}_t) \mathbf{a}'_{t+1/t}] = E[\mathbf{a}_t \mathbf{a}'_{t+1/t}] - E[\mathbf{a}_t \mathbf{a}'_{t+1/t}] = \mathbf{0}$ .

<sup>91</sup> Véase Durbin y Koopman (2001:61-62).

<sup>92</sup> Cuanto más pequeño sea el valor de  $\lambda$ , más cercanos estarán los valores de  $\mu_t$  a los de  $y_t$ , pero menos alisada será  $\mu_t$ .

<sup>93</sup> Este planteamiento es similar al expuesto en el capítulo segundo, cuando un tipo particular de funciones *splines* se utiliza para recoger el componente estacional de una serie.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$-\frac{\lambda}{2\sigma^2} \sum_{t=1}^T (\Delta^2 \mu_t)^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - \mu_t)^2,$$

y maximizar esta función es equivalente a minimizar la función señalada anteriormente, ocurriendo además que, como se comentó en los párrafos precedentes, los valores  $\mathbf{a}_1, \dots, \mathbf{a}_T$  que minimizan la función de verosimilitud de  $\mathbf{a}_1, \dots, \mathbf{a}_T, y_1, \dots, y_T$  son, precisamente, las estimaciones alisadas del estado.

Una vez obtenidas las medias condicionales  $E\left[\frac{\mathbf{a}_t}{\mathbf{I}_k}\right], k = t-1, t = 1, \dots, T$ , resulta que dichas medias son sumas ponderadas de observaciones pasadas (predicción), observaciones pasadas y presentes (filtrado) o de todas las observaciones (alisado). El estudio de tales ponderaciones ayuda a comprender mejor las propiedades de los diferentes estimadores. Por ejemplo, la estimación alisada del componente de nivel  $m_{t/T}$  en torno a  $t = T/2, m_{\left(\frac{T}{2}\right)/T}$ , debería ser una media ponderada de todas las observaciones,

$$m_{\left(\frac{T}{2}\right)/T} = \sum_{t=1}^T w_t y_t,$$

tal que las ponderaciones,  $w_t$ , decrezcan simétrica y exponencialmente a ambos lados de  $t = T/2$ <sup>94</sup>. Aquellos modelos en los que los patrones de pesos difieran de lo que parece apropiado deberían ser investigados, como sugieren Durbin y Koopman (2001:81).

### 1.2.3. Inicialización y estimación máximo verosímil

El carácter recursivo del filtro de Kalman exige obtener una distribución del estado inicial  $\mathbf{a}_0$  o, cuando menos, determinar su media  $\mathbf{a}_0$  y su varianza  $\mathbf{P}_0$  para iniciar dicho proceso recursivo<sup>95</sup>. La situación ideal es aquella en la que tanto el tipo de distribución como los parámetros  $\mathbf{a}_0$  y  $\mathbf{P}_0$  son conocidos. Pero, en general, no se conoce nada sobre la

<sup>94</sup> La idea de proporcionar menor ponderación a las observaciones más alejadas está en concordancia con el concepto de tendencia local, en contraposición al de tendencia global (Muth, 1960).

<sup>95</sup> La cuestión de la inicialización es relevante, sobre todo, para la computación, dada la influencia que la determinación del estado inicial y su varianza puede tener sobre la estabilidad o convergencia de las iteraciones. En procesos estacionarios y con muestras suficientemente grandes, las propiedades de convergencia y estabilidad del filtro de Kalman hacen que las condiciones iniciales no sean excesivamente relevantes (Vargas, 2000:35-36).

distribución de  $\mathbf{a}_0$ . En este último caso, existen diferentes procedimientos de inicialización del filtro<sup>96</sup>.

Una posibilidad es estimar la ecuación de medida por OLS y adecuar las condiciones iniciales a los correspondientes parámetros constantes estimados por OLS<sup>97</sup>. Otra posibilidad consiste en derivar la distribución inicial a partir de información externa al modelo (Steyn, 1996:25-26), bien, estableciendo una determinada distribución *a priori*, o bien, considerando la estacionariedad del estado. Si se imponen restricciones de estacionariedad sobre el estado, puede concluirse que  $\mathbf{a}_0 = \mathbf{0}$ , mientras que la varianza  $\mathbf{P}_0$  se determina a partir de la ecuación de transición resolviendo iterativamente la ecuación  $\mathbf{P}_i = \mathbf{S}\mathbf{P}_{i-1}\mathbf{S}' + \mathbf{Q}$  hasta que se alcanza la convergencia<sup>98</sup>. Puede demostrarse que

$$\text{vec}(\mathbf{P}_0) = [\mathbf{I}_{m^2} - \mathbf{S} \otimes \mathbf{S}]^{-1} \text{vec}(\mathbf{Q}),$$

donde  $\text{vec}(\mathbf{P}_0)$  y  $\text{vec}(\mathbf{Q})$  son vectores columna formados por las columnas de  $\mathbf{P}_0$  y  $\mathbf{Q}$ <sup>99</sup>.

Si no se dispone de información externa, se suele asumir que  $\mathbf{a}_0 \sim N(\mathbf{a}_0, \hat{\epsilon}\mathbf{I})$ , siendo  $\kappa$  un número suficientemente grande (Steyn, 1996:26). En esta misma línea, puede fijarse un valor arbitrario  $\mathbf{a}_0$  y considerar que  $\kappa \rightarrow \infty$ , de modo que  $\mathbf{P}_0 \rightarrow \infty$ . Se dice entonces que  $\mathbf{a}_0$  tiene una distribución *a priori* difusa. En este último caso, el proceso recursivo inherente al filtro de Kalman se inicia calculando los límites cuando  $\mathbf{P}_0 \rightarrow \infty$  de aquellas expresiones en las que aparece este término. Este proceso se denomina inicialización difusa y el filtro resultante se denomina filtro de Kalman difuso (Harvey y Peters, 1990, y Durbin y Koopman, 2001:100-101). Algunos filtros difusos fueron propuestos por Ansley y Kohn (1985, 1990) y De Jong (1991)<sup>100</sup>. Koopman (1997) y Durbin y Koopman (2001:101-120), basándose en las ideas de Ansley y Kohn (1985,

<sup>96</sup> Véase, entre otros, Bell y Hillmer (1991), Steyn (1996) y Durbin y Koopman (2001).

<sup>97</sup> El programa Eviews 3.1 utiliza este enfoque para inicializar la estimación del filtro de Kalman.

<sup>98</sup> Nótese que, bajo el supuesto de que  $\text{Var}[\mathbf{a}_t] = \text{Var}[\mathbf{a}_{t-1}] = \mathbf{P}_0$  y dado que  $\mathbf{a}_t = \mathbf{S}\mathbf{a}_{t-1} + \mathbf{u}_t$ , se obtiene que  $\mathbf{P}_0 = \mathbf{S}\mathbf{P}_0\mathbf{S}' + \mathbf{Q}$ .

<sup>99</sup> Es decir,  $\text{vec}(\mathbf{P}_0)$  es un vector columna  $m^2 \times 1$  cuyos primeros  $m$  elementos son los de la primera columna de  $\mathbf{P}_0$ , los siguientes  $m$  elementos son los de la segunda columna y, así, sucesivamente. Dado que  $\mathbf{P}_0 = \mathbf{S}\mathbf{P}_0\mathbf{S}' + \mathbf{Q}$ , entonces  $\text{vec}(\mathbf{P}_0) = (\mathbf{S} \otimes \mathbf{S})\text{vec}(\mathbf{P}_0) + \text{vec}(\mathbf{Q})$ . Por tanto,  $\text{vec}(\mathbf{P}_0) = (\mathbf{I}_{m^2} - \mathbf{S} \otimes \mathbf{S})^{-1} \text{vec}(\mathbf{Q})$ . Véase Hamilton (1994:264-265), Steyn (1996:25) y Durbin y Koopman (2001:112).

<sup>100</sup> La propuesta de De Jong (1991) implica aumentar la dimensión del vector de estado, de ahí que se conozca como filtrado y alisado aumentado.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

1990), efectúan un tratamiento más transparente del filtrado y alisado difuso y proponen los métodos denominados filtrado y alisado inicial exacto.

La inicialización difusa es el mecanismo más frecuente cuando no se conoce nada sobre la distribución de  $\mathbf{a}_0$ . Sin embargo, algunos autores opinan que asignar una varianza infinita al estado inicial es incongruente con el hecho de que los valores de la magnitud observada sean finitos. Una aproximación alternativa consiste en asumir que  $\mathbf{P}_0 = \mathbf{0}$  y, por tanto,  $\mathbf{a}_0$  es un parámetro desconocido que puede ser estimado por máxima verosimilitud (Rosenberg, 1973)<sup>101</sup>.

Por último, se pueden sacrificar una o más observaciones y usar, por ejemplo,  $f\left(\mathbf{a}_1 / \mathbf{I}_1\right)$  como distribución inicial (Harvey y Pierse, 1984, y Steyn, 1996:30). Por ejemplo, en el modelo de nivel local, y dado que  $y_1 / \alpha_1 \sim N(\mu_1, \sigma_e^2)$ , se pueden utilizar las estimaciones  $\hat{\mu}_1 = y_1$  y  $Var[\hat{\mu}_1] = \sigma_e^2$ . Obtenida la distribución de  $\alpha_1$  se puede continuar el proceso recursivo<sup>102</sup>.

Como se acaba de ver, el filtro de Kalman permite obtener recursivamente estimaciones del vector de estado; pero éstas serán, en última instancia, función no sólo de los valores iniciales sino también de otros parámetros desconocidos tales como las varianzas de los términos de perturbación de las ecuaciones de transición y de medida. Pues bien, estos parámetros pueden ser estimados por máxima verosimilitud.

La función de verosimilitud de las observaciones, es decir, del vector de vectores  $\mathbf{y} = (y_1, \dots, y_T)'$ , se puede expresar en términos de distribuciones condicionadas de la forma

$$l(\mathbf{y}) = f(y_1, \dots, y_T) = f(y_1) \prod_{t=2}^T f\left(y_t / \mathbf{I}_{t-1}\right)$$

y sabiendo que

$$f\left(y_t / \mathbf{I}_{t-1}\right) \sim N_N(\mathbf{Z}_t \mathbf{S} \mathbf{a}_{t-1}, \mathbf{F}_t),$$

donde

<sup>101</sup> El método propuesto por Rosenberg (1973) implica también aumentar la dimensión del vector de estado y, por esta razón, se encuadra en el conjunto de métodos de filtrado aumentado.

<sup>102</sup> Otra opción consiste en asumir que en el punto  $t = -m$ ,  $\mathbf{a}_{-m}$  tiene una varianza dada por  $\mathbf{P}_{-m}$  y, entonces, bajo ciertos supuestos, derivar el valor de  $\mathbf{P}_0$  (Steyn, 1996:26-29).

## Modelos estructurales de series temporales

$$\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t/t-1} \mathbf{Z}'_t + \mathbf{H},$$

y, conocidos  $\mathbf{a}_0$  y  $\mathbf{P}_0$ ,

$$f(\mathbf{y}_1) \sim N_N(\mathbf{Z}_1 \mathbf{S} \mathbf{a}_0, [\mathbf{Z}_1 \mathbf{P}_{1/0} \mathbf{Z}'_1 + \mathbf{H}]).$$

La anterior función de verosimilitud se puede escribir también en términos de  $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_T)'$ . Dada la definición de los errores de predicción,  $\mathbf{e}_t = \mathbf{y}_t - E\left[\frac{\mathbf{y}_t}{\mathbf{I}_{t-1}}\right]$ , el hecho de que el jacobiano de la transformación de  $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_T)'$  en  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$  es unitario, y teniendo en cuenta que  $\mathbf{e}_t$  es independiente de  $\mathbf{I}_{t-1}$ , se obtiene la función de densidad transformada

$$l(\mathbf{e}) = f(\mathbf{e}_1, \dots, \mathbf{e}_T) = \prod_{t=1}^T f(\mathbf{e}_t).$$

Es decir, los errores de predicción son independientes<sup>103</sup>, de forma que

$$l(\mathbf{e}) = f(\mathbf{e}_1, \dots, \mathbf{e}_T) = \prod_{t=1}^T f(\mathbf{e}_t) = \frac{\prod_{t=1}^T |\mathbf{F}_t|^{-1/2}}{(2\pi)^{NT/2}} \prod_{t=1}^T \exp\left(-\frac{1}{2} \mathbf{e}'_t \mathbf{F}_t^{-1} \mathbf{e}_t\right).$$

Entonces

$$-2 \ln l = NT \cdot \ln(2\pi) + \sum_{t=1}^T \ln |\mathbf{F}_t| + \sum_{t=1}^T \mathbf{e}'_t \mathbf{F}_t^{-1} \mathbf{e}_t$$

y, a partir de esta expresión, se pueden estimar los parámetros del modelo<sup>104</sup>.

Si se supone que  $\mathbf{a}_0$  no es estocástico, sino un parámetro fijo desconocido, puede incorporarse como un parámetro más que debe ser estimado por máxima verosimilitud. Entonces conviene expresar la estimación máximo verosímil de  $\mathbf{a}_0$  como función del resto de parámetros desconocidos y obtener la función de verosimilitud concentrada, en la que  $\mathbf{a}_0$  se sustituye por su estimación máximo verosímil<sup>105</sup>.

En la práctica, la optimización de esta función se realiza a través de algoritmos numéricos de búsqueda basados, en muchos casos, en el método de Newton, que resuelve

<sup>103</sup> Los errores de predicción son importantes desde el punto de vista estadístico porque son serialmente independientes cuando el modelo está especificado correctamente.

<sup>104</sup> Schweppe (1965) fue el primer autor que expuso cómo la función de verosimilitud podía expresarse en términos de los errores de predicción obtenidos a través del filtro de Kalman.

<sup>105</sup> En el caso de que se recurra a la inicialización difusa, el logaritmo de la función de verosimilitud no converge cuando  $\kappa \rightarrow \infty$ , por lo que es preciso redefinir la función de verosimilitud para poder estimar los parámetros (De Jong, 1991, y Durbin y Koopman, 2001:139-141).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

las ecuaciones  $\partial l\left(\frac{\mathbf{y}}{\boldsymbol{\theta}}\right)/\partial \boldsymbol{\theta} = \mathbf{0}$ , siendo  $\boldsymbol{\theta}$  el vector de parámetros que se desea estimar, usando desarrollos en serie de Taylor de primer orden<sup>106</sup>. Aunque desde un punto de vista teórico, es posible maximizar la función de verosimilitud utilizando alguno de estos algoritmos, los resultados de los trabajos realizados indican que se trata de un método singularmente complejo, incluso en los modelos más simples (véase Harvey y Peters, 1990).

### 1.2.4. Observaciones anómalas y filtro de Kalman

El funcionamiento del filtro de Kalman se puede ver afectado por la presencia de observaciones anómalas, también conocidas como *outliers*, que pueden tener efectos transitorios o, por el contrario, afectar a las observaciones posteriores de un modo permanente.

En los modelos en el espacio de los estados, una observación anómala pasajera tiene lugar cuando la perturbación de la ecuación de medida toma un valor extraño en relación con la distribución asumida, pero en el periodo siguiente la perturbación vuelve a comportarse de acuerdo con lo esperado. Este tipo de observación anómala se conoce como *outlier* aditivo (AO). Sin embargo, un valor extraño en la perturbación de la ecuación de transición no sólo altera el estado actual  $\mathbf{a}_t$ , sino que, dada la naturaleza recursiva del modelo, afecta también a los valores siguientes del vector de estado. Este tipo de observación anómala se denomina *outlier* innovativo (IO).

Un valor inesperado de  $\mathbf{y}_t$ , es decir, un valor no concordante con el supuesto de que

$$f\left(\frac{\mathbf{y}_t}{\mathbf{I}_{t-1}}\right) \sim N_N(\mathbf{Z}_t \mathbf{S} \mathbf{a}_{t-1}, \mathbf{F}_t),$$

puede ser atribuido a la presencia de un *outlier* AO o IO<sup>107</sup>. Dado que el error de predicción  $\mathbf{e}_t$  es tal que  $\mathbf{e}_t \sim N_N(\mathbf{0}, \mathbf{F}_t)$ , se podría considerar que existen *outliers* si  $|\mathbf{e}_t| > \lambda \sqrt{\mathbf{F}_t}$ , siendo  $\lambda$  una constante prefijada<sup>108</sup>.

<sup>106</sup> Algunos de los algoritmos más utilizados, tales como el método de Broyden, Fletcher, Goldfarb y Shannon (BFGS) o el algoritmo EM, son descritos por Durbin y Koopman (2001:142-148).

<sup>107</sup> El mejor indicador de la presencia de observaciones anómalas es el error de predicción. Sin embargo, bajo el enfoque estructural se puede recurrir a otro conjunto de residuos, denominados residuos auxiliares, que son las estimaciones estandarizadas de los términos de perturbación asociados a los componentes inobservables. Dichos residuos auxiliares, a pesar de estar serialmente correlacionados aunque el modelo

La presencia de observaciones anómalas tiene graves efectos sobre la estimación de filtro del estado, ya que  $\mathbf{a}_t = \mathbf{a}_{t/t-1} + \mathbf{K}_t \mathbf{e}_t$ . Pues bien, los procedimientos diseñados para minimizar, limitar o incluso eliminar estos efectos, se denominan procedimientos de robustecimiento<sup>109</sup>. Algunos de estos métodos asumen distribuciones no normales que permiten recoger algunas de estas observaciones anómalas. Ahora bien, el supuesto de distribuciones invariantes en el tiempo choca con el problema de que la distribución que recoge las anomalías no se adapta a las observaciones que no son anómalas. Otro grupo de métodos emplea ecuaciones de actualización modificadas *ad hoc*, como por ejemplo  $\mathbf{a}_t = \mathbf{a}_{t/t-1} + \mathbf{K}_t \min\{\mathbf{e}_t, \lambda \sqrt{\mathbf{F}_t}\}$ , que pretenden atenuar la influencia del error de predicción en la actualización de la estimación.

En cualquier caso, es interesante detectar si existen *outliers* AO o IO. Una forma de proceder si se detecta un valor sospechoso,  $\mathbf{y}_t$ , consiste en predecir el siguiente valor,  $\mathbf{y}_{t+1}$ , y juzgar si se ajusta a lo esperado teniendo en cuenta que

$$f\left(\mathbf{y}_{t+1} / \mathbf{I}_{t-1}\right) \sim N_N\left(\mathbf{Z}_{t+1} \mathbf{S} \mathbf{a}_{t/t-1}, \left[\mathbf{Z}_{t+1} \left[\mathbf{S} \mathbf{P}_{t/t-1} \mathbf{S}' + \mathbf{Q}\right] \mathbf{Z}_{t+1}'\right] + \mathbf{H}\right).$$

Si  $\mathbf{y}_{t+1}$  es concordante con lo esperado,  $\mathbf{y}_t$  es *outlier* AO; si, por el contrario,  $\mathbf{y}_{t+1}$  no es concordante con lo esperado,  $\mathbf{y}_t$  es *outlier* IO.

Si el outlier es AO, el valor de  $\mathbf{y}_t$  no es informativo y por tanto, la expresión

$$f\left(\mathbf{a}_t / \mathbf{I}_t\right) = \frac{f\left(\mathbf{y}_t / \mathbf{a}_t\right) f\left(\mathbf{a}_t / \mathbf{I}_{t-1}\right)}{f\left(\mathbf{y}_t / \mathbf{I}_{t-1}\right)}$$

se debe transformar en

$$f\left(\mathbf{a}_t / \mathbf{I}_t\right) = f\left(\mathbf{a}_t / \mathbf{I}_{t-1}\right),$$

---

esté correctamente especificado, informan de la presencia de observaciones anómalas y, además, proporcionan una indicación de cuál es la fuente del problema (Harvey y Koopman, 1992).

<sup>108</sup> Una revisión de los métodos de diagnóstico para modelos en el espacio de los estados, incluyendo la detección de observaciones anómalas, puede encontrarse en Harvey y Koopman (1992). Véase también Atkinson y otros (1997), Koopman y otros (2000) y Durbin y Koopman (2001:33-35, 152-154).

<sup>109</sup> Véase Meinhold y Singpurwalla (1989).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

es decir, no se debe tener en cuenta  $y_t$ . Si el outlier es IO, la función de densidad

$f\left(\mathbf{a}_t / \mathbf{a}_{t-1}\right)$  no es informativa y, por tanto,

$$f\left(\mathbf{a}_t / I_{t-1}\right) = \int_{-\infty}^{\infty} f\left(\mathbf{a}_t / \mathbf{a}_{t-1}\right) f\left(\mathbf{a}_{t-1} / I_{t-1}\right) d\mathbf{a}_{t-1} = \int_{-\infty}^{\infty} C f\left(\mathbf{a}_{t-1} / I_{t-1}\right) d\mathbf{a}_{t-1} = C,$$

siendo  $C$  una constante; es decir, la función de densidad de la predicción no es informativa, ya que la información hasta  $t-1$  no informa sobre el estado en  $t$ . En este caso, es preciso comenzar otra vez la aplicación del filtro de Kalman a partir de la observación anómala.

Por otro lado, se debe tener en cuenta que este tipo de observaciones pueden generar dudas sobre los supuestos distribucionales subyacentes en el modelo. Cuando no se trabaja bajo los supuestos de normalidad no hay garantías de que la estimación proporcionada por el filtro de Kalman coincida con la media condicional del vector de estado pero, la incorrelación entre las perturbaciones es suficiente para que el filtro de Kalman proporcione estimaciones óptimas lineales.

### 1.3. Modelos estructurales y filtro de Kalman

Como ya se ha comentado, los modelos estructurales de series temporales constituyen un caso particular de los modelos en el espacio de los estados donde los componentes típicos de la serie temporal son las variables de estado inobservables y la propia serie es la magnitud observable relacionada con el vector de variables de estado<sup>110</sup>.

Supóngase una serie temporal unidimensional  $y_t = \{y_1, \dots, y_T\}$  cuyas características más sobresalientes se pueden representar mediante un modelo estructural básico. Si se supone que  $s$  es par y se adopta la representación trigonométrica de la estacionalidad, el modelo anterior queda especificado por las ecuaciones siguientes

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \varepsilon_t, & \varepsilon_t &\sim NID(0, \sigma_\varepsilon^2) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2) \end{aligned}$$

<sup>110</sup> Un planteamiento análogo podría formularse en relación a la desagregación temporal en el análisis de series temporales, en la que la suma de los desagregados inobservables coincide con el agregado observable. En este sentido, Quilis (2001:28) señala que una de las líneas de desarrollo en el campo de la desagregación temporal consiste en plantear el problema mediante su representación en el espacio de los estados y el uso del filtro de Kalman, algoritmo que permite la generalización de muchos de los métodos de desagregación existentes además de constituir un enfoque computacionalmente más eficiente (véase, por ejemplo, Gómez y Maravall, 1994).

## Modelos estructurales de series temporales

$$\beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim NID(0, \sigma_\zeta^2)$$

$$\gamma_t = \sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{j,t}$$

$$\gamma_{j,t} = \cos\left(\frac{2\pi j}{s}\right) \gamma_{j,t-1} + \text{sen}\left(\frac{2\pi j}{s}\right) \gamma_{j,t-1}^* + \omega_{j,t},$$

$$\gamma_{j,t}^* = -\text{sen}\left(\frac{2\pi j}{s}\right) \gamma_{j,t-1} + \cos\left(\frac{2\pi j}{s}\right) \gamma_{j,t-1}^* + \omega_{j,t}^*,$$

para  $j = 1, \dots, (s/2) - 1$ , y

$$\mathbf{g}_{\frac{s}{2},t} = -\mathbf{g}_{\frac{s}{2},t-1} + \mathbf{w}_{\frac{s}{2},t}$$

donde

$$\omega_{j,t} \sim NID(0, \sigma_\omega^2),$$

$$\omega_{j,t}^* \sim NID(0, \sigma_\omega^2),$$

y

$$\text{Cov}\left[\omega_{j,t}, \omega_{j,t}^*\right] = 0.$$

Por tanto, los elementos que aportan información sobre el sistema son los  $s + 1$  términos del vector de estado

$$\mathbf{a}_t = \left( \mu_t, \beta_t, \gamma_{1,t}, \gamma_{1,t}^*, \dots, \gamma_{\frac{s}{2}-1,t}, \gamma_{\frac{s}{2}-1,t}^*, \gamma_{\frac{s}{2},t} \right)'$$

La ecuación de transición viene dada por

$$\mathbf{a}_t = \mathbf{S}\mathbf{a}_{t-1} + \mathbf{u}_t,$$

donde  $\mathbf{u}_t$  es un vector  $(s + 1) \times 1$  definido como

$$\mathbf{u}_t = \left( \eta_t, \zeta_t, \omega_{1,t}, \omega_{1,t}^*, \dots, \omega_{\frac{s}{2}-1,t}, \omega_{\frac{s}{2}-1,t}^*, \omega_{\frac{s}{2},t} \right)'$$

tal que  $\mathbf{u}_t \sim NID(0, \mathbf{Q})$ , siendo  $\mathbf{Q} = \text{diag}(\sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2 \mathbf{I}_{s-1})$ . La matriz  $\mathbf{S}$  está definida como,

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(2\pi/s) & \text{sen}(2\pi/s) & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & -\text{sen}(2\pi/s) & \cos(2\pi/s) & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cos\left(2\pi\left(\frac{s}{2}-1\right)/s\right) & \text{sen}\left(2\pi\left(\frac{s}{2}-1\right)/s\right) & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & -\text{sen}\left(2\pi\left(\frac{s}{2}-1\right)/s\right) & \cos\left(2\pi\left(\frac{s}{2}-1\right)/s\right) & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \end{pmatrix}$$

y la ecuación de medida adopta la expresión

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t,$$

donde  $\mathbf{Z}_t = (1, 0, 1, 0, \dots, 1, 0, 1)$ <sup>111</sup>.

Si se incorporan al modelo variables explicativas, la ecuación de medida quedaría especificada como

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^k \delta_j x_{j,t} + \varepsilon_t.$$

A efectos ilustrativos, si se elimina el componente estacional, entonces el modelo definido como

$$y_t = \mu_t + \sum_{j=1}^k \delta_j x_{j,t} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2),$$

donde

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim NID(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim NID(0, \sigma_\zeta^2), \end{aligned}$$

y

$$\delta_{j,t} = \delta_{j,t-1}, \quad j = 1, \dots, k,$$

podría formularse como un modelo en el espacio de los estados en el que

$$\mathbf{a}_t = (\mu_t, \beta_t, \delta_{1,t}, \dots, \delta_{k,t})'.$$

<sup>111</sup> Por ejemplo, en el caso de datos trimestrales ( $s=4$ ),  $\mathbf{a}_t = (\mu_t, \beta_t, \gamma_{1,t}, \gamma_{1,t}^*, \gamma_{2,t})'$ ,  $\mathbf{u}_t = (\mathbf{h}_t, \mathbf{z}_t, \mathbf{w}_{1,t}, \mathbf{w}_{1,t}^*, \mathbf{w}_{2,t})'$ ,  $\mathbf{Q} = \text{diag}(\sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2 \mathbf{I}_3)$  y  $\mathbf{Z}_t = (1, 0, 1, 0, 1)$ .

## Modelos estructurales de series temporales

La ecuación de transición conviene escribirla como

$$\mathbf{a}_t = \mathbf{S}\mathbf{a}_{t-1} + \mathbf{R}\mathbf{u}_t,$$

donde  $\mathbf{u}_t$  es un vector  $2 \times 1$  definido como  $\mathbf{u}_t = (\eta_t, \zeta_t)'$ , tal que  $\mathbf{u}_t \sim NID(0, \mathbf{Q})$ , siendo  $\mathbf{Q}$  una matriz definida como  $\mathbf{Q} = \text{diag}(\sigma_\eta^2, \sigma_\zeta^2)$ , y  $\mathbf{R}$  una matriz de dimensión  $(k+2) \times 2$  tal que

$$\mathbf{R}' = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \end{pmatrix};$$

es decir, la matriz  $\mathbf{R}$  contiene dos vectores columna que seleccionan los términos de perturbación correspondientes a los elementos estocásticos del vector de estado. La matriz  $\mathbf{S}$  es una matriz de dimensión  $(k+2) \times (k+2)$  definida como:

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

y la ecuación de medida adopta la expresión

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t,$$

donde  $\mathbf{Z}_t = (1, 0, x_{1,t}, \dots, x_{k,t})$ .

Para el modelo univariante de nivel local, la especificación anterior se transforma en la ecuación de medida

$$y_t = \alpha_t + \varepsilon_t,$$

con  $\alpha_t = \mu_t$ , donde  $\varepsilon_t$  es un término de perturbación tal que

$$E[\varepsilon_t] = 0,$$

$$\text{Var}[\varepsilon_t] = h$$

y

$$E[\varepsilon_t \varepsilon_s] = 0, \quad \forall s \neq t, \quad s, t = 1, \dots, T,$$

y la ecuación de transición

$$\alpha_t = \alpha_{t-1} + u_t,$$

con  $u_t = \eta_t$ , donde  $\eta_t$  es un término de perturbación tal que

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$E[\eta_t] = 0,$$

$$Var[\eta_t] = q$$

y

$$E[\eta_t \eta_s] = 0, \quad \forall s \neq t, \quad s, t = 1, \dots, T.$$

Además, se supone que

$$E[\mu_0] = m_0,$$

$$Var[\mu_0] = P_0,$$

$$E[\varepsilon_t \mu_0] = 0$$

$$E[\eta_t \mu_0] = 0$$

y

$$E[\varepsilon_t \eta_s] = 0, \quad \forall s \neq t, \quad s, t = 1, \dots, T.$$

En este último caso, la estimación del componente de nivel a través del filtro de Kalman y la estimación máximo verosímil de los hiperparámetros se efectúa como se indica a continuación.

Sea  $m_{t/t-1}$ , la estimación óptima del estado en el instante  $t$  con la información disponible hasta  $t-1$  y sea  $m_t$ , la estimación óptima del estado en el instante  $t$  con la información disponible hasta  $t$ , es decir, una vez observada la medida  $y_t$ .

Las ecuaciones de predicción son

$$m_{t/t-1} = m_{t-1}$$

$$P_{t/t-1} = P_{t-1} + q$$

y las ecuaciones de actualización son

$$m_t = m_{t/t-1} + \frac{P_{t/t-1}}{P_{t/t-1} + h} [y_t - m_{t/t-1}]$$

$$P_t = \left( 1 - \frac{P_{t/t-1}}{P_{t/t-1} + h} \right) P_{t/t-1}.$$

Dados  $m_0$  y  $P_0$ , se puede obtener el estimador óptimo de  $\mu_t$  del siguiente modo:

$$\left. \begin{array}{l} m_{t/t-1} = m_{t-1} \\ P_{t/t-1} = P_{t-1} + q \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} m_t = m_{t/t-1} + \frac{P_{t/t-1}}{P_{t/t-1} + h} [y_t - m_{t/t-1}] \\ P_t = \left( 1 - \frac{P_{t/t-1}}{P_{t/t-1} + h} \right) P_{t/t-1} \end{array} \right.$$

Por ejemplo, la predicción y la estimación filtrada del estado en el instante  $t = 1$  vienen dadas por

$$\left. \begin{array}{l} m_{1/0} = m_0 \\ P_{1/0} = P_0 + q \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} m_1 = m_{1/0} + \frac{P_{1/0}}{P_{1/0} + h} [y_1 - m_{1/0}] = m_0 + \frac{P_0 + q}{P_0 + q + h} [y_1 - m_0] \\ P_1 = \left( 1 - \frac{P_0 + q}{P_0 + q + h} \right) (P_0 + q) \end{array} \right.$$

y en el momento  $t = 2$  por

$$\left. \begin{array}{l} m_{2/1} = m_1 = m_0 + \frac{P_0 + q}{P_0 + q + h} [y_1 - m_0] \\ P_{2/1} = P_1 + q = \left( 1 - \frac{P_0 + q}{P_0 + q + h} \right) (P_0 + q) + q \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} m_2 = m_{2/1} + \frac{P_{2/1}}{P_{2/1} + h} [y_2 - m_{2/1}] \\ P_2 = \left( 1 - \frac{P_{2/1}}{P_{2/1} + h} \right) P_{2/1} \end{array} \right.$$

Las ecuaciones de la estimación alisada son

$$m_{t/T} = m_t + P_t^* [m_{t+1/T} - m_t]$$

y

$$P_{t/T} = P_t + P_t^* (P_{t+1/T} - P_{t+1/t}) P_t^* ,$$

donde

$$P_t^* = P_t P_{t+1/t}^{-1} = \frac{P_t}{P_t + q}, \quad t = T - 1, \dots, 1.$$

Por tanto,

$$m_{t/T} = m_t + \frac{P_t}{P_t + q} [m_{t+1/T} - m_t] = \left( 1 - \frac{P_t}{P_t + q} \right) m_t + \frac{P_t}{P_t + q} m_{t+1/T}.$$

Es decir, la estimación alisada del estado en el momento  $t$  es, simplemente, una media ponderada entre la estimación obtenida por el filtro de Kalman para el estado en el momento  $t$  y la estimación alisada del estado en el momento  $t + 1$ .

Si el modelo es gaussiano, es decir, si  $\varepsilon_t$ ,  $\eta_t$  y  $\mu_0$  se distribuyen como variables aleatorias normales, entonces las estimaciones filtradas  $m_t$  y  $P_t$  proporcionadas por el filtro

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

de Kalman coinciden con la media y la matriz de covarianzas de la distribución de  $\mu_t$  condicionada a la información en  $t$ .

Como se puede observar, las estimaciones del nivel son función de  $m_0$ ,  $P_0$ ,  $q$  y  $h$ . Si se supone que  $P_0 = 0$ ,  $m_0$  se puede considerar como un parámetro más que debe ser estimado junto con los hiperparámetros  $q$  y  $h$ . Si se asume que el modelo es gaussiano, estos parámetros desconocidos,  $\mathbf{y}$ , se pueden estimar por máxima verosimilitud. La función de densidad conjunta del proceso estocástico observable  $\{y_t\}$ ,  $t=1,2,\dots,T$ , se puede expresar en términos de los errores de predicción, innovaciones o residuos,  $e_t$ , que se distribuyen como normales independientes con  $E[e_t] = 0$  y  $Var[e_t] = F_t$ . Es decir,

$$L(e_1, \dots, e_T, \mathbf{y}) = \prod_{t=1}^T \left\{ \frac{1}{\sqrt{2\pi}\sqrt{F_t}} \exp\left(-\frac{1}{2} \frac{e_t^2}{F_t}\right) \right\} = \frac{1}{(2\pi)^{T/2} \sqrt{F_1} \dots \sqrt{F_T}} \exp\left(-\frac{1}{2} \sum_{t=1}^T \frac{e_t^2}{F_t}\right)$$

$$\ln L(e_1, \dots, e_T, \mathbf{y}) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln(F_t) - \frac{1}{2} \sum_{t=1}^T \frac{e_t^2}{F_t}$$

donde  $e_t = y_t - m_{t/t-1}$  y  $F_t = P_{t/t-1} + h = P_{t-1/t-1} + q + h$ .

Asignados unos valores iniciales se pueden obtener recursivamente las expresiones de  $m_{t-1}$  y  $P_{t-1}$  como funciones de  $q$  y  $h$ . Una vez expresada la función de verosimilitud en términos de estos dos parámetros se obtienen directamente sus estimaciones máximo verosímiles. La función de verosimilitud se puede reparametrizar haciendo que todas las varianzas sean proporcionales a  $h$ , es decir,

$$Var[\eta_t] = hq,$$

$$Var[\alpha_0] = hP_0$$

y

$$Var[e_t] = hf_t.$$

Entonces,

$$\begin{aligned} \ln L(e_1, \dots, e_T, \mathbf{y}^*, h) &= -\frac{T}{2} \ln(2\mathbf{p}) - \frac{1}{2} \sum_{t=1}^T \ln(hf_t) - \frac{1}{2} \sum_{t=1}^T \frac{e_t^2}{hf_t} \\ &= -\frac{T}{2} \ln(2\mathbf{p}) - \frac{T}{2} \ln(h) - \frac{1}{2} \sum_{t=1}^T \ln(f_t) - \frac{1}{2h} \sum_{t=1}^T \frac{e_t^2}{f_t} \end{aligned}$$

y se tiene que

$$\frac{\partial \ln L(e_1, \dots, e_T, \mathbf{y}^*, h)}{\partial h} = -\frac{T}{2h} + \frac{1}{2h^2} \sum_{t=1}^T \frac{e_t^2}{f_t},$$

de modo que, igualando a cero la derivada anterior, resulta que

$$\hat{h}(\mathbf{y}^*) = \frac{1}{T} \sum_{t=1}^T \frac{e_t^2}{f_t}$$

Finalmente, se obtiene la función de verosimilitud concentrada

$$\ln L(e_1, \dots, e_T, \mathbf{y}^*) = -\frac{T}{2} \ln(2p) - \frac{T}{2} \ln \left( \frac{1}{T} \sum_{t=1}^T \frac{e_t^2}{f_t} \right) - \frac{1}{2} \sum_{t=1}^T \ln(f_t) - \frac{T}{2},$$

que debe ser maximizada respecto a los elementos del vector  $\mathbf{y}^*$ .

En modelos más generales, la función de verosimilitud suele ser reparametrizada con objeto de reducir la dimensionalidad de la búsqueda numérica en el proceso de optimización. Por ejemplo, se puede optimizar respecto a  $\sigma_\varepsilon^2$  y las varianzas de los elementos del vector de estado, o bien, utilizar  $\sigma_\varepsilon^2$  y los ratios señal ruido<sup>112</sup> como parámetros que deben ser estimados. Entonces, puede obtenerse la estimación máximo verosímil de  $\sigma_\varepsilon^2$  como función del resto de parámetros y sustituir esta estimación en la función de verosimilitud. Se obtiene de este modo la función de verosimilitud concentrada, que sólo dependerá de los ratios señal ruido o de las varianzas de los elementos del vector de estado.

En el caso multivariante, la estimación del modelo formulado en el espacio de los estados puede llevarse a cabo también a través del filtro de Kalman<sup>113</sup>. Bajo el supuesto de homogeneidad, que implica que las distintas series tienen las mismas propiedades estocásticas, el modelo puede ser manejado aplicando el filtro de Kalman univariante a cada serie<sup>114</sup>; sin embargo, si los hiperparámetros son desconocidos, usar un filtro de Kalman multivariante permite obtener estimaciones alisadas más eficientes de los componentes<sup>115</sup>.

<sup>112</sup> Los ratios señal ruido se definen como el cociente entre la varianza de cada uno de los componentes y la varianza del componente irregular.

<sup>113</sup> Véase Harvey y Shephard (1993).

<sup>114</sup> Para contrastar la homogeneidad, se puede recurrir al test de multiplicadores de Lagrange diseñado por Fernández y Harvey (1990).

<sup>115</sup> Véase Harvey (1989:438-442).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

En general, dado un modelo estructural multivariante para la serie  $y_t$  de dimensión  $N \times 1$ , puede ser conveniente, en términos computacionales, tratar el vector  $N$ -dimensional de  $T$  observaciones como una serie univariante de  $T \times N$  observaciones<sup>116</sup>. Por ejemplo, sea el modelo definido por las ecuaciones

$$y_t = Z_t \mathbf{a}_t + \mathbf{e}_t,$$

$$\mathbf{a}_t = S_t \mathbf{a}_{t-1} + R_t u_t,$$

donde

$$\mathbf{e}_t \sim N(\mathbf{0}, H_t),$$

$$u_t \sim N(0, Q_t),$$

siendo

$$y_t = (y_{1,t}, \dots, y_{N,t})', \quad \mathbf{a}_t = (\alpha_{1,t}, \dots, \alpha_{m,t})', \quad \mathbf{e}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})'.$$

Además,  $Z_t = (Z_{1,t}, \dots, Z_{N,t})'$  es un vector columna  $N \times 1$  cuyos elementos son vectores fila de dimensión  $1 \times m$ , y, por último, la matriz  $H_t$ , de dimensión  $N \times N$ , se define como

$$H_t = \begin{pmatrix} \sigma_{1,t}^2 & 0 & \dots & 0 \\ 0 & \sigma_{2,t}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{N,t}^2 \end{pmatrix}.$$

Entonces, las  $T$  observaciones del vector  $y_t = (y_{1,t}, \dots, y_{N,t})'$ ,  $t = 1, \dots, T$ , se pueden escribir como una serie univariante  $\{y_{i,t}\}_{i=1, \dots, N, t=1, \dots, T}$ , definida como el conjunto de observaciones  $\{y_{1,1}, \dots, y_{N,1}, y_{1,2}, \dots, y_{N,2}, \dots, y_{1,T}, \dots, y_{N,T}\}$ . La ecuación de medida para esta representación univariante puede formularse ahora como

$$y_{i,t} = Z_{i,t} \alpha_t + \varepsilon_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T,$$

y sobre esta nueva serie se pueden aplicar los algoritmos de filtrado y alisado<sup>117</sup>.

Los modelos estructurales descritos en los párrafos precedentes se caracterizan por asumir relaciones lineales e invariantes en el tiempo y se supone, además, que las observaciones se han generado a partir de distribuciones gaussianas. Un modelo lineal para

<sup>116</sup> Véase Durbin y Koopman (2001:128-134).

<sup>117</sup> Si  $H_t$  no es diagonal, puede introducirse  $\mathbf{e}$  como elemento del vector de estado, o bien, transformar las observaciones de modo que se diagonalice la matriz  $H_t$  (Durbin y Koopman, 2001:131-132).

el sistema se justifica, entre otras razones, por su manejo más fácil y porque algunas no linealidades pueden linealizarse<sup>118</sup>. Por otro lado, Maybeck (1979:7-9) justifica también el supuesto de gaussianidad y comportamiento ruido blanco para los términos de perturbación<sup>119</sup>.

Además de la mayor facilidad para su tratamiento, los modelos estructurales de esta clase son los que con menor dificultad se relacionan con otros enfoques. Y lo cierto es que el modelo estructural lineal gaussiano, tras adecuadas transformaciones de las observaciones, permite analizar muchas de las series temporales que se encuentran en la práctica. Sin embargo, hay otras situaciones en las que una distribución no normal puede proporcionar un mejor modelo<sup>120</sup>. Por ejemplo, si las observaciones registradas son datos de recuento puede resultar más apropiada una distribución de Poisson y algunas distribuciones con colas pesadas suelen ser convenientes para tratar observaciones anómalas.

También es importante permitir no linealidades con objeto, por ejemplo, de recoger situaciones en las que la tendencia y la estacionalidad se combinan de modo multiplicativo, mientras que el componente irregular se incorpora al modelo de forma aditiva. Estas no linealidades pueden introducirse en la ecuación de medida o en la ecuación de transición.

Entonces, la representación en el espacio de los estados viene dada, por ejemplo, por

$$y_t = Z_t(\alpha_t) + \varepsilon_t$$

---

<sup>118</sup> Si las ecuaciones de transición o de medida fueran no lineales, es preciso diseñar un filtro que consiga linealizar las relaciones. Este tipo de filtros se denomina filtros de Kalman extendidos (Welch y Bishop, 2001).

<sup>119</sup> El supuesto de gaussianidad se puede justificar físicamente por el hecho de que cualquier ruido o error de sistema o medida se debe, generalmente, a un número importante de fuentes pequeñas y el teorema central del límite muestra que cuando un número de variables aleatorias independientes se expresan de forma agregada, el efecto final puede ser descrito con bastante proximidad por una función de densidad de probabilidad gaussiana, a pesar de la forma de las densidades individuales. Al igual que antes, la gaussianidad hace que las operaciones matemáticas sean tratables; pero, además, normalmente un ingeniero conocerá, como mucho, los estadísticos de primer y segundo orden de un proceso ruidoso. En ausencia de estadísticos de cualquier orden mayor, no hay solución más conveniente que asumir gaussianidad. Este supuesto de índole práctica es también aplicable a la hipótesis de comportamiento ruido blanco. Maybeck recurre a sistemas físicos para mostrar que los comportamientos ruidosos suelen recogerse adecuadamente a través de procesos ruido blanco o de filtros basados en procesos ruido blanco. Pero, además de este hecho, el autor reconoce que otros supuestos no dan lugar a formulaciones tratables del problema de estimación y control. Afortunadamente, la clase de modelos en los que las operaciones matemáticas resultan tratables también proporcionan representaciones adecuadas para muchas aplicaciones de interés.

<sup>120</sup> Incluso en el caso de que los términos de perturbación del modelo sean no gaussianos, el filtro de Kalman puede ser empleado para obtener una proyección lineal de la serie a partir de sus valores pasados. El

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$\alpha_t = S_t(\alpha_{t-1}) + u_t,$$

donde  $Z_t$  y  $S_t$  están definidas de modo que  $y_t$  depende de  $\alpha_t$  de forma no lineal y lo mismo ocurre con la relación entre  $\alpha_t$  y  $\alpha_{t-1}$ <sup>121</sup>.

### 1.4. Modelos estructurales versus modelos ARIMA

De lo comentado en las páginas anteriores, se deduce que los modelos estructurales presentan ventajas respecto a los modelos ARIMA. Algunas de ellas se comentan a continuación.

Los modelos ARIMA pueden formularse en el espacio de los estados<sup>122</sup> y muchos modelos estructurales admiten una representación ARIMA. De hecho, cuando las series temporales tienen una estructura subyacente suficientemente simple, las dos formulaciones son esencialmente equivalentes<sup>123</sup>; pero cuando tales estructuras son más complejas las diferencias entre ambos enfoques son más evidentes<sup>124</sup>.

En un modelo estructural de series temporales, cada uno de los componentes relevantes de la serie, tales como la tendencia, el ciclo o las variaciones estacionales, es formulado explícitamente y, por tanto, es posible obtener información específica sobre ellos<sup>125</sup>. Además, el investigador puede identificar e introducir cualquier característica que requiera un tratamiento especial en situaciones particulares. Ésta es, quizás, la principal ventaja del enfoque estructural frente a la aproximación ARIMA, en la que la tendencia y el componente estacional se eliminan aplicando convenientes diferencias a la serie original

---

procedimiento de estimación *quasi*-máximo verosímil proporciona estimadores consistentes y asintóticamente normales de los elementos de las matrices del sistema (Hamilton, 1994:389).

<sup>121</sup> Durbin y Koopman (2001:177-239) hacen un extenso tratamiento de los modelos estructurales no lineales y no gaussianos. Véase también Kitagawa (1987).

<sup>122</sup> Algunos ejemplos de modelos ARIMA representados en el espacio de los estados pueden encontrarse en Harvey (1989:510-511), Aoki (1990:21-38), Brockwell y Davis (1991:465-473), Hamilton (1994:374-377) y De Jong y Chu-Chun-Lin (1994), entre otros.

<sup>123</sup> La equivalencia entre la forma reducida restringida de los modelos estructurales y el enfoque ARIMA permite soslayar el principal inconveniente de éste último enfoque, a saber, la imposición a priori de patrones de comportamiento (Espasa y Cancelo, 1993:296).

<sup>124</sup> Harvey (1985) muestra algunas de estas diferencias al analizar series temporales macroeconómicas anuales. En concreto, observa que los modelos ARIMA elegidos bajo el principio de parsimonia pueden dar lugar a formulaciones que, aunque predicen bien en el corto plazo, no recogen las características más notables de la serie bajo estudio y tienen poco significado desde el punto de vista económico. Algunas de estas diferencias son también señaladas por Durbin y Koopman (2001:39, 46-49, 51-53).

<sup>125</sup> Como indica Harvey (1985), un modelo estructural no tiene por objeto reproducir el proceso generador de datos, sino obtener las características presentes en la serie temporal en términos de su descomposición en componentes que, por sí mismos, resultan interesantes para los economistas.

antes de proceder al análisis<sup>126</sup>. En definitiva, la metodología ARIMA constituye una especie de caja negra en la que el modelo adoptado depende puramente de los datos, sin análisis *a priori* de la estructura del sistema que los generó. Los modelos estructurales son, desde este punto de vista, más transparentes, ya que permiten comprobar si el comportamiento predicho por el modelo para cada componente de la serie se corresponde con lo esperado.

Debe tenerse en cuenta, además, que la eliminación de tendencia y estacionalidad puede constituir un fuerte inconveniente si, como ocurre con la estadística oficial o en muchas aplicaciones econométricas, dichos componentes tienen interés en sí mismos y no sólo interesa la predicción de la serie. Además, el requerimiento de estacionariedad del planteamiento propuesto por Box y Jenkins (1976) obliga a diferenciar las series, pero no siempre es fácil decidir el orden de integración adecuado<sup>127</sup>. En realidad, las herramientas básicas de identificación de modelos ARIMA, es decir, las funciones de autocorrelación, son bastante imprecisas y no siempre permiten decantarse claramente por un único modelo<sup>128</sup>.

Los modelos en el espacio de los estados son también más flexibles en la especificación y presentan ventajas en la fase de estimación como consecuencia de la utilización del filtro de Kalman en combinación con los procedimientos de máxima verosimilitud. La naturaleza recursiva del modelo y las técnicas computacionales para su análisis permiten incorporar directamente cambios conocidos en la estructura del sistema a

---

<sup>126</sup> Por ejemplo, cuando la longitud del periodo estacional no es fija, un modelo adecuado del componente estacional resulta más fácil de integrar en el enfoque estructural que en la metodología ARIMA. Véase, por ejemplo, Harvey, Koopman y Riani (1997).

<sup>127</sup> Parece poco realista considerar que las series temporales económicas son algunas veces no estacionarias. Más bien, habría que asumir que son ocasionalmente estacionarias. En este sentido, la hipótesis nula de los tests habituales de raíz unitaria debería ser intercambiada por la hipótesis alternativa de estacionariedad. Además, estos tests están basados en aproximaciones autorregresivas generalmente pobres, especialmente en presencia de componentes cíclicos (Harvey, 1985, y Harvey, 1997). En cambio en el enfoque estructural no es preciso decidir *a priori* el orden de integración de la serie, sino que los comportamientos cambiantes son incorporados explícitamente.

<sup>128</sup> Bajo la metodología ARIMA es frecuente encontrar que las magnitudes económicas se comportan como procesos integrados de orden uno, es decir, procesos que se caracterizan por cambios persistentes. Este hecho se debe a que los procesos integrados de orden dos caracterizados por cambios más suaves no suelen ser identificados en muestras pequeñas a partir de los correlogramas y los tests de raíz unitaria estándar. Estos cambios más suaves se suelen presentar en magnitudes económicas que responden más a largo plazo ante una determinada medida de política económica. Harvey (1985) y Harvey (1997) señalan la escasa utilidad de los correlogramas para identificar modelos complejos con muestras pequeñas, sobre todo si no se dispone de información alguna sobre cuál podría ser el modelo adecuado. El principio de parsimonia también puede conducir a una elección incorrecta.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

lo largo del tiempo. En cambio, los modelos de Box y Jenkins están basados en el supuesto de que la serie diferenciada es estacionaria, lo que incorpora mayor rigidez<sup>129</sup>. Asimismo, en el planteamiento estructural, la predicción es bastante directa<sup>130</sup> y las observaciones perdidas son más fáciles de tratar<sup>131</sup>. Brockwell y Davis (1991) consideran que la representación en el espacio de los estados y las ecuaciones recursivas propias del filtro de Kalman son ideales para el análisis de datos con valores perdidos<sup>132</sup>. En el enfoque ARIMA, sin embargo, es relativamente difícil manejar observaciones perdidas, y la incorporación de variables explicativas, efectos de calendario y cambios estructurales no es siempre tan inmediata como en la representación en el espacio de los estados<sup>133</sup>.

Por otra parte, las observaciones correspondientes a series multivariantes pueden ser manejadas por extensión directa de la formulación estructural univariante; cosa que no ocurre en la metodología ARIMA. Además, la naturaleza markoviana de los modelos en el espacio de los estados permite que los cálculos necesarios para implementarlos se realicen de forma recursiva; lo que permite manejar modelos de elevada dimensión sin un incremento desproporcionado de la tarea computacional.

En este sentido, los modelos estructurales multivariantes de series temporales pueden presentar alguna ventaja sobre la representación VAR, que en general requiere un gran número de parámetros además de los propios de la matriz de covarianzas de los

---

<sup>129</sup> Algunas observaciones que no era posible explicar con los modelos de forma reducida, y, por tanto, eran valores anómalos, sí lo son con los modelos estructurales, es decir, no aparecen como valores atípicos (De los Llanos y Valentina, 1994).

<sup>130</sup> Harvey y Todd (1983) comparan las predicciones realizadas a partir de modelos ARIMA y modelos estructurales y concluyen que hay fuertes argumentos a favor del enfoque estructural desde el punto de vista práctico. Los modelos ARIMA parecen predecir razonablemente bien en el corto plazo, pero, por ejemplo, pueden ser incapaces de generar los comportamientos cíclicos que caracterizan a algunas series económicas reales (Harvey y Jaeger, 1993, y Harvey, 1997). Cabrero y otros (2002) también realizan comparaciones predictivas en series con estacionalidad y concluyen que es necesario complementar las predicciones ARIMA —mejores para algunos horizontes de predicción, pero incapaces de captar ciertos movimientos periódicos— con las obtenidas a través de modelos estructurales.

<sup>131</sup> Para obtener las estimaciones filtradas en los instantes del tiempo correspondientes a las observaciones perdidas, basta con suponer que  $e_t=0$  y  $K_t=0$  en los valores perdidos (Harvey y Koopman, 1992, y Durbin y Koopman, 2001:23-24).

<sup>132</sup> Estos autores muestran, en concreto, las ventajas de esta aproximación frente a los modelos ARIMA en lo referente a la evaluación de la función de verosimilitud gaussiana para los valores efectivamente observados y a la estimación, con error mínimo cuadrático medio, de los valores perdidos. Véase Brockwell y Davis (1991:482-488). Jones (1993:89-94) también expone el procedimiento para tratar observaciones perdidas en la representación en el espacio de los estados.

<sup>133</sup> También es verdad que los procesos autorregresivos son fáciles de especificar y estimar y se espera que proporcionen buenas predicciones en el corto plazo en muchas situaciones.

## Modelos estructurales de series temporales

términos de perturbación. En particular, cuando algunas de las series son no estacionarias, un VAR en diferencias puede no ser una especificación adecuada puesto que diferentes series pueden tener distinto orden de integración. Entonces, la aplicación de diferencias puede provocar la aparición de componentes MA estrictamente no invertibles. En estos casos, parece más viable formular el modelo en niveles; pero cuando las series son no estacionarias, puede ser necesario recurrir a un VAR de orden aún mayor que en el caso estacionario. Un modelo estructural ofrece, en cambio, una especificación más parsimoniosa.

Además, estos últimos modelos permiten imponer restricciones de cointegración y tomar en consideración variables explicativas (Harvey, 1989:9, 468-470). Dado que en un modelo estructural multivariante la tendencia es modelada explícitamente, los movimientos de corto plazo pueden ser capturados por una representación VAR de bajo orden y la cointegración surgirá cuando haya tendencias comunes. Así, aunque los modelos estructurales son un camino para presentar las características más notables de las series observadas, se convierten en modelos de comportamiento una vez que se empiezan a incorporar restricciones introduciendo comportamientos dinámicos de corto plazo e incluyendo variables exógenas en algunas de las ecuaciones.

Como señalan Durbin y Koopman (2001:52-53), seguramente, la principal desventaja de los modelos estructurales frente a los modelos ARIMA radica en que los primeros sólo se enseñan en relativamente pocas universidades, existe escasa bibliografía en el campo estadístico-econométrico y sólo desde hace poco tiempo está disponible un *software* específico. Por supuesto, todo lo contrario ocurre en relación con la metodología Box-Jenkins.

## **Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia**

## 2. Estacionalidad y funciones *splines*

Una vez expuesta la formulación básica de los modelos estructurales de series temporales y explicado el funcionamiento del filtro de Kalman para estimar los componentes típicos de una serie temporal en cada instante del tiempo, el presente capítulo se centra en el estudio del componente estacional.

La mayoría de las series temporales económicas presentan fluctuaciones estacionales en torno a la tendencia a largo plazo que explican una parte muy importante del comportamiento de las mismas; aunque, normalmente, esta estacionalidad o, mejor dicho, la extensión del periodo asociado a los ciclos estacionales que pueden existir, suele estar oculta por el grado de agregación en los datos disponibles. Así, si se dispone de series anuales es, evidentemente, imposible detectar si existe un comportamiento estacional. Por ejemplo, en el ámbito agrario, las series semanales son imprescindibles para poder conocer con cierta precisión las leyes que rigen el funcionamiento de los mercados; en el caso de los servicios de urgencias hospitalarios, las series diarias son necesarias para detectar en qué días de la semana y del año se producen los mayores desequilibrios entre la oferta y la demanda; y, finalmente, las series horarias parecen las más apropiadas para obtener un conocimiento de la demanda de electricidad con cierto grado de profundidad<sup>1</sup>. Esta circunstancia ayuda a explicar el interés por el estudio de datos de alta frecuencia<sup>2</sup>, un campo al que la investigación econométrica había prestado atención desde el punto de vista teórico, pero no tanto desde el punto de vista aplicado, sobre todo si se compara el número de trabajos que se han ocupado de series mensuales, semanales, diarias u horarias<sup>3</sup>. Como afirman Ghysels y Osborn, «... Hay una literatura

---

<sup>1</sup> En general, cuando se estudian variables que se toman como indicadores para adoptar ciertas medidas de política económica, la periodicidad de la observación es un elemento importante. Por ejemplo, el examen semanal de la oferta monetaria puede ser suficiente para adoptar medidas sobre el tipo de interés; pero un conocimiento diario permite una mejor regulación (*fine tuning*) de la situación de liquidez. Véase Cabrero y otros (2002:9).

<sup>2</sup> El análisis de series de alta frecuencia ayuda a descubrir ciertos patrones de comportamiento que podrían enriquecer los modelos teóricos de la Economía; sin embargo, en la mayor parte de estos modelos el tiempo es irrelevante.

<sup>3</sup> En este sentido, es interesante reproducir las siguientes palabras de Engle (2000:1): «Al entrar en el *Graduate School*, aprendí que T.C. Liu acababa de estimar el primer modelo trimestral de la economía de EE.UU., Liu (1963). En breve, volvió a sorprenderme publicando un modelo macroeconómico

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

emergente que trata datos observados con [...alta...] frecuencia, tales como los semanales, diarios o intradiarios ... Esta literatura está todavía en su infancia» (Ghysels y Osborn, 2001:1).

Los modelos estructurales permiten recoger el comportamiento estacional de muchas series temporales económicas a través de variables cualitativas o términos trigonométricos. Sin embargo, cuando las observaciones se registran con alta frecuencia —por ejemplo, semanal, diaria u horaria—, estos modelos no se caracterizan por su parsimonia y, además, es posible encontrar simultáneamente fluctuaciones estacionales de distinto periodo que difícilmente pueden aislarse<sup>4</sup>. Por ello, podría ser apropiado introducir una herramienta, funciones *splines*, que pueda ser empleada conjuntamente con las otras formulaciones convencionales para completar el modelo del componente estacional y que, sobre todo, permita obtener especificaciones más sencillas de este componente. Sin embargo, antes de exponer la propuesta que en este trabajo se realiza para el tratamiento de las fluctuaciones de corto plazo en series temporales de alta frecuencia, será preciso definir qué se entiende por estacionalidad y cómo se puede modelar. A continuación se señalan algunas interpretaciones de lo que puede identificarse como variaciones estacionales y, además, se destaca cómo ha ido cambiando la percepción de estas fluctuaciones en el ámbito de la Ciencia Económica: se ha pasado de una fase en la que la estacionalidad era un ruido que debía ser eliminado porque impedía observar con claridad las fluctuaciones tendenciales y cíclicas, a otra en la que proliferan los intentos de modelar la estacionalidad como causante de un porcentaje importante de la variación de una serie temporal. Se presentan como dos alternativas excluyentes, los modelos determinísticos y estocásticos de la estacionalidad. Estos últimos constituyen la formulación propia del enfoque estructural,

---

mensual, Liu (1969). Todavía no he visto un modelo macroeconómico semanal completo, pero sospecho que podría haber alguno entre los grupos de investigación privados o gubernamentales. En el campo de las finanzas ha tenido lugar una sucesión similar, desde el análisis de datos anuales a mensuales, a semanales, a diarios, y ahora existe un gran interés por los modelos intradiarios». En los últimos años, las series financieras *tick-by-tick*, denominadas por Engle (2000) como datos de *muy alta frecuencia* —*ultra-high-frequency*—, han recibido una atención creciente. Su rasgo más distintivo es que las observaciones se registran en intervalos de tiempo cuya longitud varía aleatoriamente, como si existiese un reloj diferente para cada uno de los eventos del fenómeno observado. Kitagawa (1984) efectúa una interesante aplicación de modelos en el espacio de los estados a datos irregularmente espaciados. Como ya se mencionó en la introducción, este tipo de datos no serán objeto de análisis en este trabajo. Algunos artículos para este tipo de datos en el campo financiero se recopilan en Dunis y Zhou (1998).

<sup>4</sup> En este sentido, la interacción entre variaciones periódicas observadas en series de alta frecuencia abre nuevos interrogantes entre los analistas, puesto que las operaciones de extracción de señales pueden alterar significativamente las propiedades de los datos en las distintas frecuencias muestrales (Maravall y del Río, 2001, y Ravn y Uhlig, 2002).

que admite como punto de partida la existencia de un patrón estacional cambiante, para desembocar en el planteamiento determinístico como caso particular. Se indican también algunas de las implicaciones de la desestacionalización y la conveniencia de modelar simultáneamente todos los componentes de una serie.

Desde principios del siglo pasado, se considera en economía que una serie temporal es el resultado de la agregación —aditiva o multiplicativa— de varios componentes no observables —tendencia, ciclo, estacionalidad y componente irregular— con diferentes periodicidades. A partir de ese momento, numerosas investigaciones se han dedicado a separar el componente estacional del resto de la serie, con objeto de centrar el análisis en la tendencia y el ciclo, tradicionalmente considerados como los componentes más importantes en el análisis económico. El motivo de esta práctica era doble. Por un lado, se pensaba que no existían interacciones entre ciclo estacional y ciclo de negocios; por otro, no se consideraba que las fluctuaciones estacionales pudieran provocar incrementos de bienestar.

Sin embargo, y tal como afirma Sims (1993:9), «en un mundo donde las series temporales muestran claras fluctuaciones estacionales, agentes racionales tendrán en cuenta aquellas fluctuaciones para planear su propio comportamiento. Usar datos ajustados estacionalmente para modelar el comportamiento de tales agentes es malgastar información e introducir posiblemente graves sesgos». Sin embargo, sigue diciendo el autor, aplicar técnicas estándar no refinadas con datos sin ajustar puede dar peores resultados que si se trabaja con datos ajustados estacionalmente. Por supuesto, y a pesar de que hayan argumentos a favor de trabajar con datos ajustados estacionalmente, no puede olvidarse que la desestacionalización no es una transformación neutral con respecto a la estimación de los comportamientos a largo plazo ni, obviamente, de los de corto plazo.

Afortunadamente, cada vez se acepta con mayor unanimidad entre los economistas que la estacionalidad es responsable de una parte importante de la variabilidad mostrada por una serie temporal. De ahí que hayan proliferado en los últimos años nuevos modelos de la estacionalidad y, sobre todo, se haya producido un cambio de orientación, prestándose mayor atención al objetivo de especificar la estacionalidad, frente al objetivo de eliminarla para evitar que contamine a otros elementos.

Aunque sea posible encontrar en la literatura definiciones de los distintos componentes, es interesante destacar que suelen hacer referencia a las variaciones

## **Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia**

estacionales como fluctuaciones perfectamente periódicas y regulares. Si se analizan las causas de la estacionalidad, tal vez pueda entenderse que el componente estacional que aparece en una serie temporal sea más o menos estable. Suelen considerarse cuatro aspectos como factores motivantes de la presencia de estacionalidad en series temporales económicas<sup>5</sup>:

Efectos de calendario. La localización de los días festivos a lo largo del año puede ser una fuente exógena de variabilidad de una serie temporal. Por ejemplo, el cierre de los mercados mayoristas europeos en algunas fechas de las fiestas navideñas, se deja sentir en las exportaciones de tomate canario y en las cotizaciones alcanzadas por el fruto. También se producen fluctuaciones en estas magnitudes en torno a la festividad de Semana Santa. Los servicios de urgencias cubren la demanda sanitaria existente fuera de la jornada de trabajo de un día laborable o durante los días festivos; de ahí que los días de máxima demanda de estos servicios se concentran en los fines de semana o días festivos. Es claro, por otra parte, el efecto que tiene la jornada laborable y el carácter festivo o no de un determinado día en la programación de las actividades profesionales y, por tanto, en la demanda de energía eléctrica.

Clima. Especialmente en producciones agrarias, el clima condiciona la distribución intra-anual de las cosechas e incluso puede determinar el periodo de producción. En el caso particular de la exportación de tomate, el clima genera un vacío en la producción norte-europea y el hueco estacional es aprovechado por exportadores canarios, peninsulares y marroquíes. Además, las temperaturas frías en destino restringen la demanda de tomates y también la oferta de los países del Norte, mientras que, por ejemplo, el viento en Canarias reduce la oferta en origen. Por otro lado, los excesos de demanda observados en los centros de urgencias hospitalarios en determinados días de los meses de invierno, motivados, en parte, por los cambios en la morbilidad, pueden responder, entre otros factores, a cambios climáticos. Finalmente, las variaciones de la demanda de energía eléctrica según la época del año pueden ser explicadas, al menos parcialmente, por las oscilaciones de temperatura entre estaciones.

Decisiones sobre la temporización de las actividades. Las pagas extras en julio y diciembre o la declaración del impuesto sobre la renta en mayo y junio pueden dejarse sentir sobre determinados gastos. Las fechas elegidas para el periodo vacacional, centradas en el mes de agosto, tienen efectos sobre el nivel de actividad empresarial o la intensidad de la circulación en los lugares de origen y destino turístico. Además, no hay

---

<sup>5</sup> Granger (1978), Hylleberg (1992) y Miron (1994).

que olvidar el efecto en cadena de estas pautas de comportamiento sobre otras actividades como la sanitaria. Por ejemplo, y dependiendo del lugar en el que esté ubicado un centro de urgencias, el periodo vacacional puede provocar una disminución o aumento en la afluencia de pacientes a estos centros. Respecto a la demanda de energía eléctrica, quizás cabría mencionar el cambio de hora establecido por el gobierno con objeto de que la actividad del país aproveche en mayor medida la luz solar y permita el consecuente ahorro de consumo. En la actividad exportadora de tomate, es obvio que el momento de plantación determina el desarrollo de la planta y, en suma, el ciclo de cultivo a lo largo de la campaña. En este mismo sentido, la existencia de calendarios de aplicación de los precios de referencia determinó durante años la duración de la campaña canaria<sup>6</sup> y los cambios en la normativa al respecto han conducido a la ampliación de la zafra.

Expectativas. Los agentes económicos reaccionan anticipándose a determinados fenómenos y esas reacciones pueden motivar fluctuaciones estacionales de determinadas variables. Así, el incremento previsto de compras de ciertos artículos en épocas muy definidas del año suele conducir a incrementos de los gastos en publicidad en fechas próximas. De igual modo, la previsión de los precios es un factor relevante a la hora de planificar la distribución de la exportación de tomates canarios a Europa a lo largo de la campaña.

Algunas de las anteriores causas de estacionalidad son exógenas y además, más o menos fijas de un año a otro y, por tanto, pueden dar lugar a un patrón estacional estable. Pero parece mucho más difícil asumir que fuentes cambiantes de estacionalidad produzcan siempre la misma variación estacional. Incluso, aunque la causa de la estacionalidad se presente todos los años con la misma intensidad y distribución anual, puede ocurrir que el efecto sobre la estacionalidad de una serie temporal varíe. Por ejemplo, los avances tecnológicos hacen que, aunque las condiciones climáticas se repitan cada año, el impacto estacional sea cada vez menor. Las innovaciones técnicas en invernaderos han contribuido a acentuar el solapamiento de las producciones de tomate de los países del norte y sur de Europa. De la misma forma, el incremento en la eficiencia de la maquinaria de calefacción o aire acondicionado puede amortiguar las variaciones en el consumo de energía eléctrica asociadas a las temperaturas extremas. Las expectativas pueden también modificar el patrón estacional. Si un profesional liberal espera una época de inestabilidad en sus ingresos, tratará, posiblemente, de

---

<sup>6</sup> Véase Cáceres (2000).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

estabilizar su consumo. Asimismo, las previsiones juegan un papel crucial en la programación de los exportadores canarios de tomate: los cambios en la normativa durante el periodo de integración en la UE han generado expectativas de mayor rentabilidad que, no sólo se han traducido en un crecimiento del volumen global ofrecido, sino también en la extensión de la zafra al final de campaña. En definitiva, el cambio en el patrón estacional con el transcurso del tiempo es un fenómeno constatado y, de hecho, reconocido desde el siglo pasado.

Teniendo en cuenta que la estacionalidad no tiene que ser perfectamente periódica y estable, puede adoptarse la definición de Hylleberg (1992:4), según la cual «la estacionalidad es el movimiento intra-anual sistemático, aunque no necesariamente regular, causado, directamente, por los cambios en el clima, el calendario y la temporización de las decisiones, o, indirectamente, a través de las decisiones de producción y consumo hechas por los agentes de la economía. Estas decisiones son influidas por las dotaciones, las expectativas y las preferencias de los agentes, y por las técnicas de producción disponibles en la economía».

Una definición formal, desde el punto de vista estadístico, de estacionalidad, puede obtenerse utilizando algunos conceptos del análisis espectral, que permite descomponer la variabilidad en el tiempo de una serie en una suma de movimientos cíclicos de diferente periodo y frecuencia. Si se divide el periodo en el que se completa la fluctuación estacional en  $s$  unidades de tiempo (estaciones), en cada una de las cuales se tiene una observación del proceso, se definen las frecuencias estacionales del siguiente modo: en radianes por unidad de tiempo, como  $\mathbf{q}_j = \frac{2\mathbf{p}}{s} j$ ,  $j = 1, \dots, \lfloor s/2 \rfloor$ ; o

en ciclos por unidad de tiempo, como  $f_j = \frac{\mathbf{q}_j}{2\mathbf{p}} = \frac{j}{s}$ ,  $j = 1, \dots, \lfloor s/2 \rfloor$ . Estas frecuencias corresponden a todos los ciclos de periodo inferior a aquél en el que se completa la variación estacional que pueden detectarse con  $s$  observaciones por cada periodo estacional. Entonces, se dice que un proceso estocástico posee componente estacional si su función de densidad espectral presenta picos alrededor de las frecuencias estacionales<sup>7</sup>. En otras palabras, se dice que un proceso es estacional si las fluctuaciones asociadas a las frecuencias estacionales explican una parte importante de la variabilidad del proceso. Esta definición indica si el proceso es estacional, pero no especifica dicho componente. Por otro lado, Hylleberg (1986) advierte que la estimación del espectro

puede ser complicada y no aportar demasiada información sobre la existencia de estacionalidad y sobre el tipo de estacionalidad presente en los datos. Además, este autor considera que existen tantas vertientes en la naturaleza de la estacionalidad que es difícil aplicar una definición precisa, resultando más factible proporcionar una definición en términos amplios y considerar los diferentes modelos posibles de este fenómeno, cada uno de los cuales recurre a un tipo específico de variación estacional<sup>8</sup>.

Las fluctuaciones estacionales pueden ser de distinta naturaleza. Una serie económica observada en el tiempo puede presentar oscilaciones en torno a su tendencia más o menos estables. Por ejemplo, si se están estudiando las exportaciones semanales de tomate de Canarias a la UE, se sabe que en las semanas del verano no existe casi exportación, que el grueso de las exportaciones se concentra en las semanas entre noviembre y marzo y que el máximo se alcanza en torno a los meses de enero-febrero, y se trata de fenómenos que se repiten campaña tras campaña. Pero en diversos trabajos se ha incidido en el hecho de que las variaciones estacionales no son tan regulares como habitualmente se supone. También existen fluctuaciones estacionales aleatorias o estocásticas, con una aparente interpretación menos clara, pero tan importantes como las anteriores y difícilmente separables de éstas. Por ejemplo, si se estudia el número diario de visitas registradas en un servicio de urgencias, está claro que se repite un patrón de máxima afluencia durante los fines de semana, pero la magnitud de las variaciones no es la misma semana tras semana a lo largo del tiempo. Si, en cambio, se analiza la demanda horaria de energía eléctrica puede comprobarse que las curvas de carga diarias presentan momentos de máximo y mínimo consumo en torno a las mismas horas del día, pero, también, con distinta intensidad a lo largo del tiempo.

Puede considerarse, pues, que existen dos tipos básicos de estacionalidad. Se dice que la estacionalidad es determinística si las variaciones estacionales son fluctuaciones periódicas perfectamente regulares, es decir, estables o no cambiantes en el tiempo<sup>9</sup>; en ese caso, tales variaciones pueden recogerse correctamente a partir de *dummies* estacionales. Por otro lado, también puede suceder que la estacionalidad no sea exógena al comportamiento de la serie estudiada, sino que la magnitud de la variación

---

<sup>7</sup> Nerlove (1964) y Granger (1978).

<sup>8</sup> Hylleberg (1992) realiza una revisión histórica extensa del tratamiento de series temporales con estacionalidad. En la actualidad, existe un buen número de trabajo en los que se discuten los diferentes modelos alternativos. Véase, entre otros, Hylleberg (1986, 1992, 1994a, 1994b), Canova (1992), Franses (1996a, 1996b, 1997), Ghysels (1994), Miron (1994, 1996), Sansó (1996) y Ghysels y Osborn (2001).

<sup>9</sup> Véase, entre otros, Beaulieu y Miron (1992), Miron (1996) y Miron y Beaulieu (1996).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

estacional presente un carácter cambiante a lo largo de la muestra. Si el patrón estacional del proceso varía aleatoriamente, la formulación convencional a partir de variables cualitativas estacionales no permite recoger esa variabilidad y resultan más adecuadas otras formulaciones. Como ya se comentó en el capítulo anterior, una forma de conseguir que el patrón estacional evolucione en el tiempo es admitir que la suma de los efectos estacionales a lo largo del periodo estacional sea igual a un término de perturbación. Otra posibilidad es permitir que el efecto de cada estación evolucione de acuerdo a un camino aleatorio<sup>10</sup>. Una alternativa a esta última posibilidad la constituye la formulación trigonométrica estocástica, también definida en el capítulo precedente<sup>11</sup>. Finalmente, habría que señalar que la imposición de un tipo de estacionalidad cuando es otra la que está presente puede resultar en serios sesgos o pérdidas de información; de ahí la importancia de su correcta especificación<sup>12</sup>.

Cuando el objetivo del análisis se centra en el componente estacional de series de alta frecuencia, en las que el número de estaciones que configuran el periodo estacional puede llegar a ser considerable, las denominadas funciones *splines*, tanto determinísticas como estocásticas, constituyen una formulación de potencial significativo. Dichas funciones pueden introducirse en un modelo estructural de series temporales como un componente inobservable más, y en este contexto se han mostrado como una forma eficiente de tratar los movimientos periódicos que cambian de forma suave.

Las funciones *splines*, que constituyen una especificación más parsimoniosa que las anteriormente citadas y que se pueden emplear conjuntamente con éstas para completar la descripción del componente estacional, determinan la variación estacional registrada en una observación determinada de la serie como función de la estación a la que pertenece dicha observación. En particular, si se asume que la estacionalidad no

---

<sup>10</sup> Esta formulación, propuesta por Harrison y Stevens (1976), se identificará, de ahora en adelante, como modelo de *dummies* estocásticas.

<sup>11</sup> Otra forma de modelar la estacionalidad que ha adquirido cierta relevancia son los procesos ARMA periódicos, procesos PARMA, que admiten diferentes parámetros en el modelo ARMA para cada estación (Tiao y Grupe, 1980, Osborn, 1991, Franses, 1996b, y Ghysels y Osborn, 2001, entre otros). Es decir, a partir de una serie con  $s$  observaciones por año, pueden construirse  $s$  series, una por estación, y especificar un VAR. Esta modelización presenta problemas derivados de la dimensionalidad y, por consiguiente, la necesidad de establecer restricciones de nulidad. Estos inconvenientes se manifestarían explícitamente si se empleara este enfoque para modelar el componente estacional de las series semanales, diarias y horarias analizadas en los capítulos siguientes o, en general, de series de alta frecuencia.

<sup>12</sup> Sólo se están considerando las opciones para modelar la estacionalidad en el contexto de los modelos estructurales. Pero, por supuesto, también existen formulaciones determinísticas y estocásticas

contiene variaciones estocásticas, es decir, si  $g_t = g_j$  cuando la observación registrada en el momento  $t$  pertenece a la estación  $j$ , resulta conveniente disponer de una función  $g(j)$  tal que  $g_j = g(j) + e_j$ , siendo  $e_j$  un término de error asociado al ajuste<sup>13</sup>. Ahora bien, puede ser complicado encontrar una forma funcional  $g(j)$  —lineal, parabólica, cúbica, ...— que recoja las variaciones estacionales de las distintas estaciones manteniendo la constancia de los parámetros. Por ejemplo, si se elige una especificación lineal, puede recurrirse a variables cualitativas que permitan alterar la constante o la pendiente de la línea de regresión en tramos correspondientes a conjuntos de estaciones. Sin embargo, estas variables cualitativas producen, generalmente, rupturas en la línea de regresión, rupturas que aparecen también con especificaciones no lineales.

Las funciones *splines* constituyen un recurso para suavizar estas rupturas y reconstruir las líneas segmentadas por tramos de modo que se elimine el artificial y, seguramente, inapropiado salto en la línea de regresión<sup>14</sup>. En general, las funciones *splines* son funciones polinómicas definidas a trozos a las que se imponen restricciones que permitan transiciones sin saltos y más suaves cuanto mayor sea el grado de los polinomios.

Además, en series de alta frecuencia suelen coexistir fluctuaciones de distinto periodo que deben ser tenidas en cuenta para especificar correctamente la estacionalidad. Este hecho dificulta la apreciación de cada una de las fluctuaciones periódicas y puede obligar, como se ilustra en los capítulos siguientes, a efectuar algún proceso previo de filtrado que resulta muy sensible a la especificación de la naturaleza de los componentes de la serie. En particular, si se pretende eliminar el componente tendencial para obtener luego una aproximación a la fluctuación estacional a partir de las medias por estación, la especificación de un componente de nivel estocástico puede conducir a la eliminación de una parte significativa de tal variación estacional, que resulta absorbida, precisamente, como consecuencia de la posibilidad concedida al nivel

---

enmarcadas en el ámbito de los modelos ARIMA, siguiendo las ideas de Box y Jenkins (1976) y Bell y Hillmer (1983).

<sup>13</sup> El supuesto de estacionalidad determinística se introduce sólo con ánimo de ilustrar el argumento que se expone de forma simple. Más adelante se abordará la situación en que la estacionalidad es estocástica.

<sup>14</sup> Este tipo de funciones se explica con rigor en Suits y otros (1978) y Smith (1979). Y han sido utilizadas, por ejemplo, para determinar el tipo impositivo óptimo del impuesto sobre la renta cuando las diferencias en los niveles de renta se deben a factores aleatorios (Strawczynski, 1998) o a habilidades no observadas de los individuos (Slemrod y otros, 1994).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

de adaptarse a los cambios operados en la serie entre un instante del tiempo y el siguiente.

Por otra parte, y del mismo modo que la percepción de la tendencia en series económicas puede alterarse si se ignora la presencia de la fluctuación cíclica, esta misma variación puede incidir, a su vez, en la identificación de la variación estacional; y más factible aún resulta la incorrecta evaluación de la magnitud de una fluctuación estacional si otras fluctuaciones de periodo más corto no son especificadas explícitamente. Por ejemplo, en el caso de la demanda horaria de energía eléctrica coexisten fluctuaciones estacionales de periodo anual, semanal y diario. Si el patrón estacional diario se repite cada uno de los días del año, pero en la especificación del modelo no se tuvieron en cuenta las fluctuaciones semanal y anual, es muy posible que se llegara a la errónea conclusión de que la fluctuación estacional diaria evoluciona a lo largo de la muestra. Esta circunstancia obliga a recurrir a una especificación que permita la estimación simultánea de todas las variaciones estacionales presentes, que pueden introducirse como determinísticas o estocásticas. En este sentido, las formulaciones convencionales pueden combinarse con las *splines*, que suponen, en general, un ahorro computacional significativo y pueden, de hecho, ser la única vía posible de estimación.

En los epígrafes siguientes se exponen con detalle las características de las funciones *splines* en el ámbito del análisis de series temporales y, más concretamente, se describen los distintos procedimientos a través de los cuales estas funciones pueden utilizarse como herramienta adecuada para recoger fluctuaciones periódicas de naturaleza determinística y también estocástica en el contexto de los modelos estructurales. El capítulo concluye señalando algunos campos de aplicación que se abordarán en los capítulos siguientes y que servirán para ilustrar las ventajas metodológicas de las funciones *splines* frente a otras especificaciones de las variaciones estacionales.

### 2.1. Funciones *splines*

Las funciones *splines* constituyen una técnica de interpolación estable y flexible esencialmente utilizada para describir el patrón de un conjunto de datos puntuales generados a partir de una función desconocida particular. En el caso de que la función desconocida sea función del tiempo, la *spline*  $g(t)$  aproxima los valores  $y_i$  generados por la función desconocida  $f(t)$  mediante varios polinomios, cada uno definido sobre

un subintervalo, que se unen entre sí obedeciendo a ciertas condiciones de continuidad en el conjunto de pares de valores  $\{(t_0, y_{t_0}), \dots, (t_k, y_{t_k})\}$ , que se denomina el conjunto de nodos asociados a la rejilla  $\{t_0, \dots, t_k\}$ <sup>15</sup>. El método de la aproximación *spline* proporciona una línea suavizada definida como  $g(t) = g_i(t)$ ,  $t_{i-1} \leq t \leq t_i$ , donde  $g_i(t) = g_{i,0} + g_{i,1}t + \dots + g_{i,p}t^p$ ,  $i = 1, \dots, k$ <sup>16</sup>. Se dice que la función *spline* es de orden  $p$  cuando las  $p$  primeras derivadas existen y las  $(p-1)$  primeras son continuas. Las condiciones de continuidad en los puntos que delimitan los tramos se refieren, precisamente, a la continuidad de la función,  $g_i(t_i) = g_{i+1}(t_i)$ ,  $i = 1, \dots, k-1$ , y de sus  $(p-1)$  primeras derivadas,  $\nabla g_i(t_i) = \nabla g_{i+1}(t_i)$ ,  $\nabla^2 g_i(t_i) = \nabla^2 g_{i+1}(t_i)$ , ...,  $\nabla^{p-1} g_i(t_i) = \nabla^{p-1} g_{i+1}(t_i)$ ,  $i = 1, \dots, k-1$ <sup>17</sup>. El valor de  $p$  puede ser interpretado como el orden del alisado: cuanto mayor sea  $p$ , más alisada será la función *spline*. Sin embargo, una función *spline* con un valor de  $p$  relativamente pequeño también da lugar a una línea suavizada. Las *splines* lineales, en las que los polinomios que intervienen son siempre de grado uno, no suelen ser aproximaciones muy eficientes y en estos casos es preferible acudir a aproximaciones polinomiales de mayor orden, cuadráticas o cúbicas, siendo esta últimas la opción más frecuente<sup>18</sup>. Aunque a priori se ha considerado que los polinomios que intervienen en la formulación de la *spline* son todos del mismo grado en

<sup>15</sup> En los modelos de series temporales, los nodos se corresponden con pares de puntos cuyas coordenadas son los instantes del tiempo considerados y los valores de la magnitud estudiada,  $y_t$ , correspondientes a dichos instantes del tiempo. En general,  $t_0$  es el instante del tiempo correspondiente a la primera observación de la serie a la que se ajusta la *spline* y  $t_k$  es el instante del tiempo correspondiente a la última.

<sup>16</sup> Nótese que si se desea obtener una aproximación polinomial a una función desconocida en un intervalo determinado, generalmente es mucho más eficiente dividir el intervalo en  $k$  partes y utilizar en cada una un polinomio de grado  $n$  que, por el contrario, recurrir a un único polinomio de orden  $kn$  (de Boor, 1978:34-35).

<sup>17</sup> Nótese que si la función *spline* es lineal, puede imponerse la restricción de continuidad de la función; pero no de sus primeras derivadas, ya que, en este último caso, se estaría obligando a que la función lineal fuera la misma en cada tramo. Por la misma razón, en el caso de funciones parabólicas debe admitirse la discontinuidad de la derivada segunda; y en el caso de recurrir a polinomios de grado tres no puede imponerse la continuidad de la derivada tercera.

<sup>18</sup> Las funciones *splines* cúbicas pueden tener muchas aplicaciones en el campo de la Estadística. Koopman (1992), además de las aplicaciones específicas en el contexto de los modelos de series temporales, analiza tres aplicaciones diferentes: a) obtención de un polígono de frecuencias a partir de un histograma de frecuencias; b) regresiones por trozos (Poirier, 1973), enfoque especialmente útil para aproximar efectos no lineales en modelos de regresión; y c) regresiones a trozos con funciones *splines* cúbicas en modelos de regresión utilizando retardos distribuidos. Una aplicación concreta de estas funciones en el campo de la economía es la realizada por Lotric y Dobnikar (1999). Cabrero y otros (2002) encuentran para series de oferta monetaria en la UE, que las predicciones obtenidas con modelos estructurales que incorporan funciones *splines* cúbicas complementan adecuadamente a las predicciones obtenidas con modelos ARIMA, y concluyen que estos últimos no pueden capturar completamente ciertos patrones estacionales, que tal vez requieren la no linealidad que introducen las *splines*.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

los diferentes tramos, en ocasiones conviene relajar este supuesto y admitir polinomios de distinto grado si, con ello, se consigue una mejor aproximación<sup>19</sup>. En todo caso, en los párrafos siguientes se exponen especificaciones de estas funciones en las que los polinomios que intervienen son siempre del mismo grado.

El caso más sencillo lo constituye una *spline* lineal, en la que  $g_i(t) = g_{i,0} + g_{i,1}t$ ,  $t_{i-1} \leq t \leq t_i$ , para  $i = 1, \dots, k$ , es decir,

$$g(t) = \begin{cases} g_{1,0} + g_{1,1}t, & t_0 \leq t \leq t_1 \\ g_{2,0} + g_{2,1}t, & t_1 \leq t \leq t_2 \\ \vdots & \vdots \\ g_{k-1,0} + g_{k-1,1}t, & t_{k-2} \leq t \leq t_{k-1} \\ g_{k,0} + g_{k,1}t, & t_{k-1} \leq t \leq t_k \end{cases}$$

La función anterior está sujeta a las restricciones de continuidad

$$g_{i,0} + g_{i,1}t_i = g_{i+1,0} + g_{i+1,1}t_i, \quad i = 1, \dots, k-1.$$

En la *spline* lineal la primera derivada es, en general, discontinua en los puntos de ruptura entre dos tramos consecutivos de modo que la transición de uno a otro no es suave, aunque no se produce salto (gráfico 2.1.1)<sup>20</sup>. Una transición menos brusca se puede conseguir incrementando el grado de los polinomios en cada uno de los subintervalos e imponiendo condiciones de continuidad adicionales. Por ejemplo, a las *splines* cuadráticas, definidas por

$$g(t) = \begin{cases} g_{1,0} + g_{1,1}t + g_{1,2}t^2, & t_0 \leq t \leq t_1 \\ g_{2,0} + g_{2,1}t + g_{2,2}t^2, & t_1 \leq t \leq t_2 \\ \vdots & \vdots \\ g_{k-1,0} + g_{k-1,1}t + g_{k-1,2}t^2, & t_{k-2} \leq t \leq t_{k-1} \\ g_{k,0} + g_{k,1}t + g_{k,2}t^2, & t_{k-1} \leq t \leq t_k \end{cases}$$

se les imponen las restricciones de continuidad de la función,

$$g_{i,0} + g_{i,1}t_i + g_{i,2}t_i^2 = g_{i+1,0} + g_{i+1,1}t_i + g_{i+1,2}t_i^2, \quad i = 1, \dots, k-1,$$

y, también, restricciones de continuidad de la primera derivada,

$$g_{i,1} + 2g_{i,2}t_i = g_{i+1,1} + 2g_{i+1,2}t_i, \quad i = 1, \dots, k-1.$$

Estas *splines* cuadráticas pueden ser también demasiado restrictivas en términos de ajuste, debido a que el carácter convexo o cóncavo de la curvatura de la función no puede cambiar en el intervalo definido por dos nodos consecutivos (gráfico 2.1.1).

Finalmente, a las *splines* cúbicas, definidas por

<sup>19</sup> Esta combinación de diferentes grados de alisado da lugar a lo que se conoce como funciones *splines* híbridas.

<sup>20</sup> Con objeto de explicar los factores que determinan los beneficios de la industria aeroespacial norteamericana en el periodo 1951-1971, Poirier y Garber (1974) plantean un modelo de regresión en el que la dimensión temporal es representada por una función *spline* lineal de tres tramos de modo que pueda evaluarse el posible efecto de los conflictos bélicos durante el periodo estudiado.

$$g(t) = \begin{cases} g_{1,0} + g_{1,1}t + g_{1,2}t^2 + g_{1,3}t^3, & t_0 \leq t \leq t_1 \\ g_{2,0} + g_{2,1}t + g_{2,2}t^2 + g_{2,3}t^3, & t_1 \leq t \leq t_2 \\ \vdots & \vdots \\ g_{k-1,0} + g_{k-1,1}t + g_{k-1,2}t^2 + g_{k-1,3}t^3, & t_{k-2} \leq t \leq t_{k-1} \\ g_{k,0} + g_{k,1}t + g_{k,2}t^2 + g_{k,3}t^3, & t_{k-1} \leq t \leq t_k \end{cases}$$

se les imponen las restricciones de continuidad de la función,

$$g_{i,0} + g_{i,1}t_i + g_{i,2}t_i^2 + g_{i,3}t_i^3 = g_{i+1,0} + g_{i+1,1}t_i + g_{i+1,2}t_i^2 + g_{i+1,3}t_i^3, \quad i = 1, \dots, k-1,$$

restricciones de continuidad de la primera derivada,

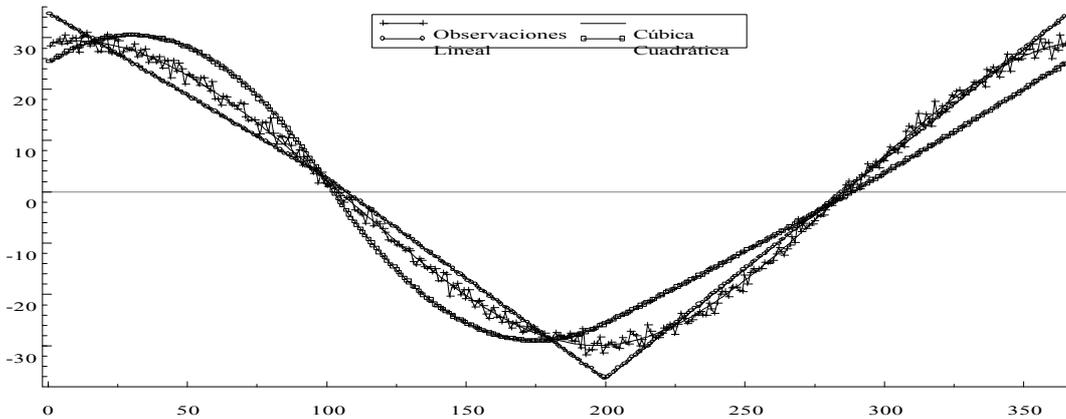
$$g_{i,1} + 2g_{i,2}t_i + 3g_{i,3}t_i^2 = g_{i+1,1} + 2g_{i+1,2}t_i + 3g_{i+1,3}t_i^2, \quad i = 1, \dots, k-1,$$

y restricciones de continuidad de la segunda derivada,

$$2g_{i,2} + 6g_{i,3}t_i = 2g_{i+1,2} + 6g_{i+1,3}t_i, \quad i = 1, \dots, k-1.$$

Estas últimas funciones no tienen el comportamiento oscilatorio característico de las funciones polinomiales de mayor grado (gráfico 2.1.1). En general, la observación de los valores  $y_i$  a los que se pretende ajustar una *spline* proporciona información útil para determinar, al menos de forma tentativa, el grado de los polinomios y las restricciones de continuidad que parezcan adecuadas.

Gráfico 2.1.1. Funciones *splines* lineal, cuadrática y cúbica



Además de elegir la forma funcional en cada segmento, la especificación de la *spline* exige fijar los puntos que delimitan estos segmentos y determinar su número<sup>21</sup>. En ocasiones, la localización de estos puntos es bastante clara y se introduce *a priori* como elemento de la especificación del modelo; en otros casos, los parámetros de localización no son fácilmente identificables y se deben introducir como parámetros

desconocidos que tendrán que ser estimados. En cualquier caso, debe tenerse en cuenta que cuanto mayor es el grado de los polinomios, menor será el número de puntos de ruptura necesarios para conseguir un buen ajuste<sup>22</sup>. Y de todas formas, la especificación de una *spline* debe ser el resultado de un compromiso entre el deseo de ajustarse a los datos y el deseo de obtener una función alisada<sup>23</sup>.

Aún asumiendo que los parámetros  $\{t_0, \dots, t_k\}$  son conocidos, es obvio que las restricciones de continuidad que se imponen a las funciones *splines* no son suficientes para que los parámetros que las definen queden determinados. Por ejemplo, en el caso de la *spline* cúbica las  $3(k-1)$  restricciones determinan un sistema de  $3(k-1)$  ecuaciones con  $4k$  parámetros que deben ser estimados o determinados imponiendo nuevas condiciones. En este sentido, existen dos planteamientos diferentes. Algunos autores sugieren imponer restricciones que signifiquen el paso por los nodos (véase de Boor, 1978, Koopman, 1992, y Harvey, Koopman y Riani, 1997)<sup>24</sup>; otros, en cambio, optan por estimar los parámetros que quedan libres (véase Marsh, 1983, 1986, Marsh y otros, 1990, y Marsh y Cormier, 2001). Los procedimientos propios de ambas aproximaciones se detallan en los apartados siguientes<sup>25</sup>.

### 2.1.1 Procedimientos que no imponen el paso por los nodos

Sin exigir que la *spline* pase por los valores observados de la serie en los puntos de ruptura —y tomando como base los trabajos de Fuller (1969), Poirier (1973, 1974, 1975, 1976) y Buse y Lim (1977)—, Suits y otros (1978) y Smith (1979) describen de forma bastante clara el procedimiento para estimar las funciones de regresión *splines*

---

<sup>21</sup> Como advierte de Boor (1978:22), la bondad del ajuste de la aproximación polinomial a una función determinada es muy sensible a la localización de estos puntos.

<sup>22</sup> La aparición de puntos de inflexión en la función que se desea aproximar no recogidos por la *spline* constituye un indicio de que es preciso añadir nuevos nodos (de Boor, 1978:303-304). También debe considerarse que incrementar el número de puntos de ruptura mejora siempre el ajuste, pero puede empeorar la capacidad predictiva. Véase, por ejemplo, Cabrero y otros (2002:17).

<sup>23</sup> Esta idea fue originalmente planteada por Whittaker (1923) y también considerada por Schoenberg (1964). Reinsch (1967) también parte de este planteamiento para la construcción de *splines* cúbicas.

<sup>24</sup> En interpolación por polinomios definidos a trozos, los puntos de interpolación y los puntos de ruptura no tienen que coincidir necesariamente. Algunos autores sugieren utilizar puntos de interpolación localizados en el punto medio del intervalo entre dos puntos de ruptura consecutivos. Véase, por ejemplo, de Boor (1978:73-80) para el caso de una *spline* cuadrática.

<sup>25</sup> Como afirma de Boor (1978:232), «... no hay mucha diferencia entre interpolación y aproximación mínimo cuadrática, si se sabe qué puntos de interpolación elegir. En cierta forma, la aproximación mínimo cuadrática es una manera muy razonable de elegir los puntos de interpolación». Además, esta última aproximación parece más adecuada para extraer una función alisada a partir de información ruidosa.

cuando los puntos de ruptura son conocidos<sup>26</sup>. Si éste fuera el caso, los parámetros libres podrían ser estimados en un modelo de regresión en el que se especifica la forma funcional de la *spline* y se incorporan las restricciones de continuidad impuestas.

En el caso de una *spline* lineal y suponiendo que los puntos de ruptura son conocidos, se dispone de un sistema de  $k-1$  ecuaciones (restricciones de continuidad de la función lineal) con  $2k$  valores desconocidos, de modo que  $k+1$  parámetros quedan libres y podrían ser estimados. Sea  $G_{(k-1) \times 1} = (g_{2,1}, g_{3,1}, \dots, g_{k-1,1}, g_{k,1})'$  el vector de parámetros expresado como función de los parámetros finalmente estimados, es decir,  $G_{(k+1) \times 1}^* = (g_{1,0}, g_{1,1}, g_{2,0}, g_{3,0}, \dots, g_{k-1,0}, g_{k,0})'$ <sup>27</sup>. En términos matriciales, el problema consiste en obtener  $G$  sabiendo que  $TG = R$ , donde

$$T_{(k-1) \times (k-1)} = \begin{bmatrix} -t_1 & 0 & \cdots & 0 & 0 \\ t_2 & -t_2 & \cdots & 0 & 0 \\ 0 & t_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & t_{k-1} & -t_{k-1} \end{bmatrix}$$

y

$$R_{(k-1) \times 1} = \begin{bmatrix} -g_{1,0} - g_{1,1}t_1 + g_{2,0} \\ -g_{2,0} + g_{3,0} \\ -g_{3,0} + g_{4,0} \\ \vdots \\ -g_{k-1,0} + g_{k,0} \end{bmatrix}$$

de modo que los parámetros del vector  $G$  pueden obtenerse teniendo en cuenta que  $G = T^{-1}R$ . Si se denota por  $a_{i,j}$  al elemento de la fila  $i$ -ésima y la columna  $j$ -ésima de la inversa de la matriz  $T$ , es decir,

$$T_{(k-1) \times (k-1)}^{-1} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k-1} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k-1,1} & a_{k-1,2} & \cdots & a_{k-1,k-1} \end{bmatrix}$$

y  $g_j$ ,  $j = 1, \dots, k-1$ , tal que

$$g_1 = g_{2,1}, g_2 = g_{3,1}, g_3 = g_{4,1}, \dots, g_{k-2} = g_{k-1,1}, g_{k-1} = g_{k,1};$$

<sup>26</sup> Por ejemplo, en el caso estudiado por Strawczynski (1998), relativo a la recaudación por el impuesto sobre la renta, el día en que cualquier cambio impositivo empieza a ser efectivo constituye un ejemplo de localización conocida de antemano.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

entonces,

$$g_j = a_{j,1}(-g_{1,0} - g_{1,1}t_1 + g_{2,0}) + a_{j,2}(-g_{2,0} + g_{3,0}) + \dots + a_{j,k-1}(-g_{k-1,0} + g_{k,0})$$

para  $j = 1, \dots, k-1$ ; y, reordenando términos,

$$g_j = \mathbf{a}_{j,0} g_{1,0} + \mathbf{a}_{j,1} g_{1,1} + \sum_{i=2}^k \mathbf{a}_{j,i} g_{i,0}, \quad j = 1, \dots, k-1,$$

donde

$$\begin{aligned} \mathbf{a}_{j,0} &= -a_{j,1} \\ \mathbf{a}_{j,1} &= -a_{j,1}t_1 \\ \mathbf{a}_{j,2} &= a_{j,1} - a_{j,2} \\ \mathbf{a}_{j,3} &= a_{j,2} - a_{j,3} \\ &\vdots \\ \mathbf{a}_{j,k-1} &= a_{j,k-2} - a_{j,k-1} \\ \mathbf{a}_{j,k} &= a_{j,k-1} \end{aligned}$$

La *spline* puede especificarse en función de los parámetros del vector  $G^*$  de la siguiente forma. Si  $t_0 \leq t \leq t_1$ :

$$g(t) = g_{1,0} + g_{1,1}t$$

Si  $t_{i-1} \leq t \leq t_i$ ,

$$\begin{aligned} g(t) &= g_{i,0} + g_{i,1}t \\ &= g_{i,0} + (\mathbf{a}_{i-1,0}g_{1,0} + \mathbf{a}_{i-1,1}g_{1,1} + \mathbf{a}_{i-1,2}g_{2,0} + \mathbf{a}_{i-1,3}g_{3,0} \\ &\quad + \dots + \mathbf{a}_{i-1,k-1}g_{k-1,0} + \mathbf{a}_{i-1,k}g_{k,0})t \end{aligned}$$

para  $i = 2, \dots, k$ . Finalmente, se tiene que

$$g(t) = [g_{1,0} + g_{1,1}t] D_{1,t} + \sum_{i=2}^k [g_{i,0} + (\mathbf{a}_{i-1,0}g_{1,0} + \mathbf{a}_{i-1,1}g_{1,1} + \mathbf{a}_{i-1,2}g_{2,0} + \mathbf{a}_{i-1,3}g_{3,0} + \dots + \mathbf{a}_{i-1,k-1}g_{k-1,0} + \mathbf{a}_{i-1,k}g_{k,0})t] D_{i,t}$$

donde  $D_{i,t} = \begin{cases} 1, & t_{i-1} \leq t < t_i \\ 0, & \text{en otro caso} \end{cases}$ , para  $i = 1, \dots, k-1$ , y  $D_{k,t} = \begin{cases} 1, & t_{k-1} \leq t \leq t_k \\ 0, & \text{en otro caso} \end{cases}$ . Y

reordenando términos

<sup>27</sup> Nótese que en este caso se ha elegido esta combinación, pero podría haberse elegido cualquier otra.

$$\begin{aligned}
 g(t) &= g_{1,0} \left[ D_{1,t} + \mathbf{a}_{1,0} t D_{2,t} + \mathbf{a}_{2,0} t D_{3,t} + \dots + \mathbf{a}_{k-2,0} t D_{k-1,t} + \mathbf{a}_{k-1,0} t D_{k,t} \right] + \\
 &g_{1,1} \left[ t D_{1,t} + \mathbf{a}_{1,1} t D_{2,t} + \mathbf{a}_{2,1} t D_{3,t} + \dots + \mathbf{a}_{k-2,1} t D_{k-1,t} + \mathbf{a}_{k-1,1} t D_{k,t} \right] + \\
 &\sum_{i=2}^k g_{i,0} \left[ (\mathbf{a}_{1,i} D_{2,t} + \mathbf{a}_{2,i} D_{3,t} + \dots + \mathbf{a}_{k-2,i} D_{k-1,t} + \mathbf{a}_{k-1,i} D_{k,t}) \right] + D_{i,t} \Big] \\
 &= g_{1,0} X_{0,t} + g_{1,1} X_{1,t} + \sum_{i=2}^k g_{i,0} X_{i,t}
 \end{aligned}$$

Entonces,  $\hat{G}^* = (\hat{g}_{1,0}, \hat{g}_{1,1}, \hat{g}_{2,0}, \hat{g}_{3,0}, \dots, \hat{g}_{k-1,0}, \hat{g}_{k,0})'$  puede obtenerse estimando la regresión

$$y_t = g_{1,0} X_{0,t} + g_{1,1} X_{1,t} + \sum_{i=2}^k g_{i,0} X_{i,t} + \mathbf{e}_t,$$

siendo  $\mathbf{e}_t$  un término de error.

Si, por el contrario, la función elegida es una *spline* cuadrática y bajo el mismo supuesto de que los puntos de ruptura son conocidos, las restricciones de continuidad de la función y de la primera derivada determinan, en este caso, un sistema de  $2(k-1)$  ecuaciones con  $3k$  parámetros que deben ser determinados. En este caso, quedan  $k+2$  parámetros libres que podrían ser estimados.

Sea  $G_{2(k-1) \times 1} = (g_{2,1}, g_{2,2}, g_{3,1}, g_{3,2}, \dots, g_{k-1,1}, g_{k-1,2}, g_{k,1}, g_{k,2})'$  el vector de parámetros expresado como función de los parámetros finalmente estimados, es decir,  $G_{(k+2) \times 1}^* = (g_{1,0}, g_{1,1}, g_{1,2}, g_{2,0}, g_{3,0}, \dots, g_{k-1,0}, g_{k,0})'$ . En términos matriciales, el problema consiste en obtener  $G$  sabiendo que  $TG = R$ , donde

$$T_{2(k-1) \times 2(k-1)} = \begin{bmatrix}
 -t_1 & -t_1^2 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
 t_2 & t_2^2 & -t_2 & -t_2^2 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & t_{k-1} & t_{k-1}^2 & -t_{k-1} & -t_{k-1}^2 \\
 -1 & -2t_1 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
 1 & 2t_2 & -1 & -2t_2 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 1 & 2t_{k-1} & -1 & -2t_{k-1}
 \end{bmatrix}$$

y

$$R_{2(k-1) \times 1} = \begin{bmatrix} -g_{1,0} - g_{1,1}t_1 - g_{1,2}t_1^2 + g_{2,0} \\ -g_{2,0} + g_{3,0} \\ \vdots \\ -g_{k-1,0} + g_{k,0} \\ -g_{1,1} - 2g_{1,2}t_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

de modo que los parámetros del vector  $G$  pueden obtenerse teniendo en cuenta que  $G = T^{-1}R$ . Si se denota por  $a_{i,j}$  al elemento de la fila  $i$ -ésima y la columna  $j$ -ésima de la inversa de la matriz  $T$ , es decir,

$$T_{2(k-1) \times 2(k-1)}^{-1} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,2k-2} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,2k-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{2k-2,1} & a_{2k-2,2} & \cdots & a_{2k-2,2k-2} \end{bmatrix}$$

y  $g_j$ ,  $j = 1, \dots, 2(k-1)$ , tal que  $g_1 = g_{2,1}$ ,  $g_2 = g_{2,2}$ ,  $g_3 = g_{3,1}$ ,  $g_4 = g_{3,2}$ ,  $g_5 = g_{4,1}$ ,  $g_6 = g_{4,2}$ ,  $\dots$ ,  $g_{2k-5} = g_{k-1,1}$ ,  $g_{2k-4} = g_{k-1,2}$ ,  $g_{2k-3} = g_{k,1}$  y  $g_{2k-2} = g_{k,2}$ ; entonces,

$$\begin{aligned} g_j = & a_{j,1}(-g_{1,0} - g_{1,1}t_1 - g_{1,2}t_1^2 + g_{2,0}) + \\ & a_{j,2}(-g_{2,0} + g_{3,0}) + \dots + a_{j,k-1}(-g_{k-1,0} + g_{k,0}) + \\ & a_{j,k}(-g_{1,1} - 2g_{1,2}t_1) \end{aligned}$$

para  $j = 1, \dots, 2(k-1)$ ; y, reordenando términos,

$$g_j = \mathbf{a}_{j,0}g_{1,0} + \mathbf{a}_{j,1}g_{1,1} + \mathbf{a}_{j,2}g_{1,2} + \sum_{i=2}^k \mathbf{a}_{j,i+1}g_{i,0}, \quad j = 1, \dots, 2(k-1),$$

donde

$$\begin{aligned} \mathbf{a}_{j,0} &= -a_{j,1} \\ \mathbf{a}_{j,1} &= -a_{j,1}t_1 - a_{j,k} \\ \mathbf{a}_{j,2} &= -a_{j,1}t_1^2 - 2a_{j,k}t_1 \\ \mathbf{a}_{j,3} &= a_{j,1} - a_{j,2} \\ \mathbf{a}_{j,4} &= a_{j,2} - a_{j,3} \\ &\vdots \\ \mathbf{a}_{j,k} &= a_{j,k-2} - a_{j,k-1} \\ \mathbf{a}_{j,k+1} &= a_{j,k-1} \end{aligned}$$

La *spline* puede especificarse en función de los parámetros del vector  $G^*$  de la siguiente forma. Si  $t_0 \leq t \leq t_1$ :

$$g(t) = g_{1,0} + g_{1,1}t + g_{1,2}t^2$$

Si  $t_{i-1} \leq t \leq t_i$ ,

$$\begin{aligned} g(t) &= g_{i,0} + g_{i,1}t + g_{i,2}t^2 \\ &= g_{i,0} + (\acute{a}_{2i-3,0}g_{1,0} + \acute{a}_{2i-3,1}g_{1,1} + \acute{a}_{2i-3,2}g_{1,2} + \acute{a}_{2i-3,3}g_{2,0} \\ &\quad + \acute{a}_{2i-3,4}g_{3,0} + \dots + \acute{a}_{2i-3,k}g_{k-1,0} + \acute{a}_{2i-3,k+1}g_{k,0})t + \\ &\quad + (\acute{a}_{2i-2,0}g_{1,0} + \acute{a}_{2i-2,1}g_{1,1} + \acute{a}_{2i-2,2}g_{1,2} + \acute{a}_{2i-2,3}g_{2,0} \\ &\quad + \acute{a}_{2i-2,4}g_{3,0} + \dots + \acute{a}_{2i-2,k}g_{k-1,0} + \acute{a}_{2i-2,k+1}g_{k,0})t^2 \end{aligned}$$

para  $i = 2, \dots, k$ . Finalmente, se tiene que

$$g(t) = [g_{1,0} + g_{1,1}t + g_{1,2}t^2] D_{1,t} + \sum_{i=2}^k \begin{bmatrix} g_{i,0} + (\acute{a}_{2i-3,0}g_{1,0} + \acute{a}_{2i-3,1}g_{1,1} + \acute{a}_{2i-3,2}g_{1,2} + \acute{a}_{2i-3,3}g_{2,0} \\ + \acute{a}_{2i-3,4}g_{3,0} + \dots + \acute{a}_{2i-3,k}g_{k-1,0} + \acute{a}_{2i-3,k+1}g_{k,0})t + \\ + (\acute{a}_{2i-2,0}g_{1,0} + \acute{a}_{2i-2,1}g_{1,1} + \acute{a}_{2i-2,2}g_{1,2} + \acute{a}_{2i-2,3}g_{2,0} \\ + \acute{a}_{2i-2,4}g_{3,0} + \dots + \acute{a}_{2i-2,k}g_{k-1,0} + \acute{a}_{2i-2,k+1}g_{k,0})t^2 \end{bmatrix} D_{i,t}$$

donde  $D_{i,t} = \begin{cases} 1, & t_{i-1} \leq t < t_i \\ 0, & \text{en otro caso} \end{cases}$ , para  $i = 1, \dots, k-1$ , y  $D_{k,t} = \begin{cases} 1, & t_{k-1} \leq t \leq t_k \\ 0, & \text{en otro caso} \end{cases}$ . Y

reordenando términos

$$\begin{aligned} g(t) &= g_{1,0} \left[ D_{1,t} + (\mathbf{a}_{1,0}t + \mathbf{a}_{2,0}t^2)D_{2,t} + \dots + (\mathbf{a}_{2k-5,0}t + \mathbf{a}_{2k-4,0}t^2)D_{k-1,t} + \right] + \\ &\quad (\mathbf{a}_{2k-3,0}t + \mathbf{a}_{2k-2,0}t^2)D_{k,t} \\ &= g_{1,1} \left[ tD_{1,t} + (\mathbf{a}_{1,1}t + \mathbf{a}_{2,1}t^2)D_{2,t} + \dots + (\mathbf{a}_{2k-5,1}t + \mathbf{a}_{2k-4,1}t^2)D_{k-1,t} + \right] + \\ &\quad (\mathbf{a}_{2k-3,1}t + \mathbf{a}_{2k-2,1}t^2)D_{k,t} \\ &= g_{1,2} \left[ t^2D_{1,t} + (\mathbf{a}_{1,2}t + \mathbf{a}_{2,2}t^2)D_{2,t} + \dots + (\mathbf{a}_{2k-5,2}t + \mathbf{a}_{2k-4,2}t^2)D_{k-1,t} + \right] + \\ &\quad (\mathbf{a}_{2k-3,2}t + \mathbf{a}_{2k-2,2}t^2)D_{k,t} \\ &= \sum_{i=2}^k g_{i,0} \left[ D_{i,t} + (\mathbf{a}_{1,i+1}t + \mathbf{a}_{2,i+1}t^2)D_{2,t} + \dots + (\mathbf{a}_{2k-5,i+1}t + \mathbf{a}_{2k-4,i+1}t^2)D_{k-1,t} + \right] \\ &\quad (\mathbf{a}_{2k-3,i+1}t + \mathbf{a}_{2k-2,i+1}t^2)D_{k,t} \\ &= g_{1,0}X_{0,t} + g_{1,1}X_{1,t} + g_{1,2}X_{2,t} + \sum_{i=2}^k g_{i,0}X_{i+1,t} \end{aligned}$$

Entonces,  $\hat{G}^* = (\hat{g}_{1,0}, \hat{g}_{1,1}, \hat{g}_{1,2}, \hat{g}_{2,0}, \hat{g}_{3,0}, \dots, \hat{g}_{k-1,0}, \hat{g}_{k,0})'$  puede obtenerse estimando la regresión

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$y_t = g_{1,0}X_{0,t} + g_{1,1}X_{1,t} + g_{1,2}X_{2,t} + \sum_{i=2}^k g_{i,0}X_{i+1,t} + e_t,$$

siendo  $e_t$  un término de error.

Finalmente, en el caso de una *spline* cúbica y asumiendo que los puntos de ruptura son conocidos, las restricciones de continuidad de la función y las de la primera y segunda derivada determinan, como ya se ha comentado, un sistema de  $3(k-1)$  ecuaciones con  $4k$  parámetros que deben ser determinados. En este caso, quedan  $k+3$  parámetros libres que podrían ser estimados. Sea  $G_{3(k-1) \times 1} = (g_{2,1}, g_{2,2}, g_{2,3}, g_{3,1}, g_{3,2}, g_{3,3}, \dots, g_{k-1,1}, g_{k-1,2}, g_{k-1,3}, g_{k,1}, g_{k,2}, g_{k,3})'$  el vector de parámetros expresado como función de los parámetros finalmente estimados, es decir,  $G_{(k+3) \times 1}^* = (g_{1,0}, g_{1,1}, g_{1,2}, g_{1,3}, g_{2,0}, g_{3,0}, \dots, g_{k-1,0}, g_{k,0})'$ . En términos matriciales, el problema consiste en obtener  $G$  sabiendo que  $TG = R$ , donde

$$T_{3(k-1) \times 3(k-1)} = \begin{bmatrix} -t_1 & -t_1^2 & -t_1^3 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ t_2 & t_2^2 & t_2^3 & -t_2 & -t_2^2 & -t_2^3 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & t_{k-1} & t_{k-1}^2 & t_{k-1}^3 & -t_{k-1} & -t_{k-1}^2 & -t_{k-1}^3 \\ -1 & -2t_1 & -3t_1^2 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2t_2 & 3t_2^2 & -1 & -2t_2 & -3t_2^2 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 1 & 2t_{k-1} & 3t_{k-1}^2 & -1 & -2t_{k-1} & -3t_{k-1}^2 \\ 0 & -2 & -6t_1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 6t_2 & 0 & -2 & -6t_2 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 2 & 6t_{k-1} & 0 & -2 & -6t_{k-1} \end{bmatrix}$$

y

$$R_{3(k-1) \times 1} = \begin{bmatrix} -g_{1,0} - g_{1,1}t_1 - g_{1,2}t_1^2 - g_{1,3}t_1^3 + g_{2,0} \\ -g_{2,0} + g_{3,0} \\ \vdots \\ -g_{k-1,0} + g_{k,0} \\ -g_{1,1} - 2g_{1,2}t_1 - 3g_{1,3}t_1^2 \\ 0 \\ \vdots \\ 0 \\ -2g_{1,2} - 6g_{1,3}t_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

de modo que los parámetros del vector  $G$  pueden obtenerse teniendo en cuenta que  $G = T^{-1}R$ . Si se denota por  $a_{i,j}$  al elemento de la fila  $i$ -ésima y la columna  $j$ -ésima de la inversa de la matriz  $T$ , es decir,

$$T_{3(k-1) \times 3(k-1)}^{-1} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,3k-3} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,3k-3} \\ \vdots & \vdots & \ddots & \vdots \\ a_{3k-3,1} & a_{3k-3,2} & \cdots & a_{3k-3,3k-3} \end{bmatrix}$$

y  $g_j$ ,  $j = 1, \dots, 3(k-1)$ , tal que  $g_1 = g_{2,1}$ ,  $g_2 = g_{2,2}$ ,  $g_3 = g_{2,3}$ ,  $g_4 = g_{3,1}$ ,  $g_5 = g_{3,2}$ ,  $g_6 = g_{3,3}$ ,  $\dots$ ,  $g_{3k-8} = g_{k-1,1}$ ,  $g_{3k-7} = g_{k-1,2}$ ,  $g_{3k-6} = g_{k-1,3}$ ,  $g_{3k-5} = g_{k,1}$ ,  $g_{3k-4} = g_{k,2}$  y  $g_{3k-3} = g_{k,3}$ ; entonces,

$$\begin{aligned} g_j = & a_{j,1}(-g_{1,0} - g_{1,1}t_1 - g_{1,2}t_1^2 - g_{1,3}t_1^3 + g_{2,0}) + \\ & a_{j,2}(-g_{2,0} + g_{3,0}) + \dots + a_{j,k-1}(-g_{k-1,0} + g_{k,0}) + \\ & a_{j,k}(-g_{1,1} - 2g_{1,2}t_1 - 3g_{1,3}t_1^2) + a_{j,2k-1}(-2g_{1,2} - 6g_{1,3}t_1) \end{aligned}$$

para  $j = 1, \dots, 3(k-1)$ ; y, reordenando términos,

$$g_j = \mathbf{a}_{j,0}g_{1,0} + \mathbf{a}_{j,1}g_{1,1} + \mathbf{a}_{j,2}g_{1,2} + \mathbf{a}_{j,3}g_{1,3} + \sum_{i=2}^k \mathbf{a}_{j,i+2}g_{i,0}, \quad j = 1, \dots, 3(k-1),$$

donde

$$\begin{aligned} \mathbf{a}_{j,0} &= -a_{j,1} \\ \mathbf{a}_{j,1} &= -a_{j,1}t_1 - 2a_{j,k} \\ \mathbf{a}_{j,2} &= -a_{j,1}t_1^2 - 2a_{j,k}t_1 - 2a_{j,2k-1} \\ \mathbf{a}_{j,3} &= -a_{j,1}t_1^3 - 3a_{j,k}t_1^2 - 6a_{j,2k-1}t_1 \\ \mathbf{a}_{j,4} &= a_{j,1} - a_{j,2} \\ \mathbf{a}_{j,5} &= a_{j,2} - a_{j,3} \\ &\vdots \\ \mathbf{a}_{j,k+1} &= a_{j,k-2} - a_{j,k-1} \\ \mathbf{a}_{j,k+2} &= a_{j,k-1} \end{aligned}$$

La *spline* puede especificarse en función de los parámetros del vector  $G^*$  de la siguiente forma. Si  $t_0 \leq t \leq t_1$ :

$$g(t) = g_{1,0} + g_{1,1}t + g_{1,2}t^2 + g_{1,3}t^3$$

Si  $t_{i-1} \leq t \leq t_i$ ,

$$\begin{aligned}
 g(t) &= g_{i,0} + g_{i,1}t + g_{i,2}t^2 + g_{i,3}t^3 \\
 &= g_{i,0} + (\mathbf{a}_{3i-5,0}g_{1,0} + \mathbf{a}_{3i-5,1}g_{1,1} + \mathbf{a}_{3i-5,2}g_{1,2} + \mathbf{a}_{3i-5,3}g_{1,3} + \\
 &\quad \mathbf{a}_{3i-5,4}g_{2,0} + \mathbf{a}_{3i-5,5}g_{3,0} + \dots + \mathbf{a}_{3i-5,k+1}g_{k-1,0} + \mathbf{a}_{3i-5,k+2}g_{k,0})t + \\
 &\quad (\mathbf{a}_{3i-4,0}g_{1,0} + \mathbf{a}_{3i-4,1}g_{1,1} + \mathbf{a}_{3i-4,2}g_{1,2} + \mathbf{a}_{3i-4,3}g_{1,3} + \\
 &\quad \mathbf{a}_{3i-4,4}g_{2,0} + \mathbf{a}_{3i-4,5}g_{3,0} + \dots + \mathbf{a}_{3i-4,k+1}g_{k-1,0} + \mathbf{a}_{3i-4,k+2}g_{k,0})t^2 + \\
 &\quad (\mathbf{a}_{3i-3,0}g_{1,0} + \mathbf{a}_{3i-3,1}g_{1,1} + \mathbf{a}_{3i-3,2}g_{1,2} + \mathbf{a}_{3i-3,3}g_{1,3} + \\
 &\quad \mathbf{a}_{3i-3,4}g_{2,0} + \mathbf{a}_{3i-3,5}g_{3,0} + \dots + \mathbf{a}_{3i-3,k+1}g_{k-1,0} + \mathbf{a}_{3i-3,k+2}g_{k,0})t^3
 \end{aligned}$$

para  $i = 2, \dots, k$ . Finalmente, se tiene que

$$g(t) = [g_{1,0} + g_{1,1}t + g_{1,2}t^2 + g_{1,3}t^3] D_{1,t} + \sum_{i=2}^k \left[ \begin{aligned} &g_{i,0} + (\mathbf{a}_{3i-5,0}g_{1,0} + \mathbf{a}_{3i-5,1}g_{1,1} + \mathbf{a}_{3i-5,2}g_{1,2} + \mathbf{a}_{3i-5,3}g_{1,3} + \\ &\mathbf{a}_{3i-5,4}g_{2,0} + \mathbf{a}_{3i-5,5}g_{3,0} + \dots + \mathbf{a}_{3i-5,k+1}g_{k-1,0} + \mathbf{a}_{3i-5,k+2}g_{k,0})t + \\ &(\mathbf{a}_{3i-4,0}g_{1,0} + \mathbf{a}_{3i-4,1}g_{1,1} + \mathbf{a}_{3i-4,2}g_{1,2} + \mathbf{a}_{3i-4,3}g_{1,3} + \\ &\mathbf{a}_{3i-4,4}g_{2,0} + \mathbf{a}_{3i-4,5}g_{3,0} + \dots + \mathbf{a}_{3i-4,k+1}g_{k-1,0} + \mathbf{a}_{3i-4,k+2}g_{k,0})t^2 + \\ &(\mathbf{a}_{3i-3,0}g_{1,0} + \mathbf{a}_{3i-3,1}g_{1,1} + \mathbf{a}_{3i-3,2}g_{1,2} + \mathbf{a}_{3i-3,3}g_{1,3} + \\ &\mathbf{a}_{3i-3,4}g_{2,0} + \mathbf{a}_{3i-3,5}g_{3,0} + \dots + \mathbf{a}_{3i-3,k+1}g_{k-1,0} + \mathbf{a}_{3i-3,k+2}g_{k,0})t^3 \end{aligned} \right] D_{i,t}$$

donde  $D_{i,t} = \begin{cases} 1, & t_{i-1} \leq t < t_i \\ 0, & \text{en otro caso} \end{cases}$ , para  $i = 1, \dots, k-1$ , y  $D_{k,t} = \begin{cases} 1, & t_{k-1} \leq t \leq t_k \\ 0, & \text{en otro caso} \end{cases}$ . Y

reordenando términos

$$\begin{aligned}
 g(t) &= g_{1,0} \left[ \begin{aligned} &D_{1,t} + (\mathbf{a}_{1,0}t + \mathbf{a}_{2,0}t^2 + \mathbf{a}_{3,0}t^3)D_{2,t} + \dots \\ &+ (\mathbf{a}_{3k-8,0}t + \mathbf{a}_{3k-7,0}t^2 + \mathbf{a}_{3k-6,0}t^3)D_{k-1,t} \\ &+ (\mathbf{a}_{3k-5,0}t + \mathbf{a}_{3k-4,0}t^2 + \mathbf{a}_{3k-3,0}t^3)D_{k,t} \end{aligned} \right] + \\
 &g_{1,1} \left[ \begin{aligned} &tD_{1,t} + (\mathbf{a}_{1,1}t + \mathbf{a}_{2,1}t^2 + \mathbf{a}_{3,1}t^3)D_{2,t} + \dots \\ &+ (\mathbf{a}_{3k-8,1}t + \mathbf{a}_{3k-7,1}t^2 + \mathbf{a}_{3k-6,1}t^3)D_{k-1,t} \\ &+ (\mathbf{a}_{3k-5,1}t + \mathbf{a}_{3k-4,1}t^2 + \mathbf{a}_{3k-3,1}t^3)D_{k,t} \end{aligned} \right] + \\
 &g_{1,2} \left[ \begin{aligned} &t^2D_{1,t} + (\mathbf{a}_{1,2}t + \mathbf{a}_{2,2}t^2 + \mathbf{a}_{3,2}t^3)D_{2,t} + \dots \\ &+ (\mathbf{a}_{3k-8,2}t + \mathbf{a}_{3k-7,2}t^2 + \mathbf{a}_{3k-6,2}t^3)D_{k-1,t} \\ &+ (\mathbf{a}_{3k-5,2}t + \mathbf{a}_{3k-4,2}t^2 + \mathbf{a}_{3k-3,2}t^3)D_{k,t} \end{aligned} \right] + \\
 &g_{1,3} \left[ \begin{aligned} &t^3D_{1,t} + (\mathbf{a}_{1,3}t + \mathbf{a}_{2,3}t^2 + \mathbf{a}_{3,3}t^3)D_{2,t} + \dots \\ &+ (\mathbf{a}_{3k-8,3}t + \mathbf{a}_{3k-7,3}t^2 + \mathbf{a}_{3k-6,3}t^3)D_{k-1,t} \\ &+ (\mathbf{a}_{3k-5,3}t + \mathbf{a}_{3k-4,3}t^2 + \mathbf{a}_{3k-3,3}t^3)D_{k,t} \end{aligned} \right] +
 \end{aligned}$$

$$\sum_{i=2}^k g_{i,0} \left[ \begin{array}{l} D_{i,t} + (\mathbf{a}_{1,i+2}t + \mathbf{a}_{2,i+2}t^2 + \mathbf{a}_{3,i+2}t^3)D_{2,t} + \dots + \\ (\mathbf{a}_{3k-8,i+2}t + \mathbf{a}_{3k-7,i+2}t^2 + \mathbf{a}_{3k-6,i+2}t^3)D_{k-1,t} + \\ (\mathbf{a}_{3k-5,i+2}t + \mathbf{a}_{3k-4,i+2}t^2 + \mathbf{a}_{3k-3,i+2}t^3)D_{k,t} \end{array} \right]$$

$$= g_{1,0}X_{0,t} + g_{1,1}X_{1,t} + g_{1,2}X_{2,t} + g_{1,3}X_{3,t} + \sum_{i=2}^k g_{i,0}X_{i+2,t}$$

Entonces,  $\hat{G}^* = (\hat{g}_{1,0}, \hat{g}_{1,1}, \hat{g}_{1,2}, \hat{g}_{1,3}, \hat{g}_{2,0}, \hat{g}_{3,0}, \dots, \hat{g}_{k-1,0}, \hat{g}_{k,0})'$  puede obtenerse estimando la regresión

$$y_t = g_{1,0}X_{0,t} + g_{1,1}X_{1,t} + g_{1,2}X_{2,t} + g_{1,3}X_{3,t} + \sum_{i=2}^k g_{i,0}X_{i+2,t} + \mathbf{e}_t$$

siendo  $\mathbf{e}_t$  un término de error.

En los párrafos anteriores se ha descrito el procedimiento para estimar las funciones de regresión *spline* lineal, cuadrática y cúbica cuando los puntos de ruptura son conocidos. Si éstos son desconocidos, entonces  $\{t_0, \dots, t_k\}$  se convierten en parámetros adicionales que deben ser estimados con el resto de parámetros libres en la especificación de la *spline*, y, por tanto, los modelos estimables antes especificados son no lineales en los parámetros, pero pueden ser estimados utilizando procedimientos de regresión mínimo cuadrática no lineal. Aunque el enfoque tradicional consiste en recurrir a la estimación por máxima verosimilitud debido a las propiedades deseables de normalidad asintótica y consistencia que presentan estos estimadores.

Robison (1964) discute la estimación de los puntos de intersección de dos funciones polinomiales de grado dos y emplea el método de la máxima verosimilitud para estimar los coeficientes incorporados en el modelo de regresión, incluyendo los parámetros de localización de los puntos de intersección. Por otro lado, Hudson (1966) examina cuatro situaciones en función de que en los puntos de ruptura se registren o no observaciones de la magnitud estudiada y dependiendo de que las especificaciones en dos tramos consecutivos tengan o no pendientes iguales en los puntos de ruptura. Gallant y Fuller (1973) realizan una importante contribución a la propuesta de Hudson; en concreto, utilizando las condiciones de continuidad de la función y de sus primeras derivadas, reparametrizan el modelo de regresión de la función *spline* para dar lugar a un modelo de regresión no lineal y sugieren obtener estimaciones mínimo cuadráticas de los parámetros utilizando el método de optimización de Gauss-Newton modificado por Hartley (Hartley, 1961).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

En esta misma línea, Marsh (1983) presenta de un modo sistemático varias técnicas de estimación de la localización de los puntos de ruptura cuando se conoce de antemano el número de éstos y el grado de la función<sup>28</sup>. En el caso de que la *spline* sea lineal el método de Gauss-Newton es el que habitualmente se utiliza en la estimación. Es decir, una vez incorporadas las restricciones de continuidad, la estimación del modelo no lineal resultante puede llevarse a cabo expresando la *spline* reespecificada en términos de su desarrollo en serie de Taylor de primer orden en un entorno del vector de valores iniciales de los parámetros desconocidos de la siguiente manera. Si  $g(t)$  es una función *spline* lineal de  $k$  tramos que aproxima los valores  $y_t$ ,  $t = 1, \dots, T$ , de modo que

$$y_t = g(t) + \mathbf{e}_t,$$

siendo  $\mathbf{e}_t$  un término de error ruido blanco, la *spline* puede reespecificarse como

$$g(t) = f(\mathbf{b})_t$$

siendo  $\mathbf{b}$  un vector definido como  $\mathbf{b} = \begin{bmatrix} G^* & N \end{bmatrix}$ , donde  $G_{(k+1) \times 1}^* = (g_{1,0}, g_{1,1}, g_{2,0}, g_{2,1}, \dots, g_{k-1,0}, g_{k,0})'$  y  $N_{(k-1) \times 1} = (t_1, \dots, t_{k-1})'$ <sup>29</sup>, y dado que  $f(\mathbf{b})_t$  es una función no lineal de los parámetros, la estimación de éstos puede efectuarse expresando  $f(\mathbf{b})_t$  como:

$$f(\mathbf{b})_t = f(\mathbf{b}_{(0)})_t + \nabla f(\mathbf{b}_{(0)})_t (\mathbf{b} - \mathbf{b}_{(0)})$$

siendo  $\mathbf{b}_{(0)}$  el vector de valores iniciales de los parámetros desconocidos y  $\nabla f(\mathbf{b})_t$  el vector de derivadas parciales de la función  $f(\mathbf{b})_t$  respecto de los parámetros del vector  $\mathbf{b}$ . Entonces, el modelo original puede especificarse como:

$$y_t = f(\mathbf{b}_{(0)})_t + \nabla f(\mathbf{b}_{(0)})_t (\mathbf{b} - \mathbf{b}_{(0)}) + \mathbf{e}_t^*,$$

siendo  $\mathbf{e}_t^*$  un nuevo término de perturbación.

Dado que  $f(\mathbf{b}_{(0)})_t$  y  $\nabla f(\mathbf{b}_{(0)})_t \mathbf{b}_{(0)}$  son valores conocidos en cada instante del tiempo, conviene expresar el modelo anterior de la forma

$$y_t - f(\mathbf{b}_{(0)})_t + \nabla f(\mathbf{b}_{(0)})_t \mathbf{b}_{(0)} = \nabla f(\mathbf{b}_{(0)})_t \mathbf{b} + \mathbf{e}_t^*.$$

<sup>28</sup> Aunque el número de puntos de ruptura es conocido de antemano, se añade flexibilidad al estimar la localización de los mismos. Véase también Marsh y otros (1990).

<sup>29</sup> En este vector no se incluyen los parámetros  $t_0$  y  $t_k$ , ya que se asume que, como ocurre en general,  $t_0 = 1$  y  $t_k = T$ , es decir, estos valores corresponden a los instantes del tiempo inicial y final del periodo al que se ajusta la *spline*.

Si definimos  $y_{t(0)}^* = y_t - f(\mathbf{b}_{(0)})_t + \nabla f(\mathbf{b}_{(0)})_t \mathbf{b}_{(0)}$ , una primera aproximación al vector de parámetros  $\mathbf{b}$  puede obtenerse estimando la regresión siguiente:

$$y_{t(0)}^* = \nabla f(\mathbf{b}_{(0)})_t \mathbf{b} + \mathbf{e}_t^*,$$

es decir,

$$\mathbf{b}_{(1)} = \left[ \left[ \nabla f(\mathbf{b}_{(0)})_t \right]' \left[ \nabla f(\mathbf{b}_{(0)})_t \right] \right]^{-1} \left[ \nabla f(\mathbf{b}_{(0)})_t \right]' y_{t(0)}^*.$$

Esta estimación preliminar obtenida a partir de los valores iniciales  $\mathbf{b}_{(0)}$  sirve de punto de partida para iniciar un proceso iterativo en el que la estimación de la  $i$ -ésima iteración viene dada por:

$$\mathbf{b}_{(i)} = \left[ \left[ \nabla f(\mathbf{b}_{(i-1)})_t \right]' \left[ \nabla f(\mathbf{b}_{(i-1)})_t \right] \right]^{-1} \left[ \nabla f(\mathbf{b}_{(i-1)})_t \right]' y_{t(i-1)}^*$$

Alternativamente, los parámetros estimados en una iteración se pueden expresar en función de las estimaciones de la iteración anterior de la siguiente forma:

$$\mathbf{b}_{(i)} = \mathbf{b}_{(i-1)} + \left[ \left[ \nabla f(\mathbf{b}_{(i-1)})_t \right]' \left[ \nabla f(\mathbf{b}_{(i-1)})_t \right] \right]^{-1} \left[ \nabla f(\mathbf{b}_{(i-1)})_t \right]' y_{t(i-1)}^{**}$$

donde  $y_{t(i-1)}^{**} = y_t - f(\mathbf{b}_{(i-1)})_t$ . El proceso concluye cuando se satisface algún criterio de convergencia. El criterio que suele establecerse es  $\|\mathbf{b}_{(i)} - \mathbf{b}_{(i-1)}\| < c$ , siendo  $c$  un valor designado de antemano; también puede realizarse un elevado número de iteraciones y decidir *a posteriori* cuando se alcanza la convergencia<sup>30</sup>.

Cuando la función *spline* es de orden superior a la lineal, conviene acudir a procedimientos de estimación en la línea del método de Newton-Raphson<sup>31</sup>. Una vez reespecificado el modelo en los términos descritos en la aproximación anterior, es decir,

$$y_t = f(\mathbf{b})_t + \mathbf{e}_t$$

siendo  $\mathbf{b}$  un vector definido como  $\mathbf{b} = [G^* \ N]$ , donde

$G_{(k+2) \times 1}^* = (g_{1,0}, g_{1,1}, g_{1,2}, g_{2,0}, g_{3,0}, \dots, g_{k-1,0}, g_{k,0})'$ , si  $g(t)$  es una *spline* cuadrática, y

<sup>30</sup> El proceso anteriormente descrito exige fijar  $\mathbf{b}_{(0)}$ . En ocasiones se dispone de información que sugiere tomar determinados valores iniciales. En general, dichos valores iniciales podrían obtenerse mediante un proceso iterativo tal que, en cada iteración, se asume como conocida una combinación diferente de puntos de ruptura y se elige, finalmente, aquel vector de parámetros que minimice la suma de cuadrados de los residuos.

<sup>31</sup> El método de Newton-Raphson no resulta útil en el caso de *splines* lineales porque las segundas derivadas requeridas se anulan.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$G_{(k+3) \times 1}^* = (g_{1,0}, g_{1,1}, g_{1,2}, g_{1,3}, g_{2,0}, \dots, g_{k-1,0}, g_{k,0})'$ , si  $g(t)$  es una *spline* cúbica, y  $N_{(k-1) \times 1} = (t_1, \dots, t_{k-1})'$ , y dado que el objetivo es minimizar

$$S(\mathbf{b}) = \sum_{t=1}^T (y_t - f(\mathbf{b})_t)^2,$$

conviene expresar  $S(\mathbf{b})$  en términos de su desarrollo en serie de Taylor hasta el segundo orden en un entorno de  $\mathbf{b}_{(0)}$ , es decir,

$$S(\mathbf{b}) \cong S(\mathbf{b}_{(0)}) + \nabla S(\mathbf{b}_{(0)}) (\mathbf{b} - \mathbf{b}_{(0)}) + \frac{(\mathbf{b} - \mathbf{b}_{(0)})' \nabla^2 S(\mathbf{b}_{(0)}) (\mathbf{b} - \mathbf{b}_{(0)})}{2!},$$

donde  $\nabla S(\mathbf{b})$  es el vector de derivadas parciales de la función  $S(\mathbf{b})$  respecto de los parámetros del vector  $\mathbf{b}$  y  $\nabla^2 S(\mathbf{b})$  es la matriz de derivadas parciales cruzadas.

El vector  $\mathbf{b}$  que minimiza  $S(\mathbf{b})$  es aquél que verifica la ecuación:

$$\frac{\partial S(\mathbf{b})}{\partial \mathbf{b}} \cong [S(\mathbf{b}_{(0)})]' + \nabla^2 S(\mathbf{b}_{(0)}) (\mathbf{b} - \mathbf{b}_{(0)}) = 0,$$

de modo que

$$\mathbf{b} = \mathbf{b}_{(0)} - [\nabla^2 S(\mathbf{b}_{(0)})]^{-1} [\nabla S(\mathbf{b}_{(0)})]',$$

donde

$$S(\mathbf{b}_{(0)}) = \sum_{t=1}^T (y_t - f(\mathbf{b}_{(0)})_t)^2,$$

$$\nabla S(\mathbf{b}_{(0)}) = -2 \sum_{t=1}^T (y_t - f(\mathbf{b}_{(0)})_t) \nabla f(\mathbf{b}_{(0)})_t$$

y

$$\nabla^2 S(\mathbf{b}_{(0)}) = 2 \left[ \sum_{t=1}^T [\nabla f(\mathbf{b}_{(0)})_t]' \nabla f(\mathbf{b}_{(0)})_t - \sum_{t=1}^T [y_t - f(\mathbf{b}_{(0)})_t] \nabla^2 f(\mathbf{b}_{(0)})_t \right].$$

Iterativamente, pueden obtenerse las estimaciones del vector  $\mathbf{b}$  dadas por:

$$\mathbf{b}_{(i)} = \mathbf{b}_{(i-1)} - [\nabla^2 S(\mathbf{b}_{(i-1)})]^{-1} [\nabla S(\mathbf{b}_{(i-1)})]',$$

donde

$$\nabla S(\mathbf{b}_{(i)}) = -2 \sum_{t=1}^T (y_t - f(\mathbf{b}_{(i)})_t) \nabla f(\mathbf{b}_{(i)})_t$$

y

$$\nabla^2 S(\mathbf{b}_{(i)}) = 2 \left[ \sum_{t=1}^T [\nabla f(\mathbf{b}_{(i)})_t]' \nabla f(\mathbf{b}_{(i)})_t - \sum_{t=1}^T [y_t - f(\mathbf{b}_{(i)})_t] \nabla^2 f(\mathbf{b}_{(i)})_t \right].$$

En el anexo 2.1 puede encontrarse una ilustración de los procedimientos de estimación no lineales descritos para los casos de una función *spline* lineal, cuadrática y cúbica definidas por dos tramos.

El problema es más complicado cuando además de su localización se desconoce el número de puntos de ruptura. En estos casos, Marsh (1983, 1986) sugiere considerar un número lo suficientemente grande de puntos de ruptura y que sea la significación de los parámetros estimados la que determine qué puntos de ruptura son adecuados. En este sentido, se puede fijar un conjunto de puntos potenciales dividiendo el rango de datos observados en un número lo suficientemente grande de intervalos de igual amplitud<sup>32</sup>. Un enfoque alternativo consiste en utilizar los valores observados de la variable que se intenta explicar a través de la función *spline* como elementos para determinar el conjunto de puntos de ruptura<sup>33</sup>.

En cualquier caso, y como se ha expuesto, cuando los puntos de ruptura no son conocidos y, por tanto, deben ser estimados conjuntamente con los parámetros que definen la función *spline*, la resolución del problema requiere recurrir a métodos de estimación no lineales. Sin embargo, la experiencia ha mostrado que en el propio proceso iterativo pueden surgir problemas que implican la finalización del proceso sin alcanzar el óptimo. Como señala Nielsen (1998: 46-47), los procesos iterativos de estimación conducen a mínimos locales y, generalmente, se obtienen mejores resultados mediante procedimientos heurísticos de ajustes sucesivos tomando, en cada caso, diferentes localizaciones como dadas<sup>34</sup>. Circunstancias como éstas aconsejan imponer cierta rigidez en el procedimiento de especificación y estimación posterior de la *spline*. En este sentido, parece recomendable efectuar la estimación de *splines* de diferente orden y con distintas combinaciones de localizaciones de los puntos de ruptura hasta encontrar aquella función que describa el comportamiento subyacente observado para el conjunto de datos sin permitir que el término de perturbación, presente en cualquier región del rango estudiado, conduzca a elegir especificaciones demasiado flexibles o laxas.

---

<sup>32</sup> Una *spline* con nodos igualmente espaciados se denomina *spline* cardinal (de Boor, 1978:322).

<sup>33</sup> Por ejemplo, si la magnitud observada es una serie temporal, podría optarse por elegir los puntos de ruptura dividiendo el intervalo de tiempo en el que se observa la serie en subintervalos de la misma amplitud. La segunda opción señalada implicaría seleccionar subintervalos del rango de valores de la serie observada y determinar los puntos de ruptura como los instantes del tiempo correspondientes a los extremos de dichos subintervalos. En este caso, ambas opciones podrían ser equivalentes.

<sup>34</sup> El objetivo de estos procedimientos es identificar la *mejor* aproximación mínimo cuadrática entre todas las posibles. Ahora bien, cuando el número o localización de los nodos es desconocido, sólo es posible

### 2.1.2. Procedimientos que imponen el paso por los nodos

Dado que el objetivo del presente epígrafe es explicar en qué consisten los procedimientos que imponen el paso por los nodos y para no ser reiterativo, sólo se ilustrará el caso de una *spline* cúbica, que es la que mejor recoge el comportamiento de una función que cambia de forma suave, como ocurre en las series que van a ser analizadas en los capítulos siguientes<sup>35</sup>.

Entonces, en el caso particular de una *spline* cúbica y asumiendo, en contraste con el planteamiento anterior, que los puntos de ruptura son conocidos y que la función pasa por los nodos, es decir,

$$g_{i,0} + g_{i,1}t_{i-1} + g_{i,2}t_{i-1}^2 + g_{i,3}t_{i-1}^3 = y_{t_{i-1}}, \quad i = 1, \dots, k,$$

y

$$g_{k,0} + g_{k,1}t_k + g_{k,2}t_k^2 + g_{k,3}t_k^3 = y_{t_k},$$

Koopman (1992) y Harvey, Koopman y Riani (1997) proponen y aplican el siguiente procedimiento para determinar la *spline*. Sea  $z_i$  la distancia entre los elementos de la rejilla del  $i$ -ésimo tramo, es decir,  $z_i = t_i - t_{i-1}$ , y sea  $\nabla^2 g_i(t)$  la derivada segunda de  $g_i(t)$  evaluada en  $t$ . Si se asume que  $\nabla^2 g_i(t_i) = a_i$ , teniendo en cuenta que los polinomios que definen la *spline* son de grado tres —por tanto, la segunda derivada de la función *spline* en el  $i$ -ésimo tramo será una función lineal— y considerando además el requisito de continuidad de la derivada segunda, resulta que:

$$\nabla^2 g_i(t) = [(t_i - t)/z_i]a_{i-1} + [(t - t_{i-1})/z_i]a_i,$$

$i = 1, \dots, k$ , siendo  $a_0, \dots, a_k$  desconocidos *a priori*<sup>36</sup>.

Una vez especificada la derivada segunda, se obtiene la función  $g_i(t)$  a partir de sencillas operaciones de integración. En primer lugar,  $\nabla_i(t) = \int \nabla_i^2(t)dt$ , es decir,

---

intentar, numéricamente, mejorar una aproximación mediante *pequeños* cambios en los parámetros que la describen y comprobar de esta forma si la aproximación propuesta es localmente óptima.

<sup>35</sup> Los desarrollos analíticos para el caso de *splines* de cualquier orden pueden consultarse en de Boor (1978).

<sup>36</sup> Adviértase que, de este modo,  $\nabla^2 g_i(t_i) = a_i$  y  $\nabla^2 g_i(t_{i-1}) = a_{i-1}$ , lo que garantiza la continuidad de la derivada segunda.

$$\nabla_i(t) = -\left[(t_i - t)^2 / 2z_i\right] a_{i-1} + \left[(t - t_{i-1})^2 / 2z_i\right] a_i + k_{1,i},$$

donde  $k_{1,i}$  es una constante que resulta de la integración. De forma similar, la función  $g_i(t)$  se obtiene integrando  $\nabla_i(t)$ , es decir,

$$g_i(t) = -\left[(t_i - t)^3 / 6z_i\right] a_{i-1} + \left[(t - t_{i-1})^3 / 6z_i\right] a_i + tk_{1,i} + k_{2,i},$$

donde  $k_{2,i}$  constituye la nueva constante de integración.

A continuación, se obliga a que la función *spline* cúbica pase por los nodos, de modo que  $g_i(t_i) = y_{t_i}$  y  $g_i(t_i - 1) = y_{t_{i-1}}$ , para  $i = 1, \dots, k$ <sup>37</sup>. Así, las constantes  $k_{1,i}$  y  $k_{2,i}$  pueden ser elegidas de la siguiente forma:

$$k_{1,i} = (y_{t_{i-1}} - y_{t_i}) / z_i - (z_i / 6)(a_i - a_{i-1})$$

y

$$k_{2,i} = y_{t_i} - (z_i^2 / 6)a_i - \left[(y_{t_i} - y_{t_{i-1}}) / z_i\right] t_i + (z_i / 6)(a_i - a_{i-1})t_i.$$

Sustituyendo las constantes  $k_{1,i}$  y  $k_{2,i}$  en la función *spline*  $g_i(t)$  se obtiene:

$$g_i(t) = \left[(t_i - t)((t_i - t)^2 - z_i^2) / 6z_i\right] a_{i-1} + \left[(t - t_{i-1})((t - t_{i-1})^2 - z_i^2) / 6z_i\right] a_i + \left[(t_i - t) / z_i\right] y_{t_{i-1}} + \left[(t - t_{i-1}) / z_i\right] y_{t_i}$$

para  $t_{i-1} \leq t \leq t_i$  y  $i = 1, \dots, k$ . De forma matricial, se tiene que  $G = W^A A + W^B Y$ , donde  $G = (g_1(t), \dots, g_k(t))'$ ,  $A = (a_0, \dots, a_k)'$ ,  $Y = (y_{t_0}, \dots, y_{t_k})'$  y  $W^A$  y  $W^B$  son matrices definidas como:

---

<sup>37</sup> Nótese que si en cada tramo se impone la restricción de que  $g_i(t_i) = y_{t_i}$  y  $g_i(t_i - 1) = y_{t_{i-1}}$ , estas restricciones implican la continuidad de la *spline*.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$W_{k \times (k+1)}^A = \begin{bmatrix} (t_1 - t) \left[ \frac{(t_1 - t)^2 - z_1^2}{6z_1} \right] & (t - t_0) \left[ \frac{(t - t_0)^2 - z_1^2}{6z_1} \right] & 0 & \dots & 0 & 0 & 0 \\ 0 & (t_2 - t) \left[ \frac{(t_2 - t)^2 - z_2^2}{6z_2} \right] & (t - t_1) \left[ \frac{(t - t_1)^2 - z_2^2}{6z_2} \right] & \dots & 0 & 0 & 0 \\ 0 & 0 & (t_3 - t) \left[ \frac{(t_3 - t)^2 - z_3^2}{6z_3} \right] & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & (t_{k-1} - t) \left[ \frac{(t_{k-1} - t)^2 - z_{k-1}^2}{6z_{k-1}} \right] & (t - t_{k-2}) \left[ \frac{(t - t_{k-2})^2 - z_{k-1}^2}{6z_{k-1}} \right] & 0 \\ 0 & 0 & 0 & \dots & 0 & (t_k - t) \left[ \frac{(t_k - t)^2 - z_k^2}{6z_k} \right] & (t - t_{k-1}) \left[ \frac{(t - t_{k-1})^2 - z_k^2}{6z_k} \right] \end{bmatrix}$$

y

$$W_{k \times (k+1)}^B = \begin{bmatrix} \frac{(t_1-t)}{z_1} & \frac{(t-t_0)}{z_1} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{(t_2-t)}{z_2} & \frac{(t-t_1)}{z_2} & \dots & 0 & 0 & 0 \\ 0 & 0 & \frac{(t_3-t)}{z_3} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{(t_{k-1}-t)}{z_{k-1}} & \frac{(t-t_{k-2})}{z_{k-1}} & 0 \\ 0 & 0 & 0 & \dots & 0 & \frac{(t_k-t)}{z_k} & \frac{(t-t_{k-1})}{z_k} \end{bmatrix}$$

Dado que  $a_0, \dots, a_k$  son parámetros desconocidos, se pueden utilizar las restricciones de continuidad de la derivada primera para expresar los parámetros del vector  $A$  en función del vector  $Y$ . Así, una vez sustituidas las constantes de integración en la derivada primera, se obtiene

$$\begin{aligned} \nabla_i(t) = & -\left[ \frac{(t_i-t)^2}{2z_i} - z_i/6 \right] a_{i-1} + \\ & \left[ \frac{(t-t_{i-1})^2}{2z_i} - z_i/6 \right] a_i + \\ & (y_{t_i} - y_{t_{i-1}}) / z_i \end{aligned}$$

e imponiendo las condiciones de continuidad de la derivada primera antes señaladas, es decir,

$$\nabla_i(t_i) = \nabla_{i+1}(t_i),$$

donde

$$\nabla_i(t_i) = (z_i/6)a_{i-1} + (z_i/3)a_i + (y_{t_i} - y_{t_{i-1}}) / z_i$$

y

$$\nabla_{i+1}(t_i) = -(z_{i+1}/3)a_i - (z_{i+1}/6)a_{i+1} + (y_{t_{i+1}} - y_{t_i}) / z_{i+1},$$

para  $i = 1, \dots, k-1$ , se tiene finalmente que:

$$\begin{aligned} (z_i/6)a_{i-1} + (z_i + z_{i+1}/3)a_i + (z_{i+1}/6)a_{i+1} = \\ (1/z_i)y_{t_{i-1}} - ((1/z_{i+1}) + (1/z_i))y_{t_i} + (1/z_{i+1})y_{t_{i+1}}, \end{aligned}$$

para  $i = 1, \dots, k-1$ . El conjunto de restricciones puede ser representado de forma matricial por la notación  $W^1 A = W^2 Y$ , donde las matrices  $W^1$  y  $W^2$  son matrices tridiagonales definidas, respectivamente, como:

$$W_{(k-1)x(k+1)}^1 = \begin{bmatrix} \frac{z_1}{6} & \frac{z_1+z_2}{3} & \frac{z_2}{6} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{z_2}{6} & \frac{z_2+z_3}{3} & \frac{z_3}{6} & \dots & 0 & 0 & 0 \\ 0 & 0 & \frac{z_3}{6} & \frac{z_3+z_4}{3} & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{z_4}{6} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \frac{z_{k-1}}{6} & \frac{z_{k-1}+z_k}{3} & \frac{z_k}{6} \end{bmatrix}$$

y

$$W_{(k-1)x(k+1)}^2 = \begin{bmatrix} \frac{1}{z_1} & -\left(\frac{1}{z_1} + \frac{1}{z_2}\right) & \frac{1}{z_2} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{1}{z_2} & -\left(\frac{1}{z_2} + \frac{1}{z_3}\right) & \frac{1}{z_3} & \dots & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{z_3} & -\left(\frac{1}{z_3} + \frac{1}{z_4}\right) & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{z_4} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \frac{1}{z_{k-1}} & -\left(\frac{1}{z_{k-1}} + \frac{1}{z_k}\right) & \frac{1}{z_k} \end{bmatrix}$$

Estas restricciones configuran un sistema de  $k - 1$  ecuaciones con  $k + 1$  valores desconocidos  $a_i$ ,  $i = 0, \dots, k$ <sup>38</sup>, de modo que la solución de este sistema exige añadir dos restricciones<sup>39</sup>. Una solución directa y, en muchos casos, apropiada consiste en tratar  $a_0$  y  $a_k$  como constantes desconocidas. Otra opción es suponer que  $a_0 = a_k = 0$ ; en este caso, la función *spline* cúbica se conoce como función *spline* natural<sup>40</sup>.

<sup>38</sup> Nótese que se está asumiendo que los parámetros  $\{t_0, \dots, t_k\}$  son conocidos.

<sup>39</sup> En Poirier (1973, 1976) se discuten las posibles restricciones que se pueden considerar.

<sup>40</sup> Además de su sonoro nombre, la interpolación *spline* natural es una aproximación poco recomendable desde un punto de vista teórico, ya que la asignación arbitraria del valor nulo al principio y final del periodo al que se ajusta la *spline* produce errores innecesarios en el ajuste (de Boor, 1978:55). Otro tipo particular de *spline* se obtiene cuando, además del paso por los nodos, se impone la condición de que la derivada primera de la *spline* coincida con la de la función que se aproxima en los puntos inicial y final del periodo al que se ajusta la *spline*. Esta interpolación, denominada *spline* cúbica completa, minimiza la expresión  $\int_{t_0}^{t_n} [\nabla^2 g(t)]^2 dt$  para el conjunto de funciones  $g$  dos veces diferenciables que pasan por los nodos (de Boor, 1978:63) y, en este sentido, puede interpretarse que describe el trazo de un dibujante forzado a cruzar por determinados puntos. Esta es la razón por la que Schoenberg (1946) eligió la palabra *spline* para este tipo de funciones.

La solución para el vector  $A$  viene dada por  $A = (W^1)^{-1}W^2Y$ , que permite determinar  $g_i(t)$  para cualquier  $t_{i-1} \leq t \leq t_i$ ,  $i = 1, \dots, k$ . Dado que el vector de parámetros  $A$  es una combinación lineal de  $y_{t_0}, \dots, y_{t_k}$ , cada función  $g_i(t)$  es también una combinación lineal de  $y_{t_0}, \dots, y_{t_k}$  que, en forma matricial, se puede expresar como:

$$G = W^A A + W^B Y = W^A (W^1)^{-1} W^2 Y + W^B Y = (W^A (W^1)^{-1} W^2 + W^B) Y = WY$$

En definitiva, conocida la matriz  $W$ , los valores de la *spline* quedan determinados como combinaciones lineales de los valores observados de la función en los puntos de ruptura<sup>41</sup>. Si  $t$  se corresponde con uno de los valores pertenecientes a  $\{t_0, \dots, t_k\}$ , por ejemplo  $t = t_i$ , todos los elementos de la fila  $i$ -ésima de la matriz  $W$  serán nulos, excepto el correspondiente a la columna  $i + 1$ , que será igual a la unidad, de modo que  $g(t_i) = y_{t_i}$ .

## 2.2. Funciones *splines* cúbicas periódicas y estacionalidad

La mayoría de las series temporales económicas presentan movimientos periódicos, tales como las variaciones estacionales, que son responsables de una parte importante de la variabilidad experimentada por dichas series. Variaciones de este tipo pueden pasar inadvertidas si el grado de agregación temporal en los datos disponibles es excesivo, pero suelen aparecer con claridad en series temporales mensuales o trimestrales. Cuando las observaciones se registran con mayor frecuencia, es decir, en intervalos de tiempo de menor duración, pueden llegar a detectarse movimientos periódicos dentro de la semana e incluso del día.

Comúnmente, un patrón periódico que se repite cada  $s$  unidades de tiempo se modela con  $s - 1$  variables cualitativas o, alternativamente, con  $s - 1$  términos trigonométricos. Sin embargo, cuando el patrón periódico se completa en un largo intervalo de tiempo o, en general, cuando el número de observaciones registradas en cada uno de estos periodos de tiempo es suficientemente alto, estos modelos no se caracterizan por su parsimonia. En estos casos, las funciones *splines* cúbicas periódicas proporcionan una formulación alternativa que requiere el empleo de un número inferior de parámetros.

---

<sup>41</sup> Para determinar los valores de  $W$ , dado un valor particular de  $t$ , se requiere la inversa de  $W^1$ , que, debido a su característica de matriz tridiagonal, se puede obtener de forma eficiente.

Con objeto de formular el patrón estacional a partir de una función *spline* cúbica periódica, puede asumirse inicialmente que la estacionalidad es determinística, es decir,  $\mathbf{g}_t = \mathbf{g}_j$ , cuando  $t$  pertenece a la estación  $j$ ,  $j = 1, \dots, s$ . Entonces, el componente estacional puede aproximarse a través de una función *spline* cúbica periódica, de forma que  $\mathbf{g}_j = \mathbf{g}(j) + \mathbf{e}_j$ , donde  $\mathbf{e}_j$  es un término de perturbación y  $\mathbf{g}(j)$  es una función polinómica de grado tres definida a trozos,  $\mathbf{g}_i(j) = \mathbf{g}_{i,0} + \mathbf{g}_{i,1}j + \mathbf{g}_{i,2}j^2 + \mathbf{g}_{i,3}j^3$ ,  $j_{i-1} \leq j \leq j_i$ , para  $i = 1, \dots, k$ , donde  $j_0$  y  $j_k$  representan la primera y última estación, respectivamente, a la que se imponen las siguientes condiciones de continuidad de la función,  $\mathbf{g}_i(j_i) = \mathbf{g}_{i+1}(j_i)$ ,  $i = 1, \dots, k-1$ , de su derivada primera,  $\nabla \mathbf{g}_i(j_i) = \nabla \mathbf{g}_{i+1}(j_i)$ ,  $i = 1, \dots, k-1$ , y de la derivada segunda,  $\nabla^2 \mathbf{g}_i(j_i) = \nabla^2 \mathbf{g}_{i+1}(j_i)$ ,  $i = 1, \dots, k-1$ . Y el carácter periódico de la *spline* se introduce mediante las tres restricciones adicionales siguientes:  $\mathbf{g}_k(j_k + 1) = \mathbf{g}_1(j_0)$ ,  $\nabla \mathbf{g}_k(j_k + 1) = \nabla \mathbf{g}_1(j_0)$  y  $\nabla^2 \mathbf{g}_k(j_k + 1) = \nabla^2 \mathbf{g}_1(j_0)$ <sup>42</sup>. A continuación se detalla el procedimiento de estimación de la *spline* en función de que se imponga o no la condición de paso por los nodos.

### 2.2.1. Procedimientos que no imponen el paso por los nodos

Si se emplean los procedimientos de estimación de la *spline* que no imponen el paso por los nodos, la redefinición de las matrices  $T$  y  $R$  es inmediata, y el vector  $G^*$  será ahora de menor dimensión. En este caso, la *spline* debe verificar las siguientes condiciones:

a) restricciones de continuidad de la función,

$$\mathbf{g}_{i,0} + \mathbf{g}_{i,1}j_i + \mathbf{g}_{i,2}j_i^2 + \mathbf{g}_{i,3}j_i^3 = \mathbf{g}_{i+1,0} + \mathbf{g}_{i+1,1}j_i + \mathbf{g}_{i+1,2}j_i^2 + \mathbf{g}_{i+1,3}j_i^3,$$

para  $i = 1, \dots, k-1$ , y

$$\mathbf{g}_{k,0} + \mathbf{g}_{k,1}(j_k + 1) + \mathbf{g}_{k,2}(j_k + 1)^2 + \mathbf{g}_{k,3}(j_k + 1)^3 = \mathbf{g}_{1,0} + \mathbf{g}_{1,1}j_0 + \mathbf{g}_{1,2}j_0^2 + \mathbf{g}_{1,3}j_0^3;$$

b) restricciones de continuidad de la derivada primera,

$$\mathbf{g}_{i,1} + 2\mathbf{g}_{i,2}j_i + 3\mathbf{g}_{i,3}j_i^2 = \mathbf{g}_{i+1,1} + 2\mathbf{g}_{i+1,2}j_i + 3\mathbf{g}_{i+1,3}j_i^2$$

para  $i = 1, \dots, k-1$ , y

<sup>42</sup> Nótese que en el caso particular de que  $(j_0, \dots, j_k) = (1, \dots, s)$  o, en general, que  $k = s-1$  —de modo que exista un punto de ruptura en cada estación—, la función *spline* se ajustará perfectamente a la variación  $\mathbf{g}_j$ ,  $j = 1, \dots, s$ ; pero no supone entonces ninguna ventaja con respecto al modelo de variables cualitativas o términos trigonométricos.

$$g_{k,1} + 2g_{k,2}(j_k + 1) + 3g_{k,3}(j_k + 1)^2 = g_{1,1} + 2g_{1,2}j_0 + 3g_{1,3}j_0^2;$$

c) restricciones de continuidad de la derivada segunda,

$$2g_{i,2} + 6g_{i,3}j_i = 2g_{i+1,2} + 6g_{i+1,3}j_i$$

para  $i = 1, \dots, k-1$ , y

$$2g_{k,2} + 6g_{k,3}(j_k + 1) = 2g_{1,2} + 6g_{1,3}j_0.$$

Una vez especificada la forma funcional en cada uno de los tramos que definen la *spline* y si se asume que los puntos de ruptura  $j_i$ ,  $i = 1, \dots, k-1$ , son conocidos, el siguiente paso es determinar los parámetros  $g_{i,0}$ ,  $g_{i,1}$ ,  $g_{i,2}$  y  $g_{i,3}$ , para  $i = 1, \dots, k$ . Dado que el número de parámetros que se desea determinar,  $4k$ , es superior al número de condiciones impuestas,  $3k$ , se puede optar por expresar el vector de parámetros  $G_{3k \times 1} = (g_{1,1}, g_{1,2}, g_{1,3}, \dots, g_{k,1}, g_{k,2}, g_{k,3})'$  como función de  $G_{k \times 1}^* = (g_{1,0}, g_{2,0}, \dots, g_{k,0})'$ , que son los parámetros finalmente estimados, y de  $j_i$ ,  $i = 0, \dots, k$ . Teniendo en cuenta el planteamiento del problema, conviene expresar el conjunto de restricciones anteriores de forma matricial como  $TG = R$ , donde

$$T_{3k \times 3k} = \begin{bmatrix} j_1 & j_1^2 & j_1^3 & -j_1 & -j_1^2 & -j_1^3 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & j_{k-1} & j_{k-1}^2 & j_{k-1}^3 & -j_{k-1} & -j_{k-1}^2 & -j_{k-1}^3 \\ -j_0 & -j_0^2 & -j_0^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & (j_k + 1) & (j_k + 1)^2 & (j_k + 1)^3 \\ 1 & 2j_1 & 3j_1^2 & -1 & -2j_1 & -3j_1^2 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 2j_{k-1} & 3j_{k-1}^2 & -1 & -2j_{k-1} & -3j_{k-1}^2 \\ -1 & -2j_0 & -3j_0^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 2(j_k + 1) & 3(j_k + 1)^2 \\ 0 & 2 & 6j_1 & 0 & -2 & -6j_1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 2 & 6j_{k-1} & 0 & -2 & -6j_{k-1} \\ 0 & -2 & -6j_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 2 & 6(j_k + 1) \end{bmatrix}$$

y

$$R_{3k \times 1} = \begin{bmatrix} -g_{1,0} + g_{2,0} \\ \vdots \\ -g_{k-1,0} + g_{k,0} \\ -g_{k,0} + g_{1,0} \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

de modo que  $G = T^{-1}R$ . Si se denota la inversa de la matriz  $T$  por

$$T^{-1}_{3k \times 3k} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,3k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,3k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{3k,1} & a_{3k,2} & \cdots & a_{3k,3k} \end{bmatrix}$$

y se define  $g_r$ ,  $r = 1, \dots, 3k$ , tal que  $g_1 = g_{1,1}$ ,  $g_2 = g_{1,2}$ ,  $g_3 = g_{1,3}$ ,  $g_4 = g_{2,1}$ ,  $g_5 = g_{2,2}$ ,  $g_6 = g_{2,3}$ ,  $\dots$ ,  $g_{3k-2} = g_{k,1}$ ,  $g_{3k-1} = g_{k,2}$  y  $g_{3k} = g_{k,3}$ , entonces

$$g_r = a_{r,1}(-g_{1,0} + g_{2,0}) + a_{r,2}(-g_{2,0} + g_{3,0}) + \dots + a_{r,k-1}(-g_{k-1,0} + g_{k,0}) + a_{r,k}(g_{1,0} - g_{k,0})$$

para  $r = 1, \dots, 3k$ . Es decir,

$$g_r = \mathbf{a}_{r,1}g_{1,0} + \mathbf{a}_{r,2}g_{2,0} + \dots + \mathbf{a}_{r,k-1}g_{k-1,0} + \mathbf{a}_{r,k}g_{k,0},$$

donde

$$\mathbf{a}_{r,1} = -a_{r,1} + a_{r,k}, \mathbf{a}_{r,2} = a_{r,1} - a_{r,2}, \dots, \mathbf{a}_{r,k-1} = a_{r,k-2} - a_{r,k-1} \text{ y } \mathbf{a}_{r,k} = a_{r,k-1} - a_{r,k},$$

para  $r = 1, \dots, 3k$ . Ahora, la *spline* puede especificarse en función de  $g_{i,0}$ ,  $i = 1, \dots, k$ , de la siguiente forma:

$$\begin{aligned} g(j) = & g_{i,0} + (\mathbf{a}_{3i-2,1}g_{1,0} + \mathbf{a}_{3i-2,2}g_{2,0} + \dots + \mathbf{a}_{3i-2,k-1}g_{k-1,0} + \mathbf{a}_{3i-2,k}g_{k,0})j + \\ & (\mathbf{a}_{3i-1,1}g_{1,0} + \mathbf{a}_{3i-1,2}g_{2,0} + \dots + \mathbf{a}_{3i-1,k-1}g_{k-1,0} + \mathbf{a}_{3i-1,k}g_{k,0})j^2 + \\ & (\mathbf{a}_{3i,1}g_{1,0} + \mathbf{a}_{3i,2}g_{2,0} + \dots + \mathbf{a}_{3i,k-1}g_{k-1,0} + \mathbf{a}_{3i,k}g_{k,0})j^3 \end{aligned}$$

si  $j_{i-1} \leq j \leq j_i$ ,  $i = 1, \dots, k$ , y finalmente como:

$$g(j) = \sum_{i=1}^k \begin{bmatrix} g_{i,0} + (\mathbf{a}_{3i-2,1}g_{1,0} + \mathbf{a}_{3i-2,2}g_{2,0} + \dots + \mathbf{a}_{3i-2,k-1}g_{k-1,0} + \mathbf{a}_{3i-2,k}g_{k,0})j \\ + (\mathbf{a}_{3i-1,1}g_{1,0} + \mathbf{a}_{3i-1,2}g_{2,0} + \dots + \mathbf{a}_{3i-1,k-1}g_{k-1,0} + \mathbf{a}_{3i-1,k}g_{k,0})j^2 \\ + (\mathbf{a}_{3i,1}g_{1,0} + \mathbf{a}_{3i,2}g_{2,0} + \dots + \mathbf{a}_{3i,k-1}g_{k-1,0} + \mathbf{a}_{3i,k}g_{k,0})j^3 \end{bmatrix} D_{i,j}$$

donde  $D_{i,j} = \begin{cases} 1, & \text{si } j_{i-1} \leq j < j_i \\ 0, & \text{resto} \end{cases}$ ,  $i = 1, \dots, k-1$ , y  $D_{k,j} = \begin{cases} 1, & \text{si } j_{k-1} \leq j \leq j_k \\ 0, & \text{resto} \end{cases}$ , y

reordenando términos, se tiene que

$$g(j) = g_{1,0}X_{1,j} + g_{2,0}X_{2,j} + \dots + g_{k-1,0}X_{k-1,j} + g_{k,0}X_{k,j}$$

donde

$$\begin{aligned} X_{i,j} = & D_{i,j} + (\mathbf{a}_{1,i}j + \mathbf{a}_{2,i}j^2 + \mathbf{a}_{3,i}j^3)D_{1,j} + (\mathbf{a}_{4,i}j + \mathbf{a}_{5,i}j^2 + \mathbf{a}_{6,i}j^3)D_{2,j} + \dots + \\ & (\mathbf{a}_{3k-2,i}j + \mathbf{a}_{3k-1,i}j^2 + \mathbf{a}_{3k,i}j^3)D_{k,j} \end{aligned}$$

para  $i = 1, \dots, k$ .

**2.2.2. Procedimientos que imponen el paso por los nodos**

En el procedimiento propuesto por Koopman (1992) y Harvey, Koopman y Riani (1997), las condiciones de comportamiento periódico hacen que no sea necesario imponer restricciones adicionales sobre los parámetros  $a_0$  y  $a_k$  para obtener un sistema de ecuaciones identificable, ya que  $a_0 = a_k$  y, además,  $y_{t_0} = y_{t_k}$ . Entonces, la *spline* puede obtenerse como función de  $\{a_1, \dots, a_k\}$  y  $\{y_{t_1}, \dots, y_{t_k}\}$  y expresarse, en forma matricial, como

$$G_{kx1} = W_{kxk} Y_{kx1},$$

donde  $Y = (y_{t_0}, \dots, y_{t_k})'$  y  $W_{kxk} = W_{kxk}^A (W_{kxk}^1)^{-1} W_{kxk}^2 + W_{kxk}^B$ . Ahora, las matrices  $W^1$  y  $W^2$  pierden su estructura tridiagonal y quedan definidas de la siguiente manera:

$$W_{(kxk)}^1 = \begin{bmatrix} \frac{z_1 + z_2}{3} & \frac{z_2}{6} & 0 & \dots & 0 & 0 & \frac{z_1}{6} \\ \frac{z_2}{6} & \frac{z_2 + z_3}{3} & \frac{z_3}{6} & \dots & 0 & 0 & 0 \\ 0 & \frac{z_3}{6} & \frac{z_3 + z_4}{3} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{z_{k-1}}{6} & \frac{z_{k-1} + z_k}{3} & \frac{z_k}{6} \\ \frac{z_1}{6} & 0 & 0 & \dots & 0 & \frac{z_k}{6} & \frac{z_k + z_1}{3} \end{bmatrix}$$

y

$$W_{(kxk)}^2 = \begin{bmatrix} -\left(\frac{1}{z_1} + \frac{1}{z_2}\right) & \frac{1}{z_2} & 0 & \dots & 0 & 0 & \frac{1}{z_1} \\ \frac{1}{z_2} & -\left(\frac{1}{z_2} + \frac{1}{z_3}\right) & \frac{1}{z_3} & \dots & 0 & 0 & 0 \\ 0 & \frac{1}{z_3} & -\left(\frac{1}{z_3} + \frac{1}{z_4}\right) & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{z_{k-1}} & -\left(\frac{1}{z_{k-1}} + \frac{1}{z_k}\right) & \frac{1}{z_k} \\ \frac{1}{z_1} & 0 & 0 & \dots & 0 & \frac{1}{z_k} & -\left(\frac{1}{z_k} + \frac{1}{z_1}\right) \end{bmatrix}$$

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

En los párrafos anteriores se ha mostrado que las variaciones estacionales y cíclicas en series temporales pueden ser modeladas por medio de funciones *splines* periódicas. Sin embargo, dado que la idea básica bajo la cual se construyen los modelos estructurales de series temporales es permitir que los componentes inobservables cambien en el tiempo, parece demasiado restrictivo exigir, *a priori*, que las funciones *splines* sean fijas. Si la estacionalidad es estocástica, dicho carácter estocástico podría recogerse a través de una función *spline* que evoluciona en el tiempo<sup>43</sup>.

Si se asume que los puntos de ruptura son conocidos y que la *spline* pasa por los nodos, entonces

$$g(j) = w_j' Y,$$

donde  $w_j = (w_{j,1}, \dots, w_{j,k})'$  es un vector cuyos parámetros resultan de la evaluación, en la estación  $j$ , de las funciones en la fila correspondiente de la matriz  $W$  definida anteriormente e  $Y = (y_{t_1}, \dots, y_{t_k})'$ .

Dado que  $g(j)$  es la variación estacional en la estación  $j$ , se tiene que

$$g(j) = w_{j,1}y_{t_1} + \dots + w_{j,k}y_{t_k},$$

y con objeto de garantizar que la suma de los efectos estacionales sea nula, es decir,

$$\sum_{j=1}^s g(j) = 0,$$

conviene eliminar la restricción que impone, por ejemplo, el paso por el último nodo ( $t_k, y_{t_k}$ ). Dado que

$$\sum_{j=1}^s g(j) = \sum_{j=1}^s w_{j,1}y_{t_1} + \dots + \sum_{j=1}^s w_{j,k-1}y_{t_{k-1}} + \sum_{j=1}^s w_{j,k}y_{t_k} = 0,$$

se tiene que

$$\sum_{j=1}^s w_{j,k}y_{t_k} = - \left[ \sum_{j=1}^s w_{j,1}y_{t_1} + \dots + \sum_{j=1}^s w_{j,k-1}y_{t_{k-1}} \right],$$

de donde

$$y_{t_k} = - \left[ \frac{\sum_{j=1}^s w_{j,1}}{\sum_{j=1}^s w_{j,k}} y_{t_1} + \dots + \frac{\sum_{j=1}^s w_{j,k-1}}{\sum_{j=1}^s w_{j,k}} y_{t_{k-1}} \right].$$

<sup>43</sup> Esta idea fue originalmente propuesta por Koopman (1992) y Harvey y Koopman (1993).

Entonces, la restricción  $\sum_{j=1}^s g(j) = 0$  puede incorporarse en el modelo definiendo

$$g(j) = w_{j,1}y_{t_1} + \dots + w_{j,k}y_{t_k} - \left[ \frac{\sum_{j=1}^s w_{j,1}}{\sum_{j=1}^s w_{j,k}} y_{t_1} + \dots + \frac{\sum_{j=1}^s w_{j,k-1}}{\sum_{j=1}^s w_{j,k}} y_{t_{k-1}} \right].$$

Y, si se define

$$w_{\bullet,i} = \sum_{j=1}^s w_{j,i},$$

para  $i = 1, \dots, k-1$ , la variación estacional en la estación  $j$  puede escribirse como

$$g(j) = \left[ w_{j,1} - w_{j,k} \frac{w_{\bullet,1}}{w_{\bullet,k}} \right] y_{t_1} + \dots + \left[ w_{j,k-1} - w_{j,k} \frac{w_{\bullet,k-1}}{w_{\bullet,k}} \right] y_{t_{k-1}},$$

o bien, de forma más compacta

$$g(j) = z_j' Y^+,$$

donde  $z_j = (z_{j,1}, \dots, z_{j,k-1})'$ , siendo  $z_{j,i} = w_{j,i} - w_{j,k} \frac{w_{\bullet,i}}{w_{\bullet,k}}$ ,  $i = 1, \dots, k-1$ , y

$$Y^+ = (y_{t_1}, \dots, y_{t_{k-1}})'$$

Para que la función *spline* así definida adquiera carácter estocástico, basta con asumir que los elementos del vector  $Y = (y_{t_1}, \dots, y_{t_k})'$  evolucionen en el tiempo como caminos aleatorios, es decir,

$$y_{t_i}(t) = y_{t_i}(t-1) + \chi_{i,t}, \quad i = 1, \dots, k,$$

o de forma vectorial,

$$Y_t = Y_{t-1} + \mathbf{c}_t,$$

donde  $\mathbf{c}_t = (\mathbf{c}_{1,t}, \dots, \mathbf{c}_{k,t})'$  es un vector de perturbaciones aleatorias incorreladas serialmente con media cero y matriz de varianzas-covarianzas  $\mathbf{S}^2 V$ .

De nuevo, la suma de los efectos estacionales debe ser nula, es decir,  $w_*' Y = 0$ , donde  $w_* = (w_{\bullet,1}, \dots, w_{\bullet,k})'$ <sup>44</sup>. Esta restricción implica que en la expresión

---

<sup>44</sup> Obsérvese que

$$w_*' Y_t = w_*' Y_{t-1} + w_*' c_t$$

se cumpla que

$$w_*' c_t = w_{\bullet,1} c_{1,t} + \dots + w_{\bullet,k-1} c_{k-1,t} + w_{\bullet,k} c_{k,t} = 0,$$

de modo que

$$c_{k,t} = -\frac{1}{w_{\bullet,k}} (w_{\bullet,1}, \dots, w_{\bullet,k-1}) (c_{1,t}, \dots, c_{k-1,t})' = -\frac{1}{w_{\bullet,k}} w_{**}' c_{*,t}.$$

Entonces, los vectores  $c_t$  y  $w_*$  pueden particionarse como  $c_t = (c_{*,t} \quad c_{k,t})'$  y

$w_* = (w_{**} \quad w_{\bullet,k})'$ . Por tanto, la matriz  $s^2 V$  también puede ser particionada como

$$s^2 \begin{bmatrix} V_* & V_k \\ V_k & V_{k,k} \end{bmatrix},$$

donde  $V_*$  es una matriz de orden  $(k-1) \times (k-1)$  definida de forma que

$$s^2 V_* = \begin{bmatrix} Var(c_{1,t}) & \dots & Cov(c_{1,t}, c_{k-1,t}) \\ \vdots & \ddots & \vdots \\ Cov(c_{k-1,t}, c_{1,t}) & \dots & Var(c_{k-1,t}) \end{bmatrix},$$

$V_k$  es un vector de orden  $(k-1)$  tal que  $s^2 V_k = \begin{bmatrix} Cov(c_{1,t}, c_{k,t}) \\ \vdots \\ Cov(c_{k-1,t}, c_{k,t}) \end{bmatrix}$  y  $V_{k,k} = Var(c_{k,t})$ .

Entonces, el vector  $V_k$  se puede expresar en función de la matriz  $V_*$  como

$V_k = -V_* (w_{**} / w_{\bullet,k})$ , teniendo en cuenta que

$$Cov(c_{j,t}, c_{k,t}) = E[c_{j,t} c_{k,t}] = -\sum_{i=1}^{k-1} \frac{w_{\bullet,i}}{w_{\bullet,k}} E[c_{j,t} c_{i,t}], \quad j = 1, \dots, k-1,$$

y, del mismo modo, se tiene que

$$V_{kk} = E[c_{k,t}^2] = E\left[\left(\sum_{i=1}^{k-1} \frac{w_{\bullet,i}}{w_{\bullet,k}} c_{i,t}\right)^2\right] = \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} \frac{w_{\bullet,i} w_{\bullet,j}}{w_{\bullet,k}^2} E[c_{i,t} c_{j,t}] = (w_{**} / w_{\bullet,k})' V_* (w_{**} / w_{\bullet,k})$$

Dado que la matriz  $V$  podría no ser simétrica, conviene definir la matriz  $V_*$  de forma que se garantice la simetría de  $V$ . Una manera de conseguirlo es definir  $V_*$  como

$$w_*' Y = \sum_{j=1}^s w_{j,1} y_{t_1} + \dots + \sum_{j=1}^s w_{j,k-1} y_{t_{k-1}} + \sum_{j=1}^s w_{j,k} y_{t_k} =$$

$$(w_{1,1} y_{t_1} + \dots + w_{1,k-1} y_{t_{k-1}} + w_{1,k} y_{t_k}) + \dots + (w_{s,1} y_{t_1} + \dots + w_{s,k-1} y_{t_{k-1}} + w_{s,k} y_{t_k}) = \sum_{j=1}^s g(j)$$

$$V_* = I - w_{**} w_{**}' / w_*' w_* = \begin{bmatrix} 1 - \frac{w_{\bullet,1}^2}{\sum_{i=1}^k w_{\bullet,i}^2} & -\frac{w_{\bullet,1} w_{\bullet,2}}{\sum_{i=1}^k w_{\bullet,i}^2} & \dots & -\frac{w_{\bullet,1} w_{\bullet,k-1}}{\sum_{i=1}^k w_{\bullet,i}^2} \\ -\frac{w_{\bullet,2} w_{\bullet,1}}{\sum_{i=1}^k w_{\bullet,i}^2} & 1 - \frac{w_{\bullet,2}^2}{\sum_{i=1}^k w_{\bullet,i}^2} & \dots & -\frac{w_{\bullet,2} w_{\bullet,k-1}}{\sum_{i=1}^k w_{\bullet,i}^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{w_{\bullet,k-1} w_{\bullet,1}}{\sum_{i=1}^k w_{\bullet,i}^2} & -\frac{w_{\bullet,k-1} w_{\bullet,2}}{\sum_{i=1}^k w_{\bullet,i}^2} & \dots & 1 - \frac{w_{\bullet,k-1}^2}{\sum_{i=1}^k w_{\bullet,i}^2} \end{bmatrix}$$

En este caso,

$$V_k = -V_*(w_{**} / w_{\bullet,k}) = w_{**}((w_{**}' w_{**} / w_*' w_*) - 1) / w_{\bullet,k} = -w_{**} (w_{\bullet,k} / w_*' w_*)$$

y

$$\begin{aligned} V_{kk} &= (w_{**} / w_{\bullet,kk})' V_*(w_{**} / w_{\bullet,k}) = w_{**}' w_{**} (1 - (w_{**}' w_{**} / w_*' w_*)) / w_{\bullet,k}^2 = \\ &= w_{**}' w_{**} / w_*' w_* = 1 - w_{\bullet,k}^2 / w_*' w_*, \end{aligned}$$

de modo que la matriz  $V$  puede expresarse como  $V = I - w_* w_*' / w_*' w_*$ .

Si alguna parte del patrón estacional evoluciona más rápidamente que otras, esta flexibilidad adicional puede recogerse incluyendo en la matriz de varianzas covarianzas  $\sigma^2 V$  elementos que recojan mayores varianzas en los nodos cercanos a los puntos de mayor variabilidad<sup>45</sup>.

### 2.3. Aplicaciones

En estos dos capítulos se han expuesto las herramientas metodológicas que permiten abordar las aplicaciones que darán contenido a los capítulos siguientes, en cada uno de los cuales se introducirán las adaptaciones de esta metodología que resulten necesarias por la especificidad del caso de estudio.

En la literatura específica del campo de las series temporales, los trabajos que utilizan series de considerable tamaño muestral pertenecen, generalmente, al ámbito financiero<sup>46</sup>. Sin embargo, las series empleadas en los capítulos siguientes para ilustrar el análisis de la estacionalidad en series de alta frecuencia corresponden, en cambio, a

<sup>45</sup> Harvey, Koopman y Riani (1997) utilizan *splines* estocásticas para recoger la evolución de la oferta monetaria semanal en el Reino Unido y encuentran que introduciendo varianzas mayores en los nodos próximos a las fiestas navideñas se obtiene un modelo que se ajusta mejor a la diferente velocidad con que cambia el patrón estacional en esa época del año. También para series diarias de agregados monetarios en la UE, Cabrero y otros (2002) utilizan *splines* estocásticas que asignan doble varianza a los nodos asociados al periodo navideño.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

contextos socioeconómicos de distinta índole. Aparte del interés propiamente económico o social que pueda tener el análisis de estas series en sus contextos respectivos, han sido utilizadas en este trabajo ya que, además de ser series de alta frecuencia, se observan con frecuencias distintas en cada caso. En concreto, se dispone de una serie de frecuencia semanal, otra diaria y, por último, una serie horaria. Además, y dado que a medida que aumenta la frecuencia es posible encontrar mayor número de fluctuaciones periódicas simultáneas, se ha iniciado el estudio analizando la serie de menor frecuencia para, a continuación, ir aumentando progresivamente el grado de complejidad en el estudio de la estacionalidad.

En el capítulo tercero se analiza la serie de exportaciones semanales de tomate canario hacia la UE. En este caso, la función *spline* actúa simplemente como una herramienta sustitutiva de las formulaciones convencionales con la ventaja de necesitar, frente a estas últimas, un número más reducido de regresores para explicar el mismo comportamiento.

El análisis del número diario de visitas registradas en un servicio de urgencias está contenido en el capítulo cuarto. En este caso, dada la frecuencia diaria con la que se observan los datos, es posible detectar la existencia de periodicidades de carácter semanal que se superponen a las observadas con periodo anual. Esta circunstancia aconseja especificar un modelo para el componente estacional que tenga en cuenta la superposición de estas fluctuaciones de distinto periodo. En concreto, y para recoger adecuadamente el componente estacional de periodo anual, ha sido necesario obtener una aproximación previa de esta fluctuación a partir de las medias por día del año de la serie original convenientemente corregida. Entonces, y admitiendo el carácter determinístico de esta variación, se ha optado por recogerla a través de una función *spline*, que se ha introducido como herramienta complementaria a la formulación trigonométrica estocástica empleada para la fluctuación estacional de periodo diario.

Finalmente, la serie de consumo horario de electricidad en España permite aplicar las funciones *splines* para modelar tres fluctuaciones paralelas de distinto periodo; en este caso diario, semanal y anual. Ahora, la complejidad analítica y computacional procede, no sólo del número de fluctuaciones superpuestas, sino, especialmente, del gran número de parámetros que precisa una formulación habitual que pretenda captar estas variaciones, ya sea a través de variables cualitativas o términos trigonométricos. En este caso, y tras los oportunos procesos de filtrado para obtener

---

<sup>46</sup> Entre otros, y como una aplicación reciente, puede citarse a Bajo y otros (2002).

aproximaciones iniciales de cada una de las variaciones estacionales, se ha recurrido, sin perder de vista el criterio de parsimonia, a dos funciones *splines* para las fluctuaciones de mayor periodo, que, junto con una formulación trigonométrica estocástica para la fluctuación diaria, han resultado adecuadas para la estimación simultánea del componente estacional.



### **3. La exportación semanal de tomate canario**

El cultivo del tomate de exportación en las Islas Canarias cuenta ya con una larga tradición<sup>1</sup>. Desde sus inicios a finales del siglo XIX y, sobre todo, en las últimas décadas del siglo XX, los agentes participantes en esta actividad se han caracterizado por un continuado esfuerzo innovador en sus estructuras productivas, así como por la introducción constante de mejoras en el empaquetado y la comercialización. El elemento motriz de este proceso modernizador ha sido la competencia nacional e internacional en los mercados europeos de destino. Sin embargo, además de las transformaciones estructurales, el sostenimiento del producto canario en mercados saturados de ofertas de diferentes orígenes —aunque cada vez más homogéneas en términos de variedades y condiciones de cultivo—, requiere adecuar la oferta a su correspondiente demanda, puesto que los incrementos desmesurados de la exportación canaria en determinados momentos en que ésta es relevante en el mercado, suelen venir acompañados de descensos en las cotizaciones. Por ello, sería deseable que los productores de Canarias planificaran sus envíos con el fin de garantizar cierto margen de rentabilidad. Desde este punto de vista, el estudio del patrón estacional de la exportación canaria de tomates se torna relevante para la toma de decisiones por parte de los productores. La competencia por épocas obliga a planificar la oferta en el tiempo, adaptándola a la demanda, aunque factores externos como la climatología pueden alterar la programación por su efecto sobre la oferta o sobre el consumo. La imposibilidad de almacenamiento concede aún más importancia a la estacionalidad<sup>2</sup>. En este capítulo se analiza, en concreto, la evolución de las exportaciones semanales de tomate canario a Europa en el periodo posterior al ingreso de España en la Unión Europea.

Ahora bien, antes de modelar las variaciones experimentadas por esta magnitud, conviene señalar algunas características de la actividad tomatera exportadora del archipiélago canario. En primer lugar, el patrón estacional de la exportación canaria se ha caracterizado históricamente por la concentración en el invierno y la desaparición en

---

<sup>1</sup> Véase Villalba (1978), Rodríguez (1986), Cortés (1989) y FEDEX (1994).

el verano, buscando las épocas de mejores precios. Por otra parte, el desarrollo de los cultivos de invernadero en los países del Norte de Europa<sup>3</sup>, el incremento de la producción peninsular y la aparición de producciones de terceros países que comparten el mismo periodo exportador que la producción canaria<sup>4</sup>, han determinado un solapamiento creciente de las diferentes ofertas en primavera y otoño. Finalmente, la plena integración española en la PAC, tras un largo periodo transitorio, ha significado la desaparición de las herramientas con que las producciones europeas se protegían comercialmente frente a la producción canaria<sup>5</sup>, y estos cambios han encontrado una respuesta inmediata en los productores de las Islas Canarias, a pesar de que los magrebíes han disfrutado también de una reducción notable de las barreras comunitarias<sup>6</sup>.

Si se tienen en cuenta estos aspectos, parece evidente que los volúmenes de tomate canario exportado en las diferentes semanas de la campaña han experimentado variaciones en los últimos años; la cuestión estriba en determinar si estas alteraciones ocurren en momentos puntuales o si se trata de cambios más continuos y aleatorios. Y en este sentido, el enfoque de los modelos estructurales constituye una propuesta adecuada para que sean los datos quienes determinen la naturaleza del patrón estacional.

La consideración de este enfoque, junto con las particulares restricciones que aparecen en el tratamiento econométrico de observaciones semanales y la existencia de periodos sin exportación de amplitud variable y, sobre todo, el empleo de las funciones *splines* como herramienta alternativa a las formulaciones convencionales del componente estacional son los ingredientes que definen la relevancia analítica del presente capítulo.

---

<sup>2</sup> Estudios como CIES (1968), Rodríguez y otros (1989), Briz (1990) y Billón (1995), señalan la relevancia de la estacionalidad.

<sup>3</sup> Véase, por ejemplo, Aldanondo (1995).

<sup>4</sup> El principal competidor extracomunitario es, sin lugar a dudas, Marruecos, que presenta ventajas competitivas en términos de costes de cultivo (Cáceres, 2000: 198-199). Véase también Parreño y Domínguez (1997).

<sup>5</sup> El sistema de precios de referencia impedía, en la práctica, que la producción canaria accediera a los mercados europeos a partir del mes de abril de cada año. Este sistema fue sustituido por otro más flexible —el sistema de precios de oferta— a partir de julio de 1991. Desde el 1 de enero de 1993 desapareció el sistema de precios de oferta y se liberalizaron las exportaciones canarias a la Comunidad, exceptuando el mantenimiento del mecanismo complementario aplicable a los intercambios (MCI), de muy escasa incidencia. La explicación detallada de estos instrumentos de protección puede consultarse en Cáceres (2000:292-305). El funcionamiento del sistema de precios de referencia se describe también en Albertos y otros (1987) y Cortés (1989).

<sup>6</sup> Desde la firma del Protocolo entre Marruecos y la UE en 1988 y hasta la plena integración española en la PAC, la producción magrebí accedía a los mercados europeos en similares condiciones que la producción canaria. Los acuerdos del GATT de 1994 significaban un endurecimiento de las condiciones

Para abordar los objetivos planteados, el capítulo se estructura como se indica a continuación. En el epígrafe siguiente se identifican los datos utilizados y se comentan algunos aspectos de interés sobre la naturaleza y tratamiento preliminar de éstos. En los epígrafes segundo y tercero se especifica y estima un modelo estructural de la serie de exportaciones semanales en el que la variación estacional, de carácter determinístico, se recoge a través de variables cualitativas o mediante funciones *splines* como formulación alternativa a las variables cualitativas estacionales. Finalmente, el epígrafe cuarto se dedica a exponer las conclusiones obtenidas sobre el patrón estacional de la exportación semanal de tomate canario y, desde un punto de vista metodológico, se hace un análisis comparativo de las dos formulaciones empleadas para recoger la estacionalidad.

### 3.1. Descripción de la serie semanal de exportaciones

En este capítulo se estudia la serie de exportaciones semanales de tomate canario a los mercados europeos, medidas en bultos de 6 kg, en el periodo comprendido entre las zafas 1986/1987 y 2000/2001<sup>7</sup>. Se considera que cada campaña se inicia en la semana 27 de un año y termina en la semana 26 del año siguiente. Los datos de la serie<sup>8</sup>, en adelante denotada por  $\{y_t\}_{t=1,\dots,780}$ , se representan en el gráfico 3.1.1.

La periodicidad semanal de las observaciones es necesaria para poder descubrir ciertos aspectos del comportamiento estacional de la magnitud económica analizada que podrían quedar ocultos como resultado de la agregación temporal implícita si se optara, por ejemplo, por datos mensuales<sup>9</sup>. Por otra parte, las asociaciones de exportadores registran semanalmente sus envíos al exterior, lo que puede interpretarse como indicio de que la semana es un periodo de tiempo que los agentes económicos consideran adecuado para percibir las reacciones del mercado y poder actuar en consecuencia.

---

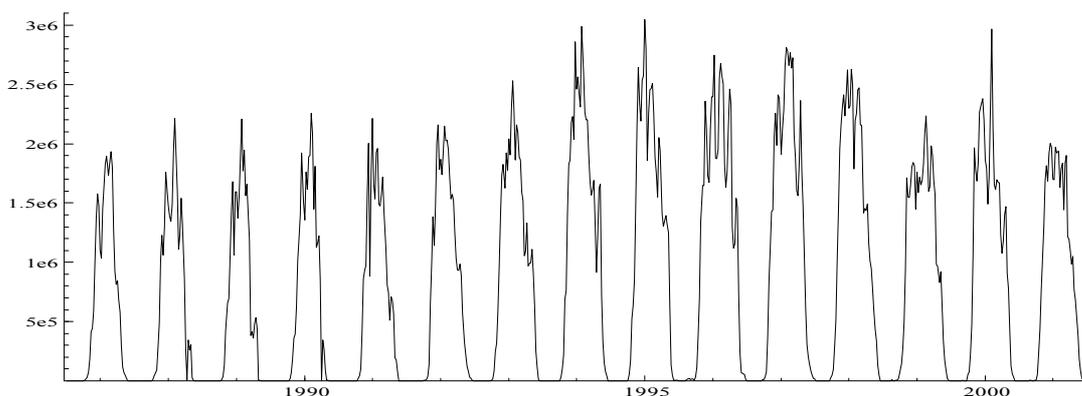
de acceso de la producción marroquí, pero estas condiciones quedaron sin efecto en virtud de los acuerdos comerciales UE -Marruecos de 1995 y 1996. Véase Cáceres (2000:278-281, 308-312).

<sup>7</sup> Los datos se han obtenido de las asociaciones provinciales de cosecheros-exportadores de Santa Cruz de Tenerife (ACETO) y Las Palmas (FEDEX). Las exportaciones a Europa se han calculado como resultado de la suma de las exportaciones al Reino Unido y las dirigidas al resto del Continente. En las semanas en las que estas fuentes no registraban dato alguno, se ha asignado un valor nulo.

<sup>8</sup> Véase el fichero *semanal.xls* del disco.

<sup>9</sup> La frecuencia semanal de observación se hace aún más necesaria si se pretende evaluar el impacto del volumen exportado sobre la cotización del fruto, dado que se trata de un mercado caracterizado por el movimiento dinámico hacia el equilibrio. Así, si los precios del tomate reaccionan ante los excesos de oferta y demanda de las últimas dos semanas es imposible detectar el proceso de ajuste hacia el equilibrio con datos mensuales. Refiriéndose, en general, a los productos hortofrutícolas españoles, Noguera (1996) advierte que la agregación temporal puede distorsionar la dinámica de las relaciones de causalidad entre precios a la producción y precios al consumo, provocando una aparente causalidad instantánea.

**Gráfico 3.1.1. Exportaciones de tomate canario a Europa (1986/1987 a 2000/2001)**



El tratamiento de esta serie semanal exige considerar problemas de índole metodológica, ya que a lo largo de un año no existen exactamente cincuenta y dos semanas, por lo que una semana de un año y la misma semana del año siguiente no recogen exactamente el mismo periodo del año. Por este motivo, en determinados años pueden incluirse observaciones correspondientes a cincuenta y tres semanas, lo que introduce un problema de heterogeneidad. Dado que en las series analizadas existía un largo periodo —el verano— sin observaciones, y teniendo en cuenta que en los años con cincuenta y tres semanas la zafra parecía comenzar una semana más tarde, se ha decidido adelantar una semana el inicio de zafra para estos años<sup>10</sup>.

Un análisis detallado de la evolución de la serie permite apreciar que en todas las campañas el patrón estacional de exportación se caracteriza por un movimiento ascendente, que se inicia con crecimientos pronunciados en octubre y que llega hasta enero-febrero, seguido de otro de descenso más suave hasta mayo o junio. Y, desde luego, no se observa que *los veranos se conviertan en inviernos*, como podría ocurrir si el comportamiento estacional estuviese dominado por un componente estocástico de elevada varianza en relación al nivel de la serie. Ahora bien, es posible distinguir periodos diferenciados por la extensión de la zafra y, sobre todo, por el volumen exportado en determinadas semanas de la campaña. La plena integración en la UE el uno de enero de 1993, y la consecuente desaparición de los precios de referencia/oferta,

---

<sup>10</sup> Podrían haberse adoptado otras soluciones al problema de heterogeneidad en los datos imponiendo restricciones de continuidad entre las semanas 52 y 53, que podrían haber sido consideradas mediante la especificación de funciones *splines*. De hecho, planteamientos de este tipo han tenido que ser aplicados en los capítulos siguientes. Pero, dado que la aparición de años con 53 semanas se debe más bien a un problema de asignación y no, obviamente, a la existencia de años con 53 semanas completas, ha parecido más oportuna la decisión adoptada, que permite la utilización de modelos convencionales de la estacionalidad y aporta, por tanto, elementos de comparación.

se traduce en un notable impulso exportador<sup>11</sup>. El crecimiento general de la exportación en este segundo periodo se trunca a partir de 1996, coincidiendo con la entrada en vigor del acuerdo comercial entre la UE y Marruecos<sup>12</sup>. A pesar de los cambios observados en el ritmo de crecimiento y en los niveles semanales medios de exportación, el patrón estacional es más o menos regular, con máximos y mínimos localizados casi siempre en torno a las mismas semanas de la zafra<sup>13</sup>.

Por tanto, y a modo de hipótesis preliminar, podría considerarse que existe un componente estacional estable alrededor de un componente tendencial variable. Sin embargo, las inestabilidades de los diferentes componentes de la serie serán analizadas mediante criterios estadísticos y sin que el investigador establezca, al menos *a priori*, patrones de cambio tan definidos. En primer lugar, se estiman modelos estructurales con las especificaciones habituales y se concluye que la estacionalidad posee un carácter determinista, que queda bien recogido a través de modelos convencionales. A pesar de ello, y con objeto de mostrar su utilidad como alternativa a la formulación anterior, se procede a incorporar una función *spline* que recoja también las variaciones estacionales. Los resultados de estos dos análisis se exponen con detalle en los epígrafes siguientes<sup>14</sup>.

### 3.2. Modelo estructural con variables cualitativas estacionales

A pesar de que la representación gráfica de la serie bajo estudio no sugiere la existencia de cambios en el patrón estacional, debe tenerse en cuenta, como señala Scott (1995), que en muchas ocasiones se hace compatible la presencia de estacionalidad estocástica y la ausencia de grandes cambios en los valores observados en cada estación, en tanto que los valores iniciales pueden ser muy grandes en relación con las varianzas. Por ello, se optó por iniciar la estrategia de identificación de la naturaleza de los

---

<sup>11</sup> Ya desde julio de 1991 —sustitución de los precios de referencia por los de oferta— se había apreciado que las campañas, que solían terminar a principios de mayo —con la entrada en vigor de dichos precios—, se prolongan hasta junio.

<sup>12</sup> Véase nota 6.

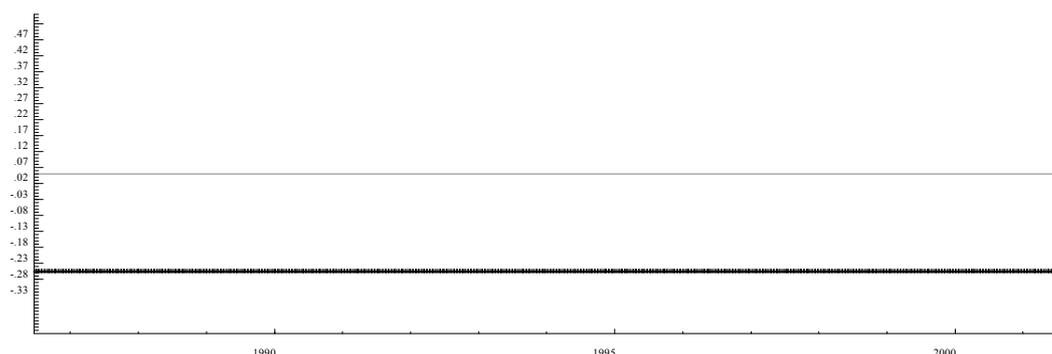
<sup>13</sup> Véase anexo 3.1.

<sup>14</sup> Se ha optado por efectuar el análisis sobre la serie en niveles. Aunque la transformación logarítmica se aplica habitualmente para eliminar los problemas de no estacionariedad en varianza, la observación del gráfico 3.1.1 no muestra que esta transformación sea necesaria. Si aún así se decidiera efectuar la transformación, podrían asignarse ceros a las exportaciones nulas. Pero en ese caso, el paso de una exportación nula a una exportación no nula en la misma semana de la zafra siguiente sería mucho más importante y podría distorsionar, de modo significativo, la percepción del patrón estacional. Precisamente, la brusquedad del cambio de nivel al principio y final de zafra inducida por la transformación logarítmica explica que la mayor parte de las intervenciones requeridas para modelar la serie transformada se ubiquen en esos periodos del año. Estos resultados se muestran en el anexo 3.2.

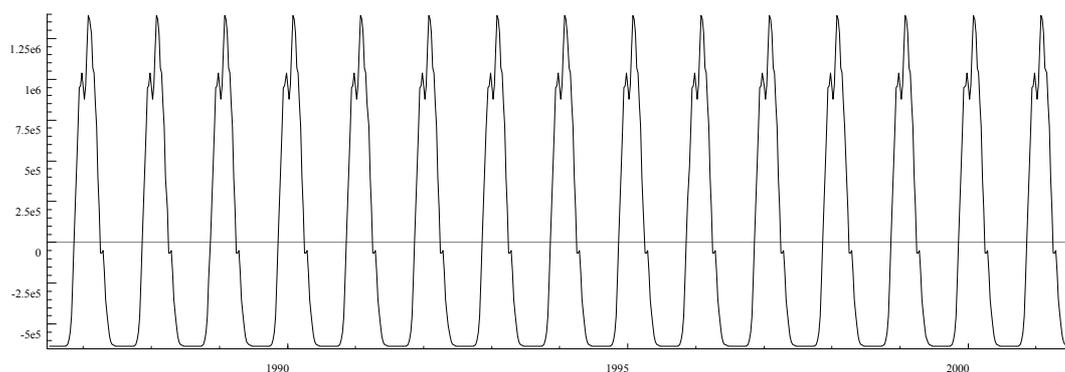
## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

componentes mediante la estimación del modelo estructural básico<sup>15</sup>. La estimación máximo verosímil de este modelo revela que la varianza de los términos de perturbación del componente estacional y de la pendiente son nulas, lo que manifiesta el carácter no estocástico de tales componentes. Este resultado permite evaluar la presencia de dichos componentes en el modelo a partir de los contrastes habituales de significación de su valor en el vector de estado final. El estadístico de significación individual de la pendiente sugiere que el modelo debe simplificarse eliminando este término en la tendencia. Respecto al componente estacional, a pesar de su comportamiento no estocástico y de la no significación estadística individual del parámetro correspondiente a algunas estaciones, el estadístico de significación conjunta del efecto estacional confirma la relevancia de este componente en el comportamiento de la serie<sup>16</sup>. La importancia de este componente se aprecia claramente si se compara el rango de variación de éste y el de la serie (gráficos 3.1.1 y 3.2.2)<sup>17</sup>.

**Gráfico 3.2.1. Pendiente del componente de nivel**



**Gráfico 3.2.2. Componente estacional**



<sup>15</sup> Véase capítulo 1:20.

<sup>16</sup> El valor de dicho estadístico, que evalúa el efecto conjunto del componente estacional, fue de 356.062. Este estadístico sigue una distribución  $\chi^2$  con  $s-1$  grados de libertad, 51 en este caso al tratarse de datos semanales.

<sup>17</sup> Aunque la presencia de observaciones nula obliga a interpretar con cautela la magnitud del componente estacional, éste representa, en términos medios, más del 75% del valor medio de la serie a lo largo de la muestra.

La representación gráfica de los componentes estimados para el modelo anterior (gráficos 3.2.1 y 3.2.2)<sup>18</sup> resulta sumamente reveladora, dado que refleja el carácter fijo del patrón estacional a lo largo de la muestra, así como la estabilidad de la pendiente en torno a un valor realmente pequeño en relación con la magnitud de las observaciones. Teniendo en cuenta estos resultados, se plantea un modelo en el que el patrón estacional es fijo, pero, con objeto de examinar las implicaciones de diferentes modelos para la tendencia, se permite que cada uno de sus elementos (nivel y pendiente) mantenga su formulación estocástica, es decir,

$$y_t = \mathbf{m}_t + \mathbf{g}_{1,t}^{anual} + \mathbf{e}_t, \mathbf{e}_t \sim NID(0, \sigma_{\epsilon}^2),$$

con

$$\mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{b}_{t-1} + \mathbf{h}_t, \mathbf{h}_t \sim NID(0, \sigma_{\eta}^2),$$

$$\mathbf{b}_t = \mathbf{b}_{t-1} + \mathbf{V}_t, \mathbf{V}_t \sim NID(0, \sigma_{\zeta}^2),$$

y

$$\mathbf{g}_{1,t}^{anual} = \sum_{j=1}^{51} z_{j,t} \mathbf{g}_{1,j}^{anual},$$

siendo  $z_{j,t}$ ,  $j = 1, \dots, 51$ , igual a 1 si la observación pertenece a la semana  $j$  del año, -1 si la observación pertenece a la semana 52 y 0 en otro caso<sup>19</sup>. De nuevo, el proceso de estimación conduce a la conclusión de que la varianza del término de perturbación que dirige el comportamiento de la pendiente es nula y el estadístico de significación individual en el vector de estado final manifiesta la no significación estadística de este componente.

Aceptando que el patrón estacional permanece inalterado con el transcurso del tiempo y que la pendiente es nula, se especificó un modelo en el que se incluyen los mismos elementos pero que difiere únicamente en la naturaleza del componente de nivel, es decir,

$$\mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{h}_t, \mathbf{h}_t \sim NID(0, \sigma_{\eta}^2).$$

<sup>18</sup> En todos los modelos de éste y de los restantes capítulos se ha utilizado la opción de alisado (*smoothed*) del programa STAMP 6.0 para obtener estimaciones de los componentes (véase Koopman y otros, 2000, y capítulo 1:42-48).

<sup>19</sup> Recuérdese que se ha considerado que cada campaña comienza en la semana 27 del año ( $j = 27$ ) y concluye en la semana 26 del año siguiente ( $j = 26$ ). En la notación del componente estacional se ha empleado el superíndice anual para indicar el periodo en el que se completa la fluctuación estacional; mientras que, el subíndice uno permite distinguir esta formulación de las especificadas en el epígrafe siguiente.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

Los resultados obtenidos para este modelo ratifican el carácter estocástico del componente de nivel<sup>20</sup> y, por tanto, nos llevan a concluir que el modelo estructural que parece adecuado debe incluir un componente de nivel estocástico, estacionalidad fija y un componente irregular (cuadros 3.2.1 y 3.2.2).

**Cuadro 3.2.1. Varianzas de los términos de perturbación**

	Irregular ( $\hat{S}_e^2$ )	Nivel ( $\hat{S}_h^2$ )
<b>Varianza</b>	7.3092x10 <sup>9</sup>	3.4810x10 <sup>10</sup>
<b>q-ratio</b> <sup>21</sup>	(0.2100)	(1.0000)

**Cuadro 3.2.2. Componente de nivel en el vector de estado final y componente estacional,  $\hat{g}_{1,j}^{anual}$**

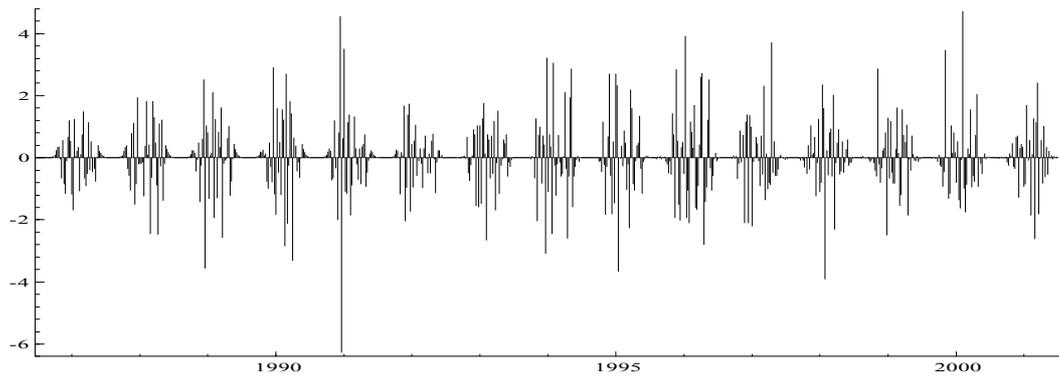
Estimación		Estimación		Estimación		Estimación	
$\hat{m}_T$	842736.624	$\hat{g}_{1,40}^{anual}$	-830923.889	$\hat{g}_{1,1}^{anual}$	1093762.800	$\hat{g}_{1,14}^{anual}$	250661.960
$\hat{g}_{1,27}^{anual}$	-844986.247	$\hat{g}_{1,41}^{anual}$	-803090.959	$\hat{g}_{1,2}^{anual}$	1053871.870	$\hat{g}_{1,15}^{anual}$	253346.690
$\hat{g}_{1,28}^{anual}$	-845911.517	$\hat{g}_{1,42}^{anual}$	-715749.629	$\hat{g}_{1,3}^{anual}$	1019639.000	$\hat{g}_{1,16}^{anual}$	145785.821(*)
$\hat{g}_{1,29}^{anual}$	-845908.454	$\hat{g}_{1,43}^{anual}$	-561990.966	$\hat{g}_{1,4}^{anual}$	1179661.330	$\hat{g}_{1,17}^{anual}$	37784.484(*)
$\hat{g}_{1,30}^{anual}$	-845905.390	$\hat{g}_{1,44}^{anual}$	-288886.835	$\hat{g}_{1,5}^{anual}$	1290264.460	$\hat{g}_{1,18}^{anual}$	-103915.853(*)
$\hat{g}_{1,31}^{anual}$	-845902.327	$\hat{g}_{1,45}^{anual}$	49357.228(*)	$\hat{g}_{1,6}^{anual}$	1323006.190	$\hat{g}_{1,19}^{anual}$	-164936.589(*)
$\hat{g}_{1,32}^{anual}$	-845899.263	$\hat{g}_{1,46}^{anual}$	307463.425	$\hat{g}_{1,7}^{anual}$	1262337.920	$\hat{g}_{1,20}^{anual}$	-397074.326
$\hat{g}_{1,33}^{anual}$	-845076.200	$\hat{g}_{1,47}^{anual}$	575378.688	$\hat{g}_{1,8}^{anual}$	1142549.180	$\hat{g}_{1,21}^{anual}$	-623716.462
$\hat{g}_{1,34}^{anual}$	-844274.470	$\hat{g}_{1,48}^{anual}$	754885.085	$\hat{g}_{1,9}^{anual}$	1046456.910	$\hat{g}_{1,22}^{anual}$	-737868.199
$\hat{g}_{1,35}^{anual}$	-844320.740	$\hat{g}_{1,49}^{anual}$	903688.082	$\hat{g}_{1,10}^{anual}$	884680.640	$\hat{g}_{1,23}^{anual}$	-797620.602
$\hat{g}_{1,36}^{anual}$	-844691.010	$\hat{g}_{1,50}^{anual}$	1005056.080	$\hat{g}_{1,11}^{anual}$	733614.503	$\hat{g}_{1,24}^{anual}$	-825465.272
$\hat{g}_{1,37}^{anual}$	-844662.613	$\hat{g}_{1,51}^{anual}$	1087419.410	$\hat{g}_{1,12}^{anual}$	552071.500	$\hat{g}_{1,25}^{anual}$	-838057.675
$\hat{g}_{1,38}^{anual}$	-844493.683	$\hat{g}_{1,52}^{anual}$	1152126.670	$\hat{g}_{1,13}^{anual}$	411962.030	$\hat{g}_{1,26}^{anual}$	-843662.012
$\hat{g}_{1,39}^{anual}$	-841840.753						

(\*) Parámetros no significativos al 95 % de confianza.

<sup>20</sup> La varianza estimada del término de perturbación del componente de nivel fue  $\hat{S}_h^2 = 3.4810 \times 10^{10}$ .

<sup>21</sup> Se ha optado por utilizar el término q-ratio, en detrimento de la expresión ratio señal ruido, debido a que el componente que tiene máxima varianza no es el componente irregular sino el término de perturbación del componente de nivel.

Gráfico 3.2.3. Error de predicción estandarizado



Respecto al componente estacional, los estadísticos de significación individual en el vector de estado final revelan la existencia de variaciones estacionales distintas de cero para la mayoría de las semanas<sup>22</sup>. Ahora bien, el examen de los residuos o errores de predicción en cada momento del tiempo y, especialmente, su representación gráfica (gráfico 3.2.3) muestran la presencia de observaciones anómalas que exigen reespecificar el modelo en los términos siguientes:

$$y_t = \mathbf{m}_t + \mathbf{g}_{1,t}^{anual} + \sum_{j-k} \mathbf{I}_{j-k} I_{j-k} + \mathbf{e}_t, \mathbf{e}_t \sim NID(0, \sigma_e^2),$$

con

$$\mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{h}_t, \mathbf{h}_t \sim NID(0, \sigma_h^2),$$

y

$$\mathbf{g}_{1,t}^{anual} = \sum_{j=1}^{51} z_{j,t} \mathbf{g}_{1,j}^{anual},$$

siendo  $z_{j,t}$ ,  $j = 1, \dots, 51$ , igual a 1 si la observación pertenece a la semana  $j$  del año, -1 si la observación pertenece a la semana 52 y 0 en otro caso, e  $I_{j-k}$  una variable de intervención que toma valor uno si la observación pertenece a la semana  $j$  del año  $k$ . El estudio de observaciones anómalas conduce finalmente al modelo cuya estimación proporciona los resultados que se muestran en los cuadros 3.2.3 a 3.2.5 y los gráficos 3.2.4 a 3.2.6.

Cuadro 3.2.3. Varianzas de los términos de perturbación

	Irregular ( $\hat{\sigma}_e^2$ )	Nivel ( $\hat{\sigma}_h^2$ )
Varianza	1.9152x10 <sup>8</sup>	2.9907x10 <sup>10</sup>
q-ratio	(0.0064)	(1.0000)

<sup>22</sup> Los resultados de los tests de significación individual de los componentes de este modelo no difieren sustancialmente de los obtenidos para el modelo finalmente propuesto, que aparecen en el cuadro 3.2.2.

Cuadro 3.2.4. Componente de nivel en el vector de estado final y componente estacional,  $\hat{g}_{1,j}^{anual}$

Estimación		Estimación		Estimación		Estimación	
$\hat{m}_T$	840642.639	$\hat{g}_{1,40}^{anual}$	-827886.630	$\hat{g}_{1,1}^{anual}$	1061852.340	$\hat{g}_{1,14}^{anual}$	222576.746
$\hat{g}_{1,27}^{anual}$	-841930.730	$\hat{g}_{1,41}^{anual}$	-800055.105	$\hat{g}_{1,2}^{anual}$	1056889.460	$\hat{g}_{1,15}^{anual}$	194096.372
$\hat{g}_{1,28}^{anual}$	-842857.404	$\hat{g}_{1,42}^{anual}$	-712715.179	$\hat{g}_{1,3}^{anual}$	1120640.630	$\hat{g}_{1,16}^{anual}$	111069.460(*)
$\hat{g}_{1,29}^{anual}$	-842855.745	$\hat{g}_{1,43}^{anual}$	-558957.920	$\hat{g}_{1,4}^{anual}$	1239232.210	$\hat{g}_{1,17}^{anual}$	40781.012(*)
$\hat{g}_{1,30}^{anual}$	-842854.086	$\hat{g}_{1,44}^{anual}$	-285855.195	$\hat{g}_{1,5}^{anual}$	1394917.890	$\hat{g}_{1,18}^{anual}$	-141886.036(*)
$\hat{\gamma}_{1,31}^{anual}$	-842852.427	$\hat{g}_{1,45}^{anual}$	-19299.264(*)	$\hat{g}_{1,6}^{anual}$	1272382.060	$\hat{g}_{1,19}^{anual}$	-211131.885
$\hat{g}_{1,32}^{anual}$	-842850.768	$\hat{g}_{1,46}^{anual}$	310492.256	$\hat{g}_{1,7}^{anual}$	1265348.490	$\hat{g}_{1,20}^{anual}$	-394082.012
$\hat{g}_{1,33}^{anual}$	-842029.109	$\hat{g}_{1,47}^{anual}$	547146.008	$\hat{g}_{1,8}^{anual}$	1145558.350	$\hat{g}_{1,21}^{anual}$	-620725.553
$\hat{g}_{1,34}^{anual}$	-841228.784	$\hat{g}_{1,48}^{anual}$	757911.108	$\hat{g}_{1,9}^{anual}$	1079810.270	$\hat{g}_{1,22}^{anual}$	-734878.694
$\hat{g}_{1,35}^{anual}$	-841276.458	$\hat{g}_{1,49}^{anual}$	906712.700	$\hat{g}_{1,10}^{anual}$	887686.999	$\hat{g}_{1,23}^{anual}$	-794632.502
$\hat{g}_{1,36}^{anual}$	-841648.133	$\hat{g}_{1,50}^{anual}$	1008079.290	$\hat{g}_{1,11}^{anual}$	736619.458	$\hat{g}_{1,24}^{anual}$	-822478.576
$\hat{g}_{1,37}^{anual}$	-841621.140	$\hat{g}_{1,51}^{anual}$	1215830.710	$\hat{g}_{1,12}^{anual}$	515267.972	$\hat{g}_{1,25}^{anual}$	-835072.384
$\hat{g}_{1,38}^{anual}$	-841453.615	$\hat{g}_{1,52}^{anual}$	1112843.870	$\hat{g}_{1,13}^{anual}$	340849.888	$\hat{g}_{1,26}^{anual}$	-840678.125
$\hat{g}_{1,39}^{anual}$	-838802.089						

(\*) Parámetros no significativos al 95 % de confianza.

Cuadro 3.2.5. Parámetros de las variables de intervención

Parámetro	Estimación	Parámetro	Estimación	Parámetro	Estimación
$\hat{I}_{51-88}$	$-7.2953 \times 10^5$	$\hat{I}_{12-95}$	$5.9711 \times 10^5$	$\hat{I}_{15-96}$	$9.3374 \times 10^5$
$\hat{I}_{14-90}$	$-5.5540 \times 10^5$	$\hat{I}_{13-95}$	$5.3608 \times 10^5$	$\hat{I}_{16-96}$	$5.6571 \times 10^5$
$\hat{I}_{51-90}$	$-1.1513 \times 10^6$	$\hat{I}_{47-95}$	$4.6890 \times 10^5$	$\hat{I}_{05-98}$	$-6.9724 \times 10^5$
$\hat{I}_{01-91}$	$5.2394 \times 10^5$	$\hat{I}_{03-96}$	$-8.2811 \times 10^5$	$\hat{I}_{45-98}$	$4.9837 \times 10^5$
$\hat{I}_{52-93}$	$6.3455 \times 10^5$	$\hat{I}_{04-96}$	$-8.4834 \times 10^5$	$\hat{I}_{45-99}$	$5.7694 \times 10^5$
$\hat{I}_{18-94}$	$6.1448 \times 10^5$	$\hat{I}_{05-96}$	$-8.2736 \times 10^5$	$\hat{I}_{06-00}$	$8.0454 \times 10^5$
$\hat{I}_{19-94}$	$7.3784 \times 10^5$	$\hat{I}_{13-96}$	$5.7563 \times 10^6$	$\hat{I}_{09-01}$	$-4.5518 \times 10^5$
$\hat{I}_{03-95}$	$-6.4167 \times 10^5$	$\hat{I}_{14-96}$	$1.0217 \times 10^5$		

Aunque *a priori* parece difícil pensar que la producción de tomate en Canarias pueda cambiar de una semana a otra, sí existen factores que pueden producir estos cambios, básicamente adelantando o retrasando la recogida del fruto. Así, la bondad de las temperaturas en el otoño es el principal responsable del incremento exportador en el principio de zafra de algunas campañas. No es extraño tampoco que los agricultores aceleren la recogida del fruto en épocas de altos precios, o simplemente, que se descuide

algo la selección y se exporten más partidas de las que se hubieran enviado en otras circunstancias. En concreto, la mayor parte de las anomalías observadas en la serie de exportaciones pueden ser explicadas por las decisiones de los exportadores a la vista del comportamiento pasado y de las previsiones futuras de las cotizaciones<sup>23</sup>, aunque a la hora de establecer relaciones entre volumen y precio debe tenerse en cuenta que los envíos registrados en origen tardan varios días en llegar a su destino y ser comercializados<sup>24</sup>. En el caso particular de la semana 14 de 1990, los envíos canarios se paralizaron con la intención de evitar que las caídas de precios terminaran suponiendo la aplicación de tasas compensatorias<sup>25</sup>. Los elevados niveles de exportación de principios de la campaña 88/89 motivaron caídas de precios que podrían justificar la notable reducción en los envíos en la semana 51 de 1988<sup>26</sup>. Un argumento similar podría utilizarse para explicar el descenso de las exportaciones en la semana 3 de 1995, en las semanas 3, 4 y 5 de 1996 o en la semana 5 de 1998. La disminución de los envíos en la semana 9 de 2001 con respecto a las semanas previas parece haber tenido un efecto positivo en las cotizaciones de las semanas siguientes, que recuperaron los niveles de principios de año. Por el contrario, los elevados volúmenes exportados en las semanas 45 de 1998 y 1999 tienen como denominador común los altos precios registrados en los inicios de campaña respectivos. También se registraron precios por encima de lo normal en las semanas previas a las semanas 1 de 1991, 18 y 19 de 1994<sup>27</sup>, 12 y 13 de 1995<sup>28</sup> y 6 de 2000. Del mismo modo, las mejoras de los precios en febrero y marzo de 1996 podrían explicar el incremento exportador en las semanas 13 a 16 de ese año<sup>29</sup>.

A pesar de las buenas cotizaciones en las semanas previas a la 51 de 1990, se registra un descenso brusco de las exportaciones que quizás puede estar asociado a los

---

<sup>23</sup> Los precios del tomate canario en los mercados mayoristas europeos pueden consultarse en la página web <http://www.mcx.es/sgcomex/home1fra.htm> de la Secretaría General de Comercio Exterior del Ministerio de Economía, así como, entre otras fuentes, en las publicaciones del SOIVRE (Servicios Oficiales de Inspección, Vigilancia y Regulación de las Exportaciones) y del ZMP (Zentrale Markt und Preisberichtsstelle). Para campañas correspondientes a los años 80 pueden consultarse también las memorias de campaña de ACETO y FEDEX Véase, además, Cáceres (2000:353).

<sup>24</sup> Según FEDEX (1994), entre cosecha, acondicionamiento y carga en barco pueden pasar hasta dos o tres días y la duración del trayecto desde Canarias a los puertos europeos es de cuatro días.

<sup>25</sup> Véase Cáceres (2000:218).

<sup>26</sup> Las bajas cotizaciones de inicio de campaña pueden ser explicadas por los altos niveles de la oferta holandesa y peninsular, los envíos desde Marruecos, los excesos de exportación de Canarias hasta mediados de diciembre y la falta de calidad de estas exportaciones (Martínez, 1989).

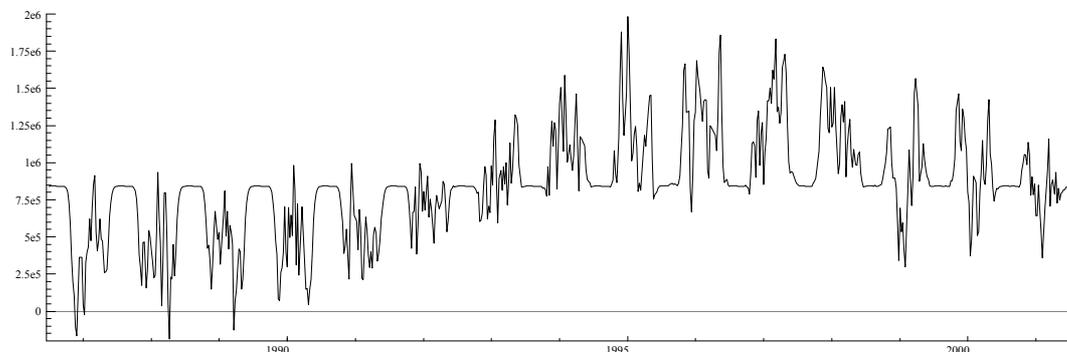
<sup>27</sup> Pascual y otros (1994) indican que en estas semanas se descuidó la selección y, por consiguiente, se envió más fruta de la debida.

<sup>28</sup> Pascual (1996a) señala de nuevo el crecimiento de los envíos, muchos sin la calidad debida para aprovechar la coyuntura, como factor que contribuyó al derrumbe de los precios al final de la campaña.

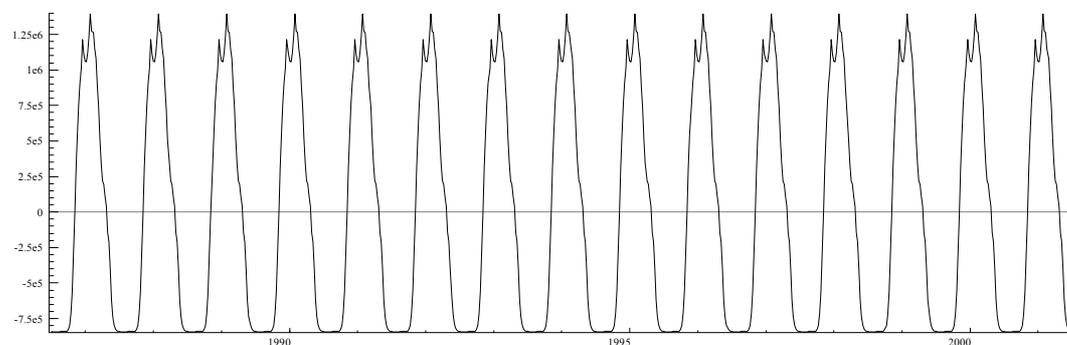
## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

elevados volúmenes enviados en la semana previa. Los incrementos registrados en las semanas 52 de 1993 o 47 de 1995 están, en cambio, relacionados con la benignidad del clima en ambos inicios de zafra<sup>30</sup>.

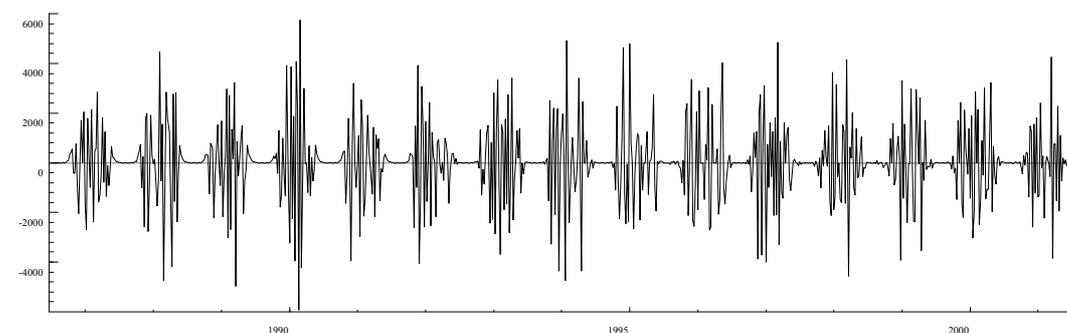
**Gráfico 3.2.4. Componente de nivel**



**Gráfico 3.2.5. Componente estacional**



**Gráfico 3.2.6. Componente irregular**



La elevada varianza del componente de nivel parece confirmar la existencia de cambios importantes en el movimiento a largo plazo de la serie que se detectaron en su representación gráfica y que fueron comentados en el primer epígrafe de este capítulo<sup>31</sup>. Por otra parte, la estimación del componente estacional confirma el patrón exportador

<sup>29</sup> Pascual (1996b) apunta a la saturación del mercado y los problemas de calidad asociados a las lluvias, el envejecimiento de los cultivos y una selección poco cuidada como responsables de los descensos de precios en abril de 1996.

<sup>30</sup> Véase Pascual y otros (1994) y Pascual (1996a).

ya descrito, con un crecimiento paulatino desde el inicio de campaña hasta los meses de enero-febrero y un descenso posterior. Este comportamiento suave de la variación estacional explica la ausencia de significación de este componente en las semanas con niveles más próximos a los niveles medios del conjunto de la campaña (cuadro 3.2.4).

### 3.3. Modelo estructural con funciones *spline*

Los resultados encontrados en el apartado anterior permiten concluir que la estacionalidad es determinística. Una vez que se dispone de una estimación del componente estacional presente en la serie, una formulación alternativa de dicho componente puede construirse mediante funciones *splines* cúbicas periódicas definidas en términos de la semana  $w$  de la campaña, donde  $w = j - 26$ , si  $j = 27, \dots, 52$ , y  $w = j + 26$ , si  $j = 1, \dots, 26$ <sup>32</sup>. Con objeto de identificar el número de tramos adecuado, se ha recurrido a la representación gráfica del componente estacional estimado en el apartado anterior (gráfico 3.2.5),  $\hat{g}_{1,w}^{anual}$ , que sugiere formular una *spline* de diez tramos con la particularidad de que en el primero parece más adecuada una especificación lineal<sup>33</sup>. Dado que la estacionalidad es determinística, la componente estacional  $g_t^{anual}$  puede denotarse por  $g_t^{anual} = \hat{g}_{1,w}^{anual}$ , si la observación en el instante  $t$  corresponde a la semana  $w$ . Entonces, la fluctuación estacional  $\hat{g}_{1,w}^{anual}$  puede ajustarse a través de la *spline*  $g(w)$ , es decir,  $\hat{g}_{1,w}^{anual} = g(w) + e_w$ , donde  $e_w$  es un término de error asociado al ajuste y  $g(w)$  es la *spline*, finalmente especificada como:

$$g(w) = \begin{cases} g_{1,0} + g_{1,1}w & , \quad w_0 \leq w \leq w_1 \\ g_{i,0} + g_{i,1}w + g_{i,2}w^2 + g_{i,3}w^3 & , \quad w_{i-1} \leq w \leq w_i, \quad i = 2, \dots, 10 \end{cases}$$

con  $w_0 = 1$  y  $w_{10} = 52$ . Esta *spline* debe verificar las siguientes condiciones:

- a) restricciones de continuidad de la función,

<sup>31</sup> Cáceres (1997, 2001) y Martín y otros (2002) proponen una especificación de componentes determinísticos cambiantes como alternativa, también válida, para recoger estas inestabilidades, aunque, obviamente, más rígida que la que aportan los modelos estructurales.

<sup>32</sup> Es obvio que la *spline* no tiene por qué ajustarse a un periodo que comience con la estación ( $j = 1$ ). De hecho, como señalan Cabrero y otros (2002:17), la restricción de periodicidad de la *spline* aconseja que el primer y último punto de ruptura se asignen a estadios estacionales con comportamiento similar. Por supuesto, si la *spline* recoge un comportamiento estacional con cambios suaves entre estaciones consecutivas, como de hecho se presupone cuando se recurre a estas funciones, la elección del punto de inicio es menos relevante.

<sup>33</sup> La elección de una forma funcional cúbica para el resto de tramos está justificada por los argumentos ya comentados en el capítulo 2:79-81.

$$g_{1,0} + g_{1,1}w_1 = g_{2,0} + g_{2,1}w_1 + g_{2,2}w_1^2 + g_{2,3}w_1^3 ,$$

$$g_{i,0} + g_{i,1}w_i + g_{i,2}w_i^2 + g_{i,3}w_i^3 = g_{i+1,0} + g_{i+1,1}w_i + g_{i+1,2}w_i^2 + g_{i+1,3}w_i^3 ,$$

para  $i = 2, \dots, 9$ , y

$$g_{10,0} + g_{10,1}(w_{10} + 1) + g_{10,2}(w_{10} + 1)^2 + g_{10,3}(w_{10} + 1)^3 = g_{1,0} + g_{1,1}w_0 ;$$

b) restricciones de continuidad de la derivada primera,

$$g_{1,1} = g_{2,1} + 2g_{2,2}w_1 + 3g_{2,3}w_1^2 ,$$

$$g_{i,1} + 2g_{i,2}w_i + 3g_{i,3}w_i^2 = g_{(i+1),1} + 2g_{(i+1),2}w_i + 3g_{(i+1),3}w_i^2 ,$$

para  $i = 2, \dots, 9$ , y

$$g_{10,1} + 2g_{10,2}(w_{10} + 1) + 3g_{10,3}(w_{10} + 1)^2 = g_{1,1} ;$$

c) restricciones de continuidad de la derivada segunda<sup>34</sup>,

$$2g_{i,2} + 6g_{i,3}w_i = 2g_{i+1,2} + 6g_{i+1,3}w_i, \quad i = 2, \dots, 9 .$$

Una vez especificada la forma funcional de la *spline* y si se asume que las localizaciones  $w_i$ ,  $i = 0, \dots, 10$ , son conocidas<sup>35</sup>, el siguiente paso es determinar los parámetros  $g_{1,0}$  y  $g_{1,1}$ , así como  $g_{i,0}$ ,  $g_{i,1}$ ,  $g_{i,2}$  y  $g_{i,3}$ ,  $i = 2, \dots, 10$ . Dado que el número de parámetros que se desea determinar, 38, es superior al número de condiciones impuestas, 28, se ha optado por expresar el vector de parámetros  $G_{28 \times 1} = (g_{1,1}, g_{2,1}, g_{2,2}, g_{2,3}, \dots, g_{10,1}, g_{10,2}, g_{10,3})'$  como función de  $G_{10 \times 1}^* = (g_{1,0}, g_{2,0}, \dots, g_{10,0})'$ , que son los parámetros finalmente estimados, y de  $w_i$ ,  $i = 0, \dots, 10$ <sup>36</sup>. Teniendo en cuenta el planteamiento del problema, conviene expresar el conjunto de restricciones anteriores de forma matricial como  $WG = R$ , donde

<sup>34</sup> No se han considerado las restricciones  $\nabla^2 g_1(w_1) = \nabla^2 g_2(w_1)$  ni  $\nabla^2 g_{10}(w_{10} + 1) = \nabla^2 g_1(w_0)$ , dada la especificación lineal del primer tramo.

<sup>35</sup> Más adelante se determinará la localización de los puntos de ruptura que delimitan cada uno de los tramos de la función *spline*, siguiendo el procedimiento descrito en el capítulo segundo.

<sup>36</sup> Otra opción consiste en imponer restricciones adicionales como, por ejemplo, el paso por los nodos (véase capítulo 2:105-109).

$$W_{28 \times 28} = \begin{bmatrix} w_1 & -w_1 & -w_1^2 & -w_1^3 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & w_2 & w_2^2 & w_2^3 & -w_2 & -w_2^2 & -w_2^3 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & w_9 & w_9^2 & w_9^3 & -w_9 & -w_9^2 & -w_9^3 \\ -w_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & (w_{10}+1) & (w_{10}+1)^2 & (w_{10}+1)^3 \\ 1 & -1 & -2w_1 & -3w_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2w_2 & 3w_2^2 & -1 & -2w_2 & -3w_2^2 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 2w_9 & 3w_9^2 & -1 & -2w_9 & -3w_9^2 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2(w_{10}+1) & 3(w_{10}+1)^2 \\ 0 & 0 & 2 & 6w_2 & 0 & -2 & -6w_2 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 6w_3 & 0 & -2 & -6w_3 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 2 & 6w_8 & 0 & -2 & -6w_8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 2 & 6w_9 & 0 & -2 & -6w_9 \end{bmatrix}$$

y

$$R_{28 \times 1} = \begin{bmatrix} -g_{1,0} + g_{2,0} \\ -g_{2,0} + g_{3,0} \\ \vdots \\ -g_{9,0} + g_{10,0} \\ -g_{10,0} + g_{1,0} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

de modo que  $G = W^{-1}R$ . Si se denota la inversa de la matriz  $W$  por

$$W_{28 \times 28}^{-1} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,28} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,28} \\ \vdots & \vdots & \ddots & \vdots \\ a_{28,1} & a_{28,2} & \cdots & a_{28,28} \end{bmatrix}$$

y  $g_r$ ,  $r = 1, \dots, 28$ , tal que  $g_1 = g_{1,1}$ ,  $g_2 = g_{2,1}$ ,  $g_3 = g_{2,2}$ ,  $g_4 = g_{2,3}$ ,  $g_5 = g_{3,1}$ ,  $g_6 = g_{3,2}$ ,  $g_7 = g_{3,3}$ ,  $\dots$ ,  $g_{26} = g_{10,1}$ ,  $g_{27} = g_{10,2}$  y  $g_{28} = g_{10,3}$ , entonces

$$g_r = a_{r,1}(-g_{1,0} + g_{2,0}) + a_{r,2}(-g_{2,0} + g_{3,0}) + \dots + a_{r,9}(-g_{9,0} + g_{10,0}) + a_{r,10}(-g_{10,0} + g_{1,0})$$

para  $r = 1, \dots, 28$ . Es decir,

$$g_r = \mathbf{a}_{r,1}g_{1,0} + \mathbf{a}_{r,2}g_{2,0} + \dots + \mathbf{a}_{r,9}g_{9,0} + \mathbf{a}_{r,10}g_{10,0},$$

donde  $\mathbf{a}_{r,1} = -a_{r,1} + a_{r,10}$ ,  $\mathbf{a}_{r,2} = a_{r,1} - a_{r,2}$ ,  $\dots$ ,  $\mathbf{a}_{r,9} = a_{r,8} - a_{r,9}$  y  $\mathbf{a}_{r,10} = a_{r,9} - a_{r,10}$ , para  $r = 1, \dots, 28$ . Ahora, la *spline* puede especificarse en función de  $g_{i,0}$ ,  $i = 1, \dots, 10$ , de la siguiente forma. Si  $w_0 \leq w \leq w_1$ :

$$g(w) = g_{1,0} + (\mathbf{a}_{1,1}g_{1,0} + \mathbf{a}_{1,2}g_{2,0} + \dots + \mathbf{a}_{1,9}g_{9,0} + \mathbf{a}_{1,10}g_{10,0})w$$

Si  $w_{i-1} \leq w \leq w_i$ :

$$\begin{aligned} g(w) = & g_{i,0} \\ & + (\mathbf{a}_{3i-4,1}g_{1,0} + \mathbf{a}_{3i-4,2}g_{2,0} + \dots + \mathbf{a}_{3i-4,9}g_{9,0} + \mathbf{a}_{3i-4,10}g_{10,0})w \\ & + (\mathbf{a}_{3i-3,1}g_{1,0} + \mathbf{a}_{3i-3,2}g_{2,0} + \dots + \mathbf{a}_{3i-3,9}g_{9,0} + \mathbf{a}_{3i-3,10}g_{10,0})w^2 \\ & + (\mathbf{a}_{3i-2,1}g_{1,0} + \mathbf{a}_{3i-2,2}g_{2,0} + \dots + \mathbf{a}_{3i-2,9}g_{9,0} + \mathbf{a}_{3i-2,10}g_{10,0})w^3 \end{aligned}$$

para  $i = 2, \dots, 10$ . Finalmente, la *spline* puede especificarse como:

$$g(w) = \left[ g_{1,0} + (a_{1,1}g_{1,0} + a_{1,2}g_{2,0} + \dots + a_{1,9}g_{9,0} + a_{1,10}g_{10,0})w \right] D_{1,w} +$$

$$\left[ \begin{aligned} &g_{2,0} + (a_{2,1}g_{1,0} + a_{2,2}g_{2,0} + \dots + a_{2,9}g_{9,0} + a_{2,10}g_{10,0})w \\ &+ (a_{3,1}g_{1,0} + a_{3,2}g_{2,0} + \dots + a_{3,9}g_{9,0} + a_{3,10}g_{10,0})w^2 \\ &+ (a_{4,1}g_{1,0} + a_{4,2}g_{2,0} + \dots + a_{4,9}g_{9,0} + a_{4,10}g_{10,0})w^3 \end{aligned} \right] D_{2,w} + \dots +$$

$$\left[ \begin{aligned} &g_{10,0} + (a_{26,1}g_{1,0} + a_{26,2}g_{2,0} + \dots + a_{26,9}g_{9,0} + a_{26,10}g_{10,0})w \\ &+ (a_{27,1}g_{1,0} + a_{27,2}g_{2,0} + \dots + a_{27,9}g_{9,0} + a_{27,10}g_{10,0})w^2 \\ &+ (a_{28,1}g_{1,0} + a_{28,2}g_{2,0} + \dots + a_{28,9}g_{9,0} + a_{28,10}g_{10,0})w^3 \end{aligned} \right] D_{10,w}$$

donde  $D_{i,w} = \begin{cases} 1, & \text{si } w_{i-1} \leq w < w_i \\ 0, & \text{resto} \end{cases}$ ,  $i = 1, 2, \dots, 9$ , y  $D_{10,w} = \begin{cases} 1, & \text{si } w_9 \leq w \leq w_{10} \\ 0, & \text{resto} \end{cases}$ . Y

reordenando términos, se tiene que

$$g(w) = g_{1,0}X_{1,w} + g_{2,0}X_{2,w} + \dots + g_{9,0}X_{9,w} + g_{10,0}X_{10,w}$$

donde

$$X_{1,w} = (1 + a_{1,1}w)D_{1,w} + (a_{2,1}w + a_{3,1}w^2 + a_{4,1}w^3)D_{2,w}$$

$$+ \dots + (a_{26,1}w + a_{27,1}w^2 + a_{28,1}w^3)D_{10,w}$$

y

$$X_{i,w} = D_{i,w} + (a_{1,i}w)D_{1,w} + (a_{2,i}w + a_{3,i}w^2 + a_{4,i}w^3)D_{2,w}$$

$$+ \dots + (a_{26,i}w + a_{27,i}w^2 + a_{28,i}w^3)D_{10,w}$$

para  $i = 2, \dots, 10$ .

Dado que los puntos de ruptura,  $w_i$ ,  $i = 1, \dots, 9$ , que delimitan los tramos de la función *spline*, no se conocen *a priori*, se ha optado por elegir como combinación óptima de localizaciones aquella que minimice la suma de cuadrados de residuos en la regresión

$$\hat{g}_{1,w}^{anual} = g(w) + \mathbf{e}_w = g_{1,0}X_{1,w} + g_{2,0}X_{2,w} + \dots + g_{9,0}X_{9,w} + g_{10,0}X_{10,w} + \mathbf{e}_w,$$

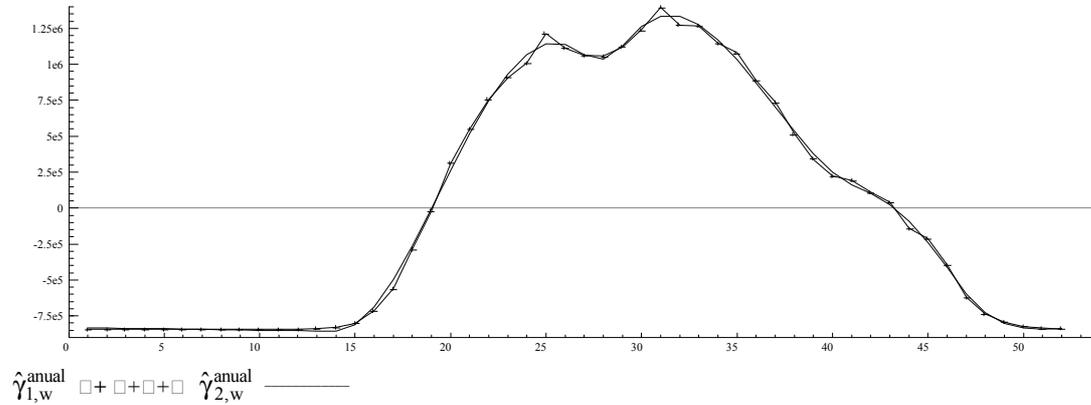
para  $w = 1, \dots, 52$ . Siguiendo este criterio, la combinación óptima de puntos de ruptura para la forma funcional elegida es  $w_1 = 14$ ,  $w_2 = 21$ ,  $w_3 = 26$ ,  $w_4 = 28$ ,  $w_5 = 30$ ,  $w_6 = 41$ ,  $w_7 = 42$ ,  $w_8 = 46$  y  $w_9 = 47$ <sup>37</sup>. La estimación mínimo cuadrática de la regresión anterior para el conjunto de localizaciones seleccionadas permite obtener una

<sup>37</sup> Recuérdese que  $w_0 = 1$  y  $w_{10} = 52$ .

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

primera estimación del componente estacional<sup>38</sup>, que se denotará por  $\hat{g}_{2,w}^{anual}$  (gráfico 3.3.1).

**Gráfico 3.3.1. Componente estacional,  $\hat{g}_{1,w}^{anual}$  y  $\hat{g}_{2,w}^{anual}$**



Ahora bien, resulta más apropiado que la variación estacional se estime conjuntamente con el resto de componentes de la serie. Con este objeto, y también para las localizaciones anteriores de los puntos de ruptura, se han generado las variables  $X_{1,t}$ , ...,  $X_{9,t}$  y  $X_{10,t}$ , que son vectores columna de dimensión  $780 \times 1$ <sup>39</sup>, definidos de forma que  $X_{i,t} = X_{i,w}$ ,  $i = 1, \dots, 10$ , si la observación registrada en el instante  $t$  pertenece a la semana  $w$  de la campaña. Dichas variables, que constituyen la especificación de la función *spline* para el conjunto de la muestra, pueden incorporarse como variables exógenas en el modelo estructural del apartado anterior sustituyendo a la especificación del componente estacional en términos de variables cualitativas. Es decir,

$$y_t = \mathbf{m}_t + g_{1,0}X_{1,t} + g_{2,0}X_{2,t} + \dots + g_{9,0}X_{9,t} + g_{10,0}X_{10,t} + \mathbf{e}_t, \quad \mathbf{e}_t \sim NID(0, \sigma_{\mathbf{e}}^2).$$

Sin embargo, este modelo no puede ser estimado ya que  $\sum_{i=1}^{10} X_{i,t} = 1, \forall t$ . Dado

que  $X_{1,t} = 1 - \sum_{i=2}^{10} X_{i,t}$ , entonces

<sup>38</sup> Las estimaciones de los parámetros del vector  $G^*$  para esta primera estimación fueron las siguientes:  $\hat{g}_{1,0} = -830314.8518$ ,  $\hat{g}_{2,0} = 17306267.85$ ,  $\hat{g}_{3,0} = 6906319.656$ ,  $\hat{g}_{4,0} = -392853645.1$ ,  $\hat{g}_{5,0} = 512420470$ ,  $\hat{g}_{6,0} = -93455711.03$ ,  $\hat{g}_{7,0} = 1068162244$ ,  $\hat{g}_{8,0} = -77687055.58$ ,  $\hat{g}_{9,0} = -1546981447$  y  $\hat{g}_{10,0} = 335358326.8$ .

<sup>39</sup> Nótese que existen 15 campañas y, a efectos de construcción de la serie, se ha considerado que a cada una de ellas corresponden 52 semanas.

## La exportación semanal de tomate canario

$$\begin{aligned}
 & g_{1,0}X_{1,t} + g_{2,0}X_{2,t} + \dots + g_{10,0}X_{10,t} \\
 &= g_{1,0} \left( 1 - \sum_{i=2}^{10} X_{i,t} \right) + g_{2,0}X_{2,t} + \dots + g_{10,0}X_{10,t} \\
 &= g_{1,0} + (g_{2,0} - g_{1,0})X_{2,t} + (g_{3,0} - g_{1,0})X_{3,t} + \dots + (g_{10,0} - g_{1,0})X_{10,t}
 \end{aligned}$$

y puede especificarse el modelo estimable siguiente, en el que se ha eliminado la variable  $X_{1,t}$ :

$$y_t = \mathbf{m}_t + g_{2,0}^* X_{2,t} + \dots + g_{9,0}^* X_{9,t} + g_{10,0}^* X_{10,t} + \mathbf{e}_t, \quad \mathbf{e}_t \sim NID(0, \sigma_e^2),$$

con

$$\mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{h}_t, \quad \mathbf{h}_t \sim NID(0, \sigma_\eta^2).$$

Los resultados de estimación de este modelo aparecen reflejados en los cuadros 3.3.1 a 3.3.4<sup>40</sup>.

**Cuadro 3.3.1. Varianzas de los términos de perturbación**

	Irregular ( $\hat{S}_e^2$ )	Nivel ( $\hat{S}_h^2$ )
<b>Varianza</b>	5.6179x10 <sup>9</sup>	3.6388x10 <sup>10</sup>
<b>q-ratio</b>	(0.1544)	(1.0000)

**Cuadro 3.3.2. Vector de estado final**

Componente	Estimación
$\hat{\mathbf{m}}_T$	831762.5

**Cuadro 3.3.3. Parámetros libres en la especificación de la *spline***

Parámetro	Estimación	Estadístico <i>t</i>	Parámetro	Estimación	Estadístico <i>t</i>
$\hat{g}_{2,0}^*$	1.9976x10 <sup>7</sup>	6.4445	$\hat{g}_{7,0}^*$	7.5149x10 <sup>8</sup>	1.4589
$\hat{g}_{3,0}^*$	-9.4822x10 <sup>5</sup> (*)	-0.0748	$\hat{g}_{8,0}^*$	4.0442x10 <sup>7</sup>	0.1648
$\hat{g}_{4,0}^*$	-3.3418x10 <sup>8</sup>	-4.0385	$\hat{g}_{9,0}^*$	-2.0843x10 <sup>9</sup>	-2.0723
$\hat{g}_{5,0}^*$	4.4028x10 <sup>8</sup>	4.7591	$\hat{g}_{10,0}^*$	4.1649x10 <sup>8</sup>	2.4948
$\hat{g}_{6,0}^*$	-7.7248x10 <sup>7</sup>	-4.8456			

(\*) Parámetros no significativos al 95 % de confianza.

<sup>40</sup> Nótese que  $g_{i,0}^* = g_{i,0} - g_{1,0}$ ,  $i = 2, \dots, 10$ . Como en el epígrafe anterior, se probó con una estimación del componente tendencial que incluía pendiente, resultando también no estocástica y no significativa en el vector de estado final.

Cuadro 3.3.4. Componente estacional  $\hat{g}_{3,w}^{anual}$

Estimación		Estimación		Estimación		Estimación	
$\hat{g}_{3,1}^{anual}$	-824679.547	$\hat{g}_{3,14}^{anual}$	-875304.531	$\hat{g}_{3,27}^{anual}$	1053628.500	$\hat{g}_{3,40}^{anual}$	303728.989
$\hat{g}_{3,2}^{anual}$	-828573.777	$\hat{g}_{3,15}^{anual}$	-829062.002	$\hat{g}_{3,28}^{anual}$	1028069.430	$\hat{g}_{3,41}^{anual}$	206388.833
$\hat{g}_{3,3}^{anual}$	-832468.006	$\hat{g}_{3,16}^{anual}$	-696353.709	$\hat{g}_{3,29}^{anual}$	1106848.350	$\hat{g}_{3,42}^{anual}$	135703.899
$\hat{g}_{3,4}^{anual}$	-836362.236	$\hat{\gamma}_{3,17}^{anual}$	-497891.278	$\hat{g}_{3,30}^{anual}$	1220167.300	$\hat{g}_{3,43}^{anual}$	51078.358
$\hat{g}_{3,5}^{anual}$	-840256.465	$\hat{g}_{3,18}^{anual}$	-254386.339	$\hat{g}_{3,31}^{anual}$	1282116.480	$\hat{g}_{3,44}^{anual}$	-61718.909
$\hat{g}_{3,6}^{anual}$	-844150.695	$\hat{g}_{3,19}^{anual}$	13449.481	$\hat{g}_{3,32}^{anual}$	1283457.020	$\hat{g}_{3,45}^{anual}$	-207321.748
$\hat{g}_{3,7}^{anual}$	-848044.924	$\hat{g}_{3,20}^{anual}$	284904.555	$\hat{\gamma}_{3,33}^{anual}$	1234117.730	$\hat{g}_{3,46}^{anual}$	-390364.006
$\hat{g}_{3,8}^{anual}$	-851939.154	$\hat{\gamma}_{3,21}^{anual}$	539267.254	$\hat{g}_{3,34}^{anual}$	1144027.440	$\hat{g}_{3,47}^{anual}$	-593650.587
$\hat{g}_{3,9}^{anual}$	-855833.383	$\hat{g}_{3,22}^{anual}$	758085.287	$\hat{g}_{3,35}^{anual}$	1023114.980	$\hat{g}_{3,48}^{anual}$	-736757.660
$\hat{g}_{3,10}^{anual}$	-859727.613	$\hat{g}_{3,23}^{anual}$	931943.715	$\hat{g}_{3,36}^{anual}$	881309.176	$\hat{g}_{3,49}^{anual}$	-813780.609
$\hat{g}_{3,11}^{anual}$	-863621.842	$\hat{g}_{3,24}^{anual}$	1053686.930	$\hat{g}_{3,37}^{anual}$	728538.846	$\hat{\gamma}_{3,50}^{anual}$	-842901.852
$\hat{g}_{3,12}^{anual}$	-867516.072	$\hat{g}_{3,25}^{anual}$	1116159.330	$\hat{g}_{3,38}^{anual}$	574732.821	$\hat{g}_{3,51}^{anual}$	-842303.810
$\hat{g}_{3,13}^{anual}$	-871410.301	$\hat{g}_{3,26}^{anual}$	1112205.310	$\hat{g}_{3,39}^{anual}$	429819.927	$\hat{g}_{3,52}^{anual}$	-830168.901

Sin embargo, y con objeto de contemplar la necesidad o no de variables de intervención, el modelo estimable se reespecificó en los términos siguientes:

$$y_t = \mathbf{m}_t + g_{2,0}^* X_{2,t} + \dots + g_{9,0}^* X_{9,t} + g_{10,0}^* X_{10,t} + \sum_{j-k} \mathbf{I}_{j-k} I_{j-k} + \mathbf{e}_t, \mathbf{e}_t \sim NID(0, \sigma_{\epsilon}^2),$$

siendo  $I_{j-k}$  una variable de intervención que toma valor uno si la observación pertenece a la semana  $j$  del año  $k$ . Después de estimar distintos modelos que diferían entre sí en función de las intervenciones consideradas, la especificación finalmente elegida es aquella cuya estimación aparece en los cuadros 3.3.5 a 3.3.9 y los gráficos 3.3.2 a 3.3.4. La estimación del componente estacional, que se denotará por  $\hat{g}_{3,w}^{anual}$ , se ha obtenido a partir de los coeficientes estimados para las variables exógenas  $X_{2,t}, \dots, X_{10,t}$  del modelo anteriormente especificado. El término independiente de la *spline*, incorporado en el componente de nivel, se ha calculado de forma que la media de las variaciones estacionales estimadas fuese nula. El componente de nivel representado, así como su estimación en el vector de estado final, también se han corregido oportunamente con el fin de que dicho término independiente no quedase recogido simultáneamente en los componentes tendencial y estacional. A pesar de esta corrección, la estimación del componente de nivel en el último momento del tiempo de la muestra es inferior a la obtenida cuando se emplean variables cualitativas para recoger el componente

estacional (cuadros 3.2.4 y 3.3.6). La razón que explica esta divergencia estriba en el efecto compensador del nivel estocástico ante las diferencias existentes en la estimación del componente estacional en la semana 26 con ambas especificaciones (cuadros 3.2.4 y 3.3.9).

**Cuadro 3.3.5. Varianzas de los términos de perturbación**

	Irregular ( $\hat{S}_e^2$ )	Nivel ( $\hat{S}_h^2$ )
<b>Varianza</b>	7.1698x10 <sup>8</sup>	2.8863x10 <sup>10</sup>
<b>q-ratio</b>	(0.0248)	(1.0000)

**Cuadro 3.3.6. Vector de estado final**

Componente	Estimación
$\hat{m}_r$	829430.7070

**Cuadro 3.3.7. Parámetros libres en la especificación de la *spline***

Parámetro	Estimación	Estadístico <i>t</i>	Parámetro	Estimación	Estadístico <i>t</i>
$\hat{g}_{2,0}^*$	1.9041x10 <sup>7</sup>	6.9940	$\hat{g}_{7,0}^*$	1.0216x10 <sup>9</sup>	2.2932
$\hat{g}_{3,0}^*$	3.8680x10 <sup>6</sup> (*)	0.3495	$\hat{g}_{8,0}^*$	-4.4641x10 <sup>7</sup> (*)	-0.2120
$\hat{g}_{4,0}^*$	-3.7011x10 <sup>8</sup>	-5.1579	$\hat{g}_{9,0}^*$	-1.7520x10 <sup>9</sup>	-2.0330
$\hat{g}_{5,0}^*$	4.9342x10 <sup>8</sup>	6.1456	$\hat{g}_{10,0}^*$	5.1503x10 <sup>6</sup> (*)	0.0352
$\hat{g}_{6,0}^*$	-9.0868x10 <sup>7</sup>	-6.4934			

(\*) Parámetros no significativos al 95 % de confianza.

Al observar que algunos de los parámetros que definen la *spline* no resultaron significativos, se llevó a cabo un contraste que permitiera obtener alguna conclusión sobre la significación conjunta de dichos parámetros. A partir de dicho contraste se decidió que las variables  $X_{2,t}, \dots, X_{10,t}$  eran, en su conjunto, relevantes y, por tanto, deberían permanecer como elementos del modelo<sup>41</sup>.

<sup>41</sup> El valor del estadístico en cuestión fue  $F=53.5742 \sim F_{7,746}$ . No obstante, téngase en cuenta que el estadístico *F* de restricciones habitualmente empleado en la estimación mínimo cuadrática de un modelo de regresión está sesgado en el caso de un modelo estructural hacia el no rechazo de la hipótesis nula debido a la naturaleza estocástica de algunos de sus componentes. Es decir, parte de la fluctuación estacional, cuando ésta no se incorpora explícitamente, puede estar siendo captada por el componente de nivel estocástico.

**Cuadro 3.3.8. Parámetros de las variables de intervención**

Parámetro	Estimación	Estadístico <i>t</i>	Parámetro	Estimación	Estadístico <i>t</i>
$\hat{I}_{51-88}$	-6.0941x10 <sup>5</sup>	-4.8869	$\hat{I}_{04-96}$	-8.4759x10 <sup>5</sup>	-4.8855
$\hat{I}_{14-90}$	-5.8312x10 <sup>5</sup>	-4.6800	$\hat{I}_{05-96}$	-7.2529x10 <sup>5</sup>	-4.7914
$\hat{I}_{51-90}$	-1.0290x10 <sup>6</sup>	-8.2515	$\hat{I}_{13-96}$	5.5264x10 <sup>5</sup>	3.6473
$\hat{I}_{01-91}$	5.2004x10 <sup>5</sup>	4.1730	$\hat{I}_{14-96}$	1.0012x10 <sup>6</sup>	5.7596
$\hat{I}_{52-93}$	5.7570x10 <sup>5</sup>	4.6154	$\hat{I}_{15-96}$	9.6808x10 <sup>5</sup>	6.3913
$\hat{I}_{18-94}$	5.4984x10 <sup>5</sup>	3.8417	$\hat{I}_{16-96}$	5.5406x10 <sup>5</sup>	4.4326
$\hat{I}_{19-94}$	7.4820x10 <sup>5</sup>	5.2152	$\hat{I}_{05-98}$	-5.9623x10 <sup>5</sup>	-4.7857
$\hat{I}_{03-95}$	-6.5378x10 <sup>5</sup>	-5.2447	$\hat{I}_{45-98}$	4.7398x10 <sup>5</sup>	3.8062
$\hat{I}_{12-95}$	5.6394x10 <sup>5</sup>	3.9488	$\hat{I}_{45-99}$	5.5472x10 <sup>5</sup>	4.4546
$\hat{I}_{13-95}$	5.0495x10 <sup>5</sup>	3.5350	$\hat{I}_{06-00}$	7.2624x10 <sup>5</sup>	5.8308
$\hat{I}_{47-95}$	4.6911x10 <sup>5</sup>	3.7646	$\hat{I}_{09-01}$	-4.0354x10 <sup>5</sup>	-3.2411
$\hat{I}_{03-96}$	-8.3224x10 <sup>5</sup>	-5.4990			

**Cuadro 3.3.9. Componente estacional  $\hat{g}_{3,w}^{anual}$**

Estimación		Estimación		Estimación		Estimación	
$\hat{g}_{3,1}^{anual}$	-824976.873	$\hat{g}_{3,14}^{anual}$	-870442.480	$\hat{g}_{3,27}^{anual}$	1070249.470	$\hat{g}_{3,40}^{anual}$	254980.632
$\hat{g}_{3,2}^{anual}$	-828474.227	$\hat{g}_{3,15}^{anual}$	-825467.955	$\hat{g}_{3,28}^{anual}$	1045133.060	$\hat{g}_{3,41}^{anual}$	164738.236
$\hat{g}_{3,3}^{anual}$	-831971.582	$\hat{g}_{3,16}^{anual}$	-696529.447	$\hat{g}_{3,29}^{anual}$	1135679.140	$\hat{g}_{3,42}^{anual}$	107660.470
$\hat{g}_{3,4}^{anual}$	-835468.936	$\hat{g}_{3,17}^{anual}$	-503096.623	$\hat{g}_{3,30}^{anual}$	1263228.440	$\hat{g}_{3,43}^{anual}$	29280.377
$\hat{g}_{3,5}^{anual}$	-838966.290	$\hat{g}_{3,18}^{anual}$	-264639.149	$\hat{g}_{3,31}^{anual}$	1331424.940	$\hat{g}_{3,44}^{anual}$	-83445.473
$\hat{g}_{3,6}^{anual}$	-842463.645	$\hat{g}_{3,19}^{anual}$	-626.692	$\hat{g}_{3,32}^{anual}$	1330473.670	$\hat{g}_{3,45}^{anual}$	-229169.460
$\hat{g}_{3,7}^{anual}$	-845960.999	$\hat{g}_{3,20}^{anual}$	269471.084	$\hat{g}_{3,33}^{anual}$	1272219.930	$\hat{g}_{3,46}^{anual}$	-406543.963
$\hat{g}_{3,8}^{anual}$	-849458.354	$\hat{g}_{3,21}^{anual}$	526184.510	$\hat{g}_{3,34}^{anual}$	1168509.050	$\hat{g}_{3,47}^{anual}$	-596680.763
$\hat{g}_{3,9}^{anual}$	-852955.708	$\hat{g}_{3,22}^{anual}$	751682.253	$\hat{g}_{3,35}^{anual}$	1031186.320	$\hat{g}_{3,48}^{anual}$	-730993.231
$\hat{g}_{3,10}^{anual}$	-856453.062	$\hat{g}_{3,23}^{anual}$	934686.298	$\hat{g}_{3,36}^{anual}$	872097.040	$\hat{g}_{3,49}^{anual}$	-805210.122
$\hat{g}_{3,11}^{anual}$	-859950.417	$\hat{g}_{3,24}^{anual}$	1065556.960	$\hat{g}_{3,37}^{anual}$	703086.527	$\hat{g}_{3,50}^{anual}$	-835524.182
$\hat{g}_{3,12}^{anual}$	-863447.771	$\hat{g}_{3,25}^{anual}$	1134654.570	$\hat{g}_{3,38}^{anual}$	536000.084	$\hat{g}_{3,51}^{anual}$	-838128.162
$\hat{g}_{3,13}^{anual}$	-866945.126	$\hat{g}_{3,26}^{anual}$	1132339.430	$\hat{g}_{3,39}^{anual}$	382683.017	$\hat{g}_{3,52}^{anual}$	-829214.809

Gráfico 3.3.2. Componente de nivel

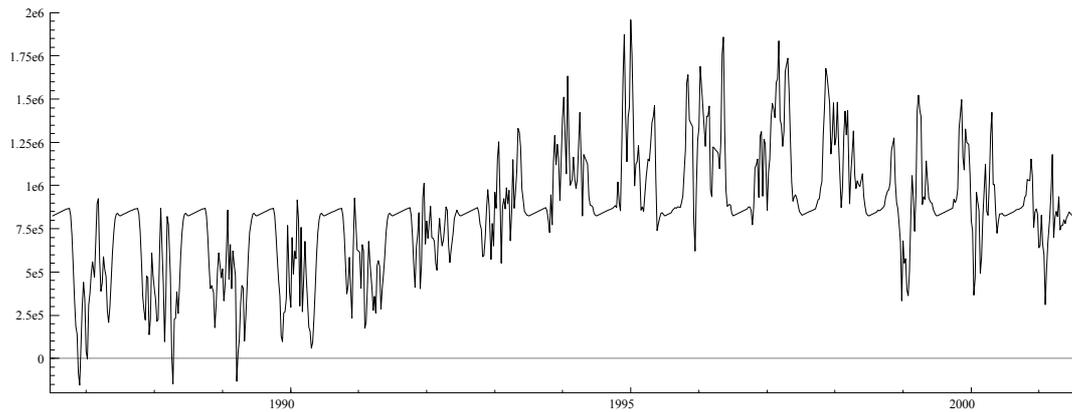
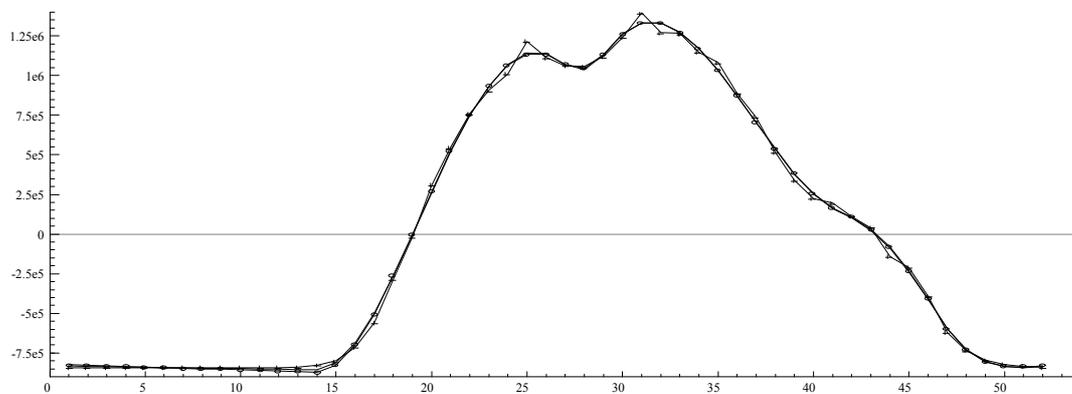
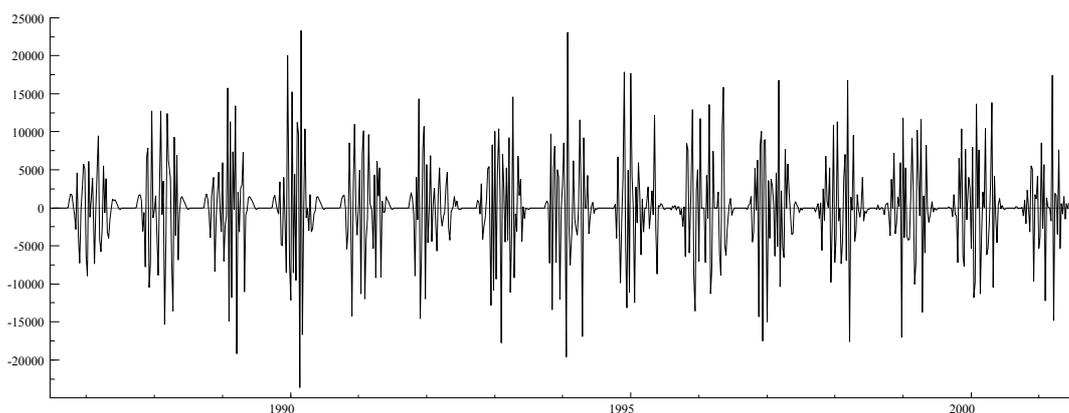


Gráfico 3.3.3. Componente estacional  $\hat{g}_{1,w}^{anual}$ ,  $\hat{g}_{2,w}^{anual}$  y  $\hat{g}_{3,w}^{anual}$



La representación conjunta de las tres estimaciones del componente estacional permite advertir su similitud. Por un lado, la *spline*  $\hat{g}_{2,w}^{anual}$ , ajustada por separado al componente estacional estimado con el modelo estructural que incluye variables cualitativas, es similar al componente estacional  $\hat{g}_{3,w}^{anual}$ , obtenido cuando se recurre a la estimación conjunta de la *spline* y del resto de componentes de la serie. Esta similitud responde, quizás, a la suavidad del patrón estacional en cada campaña y a la fuerte estabilidad de éste en las diferentes campañas. Sin embargo, y como se verá en los capítulos siguientes, las diferencias son mayores cuanto más complejo sea el comportamiento estacional. Pero, sobre todo, merece la pena destacar que la *spline* especificada, estimada de una u otra manera, recoge las variaciones estacionales de la serie de forma bastante similar a la del modelo con variables cualitativas estacionales (cuadros 3.2.4 y 3.3.9 y gráfico 3.3.3) y, por tanto, puede concluirse que las dos especificaciones son equivalentes.

### Gráfico 3.3.4. Componente irregular



Respecto a las intervenciones que se han considerado bajo las dos formulaciones del componente estacional, es importante mencionar que éstas coinciden en número y, aproximadamente, en magnitud de la estimación. En cualquier caso, es lógico que aparezcan residuos de superior magnitud cuando el componente estacional se estima a través de la *spline*, que, en realidad, constituye una aproximación funcional suave a las diferentes variaciones estacionales registradas en cada semana de la campaña<sup>42</sup>. De todos modos, la magnitud de los residuos puede considerarse despreciable en relación con la magnitud del componente aproximado<sup>43</sup>.

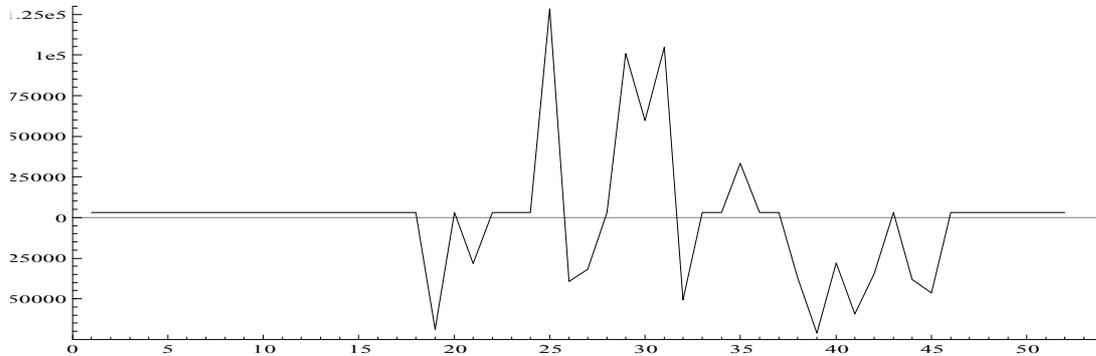
Es interesante destacar que la estimación de la fluctuación estacional es sensible a la introducción de variables de intervención, tanto cuando ésta se recoge con variables cualitativas estacionales como si se utiliza una función *spline*. Para el primero de estos casos, el gráfico 3.3.5, en el que se representan las diferencias entre las estimaciones de dicho componente con y sin intervenciones, refleja que éstas son relevantes en aquellas semanas del año en las que fue necesario introducir variables de intervención y, además, el signo de las diferencias es, lógicamente, contrario al signo de los parámetros correspondientes a las variables de intervención, dado el efecto que estas variables tienen en los niveles medios de las semanas en cuestión. Este comportamiento se reproduce, aunque de forma menos acentuada, cuando el componente estacional se modela a través de una función *spline* (gráfico 3.3.6), que impone un carácter más suave a la evolución de la fluctuación que recoge. Además, el efecto de las intervenciones está

<sup>42</sup> Ese mayor error está siendo absorbido por el componente irregular, cuya varianza ha aumentado.

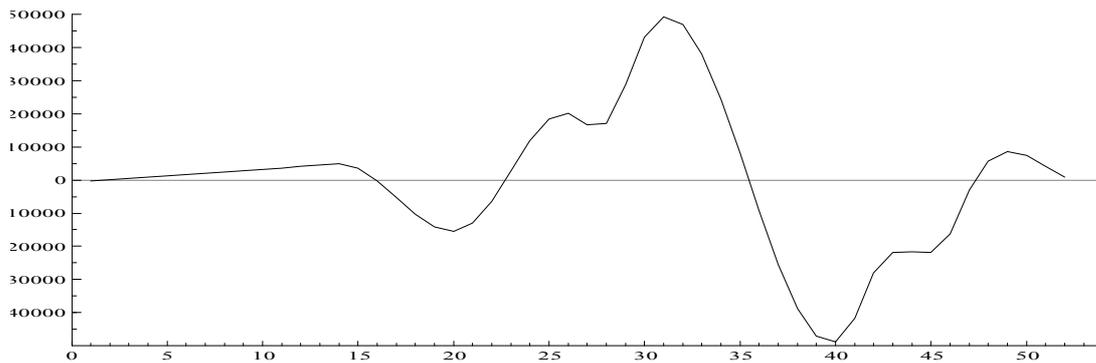
<sup>43</sup> En términos medios, durante las cincuenta y dos semanas de la campaña la diferencia entre el componente estacional estimado con variables cualitativas y el estimado a través de la *spline* representa el 6.6504% de la estimación obtenida en la primera aproximación. Y, a lo largo de toda la muestra, la magnitud del componente irregular en el modelo finalmente estimado significa, en cualquier instante del tiempo, menos del 3% del valor medio de la serie.

condicionado por la proximidad de los nodos a las semanas en las que se localizan las observaciones anómalas.

**Gráfico 3.3.5. Efecto de las intervenciones sobre la estimación del componente  $\hat{g}_{1,w}^{annual}$**



**Gráfico 3.3.6. Efecto de las intervenciones sobre la estimación del componente  $\hat{g}_{3,w}^{annual}$**



### 3.4. Conclusiones

El estudio de la serie de exportaciones de tomate canario mediante modelos estructurales muestra la inestabilidad de lo que, en el enfoque clásico de series temporales, se define como componente tendencial, cuyo comportamiento se ajusta a un modelo de nivel estocástico con raíz unitaria.

Sin embargo, el componente estacional posee un carácter no estocástico que puede recogerse adecuadamente mediante un modelo de variables cualitativas, o bien, a través de funciones *splines* periódicas, que resultan ser formulaciones equivalentes a la especificación anterior pero más parsimoniosas. A pesar de que en este caso no era necesaria la introducción de funciones *splines* para recoger adecuadamente el comportamiento estacional de la serie estudiada, se ha recurrido a ellas con objeto de ilustrar su utilidad desde un punto de vista metodológico. Sin embargo, las *splines*

## **Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia**

adquieren su verdadero potencial en aquellas ocasiones en las que no sea posible estimar el componente estacional mediante especificaciones convencionales con variables cualitativas o términos trigonométricos —por problemas computacionales asociados a la excesiva longitud del periodo estacional y, por tanto, a la elevada dimensión del vector de parámetros estimables—; pero, si ese fuera el caso, la determinación de la *spline* óptima no podría llevarse a cabo a partir del procedimiento planteado anteriormente, sino que sería preciso recurrir a la identificación tentativa de los nodos a partir de la observación de la serie original o, quizás, una vez extraídos algunos de sus componentes. Estas situaciones serán analizadas en los dos capítulos siguientes.

Finalmente, desde una perspectiva económica, el análisis anterior permite concluir que el patrón estacional es estable y las estimaciones obtenidas para cada una de las estaciones confirman la existencia de niveles mínimos a principios y finales de zafra y máximos en las semanas centrales. Por tanto, parece imponerse la conclusión de que, si bien los agentes económicos que protagonizan la actividad productiva y exportadora de tomate canario han reaccionado ante cambios en las reglas de juego del mercado, tal respuesta no se ha traducido en una modificación significativa de la distribución del volumen total exportado en la campaña entre las distintas semanas de la misma.

Por otra parte, la estimación del componente de nivel refleja con bastante claridad el comportamiento tendencial de las exportaciones que ya había sido apreciado en el gráfico original. En concreto, se advierte que tras un periodo de cierto estancamiento del nivel exportado, sucede otro de crecimiento más fuerte y en las últimas campañas se produce una relativa disminución de las exportaciones. Este comportamiento del nivel de la serie permite concluir que, a pesar de los efectos de los cambios en la normativa comercial, la actividad productora de tomate en Canarias no ha reducido notablemente su presencia en los mercados europeos. Ahora bien, obtener conclusiones sobre la rentabilidad de esta actividad exigiría analizar la estacionalidad de la serie de precios y examinar su evidente vinculación con el volumen exportado en cada semana de la campaña.

## **4. Número diario de visitas a un servicio de urgencias del Barcelonès Nord**

La atención dispensada a los pacientes en los Servicios de Urgencias Hospitalarios (SUH) es objeto de preocupación para la comunidad sanitaria y la sociedad en general. Y parece que en todos los países desarrollados, tanto si ofrecen un servicio nacional de salud con aseguramiento obligatorio y gratuito de la población, como si el modelo sanitario se articula como un sistema de mercado —como es el caso de EEUU—, se ha producido un fenómeno de mayor utilización de estos servicios<sup>1</sup>. Por supuesto, el aumento desmesurado de la demanda en los servicios de urgencias tiene consecuencias negativas para el centro en forma de falta de espacio físico, sobrecarga de los servicios de laboratorio y radiología y de los profesionales del hospital, alargamiento de la lista de espera de ingresos programados y pérdidas en productividad por las dificultades para programar el trabajo<sup>2</sup>. Y, sobre todo, se pueden producir demoras en la asistencia a pacientes que sufren situaciones de riesgo vital, habiéndose señalado una relación inversa entre la presión asistencial y la calidad de los servicios de urgencias.

Un buen número de trabajos científicos han examinado las circunstancias que influyen en el crecimiento de la frecuencia con la que se visitan los SUH<sup>3</sup>. Algunos de los factores que se han señalado tienen que ver con la accesibilidad a dichos servicios y, sobre todo, con la morbilidad, así como con determinadas características sociodemográficas. Por ejemplo, el alargamiento de la esperanza de vida y el envejecimiento progresivo de la población —con un crecimiento notable del grupo de mayores de 65 años— se asocia a una mayor prevalencia de enfermedades agudas (accidentes cerebro-vasculares, infartos de miocardio, fracturas osteoporóticas, ...) y, sobre todo, de enfermedades crónicas (obstrucción crónica al flujo aéreo, diabetes, hipertensión, insuficiencia cardíaca y otras) que sufren frecuentes descompensaciones y requieren atención urgente. Otros cambios en los patrones de morbi-mortalidad

---

<sup>1</sup> Véase Richards y otros (2000) y Derlet (2002).

<sup>2</sup> Derlet y Richards (2000), entre otros, analizan algunas de estas cuestiones.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

relacionados con la aparición o incremento notable de algunas patologías (SIDA, toxicomanías, accidentes de tráfico o laborales) contribuyen también a engrosar las listas de pacientes que visitan los centros de urgencias. Además, existen condicionantes que afectan específicamente a los centros localizados en determinadas zonas. Las migraciones laborales producen, en este sentido, un doble efecto, ya que, con frecuencia, los ancianos quedan faltos de apoyo socio familiar en los lugares de origen y, además, los emigrantes suelen traer a sus padres a su nuevo lugar de residencia cuando enferman. De igual modo, debe considerarse el caso de los hospitales de zonas costeras, que registran notables incrementos en la demanda de sus servicios durante la época estival.

En cualquier caso, los factores antes mencionados resultan insuficientes para explicar el crecimiento de la demanda de atención urgente. El aumento de las visitas a centros de urgencias ha sido mucho mayor que el crecimiento de la población anciana y, en general, de los grupos de población que, por sus factores de riesgo, parecen más propensos a efectuar estas visitas. Y las condiciones específicas de determinados centros no pueden, por supuesto, justificar una tendencia generalizada en el conjunto de todos ellos. Quizás es necesario asumir también que la oferta crea su propia demanda. Así, es difícil negar que el volumen, estructura y sistema de financiación de los servicios sanitarios influyen en la utilización de los SUH<sup>4</sup>. A este respecto, se ha comprobado que el incremento de la oferta ha sido acompañado de un incremento en la tasa de atenciones urgentes y que la relación inversa entre la utilización de los servicios sanitarios y la distancia a los mismos a la que aluden numerosos estudios se confirma también en el caso de los servicios de urgencias.

El hecho es que el crecimiento de la demanda produce desajustes en relación con la oferta de servicios que deberían ser eliminados por las evidentes consecuencias sobre los pacientes afectados, cuya salud está, *a priori*, lo suficientemente deteriorada como para requerir los servicios a los que acude. Si se acepta esta premisa, la solución a los desequilibrios debe proceder del lado de la oferta. De hecho, muchas de las intervenciones públicas han tratado de incrementar la dotación de los servicios sanitarios de modo que éstos puedan atender a un mayor número de pacientes<sup>5</sup>. Ahora

---

<sup>3</sup> Véase, por ejemplo, Andrén y Rosenqvist (1985, 1987), Padget y Brodsky (1992), Olson y Hansagi (2001) y Ovens y Chan (2001).

<sup>4</sup> Véase Derlet y Richards (2000).

<sup>5</sup> Un ejemplo de estas actuaciones lo constituyen las inversiones en reformas estructurales realizadas por el INSALUD y las Comunidades Autónomas con transferencias sanitarias en buena parte de los SUH

bien, es preciso considerar que, en última instancia, el factor determinante de la visita a los SUH es la percepción subjetiva de gravedad y ansiedad del paciente. Y este elemento de subjetividad contribuye quizás a explicar el aumento desproporcionado de la frecuencia con la que los SUH se utilizan de forma inadecuada<sup>6</sup>.

En el caso de los usuarios que acuden por propia decisión y sin que su situación revista la gravedad que se presupone en los pacientes para los que se crean estos servicios, el motivo de la visita no responde siempre a la errónea percepción del paciente sobre la gravedad de su estado de salud, que podría estar relacionada con la falta de educación sanitaria<sup>7</sup>. A veces, la visita al SUH, constituye la vía de entrada al sistema sanitario para colectivos sociales desfavorecidos<sup>8</sup> y, en muchas ocasiones, están siendo utilizados para adelantar listas de espera, ingresar pacientes desde consultas externas o realizar determinadas exploraciones y actividades programadas como si de un servicio de atención primaria se tratase.

Evidentemente, el recurso a los SUH para atender situaciones que, desde un punto de vista sanitario, no pueden calificarse como urgentes, implica que se están sustituyendo los servicios de atención primaria<sup>9</sup>, lo que puede significar un incremento de costes asistenciales y, en consecuencia, costes de oportunidad y pérdidas de bienestar para la sociedad en su conjunto. Y, sobre todo, estas visitas generan una presión innecesaria sobre una oferta limitada y reducen la calidad del servicio. De ahí que en los países desarrollados se haya puesto en práctica una serie de medidas que tratan de

---

públicos. Estas reformas no significan sólo la mejora de las instalaciones, sino que, además, permiten la incorporación de algunos cambios organizativos. Sin embargo, los problemas de masificación no se han resuelto. En el informe presentado por el Defensor del Pueblo (1988) ya se denunciaba una larga serie de problemas de estructura física y dotación en los SUH en España, principalmente relacionados con las condiciones de las salas de atención y espera. Otros estudios más recientes revelan la persistencia de estos y otros problemas (Oterino y otros, 1999). Véase también López y otros (1997).

<sup>6</sup> Es preciso aclarar que los criterios que definen la urgencia y, por lo tanto, las situaciones en que el uso de los SUH es adecuado, varían considerablemente y tanto los resultados de los estudios de ámbito internacional como nacional, evidencian, más allá de las importantes diferencias de entorno y metodología, la ausencia de consenso sobre qué constituye una urgencia inapropiada. Seguramente, la propia subjetividad del concepto de urgencia explica las divergencias en las proporciones de urgencias inadecuadas que se han encontrado en numerosos trabajos; además, los estudios que abordan estas cuestiones suelen hacer referencia a un único centro y, en consecuencia, la comparación de los resultados puede carecer de sentido o, como mínimo, debe efectuarse con extrema cautela. Véase Lowe y Bindman (1994), Foldes y otros (1994), O'Brien y otros (1996) y Gill y otros (1996).

<sup>7</sup> La repercusión de la falta de educación sanitaria en el uso inadecuado de los servicios de urgencias puede acentuarse cuando al escaso nivel de conocimientos sanitarios se le une el alarmismo generado por los medios de comunicación.

<sup>8</sup> Una parte importante de la demanda inapropiada de los servicios de urgencias hospitalarios la protagonizan personas sin hogar y sin seguro médico, que presentan condiciones sociales precarias y que acuden al hospital buscando un ingreso para cubrir sus necesidades más primarias o para asegurar la atención necesaria en su enfermedad, que no requiere estrictamente un ingreso hospitalario. Véase Padget y otros (1995).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

desviar las visitas inapropiadas hacia su lugar correcto en el sistema sanitario. Pero, como señala Gill (1999), se trata de un problema persistente con difícil solución. De hecho, el análisis de las actuaciones emprendidas con este fin muestra el escaso éxito de las intervenciones dirigidas a frenar la demanda inapropiada y aconseja, por tanto, que las políticas sobre urgencias hospitalarias se diseñen más bien con objeto de reorganizar los SUH de forma que su oferta se adecue a la demanda real<sup>10</sup>.

Es decir, ante los desajustes oferta-demanda, tal vez no sea adecuado tratar de restringir la demanda y adaptarla a una oferta fija, sino más bien adaptar la oferta a la demanda<sup>11</sup>. El conocimiento general de la demanda de los SUH es imprescindible para planificar las inversiones públicas en estructuras y personal. Pero, además de incrementar las dotaciones de recursos humanos y materiales, la correcta asignación de los recursos disponibles es otro de los elementos clave para mejorar la oferta. Ahora

---

<sup>9</sup> Véase O'Brien y otros (1999).

<sup>10</sup> Los numerosos programas de mejora de la accesibilidad a la atención primaria consistentes en aumentar el número de médicos de atención primaria no han aportado resultados concluyentes (Hilditch, 1980, Straus y otros, 1983, Maynard y Dodge, 1983, Chan y otros, 1985, Fleming y Andersen, 1986, Long y Settle, 1988, Rizos y otros, 1990, Bolívar y otros, 1996, Gill y Diamond, 1996, y Grossman y otros, 1998).

También se ha barajado la posibilidad de establecer centros de urgencias no hospitalarios con horario de veinticuatro horas que no requieran cita previa, o bien, aumentar el número de médicos del SUH. Sin embargo, no se ha contrastado si estas estrategias constituyen una alternativa eficiente o, por el contrario, sólo incrementan los costes del servicio sanitario.

La eficacia del *copago* para reducir la utilización de los SUH parece haber sido probada en determinados estudios (MacKoul y otros, 1995, y O'Grady y otros, 1985), pero sus resultados no pueden generalizarse a otras poblaciones (Selby y otros, 1996), ni existen evaluaciones del impacto a largo plazo sobre el estado de salud de las personas con menos recursos económicos (Shapiro y otros, 1986). En España, la aplicación del *copago* podría generar un importante conflicto social (Peiró y otros, 1999). Otro tipo de intervenciones consisten en la interposición de barreras o filtros que desvíen los casos no urgentes a otro nivel asistencial, pero su efectividad está oscurecida por el riesgo inherente a estas prácticas (Hansagi y otros, 1987, Hansagi y otros, 1989, Derlet y Nishio, 1990, Derlet y otros, 1992, Lowe y Bindman, 1994, Birnbaum y otros, 1994, Foldes y otros, 1994, Derlet y otros, 1995, y O'Brien y otros, 1996).

Gill y otros (2000) advierten que una atención continuada de la salud por parte del paciente está asociada a un bajo porcentaje de uso de los servicios de urgencias, lo que conlleva una reducción de costes sin necesidad de restringir el acceso a tales servicios. En este sentido, las campañas educativas parecen una opción interesante, pero su efectividad ha sido poco evaluada o es escasa (Benz y Shank, 1982, O'Shea y otros, 1984, Chande y otros, 1996, y Kaplan y otros, 1998).

Otras intervenciones han sido estudiadas por Andrén y Rosenqvist (1985, 1987) y Redelmeier y otros (1995).

<sup>11</sup> De hecho, las propias condiciones del hospital constituyen uno de los aspectos que más influye en la efectividad del servicio prestado (Miró y otros, 2000, y Miró y otros, 2001). En el caso concreto de las visitas inadecuadas, una solución es incorporar la realidad actual al sistema de forma que éstas se conviertan en una parte legítima de la carga del servicio (Driscoll y otros, 1987). En este sentido, los costes se reducen si el servicio de atención urgente es desempeñado, en primera instancia y a modo de filtro, por los médicos de atención primaria, mientras que los médicos del hospital aguardan en segunda línea y atienden sólo a los pacientes que han superado la criba de la primera fase (Dale, Green, Raid y Glucksman, 1995, Dale, Green, Raid, Glucksman y Higgs, 1995, Murphy y otros, 1996, y Dale y otros, 1996). Otra posibilidad que muestra un ahorro importante frente a los costes de la atención usual, consiste en crear *salas rápidas* de urgencias para atender a pacientes con problemas no urgentes. En cualquier caso, todas estas cuestiones requieren una reflexión profunda antes de adoptar decisiones.

bien, adoptar medidas de esta índole exige previamente conocer la demanda real de los SUH en cada momento del tiempo y, teniendo en cuenta la variedad de factores que pueden incidir en ella, cabe pensar en un comportamiento diferente según la época del año, el día de la semana o la hora del día de que se trate. De hecho, la incidencia de los desajustes entre oferta y demanda es mayor en determinadas fechas y parece evidente que estos desequilibrios pueden tener graves consecuencias, sobre todo cuando, aún siendo transitorios, se traducen en excesos de demanda que, dadas las características del servicio, pueden ocasionar daños irreparables en la salud del paciente. En suma, la estacionalidad de la demanda se convierte, pues, en el elemento clave en torno al cual debe girar la reorganización de los servicios con el fin último de aumentar la calidad del servicio prestado.

Dada la naturaleza de los datos a los que se ha tenido acceso, en este capítulo se intenta caracterizar la demanda real de un servicio específico de urgencias medida en términos del número de visitas registradas, al margen de que los pacientes estén o no haciendo un uso adecuado de estos servicios. Con este objeto, el texto se estructura como se indica a continuación. En el epígrafe siguiente se identifica el centro concreto que va a ser analizado y se señalan algunas características de los datos utilizados, haciendo especial hincapié en la descripción de las posibles fluctuaciones estacionales que aparentemente existen en la serie. En el epígrafe segundo se especifica la forma funcional de la *spline* utilizada para recoger la fluctuación estacional de periodo anual, así como el número y localización de los puntos de ruptura que delimitan los tramos que la configuran. Y se obtiene una estimación preliminar de dicha función, mientras que la estimación definitiva se obtiene conjuntamente con el resto de componentes en el modelo estructural especificado en el epígrafe tercero. El capítulo finaliza señalando las ventajas computacionales que supone el empleo de las funciones *splines* en el análisis del componente estacional y describiendo el patrón estacional de la demanda diaria registrada en el servicio de urgencias examinado, a partir del cual se sugieren nuevas líneas de estudio que podrían contribuir a mejorar la eficiencia del servicio.

### **4.1. Estacionalidad de las visitas diarias registradas**

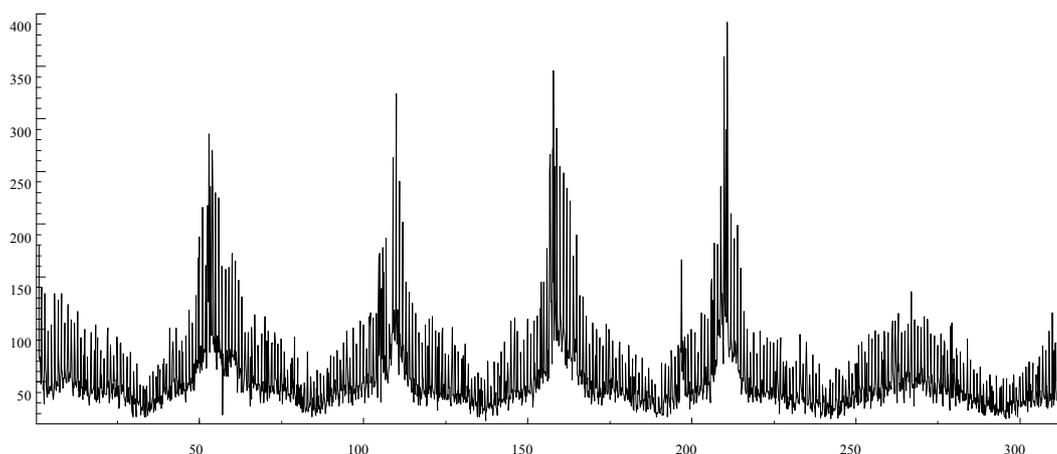
La demanda de los SUH se materializa en centros que pertenecen a distintas realidades socioeconómicas y, por tanto, en la demanda registrada en tales centros

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

inciden diferentes factores sociodemográficos<sup>12</sup>. De ahí, la necesidad de analizar de forma minuciosa el comportamiento observado en un centro concreto, de modo que se puedan formular medidas adaptadas a las particularidades de dicho centro y que puedan servir de indicadores de planificación de otros centros con características similares. El estudio realizado en este capítulo se refiere al servicio de urgencias del Barcelonès Nord, que cubre la atención de una población de 350000 habitantes.

Como indicador de la demanda real se ha utilizado el número de visitas registradas. Ahora bien, el examen de la estacionalidad de las visitas con el objetivo que se ha señalado requiere, desde un punto de vista metodológico, adoptar decisiones en cuanto al tipo de datos que va a ser objeto de estudio. La frecuencia con la que se registran las visitas puede ayudar a vislumbrar patrones estacionales o, por el contrario, puede dificultar su apreciación si el grado de agregación temporal es excesivo. Por esta razón, se ha utilizado una base diaria, dada la ausencia de datos registrados con mayor frecuencia de observación. Concretamente, se analiza el número diario de visitas registradas en el servicio de urgencias ya indicado en el periodo comprendido entre el lunes uno de enero de 1996 y el lunes treinta y uno de diciembre de 2001<sup>13</sup>. Las 2192 observaciones diarias a las que hace referencia la serie<sup>14</sup>, denotada por  $\{y_t\}_{t=1,\dots,2192}$ , aparecen representadas en el gráfico 4.1.1.

**Gráfico 4.1.1. Número diario de visitas registradas (1/01/96 al 31/12/2001)**



<sup>12</sup> Véase Quintana y otros (1999).

<sup>13</sup> Los años 1996 y 2000 fueron bisiestos, es decir, se registraron 366 observaciones frente a las 365 del resto de los años.

<sup>14</sup> Véase el fichero diario.xls del disco.

## Número diario de visitas a un servicio de urgencias del Barcelonès Nord

La observación del gráfico anterior permite apreciar un comportamiento a largo plazo que fluctúa alrededor de un nivel medio constante. Quizás, el lapso de tiempo considerado no es lo suficientemente amplio como para registrar el efecto de un incremento poblacional que, sin duda, se dejaría sentir en el número de visitas a un centro determinado de urgencias. Este efecto dependería, por supuesto, de si en ese mismo periodo ha habido cambios paralelos en la infraestructura sanitaria de la zona de modo que el número de individuos por centro no haya variado sustancialmente. En cualquier caso, asumir que el componente tendencial es constante es un supuesto demasiado rígido y parece más acorde con lo observado admitir como hipótesis de partida la naturaleza estocástica de éste y quizás otros componentes. Esta hipótesis puede introducirse adecuadamente en un modelo estructural, que parece una formulación apropiada para captar la evolución de la serie.

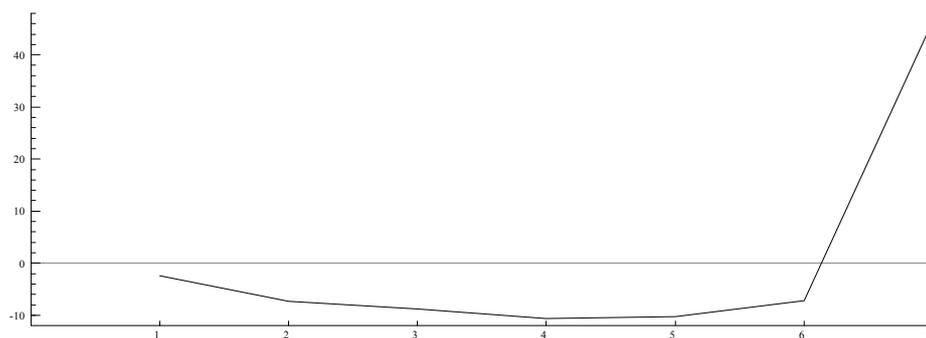
Por otro lado, la alta frecuencia con la que se observan los datos permite analizar la presencia de periodicidades de carácter semanal que se superponen a las observadas con periodo anual y que, sin embargo, no se aprecian con claridad en la serie original. Con objeto de aislar estos comportamientos estacionales se ha procedido a calcular y representar las medias por día de la semana y por día del año corregidas de modo que la suma de los efectos estacionales sea nula (gráficos 4.1.2 y 4.1.3)<sup>15</sup>. El hecho de que existan dos años bisiestos constituye una fuente de heterogeneidad que debe tratarse antes de calcular las medias por día del año. Para solucionar este inconveniente se ha optado por excluir el veintinueve de febrero de 1996 y de 2000, consiguiéndose de esta manera que en el cálculo de cada media por día del año intervengan observaciones referidas al mismo día de cada mes en todos los años<sup>16</sup>.

---

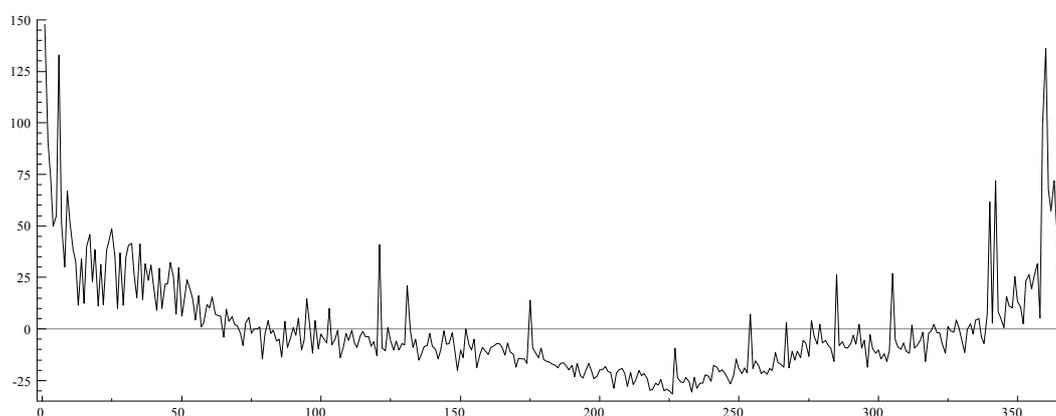
<sup>15</sup> Téngase en cuenta que estas medias sólo tienen por objeto detectar la presencia de las fluctuaciones periódicas correspondientes, sin que quepa formular juicio alguno sobre su carácter, estocástico o no.

<sup>16</sup> Otra posibilidad consiste en considerar que el veintinueve de febrero de los años 1996 y 2000 es el sexagésimo día del año correspondiente y será comparado con el sexagésimo día del resto de los años. Entonces, los días treinta y uno de diciembre de 1996 y de 2000 quedarían fuera del cálculo de los promedios. En este otro caso, las medias calculadas para cada uno de los 365 días del año eran similares. Sin embargo, el método utilizado parece más adecuado dado el carácter especial del día de fin de año en términos del número de visitas a un centro de urgencias. Otras soluciones al problema de heterogeneidad consisten en introducir valores perdidos de modo que el periodo estacional sea de la misma longitud. Esta última opción es la que eligen Cabrero y otros (2002), que, además, y sobre una base diaria, redefinen los años, meses y semanas con objeto de obtener periodos estacionales homogéneos en cuanto a duración.

### Gráfico 4.1.2. Medias por día de la semana



### Gráfico 4.1.3. Medias por día del año



Si se acepta que las medias representadas reflejan cada una de las dos variaciones estacionales a las que se ha aludido, es decir, se asume el carácter no estocástico del componente estacional, puede concluirse que las periodicidades de carácter semanal se caracterizan por mínimos registrados en torno a mediados de semana y máximos los domingos, mientras que la fluctuación de periodo anual presenta máximos a finales y principios de año y mínimos en el verano. Sin embargo, la representación de la serie original (gráfico 4.1.1) permite advertir que los máximos y mínimos no tienen lugar siempre el mismo día de la semana ni se observan siempre exactamente en la misma época del año. Por ello, y al menos en relación con la fluctuación de periodo semanal<sup>17</sup>, para la cual se dispone de un número de observaciones suficiente para captar su naturaleza evolutiva o estable, parece razonable asumir como supuesto inicial que se trata de un componente estocástico. La metodología propia de los modelos estructurales de series temporales permitirá decidir *a posteriori* sobre la naturaleza de este componente. Con respecto a la fluctuación de

<sup>17</sup> Véase anexo 4.1.

periodo anual, el reducido número de observaciones por día del año impide apreciar su inestabilidad<sup>18</sup> y conduce a imponer el carácter no estocástico de dicha fluctuación, que será recogida a través de una función *spline*. Esta especificación proporciona una formulación más parsimoniosa de tales efectos estacionales que si se recurre, por ejemplo, a 365 variables cualitativas estacionales.

Es decir, el modelo estructural para la serie analizada puede especificarse del modo siguiente:

$$y_t = \mathbf{m}_t + \mathbf{g}_t^{semanal} + \mathbf{g}_t^{anual} + \mathbf{e}_t, \quad \mathbf{e}_t \sim NID(0, \sigma_\varepsilon^2),$$

donde  $\mathbf{m}_t$  representa un componente de nivel estocástico,  $\mathbf{g}_t^{semanal}$  es la fluctuación de periodo semanal según la formulación trigonométrica estocástica y  $\mathbf{g}_t^{anual}$  es la variación de periodo anual, que se ha modelado mediante funciones *splines*<sup>19</sup>.

## 4.2. Especificación de la función *spline* para la fluctuación anual

La estimación del modelo anterior exige, previamente, identificar la forma funcional de la *spline* en cada tramo, así como el número y localización de los puntos de ruptura que delimitan los tramos. Para ello, conviene aislar la fluctuación estacional de periodo anual del resto de componentes de la serie. Evidentemente, si este componente pudiera ser estimado con una formulación convencional —por ejemplo, estocástica trigonométrica— junto al resto de componentes del modelo, el problema quedaría resuelto; pero en este caso, las funciones *splines* resultan necesarias, precisamente, porque, dado el elevado número de estaciones, las formulaciones convencionales no pueden ser empleadas o requieren utilizar un número excesivo de parámetros<sup>20</sup>.

En esta situación, una primera aproximación de la fluctuación estacional de periodo anual puede obtenerse a partir de las medias simples por día del año. Sin embargo, aunque la representación gráfica de estas medias (gráfico 4.1.3) ilustra el patrón básico de la variación estacional en cuestión, parece más correcto obtenerlo a partir del residuo de un modelo en el que se incorpore el resto de componentes de la serie. Ahora bien, el grado en el que pueda aislarse de este modo el componente  $\mathbf{g}_t^{anual}$

<sup>18</sup> Véase anexo 4.2.

<sup>19</sup> También se consideró la transformación logarítmica de los datos pero los resultados obtenidos no difieren significativamente de los finalmente presentados (anexo 4.3).

<sup>20</sup> En los programas informáticos disponibles no se contempla la posibilidad de estimar un modelo con dos fluctuaciones simultáneas de carácter estocástico. En último término, junto a la fluctuación semanal estocástica, podría introducirse la fluctuación anual mediante variables *dummies*, pero, en este caso, la limitación viene impuesta porque el número necesario de éstas es superior al admitido en los programas.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

depende de los supuestos sobre la naturaleza del resto de componentes del modelo. Dada la sensibilidad del procedimiento a estos supuestos, se optó por construir  $\mathbf{g}_t^{anual}$  a partir de las medias por día del año de los residuos del modelo

$$y_t = \mathbf{m}_t + \mathbf{g}_t^{semanal} + \mathbf{e}_t, \mathbf{e}_t \sim NID(0, \sigma_\epsilon^2),$$

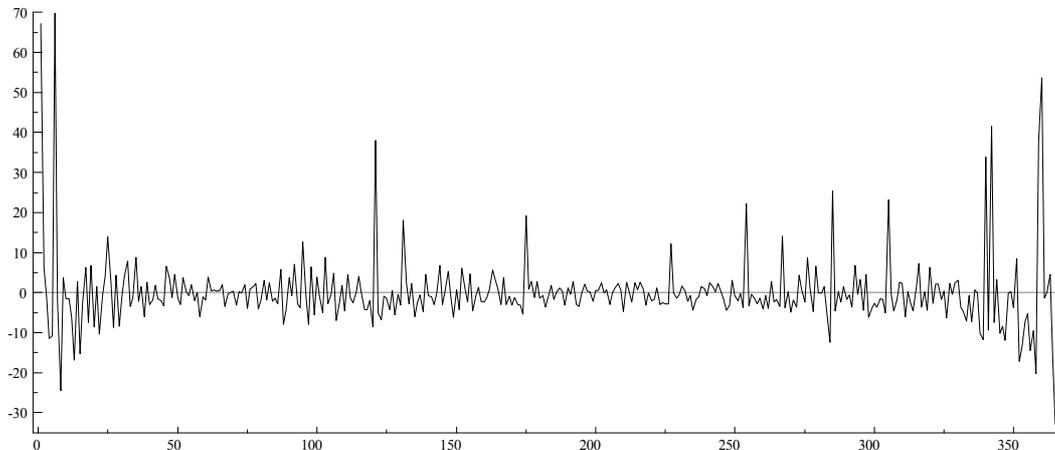
donde

$$\dot{\lambda}_t = \dot{\lambda}_{t-1} + \zeta_t, \mathbf{h}_t \sim NID(0, \sigma_\eta^2),$$

en tres situaciones diferentes<sup>21</sup>:

- suponiendo que tanto  $\mathbf{m}_t$  como  $\mathbf{g}_t^{semanal}$  poseen carácter estocástico;
- asumiendo que  $\mathbf{m}_t$  es estocástico y  $\mathbf{g}_t^{semanal}$  es determinístico;
- considerando que  $\mathbf{m}_t$  y  $\mathbf{g}_t^{semanal}$  son no estocásticos<sup>22</sup>.

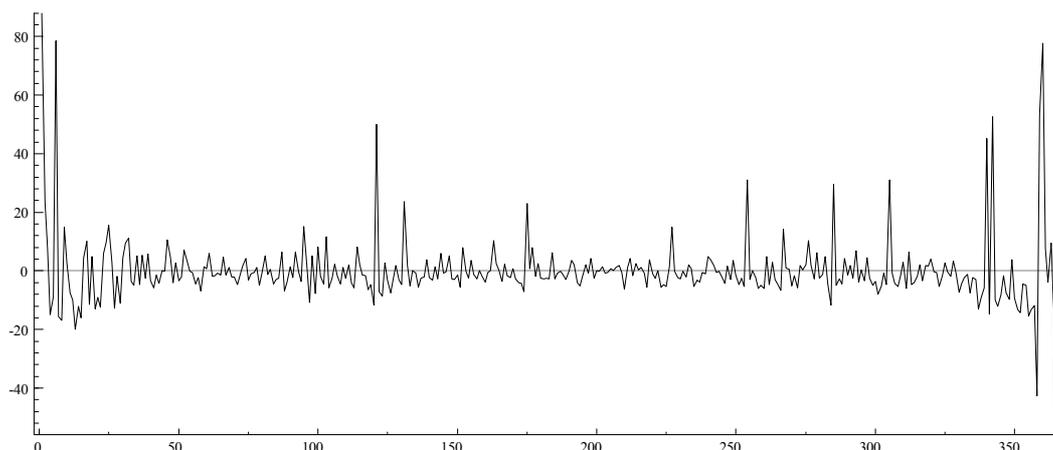
**Gráfico 4.2.1. Variación estacional de periodo anual (medias por día del año, opción a)**



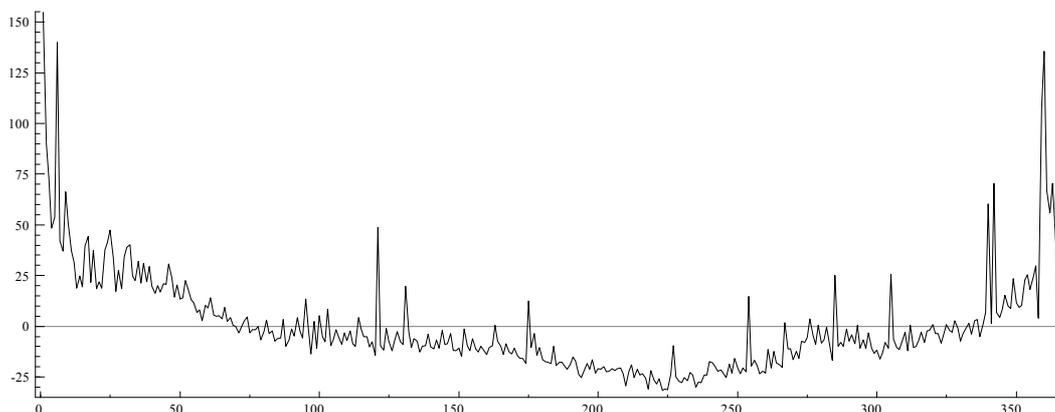
<sup>21</sup> Se probó a estimar un modelo con pendiente en el componente de nivel, pero ésta resultó no estocástica y, además, no significativa en el vector de estado final.

<sup>22</sup> Por supuesto, todas estas medias fueron corregidas de modo que la suma de los efectos estacionales fuese nula. Además, se optó por eliminar los datos correspondientes al 29 de febrero de 1996 y de 2000 antes de calcular dichas medias.

**Gráfico 4.2.2. Variación estacional de periodo anual (medias por día del año, opción b)**



**Gráfico 4.2.3. Variación estacional de periodo anual (medias por día del año, opción c)**



Como puede advertirse (gráficos 4.2.1 y 4.2.2), cuando se asume nivel estocástico en la estimación del modelo

$$y_t = \mathbf{m}_t + \mathbf{g}_t^{semanal} + \mathbf{e}_t, \mathbf{e}_t \sim NID(0, \sigma_\epsilon^2),$$

desaparece el patrón estacional de periodo anual que se observa en la serie original. Es decir, dicha variación estacional está siendo absorbida por el nivel estocástico y, por tanto, no queda recogida en el residuo del modelo estimado, aunque se refleja, en cambio, con claridad en la representación del componente de nivel. Por esta razón, las *splines* especificadas a partir de este supuesto tampoco resultarán capaces de recoger este comportamiento. Por tanto, en adelante se hará referencia a la variación estacional de periodo anual obtenida a partir de la hipótesis de que tanto el nivel como la

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

estacionalidad de periodo semanal poseen carácter determinístico (gráfico 4.2.3). Las medias por día del año obtenidas de esta forma se denotarán por  $\hat{g}_{1,d}^{anual}$ ,  $d = 1, \dots, 365$ .

Dada la variabilidad observada en el gráfico seleccionado para apreciar la fluctuación estacional de periodo anual, la determinación del número y localización de los puntos de ruptura que delimitan los tramos de la función *spline*, e incluso la identificación de la forma funcional en cada tramo, es una tarea bastante complicada. Con objeto de evitar que la componente ruidosa presente en estas medias distorsione la especificación de la *spline*, se ha procedido a ajustar un modelo estructural de nivel local<sup>23</sup> a la serie de medias por día del año antes señalada y elegir la especificación de la función *spline* a partir del nivel subyacente estimado, que se denotará por  $\hat{g}_{1,d}^{anual(m)}$  (gráfico 4.2.4). Finalmente, se consideró una función *spline* de dieciocho tramos con forma funcional cúbica en cada uno de ellos.

**Gráfico 4.2.4. Estimación del componente de nivel  $\hat{g}_{1,d}^{anual(m)}$**



Se asume que la fluctuación estacional de periodo anual es determinística; por tanto,  $g_t^{anual} = \hat{g}_{1,d}^{anual(m)}$ , si la observación en el instante  $t$  corresponde al día del año  $d$ ,  $d = 1, \dots, 365$ . Entonces, la variación estacional de periodo anual puede escribirse como  $\hat{g}_{1,d}^{anual(m)} = g(d) + e_d$ , donde  $e_d$  es un término de error asociado al ajuste y  $g(d)$  es la *spline* finalmente especificada como  $g(d) = g_i(d)$ , donde  $g_i(d) = g_{i,0} + g_{i,1}d + g_{i,2}d^2 + g_{i,3}d^3$ , con  $d_{i-1} \leq d \leq d_i$ , para  $i = 1, \dots, 18$ ,  $d_0 = 1$  y  $d_{18} = 365$ .

<sup>23</sup> Véase capítulo 1:11.

Esta *spline* debe verificar las siguientes condiciones:

a) restricciones de continuidad de la función,

$$g_{i,0} + g_{i,1}d_i + g_{i,2}d_i^2 + g_{i,3}d_i^3 = g_{i+1,0} + g_{i+1,1}d_i + g_{i+1,2}d_i^2 + g_{i+1,3}d_i^3,$$

para  $i = 1, \dots, 17$ , y

$$g_{18,0} + g_{18,1}(d_{18} + 1) + g_{18,2}(d_{18} + 1)^2 + g_{18,3}(d_{18} + 1)^3 = g_{1,0} + g_{1,1}d_0 + g_{1,2}d_0^2 + g_{1,3}d_0^3;$$

b) restricciones de continuidad de la derivada primera,

$$g_{i,1} + 2g_{i,2}d_i + 3g_{i,3}d_i^2 = g_{i+1,1} + 2g_{i+1,2}d_i + 3g_{i+1,3}d_i^2$$

para  $i = 1, \dots, 17$ , y

$$g_{18,1} + 2g_{18,2}(d_{18} + 1) + 3g_{18,3}(d_{18} + 1)^2 = g_{1,1} + 2g_{1,2}d_0 + 3g_{1,3}d_0^2;$$

c) restricciones de continuidad de la derivada segunda,

$$2g_{i,2} + 6g_{i,3}d_i = 2g_{i+1,2} + 6g_{i+1,3}d_i$$

para  $i = 1, \dots, 17$ , y

$$2g_{18,2} + 6g_{18,3}(d_{18} + 1) = 2g_{1,2} + 6g_{1,3}d_0.$$

Una vez especificada la forma funcional en cada uno de los tramos que definen la *spline* y si se asume que los puntos de ruptura  $d_i$ ,  $i = 1, \dots, 17$ , son conocidos<sup>24</sup>, el siguiente paso es determinar los parámetros  $g_{i,0}$ ,  $g_{i,1}$ ,  $g_{i,2}$  y  $g_{i,3}$ , para  $i = 1, \dots, 18$ . Dado que el número de parámetros que se desea determinar, 72, es superior al número de condiciones impuestas, 54, se puede recurrir a un proceso de estimación de los parámetros que se decida dejar como libres. En este caso, se ha optado por expresar el vector de parámetros  $G_{54 \times 1} = (g_{1,1}, g_{1,2}, g_{1,3}, \dots, g_{18,1}, g_{18,2}, g_{18,3})'$  como función de  $G_{18 \times 1}^* = (g_{1,0}, g_{2,0}, \dots, g_{18,0})'$ , que son los parámetros finalmente estimados, y de  $d_i$ ,  $i = 0, \dots, 18$ . Teniendo en cuenta el planteamiento del problema, conviene expresar el conjunto de restricciones anteriores de forma matricial como  $DG = R$ , donde

<sup>24</sup> Más adelante se determinará la localización de los puntos que delimitan cada uno de los tramos de la función *spline*, siguiendo el procedimiento descrito en el capítulo segundo.

Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$D_{54 \times 54} = \begin{bmatrix} d_1 & d_1^2 & d_1^3 & -d_1 & -d_1^2 & -d_1^3 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & d_{17} & d_{17}^2 & d_{17}^3 & -d_{17} & -d_{17}^2 & -d_{17}^3 \\ -d_0 & -d_0^2 & -d_0^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & (d_{18}+1) & (d_{18}+1)^2 & (d_{18}+1)^3 \\ 1 & 2d_1 & 3d_1^2 & -1 & -2d_1 & -3d_1^2 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 2d_{17} & 3d_{17}^2 & -1 & -2d_{17} & -3d_{17}^2 \\ -1 & -2d_0 & -3d_0^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 2(d_{18}+1) & 3(d_{18}+1)^2 \\ 0 & 2 & 6d_1 & 0 & -2 & -6d_1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 2 & 6d_{17} & 0 & -2 & -6d_{17} \\ 0 & -2 & -6d_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 2 & 6(d_{18}+1) \end{bmatrix}$$

y

$$R_{54 \times 1} = \begin{bmatrix} -g_{1,0} + g_{2,0} \\ \vdots \\ -g_{17,0} + g_{18,0} \\ -g_{18,0} + g_{1,0} \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

de modo que  $G = D^{-1}R$ . Si se denota la inversa de la matriz  $D$  por

$$D^{-1}_{54 \times 54} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,54} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,54} \\ \vdots & \vdots & \ddots & \vdots \\ a_{54,1} & a_{54,2} & \cdots & a_{54,54} \end{bmatrix}$$

y se define  $g_r$ ,  $r = 1, \dots, 54$ , tal que  $g_1 = g_{1,1}$ ,  $g_2 = g_{1,2}$ ,  $g_3 = g_{1,3}$ ,  $g_4 = g_{2,1}$ ,  $g_5 = g_{2,2}$ ,  $g_6 = g_{2,3}$ ,  $\dots$ ,  $g_{52} = g_{18,1}$ ,  $g_{53} = g_{18,2}$  y  $g_{54} = g_{18,3}$ , entonces

$$g_r = a_{r,1}(-g_{1,0} + g_{2,0}) + a_{r,2}(-g_{2,0} + g_{3,0}) + \dots + a_{r,17}(-g_{17,0} + g_{18,0}) + a_{r,18}(g_{1,0} - g_{18,0})$$

para  $r = 1, \dots, 54$ . Es decir,

$$g_r = \mathbf{a}_{r,1}g_{1,0} + \mathbf{a}_{r,2}g_{2,0} + \dots + \mathbf{a}_{r,17}g_{17,0} + \mathbf{a}_{r,18}g_{18,0},$$

donde

$$\mathbf{a}_{r,1} = -a_{r,1} + a_{r,18}, \mathbf{a}_{r,2} = a_{r,1} - a_{r,2}, \dots, \mathbf{a}_{r,17} = a_{r,16} - a_{r,17} \text{ y } \mathbf{a}_{r,18} = a_{r,17} - a_{r,18},$$

para  $r = 1, \dots, 54$ . Ahora, la *spline* puede especificarse en función de  $g_{i,0}$ ,  $i = 1, \dots, 18$ , de la siguiente forma:

$$\begin{aligned} g(d) = & g_{i,0} \\ & + (\mathbf{a}_{3i-2,1}g_{1,0} + \mathbf{a}_{3i-2,2}g_{2,0} + \dots + \mathbf{a}_{3i-2,17}g_{17,0} + \mathbf{a}_{3i-2,18}g_{18,0})d \\ & + (\mathbf{a}_{3i-1,1}g_{1,0} + \mathbf{a}_{3i-1,2}g_{2,0} + \dots + \mathbf{a}_{3i-1,17}g_{17,0} + \mathbf{a}_{3i-1,18}g_{18,0})d^2 \\ & + (\mathbf{a}_{3i,1}g_{1,0} + \mathbf{a}_{3i,2}g_{2,0} + \dots + \mathbf{a}_{3i,17}g_{17,0} + \mathbf{a}_{3i,18}g_{18,0})d^3 \end{aligned}$$

si  $d_{i-1} \leq d \leq d_i$ ,  $i = 1, \dots, 18$ ; y finalmente como:

$$g(d) = \sum_{i=1}^{18} \left[ \begin{array}{l} g_{i,0} + (\mathbf{a}_{3i-2,1}g_{1,0} + \mathbf{a}_{3i-2,2}g_{2,0} + \dots + \mathbf{a}_{3i-2,17}g_{17,0} + \mathbf{a}_{3i-2,18}g_{18,0})d \\ + (\mathbf{a}_{3i-1,1}g_{1,0} + \mathbf{a}_{3i-1,2}g_{2,0} + \dots + \mathbf{a}_{3i-1,17}g_{17,0} + \mathbf{a}_{3i-1,18}g_{18,0})d^2 \\ + (\mathbf{a}_{3i,1}g_{1,0} + \mathbf{a}_{3i,2}g_{2,0} + \dots + \mathbf{a}_{3i,17}g_{17,0} + \mathbf{a}_{3i,18}g_{18,0})d^3 \end{array} \right] D_{i,d}$$

donde  $D_{i,d} = \begin{cases} 1, & \text{si } d_{i-1} \leq d < d_i \\ 0, & \text{resto} \end{cases}$ ,  $i = 1, 2, \dots, 17$ , y  $D_{18,d} = \begin{cases} 1, & \text{si } d_{17} \leq d \leq d_{18} \\ 0, & \text{resto} \end{cases}$ . Y

reordenando términos, se tiene que

$$g(d) = g_{1,0}X_{1,d} + g_{2,0}X_{2,d} + \dots + g_{17,0}X_{17,d} + g_{18,0}X_{18,d},$$

donde

$$X_{i,d} = D_{i,d} + (\acute{a}_{1,i}d + \acute{a}_{2,i}d^2 + \acute{a}_{3,i}d^3)D_{1,d} + (\acute{a}_{4,i}d + \acute{a}_{5,i}d^2 + \acute{a}_{6,i}d^3)D_{2,d} \\ + \dots + (\acute{a}_{52,i}d + \acute{a}_{53,i}d^2 + \acute{a}_{54,i}d^3)D_{18,d}$$

para  $i = 1, \dots, 18$ .

Entonces, la regresión

$$\hat{\mathbf{g}}_{1,d}^{anual(m)} = g(d) + \mathbf{e}_d = g_{1,0}X_{1,d} + g_{2,0}X_{2,d} + \dots + g_{17,0}X_{17,d} + g_{18,0}X_{18,d} + \mathbf{e}_d$$

puede estimarse por mínimos cuadrados y obtener así el vector  $\hat{G}^*$ .

Dado que los puntos de ruptura que delimitan los tramos de la función *spline* no se conocen *a priori*, se ha optado por elegir como combinación óptima de localizaciones aquélla que minimice la suma de cuadrados de residuos en la regresión anterior. Siguiendo este criterio, y sabiendo que  $d_0 = 1$  y  $d_{18} = 365$ , la combinación óptima de los restantes puntos de ruptura para la forma funcional cúbica en cada uno de los dieciocho tramos es  $d_1 = 2$ ,  $d_2 = 3$ ,  $d_3 = 26$ ,  $d_4 = 66$ ,  $d_5 = 135$ ,  $d_6 = 138$ ,  $d_7 = 180$ ,  $d_8 = 182$ ,  $d_9 = 276$ ,  $d_{10} = 278$ ,  $d_{11} = 279$ ,  $d_{12} = 341$ ,  $d_{13} = 342$ ,  $d_{14} = 344$ ,  $d_{15} = 358$ ,  $d_{16} = 363$  y  $d_{17} = 364$ .

A partir de las estimaciones de los parámetros del vector  $G^*$  se obtuvo una segunda estimación del componente estacional de periodo anual<sup>25</sup>, que se denotará por  $\hat{\mathbf{g}}_{2,d}^{anual}$  y que aparece representada en el gráfico 4.3.3. Sin embargo, como se mostrará más adelante, resulta más apropiado que la variación estacional se estime conjuntamente

<sup>25</sup> Las estimaciones de los parámetros del vector  $G^*$  fueron las siguientes:  $\hat{g}_{1,0} = 50.8704$ ,  $\hat{g}_{2,0} = 54.8064$ ,  $\hat{g}_{3,0} = 115.6358$ ,  $\hat{g}_{4,0} = 1.1290$ ,  $\hat{g}_{5,0} = 160.3688$ ,  $\hat{g}_{6,0} = -8247.4583$ ,  $\hat{g}_{7,0} = 1109.4789$ ,  $\hat{g}_{8,0} = -25561.7063$ ,  $\hat{g}_{9,0} = 263.5888$ ,  $\hat{g}_{10,0} = 640785.9076$ ,  $\hat{g}_{11,0} = -1150794.812$ ,  $\hat{g}_{12,0} = -8539.7372$ ,  $\hat{g}_{13,0} = 13162464.25$ ,  $\hat{g}_{14,0} = -6796804.434$ ,  $\hat{g}_{15,0} = -498719.0676$ ,  $\hat{g}_{16,0} = 17891427.21$ ,  $\hat{g}_{17,0} = -449342332.4$  y  $\hat{g}_{18,0} = 640785.9076$ .

con el resto de componentes de la serie. Con este objeto y para las localizaciones señaladas, se han generado las variables  $X_{1,t}, \dots, X_{18,t}$ , que son vectores columna de dimensión  $2192 \times 1$ , definidos de forma que  $X_{i,t} = X_{i,d}$ ,  $i = 1, \dots, 18$ , si la observación registrada en el instante  $t$  pertenece al día del año  $d^{26}$ . Las variables  $X_{2,t}, \dots, X_{18,t}$  se han incorporado como variables exógenas para completar la especificación del componente estacional en el modelo estructural

$$y_t = \mathbf{m}_t + \mathbf{g}_t^{semanal} + g_{2,0}^* X_{2,t} + \dots + g_{17,0}^* X_{17,t} + g_{18,0}^* X_{18,t} + \mathbf{e}_t, \quad \mathbf{e}_t \sim NID(0, \sigma_\epsilon^2),$$

que se estima en el apartado siguiente<sup>27</sup>.

### 4.3. Estimación del modelo estructural

A pesar de que la representación gráfica de la serie sugiere empezar el proceso de estimación con un modelo más simple, se optó por iniciar la estrategia de identificación de la naturaleza de los componentes mediante la estimación de un modelo estructural con pendiente en la especificación del componente tendencial. La estimación máximo verosímil de este modelo revela que la varianza del término de perturbación de la pendiente es nula, lo que ratifica su carácter no estocástico. Una segunda estimación en la que se consideró el carácter no estocástico de la pendiente, condujo a admitir que su presencia no era necesaria dada su no significación en el vector de estado final<sup>28</sup>.

Se especificó, entonces, el modelo

$$y_t = \mathbf{m}_t + \mathbf{g}_t^{semanal} + g_{2,0}^* X_{2,t} + \dots + g_{17,0}^* X_{17,t} + g_{18,0}^* X_{18,t} + \mathbf{e}_t, \quad \mathbf{e}_t \sim NID(0, \sigma_\epsilon^2),$$

con

$$\mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{h}_t, \quad \mathbf{h}_t \sim NID(0, \sigma_\eta^2),$$

y

$$\gamma_t^{semanal} = \sum_{j=1}^3 \gamma_{j,t}^{semanal},$$

<sup>26</sup> En la observación correspondiente al 29 de febrero de 1996 y de 2000, las variables  $X_{1,t}, \dots, X_{18,t}$ , se han definido de manera que la *spline* estimada para esa fecha coincida con la estimación obtenida para el día anterior.

<sup>27</sup> Nótese que se ha eliminado la variable  $X_{1,t}$ , dado que  $\sum_{i=1}^{18} X_{i,t} = 1, \forall t$ . Y esta restricción se ha incorporado en el modelo estructural estimado.

<sup>28</sup> La estimación de la pendiente fue  $\hat{\mathbf{b}} = 0.0089562$  y el estadístico de significación individual  $t = 0.1553$ .

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

donde  $g_{j,t}^{semanal}$  es generado por

$$\begin{pmatrix} \gamma_{j,t}^{semanal} \\ \gamma_{j,t}^{semanal*} \end{pmatrix} = \begin{pmatrix} \cos \lambda_j & \text{sen } \lambda_j \\ -\text{sen } \lambda_j & \cos \lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1}^{semanal} \\ \gamma_{j,t-1}^{semanal*} \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix},$$

$$I_j = \frac{2pj}{7},$$

$$\omega_{j,t}, \omega_{j,t}^* \sim NID(0, \sigma_\omega^2),$$

para  $j = 1, 2, 3$  y  $t = 1, \dots, 2192$ .

La estimación de este modelo (cuadros 4.3.1 a 4.3.4) muestra, por una parte, el carácter estocástico tanto para el componente de nivel como para el componente estacional de periodicidad semanal<sup>29</sup> y, por otra, la significación individual de buena parte de los regresores que definen la función *spline*<sup>30</sup>. Por tanto, el modelo estructural que parece adecuado debe incluir un componente de nivel estocástico, estacionalidad de periodo semanal estocástica y componente irregular, además de la variación estacional de periodo anual recogida con la *spline*, que se denotará por  $\gamma_{3,d}^{anual}$ .

**Cuadro 4.3.1. Varianzas de los términos de perturbación**

	Irregular ( $\hat{S}_e^2$ )	Nivel ( $\hat{S}_h^2$ )	Estacionalidad semanal ( $\hat{S}_w^2$ )
<b>Varianza</b>	208.84	7.1914	1.4042
<b>Ratio señal-ruido<sup>31</sup></b>	(1.0000)	(0.0344)	(0.0067)

<sup>29</sup> A pesar del reducido valor del q-ratio correspondiente a la fluctuación estacional de periodo semanal, esta variación parece estar sujeta a cambios lentos pero permanentes, por lo que, como advierte Scott (1995), conviene mantener su formulación estocástica.

<sup>30</sup> Al observar que algunos de los parámetros que definen la *spline* no resultaron significativos, se llevó a cabo un contraste que permitiera obtener alguna conclusión sobre la significación conjunta de dichos parámetros. A partir de dicho contraste,  $F=16.1088 \sim F_{17, 2088}$ , se decidió que las variables  $X_{2,t}, \dots, X_{18,t}$ , eran, en su conjunto, relevantes y, por tanto, deberían permanecer como elementos del modelo que recogen una significativa variación estacional de periodo anual. No obstante, debe recordarse el sentido del sesgo de este contraste, ya señalado en la nota 41 del capítulo tercero y motivado por el hecho de que las fluctuaciones no incorporadas explícitamente pueden ser absorbidas por el componente de nivel estocástico. De hecho, y como ya se comentó en el epígrafe 4.2, esto es lo que ocurría cuando se trataba de extraer las medias por día del año a partir de un modelo estructural de nivel local.

<sup>31</sup> En este caso se habla de ratio señal-ruido porque la varianza mayor es la del término de perturbación del componente irregular y los ratios miden el cociente entre la varianza de cada uno de los términos de perturbación de los componentes y la del componente ruidoso.

**Cuadro 4.3.2. Vector de estado final**

Componente	Estimación	Componente	Estimación
$\hat{m}_T$	37.619	$\hat{g}_{2,T}^{semanal*}$	-23.625
$\hat{g}_{1,T}^{semanal}$	14.609	$\hat{g}_{3,T}^{semanal}$	-23.770
$\hat{g}_{1,T}^{semanal*}$	-14.291	$\hat{g}_{3,T}^{semanal*}$	-13.628
$\hat{g}_{2,T}^{semanal}$	-22.427		

El valor del estadístico del contraste de significación conjunta del componente estacional de periodo semanal fue  $c_6^2 = 101.858$ .

**Cuadro 4.3.3. Parámetros libres en la especificación de la *spline***

Parámetro	Estimación	Estadístico <i>t</i>	Parámetro	Estimación	Estadístico <i>t</i>
$\hat{g}_{2,0}^*$	301.86	6.9179	$\hat{g}_{11,0}^*$	-681070	-0.3738
$\hat{g}_{3,0}^*$	87.93	12.4300	$\hat{g}_{12,0}^*$	-14308	-2.3835
$\hat{g}_{4,0}^*$	-13.37	-0.8122	$\hat{g}_{13,0}^*$	$19.31 \times 10^6$	2.1832
$\hat{g}_{5,0}^*$	147.29	3.1346	$\hat{g}_{14,0}^*$	$-74.42 \times 10^5$	-1.2647
$\hat{g}_{6,0}^*$	-9122.70	-1.3877	$\hat{g}_{15,0}^*$	$-18.77 \times 10^5$	-4.1077
$\hat{g}_{7,0}^*$	1239.40	1.1885	$\hat{g}_{16,0}^*$	$46.97 \times 10^6$	12.1190
$\hat{g}_{8,0}^*$	-28039	-1.2273	$\hat{g}_{17,0}^*$	$-12.74 \times 10^8$	-12.9510
$\hat{g}_{9,0}^*$	249.38	0.4006	$\hat{g}_{18,0}^*$	$10.65 \times 10^8$	11.1090
$\hat{g}_{10,0}^*$	466290	0.5485			

**Cuadro 4.3.4. Componente estacional de periodo semanal en los últimos siete días de la muestra**

Día	Estimación	Día	Estimación
Martes, 25-12-01	-4.4798	Sábado, 29-12-01	-4.8312
Miércoles, 26-12-01	9.1618	Domingo, 30-12-01	75.5517
Jueves, 27-12-01	-22.7478	Lunes, 31-12-01	-31.5810
Viernes, 28-12-01	-20.7385		

Pero antes de adoptar conclusiones definitivas sobre la bondad del modelo, deben examinarse las innovaciones o errores de predicción y los residuos de los componentes estocásticos. El examen de los residuos de los componentes de nivel e irregular en cada momento del tiempo y, especialmente, sus representaciones gráficas muestran la presencia de observaciones anómalas, que exigen reespecificar el modelo en los términos siguientes:

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$y_t = \mathbf{m}_t + \mathbf{g}_t^{semanal} + g_{2,0}^* X_{2,t} + \dots + g_{17,0}^* X_{17,t} + g_{18,0}^* X_{18,t} + \sum_{j-k-\ell} I_{j-k-\ell} I_{j-k-\ell} + \mathbf{e}_t,$$

donde  $\mathbf{e}_t \sim NID(0, \sigma_e^2)$ , siendo  $I_{j-k-\ell}$  una variable de intervención que toma valor uno si la observación pertenece al día  $j$  del mes  $k$  del año  $\ell$ . Después de estimar distintos modelos que diferían entre sí en función de las intervenciones consideradas, la especificación elegida es aquella cuya estimación proporcionó los resultados que aparecen en los cuadros 4.3.5 a 4.3.9 y los gráficos 4.3.1 a 4.3.5.

**Cuadro 4.3.5. Varianzas de los términos de perturbación**

	Irregular ( $\hat{\sigma}_e^2$ )	Nivel ( $\hat{\sigma}_h^2$ )	Estacionalidad semanal ( $\hat{\sigma}_w^2$ )
<b>Varianza</b>	73.2690	3.7292	0.6042
<b>Ratio señal-ruido</b>	(1.0000)	(0.0509)	(0.0082)

**Cuadro 4.3.6. Vector de estado final**

Componente	Estimación	Componente	Estimación
$\hat{\mathbf{m}}_T$	21.955	$\hat{\mathbf{g}}_{2,T}^{semanal*}$	-11.863
$\hat{\mathbf{g}}_{1,T}^{semanal}$	11.933	$\hat{\mathbf{g}}_{3,T}^{semanal}$	-23.030
$\hat{\mathbf{g}}_{1,T}^{semanal*}$	-13.439	$\hat{\mathbf{g}}_{3,T}^{semanal*}$	-6.349
$\hat{\mathbf{g}}_{2,T}^{semanal}$	-11.700		

El valor del estadístico del contraste de significación conjunta del componente estacional de periodo semanal fue  $\mathbf{c}_6^2 = 103.427$ .

**Cuadro 4.3.7. Parámetros libres en la especificación de la *spline***

Parámetro	Estimación	Parámetro	Estimación	Parámetro	Estimación
$\hat{g}_{2,0}^*$	208.120	$\hat{g}_{8,0}^*$	-26813(*)	$\hat{g}_{14,0}^*$	908.01x10 <sup>4</sup>
$\hat{g}_{3,0}^*$	34.229	$\hat{g}_{9,0}^*$	167.61(*)	$\hat{g}_{15,0}^*$	-153.99x10 <sup>4</sup>
$\hat{g}_{4,0}^*$	-25.779	$\hat{g}_{10,0}^*$	102.51x10 <sup>4</sup> (*)	$\hat{g}_{16,0}^*$	238.44x10 <sup>5</sup>
$\hat{g}_{5,0}^*$	95.739	$\hat{g}_{11,0}^*$	-207.75x10 <sup>4</sup> (*)	$\hat{g}_{17,0}^*$	-641.22x10 <sup>6</sup>
$\hat{g}_{6,0}^*$	-7332.8(*)	$\hat{g}_{12,0}^*$	-696.49(*)	$\hat{g}_{18,0}^*$	588.15x10 <sup>6</sup>
$\hat{g}_{7,0}^*$	1106.4(*)	$\hat{g}_{13,0}^*$	-109.85x10 <sup>5</sup> (*)		

(\*) Parámetros no significativos al 95 % de confianza.

Número diario de visitas a un servicio de urgencias del Barcelonès Nord

Cuadro 4.3.8. Parámetros de las variables de intervención

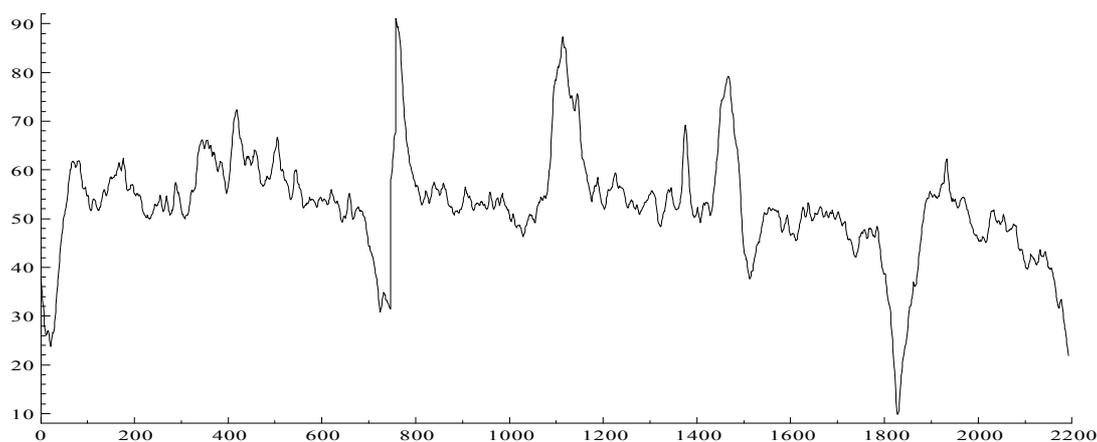
Parámetro	Estimación	Parámetro	Estimación	Parámetro	Estimación
$\hat{I}_{08-04-96(f)}$	54.953	$\hat{I}_{11-01-98(d)}$	-51.302	$\hat{I}_{11-09-99(f)}$	43.358
$\hat{I}_{01-05-96(f)}$	65.629	$\hat{I}_{25-01-98(d)}$	101.810	$\hat{I}_{30-09-99}$	41.177
$\hat{I}_{27-05-96(f)}$	52.926	$\hat{I}_{27-01-98}$	23.306	$\hat{I}_{01-10-99}$	30.671
$\hat{I}_{01-11-96(f)}$	50.339	$\hat{I}_{01-02-98(d)}$	144.170	$\hat{I}_{03-10-99(d)}$	66.402
$\hat{I}_{06-12-96(f)}$	103.440	$\hat{I}_{07-02-98(f)}$	-40.402	$\hat{I}_{06-12-99(f)}$	90.611
$\hat{I}_{08-12-96(f)}$	53.696	$\hat{I}_{08-02-98(d)}$	73.328	$\hat{I}_{08-12-99(f)}$	65.894
$\hat{I}_{15-12-96(f)}$	73.571	$\hat{I}_{15-02-98(d)}$	49.686	$\hat{I}_{25-12-99(f)}$	120.760
$\hat{I}_{25-12-96(f)}$	75.426	$\hat{I}_{10-04-98(f)}$	48.740	$\hat{I}_{31-12-99(f)}$	-58.227
$\hat{I}_{26-12-96(f)}$	95.622	$\hat{I}_{13-04-98(f)}$	61.265	$\hat{I}_{01-01-00(f)}$	124.970
$\hat{I}_{29-12-96(d)}$	104.480	$\hat{I}_{01-05-98(f)}$	42.558	$\hat{I}_{02-02-00(d)}$	125.910
$\hat{I}_{01-01-97(f)}$	92.982	$\hat{I}_{11-05-98(f)}$	52.372	$\hat{I}_{06-01-00(f)}$	181.330
$\hat{I}_{05-01-97(d)}$	90.607	$\hat{I}_{01-06-98(f)}$	64.130	$\hat{I}_{09-01-00(d)}$	186.370
$\hat{I}_{06-01-97(f)}$	127.470	$\hat{I}_{12-10-98(f)}$	67.267	$\hat{I}_{12-01-00}$	-25.436
$\hat{I}_{12-01-97(d)}$	67.802	$\hat{I}_{08-12-98(f)}$	86.674	$\hat{I}_{24-04-00(f)}$	50.538
$\hat{I}_{19-01-97(d)}$	64.758	$\hat{I}_{24-12-98(f)}$	-42.484	$\hat{I}_{01-05-00(f)}$	55.861
$\hat{I}_{27-01-97}$	-62.930	$\hat{I}_{25-12-98(f)}$	76.007	$\hat{I}_{12-06-00(f)}$	56.285
$\hat{I}_{01-05-97(f)}$	45.970	$\hat{I}_{26-12-98(f)}$	128.540	$\hat{I}_{08-12-00(f)}$	63.880
$\hat{I}_{01-11-97(f)}$	52.109	$\hat{I}_{27-12-98(d)}$	34.195	$\hat{I}_{27-12-00}$	-21.766
$\hat{I}_{06-12-97(f)}$	66.510	$\hat{I}_{31-12-98(f)}$	-63.291	$\hat{I}_{06-01-01(f)}$	39.837
$\hat{I}_{08-12-97(f)}$	66.801	$\hat{I}_{01-01-99(f)}$	123.460	$\hat{I}_{01-05-01(f)}$	58.598
$\hat{I}_{25-12-97(f)}$	42.341	$\hat{I}_{03-01-99(d)}$	111.830	$\hat{I}_{11-05-01(f)}$	44.397
$\hat{I}_{26-12-97(f)}$	87.755	$\hat{I}_{06-01-99(f)}$	141.510	$\hat{I}_{04-06-01(f)}$	51.601
$\hat{I}_{31-12-97(f)}$	-56.996	$\hat{I}_{10-01-99(d)}$	66.909	$\hat{I}_{06-12-01(f)}$	50.016
$\hat{I}_{06-01-98(f)}$	83.360	$\hat{I}_{05-04-99(f)}$	56.681	$\hat{I}_{08-12-01(f)}$	50.203
$\hat{I}_{11-01-98(d)}$	50.483	$\hat{I}_{01-05-99(f)}$	58.168	$\hat{I}_{26-12-01(f)}$	92.164
$\hat{I}_{09-01-98(v)}$	26.481	$\hat{I}_{24-05-99(f)}$	44.969	$\hat{I}_{30-12-01(d)}$	64.919

En las intervenciones se ha indicado si ese día fue festivo (f), domingo (d) o víspera de festivo (v).

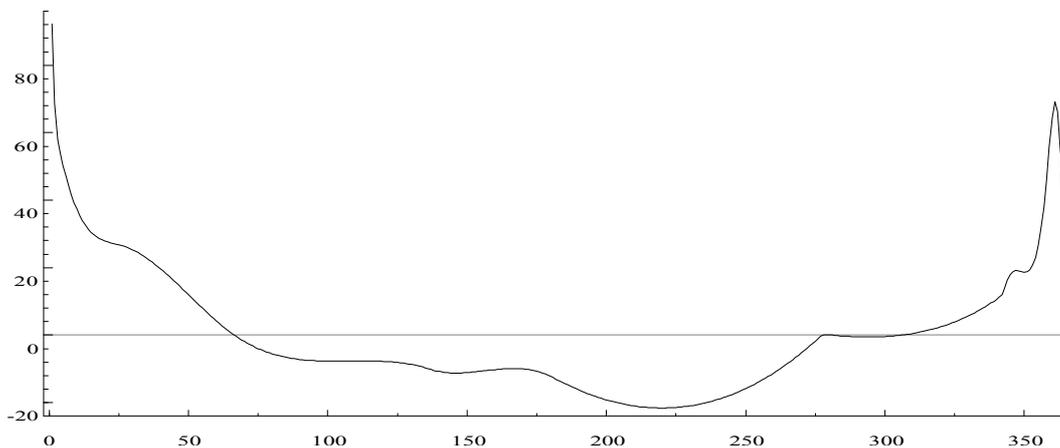
**Cuadro 4.3.9. Componente estacional de periodo semanal en los últimos siete días de la muestra**

Día	Estimación	Día	Estimación
Martes, 25-12-01	5.5326	Sábado, 29-12-01	-3.6387
Miércoles, 26-12-01	-9.8179	Domingo, 30-12-01	55.5455
Jueves, 27-12-01	-15.9411	Lunes, 31-12-01	-22.7905
Viernes, 28-12-01	-10.3808		

**Gráfico 4.3.1. Componente de nivel**



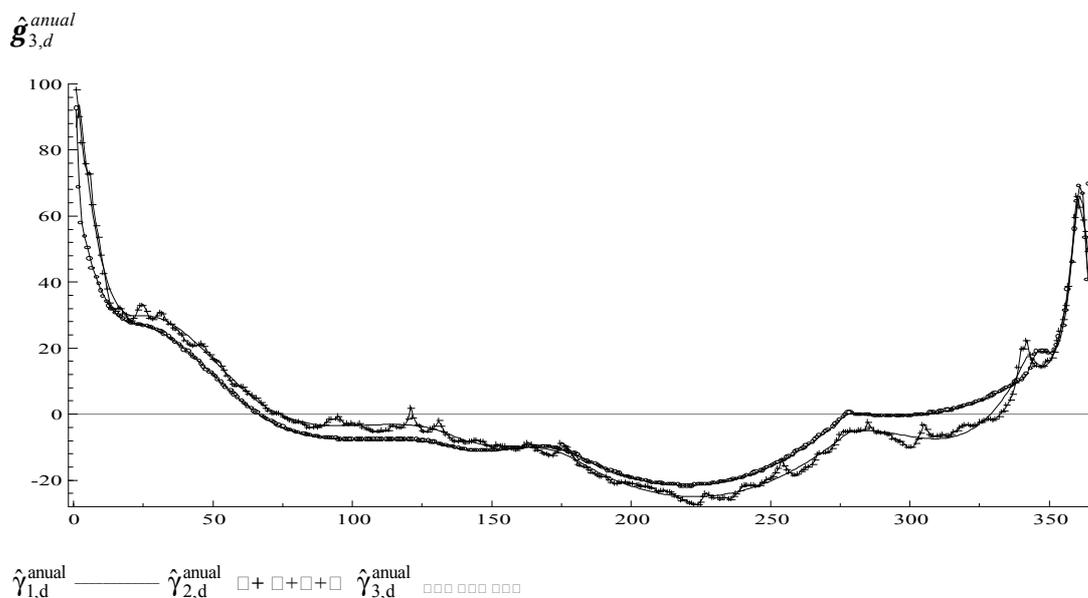
**Gráfico 4.3.2. Estacionalidad anual (año no bisiesto),  $\hat{g}_{3,d}^{anual}$**



La estimación de la fluctuación estacional anual,  $\hat{g}_{3,d}^{anual}$ , se ha obtenido a partir de los coeficientes estimados para las variables exógenas  $X_{2,t}, \dots, X_{18,t}$ . El término independiente de la *spline*, incorporado en el componente de nivel, se ha calculado de forma que la media de las fluctuaciones estacionales de periodo anual estimadas fuese nula. El componente de nivel representado, así como su estimación en el vector de

estado final (cuadro 4.3.6)<sup>32</sup>, también se han corregido oportunamente con el fin de que dicho término independiente no quedase recogido simultáneamente en los componentes tendencial y estacional. La comparación del componente de nivel representado con la serie original muestra el peso del componente estacional (gráficos 4.1.1 y 4.3.1)<sup>33</sup>.

**Gráfico 4.3.3. Variación estacional de periodo anual (año no bisiesto),  $\hat{g}_{1,d}^{anual}$ ,  $\hat{g}_{2,d}^{anual}$  y**



Nótese que en el gráfico 4.3.3. se han representado tres aproximaciones del mismo componente estacional de periodo anual. Se trata de las medias por día del año denotadas por  $\hat{g}_{1,d}^{anual}$ , la estimación aislada del componente,  $\hat{g}_{2,d}^{anual}$ , y la estimación fruto de la consideración conjunta de todos los componentes observados en la serie,  $\hat{g}_{3,d}^{anual}$ . Al contrario de lo que ocurría en la serie semanal del capítulo anterior, estas tres aproximaciones muestran ciertas diferencias. Por un lado, las diferencias observadas entre las medias por día del año y las dos *splines* estimadas pueden ser achacadas al carácter menos suave del componente en cuestión. Además, parece que la estimación conjunta proporciona una mejor aproximación de la fluctuación estacional de periodo anual, que puede ser explicada por la propia naturaleza del componente estacional. De hecho, el comportamiento tan fluctuante de dicho componente condujo a especificar la *spline* a partir del nivel subyacente observado en las medias por día del año. Por tanto,

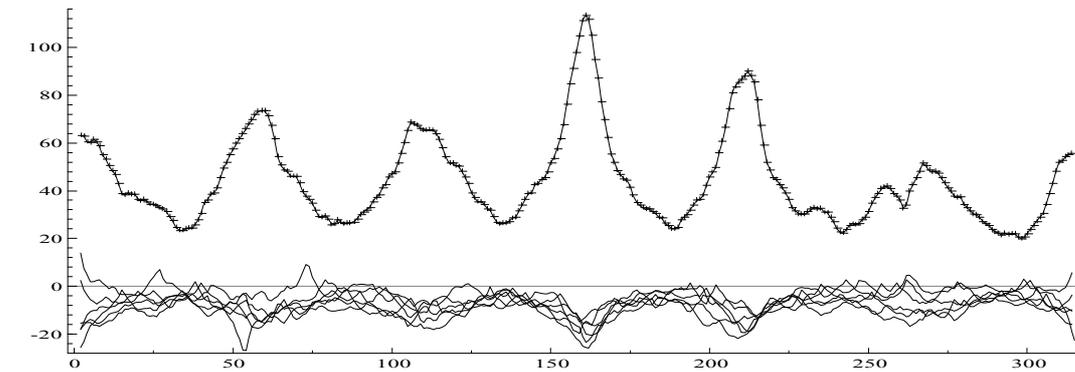
<sup>32</sup> La corrección correspondiente se efectuó también en la estimación del componente de nivel en el modelo sin intervenciones (cuadro 4.3.2).

<sup>33</sup> En términos medios, los efectos estacionales anual y semanal representan un 30.6860% y un 25.3463%, respectivamente, del nivel de la serie.

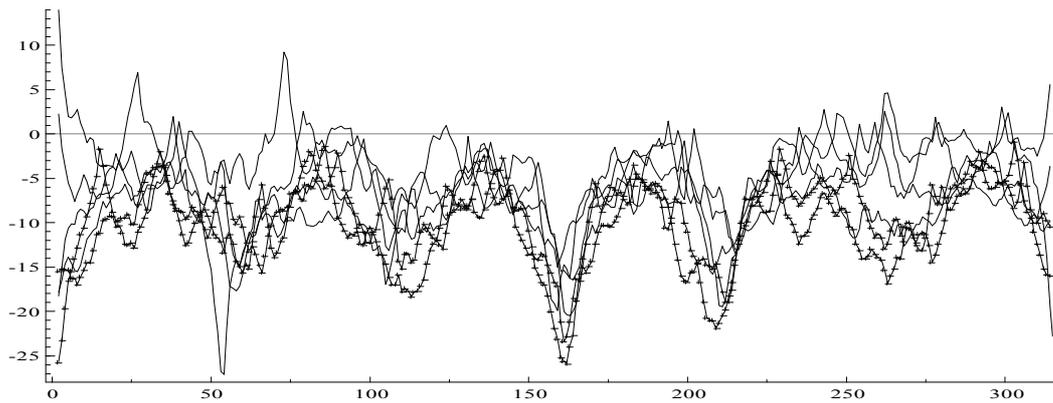
## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

cuando la *spline* se estima como función ajustada a la evolución de este nivel, sin tener en cuenta el resto de componentes de la serie, la especificación es mucho más rígida que cuando tal *spline* se estima, conjuntamente con el resto de componentes de la serie, para recoger un comportamiento estacional no suavizado artificialmente.

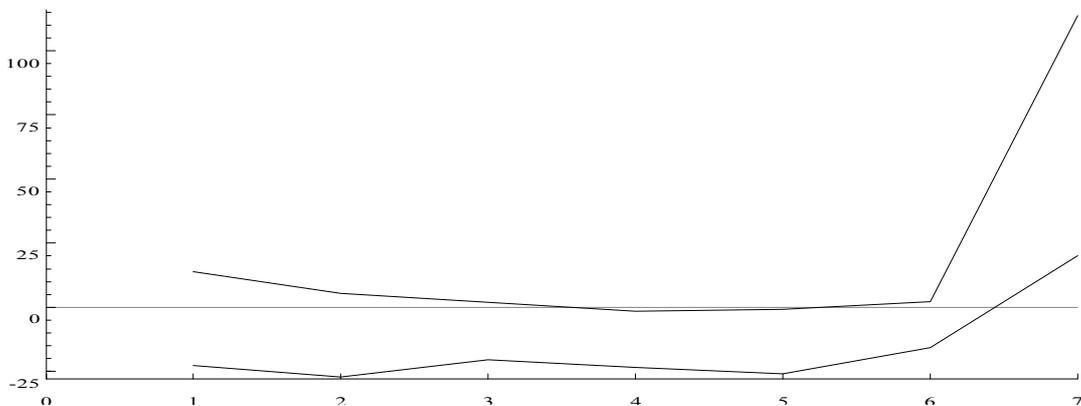
**Gráfico 4.3.4. Estacionalidad semanal (1996 - 2001)**



Domingo  $\square + \square + \square + \square$

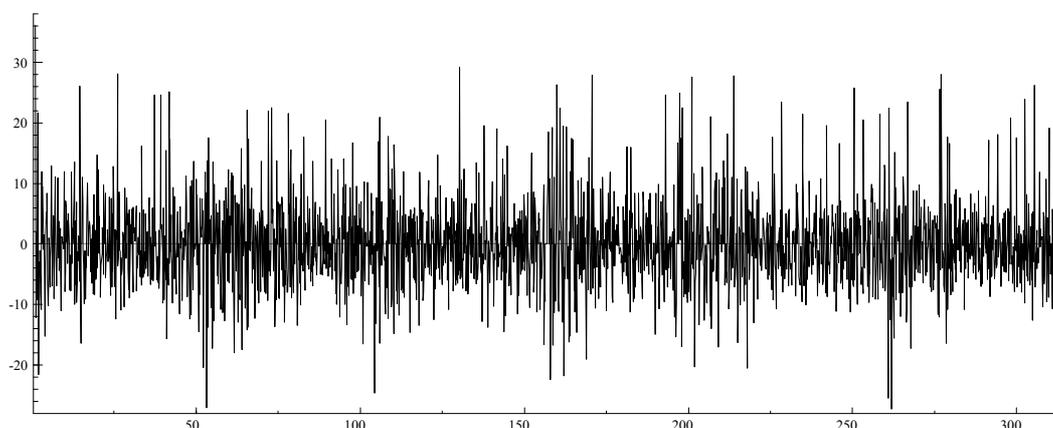


Jueves y viernes  $\square + \square + \square + \square$



En el gráfico 4.3.4, y con objeto de ilustrar el carácter estocástico de la fluctuación estacional semanal, se representa la evolución de la estimación de esta fluctuación correspondiente a cada uno de los siete días de la semana a lo largo de la muestra. Dado el nivel diferenciado de los domingos, se ha optado por representar separadamente las estimaciones correspondientes a los demás días de la semana. Se han representado también los valores mínimos y máximos del componente estimado en cada uno de los siete días. Estas representaciones muestran que la variación del patrón estacional es especialmente acentuada los domingos<sup>34</sup>.

**Gráfico 4.3.5. Componente irregular**



Por otro lado, la consideración de observaciones anómalas en la especificación y estimación del modelo, no altera las conclusiones sobre el carácter estocástico o no de los distintos componentes; sin embargo, es interesante destacar la disminución de las estimaciones de las distintas varianzas, pero sobre todo, la de la varianza del componente irregular, como era de esperar dada la naturaleza de las intervenciones incorporadas. La reducción en la varianza del término de perturbación que dirige la variación estacional de periodo semanal induce a cuestionar su carácter estocástico. Sin embargo, si se asumía que dicha variación era fija, el gráfico del componente de nivel sugería que dicho componente estaba absorbiendo, de nuevo, parte del comportamiento estacional. Todo ello condujo a mantener el carácter estocástico de la variación

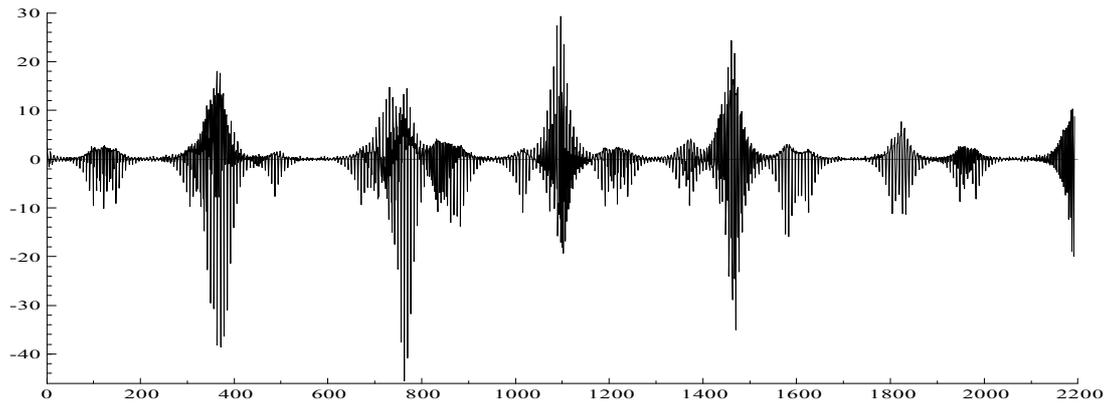
---

<sup>34</sup> Esta circunstancia explica que, cuando se analiza la transformación logarítmica de la serie, la variabilidad observada los domingos se amortigüe hasta el punto que la fluctuación estacional de periodo semanal adquiera una regularidad tal que pueda considerarse determinística. De hecho, la varianza estimada del término de perturbación de este componente fue nula (véase anexo 4.3). Este resultado no es, sin embargo, extraño, ya que los cambios en los logaritmos de series de datos

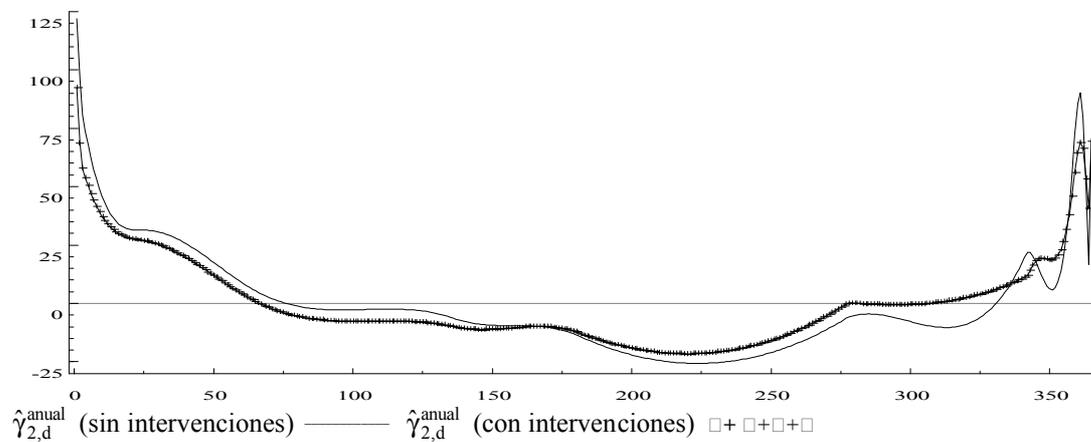
## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

estacional de periodo semanal, cuya estimación sí que se vio afectada por la incorporación de tales intervenciones. En el gráfico 4.3.6, en el que se representan las diferencias entre las estimaciones obtenidas en uno y otro caso, se aprecia, como era de esperar, que el efecto se concentra en aquellos periodos en los que se registró mayor número de intervenciones y las observaciones anómalas eran de mayor magnitud.

**Gráfico 4.3.6. Diferencias en las estimaciones de la fluctuación estacional de periodo semanal con y sin intervenciones**



**Gráfico 4.3.7. Estimaciones de la fluctuación estacional de periodo anual con y sin intervenciones**



Sin embargo, la estimación de la fluctuación estacional de periodo anual es aún más sensible a la introducción de variables de intervención, ya que ésta puede suponer, en el caso de que varias intervenciones correspondan al mismo día del año, una alteración significativa de los niveles medios registrados en ese día. El gráfico 4.3.7, en el que se representan las dos estimaciones para un año no bisiesto, refleja las diferencias

---

económicos pueden parecer más cercanos a un comportamiento no estocástico que los cambios en las series en niveles. En este sentido se pronuncian Banerjee y otros (1993:29).

entre ellas. En particular, la localización más frecuente de intervenciones, generalmente de signo positivo, en torno a principios y finales de año, se traduce en estimaciones inferiores de la fluctuación estacional en cuestión cuando tales intervenciones son incorporadas en el modelo. Y, más específicamente, merece la pena destacar lo que ocurre entre los días seis y ocho de diciembre, en los que la introducción de intervenciones suaviza las diferencias en la magnitud de la variación estacional anual correspondiente a los días de este *punte*. También resulta lógico que la estimación correspondiente al día uno de mayo sea inferior cuando se introducen variables de intervención, aunque la magnitud de la diferencia sea más pequeña que la que en principio podría esperarse teniendo en cuenta los valores estimados de los parámetros que acompañan a las intervenciones correspondientes a ese día. La explicación de este fenómeno debe buscarse en la especificación de la función *spline* para recoger la fluctuación anual, que supone una suavidad en el comportamiento estacional que puede no cumplirse en todas las estaciones (días) consideradas. De hecho, sólo con esta especificación puede admitirse la presencia de intervenciones localizadas en todos los días uno de mayo de la muestra, lo que no debe interpretarse, obviamente, como reflejo de anomalías en cada uno de esos días en relación con la magnitud de la variación estacional anual determinística correspondiente al uno de mayo, sino que, más bien, es consecuencia de las diferencias entre los niveles medios registrados en ese día en particular con respecto a los días circundantes.

Atendiendo a la fecha en la que fue preciso introducir intervenciones, es interesante destacar que la mayoría de ellas tienen lugar en domingo o en cualquier otro día de la semana que correspondiese a alguna festividad de carácter nacional o local. En estos casos, el parámetro estimado casi siempre resultó de signo positivo<sup>35</sup>, lo que significa un incremento en el número de visitas con respecto a lo predicho por el modelo que incorpora nivel estocástico y las fluctuaciones estacionales de periodo anual y semanal consideradas. Además, y dependiendo del día de la semana en el que se celebre la festividad, en la víspera de ese día se registran en ocasiones visitas al centro superiores a lo esperado. Este comportamiento se observa incluso en días que, no siendo festividad ni víspera, se convierten en no laborables. También, y dado el carácter determinista que se ha considerado para el componente estacional de periodo anual, es

---

<sup>35</sup> Sólo en nueve de las setenta y ocho intervenciones finalmente consideradas, los parámetros correspondientes eran de signo negativo, como reflejo de una notable disminución en el número de visitas registradas en esos días respecto a lo predicho por el modelo.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

lógico que el número de visitas en las festividades del viernes santo y lunes de pascua, de carácter cambiante, deba ser recogido mediante variables de intervención.

Como ya se comentó, la mayor concentración de intervenciones corresponde a los meses de finales y principios de año, fechas en las que ya se habían observado máximas afluencias al servicio de urgencias. Y, aunque este periodo tiene un comportamiento diferente al del resto del año, es posible establecer diferencias entre los meses que lo componen. Concretamente, y corroborando el comportamiento observado en el gráfico original de la serie, las fluctuaciones observadas en los meses de enero y febrero son más persistentes pero de menor cuantía que las observadas en los meses de finales de año. Fuera de este periodo, las observaciones anómalas correspondientes a los domingos se registran con mucha menor frecuencia.

En 1998 se registró un número de intervenciones mayor al del resto de los años considerados. Sin embargo, este hecho no permite emitir juicio alguno sobre la naturaleza de la variación estacional de periodo anual, que se intentó evaluar, a pesar del reducido número de años disponible, mediante la estimación del modelo

$$y_t^a = \mathbf{m}_t + \mathbf{g}_t^{anual} + \mathbf{e}_t, \quad \mathbf{e}_t \sim NID(0, \sigma_e^2),$$

donde  $\{y_t^a\}$  es el proceso resultante de filtrar la serie original de la estacionalidad estocástica de periodo diario, así como de los efectos de las variables de intervención estimados en el modelo finalmente propuesto,  $\mathbf{m}_t$  es un componente de nivel estocástico sin término de pendiente y  $\mathbf{g}_t^{anual}$  es el componente estacional estocástico de periodo anual. Dicho modelo sólo pudo ser estimado incorporando el componente  $\mathbf{g}_t^{anual}$  con la formulación de *dummies* estocásticas, es decir,  $\mathbf{g}_t^{anual} = \mathbf{g}_{d,t}^{anual}$ , si la observación en el instante  $t$  corresponde a la hora del año  $d$ , y  $\mathbf{g}_{d,t}^{anual} = \mathbf{g}_{d,t-1}^{anual} + \mathbf{w}_{d,t}$ ,  $d = 1, \dots, 365$ , donde  $\mathbf{w}_{d,t}$  sigue siendo un término de perturbación aleatoria tal que  $\mathbf{w}_{d,t} \sim NID(0, \sigma_w^2)$ . En ese caso, la varianza del término de perturbación de dicho componente fue nula, lo que indica su carácter determinístico<sup>36</sup>.

### 4.4. Conclusiones

El estudio de la serie diaria del número de visitas registradas en un centro de urgencias muestra la inestabilidad de lo que, en el enfoque clásico de series temporales,

---

<sup>36</sup> Antes de efectuar la estimación, se eliminaron las dos observaciones correspondientes al veintinueve de febrero de 1996 y 2000.

se define como componente tendencial, cuyo comportamiento se ajusta a un modelo de nivel estocástico con raíz unitaria. Respecto al componente estacional, existe una variación de periodo semanal que posee también un carácter estocástico. También se aprecia una variación estacional de periodo anual que se ha considerado determinística y ha sido modelada satisfactoriamente a través de una función *spline* cúbica de dieciocho tramos.

Desde el punto de vista metodológico, esta aplicación ilustra el verdadero potencial de las funciones *splines* para recoger variaciones estacionales. El recurso a estas funciones permite la estimación conjunta de variaciones periódicas de distinto periodo, que ni siquiera asumiendo el carácter no estocástico de la fluctuación de periodo anual podría ser llevada a cabo con los programas informáticos disponibles siguiendo especificaciones convencionales, salvo que se incluya un número extraordinariamente elevado de regresores que, por si mismo, puede causar problemas computacionales.

En cualquier caso, las variaciones estacionales registradas indican que los días de la semana con mayor demanda son los domingos y el menor número de visitas suele registrarse los jueves y viernes. Considerando en su conjunto los 365 días del año, las épocas de mayor saturación de este centro corresponden a principios y finales de año, mientras que en el verano se reducen notablemente las visitas, quizás como consecuencia de la disminución de la población de la zona en la que está ubicado el centro estudiado durante el periodo vacacional. Esta información debería ser tenida en cuenta por los gestores de los centros de salud como uno de los elementos que sirva de guía para reasignar los recursos disponibles.

En este sentido, no cabe duda que una base de datos horaria proporcionaría una información más rica de la demanda. Del mismo modo que el análisis diario puede indicar la conveniencia de redistribuir los recursos en función de que el día sea laborable o no, el estudio por hora podría ser válido para establecer una distribución óptima de los turnos del personal, condicionada por la organización y flexibilidad del centro<sup>37</sup>. Las dotaciones de medios materiales y humanos podrían incluso adaptarse mejor a la realidad social del entorno del centro en cuestión si se dispusiera de

---

<sup>37</sup> Por ejemplo, en el informe del Defensor del Pueblo se señala como hecho constatado el desajuste en la demanda de los servicios prestados por el personal de enfermería, que experimenta un fortísimo incremento los domingos y festivos de nueve a catorce horas y de diecisiete a veintiuna horas, o por los asistentes sociales, cuya actuación es necesaria fuera del horario que tienen establecido en los centros hospitalarios.

## **Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia**

información más exhaustiva, como puede ser la relativa a la incidencia horaria de cada una de las diferentes patologías presentadas por los pacientes que acuden a ese servicio de urgencias.

Si el ámbito de estudio se extiende a otros centros, cabría la posibilidad de evaluar la conveniencia de desviar la demanda de servicios de urgencia registrada en centros saturados, en determinados días a determinadas horas, hacia otros que en esos mismos momentos pueden presentar exceso de capacidad para atender exclusivamente su propia demanda. Estos desequilibrios pueden depender, entre otros factores, de la localización geográfica y del contexto socioeconómico de los centros y, en cualquier caso, debería tenerse en cuenta que la disparidad entre la capacidad de las estructuras y el número de visitas determina inevitablemente una asistencia precaria o, *en el mejor de los casos*, supone un despilfarro de recursos.

## **5. Demanda horaria de energía eléctrica en España**

Dadas las características técnicas particulares de la energía eléctrica, que la hacen diferente de otros bienes industriales y servicios, cualquier estudio sobre este mercado debe tener en cuenta la naturaleza del producto, las restricciones técnicas que conlleva su producción y distribución y sus consecuencias sobre la estructura productiva del sector, así como el grado de regulación del mismo.

La estructura de la industria de energía eléctrica está compuesta por tres segmentos distintos: generación, transmisión y distribución. La electricidad se obtiene aplicando energía mecánica a un generador mediante diferentes tecnologías<sup>1</sup>. La energía generada debe ser transmitida desde los centros de producción a los centros de consumo del mercado<sup>2</sup>. Por último, en la fase de distribución la energía transportada a los centros de consumo se acerca a los consumidores individuales, geográficamente más o menos dispersos<sup>3</sup>. Todos estos componentes del sistema eléctrico son altamente interdependientes, de forma que dicho sistema sólo puede operar eficientemente si existe la necesaria coordinación, tanto a corto como a largo plazo, de forma que la energía demandada pueda ser satisfecha en todo momento con las características de calidad adecuadas.

Si se tiene en cuenta que la electricidad no es almacenable, el equilibrio del sistema eléctrico requiere que la suma de las demandas de energía se iguale a la energía generada ofertada menos las pérdidas de transporte en todos los puntos de consumo. El grado de utilización de la capacidad de las plantas debe ser mayor que el estrictamente necesario para satisfacer la demanda esperada porque esta última puede aumentar

---

<sup>1</sup> Según la tecnología generadora de energía eléctrica, se puede distinguir entre plantas hidroeléctricas, con turbinas de gas, con generación térmica —por ejemplo, las plantas nucleares—, y plantas de combustible de origen fósil, como las que utilizan carbón (Rodríguez y Castro, 1994).

<sup>2</sup> El sistema de transporte de una determinada empresa está generalmente interconectado con los de otras empresas y con los de los países limítrofes, con objeto de posibilitar intercambios de energía que puedan mejorar económicamente la explotación del sistema, así como garantizar el suministro.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

rápidamente en cualquier momento o porque alguno de los generadores puede sufrir una indisponibilidad fortuita que conduzca a su desconexión brusca. La necesidad de un relativo exceso de oferta se justifica además por la imposibilidad de reajustar la producción, con la suficiente celeridad, poniendo en marcha centrales que están paradas, aun cuando los tiempos de respuesta sean muy distintos según el tipo de central<sup>4</sup>. Más específicamente, es necesaria una *reserva giratoria*, es decir, un exceso de capacidad de generación en disposición de incrementar la producción efectiva en cualquier momento. En este sentido, la magnitud del sistema eléctrico en su conjunto constituye un elemento clave para la eficiencia y calidad del servicio.

Una gran dimensión del sistema eléctrico supone una interesante ventaja económica, pero el incremento de tal dimensión puede lograrse sin que las empresas sean individualmente más grandes, sino a través de la interconexión entre ellas, que permite reducir las inversiones en potencia instalada. Además, los grandes sistemas interconectados tienen la posibilidad de reducir los costes variables de generación por medio del intercambio de energía eléctrica entre las empresas que forman parte de ellos, y rebajar los costes asociados a las incertidumbres aleatorias de demanda<sup>5</sup>.

Los condicionantes técnicos enumerados implican que la industria y el mercado eléctrico tengan características muy peculiares desde el punto de vista económico<sup>6</sup>. El

---

<sup>3</sup> En esta fase, es frecuente separar la mera función de transporte de la de suministro o abastecimiento, que comprende todas las actividades relacionadas con la venta y facturación de electricidad a los usuarios finales —contratación, lectura, asesoramiento al cliente, facturación, cobro.

<sup>4</sup> En particular, las centrales térmicas requieren tiempos de arranque de varias horas, mientras que en el caso de las hidráulicas estos tiempos son de minutos; aunque en ningún caso son lo suficientemente cortos como para evitar un colapso, al menos parcial, del sistema eléctrico.

<sup>5</sup> En este sentido, el informe «El mercado eléctrico español ante el 2003. Garantía de Suministro y Eficiencia», del Club Español de la Energía y la Asociación Empresarial Eléctrica UNESA, justifica la necesidad de acometer un nuevo ciclo inversor dada la disminución del margen de reserva en la capacidad de generación de energía eléctrica en España de 1.3 a 1.05 entre 1995 y 2001.

<sup>6</sup> En primer lugar, el hecho de que las empresas eléctricas tengan que realizar inversiones con un alto grado de especificidad, de forma que los costes asumidos puedan considerarse como irrecuperables, junto al hecho de tener que proveer electricidad a todos los usuarios de un país con una tarifa común, independientemente de la localización geográfica de los mismos, conduce a una fuerte integración vertical. Esto también puede generar barreras de entrada a nuevos competidores. En segundo lugar, dadas sus características técnicas, las actividades de transmisión y distribución gozan de las características de monopolio natural, ya que existen sustanciales economías de escala asociadas con el uso de líneas de alta tensión empleadas en la fase de transmisión. Además, es importante señalar las notables pérdidas de eficiencia que surgirían si las redes son duplicadas. En tercer lugar, por lo que respecta al sistema de generación existe una situación de oligopolio donde, si no se toma ninguna medida, pueden surgir problemas de colusión y comportamiento estratégico. En cuarto lugar, la obligación de garantizar el suministro por su naturaleza de servicio público, la necesidad de un equilibrio continuo entre oferta y demanda y las externalidades medioambientales asociadas con los principales combustibles utilizados en la generación, hacen imprescindible la regulación del sector. Por último, cabe señalar que el sector eléctrico produce importantes efectos socioeconómicos. No debe ignorarse la dependencia de todos los sectores productivos respecto a los sectores energéticos en general y del eléctrico en particular (Sanz y Calvo, 1997, realizan un análisis, para el conjunto de la economía

sector eléctrico español funciona como un sistema integrado en lo referente a la toma de decisiones a corto y largo plazo. Las necesidades de capacidad de generación se definen en el Plan Energético Nacional (PEN). El diseño y la capacidad de la red de transporte son determinados por la Red Eléctrica de España (REE)<sup>7</sup>. Por su parte, y con objeto de lograr una explotación unificada de los medios de generación disponibles que garanticen el suministro incurriendo en el mínimo coste, la propia REE gestiona un *pool* centralizado de energía y adopta las decisiones correspondientes en lo que se denomina *despacho centralizado*. En concreto, se simula un mercado de competencia perfecta en generación en el que la oferta está constituida por la capacidad potencial de generación de cada una de las unidades del sistema, ordenadas de forma creciente por su coste marginal de funcionamiento, y la demanda refleja las necesidades del conjunto de clientes conectados al sistema. En cada momento horario, el centro de explotación asigna a cada central el volumen de energía que ésta debe generar para cubrir la demanda existente, computando el número de unidades que componen la oferta y el coste marginal de las mismas.

La existencia de este proceso de explotación unificada con una estructura en la que conviven diferentes empresas integradas con generación y distribución, y una empresa (Endesa) especializada únicamente en generación, exige la realización de intercambios de energía, de tal forma que, en cada uno de los tramos horarios, la producción programada para las centrales de cada una de las empresas, más o menos la energía adquirida o cedida, coincida con la demanda. Para ello, cada hora se forma un *pool* con la energía sobrante de las empresas excedentarias, es decir, aquéllas a las que la programación ha asignado una producción en sus centrales que supera los niveles de energía necesarios para cubrir la demanda de su mercado. Y las empresas deficitarias obtienen del *pool* la energía que precisan para cubrir su mercado<sup>8</sup>.

---

española, de los requerimientos eléctricos directos e indirectos de cada sector). Así, determinadas medidas sobre el sector eléctrico han sido utilizadas con frecuencia como instrumentos de aplicación de ciertas políticas económicas tales como el control del nivel de precios o la protección de determinados sectores industriales en declive, favoreciendo el uso de ciertos combustibles como, por ejemplo, el carbón. Además, la determinación de la tarifa eléctrica puede afectar significativamente a la distribución de la renta.

<sup>7</sup> Constituida por todas las grandes empresas del sector, es la responsable de la transmisión de la energía eléctrica.

<sup>8</sup> Para computar estos intercambios, a los excedentes proporcionados por las centrales con exceso de generación, se les asigna el coste marginal de sus energías programadas más caras. Por otro lado, la energía obtenida del *pool* por las centrales deficitarias se computa al precio del *pool*, que coincidirá con el precio medio de las energías entregadas por las empresas excedentarias. Posteriormente, dado que pueden existir empresas con energías cuyo coste variable sea superior al medio del *pool*, se supone que dichas empresas entregan a éste dichas energías a ese precio y la recompran al nuevo precio medio del

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

La normativa que desarrolla todos los aspectos de la retribución de las empresas eléctricas se encuentra recogida en el Marco Legal Estable (MLE), fijado por R.D. 1538/1987, que regula el entorno económico en el que las empresas deben desarrollar su actividad, estableciendo el procedimiento de determinación de los ingresos de estas empresas. Concretamente, se conforma la metodología de cálculo del precio medio de venta de la energía eléctrica y la del sistema de compensaciones interempresas. Respecto al precio medio de venta, la premisa fundamental del MLE es que cubra exactamente el coste de servicio. Este coste de servicio se obtiene agregando todos los costes reconocidos, para todos los elementos y todas las empresas, calculándose unos costes para todo el sistema, y dividiéndose por la demanda prevista. Este procedimiento asegura que, para el nivel de demanda previsto, los ingresos del conjunto del sistema cubren los costes estándar en los que incurre el sector para la prestación del servicio.

El proceso de determinación del precio medio de venta explica por sí solo el interés en la previsión de la demanda. Pero son, sobre todo, las rigideces de la oferta para adaptarse a cambios súbitos en la demanda las que justifican también la conveniencia de anticipar la demanda, de modo que la capacidad de generación y las redes de transporte y distribución puedan coordinarse y conseguir el mantenimiento continuo e instantáneo de la igualdad global entre la demanda y la oferta del sistema. Ahora bien, dada la imposibilidad de almacenamiento de la oferta y teniendo en cuenta, además, el carácter estocástico de la demanda y las notables variaciones que ésta experimenta según la época del año (calefacción en invierno, aire acondicionado en verano, ...), según el día de la semana (días laborables, días festivos) y, por supuesto, según la hora del día (horas laborables, llegada de la noche, horas de madrugada), parece interesante examinar las variaciones de la demanda de energía eléctrica utilizando datos que, temporalmente, estén lo más desagregados posible. En definitiva, aumentar el grado de eficiencia técnica y distributiva, tanto en términos estáticos como dinámicos, requiere un conocimiento de la estacionalidad de la demanda y, como ya se ha manifestado, el ámbito temporal horario resulta el más apropiado para este análisis. De hecho, quizás el carácter marcadamente estacional de la demanda horaria de energía eléctrica sea uno de los factores que explique los problemas de suministro en horas punta a los que aluden los medios de comunicación en determinadas épocas del año y

---

*pool* resultante. Este proceso iterativo finaliza cuando ninguna empresa utiliza, para la cobertura de su demanda, energía con un coste variable superior al del *pool*.

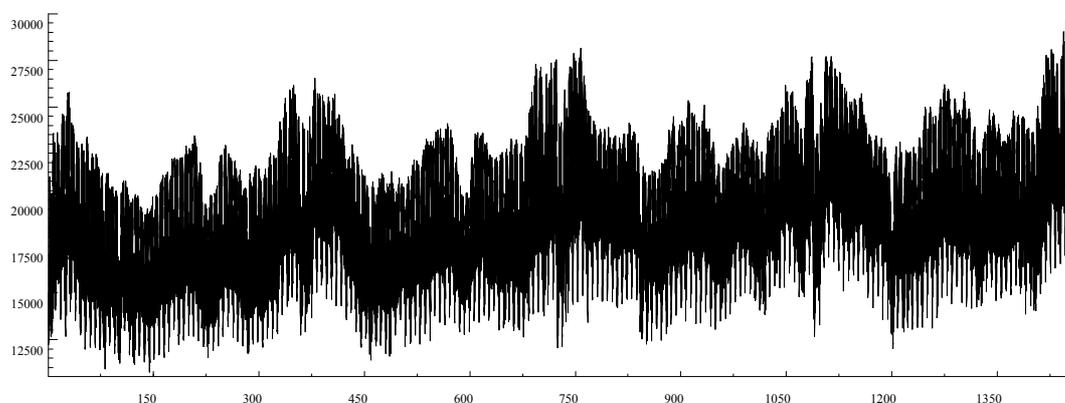
zonas geográficas<sup>9</sup>. El impacto de la estacionalidad se pone también de manifiesto en estudios como el realizado por parte del Ministerio de Economía con objeto de evaluar la rentabilidad de determinadas tecnologías de ahorro energético en el sector hotelero según zonas geográficas, dada la evidente influencia del clima así como del grado y estacionalidad de la demanda de alojamiento sobre los consumos<sup>10</sup>.

### 5.1 Descripción de la demanda horaria de energía eléctrica

En este capítulo se estudia la serie de demanda de energía eléctrica en España, medida en Mwh, en el periodo comprendido entre la hora uno del jueves uno de enero de 1998 y la hora veinticuatro del lunes treinta y uno de diciembre de 2001<sup>11</sup>. Los datos de la serie<sup>12</sup>, en adelante denotada por  $\{y_t\}_{t=1, \dots, 35064}$ , se representan en el gráfico 5.1.1, que también contiene la serie de medias móviles de periodo 168, obtenidas con objeto de captar mejor la evolución registrada en cada uno de los cuatro años al margen de fluctuaciones semanales o diarias.

**Gráfico 5.1.1**

**Demanda horaria de energía eléctrica en España,  $\{y_t\}$**



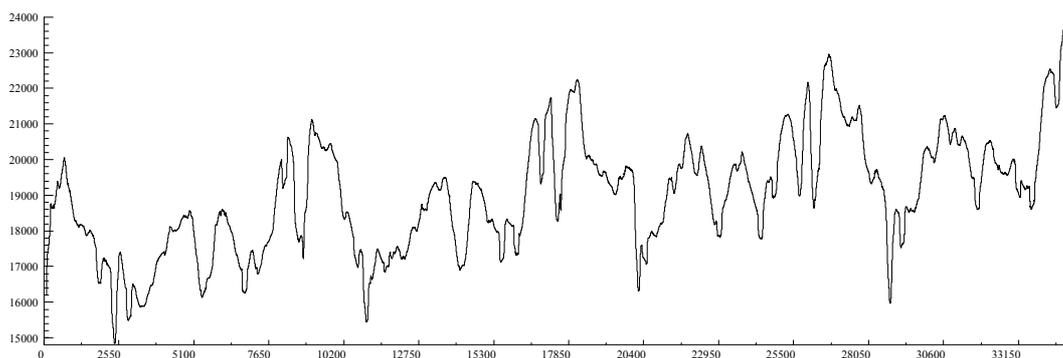
<sup>9</sup> Véase *El País Negocios*, 21 de julio de 2002:2. En la prensa regional de Canarias (*Diario de Avisos*, 26 de julio de 2002:27), la Asociación Hotelera y Extrahotelera de Tenerife, La Palma, La Gomera y el Hierro (ASHOTEL) pide a los hoteleros del Sur de la isla de Tenerife que desconecten el aire acondicionado entre las diecinueve y las veintidós horas para ahorrar energía y evitar cortes en el suministro.

<sup>10</sup> Véase IDAE (2001).

<sup>11</sup> Los datos utilizados corresponden a la información pública proporcionada por el Operador del Mercado de Electricidad (OMEL) y disponible en <http://www.omel.com>, donde también se definen las magnitudes asociadas a la demanda de energía eléctrica. En concreto, el concepto de demanda manejado incluye las siguientes unidades de adquisición: distribuidoras, comercializadoras para consumidores nacionales y para exportación, consumidores cualificados, agentes externos para exportación y contratos de exportación de la REE. Se consideran también las unidades de bombeo en consumo.

<sup>12</sup> Véase el fichero horario.xls del disco.

### Medias móviles de periodo 168 de $\{v_t\}$



Como consecuencia del cambio de hora que cada año se efectúa en los meses de marzo (se adelanta la hora) y octubre (se atrasa la hora), la fuente oficial de datos registra la demanda de las veintitrés horas de los días veintinueve de marzo de 1998, veintiocho de marzo de 1999, veintiséis de marzo de 2000 y veinticinco de marzo de 2001; mientras que en los días veinticinco de octubre de 1998, treinta y uno de octubre de 1999, veintinueve de octubre de 2000 y veintiocho de octubre de 2001 se registran datos correspondientes a veinticinco horas. Para evitar la heterogeneidad que ocasiona el cambio de hora oficial, dicho cambio no ha sido tenido en cuenta a la hora de asignar los datos disponibles a las horas del día correspondientes<sup>13</sup>. A pesar de la transformación realizada en los datos, cabe pensar que el cambio de hora oficial tiene un efecto sobre la demanda. De hecho, ésta es la hipótesis que justifica que el cambio de hora se produzca, buscando un supuesto ahorro energético. Dicho efecto será tenido en cuenta en la especificación posterior del modelo que se propondrá para recoger las fluctuaciones de la demanda en el periodo estudiado.

Para facilitar la identificación de los datos en cualquiera de las etapas del análisis que posteriormente será llevado a cabo, conviene señalar que las 35064 observaciones horarias registradas entre la hora uno del jueves uno de enero de 1998 y la hora veinticuatro del lunes treinta y uno de diciembre de 2001 se distribuyen de la siguiente forma. Las observaciones 1 a 8760, 8761 a 17520, 17521 a 26304 y 26305 a 35064

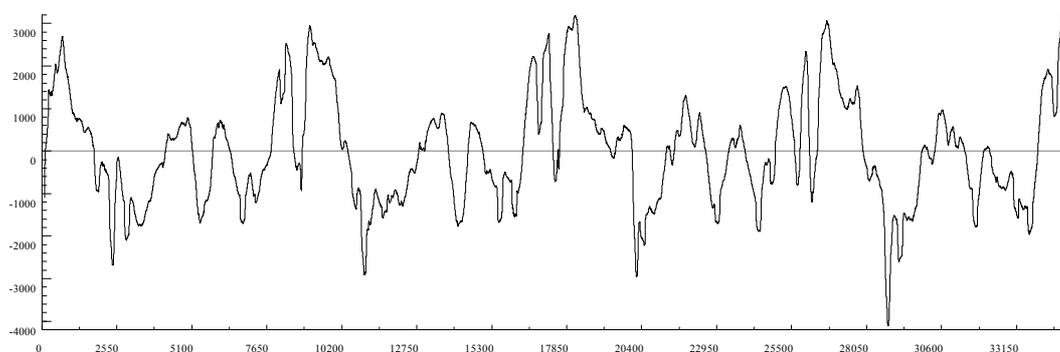
<sup>13</sup> En concreto, cuando la hora cambia en un día de marzo, se supone que a las dos de la mañana el reloj se adelanta a las tres, de forma que ese día sólo se registran veintitrés horas. Pues bien, esas veintitrés horas y la primera hora del día siguiente al del cambio de hora se han considerado las veinticuatro horas correspondientes al día del cambio. Si el cambio de hora tiene lugar en un día de octubre, se supone que a las tres de la mañana el reloj se atrasa a las dos, de modo que en ese día se registran veinticinco horas. En este caso, la hora uno se considera la hora veinticuatro del día anterior y sólo las veinticuatro últimas horas registradas se asignan al día del cambio. Así, se dispone finalmente de una serie que registra demandas correspondientes a veinticuatro horas para cada uno de los días del periodo estudiado.

corresponden, respectivamente, a los años 1998, 1999, 2000<sup>14</sup> y 2001; y equivalen a 1461 días o, definiendo la semana de lunes a domingo, a 4 días, 208 semanas y 1 día.

La representación gráfica de la serie muestra la gran inestabilidad de la demanda horaria de energía eléctrica y la presencia de un cierto crecimiento a largo plazo; si bien, este comportamiento puede estar oscurecido por la presencia de un componente estacional dominante en el que cabe pensar que coexisten, al menos, tres fluctuaciones de periodos anual, semanal y diario, que deberían ser tenidas en cuenta a la hora de describir la evolución de esta magnitud. Por tanto, y a modo de hipótesis preliminar, podría considerarse que existe un componente tendencial y un componente estacional integrado por tres fluctuaciones superpuestas de distinto periodo. Con objeto de identificar la presencia de estas fluctuaciones se ha seguido el proceso que se describe a continuación.

En primer lugar, y con el fin de evitar que el crecimiento de la serie distorsione la observación de las fluctuaciones estacionales, se procedió a extraer, previamente, el componente tendencial<sup>15</sup>. Una vez realizada esta transformación, el cálculo de medias móviles de periodo 168,  $\{mm(168)_t\}_{t=85,\dots,34980}$ , permite eliminar, o cuando menos suavizar, las variaciones de periodos 24 y 168. El gráfico de la serie de medias móviles correspondientes a cada año permite apreciar la existencia de un patrón estacional regular en cada uno de los cuatro años (gráfico 5.1.2) y, en consecuencia, asumir que esta variación estacional posee carácter determinista. Además, el reducido número de años disponible impediría detectar una evolución estocástica suave.

**Gráfico 5.1.2. Medias móviles de periodo 168,  $\{mm(168)_t\}$**



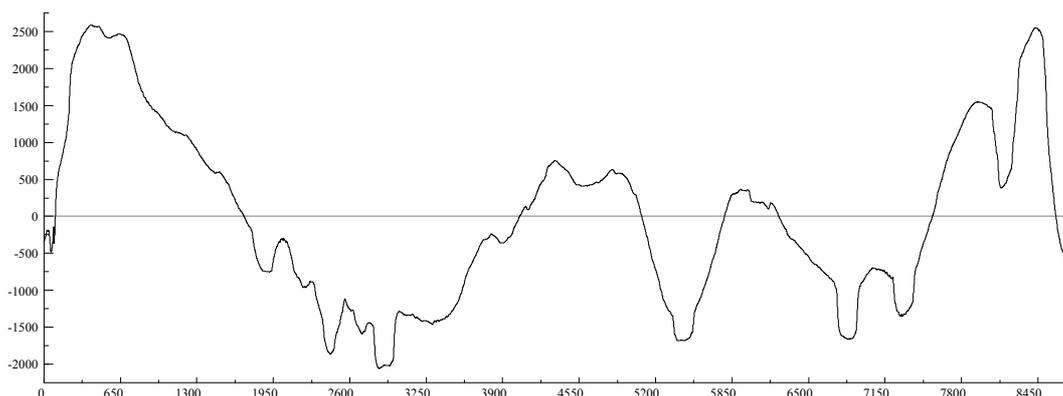
<sup>14</sup> El año 2000 es bisiesto y, por tanto, contiene 8784 observaciones. En concreto, los veinticuatro datos comprendidos entre las observaciones 18937 y 18960, ambos inclusive, corresponden al veintinueve de febrero de 2000.

<sup>15</sup> El procedimiento de extracción de este componente se detalla en el epígrafe siguiente.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

El cálculo de las medias por hora del año a partir de la serie de medias móviles de periodo 168, que se denotarán por  $\{\hat{g}_{1,h^a}^{anual}\}_{h^a=1,\dots,8760}$  (gráfico 5.1.3)<sup>16</sup>, permite advertir que, tras los reducidos niveles de demanda de los primeros días de enero, el patrón estacional de periodo anual se caracteriza por un fuerte crecimiento hasta mediados de dicho mes y experimenta a continuación una lenta y prolongada caída hasta finales del mes de abril; se produce entonces una recuperación suave hasta finales de julio y una brusca caída que conduce a niveles mínimos a mediados del mes de agosto; después del verano se inicia una fase de fluctuaciones más acentuadas con crecimiento en septiembre, descenso hasta mediados de octubre, máximos locales a finales de noviembre y mediados de diciembre y mínimos locales a principios y finales del mes de diciembre, coincidiendo con las fiestas navideñas<sup>17</sup>.

**Gráfico 5.1.3. Medias por hora del año,  $\{\hat{g}_{1,h^a}^{anual}\}$**



Si a los valores de la serie original filtrada de componente tendencial se le sustrae una aproximación de la variación estacional anual<sup>18</sup> y, después de calcular medias móviles de periodo 24 con objeto de eliminar la fluctuación diaria, se representan los valores de la serie resultante en cada semana,  $\{mm(24)_t\}_{t=25,\dots,35040}$  (gráfico 5.1.4), se detecta la presencia de un patrón más o menos regular. Este comportamiento se observa con menos claridad si se representan los datos de esta

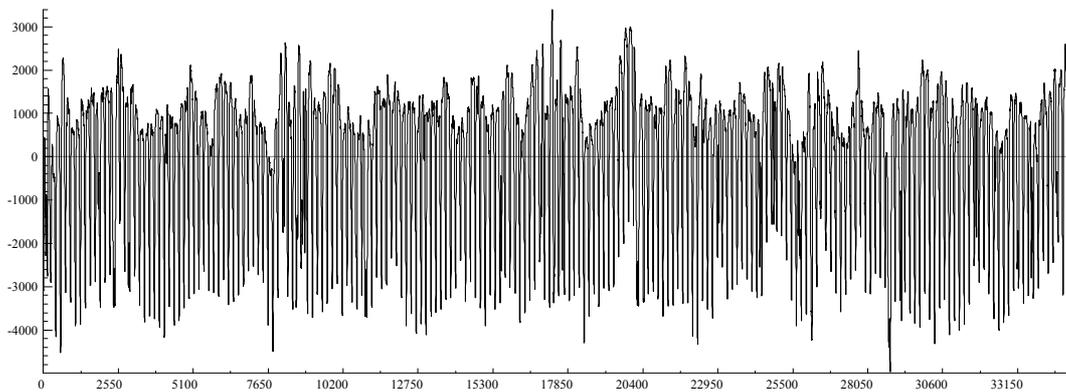
<sup>16</sup> Para el cálculo de estas medias, se han suprimido las veinticuatro observaciones correspondientes al veintinueve de febrero de 2000. Las medias se han corregido de forma que la suma de efectos estacionales a lo largo de las 8760 horas del año fuese nula.

<sup>17</sup> Los datos utilizados resultan de la agregación de demandas registradas en las diferentes zonas del territorio nacional, que presentan diferentes patrones estacionales a lo largo del año. Ello explica que el patrón obtenido en este trabajo no se corresponda con el propio de ninguna área geográfica concreta.

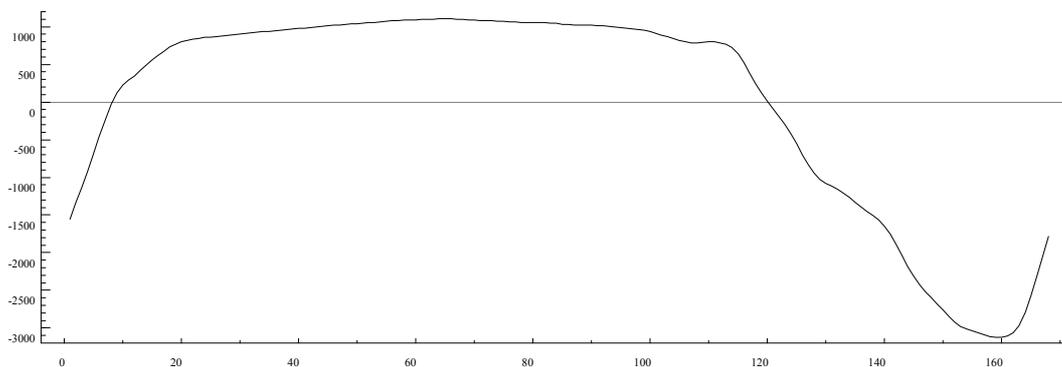
<sup>18</sup> En el epígrafe siguiente se obtiene una estimación de la fluctuación estacional de periodo anual a partir de una función *spline*.

misma serie correspondientes a cada estación (anexo 5.1). A pesar de ello, se ha optado por asumir, *a priori*, que la fluctuación estacional de periodo 168 posee carácter determinista, aunque posteriormente se lleve a cabo un análisis confirmatorio de la veracidad de este supuesto. Las medias por hora de la semana, que se denotarán por  $\{\hat{g}_{1,h^s}^{semanal}\}_{h^s=1,\dots,168}$ , calculadas a partir de la serie de medias móviles de periodo 24 señalada anteriormente (gráfico 5.1.5), reflejan, como cabía esperar, que la demanda de energía eléctrica es mínima durante los fines de semana y máxima a lo largo de los días laborables. Después del crecimiento continuo apreciado a lo largo del lunes, se registran las demandas más altas y estables, y en la tarde de los viernes se inicia el descenso que culmina con el mínimo absoluto localizado en la tarde de los domingos.

**Gráfico 5.1.4. Medias móviles de periodo 24,  $\{mm(24)_i\}$**



**Gráfico 5.1.5. Medias por hora de la semana,  $\{\hat{g}_{1,h^s}^{semanal}\}$**

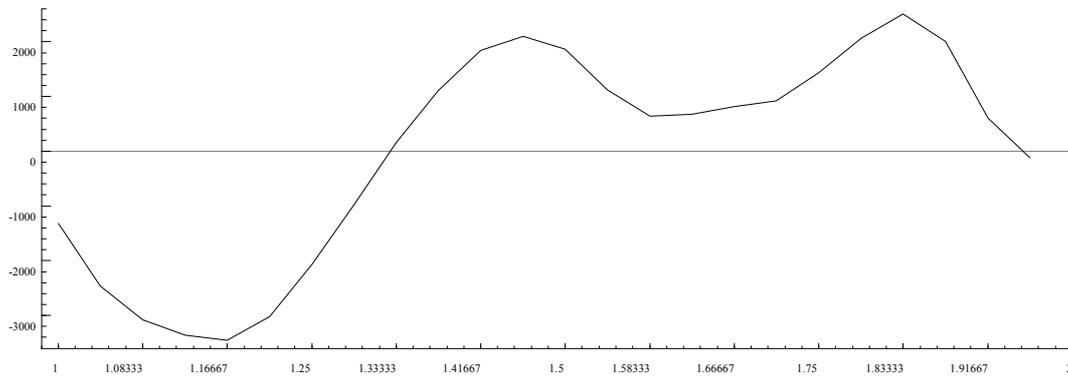


Finalmente, en cuanto a la fluctuación estacional de periodo 24, es más difícil observar si el patrón estacional cambia o permanece relativamente estable, aunque los gráficos de la evolución por hora del día de la serie original una vez filtrada de

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

tendencia y de las fluctuaciones estacionales de periodo anual y semanal estimadas parecen indicar la variabilidad de esta fluctuación (anexo 5.2). Por tanto, se asume, en principio, que posee un carácter estocástico y que sea el proceso de estimación el que conduzca a rechazar esa hipótesis, si fuera el caso. De todos modos, y con objeto de caracterizar este comportamiento, se ha procedido al cálculo de las medias por hora del día sobre la serie filtrada anteriormente comentada. La serie resultante se denotará por  $\{\hat{g}_{1,h^d}^{diaria}\}_{h^d=1,\dots,24}$  (gráfico 5.1.6)<sup>19</sup>. A partir del gráfico de estas medias, puede concluirse que la demanda de energía eléctrica es más baja durante la madrugada, situándose los mínimos en torno a las cuatro o cinco de la mañana; mientras que los niveles de demanda crecen conforme avanza la jornada laborable, registrándose un máximo alrededor del medio día y un descenso durante el periodo de la tarde en el que se interrumpe dicha jornada; el máximo absoluto tiene lugar en torno a las veintiuna horas, lo que parece estar relacionado con la desaparición de la luz solar<sup>20</sup>.

**Gráfico 5.1.6. Medias por hora del día,  $\{\hat{g}_{1,h^d}^{diaria}\}$**



La variabilidad observada en algunos de los componentes de la serie aconseja recurrir a los modelos estructurales como herramienta estadística para recoger estos comportamientos. Concretamente, los comentarios anteriores sugieren especificar el modelo siguiente:

$$y_t = m_t + g_t^{diaria} + g_t^{semanal} + g_t^{anual} + e_t, \quad e_t \sim NID(0, \sigma_e^2),$$

<sup>19</sup> En el epígrafe siguiente se muestra la especificación del modelo a partir del cual se han obtenido estas medias por hora del día. En dicho modelo, las fluctuaciones estacionales de periodo anual y semanal se han incorporado a través de las correspondientes funciones *splines*.

<sup>20</sup> Nótese que es difícil asumir que este comportamiento sea estable a lo largo del año, y más bien puede pensarse que el máximo se registre durante una hora más temprana durante el periodo de invierno. Esta circunstancia podría ayudar a confirmar que la fluctuación estacional de periodo diario posee un

donde  $\mathbf{m}_t$  representa un componente de nivel estocástico,  $\mathbf{g}_t^{diaria}$  es la fluctuación de periodo diario según la formulación estocástica trigonométrica y  $\mathbf{g}_t^{semanal}$  y  $\mathbf{g}_t^{anual}$  representan las variaciones de periodo semanal y anual, respectivamente, que se han modelado mediante las correspondientes funciones *splines*. Con el fin de medir el efecto del cambio de hora oficial, podría añadirse en la especificación del modelo anterior una variable *dummy* escalón que tome valor uno durante el periodo de marzo a octubre adecuado en cada año.

## 5.2. Especificación y estimación del modelo estructural

De acuerdo con lo comentado en el apartado anterior, la evolución de la demanda horaria de energía eléctrica podría recogerse mediante el modelo que se especifica a continuación:

$$y_t = \mathbf{m}_t + \mathbf{g}_t^{diaria} + \mathbf{g}_t^{semanal} + \mathbf{g}_t^{anual} + \mathbf{d}CH_t + \mathbf{e}_t, \quad \mathbf{e}_t \sim NID(0, \sigma_e^2),$$

donde

$$CH_t = \begin{cases} 1, & \text{si} \begin{cases} 2091 \leq t \leq 7130 \\ 10827 \leq t \leq 16034 \\ 19563 \leq t \leq 24770 \\ 28299 \leq t \leq 33506 \end{cases} \\ 0, & \text{en el resto} \end{cases}$$

asumiendo que el cambio de hora oficial se produce entre las dos y las tres de la mañana.

La estimación del modelo anterior no es, precisamente, una tarea sencilla desde el punto de vista computacional. Además de los problemas derivados de la gran longitud de la serie y del número de parámetros que interviene en la formulación del componente estacional definido, los programas informáticos disponibles no recogen la posibilidad de una especificación como la que en este caso parece apropiada y, mucho menos, permiten la estimación conjunta de dos o más fluctuaciones estacionales de carácter estocástico. Por ello, se asume, en principio, que  $\{\mathbf{g}_t^{semanal}\}$  y  $\{\mathbf{g}_t^{anual}\}$  poseen naturaleza determinística. Ahora bien, la longitud del periodo asociado a estas dos fluctuaciones estacionales impide la estimación de un modelo en el que ambas fueran recogidas a través de variables cualitativas. De ahí que, como se ha indicado, se recurra

---

carácter estocástico, si bien parte de esta variabilidad podría ser captada a través de la fluctuación de periodo anual.

a la formulación de estas variaciones en términos de funciones *splines*. Sin embargo, la especificación de dichas funciones exige aislar cada una de las fluctuaciones estacionales presentes en la serie. A continuación se describe el proceso de extracción de cada una de ellas.

Dado el crecimiento observado en la serie, que quizás pueda ser modelado mediante un nivel estocástico, con o sin término específico de pendiente, conviene evitar que dicho crecimiento distorsione la extracción de las fluctuaciones estacionales consideradas. Con tal fin, es apropiado eliminar previamente, como ya se ha comentado, el componente tendencial. Ahora bien, la formulación de un nivel estocástico podría absorber parte de la variabilidad aparejada a las variaciones estacionales. De este modo, se optó por estimar un modelo con tendencial lineal determinística, es decir,

$$y_t = \mathbf{a} + \mathbf{b}t + v_t$$

y extraer las correspondientes fluctuaciones estacionales a partir del residuo estimado del modelo anterior<sup>21</sup>,  $\{\hat{v}_t\}_{t=1,\dots,35064}$ .

### 5.2.1. Extracción de la fluctuación de periodo anual, $\{\mathbf{g}_t^{anual}\}$

Considerando la naturaleza determinística de la fluctuación estacional de periodo anual y teniendo en cuenta que el posible efecto del cambio de hora oficial será recogido a través de la magnitud del componente estacional en las horas del año afectadas por dicho cambio, una primera aproximación al componente  $\mathbf{g}_t^{anual}$  puede obtenerse calculando: 1) la serie de medias móviles de periodo 168 del residuo antes comentado —que, presumiblemente, anulará los efectos estacionales de periodo 168 y, también, aquellos otros de periodo 24—, denotada por  $\{mm(168)_t\}_{t=1,\dots,35064}$ ; 2) la serie de medias por hora del año<sup>22</sup>; 3) la serie de medias por hora del año corregida de forma que la suma de las variaciones estacionales a lo largo del año sea nula. La serie así obtenida se denotará por  $\{\hat{\mathbf{g}}_{1,t}^{anual}\} = \{\hat{\mathbf{g}}_{1,h^a}^{anual}\}_{h^a=1,\dots,8760}$ , si la observación en el instante  $t$  corresponde a la hora del año  $h^a$  (gráfico 5.1.3)<sup>23</sup>.

<sup>21</sup> Las estimaciones mínimo cuadráticas de los parámetros de este modelo fueron  $\hat{\mathbf{a}} = 17295.0962$  y  $\hat{\mathbf{b}} = 0.09711965$ .

<sup>22</sup> En el cálculo de estas medias, se han suprimido las observaciones correspondientes al veintinueve de febrero de 2000.

<sup>23</sup> Por supuesto, estas medias constituyen una aproximación a la variación estacional de periodo anual correspondiente a un año no bisiesto. En un año bisiesto, es preciso incorporar supuestos adicionales que serán explicados más adelante.

## Demanda horaria de energía eléctrica en España

El componente  $\hat{g}_{1,h^a}^{anual}$  será modelado a través de la función *spline*  $g^a(h^a)$ , definida como  $g^a(h^a) = g_i^a(h^a)$ , donde  $g_i^a(h^a) = g_{i,0}^a + g_{i,1}^a h^a + g_{i,2}^a (h^a)^2 + g_{i,3}^a (h^a)^3$ , con  $h_{i-1}^a \leq h^a \leq h_i^a$ , para  $i = 1, \dots, 13$ ,  $h^a = 1, \dots, 8760$ , y  $h_0^a = 1$  y  $h_{13}^a = 8760$ . Esta *spline* debe verificar las siguientes condiciones:

a) restricciones de continuidad de la función,

$$g_{i,0}^a + g_{i,1}^a h_i^a + g_{i,2}^a (h_i^a)^2 + g_{i,3}^a (h_i^a)^3 = g_{i+1,0}^a + g_{i+1,1}^a h_i^a + g_{i+1,2}^a (h_i^a)^2 + g_{i+1,3}^a (h_i^a)^3,$$

para  $i = 1, \dots, 12$ , y

$$g_{13,0}^a + g_{13,1}^a (h_{13}^a + 1) + g_{13,2}^a (h_{13}^a + 1)^2 + g_{13,3}^a (h_{13}^a + 1)^3 = g_{1,0}^a + g_{1,1}^a h_0^a + g_{1,2}^a (h_0^a)^2 + g_{1,3}^a (h_0^a)^3;$$

b) restricciones de continuidad de la derivada primera,

$$g_{i,1}^a + 2g_{i,2}^a h_i^a + 3g_{i,3}^a (h_i^a)^2 = g_{i+1,1}^a + 2g_{i+1,2}^a h_i^a + 3g_{i+1,3}^a (h_i^a)^2,$$

para  $i = 1, \dots, 12$ , y

$$g_{13,1}^a + 2g_{13,2}^a (h_{13}^a + 1) + 3g_{13,3}^a (h_{13}^a + 1)^2 = g_{1,1}^a + 2g_{1,2}^a h_0^a + 3g_{1,3}^a (h_0^a)^2;$$

c) restricciones de continuidad de la derivada segunda,

$$2g_{i,2}^a + 6g_{i,3}^a h_i^a = 2g_{i+1,2}^a + 6g_{i+1,3}^a h_i^a,$$

para  $i = 1, \dots, 12$ , y

$$2g_{13,2}^a + 6g_{13,3}^a (h_{13}^a + 1) = 2g_{1,2}^a + 6g_{1,3}^a h_0^a.$$

Dado que la *spline* anual es de trece tramos y se decidió dejar como parámetros libres  $g_{i,0}^a$ ,  $i = 1, \dots, 13$ , entonces, los parámetros del vector  $G_{39 \times 1}^a = (g_{1,1}^a, g_{1,2}^a, g_{1,3}^a, \dots, g_{13,1}^a, g_{13,2}^a, g_{13,3}^a)'$  pueden expresarse como

$$g_r = a_{r,1}(-g_{1,0}^a + g_{2,0}^a) + a_{r,2}(-g_{2,0}^a + g_{3,0}^a) + \dots + a_{r,12}(-g_{12,0}^a + g_{13,0}^a) + a_{r,13}(g_{1,0}^a - g_{13,0}^a),$$

para  $r = 1, \dots, 39$ , siendo  $g_1^a = g_{1,1}^a$ ,  $g_2^a = g_{1,2}^a$ ,  $g_3^a = g_{1,3}^a$ ,  $\dots$ ,  $g_{37}^a = g_{13,1}^a$ ,  $g_{38}^a = g_{13,2}^a$  y  $g_{39}^a = g_{13,3}^a$ , y siendo  $a_{j,k}^a$  el elemento de la fila  $j$  y la columna  $k$  de la inversa de la matriz  $H^a$ , definida como

Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$H_{39 \times 39}^a = \begin{bmatrix} h_1^a & (h_1^a)^2 & (h_1^a)^3 & -h_1^a & -(h_1^a)^2 & -(h_1^a)^3 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & h_{12}^a & (h_{12}^a)^2 & (h_{12}^a)^3 & -h_{12}^a & -(h_{12}^a)^2 & -(h_{12}^a)^3 \\ -h_0^a & -(h_0^a)^2 & -(h_0^a)^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & (h_{13}^a + 1) & (h_{13}^a + 1)^2 & (h_{13}^a + 1)^3 \\ 1 & 2h_1^a & 3(h_1^a)^2 & -1 & -2h_1^a & -3(h_1^a)^2 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 2h_{12}^a & 3(h_{12}^a)^2 & -1 & -2h_{12}^a & -3(h_{12}^a)^2 \\ -1 & -2h_0^a & -3(h_0^a)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 2(h_{13}^a + 1) & 3(h_{13}^a + 1)^2 \\ 0 & 2 & 6h_1^a & 0 & -2 & -6h_1^a & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 2 & 6h_{12}^a & 0 & -2 & -6h_{12}^a \\ 0 & -2 & -6h_0^a & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 2 & 6(h_{13}^a + 1) \end{bmatrix}$$

Es decir,

$$g_r^a = \mathbf{a}_{r,1}^a g_{1,0}^a + \mathbf{a}_{r,2}^a g_{2,0}^a + \dots + \mathbf{a}_{r,12}^a g_{12,0}^a + \mathbf{a}_{r,13}^a g_{13,0}^a,$$

para  $r = 1, \dots, 39$ , donde

$$\mathbf{a}_{r,1}^a = -a_{r,1}^a + a_{r,13}^a, \mathbf{a}_{r,2}^a = a_{r,1}^a - a_{r,2}^a, \dots, \mathbf{a}_{r,12}^a = a_{r,11}^a - a_{r,12}^a \text{ y } \mathbf{a}_{r,13}^a = a_{r,12}^a - a_{r,13}^a,$$

para  $r = 1, \dots, 39$ . Y entonces,

$$\begin{aligned} g^a(h^a) &= [g_{1,0}^a + g_{1,1}^a h^a + g_{1,2}^a (h^a)^2 + g_{1,3}^a (h^a)^3] D_{1,h^a} + \\ &\quad [g_{2,0}^a + g_{2,1}^a h^a + g_{2,2}^a (h^a)^2 + g_{2,3}^a (h^a)^3] D_{2,h^a} + \dots + \\ &\quad [g_{13,0}^a + g_{13,1}^a h^a + g_{13,2}^a (h^a)^2 + g_{13,3}^a (h^a)^3] D_{13,h^a} \\ &= g_{1,0}^a X_{1,h^a} + g_{2,0}^a X_{2,h^a} + \dots + g_{13,0}^a X_{13,h^a} \end{aligned}$$

$$\text{donde } D_{i,h^a} = \begin{cases} 1, & \text{si } h_{i-1}^a \leq h^a < h_i^a, \quad i = 1, 2, \dots, 12, \text{ y } D_{13,h^a} = \begin{cases} 1, & \text{si } h_{12}^a \leq h^a \leq h_{13}^a, \\ 0, & \text{resto} \end{cases} \end{cases}, \text{ y}$$

$$\begin{aligned} X_{i,h^a} &= D_{i,h^a} + (\acute{a}_{1,i}^a h^a + \acute{a}_{2,i}^a (h^a)^2 + \acute{a}_{3,i}^a (h^a)^3) D_{1,h^a} + \\ &\quad (\acute{a}_{4,i}^a h^a + \acute{a}_{5,i}^a (h^a)^2 + \acute{a}_{6,i}^a (h^a)^3) D_{2,h^a} + \dots + \\ &\quad (\acute{a}_{37,i}^a h^a + \acute{a}_{38,i}^a (h^a)^2 + \acute{a}_{39,i}^a (h^a)^3) D_{13,h^a} \end{aligned}$$

para  $i = 1, \dots, 13$ .

Con la *spline*  $g^a(h^a)$  definida en términos de los regresores anteriores, puede estimarse el modelo  $\hat{g}_{1,h^a}^{anual} = g^a(h^a) + \mathbf{e}_{h^a}$ , donde  $\mathbf{e}_{h^a}$  es un término de error asociado al ajuste, y obtener otra aproximación al componente estacional de periodo anual, que se denotará por  $\hat{g}_{2,h^a}^{anual} = \hat{g}^a(h^a)$ ,  $h^a = 1, \dots, 8760$ .

La localización de los puntos de ruptura  $h_i^a$ ,  $i = 1, \dots, 12$ , que minimiza la suma de cuadrados de residuos en la estimación del modelo  $\hat{g}_{1,h^a}^{anual} = g^a(h^a) + \mathbf{e}_{h^a}$  fue la siguiente:  $h_1^a = 323$ ,  $h_2^a = 468$ ,  $h_3^a = 3292$ ,  $h_4^a = 5354$ ,  $h_5^a = 5403$ ,  $h_6^a = 5536$ ,  $h_7^a = 7476$ ,  $h_8^a = 8141$ ,  $h_9^a = 8142$ ,  $h_{10}^a = 8255$ ,  $h_{11}^a = 8529$  y  $h_{12}^a = 8530$ . Sin embargo, la contigüidad de los valores  $h_8^a$  y  $h_9^a$ , así como  $h_{11}^a$  y  $h_{12}^a$ , originaba problemas de proximidad a la multicolinealidad que no podían ser resueltos de manera que pudiera estimarse el modelo estructural con los regresores necesarios para recoger la *spline*. En este sentido, y teniendo en cuenta que no se puede esperar localizar cada nodo de forma

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

óptima, sino más bien obtener una distribución óptima de los nodos<sup>24</sup>, se optó por separar, a la distancia necesaria para la estimación, las posiciones de los puntos antes señalados, de modo que en la combinación final se consideró que  $h_8^a = 8135$ ,  $h_9^a = 8145$ ,  $h_{11}^a = 8525$  y  $h_{12}^a = 8535$ <sup>25</sup>.

Ahora bien, la *spline* anterior recoge la variación estacional de periodo anual en años no bisiestos, pero es preciso evaluar también este componente para el caso de años bisiestos. Para ello, si se denota por  $h^a$  a la hora  $h^a$ -ésima de un año no bisiesto,  $h^a = 1, \dots, 8760$ , y por  $h^{ab}$  a la hora  $h^{ab}$ -ésima de un año bisiesto,  $h^{ab} = 1, \dots, 8784$ , la *spline* anual correspondiente a un año bisiesto,  $g^{ab}$ , puede definirse como

$$g^{ab}(h^{ab}) = \begin{cases} g^a(h^{ab}) & , h^{ab} = 1, \dots, 1416 \\ g^{29F}(h^{ab}) & , h^{ab} = 1417, \dots, 1440 \\ g^a(h^{ab} - 24) & , h^{ab} = 1441, \dots, 8784 \end{cases}$$

siendo  $g^a$  la *spline* antes definida para un año no bisiesto y  $g^{29F}$  una función polinómica específica para recoger la evolución de la demanda en las horas correspondientes al veintinueve de febrero de 2000.

Entonces, la cuestión radica en determinar  $g^{29F}(h^{ab})$ ,  $h^{ab} = 1417, \dots, 1440$ . El problema puede resolverse especificando  $g^{29F}(h^{ab})$  como

$$g^{29F}(h^{ab}) = g_0^{29F} + g_1^{29F} h^{ab} + g_2^{29F} (h^{ab})^2 + g_3^{29F} (h^{ab})^3,$$

donde  $h_1^{ab} \leq h^{ab} \leq h_2^{ab}$ , con  $h_1^{ab} = 1417$  y  $h_2^{ab} = 1440$ .

Para garantizar la continuidad de la *spline*, parece conveniente imponer restricciones de continuidad con la hora veinticuatro del veintiocho de febrero ( $h^a = 1416$ ,  $h^{ab} = 1416$ ) y con la hora uno del uno de marzo ( $h^a = 1417$ ,  $h^{ab} = 1441$ ). Es decir,

$$g_0^{29F} + g_1^{29F} (h_1^{ab} - 1) + g_2^{29F} (h_1^{ab} - 1)^2 + g_3^{29F} (h_1^{ab} - 1)^3 = g^a(1416)$$

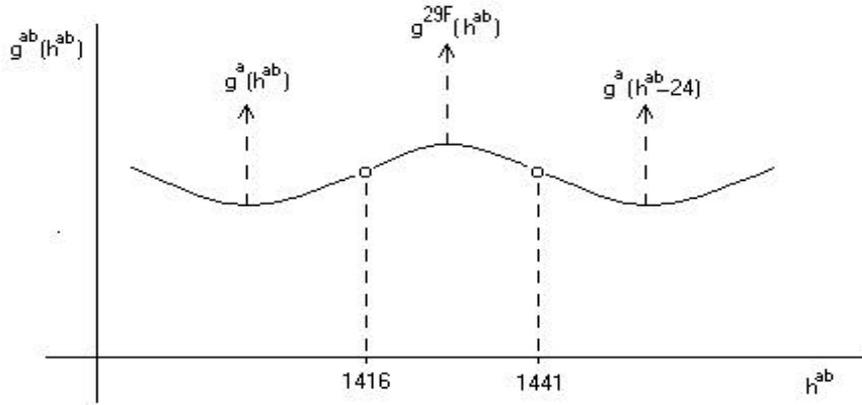
y

$$g_0^{29F} + g_1^{29F} (h_2^{ab} + 1) + g_2^{29F} (h_2^{ab} + 1)^2 + g_3^{29F} (h_2^{ab} + 1)^3 = g^a(1417)$$

<sup>24</sup> Como señala de Boor (1978:181), en estas circunstancias es posible tomarse ciertas libertades con la localización de nodos individuales con objeto de hacer más fácil el análisis.

<sup>25</sup> Las estimaciones de los parámetros libres en la especificación de la *spline* anual fueron las siguientes:  $\hat{g}_{1,0}^a = -557.8963$ ,  $\hat{g}_{2,0}^a = -6930.6272$ ,  $\hat{g}_{3,0}^a = 3088.8095$ ,  $\hat{g}_{4,0}^a = 55648.1559$ ,  $\hat{g}_{5,0}^a = -85535485.35$ ,

Gráficamente,



En la *spline* anual de un año no bisiesto,  $h^a = 1416$  y  $h^a = 1417$  pertenecen al tramo tercero. Por tanto,

$$g^a(1416) = g_{3,0}^a + g_{3,1}^a(h_1^{ab} - 1) + g_{3,2}^a(h_1^{ab} - 1)^2 + g_{3,3}^a(h_1^{ab} - 1)^3$$

y

$$g^a(1417) = g_{3,0}^a + g_{3,1}^a h_1^{ab} + g_{3,2}^a(h_1^{ab})^2 + g_{3,3}^a(h_1^{ab})^3$$

Así pues, se tiene que

$$g_0^{29F} + g_1^{29F}(h_1^{ab} - 1) + g_2^{29F}(h_1^{ab} - 1)^2 + g_3^{29F}(h_1^{ab} - 1)^3 = g_{3,0}^a + g_{3,1}^a(h_1^{ab} - 1) + g_{3,2}^a(h_1^{ab} - 1)^2 + g_{3,3}^a(h_1^{ab} - 1)^3$$

y

$$g_0^{29F} + g_1^{29F}(h_2^{ab} + 1) + g_2^{29F}(h_2^{ab} + 1)^2 + g_3^{29F}(h_2^{ab} + 1)^3 = g_{3,0}^a + g_{3,1}^a h_1^{ab} + g_{3,2}^a(h_1^{ab})^2 + g_{3,3}^a(h_1^{ab})^3$$

Además,

$$g_{3,1}^a = \mathbf{a}_{7,1}^a g_{1,0}^a + \mathbf{a}_{7,2}^a g_{2,0}^a + \dots + \mathbf{a}_{7,12}^a g_{12,0}^a + \mathbf{a}_{7,13}^a g_{13,0}^a,$$

$$g_{3,2}^a = \mathbf{a}_{8,1}^a g_{1,0}^a + \mathbf{a}_{8,2}^a g_{2,0}^a + \dots + \mathbf{a}_{8,12}^a g_{12,0}^a + \mathbf{a}_{8,13}^a g_{13,0}^a$$

y

$$g_{3,3}^a = \mathbf{a}_{9,1}^a g_{1,0}^a + \mathbf{a}_{9,2}^a g_{2,0}^a + \dots + \mathbf{a}_{9,12}^a g_{12,0}^a + \mathbf{a}_{9,13}^a g_{13,0}^a.$$

Por tanto,

$$\begin{aligned} & g_0^{29F} + g_1^{29F}(h_1^{ab} - 1) + g_2^{29F}(h_1^{ab} - 1)^2 + g_3^{29F}(h_1^{ab} - 1)^3 = \\ & g_{3,0}^a + (h_1^{ab} - 1) \left[ \hat{a}_{7,1}^a g_{1,0}^a + \hat{a}_{7,2}^a g_{2,0}^a + \dots + \hat{a}_{7,12}^a g_{12,0}^a + \hat{a}_{7,13}^a g_{13,0}^a \right] \\ & + (h_1^{ab} - 1)^2 \left[ \hat{a}_{8,1}^a g_{1,0}^a + \hat{a}_{8,2}^a g_{2,0}^a + \dots + \hat{a}_{8,12}^a g_{12,0}^a + \hat{a}_{8,13}^a g_{13,0}^a \right] \\ & + (h_1^{ab} - 1)^3 \left[ \hat{a}_{9,1}^a g_{1,0}^a + \hat{a}_{9,2}^a g_{2,0}^a + \dots + \hat{a}_{9,12}^a g_{12,0}^a + \hat{a}_{9,13}^a g_{13,0}^a \right] \end{aligned}$$

---


$$\hat{g}_{6,0}^a = 35701652.84, \quad \hat{g}_{7,0}^a = -749665.4254, \quad \hat{g}_{8,0}^a = 8409667.917, \quad \hat{g}_{9,0}^a = -4545843143,$$

y

$$\begin{aligned} & g_0^{29F} + g_1^{29F} (h_2^{ab} + 1) + g_2^{29F} (h_2^{ab} + 1)^2 + g_3^{29F} (h_2^{ab} + 1)^3 = \\ & g_{3,0}^a + h_1^{ab} \left[ \hat{a}_{7,1}^a g_{1,0}^a + \hat{a}_{7,2}^a g_{2,0}^a + \dots + \hat{a}_{7,12}^a g_{12,0}^a + \hat{a}_{7,13}^a g_{13,0}^a \right] \\ & + (h_1^{ab})^2 \left[ \hat{a}_{8,1}^a g_{1,0}^a + \hat{a}_{8,2}^a g_{2,0}^a + \dots + \hat{a}_{8,12}^a g_{12,0}^a + \hat{a}_{8,13}^a g_{13,0}^a \right] \\ & + (h_1^{ab})^3 \left[ \hat{a}_{9,1}^a g_{1,0}^a + \hat{a}_{9,2}^a g_{2,0}^a + \dots + \hat{a}_{9,12}^a g_{12,0}^a + \hat{a}_{9,13}^a g_{13,0}^a \right] \end{aligned}$$

Si se imponen estas restricciones de continuidad, quedan libres dos de los cuatro parámetros de la función  $g^{29F}(h^{ab})$ . Por ejemplo, los parámetros del vector  $G_{2 \times 1}^{29F} = (g_2^{29F}, g_3^{29F})'$  pueden expresarse como función de  $g_0^{29F}$  y  $g_1^{29F}$ . En concreto, se tiene que  $H^{29F} G^{29F} = R^{29F}$ , donde

$$H^{29F} = \begin{bmatrix} (h_1^{ab} - 1)^2 & (h_1^{ab} - 1)^3 \\ (h_2^{ab} + 1)^2 & (h_2^{ab} + 1)^3 \end{bmatrix}$$

y

$$R^{29F} = \begin{bmatrix} r_1^{29F} \\ r_2^{29F} \end{bmatrix},$$

siendo

$$\begin{aligned} r_1^{29F} &= -g_0^{29F} - g_1^{29F} (h_1^{ab} - 1) + g_{3,0}^a + \\ & (h_1^{ab} - 1) \left[ \hat{a}_{7,1}^a g_{1,0}^a + \hat{a}_{7,2}^a g_{2,0}^a + \dots + \hat{a}_{7,12}^a g_{12,0}^a + \hat{a}_{7,13}^a g_{13,0}^a \right] + \\ & (h_1^{ab} - 1)^2 \left[ \hat{a}_{8,1}^a g_{1,0}^a + \hat{a}_{8,2}^a g_{2,0}^a + \dots + \hat{a}_{8,12}^a g_{12,0}^a + \hat{a}_{8,13}^a g_{13,0}^a \right] + \\ & (h_1^{ab} - 1)^3 \left[ \hat{a}_{9,1}^a g_{1,0}^a + \hat{a}_{9,2}^a g_{2,0}^a + \dots + \hat{a}_{9,12}^a g_{12,0}^a + \hat{a}_{9,13}^a g_{13,0}^a \right] + \\ & = -g_0^{29F} - g_1^{29F} (h_1^{ab} - 1) + \\ & \left[ (h_1^{ab} - 1) \hat{a}_{7,1}^a + (h_1^{ab} - 1)^2 \hat{a}_{8,1}^a + (h_1^{ab} - 1)^3 \hat{a}_{9,1}^a \right] g_{1,0}^a + \dots + \\ & \left[ 1 + (h_1^{ab} - 1) \hat{a}_{7,3}^a + (h_1^{ab} - 1)^2 \hat{a}_{8,3}^a + (h_1^{ab} - 1)^3 \hat{a}_{9,3}^a \right] g_{3,0}^a + \dots + \\ & \left[ (h_1^{ab} - 1) \hat{a}_{7,13}^a + (h_1^{ab} - 1)^2 \hat{a}_{8,13}^a + (h_1^{ab} - 1)^3 \hat{a}_{9,13}^a \right] g_{13,0}^a \\ & = -g_0^{29F} - g_1^{29F} (h_1^{ab} - 1) + \ddot{e}_{1,1} g_{1,0}^a + \dots + \ddot{e}_{1,3} g_{3,0}^a + \dots + \ddot{e}_{1,13} g_{13,0}^a \end{aligned}$$

y

---


$$\hat{g}_{10,0}^a = 405507886.8, \hat{g}_{11,0}^a = 71473078.83, \hat{g}_{12,0}^a = -4071218211 \text{ y } \hat{g}_{13,0}^a = 34862484.98.$$

$$\begin{aligned}
 r_2^{29F} &= -g_0^{29F} - g_1^{29F} (h_2^{ab} + 1) + g_{3,0}^a + \\
 &\quad h_1^{ab} \left[ \dot{a}_{7,1}^a g_{1,0}^a + \dot{a}_{7,2}^a g_{2,0}^a + \dots + \dot{a}_{7,12}^a g_{12,0}^a + \dot{a}_{7,13}^a g_{13,0}^a \right] + \\
 &\quad (h_1^{ab})^2 \left[ \dot{a}_{8,1}^a g_{1,0}^a + \dot{a}_{8,2}^a g_{2,0}^a + \dots + \dot{a}_{8,12}^a g_{12,0}^a + \dot{a}_{8,13}^a g_{13,0}^a \right] + \\
 &\quad (h_1^{ab})^3 \left[ \dot{a}_{9,1}^a g_{1,0}^a + \dot{a}_{9,2}^a g_{2,0}^a + \dots + \dot{a}_{9,12}^a g_{12,0}^a + \dot{a}_{9,13}^a g_{13,0}^a \right] \\
 &= -g_0^{29F} - g_1^{29F} (h_2^{ab} + 1) + \\
 &\quad \left[ h_1^{ab} \dot{a}_{7,1}^a + (h_1^{ab})^2 \dot{a}_{8,1}^a + (h_1^{ab})^3 \dot{a}_{9,1}^a \right] g_{1,0}^a + \dots + \\
 &\quad \left[ 1 + h_1^{ab} \dot{a}_{7,3}^a + (h_1^{ab})^2 \dot{a}_{8,3}^a + (h_1^{ab})^3 \dot{a}_{9,3}^a \right] g_{3,0}^a + \dots + \\
 &\quad \left[ h_1^{ab} \dot{a}_{7,13}^a + (h_1^{ab})^2 \dot{a}_{8,13}^a + (h_1^{ab})^3 \dot{a}_{9,13}^a \right] g_{13,0}^a \\
 &= -g_0^{29F} - g_1^{29F} (h_2^{ab} + 1) + \ddot{e}_{2,1} g_{1,0}^a + \dots + \ddot{e}_{2,3} g_{3,0}^a + \dots + \ddot{e}_{2,13} g_{13,0}^a
 \end{aligned}$$

De modo que  $G^{29F} = (H^{29F})^{-1} R^{29F}$ . Si se denota la inversa de la matriz  $H^{29F}$  por

$$(H^{29F})^{-1} = \begin{bmatrix} \mathbf{d}_{1,1} & \mathbf{d}_{1,2} \\ \mathbf{d}_{2,1} & \mathbf{d}_{2,2} \end{bmatrix},$$

resulta que

$$\begin{aligned}
 g_2^{29F} &= \ddot{a}_{1,1} r_1^{29F} + \ddot{a}_{1,2} r_2^{29F} \\
 &= -(\ddot{a}_{1,1} + \ddot{a}_{1,2}) g_0^{29F} - \left[ \ddot{a}_{1,1} (h_1^{ab} - 1) + \ddot{a}_{1,2} (h_2^{ab} + 1) \right] g_1^{29F} \\
 &\quad + (\ddot{a}_{1,1} \ddot{e}_{1,1} + \ddot{a}_{1,2} \ddot{e}_{2,1}) g_{1,0}^a + \dots + (\ddot{a}_{1,1} \ddot{e}_{1,13} + \ddot{a}_{1,2} \ddot{e}_{2,13}) g_{13,0}^a
 \end{aligned}$$

y

$$\begin{aligned}
 g_3^{29F} &= \ddot{a}_{2,1} r_1^{29F} + \ddot{a}_{2,2} r_2^{29F} \\
 &= -(\ddot{a}_{2,1} + \ddot{a}_{2,2}) g_0^{29F} - \left[ \ddot{a}_{2,1} (h_1^{ab} - 1) + \ddot{a}_{2,2} (h_2^{ab} + 1) \right] g_1^{29F} \\
 &\quad + (\ddot{a}_{2,1} \ddot{e}_{1,1} + \ddot{a}_{2,2} \ddot{e}_{2,1}) g_{1,0}^a + \dots + (\ddot{a}_{2,1} \ddot{e}_{1,13} + \ddot{a}_{2,2} \ddot{e}_{2,13}) g_{13,0}^a
 \end{aligned}$$

En definitiva, se tiene que

$$\begin{aligned}
 g^{29F} (h^{ab}) &= g_0^{29F} + g_1^{29F} h^{ab} + \\
 &\quad \left[ -(\ddot{a}_{1,1} + \ddot{a}_{1,2}) g_0^{29F} - \left[ \ddot{a}_{1,1} (h_1^{ab} - 1) + \ddot{a}_{1,2} (h_2^{ab} + 1) \right] g_1^{29F} \right] (h^{ab})^2 + \\
 &\quad \left[ (\ddot{a}_{1,1} \ddot{e}_{1,1} + \ddot{a}_{1,2} \ddot{e}_{2,1}) g_{1,0}^a + \dots + (\ddot{a}_{1,1} \ddot{e}_{1,13} + \ddot{a}_{1,2} \ddot{e}_{2,13}) g_{13,0}^a \right] \\
 &\quad \left[ -(\ddot{a}_{2,1} + \ddot{a}_{2,2}) g_0^{29F} - \left[ \ddot{a}_{2,1} (h_1^{ab} - 1) + \ddot{a}_{2,2} (h_2^{ab} + 1) \right] g_1^{29F} \right] (h^{ab})^3 \\
 &\quad \left[ (\ddot{a}_{2,1} \ddot{e}_{1,1} + \ddot{a}_{2,2} \ddot{e}_{2,1}) g_{1,0}^a + \dots + (\ddot{a}_{2,1} \ddot{e}_{1,13} + \ddot{a}_{2,2} \ddot{e}_{2,13}) g_{13,0}^a \right]
 \end{aligned}$$

o también,

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$\begin{aligned}
 g^{29F}(h^{ab}) &= \left[ 1 - (\ddot{a}_{1,1} + \ddot{a}_{1,2})(h^{ab})^2 - (\ddot{a}_{2,1} + \ddot{a}_{2,2})(h^{ab})^3 \right] g_0^{29F} + \\
 &\quad \left[ h^{ab} - ((h_1^{ab} - 1)\ddot{a}_{1,1} + (h_2^{ab} + 1)\ddot{a}_{1,2})(h^{ab})^2 \right. \\
 &\quad \left. - ((h_1^{ab} - 1)\ddot{a}_{2,1} + (h_2^{ab} + 1)\ddot{a}_{2,2})(h^{ab})^3 \right] g_1^{29F} + \\
 &\quad \left[ (\ddot{a}_{1,1}\ddot{e}_{1,1} + \ddot{a}_{1,2}\ddot{e}_{2,1})(h^{ab})^2 + (\ddot{a}_{2,1}\ddot{e}_{1,1} + \ddot{a}_{2,2}\ddot{e}_{2,1})(h^{ab})^3 \right] g_{1,0}^a + \dots + \\
 &\quad \left[ (\ddot{a}_{1,1}\ddot{e}_{1,13} + \ddot{a}_{1,2}\ddot{e}_{2,13})(h^{ab})^2 + (\ddot{a}_{2,1}\ddot{e}_{1,13} + \ddot{a}_{2,2}\ddot{e}_{2,13})(h^{ab})^3 \right] g_{13,0}^a \\
 &= g_{1,0}^a X_{1,h^{ab}}^{29F} + \dots + g_{13,0}^a X_{13,h^{ab}}^{29F} + g_0^{29F} X_{0(29F),h^{ab}}^{29F} + g_1^{29F} X_{1(29F),h^{ab}}^{29F}
 \end{aligned}$$

La *spline* así definida proporciona una aproximación a la variación estacional de periodo anual que puede denotarse como  $g_{3,t}^{anual} = g^a(h^a)$ , si la observación en el instante  $t$  corresponde a la hora  $h^a$  de un año no bisiesto, o bien,  $g_{3,t}^{anual} = g^{ab}(h^{ab})$ , si la observación en el instante  $t$  corresponde a la hora  $h^{ab}$  de un año bisiesto.

La fluctuación estacional obtenida de esta forma, puede incorporarse en el modelo estructural como función de un conjunto de variables explicativas, es decir,

$$g_{3,t}^{anual} = g_{1,0}^a X_{1,t}^a + \dots + g_{13,0}^a X_{13,t}^a + g_0^{29F} X_{0(29F),t}^{29F} + g_1^{29F} X_{1(29F),t}^{29F},$$

donde las variables  $X_{i,t}^a$ ,  $i = 1, \dots, 13$ ,  $X_{j(29F),t}^{29F}$ ,  $j = 0, 1$ , tienen la definición siguiente. Si la observación en el instante  $t$  corresponde a la hora  $h^a$  de un año no bisiesto,  $h^a = 1, \dots, 8760$ ,  $X_{i,t}^a = X_{i,h^a}^a$  y  $X_{j(29F),t}^{29F} = 0$ . Si la observación en el instante  $t$  corresponde a la hora  $h^{ab}$  de un año bisiesto,  $h^{ab} = 1, \dots, 8784$ ,

$$X_{i,t}^a = \begin{cases} X_{i,h^{ab}}^a & , \quad \text{si } h^{ab} = 1, \dots, 1416 \\ X_{i,h^{ab}}^{29F} & , \quad \text{si } h^{ab} = 1417, \dots, 1440 \\ X_{i,(h^{ab}-24)}^a & , \quad \text{si } h^{ab} = 1441, \dots, 8784 \end{cases}$$

y

$$X_{j(29F),t}^{29F} = \begin{cases} 0 & , \quad \text{si } h^{ab} = 1, \dots, 1416 \\ X_{j(29F),h^{ab}}^{29F} & , \quad \text{si } h^{ab} = 1417, \dots, 1440 \\ 0 & , \quad \text{si } h^{ab} = 1441, \dots, 8784 \end{cases}$$

Pero antes de estimar el modelo estructural, es preciso especificar la *spline* que recoge la variación estacional de periodo semanal. Para ello, conviene eliminar la fluctuación de periodo anual, que puede aproximarse a partir de la estimación de  $g^a(h^a)$ ,  $\hat{g}^a(h^a)$ , que se había denotado por  $\hat{g}_{2,t}^{anual} = \hat{g}^a(h^a)$  en el caso de un año no bisiesto. Pero, para el caso de un año bisiesto, no se ha obtenido una estimación  $\hat{g}^{ab}(h^{ab})$  de  $g^{ab}(h^{ab})$ , que ya se ha definido como

$$g^{ab}(h^{ab}) = \begin{cases} g^a(h^{ab}) & , \quad h^{ab} = 1, \dots, 1416 \\ g^{29F}(h^{ab}) & , \quad h^{ab} = 1417, \dots, 1440 \\ g^a(h^{ab} - 24), & h^{ab} = 1441, \dots, 8784 \end{cases}$$

que permita obtener  $\hat{g}_{2,t}^{anual}$ ,  $t = 1, \dots, 35064$ , tal que  $\hat{g}_{2,t}^{anual} = \hat{g}^a(h^a)$ , si la observación en el instante  $t$  corresponde a la hora  $h^a$  de un año no bisiesto, y  $\hat{g}_{2,t}^{anual} = \hat{g}^{ab}(h^{ab})$ , si la observación en el instante  $t$  corresponde a la hora  $h^{ab}$  de un año bisiesto<sup>26</sup>.

Si se denota por  $\hat{g}^a(1416)$  y  $\hat{g}^a(1417)$  a las estimaciones  $\hat{g}_{2,h^a}^{anual} = \hat{g}^a(h^a)$  en  $h^a = 1416$  y  $h^a = 1417$ , respectivamente, una estimación  $\hat{g}_{2,h^{ab}}^{anual} = \hat{g}^{ab}(h^{ab})$  puede obtenerse imponiendo las dos restricciones siguientes:

$$g_0^{29F} + g_1^{29F}(h_1^{ab} - 1) + g_2^{29F}(h_1^{ab} - 1)^2 + g_3^{29F}(h_1^{ab} - 1)^3 = \hat{g}^a(1416)$$

y

$$g_0^{29F} + g_1^{29F}(h_2^{ab} + 1) + g_2^{29F}(h_2^{ab} + 1)^2 + g_3^{29F}(h_2^{ab} + 1)^3 = \hat{g}^a(1417)$$

Se tiene ahora que  $H^{29F} G^{29F} = R^{29F*}$ , donde  $H^{29F}$  y  $G^{29F}$  están definidos como se

indicó anteriormente y  $R^{29F*} = \begin{bmatrix} r_1^{29F*} \\ r_2^{29F*} \end{bmatrix}$ , siendo

$$r_1^{29F*} = \hat{g}^a(1416) - g_0^{29F} - g_1^{29F}(h_1^{ab} - 1)$$

y

$$r_2^{29F*} = \hat{g}^a(1417) - g_0^{29F} - g_1^{29F}(h_2^{ab} + 1).$$

De modo que,

$$\begin{aligned} g_2^{29F} &= \delta_{1,1} r_1^{29F*} + d_{1,2} r_2^{29F*} \\ &= \delta_{1,1} [\hat{g}^a(1416) - g_0^{29F} - g_1^{29F}(h_1^{ab} - 1)] + \ddot{a}_{1,2} [\hat{g}^a(1417) - g_0^{29F} - g_1^{29F}(h_2^{ab} + 1)] \end{aligned}$$

y

$$\begin{aligned} g_3^{29F} &= \delta_{2,1} r_1^{29F*} + \ddot{a}_{2,2} r_2^{29F*} \\ &= \delta_{2,1} [\hat{g}^a(1416) - g_0^{29F} - g_1^{29F}(h_1^{ab} - 1)] + \ddot{a}_{2,2} [\hat{g}^a(1417) - g_0^{29F} - g_1^{29F}(h_2^{ab} + 1)] \end{aligned}$$

Entonces,

---

<sup>26</sup> Nótese que en el caso de un año bisiesto, la *spline* anual estimada no es directamente el componente estacional, ya que las variaciones estacionales deberían ser corregidas de forma que su suma a lo largo de ese año fuese nula.

$$\begin{aligned}
 g^{29F}(h^{ab}) &= g_0^{29F} + g_1^{29F} h^{ab} + \\
 &\quad \left[ \hat{g}^a(1416)\delta_{1,1} + \hat{g}^a(1417)\ddot{a}_{1,2} - g_0^{29F}(\ddot{a}_{1,1} + \ddot{a}_{1,2}) \right] (h^{ab})^2 + \\
 &\quad \left[ -g_1^{29F}(\ddot{a}_{1,1}(h_1^{ab} - 1) + \ddot{a}_{1,2}(h_2^{ab} + 1)) \right] \\
 &\quad \left[ \hat{g}^a(1416)\delta_{2,1} + \hat{g}^a(1417)\ddot{a}_{2,2} - g_0^{29F}(\ddot{a}_{2,1} + \ddot{a}_{2,2}) \right] (h^{ab})^3 \\
 &\quad \left[ -g_1^{29F}(\ddot{a}_{2,1}(h_1^{ab} - 1) + \ddot{a}_{2,2}(h_2^{ab} + 1)) \right] \\
 &= \left[ 1 - (\ddot{a}_{1,1} + \ddot{a}_{1,2})(h^{ab})^2 - (\ddot{a}_{2,1} + \ddot{a}_{2,2})(h^{ab})^3 \right] g_0^{29F} + \\
 &\quad \left[ h^{ab} - ((h_1^{ab} - 1)\ddot{a}_{1,1} + (h_2^{ab} + 1)\ddot{a}_{1,2})(h^{ab})^2 \right] g_1^{29F} + \\
 &\quad \left[ -((h_1^{ab} - 1)\ddot{a}_{2,1} + (h_2^{ab} + 1)\ddot{a}_{2,2})(h^{ab})^3 \right] \\
 &\quad \left[ \hat{g}^a(1416)\mathbf{d}_{1,1} + \hat{g}^a(1417)\mathbf{d}_{1,2} \right] (h^{ab})^2 + \left[ \hat{g}^a(1416)\mathbf{d}_{2,1} + \hat{g}^a(1417)\mathbf{d}_{2,2} \right] (h^{ab})^3 \\
 &= g_0^{29F} X_{0(29F),h^{ab}}^{29F} + g_1^{29F} X_{1(29F),h^{ab}}^{29F} + \\
 &\quad \left[ \hat{g}^a(1416)\mathbf{d}_{1,1} + \hat{g}^a(1417)\mathbf{d}_{1,2} \right] (h^{ab})^2 + \left[ \hat{g}^a(1416)\mathbf{d}_{2,1} + \hat{g}^a(1417)\mathbf{d}_{2,2} \right] (h^{ab})^3
 \end{aligned}$$

Y los valores de la *spline* anual en las observaciones correspondientes al veintinueve de febrero se pueden obtener estimando el modelo

$$F(h^{ab}) = g_0^{29F} X_{0(29F),h^{ab}}^{29F} + g_1^{29F} X_{1(29F),h^{ab}}^{29F} + \mathbf{e}_{h^{ab}},$$

donde

$$\begin{aligned}
 F(h^{ab}) &= mm(168)_{h^{ab}}^* - \left[ \hat{g}^a(1416)\mathbf{d}_{1,1} + \hat{g}^a(1417)\mathbf{d}_{1,2} \right] (h^{ab})^2 - \\
 &\quad \left[ \hat{g}^a(1416)\mathbf{d}_{2,1} + \hat{g}^a(1417)\mathbf{d}_{2,2} \right] (h^{ab})^3
 \end{aligned}$$

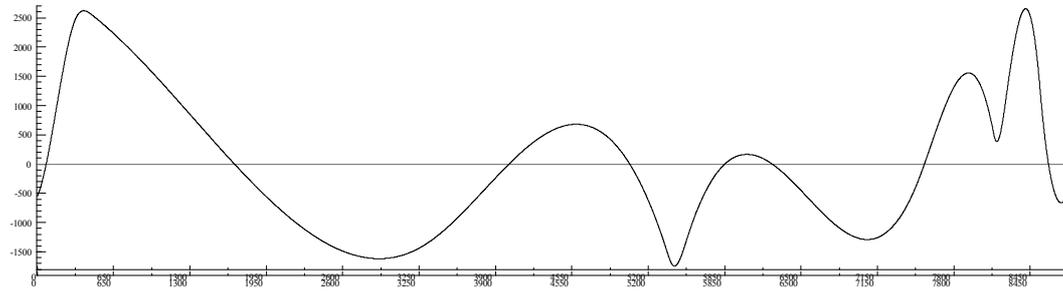
siendo  $mm(168)_{h^{ab}}^*$ ,  $h^{ab} = 1417, \dots, 1440$ , los valores de la serie  $\{mm(168)_t\}$  en los instantes del tiempo  $t$  correspondientes a las veinticuatro observaciones del veintinueve de febrero de 2000, corregidos por el mismo factor con el que se corrigieron las medias por hora del año para que su suma a lo largo de un año no bisiesto fuese nula<sup>27</sup>.

En el gráfico 5.2.1 se muestra la *spline* anual estimada para un año no bisiesto y para uno bisiesto. Para ilustrar con mayor claridad las condiciones de continuidad de la *spline* especificada para el veintinueve de febrero del año 2000, se representa también la *spline* estimada en un entorno de ese día.

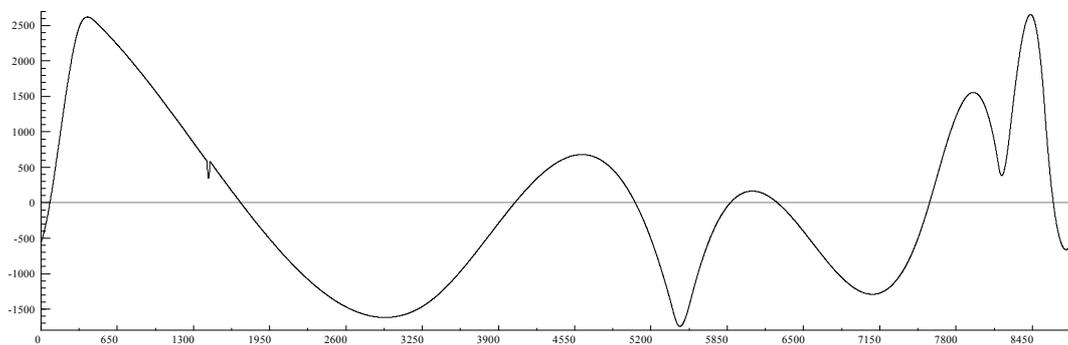
<sup>27</sup> Las estimaciones de los parámetros del modelo anterior fueron las siguientes:  $\hat{g}_0^{29F} = 12922874.72$  y  $\hat{g}_1^{29F} = -24998.54537$ .

Gráfico 5.2.1. Variación estacional de periodo anual estimada,  $\hat{g}_{2,h^a}^{anual}$  y  $\hat{g}_{2,h^{ab}}^{anual}$

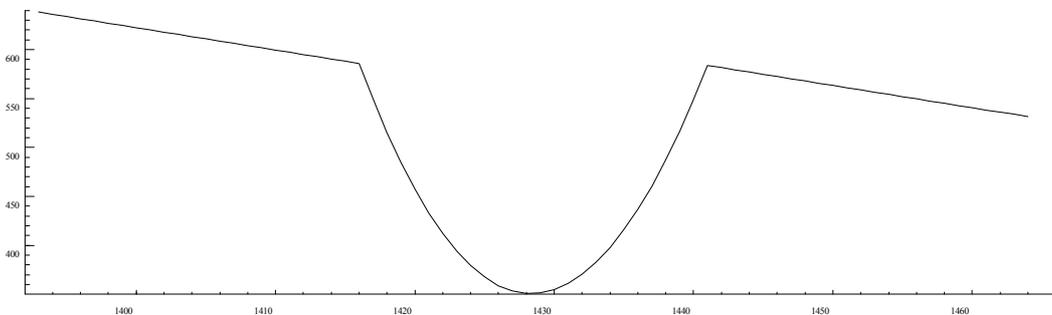
Año no bisiesto



Año bisiesto



29 de febrero



### 5.2.2 Extracción de la fluctuación de periodo semanal, $\{g_t^{semanal}\}$

Una vez obtenida la estimación  $\{g_{2,t}^{anual}\}_{t=1,\dots,35064}$ , puede obtenerse una primera aproximación al componente  $g_t^{semanal}$  calculando: 1) la serie de diferencias entre los valores de la serie original filtrada de una tendencia lineal determinística y los valores de la variación estacional de periodo anual estimada anteriormente, es decir,  $\{\hat{v}_t - \hat{g}_{2,t}^{anual}\}_{t=1,\dots,35064}$ ; 2) la serie de medias móviles de periodo veinticuatro, que, presumiblemente, anulará los efectos estacionales de periodo diario, y que se había denotado por  $\{mm(24)_t\}$ ; 3) la serie de medias por hora de la semana; 4) la serie de

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

medias por hora de la semana corregida de forma que la suma de las variaciones estacionales a lo largo de la semana sea nula. La serie así obtenida se denotará por  $\{\hat{\mathbf{g}}_{1,h^s}^{semanal}\}$ , tal que  $\{\hat{\mathbf{g}}_{1,t}^{semanal}\} = \{\hat{\mathbf{g}}_{1,h^s}^{semanal}\}_{h^s=1,\dots,168}$  si la observación en el instante  $t$  corresponde a la hora de la semana  $h^s$  (gráfico 5.1.5).

El componente  $\hat{\mathbf{g}}_{1,h^s}^{semanal}$  será modelado a través de la función *spline*  $g^s(h^s)$ , definida como  $g^s(h^s) = g_i^s(h^s)$ , donde  $g_i^s(h^s) = g_{i,0}^s + g_{i,1}^s h^s + g_{i,2}^s (h^s)^2 + g_{i,3}^s (h^s)^3$ , con  $h_{i-1}^s \leq h^s \leq h_i^s$ , para  $i = 1, \dots, 6$ ,  $h^s = 1, \dots, 168$ , y  $h_0^s = 1$  y  $h_6^s = 168$ . Esta *spline* debe verificar las siguientes condiciones:

a) restricciones de continuidad de la función,

$$g_{i,0}^s + g_{i,1}^s h_i^s + g_{i,2}^s (h_i^s)^2 + g_{i,3}^s (h_i^s)^3 = g_{i+1,0}^s + g_{i+1,1}^s h_i^s + g_{i+1,2}^s (h_i^s)^2 + g_{i+1,3}^s (h_i^s)^3,$$

para  $i = 1, \dots, 5$ , y

$$g_{6,0}^s + g_{6,1}^s (h_6^s + 1) + g_{6,2}^s (h_6^s + 1)^2 + g_{6,3}^s (h_6^s + 1)^3 = g_{1,0}^s + g_{1,1}^s h_0^s + g_{1,2}^s (h_0^s)^2 + g_{1,3}^s (h_0^s)^3;$$

b) restricciones de continuidad de la derivada primera,

$$g_{i,1}^s + 2g_{i,2}^s h_i^s + 3g_{i,3}^s (h_i^s)^2 = g_{i+1,1}^s + 2g_{i+1,2}^s h_i^s + 3g_{i+1,3}^s (h_i^s)^2,$$

para  $i = 1, \dots, 5$ , y

$$g_{6,1}^s + 2g_{6,2}^s (h_6^s + 1) + 3g_{6,3}^s (h_6^s + 1)^2 = g_{1,1}^s + 2g_{1,2}^s h_0^s + 3g_{1,3}^s (h_0^s)^2;$$

c) restricciones de continuidad de la derivada segunda,

$$2g_{i,2}^s + 6g_{i,3}^s h_i^s = 2g_{i+1,2}^s + 6g_{i+1,3}^s h_i^s,$$

para  $i = 1, \dots, 5$ , y

$$2g_{6,2}^s + 6g_{6,3}^s (h_6^s + 1) = 2g_{1,2}^s + 6g_{1,3}^s h_0^s.$$

Dado que la *spline* semanal es de seis tramos y se dejan como parámetros libres  $g_{i,0}^s$ ,  $i = 1, \dots, 6$ , entonces, los parámetros del vector

$G_{18 \times 1}^s = (g_{1,1}^s, g_{1,2}^s, g_{1,3}^s, \dots, g_{6,1}^s, g_{6,2}^s, g_{6,3}^s)'$  pueden expresarse como

$$g_r^s = a_{r,1}^s (-g_{1,0}^s + g_{2,0}^s) + a_{r,2}^s (-g_{2,0}^s + g_{3,0}^s) + \dots + a_{r,5}^s (-g_{5,0}^s + g_{6,0}^s) + a_{r,6}^s (g_{1,0}^s - g_{6,0}^s),$$

para  $r = 1, \dots, 18$ , siendo  $g_1^s = g_{1,1}^s$ ,  $g_2^s = g_{1,2}^s$ ,  $g_3^s = g_{1,3}^s$ ,  $\dots$ ,  $g_{16}^s = g_{6,1}^s$ ,  $g_{17}^s = g_{6,2}^s$  y

$g_{18}^s = g_{6,3}^s$ , y siendo  $a_{j,k}^s$  el elemento de la fila  $j$  y la columna  $k$  de la inversa de la matriz

$H^s$ , definida como

$$H_{18 \times 18}^s = \begin{bmatrix} h_1^s & (h_1^s)^2 & (h_1^s)^3 & -h_1^s & -(h_1^s)^2 & -(h_1^s)^3 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & h_5^s & (h_5^s)^2 & (h_5^s)^3 & -h_5^s & -(h_5^s)^2 & -(h_5^s)^3 \\ -h_0^s & -(h_0^s)^2 & -(h_0^s)^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & (h_6^s + 1) & (h_6^s + 1)^2 & (h_6^s + 1)^3 \\ 1 & 2h_1^s & 3(h_1^s)^2 & -1 & -2h_1^s & -3(h_1^s)^2 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 2h_5^s & 3(h_5^s)^2 & -1 & -2h_5^s & -3(h_5^s)^2 \\ -1 & -2h_0^s & -3(h_0^s)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 2(h_6^s + 1) & 3(h_6^s + 1)^2 \\ 0 & 2 & 6h_1^s & 0 & -2 & -6h_1^s & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 2 & 6h_5^s & 0 & -2 & -6h_5^s \\ 0 & -2 & -6h_0^s & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 2 & 6(h_6^s + 1) \end{bmatrix}$$

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

Es decir,

$$\mathbf{g}_r^s = \mathbf{a}_{r,1}^s \mathbf{g}_{1,0}^s + \mathbf{a}_{r,2}^s \mathbf{g}_{2,0}^s + \dots + \mathbf{a}_{r,5}^s \mathbf{g}_{5,0}^s + \mathbf{a}_{r,6}^s \mathbf{g}_{6,0}^s,$$

para  $r = 1, \dots, 18$ , donde

$$\mathbf{a}_{r,1}^s = -\mathbf{a}_{r,1}^s + \mathbf{a}_{r,6}^s, \mathbf{a}_{r,2}^s = \mathbf{a}_{r,1}^s - \mathbf{a}_{r,2}^s, \dots, \mathbf{a}_{r,5}^s = \mathbf{a}_{r,4}^s - \mathbf{a}_{r,5}^s \text{ y } \mathbf{a}_{r,6}^s = \mathbf{a}_{r,5}^s - \mathbf{a}_{r,6}^s,$$

para  $r = 1, \dots, 18$ . Y entonces,

$$\begin{aligned} g^s(h^s) &= \left[ \mathbf{g}_{1,0}^s + \mathbf{g}_{1,1}^s h^s + \mathbf{g}_{1,2}^s (h^s)^2 + \mathbf{g}_{1,3}^s (h^s)^3 \right] D_{1,h^s} + \\ &\quad \left[ \mathbf{g}_{2,0}^s + \mathbf{g}_{2,1}^s h^s + \mathbf{g}_{2,2}^s (h^s)^2 + \mathbf{g}_{2,3}^s (h^s)^3 \right] D_{2,h^s} + \dots + \\ &\quad \left[ \mathbf{g}_{6,0}^s + \mathbf{g}_{6,1}^s h^s + \mathbf{g}_{6,2}^s (h^s)^2 + \mathbf{g}_{6,3}^s (h^s)^3 \right] D_{6,h^s} \\ &= \mathbf{g}_{1,0}^s X_{1,h^s} + \mathbf{g}_{2,0}^s X_{2,h^s} + \dots + \mathbf{g}_{6,0}^s X_{6,h^s} \end{aligned}$$

$$\text{donde } D_{i,h^s} = \begin{cases} 1, & \text{si } h_{i-1}^s \leq h^s < h_i^s, \quad i = 1, \dots, 5, \text{ y } D_{6,h^s} = \begin{cases} 1, & \text{si } h_5^s \leq h^s \leq h_6^s, \\ 0, & \text{resto} \end{cases} \text{ y} \end{cases}$$

$$\begin{aligned} X_{i,h^s} &= D_{i,h^s} + \left( \mathbf{a}_{1,i}^s h^s + \mathbf{a}_{2,i}^s (h^s)^2 + \mathbf{a}_{3,i}^s (h^s)^3 \right) D_{1,h^s} + \\ &\quad \left( \mathbf{a}_{4,i}^s h^s + \mathbf{a}_{5,i}^s (h^s)^2 + \mathbf{a}_{6,i}^s (h^s)^3 \right) D_{2,h^s} + \dots + \\ &\quad \left( \mathbf{a}_{16,i}^s h^s + \mathbf{a}_{17,i}^s (h^s)^2 + \mathbf{a}_{18,i}^s (h^s)^3 \right) D_{6,h^s} \end{aligned}$$

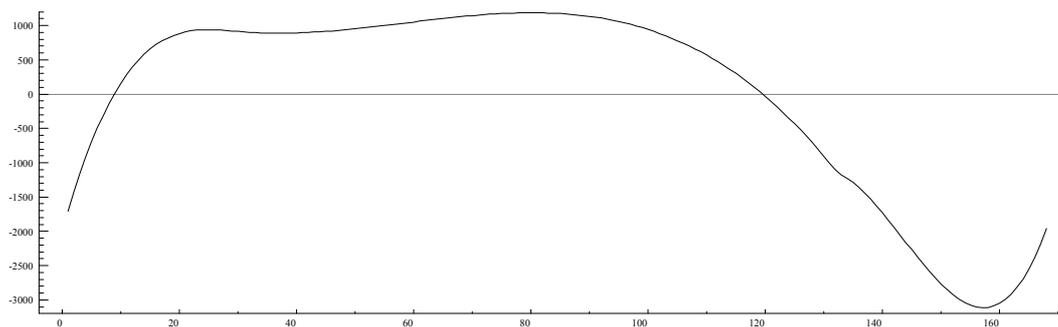
para  $i = 1, \dots, 6$ .

La localización de los puntos de ruptura,  $h_i^s$ ,  $i = 1, \dots, 5$ , que minimiza la suma de cuadrados de residuos en la estimación del modelo  $\hat{\mathbf{g}}_{1,h^s}^{\text{semanal}} = g^s(h^s) + \mathbf{e}_{h^s}$ , donde  $\mathbf{e}_{h^s}$  es un término de error asociado al ajuste, fue la siguiente:  $h_1^s = 2$ ,  $h_2^s = 31$ ,  $h_3^s = 131$ ,  $h_4^s = 133$  y  $h_5^s = 134$ .

Con la *spline*  $g^s(h^s)$  definida en términos de los regresores anteriores, puede estimarse el modelo  $\hat{\mathbf{g}}_{1,h^s}^{\text{semanal}} = g^s(h^s) + \mathbf{e}_{h^s}$  y obtener una segunda aproximación al componente estacional de periodo semanal, que se denotará por  $\hat{\mathbf{g}}_{2,h^s}^{\text{semanal}} = \hat{g}^s(h^s)$ ,  $h^s = 1, \dots, 168$ <sup>28</sup>. Es decir, se obtiene  $\hat{\mathbf{g}}_{2,t}^{\text{semanal}}$ ,  $t = 1, \dots, 35064$ , tal que  $\hat{\mathbf{g}}_{2,t}^{\text{semanal}} = \hat{g}^s(h^s)$ , si la observación en el instante  $t$  corresponde a la hora de la semana  $h^s$ . En el gráfico 5.2.2 se representa dicha estimación para una semana en particular.

<sup>28</sup> Las estimaciones de los parámetros libres en la especificación de la *spline* semanal fueron las siguientes:  $\hat{g}_{1,0}^s = -1947.1371$ ,  $\hat{g}_{2,0}^s = -2014.8875$ ,  $\hat{g}_{3,0}^s = 1909.4849$ ,  $\hat{g}_{4,0}^s = 179354.9594$ ,  $\hat{g}_{5,0}^s = 13074751.17$  y  $\hat{g}_{6,0}^s = -560388.7523$ .

Gráfico 5.2.2. Variación estacional de periodo semanal estimada,  $\hat{g}_{2,h^s}^{semanal}$



Por otra parte, si el conjunto de regresores que definen la *spline*  $g^s(h^s)$  se incorporan en el modelo estructural, puede obtenerse otra aproximación a la fluctuación estacional de periodo semanal definida como

$$g_{3,t}^{semanal} = g_{1,0}^s X_{1,t}^s + \dots + g_{6,0}^s X_{6,t}^s,$$

donde  $X_{i,t}^s = X_{i,h^s}^s$ ,  $i = 1, 2, \dots, 6$ , si la observación en el instante  $t$  corresponde a la hora de la semana  $h^s$ .

### 5.2.3. Extracción de la fluctuación de periodo diario, $\{g_t^{diaria}\}$

Con los regresores definidos para las funciones *splines* que captan las variaciones estacionales de periodo anual y semanal, es posible estimar el modelo estructural originalmente especificado. Sin embargo, y a efectos ilustrativos, se ha optado por tratar de aislar también la fluctuación estacional de periodo diario,  $g_t^{diaria}$ , y recogerla mediante una función *spline*. Por otra parte, y ante la posibilidad de que la estacionalidad de periodo semanal pudiera ser estocástica, esta operación de especificación de  $g_t^{diaria}$  permite elegir una de estas dos variaciones estacionales como candidata a ser estocástica y recoger las otras dos mediante los regresores apropiados.

Para obtener una primera aproximación de  $g_t^{diaria}$ , se ha introducido la hipótesis de que la fluctuación estacional de periodo diario es determinística y se ha procedido a estimar el modelo

$$y_t = m_t + g_t^{diaria} + g_t^{semanal} + g_t^{anual} + e_t, \quad e_t \sim NID(0, \sigma_e^2),$$

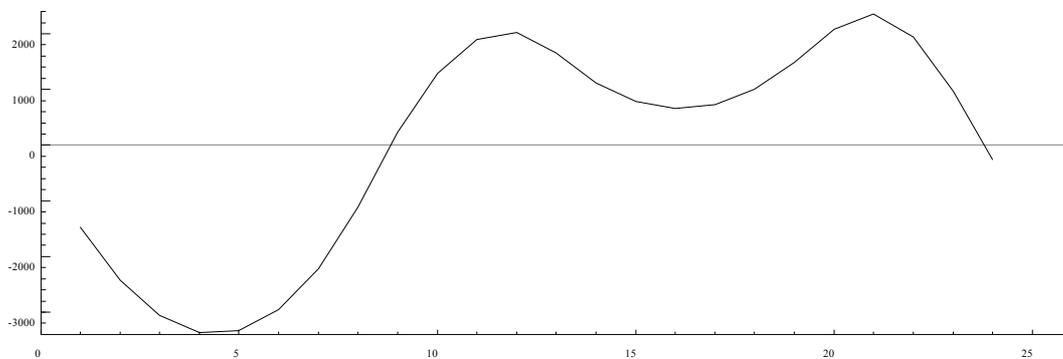
donde  $m_t$  representa una tendencia lineal determinística,  $g_t^{diaria}$  es la fluctuación de periodo diario modelada a través de variables cualitativas estacionales y  $g_t^{semanal}$  y  $g_t^{anual}$

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

representan las variaciones de periodo semanal y anual, respectivamente, que se han incorporado mediante los regresores adecuados para las *splines* correspondientes. La estimación de la estacionalidad de periodo diario,  $\hat{\mathbf{g}}_{1,h^d}^{diaria}$ , es la que se muestra en el gráfico 5.1.6.

Y dado que, bajo el supuesto de que  $\mathbf{g}_i^{diaria}$  es determinística,  $\hat{\mathbf{g}}_{1,t}^{diaria} = \mathbf{g}_{1,h^d}^{diaria}$ , si la observación en el instante  $t$  corresponde a la hora del día  $h^d$ ,  $h^d = 1, \dots, 24$ , el componente diario fue especificado a través de la función *spline*  $g^d(h^d)$ , definida como  $g^d(h^d) = g_i^d(h^d)$ , donde  $g_i^d(h^d) = g_{i,0}^d + g_{i,1}^d h^d + g_{i,2}^d (h^d)^2 + g_{i,3}^d (h^d)^3$ , con  $h_{i-1}^d \leq h^d \leq h_i^d$ , para  $i = 1, \dots, 6$ ,  $h^d = 1, \dots, 24$ , y  $h_0^d = 1$  y  $h_6^d = 24$ , y sujeta a las conocidas restricciones de continuidad. La combinación elegida para los puntos de ruptura fue  $h_1^d = 8$ ,  $h_2^d = 9$ ,  $h_3^d = 13$ ,  $h_4^d = 19$  y  $h_5^d = 21$ . Con estos puntos de ruptura, la variación estacional de periodo diario puede aproximarse ajustando las medias denotadas por  $\hat{\mathbf{g}}_{1,h^d}^{diaria}$  a través de la *spline*  $g^d(h^d)$ , expresada en términos de un conjunto de regresores  $X_{i,h^d}^d$ ,  $i = 1, \dots, 6$ , cuya definición es similar a la del caso semanal. Estimando el modelo  $\hat{\mathbf{g}}_{1,h^d}^{diaria} = g^d(h^d) + \mathbf{e}_{h^d}$  se obtiene  $\hat{\mathbf{g}}_{2,h^d}^{diaria}$ , tal que  $\hat{\mathbf{g}}_{2,t}^{diaria} = \hat{\mathbf{g}}_{2,h^d}^{diaria} = \hat{g}^d(h^d)$ ,  $t = 1, \dots, 35064$ , si la observación en el instante  $t$  corresponde a la hora del día  $h^d$ ,  $h^d = 1, \dots, 24$ . Esta aproximación se representa en el gráfico 5.2.3 para un día cualquiera<sup>29</sup>.

**Gráfico 5.2.3. Variación estacional de periodo diario,  $\hat{\mathbf{g}}_{2,h^d}^{diaria}$**



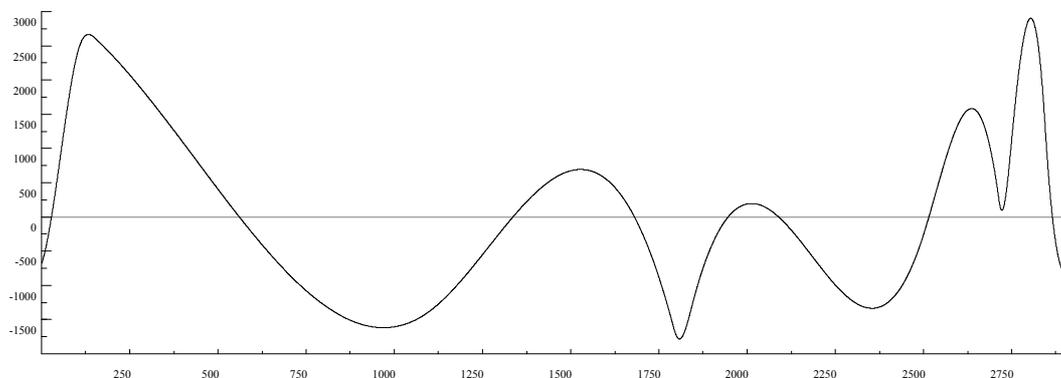
<sup>29</sup> Las estimaciones de los parámetros libres en la especificación de la *spline* diaria fueron las siguientes:  $\hat{g}_{1,0}^d = -217.8302$ ,  $\hat{g}_{2,0}^d = 68862.2219$ ,  $\hat{g}_{3,0}^d = -24882.7457$ ,  $\hat{g}_{4,0}^d = 26731.5634$ ,  $\hat{g}_{5,0}^d = 620824.8211$  y  $\hat{g}_{6,0}^d = -621096.6355$ .

Por otra parte, la fluctuación estacional de periodo diario puede aproximarse como  $g_{3,h^d}^{diaria} = g^d(h^d)$  e incorporarse en un modelo estructural como función de las variables explicativas  $X_{i,t}^d = X_{i,h^d}^d$ ,  $i=1, 2, \dots, 6$ , si la observación en el instante  $t$  corresponde a la hora del día  $h^d$ , es decir,

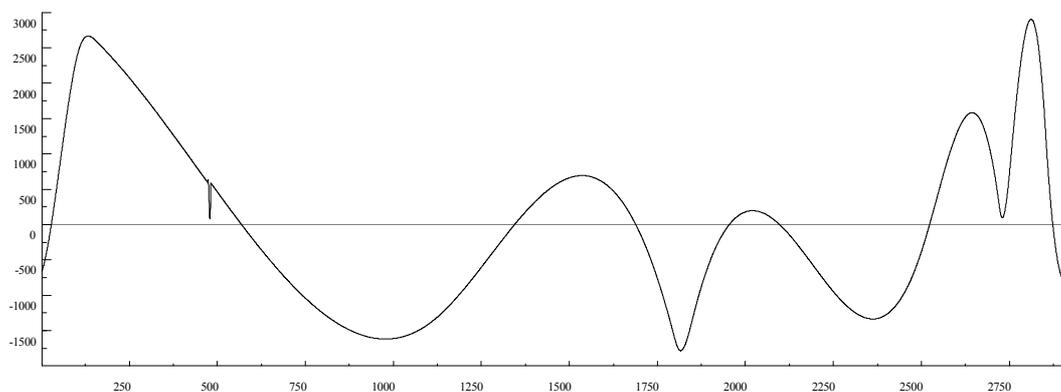
$$g_{3,t}^{diaria} = g_{1,0}^d X_{1,t}^d + \dots + g_{6,0}^d X_{6,t}^d.$$

Con objeto de comprobar que los regresores propios de cada una de las *splines* recogían adecuadamente las fluctuaciones estacionales correspondientes, se procedió a estimar un modelo para la serie original que incluyera un componente tendencial determinístico y las tres fluctuaciones estacionales recogidas a través de las respectivas funciones *splines* (gráficos 5.2.4 a 5.2.6). A partir de dichos gráficos, se advierte que, efectivamente, el comportamiento descrito por las tres *splines* se ajusta al obtenido mediante  $\hat{g}_{2,h^a}^{anual}$ ,  $\hat{g}_{2,h^s}^{semanal}$  y  $\hat{g}_{2,h^d}^{diaria}$  para cada una de las tres fluctuaciones periódicas que configuran el componente estacional de la serie.

**Gráfico 5.2.4. Spline anual estimada en un modelo con componentes determinísticos  
Año no bisiesto**



**Año bisiesto**



29 de febrero

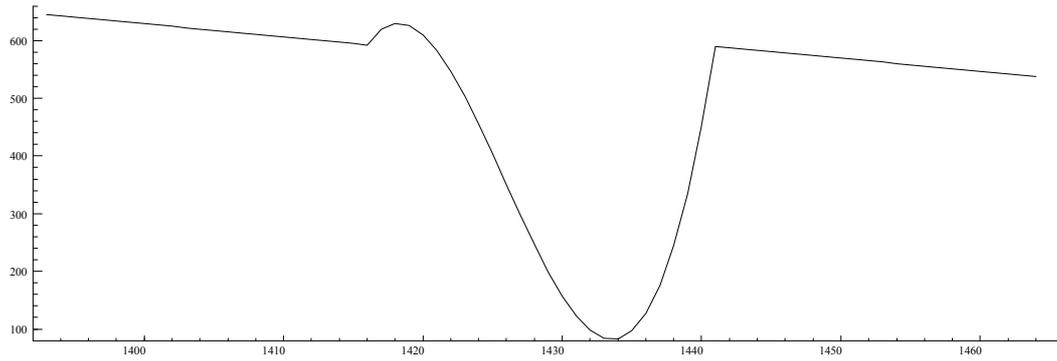


Gráfico 5.2.5. *Spline* semanal estimada en un modelo con componentes determinísticos

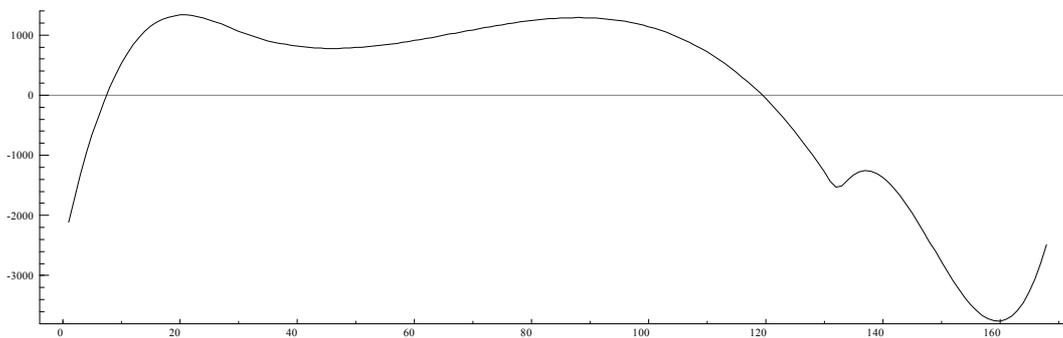
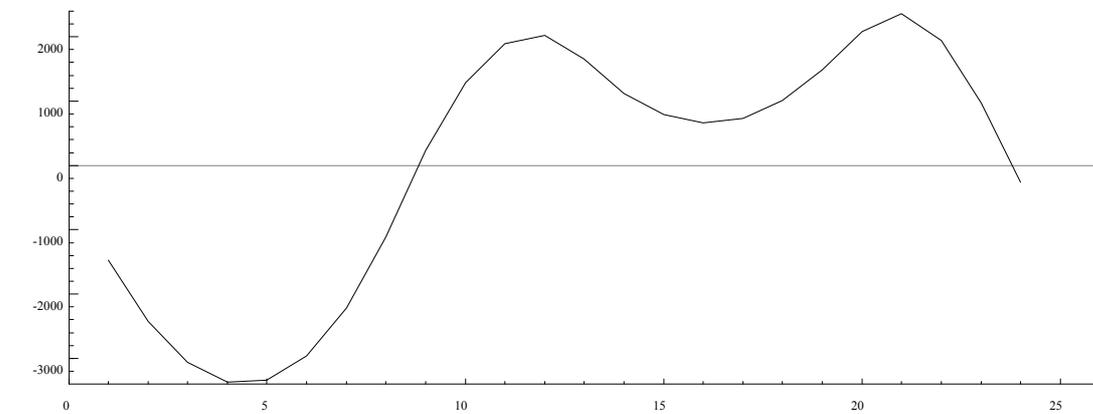


Gráfico 5.2.6. *Spline* diaria estimada en un modelo con componentes determinísticos



#### 5.2.4. Estimación del modelo estructural

Una vez identificados los regresores que definen las funciones *splines* correspondientes a las fluctuaciones estacionales de periodo anual y semanal, el modelo estructural que parece apropiado para recoger la evolución de la serie de demanda horaria de energía eléctrica es el que se especifica a continuación:

$$y_t = \mathbf{m}_t + \mathbf{g}_t^{diaria} + \mathbf{g}_{1,0}^a X_{1,t}^a + \dots + \mathbf{g}_{13,0}^a X_{13,t}^a + \mathbf{g}_0^{29F} X_{0(29F),t}^{29F} + \mathbf{g}_1^{29F} X_{1(29F),t}^{29F} + \mathbf{g}_{1,0}^s X_{1,t}^s + \dots + \mathbf{g}_{6,0}^s X_{6,t}^s + \mathbf{dCH}_t + \mathbf{e}_t$$

con  $\mathbf{e}_t \sim NID(0, \sigma_\varepsilon^2)$ .

Sin embargo, este modelo no puede ser estimado ya que  $\sum_{i=1}^{13} X_{i,t}^a = 1, \forall t$ <sup>30</sup>, y

$\sum_{i=1}^6 X_{i,t}^s = 1, \forall t$ . Dado que  $X_{1,t}^a = 1 - \sum_{i=2}^{13} X_{i,t}^a$ , entonces

$$\begin{aligned} & \mathbf{g}_{1,0}^a X_{1,t}^a + \mathbf{g}_{2,0}^a X_{2,t}^a + \dots + \mathbf{g}_{13,0}^a X_{13,t}^a \\ &= \mathbf{g}_{1,0}^a \left( 1 - \sum_{i=2}^{13} X_{i,t}^a \right) + \mathbf{g}_{2,0}^a X_{2,t}^a + \dots + \mathbf{g}_{13,0}^a X_{13,t}^a \\ &= \mathbf{g}_{1,0}^a + (\mathbf{g}_{2,0}^a - \mathbf{g}_{1,0}^a) X_{2,t}^a + (\mathbf{g}_{3,0}^a - \mathbf{g}_{1,0}^a) X_{3,t}^a + \dots + (\mathbf{g}_{13,0}^a - \mathbf{g}_{1,0}^a) X_{13,t}^a \end{aligned}$$

Del mismo modo, dado que  $X_{1,t}^s = 1 - \sum_{i=2}^6 X_{i,t}^s$ , entonces

$$\begin{aligned} & \mathbf{g}_{1,0}^s X_{1,t}^s + \mathbf{g}_{2,0}^s X_{2,t}^s + \dots + \mathbf{g}_{6,0}^s X_{6,t}^s \\ &= \mathbf{g}_{1,0}^s \left( 1 - \sum_{i=2}^6 X_{i,t}^s \right) + \mathbf{g}_{2,0}^s X_{2,t}^s + \dots + \mathbf{g}_{6,0}^s X_{6,t}^s \\ &= \mathbf{g}_{1,0}^s + (\mathbf{g}_{2,0}^s - \mathbf{g}_{1,0}^s) X_{2,t}^s + (\mathbf{g}_{3,0}^s - \mathbf{g}_{1,0}^s) X_{3,t}^s + \dots + (\mathbf{g}_{6,0}^s - \mathbf{g}_{1,0}^s) X_{6,t}^s \end{aligned}$$

Por tanto, puede especificarse el modelo estimable siguiente:

$$y_t = \mathbf{m}_t + \mathbf{g}_t^{diaria} + \mathbf{g}_{2,0}^{a*} X_{2,t}^a + \dots + \mathbf{g}_{13,0}^{a*} X_{13,t}^a + \mathbf{g}_0^{29F} X_{0(29F),t}^{29F} + \mathbf{g}_1^{29F} X_{1(29F),t}^{29F} + \mathbf{g}_{2,0}^{s*} X_{2,t}^s + \dots + \mathbf{g}_{6,0}^{s*} X_{6,t}^s + \mathbf{dCH}_t + \mathbf{e}_t$$

con  $\mathbf{e}_t \sim NID(0, \sigma_\varepsilon^2)$  y donde, inicialmente,  $\mathbf{m}_t$  se ha especificado como un componente de nivel estocástico con pendiente estocástica y  $\mathbf{g}_t^{diaria}$  recoge la fluctuación estacional de periodo diario mediante formulación estocástica trigonométrica<sup>31</sup>.

La estimación de este modelo revela la naturaleza estocástica de la fluctuación estacional de periodo diario y que la varianza del término de pendiente del componente de nivel es nula, lo que implica, en principio, la naturaleza determinística de este término. Se especificó, entonces, un modelo con pendiente determinística, cuya

<sup>30</sup> Nótese que esta igualdad no es estricta en las observaciones correspondientes al 29 de febrero de 2000.

<sup>31</sup> Como ya sucedió en el capítulo anterior, la transformación logarítmica no proporcionaba mejores resultados que los finalmente presentados en este capítulo (anexo 5.5).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

estimación en el vector de estado final reveló su no significación estadística y condujo a especificar un modelo con componente de nivel estocástico sin término de pendiente<sup>32</sup>. Esta especificación merece ser comentada dado que el propio gráfico de la serie muestra un cierto crecimiento que ha sido tenido en cuenta a lo largo del proceso y, por tanto, llama la atención la ausencia de pendiente. Es preciso señalar además que, aunque los contrastes habituales de significación estadística conduzcan a eliminar este término, su magnitud podría ser suficiente para provocar la divergencia del filtro de Kalman, que se manifiesta en un distanciamiento, a veces cíclico, entre observaciones y predicciones<sup>33</sup>. Sin embargo, tal distanciamiento no se produce en el modelo especificado sin término de pendiente<sup>34</sup>.

Entonces, el modelo considerado fue

$$y_t = \mathbf{m}_t + \mathbf{g}_t^{diaria} + \mathbf{g}_{1,0}^a X_{1,t}^a + \dots + \mathbf{g}_{13,0}^a X_{13,t}^a + \mathbf{g}_0^{29F} X_{0(29F),t}^{29F} + \mathbf{g}_1^{29F} X_{1(29F),t}^{29F} + \mathbf{g}_{1,0}^s X_{1,t}^s + \dots + \mathbf{g}_{6,0}^s X_{6,t}^s + \mathbf{dCH}_t + \mathbf{e}_t$$

donde  $\mathbf{e}_t \sim NID(0, \sigma_\varepsilon^2)$ , con

$$\mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{h}_t, \quad \mathbf{h}_t \sim NID(0, \sigma_\eta^2),$$

y

$$\gamma_t^{diaria} = \sum_{j=1}^{12} \gamma_{j,t},$$

donde cada término  $\mathbf{g}_{j,t}$  es generado por

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} = \begin{pmatrix} \cos \lambda_j & \text{sen } \lambda_j \\ -\text{sen } \lambda_j & \cos \lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix}, \quad \mathbf{I}_j = \frac{2\mathbf{p}j}{24}, \quad \omega_{j,t}, \omega_{j,t}^* \sim NID(0, \sigma_\omega^2),$$

para  $j = 1, \dots, 12$  y  $t = 1, \dots, 35064$ , cuya estimación ofrece los resultados que aparecen en los cuadros 5.2.1 a 5.2.4 y gráfico 5.2.7.

<sup>32</sup> Es preciso señalar que para que algunos de los modelos anteriores pudieran ser estimados, fue necesario fijar un número de iteraciones hasta alcanzar la convergencia inferior al establecido por defecto en el programa *STAMP*. Aún así, en el modelo con pendiente estocástica se alcanzó la convergencia fuerte en once iteraciones, mientras que, en el de pendiente determinística se obtuvo convergencia muy fuerte en cinco iteraciones. Con objeto de afianzar la conclusión sobre la naturaleza del componente tendencial, se efectuó una estimación alternativa, en la que, una vez filtrada la serie original de las fluctuaciones estacionales de periodo anual, semanal y diario, recogidas a través de las respectivas funciones *splines*, se especificó un modelo que sólo incluía componente de nivel estocástico con pendiente estocástica. La estimación de este modelo sugería, de nuevo, que la pendiente no era estocástica y, además, la estimación del modelo con pendiente fija revelaba que ésta no era significativa.

<sup>33</sup> Véase Vargas (1999:34).

<sup>34</sup> Véase, más adelante, los gráficos 5.2.7 y 5.2.12.

**Cuadro 5.2.1. Varianzas de los términos de perturbación**

	Irregular ( $\hat{S}_e^2$ )	Nivel ( $\hat{S}_h^2$ )	Estacionalidad diaria ( $\hat{S}_w^2$ )
<b>Varianza</b>	0.0000	160070	9.5773
<b>q-ratios</b>	(0.0000)	(1.0000)	(0.0001)

Nótese que la varianza del componente irregular es nula; sin embargo, esta circunstancia no debe interpretarse como ausencia de comportamientos no sistemáticos, que, de hecho, se aprecian con claridad en la serie y que quedan recogidos a través de los otros componentes inobservables de naturaleza estocástica, especialmente del componente de nivel cuyo término de perturbación es el que presenta mayor varianza.

**Cuadro 5.2.2. Vector de estado final<sup>35</sup>**

Componente	Estimación	Componente	Estimación
$\hat{m}_T$	18681.4611	$\hat{g}_{6,T}^*$	-24.130(*)
$\hat{g}_{1,T}$	-85.681(*)	$\hat{g}_{7,T}$	30.100(*)
$\hat{g}_{1,T}^*$	-2403.6	$\hat{g}_{7,T}^*$	44.516(*)
$\hat{g}_{2,T}$	1138.4	$\hat{g}_{8,T}$	-48.757(*)
$\hat{g}_{2,T}^*$	-1111.2	$\hat{g}_{8,T}^*$	84.210
$\hat{g}_{3,T}$	-596.09	$\hat{g}_{9,T}$	93.279
$\hat{g}_{3,T}^*$	-89.782	$\hat{g}_{9,T}^*$	120.96
$\hat{g}_{4,T}$	133.76	$\hat{g}_{10,T}$	104.92
$\hat{g}_{4,T}^*$	276.81	$\hat{g}_{10,T}^*$	-32.202(*)
$\hat{g}_{5,T}$	211.50	$\hat{g}_{11,T}$	67.751
$\hat{g}_{5,T}^*$	187.42	$\hat{g}_{11,T}^*$	-10.393(*)
$\hat{g}_{6,T}$	185.81	$\hat{g}_{12,T}$	-12.099(*)

(\*) Parámetros no significativos al 90 % de confianza.

El valor del estadístico del contraste de significación conjunta del componente estacional de periodo diario fue  $c_{23}^2 = 1960.79$ .

<sup>35</sup> Nótese que  $\hat{g}_{12,t} = -\hat{g}_{12,t-1} + w_{12,t}$ , por lo que  $\hat{g}_{12,t}^*$  no interviene en el proceso recursivo de determinación de los términos  $\hat{y}_{12,t}$  (véase capítulo 1:17).

**Cuadro 5.2.3. Componente estacional de periodo diario (estimación correspondiente al lunes 31 de diciembre de 2001)**

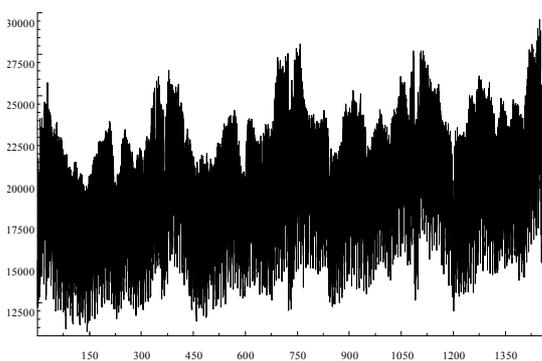
Hora del día	Estimación	Hora del día	Estimación	Hora del día	Estimación
1	-252.2775	9	-1150.7601	17	-29.3460
2	-1947.1366	10	37.4587	18	683.2072
3	-2887.9349	11	1491.6389	19	3033.5662
4	-3173.5419	12	1780.9657	20	3421.7037
5	-3370.9540	13	1682.6747	21	3048.5763
6	-3395.9689	14	1174.1238	22	2260.0364
7	-2799.8942	15	308.5995	23	1071.6638
8	-2200.4757	16	-10.1012	24	1223.0096

**Cuadro 5.2.4. Parámetros de las variables explicativas**

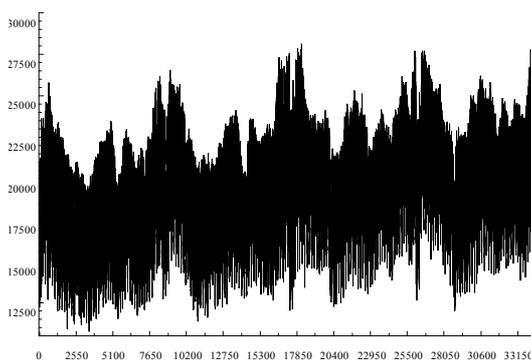
Parámetro	Estimación	Estadístico <i>t</i>	Parámetro	Estimación	Estadístico <i>t</i>
$\hat{\xi}_{2,0}^{a*}$	-7310.3	-0.82178	$\hat{\xi}_{12,0}^{a*}$	-2647x10 <sup>6</sup>	-0.39943
$\hat{\xi}_{3,0}^{a*}$	4124.7	0.72002	$\hat{\xi}_{13,0}^{a*}$	18546x10 <sup>3</sup>	0.081185
$\hat{\xi}_{4,0}^{a*}$	57260	0.58442	$\hat{\xi}_0^{29F}$	-1762.4x10 <sup>6</sup>	-1.0308
$\hat{\xi}_{5,0}^{a*}$	-99573x10 <sup>3</sup>	-0.96892	$\hat{\xi}_1^{29F}$	3691.4x10 <sup>3</sup>	1.0828
$\hat{\xi}_{6,0}^{a*}$	42372x10 <sup>3</sup>	0.99104	$\hat{\xi}_{2,0}^{s*}$	-107.01	-38.99
$\hat{\xi}_{7,0}^{a*}$	-1010.6x10 <sup>3</sup>	-0.99926	$\hat{\xi}_{3,0}^{s*}$	6230.3	29.991
$\hat{\xi}_{8,0}^{a*}$	14730x10 <sup>3</sup>	1.4773	$\hat{\xi}_{4,0}^{s*}$	673.86x10 <sup>3</sup>	14.772
$\hat{\xi}_{9,0}^{a*}$	-11421x10 <sup>6</sup>	-1.8868	$\hat{\xi}_{5,0}^{s*}$	47066x10 <sup>3</sup>	15.829
$\hat{\xi}_{10,0}^{a*}$	1211.8x10 <sup>6</sup>	1.604	$\hat{\xi}_{6,0}^{s*}$	-1155x10 <sup>3</sup>	-25.403
$\hat{\xi}_{11,0}^{a*}$	-62209x10 <sup>3</sup>	-0.26394	$\hat{d}$	-306.72	-2.0732

**Gráfico 5.2.7. Valores observados y ajustados**

**Observados**



**Ajustados**



A pesar de que las diferencias entre las observaciones de la serie y los valores predichos por el modelo eran de magnitud muy reducida<sup>36</sup> (gráfico 5.2.7), el análisis de

<sup>36</sup> Los errores de predicción del modelo anterior oscilaban entre -0.00914157 y 0.00914363, cifras que representan menos de un 0.00005% del valor medio de la serie en el periodo examinado.

los residuos del modelo anteriormente estimado condujo a la incorporación de variables de intervención para recoger observaciones anómalas detectadas en el componente de nivel, que se especificó como:

$$m_t = m_{t-1} + \sum_{d-h^d} I_{d-h^d} I_{d-h^d} + h_t, \quad h_t \sim NID(0, \sigma_\eta^2),$$

donde  $I_{d-h^d}$  es una variable de intervención que toma el valor uno si la observación corresponde a hora  $h^d$  del día  $d$  y cero en cualquier otro caso<sup>37</sup>. Los resultados de la estimación del modelo finalmente especificado son los que aparecen en los cuadros 5.2.5 a 5.2.9 y los gráficos 5.2.8 a 5.2.12.

**Cuadro 5.2.5. Varianzas de los términos de perturbación**

	Irregular ( $\hat{S}_e^2$ )	Nivel ( $\hat{S}_h^2$ )	Estacionalidad diaria ( $\hat{S}_w^2$ )
<b>Varianza</b>	0.0000	159190	9.5938
<b>q-ratios</b>	(0.0000)	(1.0000)	(0.0001)

**Cuadro 5.2.6. Vector de estado final**

Componente	Estimación	Componente	Estimación
$\hat{m}_T$	19570.096	$\hat{g}_{6,T}^*$	-24.242(*)
$\hat{g}_{1,T}$	-84.117(*)	$\hat{g}_{7,T}$	30.199(*)
$\hat{g}_{1,T}^*$	-2404.200	$\hat{g}_{7,T}^*$	44.599(*)
$\hat{g}_{2,T}$	1138.900	$\hat{g}_{8,T}$	-48.651(*)
$\hat{g}_{2,T}^*$	-1110.600	$\hat{g}_{8,T}^*$	84.182
$\hat{g}_{3,T}$	-596.040	$\hat{g}_{9,T}$	93.347
$\hat{g}_{3,T}^*$	-90.157	$\hat{g}_{9,T}^*$	120.950
$\hat{g}_{4,T}$	134.000	$\hat{g}_{10,T}$	104.970
$\hat{g}_{4,T}^*$	276.830	$\hat{g}_{10,T}^*$	-32.215(*)
$\hat{g}_{5,T}$	211.530	$\hat{g}_{11,T}$	67.823
$\hat{g}_{5,T}^*$	187.660	$\hat{g}_{11,T}^*$	-10.405(*)
$\hat{g}_{6,T}$	185.910	$\hat{g}_{12,T}$	-12.091(*)

(\*) Parámetros no significativos al 90 % de confianza.

El valor del estadístico del contraste de significación conjunta del componente estacional de periodo diario fue  $c_{23}^2 = 1965.26$ .

<sup>37</sup> Se ha tomado la decisión de intervenir sólo cuando las desviaciones respecto a lo predicho por el modelo son particularmente grandes. Este criterio se adopta en general en series de alta frecuencia, en las que suelen aparecer numerosas observaciones anómalas. Véase, por ejemplo, Cabrero y otros (2002).

**Cuadro 5.2.7. Parámetros de las variables explicativas**

Parámetro	Estimación	Estadístico $t$	Parámetro	Estimación	Estadístico $t$
$\hat{\xi}_{2,0}^{a*}$	-6972	-0.78584	$\hat{\xi}_{12,0}^{a*}$	$-1885.8 \times 10^6$	-0.28523
$\hat{\xi}_{3,0}^{a*}$	2970.1	0.51972	$\hat{\xi}_{13,0}^{a*}$	$-16353 \times 10^3$	-0.071737
$\hat{\xi}_{4,0}^{a*}$	67040	0.68607	$\hat{\xi}_0^{29F}$	$-1766.7 \times 10^6$	-1.0361
$\hat{\xi}_{5,0}^{a*}$	$-101540 \times 10^3$	-0.99074	$\hat{\xi}_1^{29F}$	$3700.4 \times 10^3$	1.0333
$\hat{\xi}_{6,0}^{a*}$	$43018 \times 10^3$	1.0089	$\hat{\xi}_{2,0}^{s*}$	-107.61	-39.3
$\hat{\xi}_{7,0}^{a*}$	$-1011.4 \times 10^3$	-1.0028	$\hat{\xi}_{3,0}^{s*}$	6271.8	30.267
$\hat{\xi}_{8,0}^{a*}$	$14768 \times 10^3$	1.4853	$\hat{\xi}_{4,0}^{s*}$	$678.33 \times 10^3$	14.91
$\hat{\xi}_{9,0}^{a*}$	$-11517 \times 10^6$	-1.9079	$\hat{\xi}_{5,0}^{s*}$	$47344 \times 10^3$	15.965
$\hat{\xi}_{10,0}^{a*}$	$1228.1 \times 10^6$	1.63	$\hat{\xi}_{6,0}^{s*}$	$-1159.5 \times 10^3$	-25.568
$\hat{\xi}_{11,0}^{a*}$	$-72707 \times 10^3$	-0.30931	$\hat{d}$	-306.62	-2.0779

Al observar que algunos de los parámetros que definen la *spline* que recoge la fluctuación anual no resultaron significativos, se llevó a cabo un contraste que permitiera obtener alguna conclusión sobre la significación conjunta de dichos parámetros. Aunque dicho contraste,  $F=0.74211 \sim F_{14, 35036}$ , sugiere que las variables  $X_{2,t}^a, \dots, X_{13,t}^a, X_{0(29F),t}^{29F}, X_{1(29F),t}^{29F}$ , eran, en su conjunto, irrelevantes desde un punto de vista estadístico, se decidió mantenerlas en el modelo porque, desde el punto de vista económico, recogen un significativo comportamiento estacional de periodo anual<sup>38</sup>.

Habría que hacer notar que las fluctuaciones anuales en la demanda de energía eléctrica están más directamente relacionadas con factores ambientales, como la temperatura, que con la propia época del año y que quizás este hecho contribuye a la escasa significación estadística de los regresores que definen la *spline* anual. Esta circunstancia aconsejaría introducir la temperatura como variable explicativa adicional de la demanda horaria de energía eléctrica<sup>39</sup>.

Nótese en el cuadro 5.2.7 el signo negativo del parámetro  $\hat{d}$ , lo que refleja, en principio, el ahorro de energía que se produce como consecuencia del cambio de hora

<sup>38</sup> Recuérdese el sentido del sesgo del estadístico  $F$  de restricciones comentado en los capítulos tercero y cuarto. En cualquier caso, el hecho de que se mantengan variables irrelevantes en la especificación del modelo puede ocasionar problemas de estimación. Sin embargo, la estimación del modelo sin el conjunto de regresores que definen la fluctuación anual reflejaba la ausencia de efectos sustanciales sobre la estimación de los parámetros del modelo propuesto.

<sup>39</sup> Un indicador de la existencia de un efecto de la temperatura sobre la demanda de energía eléctrica es la publicación por parte de organismos oficiales de datos corregidos de dicho efecto (véase Informes sobre la Explotación del Sistema Eléctrico Nacional. Delegación del Gobierno en la Explotación del Sistema Eléctrico).

oficial. Ahora bien, esta interpretación debe hacerse con suma cautela en el sentido de que no es directa la imputación de este efecto al cambio de hora, sino que tal vez podría estar asociado a la propia fluctuación de periodo anual y, más en concreto, con aspectos relacionados con ésta como, por ejemplo, el menor uso de aparatos de calefacción durante la época de temperaturas más suaves en la que tiene lugar el cambio de hora a lo largo de los cuatro años.

**Cuadro 5.2.8. Parámetros de las variables de intervención**

Parámetro	Estimación	Estadístico $t$	Parámetro	Estimación	Estadístico $t$
$\hat{I}_{366-1}$	1954.4	4.6819	$\hat{I}_{1097-9}$	-1885.2	-4.5161
$\hat{I}_{410-9}$	-2216.9	-5.3127	$\hat{I}_{1111-1}$	-2445.0	-5.8562
$\hat{I}_{731-1}$	2379.2	5.6993	$\hat{I}_{1118-1}$	-2695.6	-6.4564

De las intervenciones consideradas cabe destacar aquellas dos que hacen referencia a la una de la mañana del viernes uno de enero de 1999 y del sábado uno de enero de 2000 (intervenciones 366-1 y 731-1). En estos dos casos, el signo positivo de los parámetros que acompañan a las variables de intervención queda explicado por el previsible incremento de la demanda de energía eléctrica en la madrugada del primer día de año. En el primero de año de 2001 y a las nueve de la mañana es necesario introducir otra variable de intervención (1097-9) cuyo parámetro, en este caso negativo, queda justificado por tratarse de un lunes festivo; es decir, los niveles de demanda de energía están por debajo de los correspondientes al inicio de la semana y de la jornada laborable. El resto de variables de intervención hacen referencia a las nueve de la mañana del domingo catorce de febrero de 1999 y a la una de la mañana de los lunes quince y veintidós de enero de 2001. En los tres casos la demanda registrada parece haber caído por debajo del nivel previsto por el modelo.

**Cuadro 5.2.9. Componente estacional de periodo diario (estimación correspondiente al lunes 31 de diciembre de 2001)**

Hora del día	Estimación	Hora del día	Estimación	Hora del día	Estimación
1	-250.4753	9	-1153.5564	17	-29.5010
2	-1945.8889	10	35.0776	18	683.0534
3	-2887.2208	11	1490.3848	19	3033.6462
4	-3173.2689	12	1780.1023	20	3422.2334
5	-3370.9605	13	1682.0418	21	3049.6947
6	-3396.3067	14	1174.0108	22	2261.2652
7	-2800.9160	15	308.9934	23	1073.3528
8	-2202.6350	16	-10.0303	24	1225.7487

Gráfico 5.2.8. Componente de nivel

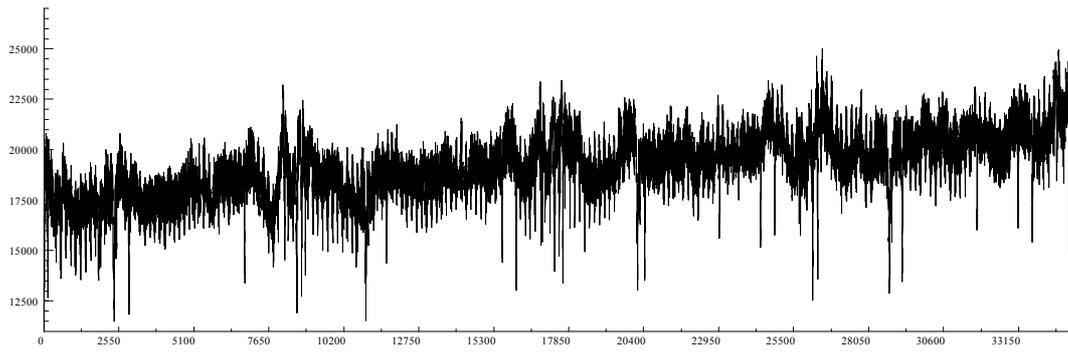
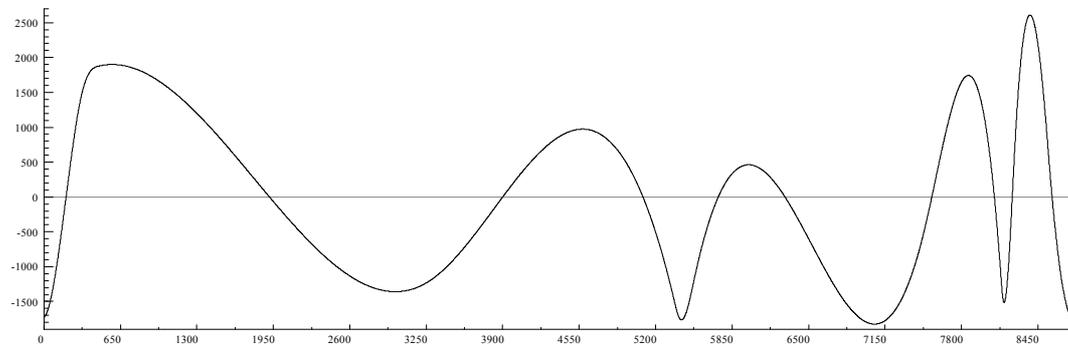
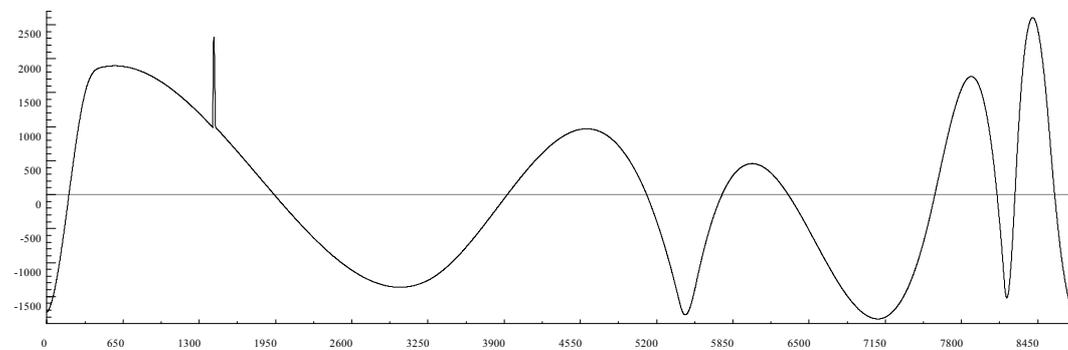


Gráfico 5.2.9. Estacionalidad anual

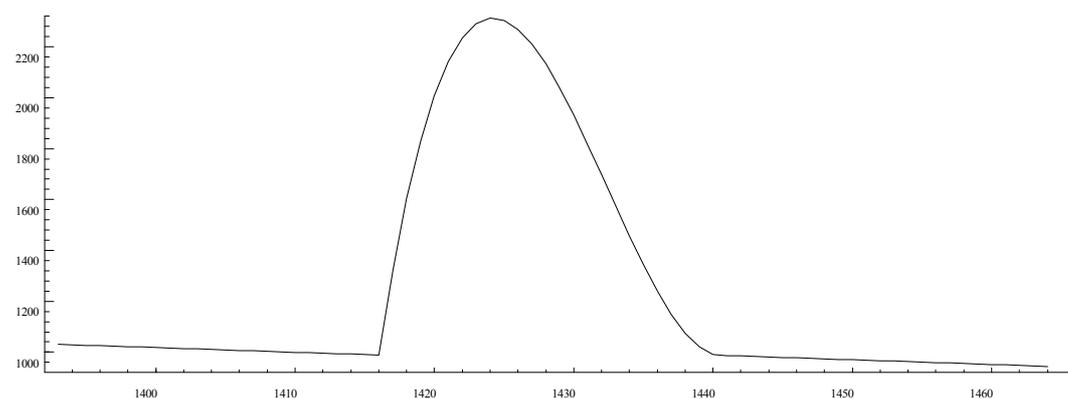
Año no bisiesto



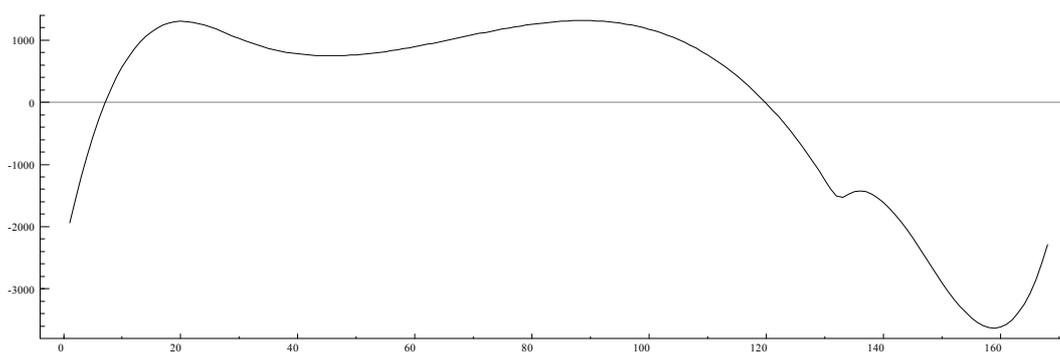
Año bisiesto



29 de febrero

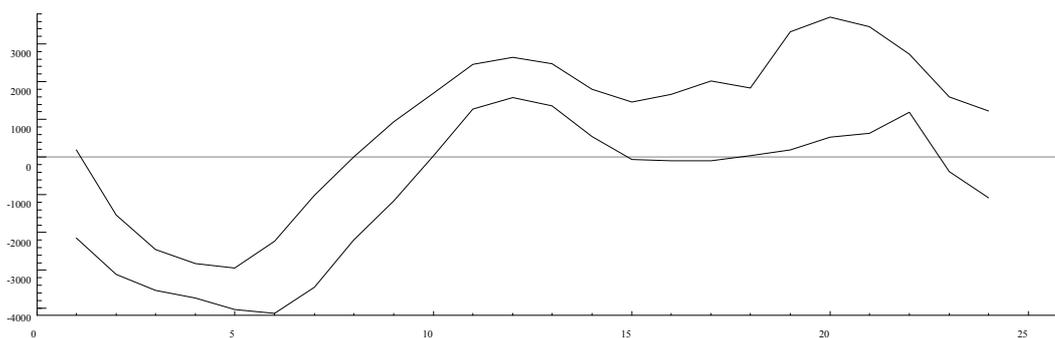


**Gráfico 5.2.10. Estacionalidad semanal**



Las estimaciones de las fluctuaciones estacionales de periodo anual,  $\hat{g}_{3,h^a}^{anual}$ , y semanal,  $\hat{g}_{3,h^s}^{semanal}$ , se han obtenido a partir de los coeficientes estimados para las variables exógenas  $X_{2,t}^{anual}$ , ...,  $X_{13,t}^{anual}$ ,  $X_{0(29F),t}^{29F}$  y  $X_{1(29F),t}^{29F}$ , en el primer caso, y  $X_{2,t}^{semanal}$ , ...,  $X_{6,t}^{semanal}$ , en el caso semanal. El término independiente de cada una de las *splines*, incorporado en el componente de nivel, se ha calculado de forma que las medias de las fluctuaciones estacionales de periodo anual y semanal estimadas fuesen nulas<sup>40</sup>. El componente de nivel representado, así como su estimación en el vector de estado final (cuadro 5.2.6)<sup>41</sup>, también se han corregido oportunamente con el fin de que dichos términos independientes no quedasen recogidos simultáneamente en los componentes tendencial y estacional.

**Gráfico 5.2.11. Estacionalidad diaria**



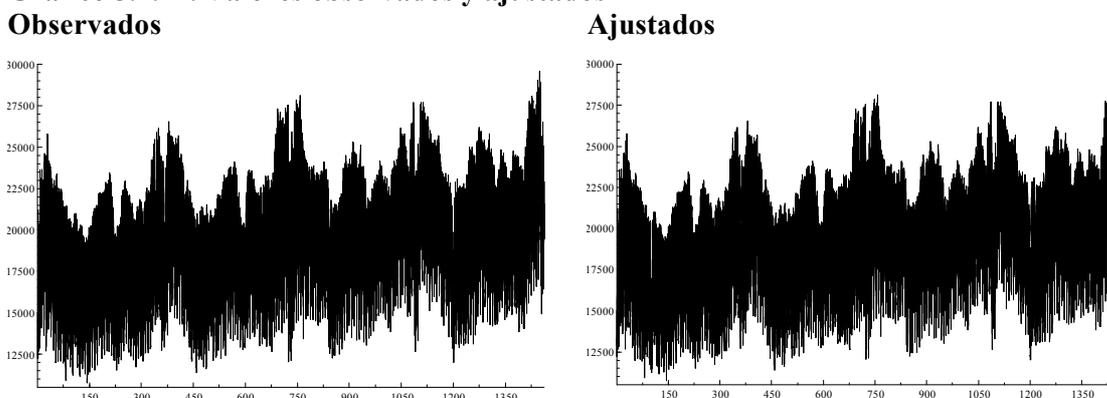
<sup>40</sup> En relación con el nivel de la serie, la variación estacional de periodo anual significa como máximo el 15.5379% y, en términos medios, este porcentaje es del 5.0819%. En el caso semanal, estos porcentajes ascienden al 24.9951% y 6.8734%, respectivamente. Estas cifras pueden interpretarse como reflejo de la mayor importancia relativa de la variación de periodo semanal.

<sup>41</sup> La corrección adecuada se efectuó también en la estimación del componente de nivel en el modelo sin intervenciones (cuadro 5.2.3).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

La fluctuación estacional diaria estimada para el treinta y uno de diciembre de 2001 (cuadro 5.2.9) muestra un patrón similar al observado en el gráfico 5.1.6. En cualquier caso, las diferencias existentes son, evidentemente, resultado de la naturaleza estocástica otorgada a este componente en la especificación del modelo. En el gráfico 5.2.11 se han representado los valores mínimos y máximos del componente estimado en cada una de las veinticuatro horas del día. El carácter estocástico de la fluctuación de periodo diario se refleja con mayor claridad si se representa la evolución de este componente para cada una de las veinticuatro horas del día a lo largo de toda la muestra (anexo 5.3)<sup>42</sup>.

**Gráfico 5.2.12. Valores observados y ajustados**



La magnitud de los errores de predicción del modelo estimado parece indicar que éste es apropiado<sup>43</sup>. Ahora bien, dado que en el proceso de estimación se ha considerado que la fluctuación estacional de periodo semanal es determinística y que, como se comentó en el epígrafe 5.1, la naturaleza de esta fluctuación podría ser estocástica, se probó a estimar un modelo que permitiera evaluar el carácter estocástico de dicha variación. En concreto, se optó por eliminar la estacionalidad estocástica de periodo diario, así como el resto de componentes de naturaleza determinística, excepto, por supuesto, la fluctuación semanal recogida a través de los regresores de la *spline*

<sup>42</sup> El efecto estacional de periodo diaria llega a significar un 25.9674% del nivel de la serie y, en términos medios, esta variación representa un 8.8326% del nivel anterior a lo largo de la muestra. Y si se considera el componente estacional en su conjunto, como agregación de las variaciones de periodo anual, semanal y diario, este último porcentaje asciende al 20.7879%.

<sup>43</sup> Dichos errores oscilaban entre -0.00906235 y 0.00906641, cifras que representan algo menos del 0.00005% del valor medio de la serie estudiada.

correspondiente<sup>44</sup>. Si se denota por  $\{y_t^s\}$  al proceso transformado, el modelo estimado fue:

$$y_t^s = \mathbf{m}_t + \mathbf{g}_t^{semanal} + \mathbf{e}_t, \quad \mathbf{e}_t \sim NID(0, \sigma_e^2),$$

donde  $\mathbf{m}_t$  es un componente de nivel estocástico sin término de pendiente y  $\mathbf{g}_t^{semanal}$  es el componente estacional estocástico de periodo semanal. Dicho modelo sólo pudo ser estimado incorporando el componente  $\mathbf{g}_t^{semanal}$  con la formulación de *dummies* estocásticas, es decir,  $\mathbf{g}_t^{semanal} = \mathbf{g}_{h^s, t}^{semanal}$ , si la observación en el instante  $t$  corresponde a la hora de la semana  $h^s$ , y  $\mathbf{g}_{h^s, t}^{semanal} = \mathbf{g}_{h^s, t-1}^{semanal} + \mathbf{w}_{h^s, t}$ ,  $h^s = 1, \dots, 168$ , donde  $\mathbf{w}_{h^s, t}$  sigue siendo un término de perturbación aleatoria tal que  $\mathbf{w}_{h^s, t} \sim NID(0, \sigma_w^2)$ . En ese caso, la varianza de dicho componente no fue nula, lo que indica su carácter estocástico (cuadro 5.2.10). Sin embargo, las estimaciones de los términos  $\mathbf{g}_{h^s, t}^{semanal}$  para cada una de las horas de la semana mostraban una relativa estabilidad (anexo 5.4), por lo que cabría pensar que el error que se comete al considerarla determinística no es excesivo. En cualquier caso, no se advierte que la variabilidad de la fluctuación estacional de periodo semanal sea mayor que la que se observa para la fluctuación diaria (anexos 5.3 y 5.4), por lo que no resulta oportuno intercambiar la naturaleza de estas dos variaciones en la especificación del modelo, y mucho menos si se tiene en cuenta el incremento de complejidad computacional aparejado a la consideración de la naturaleza estocástica de la variación semanal<sup>45</sup>. Por tanto, aunque lo adecuado es incorporar las fluctuaciones diaria y semanal con naturaleza estocástica, este argumento justifica la decisión tomada

---

<sup>44</sup> Se intentó también estudiar el carácter estocástico de la fluctuación de periodo semanal a partir de un modelo que contuviera, además de nivel estocástico y la *dummy* para el cambio de hora, dos funciones *splines* para las fluctuaciones anual y diaria, así como una formulación trigonométrica o de *dummies* estocásticas para la fluctuación semanal. Sin embargo, en ninguno de los dos casos, el modelo pudo ser estimado satisfactoriamente.

<sup>45</sup> Los resultados de la estimación de ambos modelos (cuadros 5.2.4 y 5.2.9) podrían llevar a pensar que la varianza del término de perturbación de la fluctuación estacional diaria es mayor que la correspondiente a la variación semanal. Sin embargo, estas varianzas no son directamente comparables dado que la variable dependiente y los regresores incorporados son diferentes. Otro elemento que dificulta la comparación es que las fluctuaciones consideradas como estocásticas han sido formuladas de forma diferente; en el primer caso se utilizó la especificación trigonométrica y en el segundo se recurrió a la formulación de *dummies* estocásticas. Por ello, y a efectos comparativos con la necesaria cautela, se procedió a estimar el modelo propuesto para la demanda horaria utilizando *dummies* estocásticas para recoger el componente estacional de periodo diario; encontrándose que la varianza del término de perturbación de dicho componente fue 1267.1, mayor que la varianza del término de perturbación del componente estacional de periodo semanal cuando éste se incorporaba con la misma formulación de *dummies* estocásticas.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

dadas las limitaciones computacionales a las que se ha aludido para estimar conjuntamente dos fluctuaciones estacionales de naturaleza estocástica.

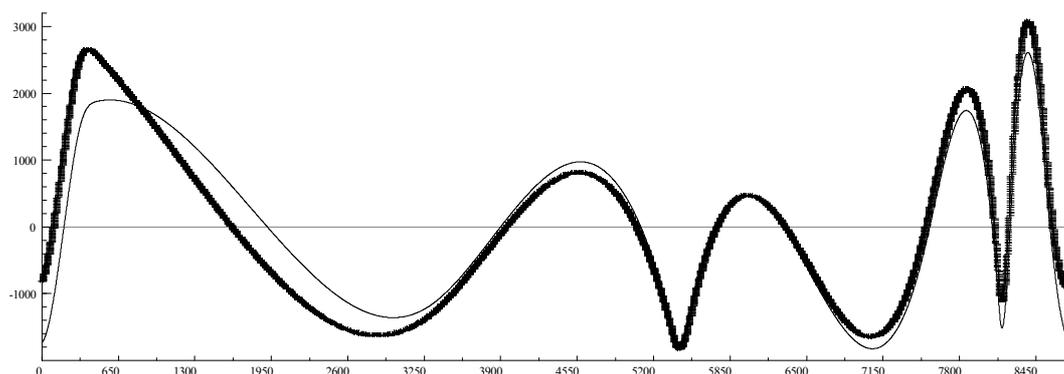
**Cuadro 5.2.10. Varianzas de los términos de perturbación**

	Irregular ( $\hat{S}_e^2$ )	Nivel ( $\hat{S}_h^2$ )	Estacionalidad semanal ( $\hat{S}_w^2$ )
<b>Varianza</b>	0.0000	63462	237.5
<b>q-ratios</b>	(0.0000)	(1.0000)	(0.0037)

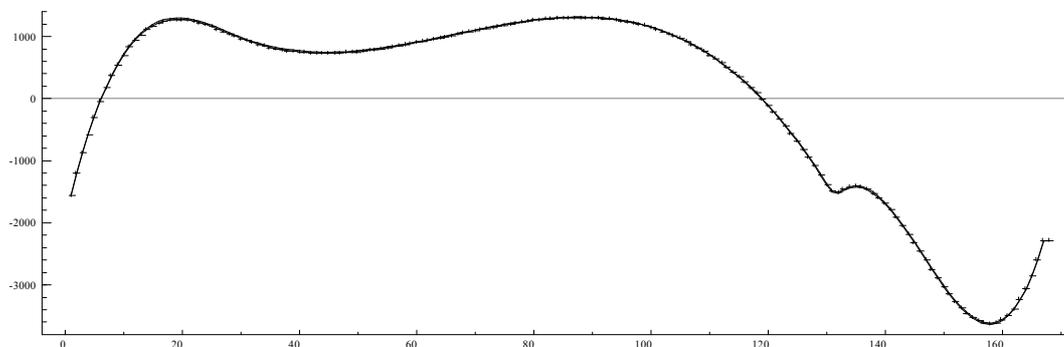
Con respecto al modelo finalmente elegido, en el que la variación estacional de periodo diario se consideró estocástica, merece la pena comentar el efecto de las variables de intervención sobre la estimación de esta fluctuación, así como sobre el resto de las variaciones periódicas detectadas en la serie, especialmente la anual. La introducción de dos variables de intervención correspondientes a la misma hora del año (366-1 y 731-1) y con signo positivo en los respectivos parámetros estimados produce una significativa reducción de la fluctuación anual estimada en las primeras horas del año con respecto a la estimación obtenida sin incorporar variables de intervención. El tipo de variable de intervención en los dos casos anteriores, el signo de los parámetros del resto de las intervenciones consideradas, también ubicadas en los primeros meses del año, y la propia naturaleza de la *spline* pueden ayudar a explicar el impacto que estas intervenciones generan en las estimaciones correspondientes al resto de horas del año. De hecho, pueden ser las restricciones de continuidad y periodicidad las que obligan a que el ajuste se produzca en las horas de principio y final de año, mientras que a mediados de éste las dos estimaciones son prácticamente coincidentes (gráfico 5.2.13).

Por otro lado, la estimación de la fluctuación semanal es mucho menos sensible a la introducción de variables de intervención, lo que parece lógico si se tiene en cuenta el reducido número de intervenciones en comparación con el elevado número de observaciones semanales correspondientes a cada hora de la semana, por lo que difícilmente se producen alteraciones relevantes en los niveles medios por hora de la semana. Esta similitud de comportamientos se refleja claramente en el gráfico 5.2.14, en el que las dos estimaciones están prácticamente superpuestas. Por el contrario, la naturaleza estocástica de la fluctuación diaria justifica la sensibilidad de la estimación únicamente en el entorno de los instantes del tiempo en los que se localizan las variables de intervención (gráfico 5.2.15).

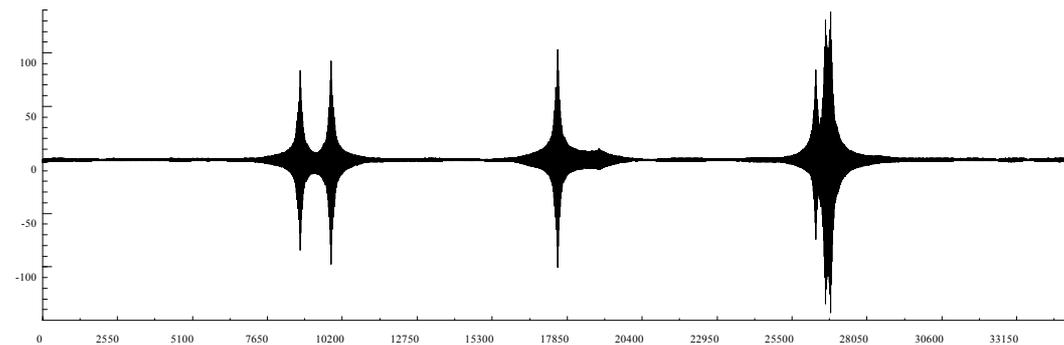
**Gráfico 5.2.13. Estimaciones de la fluctuación estacional de periodo anual con y sin intervenciones**



**Gráfico 5.2.14. Estimaciones de la fluctuación estacional de periodo semanal con y sin intervenciones**



**Gráfico 5.2.15. Estimaciones de la fluctuación estacional de periodo diario con y sin intervenciones**



Lógicamente, las variaciones en las estimaciones de las fluctuaciones estacionales deben tener su contrapartida en la estimación del nivel. De hecho, en el último instante de la muestra no existen diferencias significativas en las estimaciones de las fluctuaciones semanal y diaria, pero sí se produce un descenso de magnitud considerable en la *spline* anual estimada, causado fundamentalmente por las ya comentadas restricciones de periodicidad. Esta caída es compensada por el incremento

experimentado en la estimación del nivel al final de la serie, de modo que el valor ajustado permanece prácticamente inalterado.

### 5.3. Conclusiones

Desde el punto de vista metodológico, se ha mostrado la utilidad de la combinación de funciones *splines* con otras formulaciones específicas del componente estacional para caracterizar fluctuaciones periódicas que se superponen en el tiempo. Ahora bien, a la hora de aplicar esta metodología han surgido problemas de índole computacional. Sin duda, lo más correcto es plantear como punto de partida en la especificación del modelo la naturaleza estocástica para cada una de las fluctuaciones periódicas detectadas; sin embargo, las limitaciones computacionales obligan a adoptar supuestos simplificadores y confirmar, *a posteriori*, la veracidad de éstos o, en caso de refutación, evaluar la conveniencia de reespecificar el modelo en términos de reducción de la magnitud del error cometido y de incremento de la complejidad.

Si, teniendo en cuenta las limitaciones anteriormente señaladas, se asume que el modelo finalmente propuesto es aceptable, los resultados de la estimación del mismo desvelan algunas características relevantes de la demanda horaria de energía eléctrica. En primer lugar, destaca su carácter enormemente complejo y fluctuante, lo que, de entrada, supone un *handicap* para su predicción; en cualquier caso, y desde este mismo punto de vista, la consideración por separado de las fluctuaciones estacionales de periodo diario, semanal y anual constituye una guía para planificar la oferta y, en particular, permite identificar los puntos álgidos de la demanda según la hora del día, la semana o el año, así como los momentos en los que las demandas son mucho menores. En concreto, la evolución de la demanda presenta una fuerte estacionalidad en el plano anual que, posiblemente motivada por factores climatológicos y sociales tales como la temperatura, la luminosidad o las vacaciones, justifica las diferencias entre el invierno y el verano. A corto plazo, se aprecia que la demanda en cada una de las horas de los días festivos es más moderada que en los días laborables. En general, a lo largo de un día, la demanda es más baja en horas nocturnas (valle), comenzando a subir en las primeras horas de la mañana y manteniéndose a partir del medio día (llano), para crecer de nuevo por la tarde en aquellas horas en que los usos domésticos coinciden con los usos industriales (punta).

Aunque una base de datos con frecuencia superior a la horaria podría proporcionar información adicional, ésta no se encuentra disponible y, en caso de

existir, las fluctuaciones analizadas parecen más relevantes que las que pudieran registrarse, por ejemplo, en los diferentes minutos de una hora. Sin embargo, y teniendo en cuenta que el consumo de energía eléctrica depende, entre otros factores, del clima —que influye en las necesidades de calefacción y refrigeración—, de las diferencias en los usos horarios (Península-Canarias) —que inciden en las necesidades lumínicas— y de las principales actividades económicas desarrolladas, parece razonable pensar que existen diferencias en el perfil estacional anual de las distintas zonas geográficas del territorio nacional. Por tanto, la disponibilidad de datos más desagregados espacialmente permitiría efectuar, con la metodología ya descrita, un análisis que facilite la gestión centralizada por parte de la REE, de modo que las interconexiones entre las centrales ayuden a lograr mayor eficiencia en el suministro. A partir de este análisis podrían detectarse comportamientos diferenciados en los patrones de consumo semanal y diario, pero sobre todo, por la influencia de los factores a los que se ha aludido, cabe esperar que sea preciso utilizar diferentes especificaciones para la fluctuación anual.

Además, y como ya se ha comentado a lo largo del texto, las fluctuaciones anuales en la demanda de energía eléctrica podrían estar más directamente relacionadas con la temperatura que con la propia época del año. Esta circunstancia aconsejaría introducir la temperatura como variable explicativa adicional de la demanda horaria de energía eléctrica. Obviamente, la sustitución de la *spline* anual por este regresor supone una simplificación metodológica que impide ilustrar, por una parte, la utilidad de estas funciones para recoger la fluctuación estacional de periodo anual presente en una serie horaria y, por otra, el procedimiento de combinación de dos funciones *splines* para recoger variaciones periódicas simultáneas. Pero, si el objetivo no es tanto metodológico como explicativo de las variaciones registradas en la demanda horaria, la introducción de la temperatura, mediante una especificación lineal o no, puede resultar más adecuada.

Finalmente, el conocimiento del comportamiento estacional de la demanda por sectores económicos y zonas geográficas, además de permitir una planificación precisa de la generación de energía, proporciona la información necesaria para efectuar intercambios de fuentes de generación de energía que cubran la demanda con menor coste económico y medio ambiental.



## Conclusiones y extensiones

La esencia de los modelos estructurales es permitir que las señales subyacentes que determinan la evolución de la serie temporal, es decir, tendencia, ciclo y variaciones estacionales, posean, en principio, naturaleza estocástica. De este modo, y sin pretender identificar las razones concretas que explican las variaciones de la serie, las reacciones que en el corto y largo plazo experimentan estas señales ante cambios en los elementos que configuran el entorno de la magnitud estudiada pueden captarse de manera suficientemente flexible y automática con las especificaciones propias de este enfoque.

En consecuencia, las ventajas que ofrece este planteamiento frente a otras especificaciones son más acentuadas cuando las inestabilidades son más pronunciadas y, sobre todo, más continuas. Aunque la observación de la serie es el primer indicador de la variabilidad de ésta, el filtro de Kalman y la estimación máximo verosímil descomponen esa variabilidad y la asignan a los diferentes componentes de la serie en función de la magnitud estimada de las varianzas de los respectivos términos de perturbación. A través de esta descomposición, el analista es capaz de identificar qué señales son responsables de la evolución estocástica de la serie y qué otras obedecen a patrones determinísticos. Ahora bien, la naturaleza estocástica de alguno de los componentes depende de la longitud en el tiempo del periodo analizado y de la frecuencia con la que se registran las observaciones.

Teniendo en cuenta que el análisis efectuado con un cierto nivel de desagregación temporal proporciona información adicional relevante para los agentes económicos que actúan en el mercado, en este trabajo se ha pretendido proporcionar una herramienta metodológica que, integrada en el ámbito de los modelos estructurales, permite analizar el componente estacional de un amplio conjunto de series temporales económicas.

Cuando el componente estacional puede ser estimado a través de especificaciones convencionales, las funciones *splines* pueden actuar como formulación sustitutiva de las primeras, utilizando, sin embargo, un número más reducido de regresores para explicar el mismo comportamiento, como sucedió en el caso de la serie

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

semanal de exportaciones de tomate. Previsiblemente, la simplificación en la especificación del modelo será más acentuada cuanto mayor sea la longitud del periodo estacional, en tanto que cuando aumenta ésta, desciende el porcentaje que, en relación al número de estaciones, representa el número de regresores correspondientes a la *spline* e incorporados en el modelo estructural finalmente estimado. En las series analizadas, la *spline* anual se especificó introduciendo nueve regresores en el caso de la serie semanal, diecisiete para la serie diaria y catorce en la serie horaria, mientras que el número de estaciones por año es de cincuenta y dos, trescientas sesenta y cinco —trescientas sesenta y seis en un año bisiesto— y ocho mil setecientos sesenta —ocho mil setecientos ochenta y cuatro en un año bisiesto—, respectivamente; la *spline* semanal de la serie horaria necesitó cinco regresores para captar una fluctuación que se completa en ciento sesenta y ocho horas.

Como se pone de manifiesto en las series diaria y horaria analizadas, las funciones *splines* adquieren su verdadero potencial en aquellas ocasiones en las que las especificaciones convencionales no resultan apropiadas. Especialmente en el caso de series de alta frecuencia, la superposición de fluctuaciones estacionales de distinto periodo es difícil de abordar con el *software* estadístico disponible, que, por una parte, puede no contemplar la posibilidad de especificar un modelo con esta característica si dos o más de estas fluctuaciones poseen naturaleza estocástica y, por otra, puede carecer de capacidad computacional para admitir el número de regresores que exige la especificación determinística de tales variaciones. En estos casos, las *splines* se convierten en una alternativa necesaria y actúan como complemento a formulaciones determinísticas o estocásticas más habituales. Además, la estimación conjunta de fluctuaciones periódicas exige la especificación adecuada del modelo y, para ello, conviene aislar cada una de estas fluctuaciones. En este sentido, las funciones *splines* pueden ser un instrumento útil para obtener, cuando menos, aproximaciones iniciales de estos comportamientos.

En la serie diaria de visitas a un servicio de urgencias, la frecuencia de observación permitía considerar la presencia de dos fluctuaciones estacionales, y, dada la longitud del periodo asociado a la variación anual, la especificación de la *spline* correspondiente pudo obtenerse a partir de una aproximación inicial a esta fluctuación basada en el cálculo de las medias por día del año de la serie filtrada de la variación estacional de periodo semanal estimada mediante variables cualitativas. Y, sólo una vez especificada la *spline* anual, en la estimación conjunta del modelo estructural se

incorporó la formulación estocástica tanto del nivel como de la fluctuación estacional de menor periodo.

El procedimiento anterior no pudo ser aplicado, sin embargo, para la serie horaria de demanda de energía eléctrica debido a la complejidad adicional que supone la presencia de tres fluctuaciones periódicas superpuestas. En este caso, la longitud del periodo asociado a dos de estas fluctuaciones aconsejaba, al menos *a priori*, especificar un modelo en el que las variaciones estacionales de periodos anual y semanal fueran consideradas como determinísticas y recogidas a través de las correspondientes *splines*, reservándose la naturaleza estocástica para la variación de periodo diario. Por tanto, la especificación de ambas funciones *splines* exigía obtener una aproximación preliminar de cada una de las fluctuaciones estacionales. En el caso de la *spline* anual, esta aproximación se obtuvo a partir de las medias por hora del año calculadas sobre las medias móviles de periodo ciento sesenta y ocho de la serie filtrada de tendencia lineal determinística. Mientras que, para obtener una aproximación de la variación semanal, fue preciso filtrar la serie original de la estimación de la variación anual obtenida a través de la *spline* y luego calcular medias por hora de la semana sobre la serie de medias móviles de periodo veinticuatro de la serie anterior.

Además, la tres series examinadas poseen un componente estacional que, dada la frecuencia de observación, introduce elementos que dificultan la comparación de fluctuaciones de periodo anual en años diferentes. Cuando se dispone de datos semanales, esta fluctuación no se completa en un número entero de semanas, de modo que a cada año se asignan cincuenta y dos o cincuenta y tres. Pero las especiales características del patrón exportador de tomate canario permiten efectuar un ajuste artificial adecuado en la extensión del periodo sin exportación de aquellas campañas a las que se habían asignado más observaciones. Dicho ajuste no implica, sin embargo, la comparación de periodos del año más heterogéneos que si se prescinde de tal corrección y, al propio tiempo, hace posible la especificación del componente estacional con formulaciones convencionales o con funciones *splines*. Estas últimas vuelven a convertirse en una aproximación más útil que las primeras cuando ajustes como el anterior carecen de sentido, como sucede con las otras dos series o, en general, en series diarias u horarias que registren variaciones estacionales de periodo anual en años bisiestos y no bisiestos. En estos casos, el periodo en el que se completa la fluctuación estacional no es de longitud fija y la especificación de la *spline*, que determina la variación estacional en una observación correspondiente a una determinada estación

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

como función del número de estaciones que han transcurrido desde el momento en que se considera que se inicia la fluctuación estacional, introduce una flexibilidad que hace posible incorporar las adaptaciones específicas que cada caso requiere.

En el caso de la serie diaria, la solución al problema de heterogeneidad causado por la presencia de dos años bisiestos fue bastante inmediata y consistió simplemente en considerar que la magnitud de la variación estacional de periodo anual correspondiente al veintinueve de febrero de 1996 y 2000 era igual a la estimada a través de la *spline* para el día anterior; lo que obligaba, por supuesto, a corregir oportunamente los valores ajustados por la *spline* con objeto de que estas estimaciones pudieran ser interpretadas como componente estacional. De nuevo, la complejidad es mayor en el caso de la serie horaria. A pesar de la presencia de un solo año bisiesto, el cambio en la longitud del periodo en el que se completa la fluctuación anual es mayor y las observaciones correspondientes a las veinticuatro horas del día veintinueve de febrero de 2000 requerían un tratamiento específico. La solución adoptada consistió en introducir un tramo adicional en la especificación de la *spline* correspondiente a un año no bisiesto y diseñar procedimientos capaces de proporcionar una estimación de la variación estacional anual en un año bisiesto; bien, a partir de los datos correspondientes al 29 de febrero de 2000 y de las medias para las restantes horas del año —filtradas de la variación estacional semanal y diaria—, o bien, mediante la definición de regresores adecuados que, incorporados en el modelo estructural, ofrezcan la posibilidad de estimar esta variación anual junto al resto de componentes de la serie.

Las conclusiones metodológicas de la investigación emprendida sugieren nuevas líneas de trabajo. En primer lugar, algunas de las limitaciones en la aplicación de la propuesta realizada podrían ser superadas si se dispusiera de un periodo de observación de alta frecuencia más largo. En series que posean esta característica sería factible detectar el carácter estocástico de fluctuaciones estacionales que, observadas en periodos más cortos, pueden ser consideradas como determinísticas, ya que no existe margen temporal para observar inestabilidades suaves, como tal vez ha ocurrido en las series analizadas en este trabajo. Aunque, desde el punto de vista práctico, la posibilidad que ofrecen los datos no puede materializarse sin superar previamente las limitaciones de los programas informáticos disponibles, que impiden la estimación de formulaciones estocásticas para recoger variaciones estacionales de periodo muy largo —por ejemplo, el periodo anual en series horarias— y que, como ya se ha comentado, ni siquiera contemplan la estimación de dos o más fluctuaciones estacionales estocásticas, sea cual

sea el periodo asociado a ellas. En este sentido, y aunque cambios en el diseño de los programas pueden ser suficientes para considerar conjuntamente varias fluctuaciones estacionales, la implementación de funciones *splines* estocásticas constituiría, sin duda, una alternativa computacionalmente eficiente.

Por otro lado, una serie mucho más larga permitiría detectar la existencia de comportamientos cíclicos, ausentes en cada una de las series analizadas. En este último caso se podría trabajar con un modelo estructural en el que interviniesen todos los componentes inobservables típicos de una serie temporal y, al mismo tiempo, sería posible advertir otra de las ventajas de las funciones *splines*. Aunque en la exposición se ha enfocado la utilidad de las *splines* periódicas hacia la especificación del componente estacional, estas funciones son, obviamente, apropiadas para recoger cualquier fluctuación periódica y, por tanto, también para el componente cíclico. Las formulaciones habituales de este componente a partir de funciones trigonométricas suponen, por las propias características de estas funciones, que los ciclos son simétricos, a pesar de que existen argumentos que ponen en duda este comportamiento. Sin embargo, la asimetría puede recogerse con facilidad si se recurre a las funciones *splines*.

Tampoco debe perderse de vista el uso cada vez más generalizado de las denominadas series de muy alta frecuencia, en las que, quizás, es necesario replantear el concepto de estacionalidad. Ahora bien, siempre que se conozca el instante del tiempo en el que se registra la observación y éste pueda ponerse en relación con un punto de referencia que sea el inicio del periodo en el que transcurre la fluctuación estacional, las funciones *splines* constituyen tal vez el enfoque más apropiado para modelar un componente estacional en el que no se disponga de datos correspondientes a algunas de las estaciones.

La investigación realizada posee una orientación metodológica y los ámbitos de aplicación se han elegido con la pretensión de ilustrar la utilidad de las funciones *splines* como instrumento capaz de recoger el comportamiento estacional de series de alta frecuencia y mostrar cómo el recurso a estas funciones en el marco de los modelos estructurales es más ventajoso frente a otras especificaciones cuanto mayor es la frecuencia de observación. Por esta razón, la profundidad del análisis sustantivo de cada una de las series consideradas a partir del conocimiento cualitativo de la realidad bajo estudio en cada caso ha quedado encorsetada o supeditada por la persecución de la claridad expositiva del funcionamiento de las herramientas examinadas.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

Ahora bien, como ya se ha comentado, parece interesante abordar cuestiones como la estimación de la relación entre oferta y precio del tomate canario exportado a Europa, la extensión del estudio del número de visitas a diversos centros de urgencias del territorio nacional o la consideración de regresores como la temperatura para explicar el componente estacional de la demanda horaria de energía eléctrica en España.

Algunas de estas líneas de trabajo exigirían recurrir a los modelos estructurales multivariantes. En particular, el estudio de la relación entre exportaciones de tomate y cotizaciones podría reportar interesantes conclusiones si se examina la existencia de tendencias comunes y, sobre todo, de paralelismos en el comportamiento estacional de ambas series.

Además, puede ser informativo estudiar la presencia de efectos de calendario, sobre todo en las dos últimas series estudiadas. Con respecto a la serie de exportaciones, es preciso tener en cuenta que los envíos de tomate canario a Europa se producen cada semana en determinados días en función, básicamente, del grado de madurez de la fruta y de los precios que se estén registrando. Los productores intentan planificar la producción a lo largo de la campaña, pero, obviamente, se trata de un producto agrario, cuya oferta no puede programarse con precisión y es difícil que se produzcan cambios en los volúmenes de exportación en función de que en una semana concreta exista un número dado de días festivos en los mercados de destino. Más bien, los desajustes oferta-demanda que aparezcan como consecuencia de la existencia de días festivos en los que se cierran los mercados, pueden originar variaciones relevantes en la serie de precios.

En las series diaria y horaria examinadas en los capítulos anteriores, determinadas fiestas móviles o, incluso, fiestas con fecha fija en el calendario no corresponden siempre al mismo estadio estacional de una determinada fluctuación periódica, por ejemplo semanal. El impacto diferenciado de estas fiestas puede conducir a que aparezcan observaciones anómalas. En este sentido, algunas de las intervenciones utilizadas en la serie diaria de visitas a un centro de urgencias se introdujeron para recoger el comportamiento diferente de los días primero de mayo, así como de varios lunes de Pascua, día festivo en Barcelona. Cuando este tipo de efectos no son suficientemente acentuados como para exigir el recurso a *dummies* determinísticas, probablemente estén siendo absorbidos por la parte estocástica del modelo estructural especificado, como pudo haber ocurrido en la serie horaria de demanda de energía eléctrica. En cualquier caso, los efectos registrados en las cercanías del periodo de

Semana Santa deberían ser explícitamente examinados, sobre todo si se extiende la longitud del periodo de análisis. En el caso de otras fiestas, la coexistencia de fluctuaciones anual y semanal puede causar problemas de multicolinealidad si se intentan capturar determinados efectos de calendario a través de *dummies*. Por ejemplo, el día de fin de año no es siempre el mismo día de la semana, pero si se intenta recoger este efecto con variables de intervención, existirá una evidente correlación con el efecto estacional anual correspondiente a ese día. Todas estas consideraciones convierten el examen de los efectos de calendario en otro sugerente campo de estudio.



## **Anexos**



## Anexo 1.1. Ejemplos de la especificación del modelo de tendencias comunes

A continuación se especifican las tendencias comunes en un modelo de nivel local con deriva en dos situaciones diferentes.

### Ejemplo 1

Sea un vector de variables observables de dimensión  $2 \times 1$   $\mathbf{y}_t = (y_{1,t}, y_{2,t})'$ ,  $N = 2$ , y una tendencia común,  $k = 1$ . Entonces, el modelo de nivel local con deriva se puede representar de la forma

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{q}_{2,1} \end{pmatrix} \mathbf{m}_t^* + \begin{pmatrix} 0 \\ \mathbf{m}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{e}_{1,t} \\ \mathbf{e}_{2,t} \end{pmatrix}$$

donde  $\mathbf{m}_t^* = \mathbf{m}_{t-1}^* + \mathbf{b} + \mathbf{h}_t^*$ . Es decir,

$$y_{1,t} = \mathbf{m}_t^* + \mathbf{e}_{1,t}$$

$$y_{2,t} = \mathbf{q}_{2,1} \mathbf{m}_t^* + \mathbf{m}_2 + \mathbf{e}_{2,t}$$

donde  $\mathbf{m}_t^*$  se ha definido anteriormente. Por tanto,

$$y_{1,t} - \frac{1}{\mathbf{q}_{2,1}} y_{2,t} = -\frac{\mathbf{m}_2}{\mathbf{q}_{2,1}} + \mathbf{e}_{1,t} - \frac{1}{\mathbf{q}_{2,1}} \mathbf{e}_{2,t}$$

es una combinación lineal estacionaria, lo que implica cointegración. El vector de cointegración  $\begin{pmatrix} 1, -\frac{1}{\mathbf{q}_{2,1}} \end{pmatrix}$  es tal que  $\begin{pmatrix} 1, -\frac{1}{\mathbf{q}_{2,1}} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{q}_{2,1} \end{pmatrix} = 0$ .

### Ejemplo 2

Sea un vector de variables observables de dimensión  $3 \times 1$   $\mathbf{y}_t = (y_{1,t}, y_{2,t}, y_{3,t})'$ , con  $N = 3$  y dos tendencias comunes,  $k = 2$ . Entonces, el modelo de nivel local con deriva se puede representar de la forma

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \mathbf{q}_{2,1} & 1 \\ \mathbf{q}_{3,1} & \mathbf{q}_{3,2} \end{pmatrix} \begin{pmatrix} \mathbf{m}_{1,t}^* \\ \mathbf{m}_{2,t}^* \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \mathbf{m}_3 \end{pmatrix} + \begin{pmatrix} \mathbf{e}_{1,t} \\ \mathbf{e}_{2,t} \\ \mathbf{e}_{3,t} \end{pmatrix}$$

donde

$$\begin{pmatrix} \mathbf{m}_{1,t}^* \\ \mathbf{m}_{2,t}^* \end{pmatrix} = \begin{pmatrix} \mathbf{m}_{1,t-1}^* \\ \mathbf{m}_{2,t-1}^* \end{pmatrix} + \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{h}_{1,t} \\ \mathbf{h}_{2,t} \end{pmatrix}$$

Es decir,

$$y_{1,t} = \mathbf{m}_{1,t}^* + \mathbf{e}_{1,t},$$

$$y_{2,t} = \mathbf{q}_{2,1} \mathbf{m}_{1,t}^* + \mathbf{m}_{2,t}^* + \mathbf{e}_{2,t}$$

y

$$y_{3,t} = \mathbf{q}_{3,1} \mathbf{m}_{1,t}^* + \mathbf{q}_{3,2} \mathbf{m}_{2,t}^* + \mathbf{m}_3 + \mathbf{e}_{3,t},$$

donde  $\mathbf{m}_{1,t}^* = \mathbf{m}_{1,t-1}^* + \mathbf{b}_1 + \mathbf{h}_{1,t}^*$  y  $\mathbf{m}_{2,t}^* = \mathbf{m}_{2,t-1}^* + \mathbf{b}_2 + \mathbf{h}_{2,t}^*$ . Por tanto, la matriz de vectores de cointegración  $\mathbf{A}_{1 \times 3}$  será tal que  $\mathbf{A}\mathbf{Q} = \mathbf{0}$ , es decir,

$$\begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \mathbf{q}_{2,1} & 1 \\ \mathbf{q}_{3,1} & \mathbf{q}_{3,2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}.$$

De modo que  $a_1 + a_2 \mathbf{q}_{2,1} + a_3 \mathbf{q}_{3,1} = 0$  y  $a_2 + a_3 \mathbf{q}_{3,2} = 0$ . Si se toma  $a_1 = 1$ , resulta que  $a_2 = -a_3 \mathbf{q}_{3,2}$  y  $1 - a_3 \mathbf{q}_{3,2} \mathbf{q}_{2,1} + a_3 \mathbf{q}_{3,1} = 0$ , lo que implica que  $a_3 = 1 / (\mathbf{q}_{3,2} \mathbf{q}_{2,1} - \mathbf{q}_{3,1})$ . Por tanto, el vector de cointegración viene dado por

$$\begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix} = \begin{pmatrix} 1, & -\frac{\mathbf{q}_{3,2}}{\mathbf{q}_{3,2} \mathbf{q}_{2,1} - \mathbf{q}_{3,1}}, & \frac{1}{\mathbf{q}_{3,2} \mathbf{q}_{2,1} - \mathbf{q}_{3,1}} \end{pmatrix}.$$

## Anexo 1.2. Propiedades estadísticas de un modelo en el espacio de los estados

Sea un sistema descrito por un vector  $m \times 1$  de estados  $\mathbf{a}_t$ , en los instantes del tiempo  $t = 1, \dots, T$  y sea el vector  $N \times 1$  de variables observables  $\mathbf{y}_t$ . El modelo básico de la representación en el espacio de los estados consta de la ecuación de transición,

$$\mathbf{a}_t = \mathbf{S}\mathbf{a}_{t-1} + \mathbf{u}_t, \quad t = 1, \dots, T,$$

y de la ecuación de medida

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{a}_t + \mathbf{e}_t, \quad t = 1, \dots, T,$$

donde  $\mathbf{S}$  es una matriz de transición  $m \times m$ ,  $\mathbf{u}_t$  es el vector  $m \times 1$  de perturbaciones de la ecuación de transición tal que  $\mathbf{u}_t \sim i.i.d. N_m(\mathbf{0}, \mathbf{Q})$ ,  $\mathbf{Z}_t$  es una matriz de medida  $N \times m$ , y  $\mathbf{e}_t$  es el vector  $N \times 1$  de perturbaciones de la ecuación de medida tal que  $\mathbf{e}_t \sim i.i.d. N_N(\mathbf{0}, \mathbf{H})$ .

Se asume que  $E[\mathbf{a}_0] = \mathbf{a}_0$ ;  $Var[\mathbf{a}_0] = \mathbf{P}_0$ ;  $E[\mathbf{e}_t \mathbf{u}_s'] = \mathbf{0}$ ,  $\forall t, s = 1, \dots, T$ ;  $E[\mathbf{e}_t \mathbf{a}_0'] = \mathbf{0}$ ,  $t = 1, \dots, T$  y  $E[\mathbf{u}_t \mathbf{a}_0'] = \mathbf{0}$ ,  $t = 1, \dots, T$ .

Por sustitución recursiva, la ecuación de transición se puede expresar como

$$\mathbf{a}_t = \mathbf{S}\mathbf{a}_{t-1} + \mathbf{u}_t = \mathbf{S}(\mathbf{S}\mathbf{a}_{t-2} + \mathbf{u}_{t-1}) + \mathbf{u}_t = \mathbf{S}^2\mathbf{a}_{t-2} + \mathbf{S}\mathbf{u}_{t-1} + \mathbf{u}_t = \dots = \mathbf{S}^t\mathbf{a}_0 + \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{u}_{t-i}$$

y la ecuación de medida como

$$\mathbf{y}_t = \mathbf{Z}_t \left[ \mathbf{S}^t \mathbf{a}_0 + \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{u}_{t-i} \right] + \mathbf{e}_t$$

A) Momentos de primer y segundo orden de las variables de estado  $\mathbf{a}_t$  y de las variables observables  $\mathbf{y}_t$

$$- E[\mathbf{a}_t] = \mathbf{S}^t E[\mathbf{a}_0] = \mathbf{S}^t \mathbf{a}_0$$

$$- E[\mathbf{y}_t] = \mathbf{Z}_t \mathbf{S}^t E[\mathbf{a}_0] = \mathbf{Z}_t \mathbf{S}^t \mathbf{a}_0$$

$$- \text{Para calcular } Cov[\mathbf{a}_t, \mathbf{a}_{t-k}] = E\left[ (\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])' \right], \text{ se puede tener en}$$

cuenta que

$$\mathbf{a}_t = \mathbf{S}^k \mathbf{a}_{t-k} + \sum_{i=0}^{k-1} \mathbf{S}^i \mathbf{u}_{t-i}$$

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

y, entonces,

$$\begin{aligned}
 E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] &= E\left[\left(\mathbf{S}^k \mathbf{a}_{t-k} + \sum_{i=0}^{k-1} \mathbf{S}^i \mathbf{u}_{t-i} - \mathbf{S}^k E[\mathbf{a}_{t-k}]\right)(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \\
 &= E\left[\left(\mathbf{S}^k (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}]) + \sum_{i=0}^{k-1} \mathbf{S}^i \mathbf{u}_{t-i}\right)(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \\
 &= E\left[\mathbf{S}^k (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \\
 &\quad + E\left[\left(\sum_{i=0}^{k-1} \mathbf{S}^i \mathbf{u}_{t-i}\right)(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right]
 \end{aligned}$$

Considerando entonces que

$$\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}] = \mathbf{S}^{t-k} \mathbf{a}_0 + \sum_{i=0}^{t-k-1} \mathbf{S}^i \mathbf{u}_{t-k-i} - \mathbf{S}^{t-k} \mathbf{a}_0 = \mathbf{S}^{t-k} (\mathbf{a}_0 - \mathbf{a}_0) + \sum_{i=0}^{t-k-1} \mathbf{S}^i \mathbf{u}_{t-k-i},$$

resulta que

$$\begin{aligned}
 &E\left[(\mathbf{a}_t - E[\mathbf{a}_t])(\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])'\right] \\
 &= E\left[\left(\mathbf{S}^k \left(\mathbf{S}^{t-k} (\mathbf{a}_0 - \mathbf{a}_0) + \sum_{i=0}^{t-k-1} \mathbf{S}^i \mathbf{u}_{t-k-i}\right)\right) \left(\mathbf{S}^{t-k} (\mathbf{a}_0 - \mathbf{a}_0) + \sum_{i=0}^{t-k-1} \mathbf{S}^i \mathbf{u}_{t-k-i}\right)'\right] \\
 &\quad + E\left[\left(\sum_{i=0}^{k-1} \mathbf{S}^i \mathbf{u}_{t-i}\right) \left(\mathbf{S}^{t-k} (\mathbf{a}_0 - \mathbf{a}_0) + \sum_{i=0}^{t-k-1} \mathbf{S}^i \mathbf{u}_{t-k-i}\right)'\right] \\
 &= \mathbf{S}^t E\left[(\mathbf{a}_0 - \mathbf{a}_0)(\mathbf{a}_0 - \mathbf{a}_0)'\right] (\mathbf{S}^{t-k})' + \mathbf{S}^k \sum_{i=0}^{t-k-1} \mathbf{S}^i E\left[u_{t-k-i} (\mathbf{a}_0 - \mathbf{a}_0)'\right] (\mathbf{S}^{t-k})' \\
 &\quad + \mathbf{S}^t \sum_{i=0}^{t-k-1} E[(\mathbf{a}_0 - \mathbf{a}_0) u'_{t-k-i}] (\mathbf{S}^i)' \\
 &\quad + \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} E[u_{t-k-i} u'_{t-k-i}] (\mathbf{S}^i)' + \sum_{\substack{i,j=0 \\ i \neq j}}^{t-k-1} \mathbf{S}^{k+i} E[u_{t-k-i} u'_{t-k-j}] (\mathbf{S}^j)' \\
 &\quad + \sum_{i=0}^{k-1} \mathbf{S}^i E\left[u_{t-i} (\mathbf{a}_0 - \mathbf{a}_0)'\right] (\mathbf{S}^{t-k})' + \sum_{i=0}^{k-1} \sum_{j=0}^{t-k-1} \mathbf{S}^i E[u_{t-i} u'_{t-k-j}] (\mathbf{S}^j)' \\
 &= \mathbf{S}^t \text{Var}[\mathbf{a}_0] (\mathbf{S}^{t-k})' + \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \text{Var}[u_{t-k-i}] (\mathbf{S}^i)' = \mathbf{S}^t \mathbf{P}_0 (\mathbf{S}^{t-k})' + \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \mathbf{Q} (\mathbf{S}^i)' \\
 &= \left[ \mathbf{S}^t \mathbf{P}_0 (\mathbf{S}^{t-k})' + \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \mathbf{Q} (\mathbf{S}^i)' \right] = \mathbf{P}_{t,t-k}
 \end{aligned}$$

$$\begin{aligned}
& - E \left[ (y_t - E[y_t])(y_{t-k} - E[y_{t-k}])' \right] \\
& = E \left[ (Z_t \mathbf{a}_t + \mathbf{e}_t - Z_t E[\mathbf{a}_t])(Z_{t-k} \mathbf{a}_{t-k} + \mathbf{e}_{t-k} - Z_{t-k} E[\mathbf{a}_{t-k}])' \right] \\
& = E \left[ (Z_t (\mathbf{a}_t - E[\mathbf{a}_t]) + \mathbf{e}_t)(Z_{t-k} (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}]) + \mathbf{e}_{t-k})' \right] \\
& = E \left[ Z_t (\mathbf{a}_t - E[\mathbf{a}_t]) (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])' Z_{t-k}' \right] + E \left[ \mathbf{e}_t (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])' Z_{t-k}' \right] \\
& \quad + E \left[ Z_t (\mathbf{a}_t - E[\mathbf{a}_t]) \mathbf{e}_{t-k}' \right] + E \left[ \mathbf{e}_t \mathbf{e}_{t-k}' \right] \\
& = Z_t E \left[ (\mathbf{a}_t - E[\mathbf{a}_t]) (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])' \right] Z_{t-k}' + E \left[ \mathbf{e}_t (\mathbf{a}_{t-k} - E[\mathbf{a}_{t-k}])' \right] Z_{t-k}' \\
& \quad + Z_t E \left[ (\mathbf{a}_t - E[\mathbf{a}_t]) \mathbf{e}_{t-k}' \right] + E \left[ \mathbf{e}_t \mathbf{e}_{t-k}' \right] \\
& = Z_t \mathbf{P}_{t,t-k} Z_{t-k}', \text{ si } k \neq 0 \\
& - E \left[ (y_t - E[y_t])(y_t - E[y_t])' \right] \\
& = E \left[ (Z_t \mathbf{a}_t + \mathbf{e}_t - Z_t E[\mathbf{a}_t])(Z_t \mathbf{a}_t + \mathbf{e}_t - Z_t E[\mathbf{a}_t])' \right] \\
& = E \left[ (Z_t (\mathbf{a}_t - E[\mathbf{a}_t]) + \mathbf{e}_t)(Z_t (\mathbf{a}_t - E[\mathbf{a}_t]) + \mathbf{e}_t)' \right] \\
& = E \left[ Z_t (\mathbf{a}_t - E[\mathbf{a}_t]) (\mathbf{a}_t - E[\mathbf{a}_t])' Z_t' \right] + E \left[ \mathbf{e}_t (\mathbf{a}_t - E[\mathbf{a}_t])' Z_t' \right] \\
& \quad + E \left[ Z_t (\mathbf{a}_t - E[\mathbf{a}_t]) \mathbf{e}_t' \right] + E \left[ \mathbf{e}_t \mathbf{e}_t' \right] \\
& = Z_t E \left[ (\mathbf{a}_t - E[\mathbf{a}_t]) (\mathbf{a}_t - E[\mathbf{a}_t])' \right] Z_t' + E \left[ \mathbf{e}_t (\mathbf{a}_t - E[\mathbf{a}_t])' \right] Z_t' \\
& \quad + Z_t E \left[ (\mathbf{a}_t - E[\mathbf{a}_t]) \mathbf{e}_t' \right] + E \left[ \mathbf{e}_t \mathbf{e}_t' \right] \\
& = Z_t \mathbf{P}_t Z_t' + \mathbf{H}
\end{aligned}$$

Sustituyendo  $\mathbf{P}_{t,t-k}$  y  $\mathbf{P}_t$  por sus valores respectivos, se tiene que

$$\begin{aligned}
& E \left[ (y_t - E[y_t])(y_{t-k} - E[y_{t-k}])' \right] \\
& = Z_t \left[ \mathbf{S}' \mathbf{P}_0 (\mathbf{S}^{t-k})' + \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \mathbf{Q} (\mathbf{S}^i)' \right] (Z_{t-k}') \\
& = \left[ Z_t \mathbf{S}' \mathbf{P}_0 (Z_{t-k} \mathbf{S}^{t-k})' + Z_t \sum_{i=0}^{t-k-1} \mathbf{S}^{k+i} \mathbf{Q} (Z_{t-k} \mathbf{S}^i)' \right], \text{ si } k \neq 0
\end{aligned}$$

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

y que

$$\begin{aligned} & E\left[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_t - E[\mathbf{y}_t])'\right] \\ &= \left[ \mathbf{Z}_t \left[ \mathbf{S}' \mathbf{P}_0 (\mathbf{S}^t)' + \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{Q} (\mathbf{S}^i)' \right] \mathbf{Z}_t' + \mathbf{H} \right] \\ &= \left[ \mathbf{Z}_t \mathbf{S}' \mathbf{P}_0 (\mathbf{Z}_t \mathbf{S}^t)' + \mathbf{Z}_t \sum_{i=0}^{t-1} \mathbf{S}^i \mathbf{Q} (\mathbf{Z}_t \mathbf{S}^i)' + \mathbf{H} \right] \end{aligned}$$

### B) Función de verosimilitud

Sean los vectores de vectores aleatorios  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$   $N \cdot T \times 1$ ,  $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_T)'$   $m \cdot T \times 1$  y  $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_T)'$   $N \cdot T \times 1$ , y la matriz de matrices  $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_T)$   $NT \times mT$ . Entonces  $\mathbf{y} = \mathbf{Z}\mathbf{a} + \mathbf{e}$ , donde  $\mathbf{e} \sim N_{NT}(\mathbf{0}, (\mathbf{I}_T \otimes \mathbf{H}))$ , con  $\text{Var}[\mathbf{e}]$  una matriz  $T \times T$  de matrices de orden  $N \times N$  definida como

$$\text{Var}[\mathbf{e}] = \begin{bmatrix} \mathbf{H} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H} \end{bmatrix}$$

Por otro lado, se tiene que  $\mathbf{a} = \mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{S}_2 \mathbf{u}$ , donde  $\mathbf{S}_1 = \text{diag}(\mathbf{S}, \mathbf{S}^2, \dots, \mathbf{S}^T)$ ,  $\mathbf{i}_T$  es el vector unitario  $T \times 1$ ,

$$\mathbf{S}_2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{S} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}^{T-1} & \mathbf{S}^{T-2} & \dots & \mathbf{I} \end{bmatrix}$$

y  $\mathbf{u} \sim N_{mT}(\mathbf{0}, (\mathbf{I}_T \otimes \mathbf{Q}))$ .

Nótese que

$$\begin{aligned} \mathbf{a}_1 &= \mathbf{S}\mathbf{a}_0 + \mathbf{u}_1, \\ \mathbf{a}_2 &= \mathbf{S}^2\mathbf{a}_0 + \mathbf{u}_2 + \mathbf{S}\mathbf{u}_1, \dots, \\ \mathbf{a}_T &= \mathbf{S}^T\mathbf{a}_0 + \mathbf{u}_T + \mathbf{S}\mathbf{u}_{T-1} + \dots + \mathbf{S}^{T-1}\mathbf{u}_1. \end{aligned}$$

Por tanto,

$$E[\mathbf{y}] = \mathbf{Z}E[\mathbf{a}] = \mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) = \mathbf{Z}[E[\mathbf{a}'_1], \dots, E[\mathbf{a}'_T]]' = \mathbf{Z}[(\mathbf{S}\mathbf{a}_0)', \dots, (\mathbf{S}^T \mathbf{a}_0)']' \equiv \mathbf{m}$$

Para calcular la varianza del vector de vectores  $\mathbf{y}$  conviene expresarlo de la forma

$$\mathbf{y} = \mathbf{Z}[\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{S}_2\mathbf{u}] + \mathbf{e}$$

Entonces,

$$\begin{aligned} & E[(\mathbf{y} - \mathbf{m})(\mathbf{y} - \mathbf{m})'] \\ &= E[(\mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) - \mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{Z}\mathbf{S}_2\mathbf{u} + \mathbf{e})(\mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) - \mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \otimes \mathbf{a}_0) + \mathbf{Z}\mathbf{S}_2\mathbf{u} + \mathbf{e})'] \\ &= E[\mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \mathbf{i}_T' \otimes (\mathbf{a}_0 - \mathbf{a}_0)(\mathbf{a}_0 - \mathbf{a}_0)')\mathbf{S}_1'\mathbf{Z}' + E[\mathbf{Z}\mathbf{S}_2\mathbf{u}\mathbf{u}'\mathbf{S}_2'\mathbf{Z}'] + E[\mathbf{e}\mathbf{e}']] \\ &= \mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \mathbf{i}_T' \otimes E[(\mathbf{a}_0 - \mathbf{a}_0)(\mathbf{a}_0 - \mathbf{a}_0)'])\mathbf{S}_1'\mathbf{Z}' + \mathbf{Z}\mathbf{S}_2E[\mathbf{u}\mathbf{u}']\mathbf{S}_2'\mathbf{Z}' + E[\mathbf{e}\mathbf{e}'] \\ &= \mathbf{Z}\mathbf{S}_1(\mathbf{i}_T \mathbf{i}_T' \otimes \mathbf{P}_0)\mathbf{S}_1'\mathbf{Z}' + \mathbf{Z}\mathbf{S}_2(\mathbf{I}_T \otimes \mathbf{Q})\mathbf{S}_2'\mathbf{Z}' + (\mathbf{I}_T \otimes \mathbf{H}) \equiv \mathbf{S} \end{aligned}$$

Como  $\mathbf{y}$  es normal multivariante de dimensión  $NT$ , con media  $\mathbf{m}$  y matriz de covarianzas  $\mathbf{S}$ , la función de verosimilitud es

$$L(\mathbf{y}) = \frac{\sqrt{|\mathbf{S}|^{-1}}}{(2\pi)^{NT/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{m})\mathbf{S}^{-1}(\mathbf{y} - \mathbf{m})'\right\}$$

y su logaritmo neperiano

$$\ln L(\mathbf{y}) = -\frac{1}{2} \ln |\mathbf{S}| - \frac{NT}{2} \ln 2\pi - \frac{1}{2}(\mathbf{y} - \mathbf{m})\mathbf{S}^{-1}(\mathbf{y} - \mathbf{m})'$$

A partir de la función de verosimilitud pueden derivarse estimaciones máximo verosímiles de los parámetros del modelo, aunque pueden surgir problemas de identificación. Si no se imponen restricciones en las matrices del sistema, los parámetros de la representación en el espacio de los estados pueden quedar sin identificar; es decir, puede ocurrir que más de un conjunto de valores de los parámetros de lugar a idéntico valor de la

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

función de verosimilitud y, por tanto, los datos no ofrecen una guía para elegir tales valores (Hamilton, 1994:387-388)<sup>1</sup>.

---

<sup>1</sup> Se dice que un vector de parámetros  $\theta$ , del cual depende la función de verosimilitud  $f(y; \theta)$  está identificado cuando  $f(y; \theta_1) = f(y; \theta_2) \Rightarrow \theta_1 = \theta_2$ . Pues bien, se puede demostrar que, bajo ciertas condiciones iniciales, cualquier transformación lineal del vector de estado no cambia las covarianzas ni la media de  $y_t$  y, por tanto, no modifica la función de verosimilitud. Esto quiere decir que los parámetros del sistema no pueden ser identificados por la función de verosimilitud (Steyn, 1996:12). Otter y Mittnik (1989) derivan las condiciones bajo las cuales la introducción de ciertas restricciones conduce a la identificación (Steyn, 1996:12-14).

## Anexo 1.3. Distribución del estado condicionada a las observaciones, $f(\mathbf{a}_t / y_t)$

Si

$$\begin{bmatrix} \mathbf{a}_t \\ y_t \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{S}' \mathbf{a}_0 \\ \mathbf{Z}_t \mathbf{S}' \mathbf{a}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{P}_t & \mathbf{P}_t \mathbf{Z}_t' \\ \mathbf{Z}_t \mathbf{P}_t & \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H} \end{bmatrix} \right),$$

entonces  $f(\mathbf{a}_t / y_t)$  sigue una distribución normal  $m$ -variante con vector de medias dado por

$$E \left[ \mathbf{a}_t / y_t \right] = E[\mathbf{a}_t] + \text{Cov}[\mathbf{a}_t, y_t] (\text{Var}[y_t])^{-1} (y_t - E[y_t]),$$

es decir,

$$E \left[ \mathbf{a}_t / y_t \right] = \mathbf{S}' \mathbf{a}_0 + \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H})^{-1} (y_t - \mathbf{Z}_t \mathbf{S}' \mathbf{a}_0).$$

Para obtener la matriz de varianzas-covarianzas, conviene tener en cuenta que si se definen las matrices

$$\mathbf{M}_t = \begin{bmatrix} \mathbf{P}_t & \mathbf{P}_t \mathbf{Z}_t' \\ \mathbf{Z}_t \mathbf{P}_t & \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H} \end{bmatrix}$$

y

$$\mathbf{W}_t = \mathbf{M}_t^{-1} = \begin{bmatrix} \mathbf{W}_{1,t} & \mathbf{W}_{3,t} \\ \mathbf{W}_{3,t}' & \mathbf{W}_{2,t} \end{bmatrix},$$

entonces<sup>1</sup>  $\text{Var} \left[ \mathbf{a}_t / y_t \right] = \mathbf{W}_{1,t}^{-1}$ , donde<sup>2</sup>

<sup>1</sup> Véase Zoroa y Zoroa (1991:277-279).

<sup>2</sup> Véase Johnston (1987:160).

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

$$\begin{aligned}
 W_{1,t} &= \text{Var}[\mathbf{a}_t]^{-1} \\
 &+ \text{Var}[\mathbf{a}_t]^{-1} \text{Cov}[\mathbf{a}_t, \mathbf{y}_t] \left( \text{Var}[\mathbf{y}_t] - \text{Cov}[\mathbf{y}_t, \mathbf{a}_t] \text{Var}[\mathbf{a}_t]^{-1} \text{Cov}[\mathbf{a}_t, \mathbf{y}_t] \right)^{-1} \text{Cov}[\mathbf{y}_t, \mathbf{a}_t] \left( \text{Var}[\mathbf{a}_t] \right)^{-1} \\
 &= \left[ \mathbf{P}_t^{-1} + \mathbf{P}_t^{-1} \mathbf{P}_t \mathbf{Z}_t' \left[ \mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H} - \mathbf{Z}_t \mathbf{P}_t \mathbf{P}_t^{-1} \mathbf{P}_t \mathbf{Z}_t' \right]^{-1} \mathbf{Z}_t \mathbf{P}_t \mathbf{P}_t^{-1} \right] \\
 &= \left[ \mathbf{P}_t^{-1} + \mathbf{Z}_t' \mathbf{H}^{-1} \mathbf{Z}_t \right]
 \end{aligned}$$

Por tanto,  $f\left(\frac{\mathbf{a}_t}{\mathbf{y}_t}\right)$  sigue una distribución normal  $m$ -variante con media

$$E\left[\frac{\mathbf{a}_t}{\mathbf{y}_t}\right] = \mathbf{S}' \mathbf{a}_0 + \mathbf{P}_t \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{P}_t \mathbf{Z}_t' + \mathbf{H})^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{S}' \mathbf{a}_0)$$

y matriz de covarianzas

$$\text{Var}\left[\frac{\mathbf{a}_t}{\mathbf{y}_t}\right] = \left[ \mathbf{P}_t^{-1} + \mathbf{Z}_t' \mathbf{H}^{-1} \mathbf{Z}_t \right]^{-1}.$$

Si  $\mathbf{a}_t$  es no estacionario y  $\mathbf{Z}_t$  es de rango pleno por columnas, se puede demostrar que  $f\left(\frac{\mathbf{a}_t}{\mathbf{y}_t}\right)$  sigue una distribución normal  $m$ -variante con media

$$E\left[\frac{\mathbf{a}_t}{\mathbf{y}_t}\right] = (\mathbf{Z}_t' \mathbf{Z}_t)^{-1} \mathbf{Z}_t' \mathbf{y}_t$$

y matriz de covarianzas

$$\text{Var}\left[\frac{\mathbf{a}_t}{\mathbf{y}_t}\right] = (\mathbf{Z}_t' \mathbf{Z}_t)^{-1} \mathbf{Z}_t' \mathbf{H} \mathbf{Z}_t (\mathbf{Z}_t' \mathbf{Z}_t)^{-1}.$$

Nótese que

$$\mathbf{Z}_t \mathbf{a}_t = \mathbf{y}_t - \mathbf{e}_t.$$

Por tanto,

$$\begin{aligned}
 \mathbf{Z}_t' \mathbf{Z}_t \mathbf{a}_t &= \mathbf{Z}_t' \mathbf{y}_t - \mathbf{Z}_t' \mathbf{e}_t \\
 (\mathbf{Z}_t' \mathbf{Z}_t)^{-1} \mathbf{Z}_t' \mathbf{Z}_t \mathbf{a}_t &= (\mathbf{Z}_t' \mathbf{Z}_t)^{-1} \mathbf{Z}_t' \mathbf{y}_t - (\mathbf{Z}_t' \mathbf{Z}_t)^{-1} \mathbf{Z}_t' \mathbf{e}_t \\
 \mathbf{a}_t &= (\mathbf{Z}_t' \mathbf{Z}_t)^{-1} \mathbf{Z}_t' \mathbf{y}_t - (\mathbf{Z}_t' \mathbf{Z}_t)^{-1} \mathbf{Z}_t' \mathbf{e}_t
 \end{aligned}$$

de modo que

$$E\left[\frac{\mathbf{a}_t}{\mathbf{y}_t}\right] = (\mathbf{Z}_t' \mathbf{Z}_t)^{-1} \mathbf{Z}_t' \mathbf{y}_t$$

y

$$\begin{aligned} \text{Var}\left[\frac{\mathbf{a}_t}{\mathbf{y}_t}\right] &= E\left[\left((\mathbf{Z}'_t\mathbf{Z}_t)^{-1}\mathbf{Z}'_t\mathbf{e}_t\right)\left(\mathbf{e}'_t\mathbf{Z}_t(\mathbf{Z}'_t\mathbf{Z}_t)^{-1}\right)\right] \\ &= (\mathbf{Z}'_t\mathbf{Z}_t)^{-1}\mathbf{Z}'_tE[\mathbf{e}_t\mathbf{e}'_t]\mathbf{Z}_t(\mathbf{Z}'_t\mathbf{Z}_t)^{-1} \\ &= (\mathbf{Z}'_t\mathbf{Z}_t)^{-1}\mathbf{Z}'_t\mathbf{H}\mathbf{Z}_t(\mathbf{Z}'_t\mathbf{Z}_t)^{-1} \end{aligned}$$

## **Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia**

## Anexo 2.1. Ilustración de los procedimientos de estimación de la localización de los nodos

Sea  $y_t$  una función desconocida, cuyos valores se observan en los instantes del tiempo  $t = 1, \dots, T$ , que se aproximan a través de la *spline*  $g(t)$ , definida por dos tramos,  $t_0 \leq t \leq t_1$  y  $t_1 \leq t \leq t_2$ , con punto de ruptura  $t_1$  desconocido<sup>1</sup>, de modo que

$$y_t = g(t) + \mathbf{e}_t,$$

siendo  $\mathbf{e}_t$  un término de error ruido blanco. A continuación se ilustrará el procedimiento de estimación no lineal de los parámetros libres de dicha *spline* en los casos lineal, cuadrático y cúbico.

### Caso I

Sea  $g(t)$  la función *spline* lineal definida por dos tramos, es decir,

$$g(t) = \begin{cases} g_{1,0} + g_{1,1}t, & t_0 \leq t \leq t_1 \\ g_{2,0} + g_{2,1}t, & t_1 \leq t \leq t_2 \end{cases}$$

Esta función puede expresarse como

$$g(t) = (g_{1,0} + g_{1,1}t)D_{1,t} + (g_{2,0} + g_{2,1}t)D_{2,t}$$

donde  $D_{1,t} = \begin{cases} 1, & t_0 \leq t < t_1 \\ 0, & \text{resto} \end{cases}$ ,  $D_{2,t} = \begin{cases} 1, & t_1 \leq t \leq t_2 \\ 0, & \text{resto} \end{cases}$ . La función *spline* lineal está sujeta

a la restricción de continuidad

$$g_{1,0} + g_{1,1}t_1 = g_{2,0} + g_{2,1}t_1$$

o también,

$$g_{2,1} = g_{1,1} + \frac{g_{1,0} - g_{2,0}}{t_1}$$

Si se introduce la restricción anterior, la función puede expresarse como

---

<sup>1</sup> Nótese que  $t_0 = 1$  y  $t_2 = T$ .

$$\begin{aligned}
 g(t) &= [g_{1,0} + g_{1,1}t] D_{1,t} + \left[ g_{2,0} + \left( g_{1,1} + \frac{g_{1,0} - g_{2,0}}{t_1} \right) t \right] D_{2,t} \\
 &= g_{1,0} \left( D_{1,t} + \frac{t}{t_1} D_{2,t} \right) + g_{2,0} \left( 1 - \frac{t}{t_1} \right) D_{2,t} + g_{1,1}t \\
 &= f(\mathbf{b})_t
 \end{aligned}$$

siendo  $\mathbf{b}$  el vector de parámetros  $[g_{1,0}, g_{1,1}, g_{2,0}, t_1]'$ . Una primera estimación de este vector puede obtenerse como

$$\mathbf{b}_{(1)} = \left[ \left[ \nabla f(\mathbf{b}_{(0)})_t \right]' \left[ \nabla f(\mathbf{b}_{(0)})_t \right] \right]^{-1} \left[ \nabla f(\mathbf{b}_{(0)})_t \right]' y_{t(0)}^*,$$

donde

$$f(\mathbf{b}_{(0)})_t = g_{1,0(0)} \left( D_{1,t(0)} + \frac{t}{t_{1(0)}} D_{2,t(0)} \right) + g_{2,0(0)} \left( 1 - \frac{t}{t_{1(0)}} \right) D_{2,t(0)} + g_{1,1(0)} t,$$

$$\text{con } D_{1,t(0)} = \begin{cases} 1, & t_0 \leq t < t_{1(0)} \\ 0, & \text{resto} \end{cases} \text{ y } D_{2,t(0)} = \begin{cases} 1, & t_{1(0)} \leq t \leq t_2 \\ 0, & \text{resto} \end{cases},$$

$$\nabla f(\mathbf{b}_{(0)})_t = \left[ \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,0}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,1}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{2,0}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial t_1} \right],$$

con

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,0}} = D_{1,t(0)} + \frac{t}{t_{1(0)}} D_{2,t(0)},$$

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,1}} = t,$$

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{2,0}} = \left( 1 - \frac{t}{t_{1(0)}} \right) D_{2,t(0)}$$

y

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial t_1} = (g_{2,0(0)} - g_{1,0(0)}) \frac{t}{(t_{1(0)})^2} D_{2,t(0)};$$

y

$$\begin{aligned}
 y_{t(0)}^* &= y_t - f(\mathbf{b}_{(0)})_t + \nabla f(\mathbf{b}_{(0)})_t \mathbf{b}_{(0)} \\
 &= y_t + (g_{2,0(0)} - g_{1,0(0)}) \frac{t}{t_{1(0)}} D_{2,t(0)}
 \end{aligned}$$

Las estimaciones de la  $i$ -ésima iteración vienen dadas por:

$$\mathbf{b}_{(i)} = \left[ \left[ \nabla f(\mathbf{b}_{(i-1)})_t \right] \left[ \nabla f(\mathbf{b}_{(i-1)})_t \right] \right]^{-1} \left[ \nabla f(\mathbf{b}_{(i-1)})_t \right]' y_{t(i-1)}^* .$$

### Caso II

Sea  $g(t)$  una función *spline* cuadrática definida por dos tramos, es decir,

$$g(t) = \begin{cases} g_{1,0} + g_{1,1}t + g_{1,2}t^2, & t_0 \leq t \leq t_1 \\ g_{2,0} + g_{2,1}t + g_{2,2}t^2, & t_1 \leq t \leq t_2 \end{cases}$$

Esta función puede expresarse como

$$g(t) = (g_{1,0} + g_{1,1}t + g_{1,2}t^2)D_{1,t} + (g_{2,0} + g_{2,1}t + g_{2,2}t^2)D_{2,t}$$

donde  $D_{1,t} = \begin{cases} 1, & t_0 \leq t < t_1 \\ 0, & \text{resto} \end{cases}$  y  $D_{2,t} = \begin{cases} 1, & t_1 \leq t \leq t_2 \\ 0, & \text{resto} \end{cases}$ . La función *spline* está sujeta a las

dos restricciones de continuidad

$$\begin{aligned} g_{1,0} + g_{1,1}t_1 + g_{1,2}t_1^2 &= g_{2,0} + g_{2,1}t_1 + g_{2,2}t_1^2 \\ g_{1,1} + 2g_{1,2}t_1 &= g_{2,1} + 2g_{2,2}t_1 \end{aligned}$$

Introduciendo estas restricciones, la función *spline* puede expresarse como

$$\begin{aligned} g(t) &= [g_{1,0} + g_{1,1}t + g_{1,2}t^2]D_{1,t} \\ &+ \left[ g_{2,0} + \left( \frac{2(g_{1,0} - g_{2,0})}{t_1} + g_{1,1} \right) t + \left( \frac{(g_{2,0} - g_{1,0})}{t_1^2} + g_{1,2} \right) t^2 \right] D_{2,t} \\ &= g_{1,0} \left[ D_{1,t} + \left( \frac{2t}{t_1} - \frac{t^2}{t_1^2} \right) D_{2,t} \right] + g_{2,0} \left( 1 - \frac{2t}{t_1} + \frac{t^2}{t_1^2} \right) D_{2,t} + g_{1,1}t + g_{1,2}t^2 \\ &= f(\mathbf{b})_t \end{aligned}$$

siendo  $\mathbf{b}$  el vector de parámetros  $[g_{1,0}, g_{1,1}, g_{1,2}, g_{2,0}, t_1]'$ , que puede estimarse, en una primera aproximación, como

$$\mathbf{b}_{(1)} = \mathbf{b}_{(0)} - \left[ \nabla^2 S(\mathbf{b}_{(0)}) \right]^{-1} \left[ \nabla S(\mathbf{b}_{(0)}) \right]'$$

siendo  $S(\mathbf{b}) = \sum_{t=1}^T (y_t - f(\mathbf{b})_t)^2$ , de modo que  $S(\mathbf{b}_{(0)})$ ,  $\nabla S(\mathbf{b}_{(0)})$  y  $\nabla^2 S(\mathbf{b}_{(0)})$  son

funciones de  $f(\mathbf{b}_{(0)})_t$ ,  $\nabla f(\mathbf{b}_{(0)})_t$  y  $\nabla^2 f(\mathbf{b}_{(0)})_t$ , definidas como

$$\begin{aligned} f(\mathbf{b}_{(0)})_t &= g_{1,0(0)} \left[ D_{1,t(0)} + \left( \frac{2t}{t_{1(0)}} - \frac{t^2}{t_{1(0)}^2} \right) D_{2,t(0)} \right] + g_{2,0(0)} \left( 1 - \frac{2t}{t_{1(0)}} + \frac{t^2}{t_{1(0)}^2} \right) D_{2,t(0)} \\ &+ g_{1,1(0)} t + g_{1,2(0)} t^2 \end{aligned}$$

y

$$\nabla f(\mathbf{b}_{(0)})_t = \left[ \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,0}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,1}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,2}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{2,0}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial t_1} \right],$$

con

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,0}} = D_{1,t(0)} + \left( \frac{2t}{t_{1(0)}} - \frac{t^2}{t_{1(0)}^2} \right) D_{2,t(0)},$$

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,1}} = t,$$

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,2}} = t^2,$$

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{2,0}} = \left( 1 - \frac{2t}{t_{1(0)}} + \frac{t^2}{t_{1(0)}^2} \right) D_{2,t(0)}$$

y

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial t_1} = (g_{2,0(0)} - g_{1,0(0)}) \left( \frac{2t}{t_{1(0)}^2} - \frac{2t^2}{t_{1(0)}^3} \right) D_{2,t(0)};$$

y

$$\nabla^2 f(\mathbf{b}_{(0)})_t = \begin{bmatrix} \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 g_{1,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,1} \partial g_{1,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,2} \partial g_{1,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{2,0} \partial g_{1,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial t_1 \partial g_{1,0}} \\ \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,0} \partial g_{1,1}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 g_{1,1}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,2} \partial g_{1,1}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{2,0} \partial g_{1,1}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial t_1 \partial g_{1,1}} \\ \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,0} \partial g_{1,2}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,1} \partial g_{1,2}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 g_{1,2}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{2,0} \partial g_{1,2}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial t_1 \partial g_{1,2}} \\ \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,0} \partial g_{2,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,1} \partial g_{2,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,2} \partial g_{2,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 g_{2,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial t_1 \partial g_{2,0}} \\ \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,0} \partial t_1} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,1} \partial t_1} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,2} \partial t_1} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{2,0} \partial t_1} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 t_1} \end{bmatrix} =$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & \left(-\frac{2t}{t_{1(0)}^2} + \frac{2t^2}{t_{1(0)}^3}\right)D_{2,t(0)} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \left(\frac{2t}{t_{1(0)}^2} - \frac{2t^2}{t_{1(0)}^3}\right)D_{2,t(0)} \\ \left(-\frac{2t}{t_{1(0)}^2} + \frac{2t^2}{t_{1(0)}^3}\right)D_{2,t(0)} & 0 & 0 & \left(\frac{2t}{t_{1(0)}^2} - \frac{2t^2}{t_{1(0)}^3}\right)D_{2,t(0)} & \left(g_{2,0(0)} - g_{1,0(0)}\right) \left(-\frac{4t}{t_{1(0)}^3} + \frac{6t^2}{t_{1(0)}^4}\right)D_{2,t(0)} \end{bmatrix}$$

Iterativamente, se pueden ir obteniendo estimaciones del vector  $\mathbf{b}$  dadas por

$$\mathbf{b}_{(i)} = \mathbf{b}_{(i-1)} - \left[\nabla^2 S(\mathbf{b}_{(i-1)})\right]^{-1} \left[\nabla S(\mathbf{b}_{(i-1)})\right]'$$

### Caso III

Sea  $g(t)$  la función *spline* cúbica definida por dos tramos, es decir,

$$g(t) = \begin{cases} g_{1,0} + g_{1,1}t + g_{1,2}t^2 + g_{1,3}t^3, & t_0 \leq t \leq t_1 \\ g_{2,0} + g_{2,1}t + g_{2,2}t^2 + g_{2,3}t^3, & t_1 \leq t \leq t_2 \end{cases}$$

Esta función puede expresarse como

$$g(t) = (g_{1,0} + g_{1,1}t + g_{1,2}t^2 + g_{1,3}t^3)D_{1,t} + (g_{2,0} + g_{2,1}t + g_{2,2}t^2 + g_{2,3}t^3)D_{2,t}$$

donde  $D_{1,t} = \begin{cases} 1, & t_0 \leq t < t_1 \\ 0, & \text{resto} \end{cases}$  y  $D_{2,t} = \begin{cases} 1, & t_1 \leq t \leq t_2 \\ 0, & \text{resto} \end{cases}$ . La función *spline* está sujeta a las

tres restricciones de continuidad

$$g_{1,0} + g_{1,1}t_1 + g_{1,2}t_1^2 + g_{1,3}t_1^3 = g_{2,0} + g_{2,1}t_1 + g_{2,2}t_1^2 + g_{2,3}t_1^3,$$

$$g_{1,1} + 2g_{1,2}t_1 + 3g_{1,3}t_1^2 = g_{2,1} + 2g_{2,2}t_1 + 3g_{2,3}t_1^2$$

y

$$2g_{1,2} + 6g_{1,3}t_1 = 2g_{2,2} + 6g_{2,3}t_1.$$

Introduciendo estas restricciones, la función *spline* puede expresarse como

$$\begin{aligned} g(t) &= \left[ g_{1,0} + g_{1,1}t + g_{1,2}t^2 + g_{1,3}t^3 \right] D_{1,t} \\ &+ \left[ g_{2,0} + \left( g_{1,1} + \frac{3(g_{1,0} - g_{2,0})}{t_1} \right) t + \left( g_{1,2} - \frac{3(g_{1,0} - g_{2,0})}{t_1^2} \right) t^2 + \left( g_{1,3} + \frac{(g_{1,0} - g_{2,0})}{t_1^3} \right) t^3 \right] D_{2,t} \\ &= g_{1,0} \left[ D_{1,t} + \left( \frac{3t}{t_1} - \frac{3t^2}{t_1^2} + \frac{t^3}{t_1^3} \right) D_{2,t} \right] + g_{2,0} \left( 1 - \frac{3t}{t_1} + \frac{3t^2}{t_1^2} - \frac{t^3}{t_1^3} \right) D_{2,t} \\ &+ g_{1,1}t + g_{1,2}t^2 + g_{1,3}t^3 = f(\mathbf{b})_t \end{aligned}$$

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

siendo  $\mathbf{b}$  el vector de parámetros  $[g_{1,0}, g_{1,1}, g_{1,2}, g_{1,3}, g_{2,0}, t_1]'$ , que puede estimarse como

$$\mathbf{b}_{(1)} = \mathbf{b}_{(0)} - [\nabla^2 S(\mathbf{b}_{(0)})]^{-1} [\nabla S(\mathbf{b}_{(0)})]'$$

siendo  $S(\mathbf{b}) = \sum_{t=1}^T (y_t - f(\mathbf{b})_t)^2$ , de modo que  $S(\mathbf{b}_{(0)})$ ,  $\nabla S(\mathbf{b}_{(0)})$  y  $\nabla^2 S(\mathbf{b}_{(0)})$  son

funciones de  $f(\mathbf{b}_{(0)})_t$ ,  $\nabla f(\mathbf{b}_{(0)})_t$  y  $\nabla^2 f(\mathbf{b}_{(0)})_t$ , definidas como

$$f(\mathbf{b}_{(0)})_t = g_{1,0(0)} \left[ D_{1,t(0)} + \left( \frac{3t}{t_{1(0)}} - \frac{3t^2}{t_{1(0)}^2} + \frac{t^3}{t_{1(0)}^3} \right) D_{2,t(0)} \right] + g_{2,0(0)} \left( 1 - \frac{3t}{t_{1(0)}} + \frac{3t^2}{t_{1(0)}^2} - \frac{t^3}{t_{1(0)}^3} \right) D_{2,t(0)} \\ + g_{1,1(0)} t + g_{1,2(0)} t^2 + g_{1,3(0)} t^3$$

y

$$\nabla f(\mathbf{b}_{(0)})_t = \left[ \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,0}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,1}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,2}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,3}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{2,0}}, \frac{\partial f(\mathbf{b}_{(0)})_t}{\partial t_1} \right],$$

con

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,0}} = D_{1,t(0)} + \left( \frac{3t}{t_{1(0)}} - \frac{3t^2}{t_{1(0)}^2} + \frac{t^3}{t_{1(0)}^3} \right) D_{2,t(0)},$$

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,1}} = t,$$

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,2}} = t^2,$$

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{1,3}} = t^3,$$

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial g_{2,0}} = \left( 1 - \frac{3t}{t_{1(0)}} + \frac{3t^2}{t_{1(0)}^2} - \frac{t^3}{t_{1(0)}^3} \right) D_{2,t(0)}$$

y

$$\frac{\partial f(\mathbf{b}_{(0)})_t}{\partial t_1} = (g_{1,0(0)} - g_{2,0(0)}) \left( -\frac{3t}{t_{1(0)}^2} + \frac{6t^2}{t_{1(0)}^3} - \frac{3t^3}{t_{1(0)}^4} \right) D_{2,t(0)},$$

y

$$\nabla^2 f(\mathbf{b}_{(0)})_t = \begin{bmatrix} \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 g_{1,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,1} \partial g_{1,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,2} \partial g_{1,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,3} \partial g_{1,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{2,0} \partial g_{1,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial t_1 \partial g_{1,0}} \\ \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,0} \partial g_{1,1}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 g_{1,1}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,2} \partial g_{1,1}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,3} \partial g_{1,1}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{2,0} \partial g_{1,1}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial t_1 \partial g_{1,1}} \\ \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,0} \partial g_{1,2}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,1} \partial g_{1,2}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 g_{1,2}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,3} \partial g_{1,2}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{2,0} \partial g_{1,2}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial t_1 \partial g_{1,2}} \\ \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,0} \partial g_{1,3}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,1} \partial g_{1,3}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,2} \partial g_{1,3}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 g_{1,3}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{2,0} \partial g_{1,3}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial t_1 \partial g_{1,3}} \\ \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,0} \partial g_{2,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,1} \partial g_{2,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,2} \partial g_{2,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,3} \partial g_{2,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 g_{2,0}} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial t_1 \partial g_{2,0}} \\ \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,0} \partial t_1} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,1} \partial t_1} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,2} \partial t_1} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{1,3} \partial t_1} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial g_{2,0} \partial t_1} & \frac{\partial^2 f(\mathbf{b}_{(0)})_t}{\partial^2 t_1} \end{bmatrix} =$$

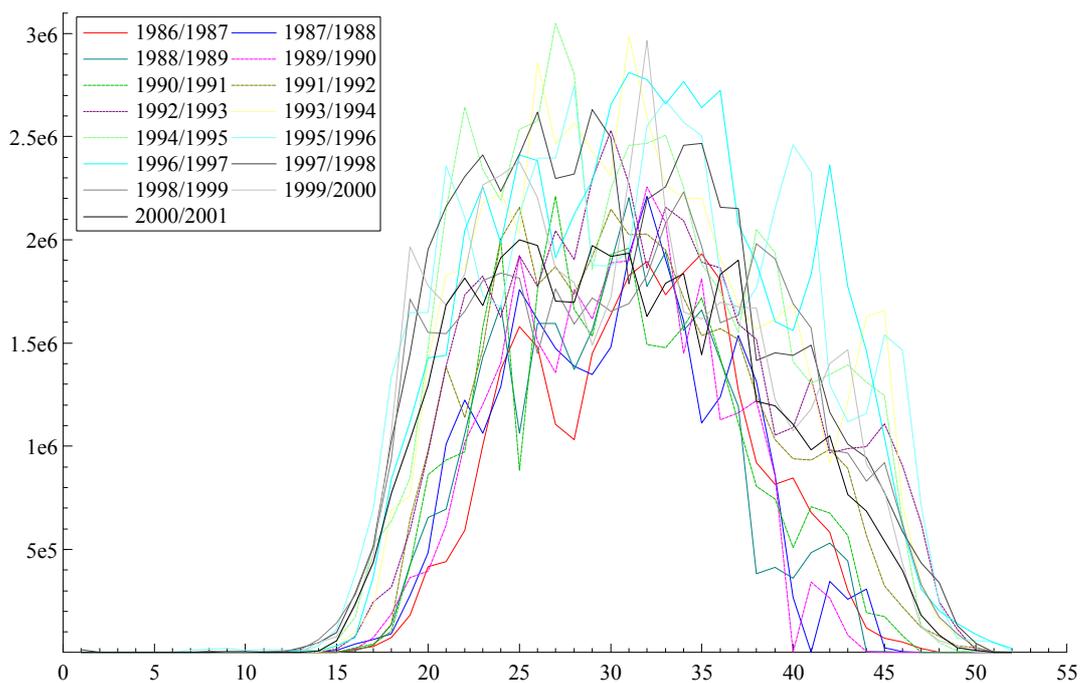
$$\begin{bmatrix} 0 & 0 & 0 & 0 & \left( -\frac{3t}{t_{1(0)}^2} + \frac{6t^2}{t_{1(0)}^3} - \frac{3t^3}{t_{1(0)}^4} \right) D_{2,t(0)} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \left( \frac{3t}{t_{1(0)}^2} - \frac{6t^2}{t_{1(0)}^3} + \frac{3t^3}{t_{1(0)}^4} \right) D_{2,t(0)} \\ \left( -\frac{3t}{t_{1(0)}^2} + \frac{6t^2}{t_{1(0)}^3} - \frac{3t^3}{t_{1(0)}^4} \right) D_{2,t(0)} & 0 & 0 & 0 & \left( \frac{3t}{t_{1(0)}^2} - \frac{6t^2}{t_{1(0)}^3} + \frac{3t^3}{t_{1(0)}^4} \right) D_{2,t(0)} & \left( \frac{6t}{t_{1(0)}^3} - \frac{18t^2}{t_{1(0)}^4} + \frac{12t^3}{t_{1(0)}^5} \right) (g_{1,0(0)} - g_{2,0(0)}) D_{2,t(0)} \end{bmatrix}$$

Iterativamente, se pueden ir calculando estimaciones del vector  $\mathbf{b}$  dadas por:

$$\mathbf{b}_{(i)} = \mathbf{b}_{(i-1)} - [\nabla^2 S(\mathbf{b}_{(i-1)})]^{-1} [\nabla S(\mathbf{b}_{(i-1)})]'$$



### Anexo 3.1. Evolución por campaña de las exportaciones semanales de tomate canario a Europa (1986/1987 a 2000/2001)



## Anexo 3.2. Resultados para la transformación logarítmica de la serie de exportaciones semanales

### A. Estacionalidad determinística (variables cualitativas)

#### A.1) Resultados del modelo sin intervenciones:

##### Varianzas de los términos de perturbación

	Irregular ( $\hat{s}_e^2$ )	Nivel ( $\hat{s}_h^2$ )
Varianza	0.94109	1.7097
q-ratio	(0.5504)	(1.0000)

##### Componente de nivel en el vector de estado final y componente estacional, $\hat{g}_{1,j}^{anual}$

Estimación		Estimación		Estimación		Estimación	
$\hat{m}_T$	7.3923	$\hat{g}_{1,40}^{anual}$	-3.2817	$\hat{g}_{1,1}^{anual}$	5.0356	$\hat{g}_{1,14}^{anual}$	3.5116
$\hat{g}_{1,27}^{anual}$	-8.3567	$\hat{g}_{1,41}^{anual}$	-0.6674(*)	$\hat{g}_{1,2}^{anual}$	5.0118	$\hat{g}_{1,15}^{anual}$	4.0694
$\hat{g}_{1,28}^{anual}$	-9.4528	$\hat{g}_{1,42}^{anual}$	1.7270	$\hat{g}_{1,3}^{anual}$	5.0104	$\hat{g}_{1,16}^{anual}$	4.2844
$\hat{g}_{1,29}^{anual}$	-9.4512	$\hat{g}_{1,43}^{anual}$	2.6125	$\hat{g}_{1,4}^{anual}$	5.0999	$\hat{g}_{1,17}^{anual}$	4.0830
$\hat{g}_{1,30}^{anual}$	-9.4496	$\hat{g}_{1,44}^{anual}$	3.3881	$\hat{g}_{1,5}^{anual}$	5.1558	$\hat{g}_{1,18}^{anual}$	3.3914
$\hat{g}_{1,31}^{anual}$	-9.4479	$\hat{g}_{1,45}^{anual}$	4.0625	$\hat{g}_{1,6}^{anual}$	5.1670	$\hat{g}_{1,19}^{anual}$	3.0675
$\hat{g}_{1,32}^{anual}$	-9.4463	$\hat{g}_{1,46}^{anual}$	4.4171	$\hat{g}_{1,7}^{anual}$	5.1471	$\hat{g}_{1,20}^{anual}$	1.9107
$\hat{g}_{1,33}^{anual}$	-8.8168	$\hat{g}_{1,47}^{anual}$	4.6447	$\hat{g}_{1,8}^{anual}$	5.0827	$\hat{g}_{1,21}^{anual}$	-0.3715(*)
$\hat{g}_{1,34}^{anual}$	-8.1982	$\hat{g}_{1,48}^{anual}$	4.7957	$\hat{g}_{1,9}^{anual}$	5.0306	$\hat{g}_{1,22}^{anual}$	-1.5024
$\hat{g}_{1,35}^{anual}$	-8.2311	$\hat{g}_{1,49}^{anual}$	4.9203	$\hat{g}_{1,10}^{anual}$	4.9444	$\hat{g}_{1,23}^{anual}$	-2.1244
$\hat{g}_{1,36}^{anual}$	-7.7341	$\hat{g}_{1,50}^{anual}$	4.9983	$\hat{g}_{1,11}^{anual}$	4.8586	$\hat{g}_{1,24}^{anual}$	-3.3268
$\hat{g}_{1,37}^{anual}$	-8.2485	$\hat{g}_{1,51}^{anual}$	5.0233	$\hat{g}_{1,12}^{anual}$	4.6860	$\hat{g}_{1,25}^{anual}$	-5.7643
$\hat{g}_{1,38}^{anual}$	-8.2097	$\hat{g}_{1,52}^{anual}$	5.0688	$\hat{g}_{1,13}^{anual}$	4.5683	$\hat{g}_{1,26}^{anual}$	-7.2733
$\hat{g}_{1,39}^{anual}$	-5.4196						

#### A.2) Resultados del modelo con intervenciones:

##### Varianzas de los términos de perturbación

	Irregular ( $\hat{s}_e^2$ )	Nivel ( $\hat{s}_h^2$ )
Varianza	$2.2698 \times 10^{-6}$	0.057012
q-ratio	(0.0000)	(1.0000)

Componente de nivel en el vector de estado final y componente estacional,  $\hat{g}_{1,j}^{anual}$

Estimación		Estimación		Estimación		Estimación	
$\hat{m}_T$	2.4443	$\hat{g}_{1,40}^{anual}$	-2.6662	$\hat{g}_{1,1}^{anual}$	2.2443(*)	$\hat{g}_{1,14}^{anual}$	1.3440(*)
$\hat{g}_{1,27}^{anual}$	-2.5145	$\hat{g}_{1,41}^{anual}$	-1.9327(*)	$\hat{g}_{1,2}^{anual}$	2.1967	$\hat{g}_{1,15}^{anual}$	1.3001(*)
$\hat{g}_{1,28}^{anual}$	-2.5400	$\hat{g}_{1,42}^{anual}$	-0.8030	$\hat{g}_{1,3}^{anual}$	2.1716(*)	$\hat{g}_{1,16}^{anual}$	1.1368(*)
$\hat{g}_{1,29}^{anual}$	-2.5621	$\hat{g}_{1,43}^{anual}$	0.0587	$\hat{g}_{1,4}^{anual}$	2.2374(*)	$\hat{g}_{1,17}^{anual}$	0.9753(*)
$\hat{g}_{1,30}^{anual}$	-2.5843	$\hat{g}_{1,44}^{anual}$	0.8105(*)	$\hat{g}_{1,5}^{anual}$	2.2695	$\hat{g}_{1,18}^{anual}$	0.7195(*)
$\hat{g}_{1,31}^{anual}$	-2.6064	$\hat{g}_{1,45}^{anual}$	1.4612(*)	$\hat{g}_{1,6}^{anual}$	2.2570(*)	$\hat{g}_{1,19}^{anual}$	0.5760(*)
$\hat{g}_{1,32}^{anual}$	-2.6285	$\hat{g}_{1,46}^{anual}$	1.7920(*)	$\hat{g}_{1,7}^{anual}$	2.2133(*)	$\hat{g}_{1,20}^{anual}$	0.1009(*)
$\hat{g}_{1,33}^{anual}$	-2.6522	$\hat{g}_{1,47}^{anual}$	1.9958	$\hat{g}_{1,8}^{anual}$	2.1251(*)	$\hat{g}_{1,21}^{anual}$	-0.6413(*)
$\hat{g}_{1,34}^{anual}$	-2.6546(*)	$\hat{g}_{1,48}^{anual}$	2.1231(*)	$\hat{g}_{1,9}^{anual}$	2.0493(*)	$\hat{g}_{1,22}^{anual}$	-1.1710(*)
$\hat{g}_{1,35}^{anual}$	-2.6642(*)	$\hat{g}_{1,49}^{anual}$	2.2239(*)	$\hat{g}_{1,10}^{anual}$	1.9394(*)	$\hat{g}_{1,23}^{anual}$	-1.7441(*)
$\hat{g}_{1,36}^{anual}$	-2.7295(*)	$\hat{g}_{1,50}^{anual}$	2.2782(*)	$\hat{g}_{1,11}^{anual}$	1.8298(*)	$\hat{g}_{1,24}^{anual}$	-2.1211(*)
$\hat{g}_{1,37}^{anual}$	-2.7301	$\hat{g}_{1,51}^{anual}$	2.3325(*)	$\hat{g}_{1,12}^{anual}$	1.6334(*)	$\hat{g}_{1,25}^{anual}$	-2.2916(*)
$\hat{g}_{1,38}^{anual}$	-2.7581	$\hat{g}_{1,52}^{anual}$	2.3013(*)	$\hat{g}_{1,13}^{anual}$	1.4921(*)	$\hat{g}_{1,26}^{anual}$	-2.4443(*)
$\hat{g}_{1,39}^{anual}$	-2.7495						

(\*) parámetros no significativos al 95 % de confianza.

Párametros de las variables de intervención en el componente irregular

Parámetro	Estimación	Parámetro	Estimación	Parámetro	Estimación
$\hat{I}_{14-88}$	-0.8360	$\hat{I}_{24-93}$	-1.8418	$\hat{I}_{37-98}$	8.2693
$\hat{I}_{15-88}$	-5.3165	$\hat{I}_{40-93}$	6.4678	$\hat{I}_{38-98}$	8.9142
$\hat{I}_{19-88}$	-2.4022	$\hat{I}_{39-94}$	7.1993	$\hat{I}_{39-98}$	9.6746
$\hat{I}_{20-88}$	-3.7026	$\hat{I}_{40-94}$	8.6754	$\hat{I}_{24-99}$	-0.9806
$\hat{I}_{20-89}$	-1.9862	$\hat{I}_{23-95}$	-1.0910	$\hat{I}_{25-99}$	-2.2539
$\hat{I}_{14-90}$	-13.1500	$\hat{I}_{24-95}$	-0.9794	$\hat{I}_{36-99}$	8.0388
$\hat{I}_{17-90}$	-0.9536	$\hat{I}_{27-95}$	-1.0975	$\hat{I}_{39-99}$	7.5924
$\hat{I}_{18-90}$	-3.4937	$\hat{I}_{39-95}$	-1.3196	$\hat{I}_{35-00}$	8.2532
$\hat{I}_{19-90}$	-4.1531	$\hat{I}_{26-96}$	-1.7283	$\hat{I}_{36-00}$	8.2934
$\hat{I}_{41-90}$	6.2115	$\hat{I}_{25-98}$	-3.1305	$\hat{I}_{39-00}$	7.9289
$\hat{I}_{51-90}$	-0.7956	$\hat{I}_{26-98}$	-5.9783	$\hat{I}_{40-00}$	8.8011
$\hat{I}_{25-92}$	-2.0358	$\hat{I}_{34-98}$	8.9583		

**Párametros de las variables de intervención en el componente de nivel**

Parámetro	Estimación	Parámetro	Estimación	Parámetro	Estimación
$\hat{I}_{41-86}$	7.3092	$\hat{I}_{40-91}$	7.5909	$\hat{I}_{39-96}$	9.1251
$\hat{I}_{22-87}$	-9.3752	$\hat{I}_{26-92}$	-9.6837	$\hat{I}_{28-97}$	-9.4333
$\hat{I}_{41-87}$	8.5576	$\hat{I}_{40-92}$	8.2524	$\hat{I}_{39-97}$	10.0810
$\hat{I}_{21-88}$	-11.2770	$\hat{I}_{25-93}$	-11.0810	$\hat{I}_{27-98}$	-10.3270
$\hat{I}_{42-88}$	8.1122	$\hat{I}_{41-93}$	9.4723	$\hat{I}_{40-98}$	10.9840
$\hat{I}_{18-89}$	-4.3536	$\hat{I}_{25-94}$	-10.4650	$\hat{I}_{26-99}$	-10.7310
$\hat{I}_{21-89}$	-7.3123	$\hat{I}_{41-94}$	10.1420	$\hat{I}_{40-99}$	10.7670
$\hat{I}_{42-89}$	8.5708	$\hat{I}_{28-95}$	-8.0795	$\hat{I}_{24-00}$	-10.0270
$\hat{I}_{20-90}$	-11.4490	$\hat{I}_{33-95}$	9.4411	$\hat{I}_{41-00}$	10.1080
$\hat{I}_{42-90}$	8.1433	$\hat{I}_{27-96}$	-10.6310	$\hat{I}_{25-01}$	-9.1959
$\hat{I}_{21-91}$	-10.5540				

Para la especificación de una intervención en el componente de nivel, véase capítulo 1:21.

**B) Estacionalidad determinística (spline)****B.1) Resultados del modelo sin intervenciones:****Varianzas de los términos de perturbación**

	Irregular ( $\hat{s}_e^2$ )	Nivel ( $\hat{s}_h^2$ )
Varianza	0.57380	2.5780
q-ratio	(0.2226)	(1.0000)

**Cuadro b.1.2. Estimación del nivel en el vector de estado final**

Componente	Estimación
$\hat{m}_T$	7.9307

**Párametros libres en la especificación de la spline**

Parámetro	Estimación	Estadístico $t$	Parámetro	Estimación	Estadístico $t$
$\hat{g}_{2,0}^*$	3444.70	6.3785	$\hat{g}_{7,0}^*$	767.02	1.1474
$\hat{g}_{3,0}^*$	-182.49	-4.8226	$\hat{g}_{8,0}^*$	-2625.00	-0.3433
$\hat{g}_{4,0}^*$	73.94	0.8888	$\hat{g}_{9,0}^*$	1254.30	0.4097
$\hat{g}_{5,0}^*$	131.46	0.1152	$\hat{g}_{10,0}^*$	-6118.30	-3.4913
$\hat{g}_{6,0}^*$	-109.90	-0.5360			

(\*) Parámetros no significativos al 95 % de confianza.

Componente estacional  $\hat{g}_{3,w}^{anual}$

Estimación		Estimación		Estimación		Estimación	
$\hat{g}_{3,1}^{anual}$	0.3179	$\hat{g}_{3,14}^{anual}$	4.4510	$\hat{g}_{3,27}^{anual}$	12.6465	$\hat{g}_{3,40}^{anual}$	12.0527
$\hat{g}_{3,2}^{anual}$	0.6359	$\hat{g}_{3,15}^{anual}$	6.4929	$\hat{g}_{3,28}^{anual}$	12.7686	$\hat{g}_{3,41}^{anual}$	12.0094
$\hat{g}_{3,3}^{anual}$	0.9538	$\hat{g}_{3,16}^{anual}$	8.8305	$\hat{g}_{3,29}^{anual}$	12.8824	$\hat{g}_{3,42}^{anual}$	11.8851
$\hat{g}_{3,4}^{anual}$	1.2717	$\hat{g}_{3,17}^{anual}$	10.4969	$\hat{g}_{3,30}^{anual}$	12.9694	$\hat{\gamma}_{3,43}^{anual}$	11.6085
$\hat{g}_{3,5}^{anual}$	1.5897	$\hat{g}_{3,18}^{anual}$	11.6002	$\hat{g}_{3,31}^{anual}$	13.0090	$\hat{g}_{3,44}^{anual}$	11.1080
$\hat{g}_{3,6}^{anual}$	1.9076	$\hat{g}_{3,19}^{anual}$	12.2483	$\hat{g}_{3,32}^{anual}$	12.9801	$\hat{\gamma}_{3,45}^{anual}$	10.3522
$\hat{g}_{3,7}^{anual}$	2.2255	$\hat{g}_{3,20}^{anual}$	12.5490	$\hat{g}_{3,33}^{anual}$	12.8919	$\hat{g}_{3,46}^{anual}$	9.4262
$\hat{g}_{3,8}^{anual}$	2.5435	$\hat{g}_{3,21}^{anual}$	12.6103	$\hat{g}_{3,34}^{anual}$	12.7617	$\hat{g}_{3,47}^{anual}$	8.2845
$\hat{g}_{3,9}^{anual}$	2.8614	$\hat{g}_{3,22}^{anual}$	12.5402	$\hat{g}_{3,35}^{anual}$	12.6069	$\hat{g}_{3,48}^{anual}$	6.8394
$\hat{g}_{3,10}^{anual}$	3.1793	$\hat{g}_{3,23}^{anual}$	12.4466	$\hat{g}_{3,36}^{anual}$	12.4446	$\hat{g}_{3,49}^{anual}$	5.0696
$\hat{g}_{3,11}^{anual}$	3.4972	$\hat{g}_{3,24}^{anual}$	12.4163	$\hat{g}_{3,37}^{anual}$	12.2922	$\hat{g}_{3,50}^{anual}$	3.2204
$\hat{g}_{3,12}^{anual}$	3.8152	$\hat{g}_{3,25}^{anual}$	12.4519	$\hat{g}_{3,38}^{anual}$	12.1670	$\hat{g}_{3,51}^{anual}$	1.6040
$\hat{g}_{3,13}^{anual}$	4.1331	$\hat{g}_{3,26}^{anual}$	12.5348	$\hat{g}_{3,39}^{anual}$	12.0863	$\hat{g}_{3,52}^{anual}$	0.5325

b.2) Resultados del modelo con intervenciones:

Varianzas de los términos de perturbación

	Irregular ( $\hat{s}_e^2$ )	Nivel ( $\hat{s}_h^2$ )
Varianza	$3.53038 \times 10^{-6}$	0.055825
q-ratio	(0.0001)	(1.0000)

Estimación del nivel en el vector de estado final

Componente	Estimación
$\hat{m}_r$	2.4574

Parámetros libres en la especificación de la *spline*

Parámetro	Estimación	Estadístico t	Parámetro	Estimación	Estadístico t
$\hat{g}_{2,0}^*$	1603.10	19.1020	$\hat{g}_{7,0}^*$	72.58	0.7590
$\hat{g}_{3,0}^*$	-56.49	-9.8410	$\hat{g}_{8,0}^*$	4315.30	3.8787
$\hat{g}_{4,0}^*$	-39.39	-3.2883	$\hat{g}_{9,0}^*$	-2796.90	-6.0746
$\hat{g}_{5,0}^*$	451.05	2.7783	$\hat{g}_{10,0}^*$	596.62	2.1270
$\hat{g}_{6,0}^*$	-56.30	-1.9298			

(\*) Parámetros no significativos al 95 % de confianza.

Componente estacional  $\hat{g}_{3,w}^{anual}$ 

Estimación		Estimación		Estimación		Estimación	
$\hat{g}_{3,1}^{anual}$	-2.5101	$\hat{g}_{3,14}^{anual}$	-2.7428	$\hat{g}_{3,27}^{anual}$	2.2345	$\hat{g}_{3,40}^{anual}$	1.3716
$\hat{g}_{3,2}^{anual}$	-2.5280	$\hat{g}_{3,15}^{anual}$	-1.9014	$\hat{g}_{3,28}^{anual}$	2.2016	$\hat{g}_{3,41}^{anual}$	1.2382
$\hat{g}_{3,3}^{anual}$	-2.5459	$\hat{g}_{3,16}^{anual}$	-0.8016	$\hat{g}_{3,29}^{anual}$	2.1854	$\hat{g}_{3,42}^{anual}$	1.1009
$\hat{g}_{3,4}^{anual}$	-2.5638	$\hat{g}_{3,17}^{anual}$	0.0874	$\hat{g}_{3,30}^{anual}$	2.1999	$\hat{\gamma}_{3,43}^{anual}$	0.9538
$\hat{g}_{3,5}^{anual}$	-2.5817	$\hat{g}_{3,18}^{anual}$	0.7880	$\hat{g}_{3,31}^{anual}$	2.2410	$\hat{g}_{3,44}^{anual}$	0.7913
$\hat{g}_{3,6}^{anual}$	-2.5996	$\hat{g}_{3,19}^{anual}$	1.3226	$\hat{g}_{3,32}^{anual}$	2.2490	$\hat{\gamma}_{3,45}^{anual}$	0.5576
$\hat{g}_{3,7}^{anual}$	-2.6175	$\hat{g}_{3,20}^{anual}$	1.7139	$\hat{g}_{3,33}^{anual}$	2.2140	$\hat{g}_{3,46}^{anual}$	0.0761
$\hat{g}_{3,8}^{anual}$	-2.6354	$\hat{g}_{3,21}^{anual}$	1.9842	$\hat{g}_{3,34}^{anual}$	2.1432	$\hat{g}_{3,47}^{anual}$	-0.5677
$\hat{g}_{3,9}^{anual}$	-2.6533	$\hat{g}_{3,22}^{anual}$	2.1560	$\hat{g}_{3,35}^{anual}$	2.0440	$\hat{g}_{3,48}^{anual}$	-1.2102
$\hat{g}_{3,10}^{anual}$	-2.6712	$\hat{g}_{3,23}^{anual}$	2.2518	$\hat{g}_{3,36}^{anual}$	1.9235	$\hat{g}_{3,49}^{anual}$	-1.7184
$\hat{g}_{3,11}^{anual}$	-2.6891	$\hat{g}_{3,24}^{anual}$	2.2927	$\hat{g}_{3,37}^{anual}$	1.7890	$\hat{g}_{3,50}^{anual}$	-2.0820
$\hat{g}_{3,12}^{anual}$	-2.7070	$\hat{g}_{3,25}^{anual}$	2.2940	$\hat{g}_{3,38}^{anual}$	1.6477	$\hat{g}_{3,51}^{anual}$	-2.3215
$\hat{g}_{3,13}^{anual}$	-2.7249	$\hat{g}_{3,26}^{anual}$	2.2700	$\hat{g}_{3,39}^{anual}$	1.5069	$\hat{g}_{3,52}^{anual}$	-2.4574

## Parámetros de las variables de intervención en el componente irregular

Parámetro	Estimación	Parámetro	Estimación	Parámetro	Estimación
$\hat{I}_{14-88}$	-0.8657	$\hat{I}_{24-93}$	-1.8551	$\hat{I}_{37-98}$	8.2866
$\hat{I}_{15-88}$	-5.2737	$\hat{I}_{40-93}$	6.5690	$\hat{I}_{38-98}$	8.9215
$\hat{I}_{19-88}$	-2.3120	$\hat{I}_{39-94}$	7.2257	$\hat{I}_{39-98}$	9.7083
$\hat{I}_{20-88}$	-3.6059	$\hat{I}_{40-94}$	8.8030	$\hat{I}_{24-99}$	-0.9939
$\hat{I}_{20-89}$	-1.9797	$\hat{I}_{23-95}$	-1.1528	$\hat{I}_{25-99}$	-2.1983
$\hat{I}_{14-90}$	-13.2020	$\hat{I}_{24-95}$	-1.0514	$\hat{I}_{36-99}$	8.0064
$\hat{I}_{17-90}$	-0.9681	$\hat{I}_{27-95}$	-1.1150	$\hat{I}_{39-99}$	7.6188
$\hat{I}_{18-90}$	-3.6014	$\hat{I}_{39-95}$	-1.3570	$\hat{I}_{35-00}$	8.2687
$\hat{I}_{19-90}$	-4.1707	$\hat{I}_{26-96}$	-1.7452	$\hat{I}_{36-00}$	8.2687
$\hat{I}_{41-90}$	6.1036	$\hat{I}_{25-98}$	-3.0616	$\hat{I}_{39-00}$	7.9553
$\hat{I}_{51-90}$	-0.7656	$\hat{I}_{26-98}$	-5.9263	$\hat{I}_{40-00}$	8.9287
$\hat{I}_{25-92}$	-1.9669	$\hat{I}_{34-98}$	8.9619		

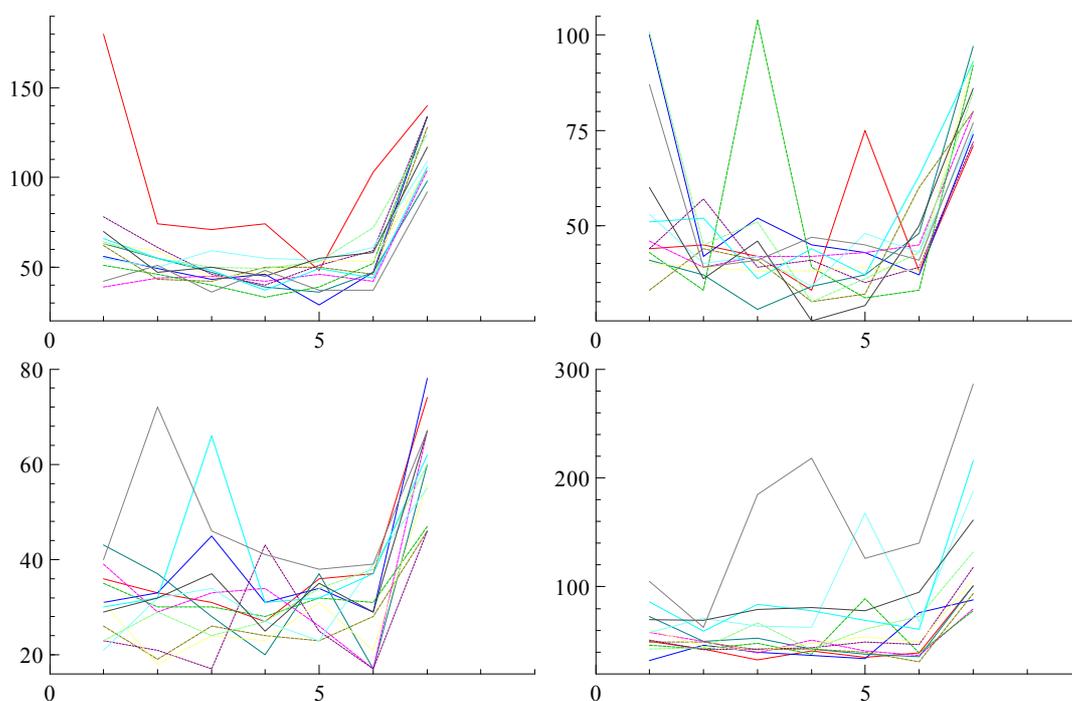
**Parámetros de las variables de intervención en el componente de nivel**

Parámetro	Estimación	Parámetro	Estimación	Parámetro	Estimación
$\hat{I}_{41-86}$	7.2013	$\hat{I}_{40-91}$	7.6920	$\hat{I}_{39-96}$	9.1516
$\hat{I}_{22-87}$	-9.2623	$\hat{I}_{26-92}$	-9.6316	$\hat{I}_{28-97}$	-9.4409
$\hat{I}_{41-87}$	8.4498	$\hat{I}_{40-92}$	8.3536	$\hat{I}_{39-97}$	10.1070
$\hat{I}_{21-88}$	-11.2790	$\hat{I}_{25-93}$	-11.0260	$\hat{I}_{27-98}$	-10.2920
$\hat{I}_{42-88}$	8.1420	$\hat{I}_{41-93}$	9.4656	$\hat{I}_{40-98}$	11.1180
$\hat{I}_{18-89}$	-4.4469	$\hat{I}_{25-94}$	-10.3960	$\hat{I}_{26-99}$	-10.6920
$\hat{I}_{21-89}$	-7.4042	$\hat{I}_{41-94}$	10.1610	$\hat{I}_{40-99}$	10.8950
$\hat{I}_{42-89}$	8.6006	$\hat{I}_{28-95}$	-8.1047	$\hat{I}_{24-00}$	-10.0410
$\hat{I}_{20-90}$	-11.4600	$\hat{I}_{33-95}$	9.4353	$\hat{I}_{41-00}$	10.1270
$\hat{I}_{42-90}$	8.0653	$\hat{I}_{27-96}$	-10.6660	$\hat{I}_{25-01}$	-9.1270
$\hat{I}_{21-91}$	-10.6530				

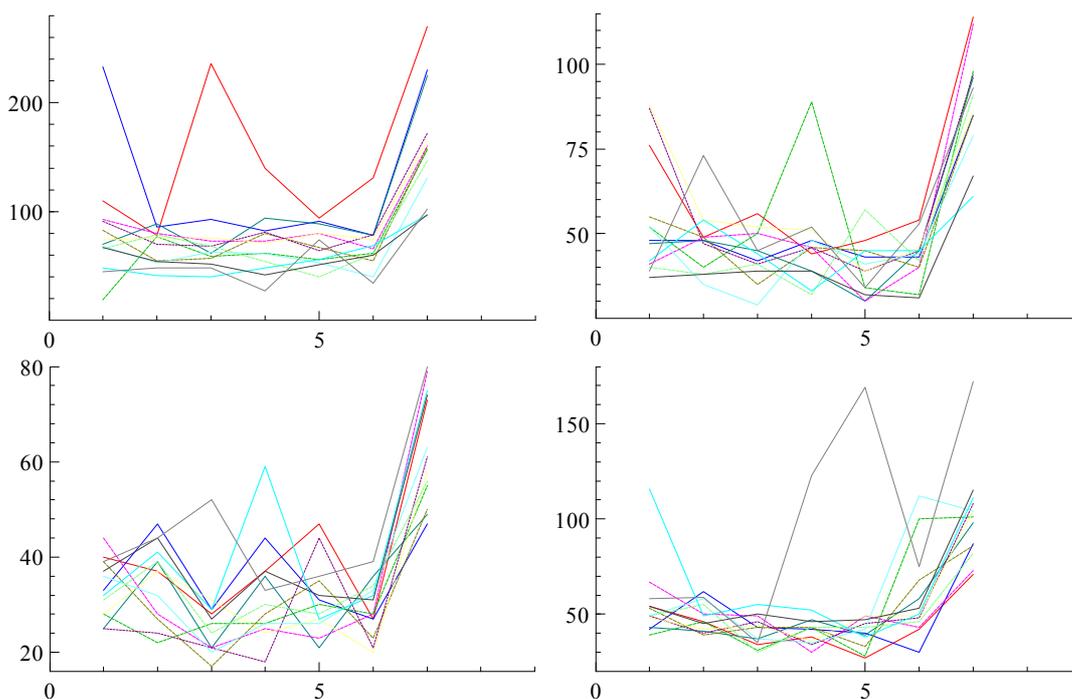
Para la especificación de una intervención en el componente de nivel, véase capítulo 1:21.

# Anexo 4.1 Evolución semanal del número diario de visitas registradas en el servicio de urgencias de Barcelonès Nord

Año 1996

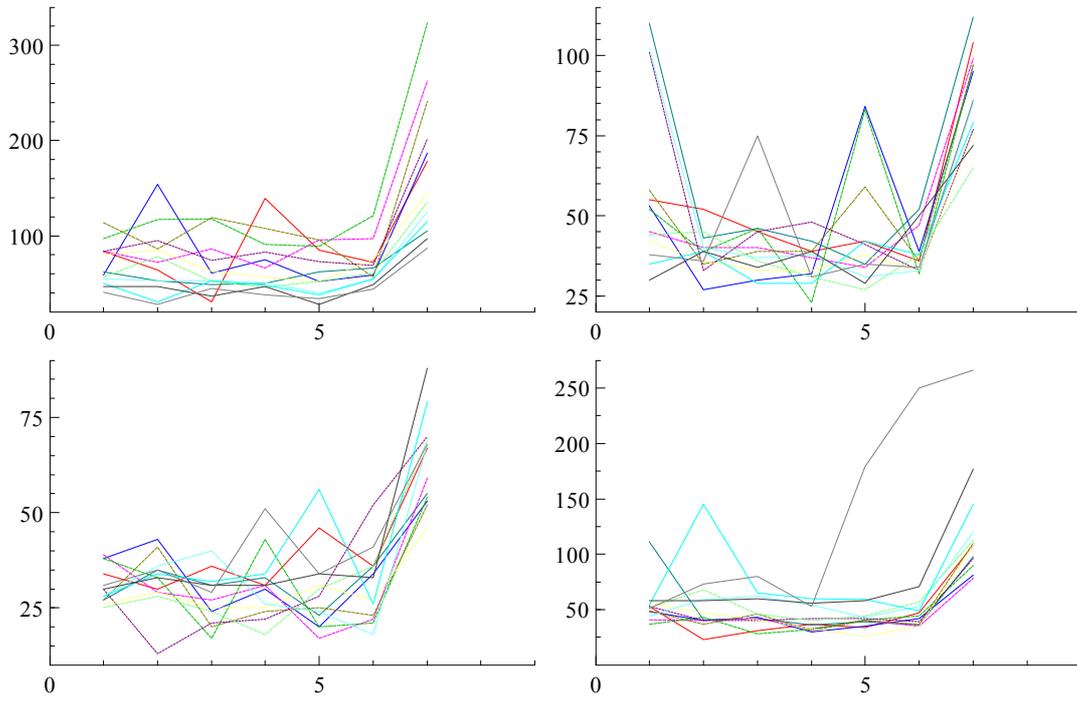


Año 1997

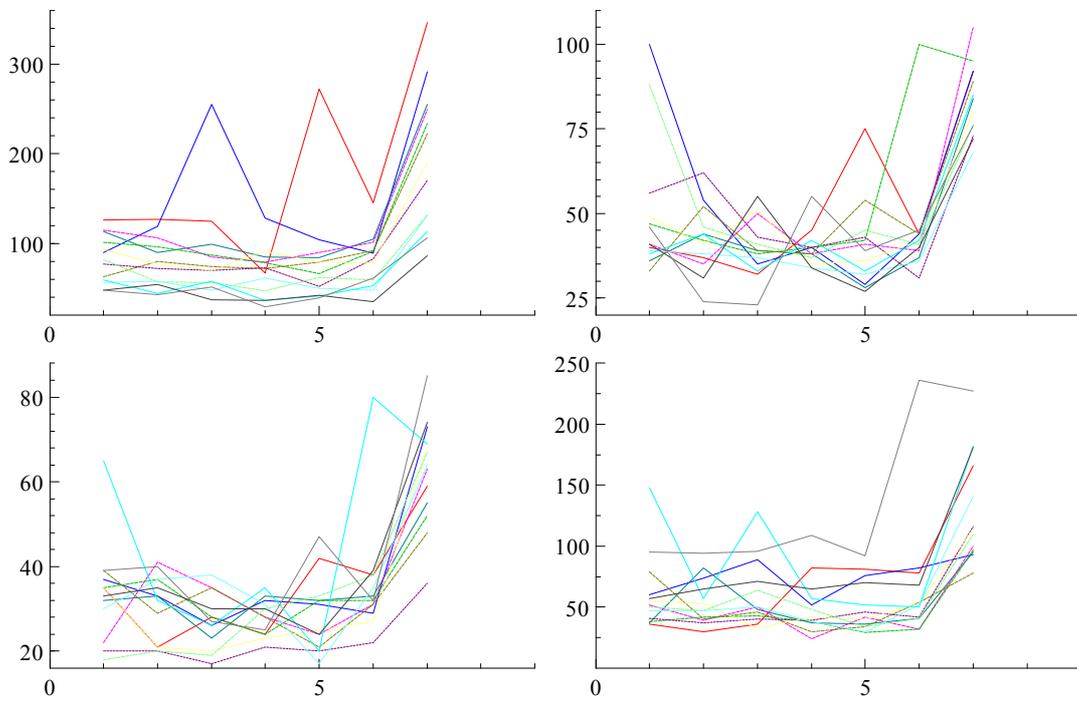


# Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

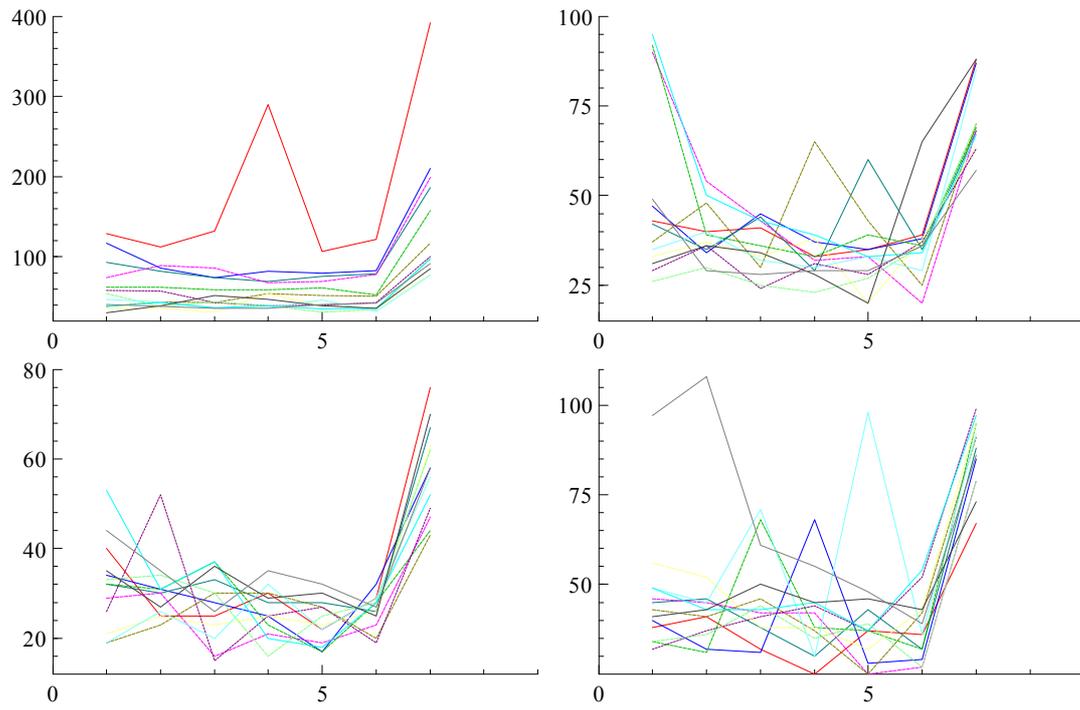
## Año 1998



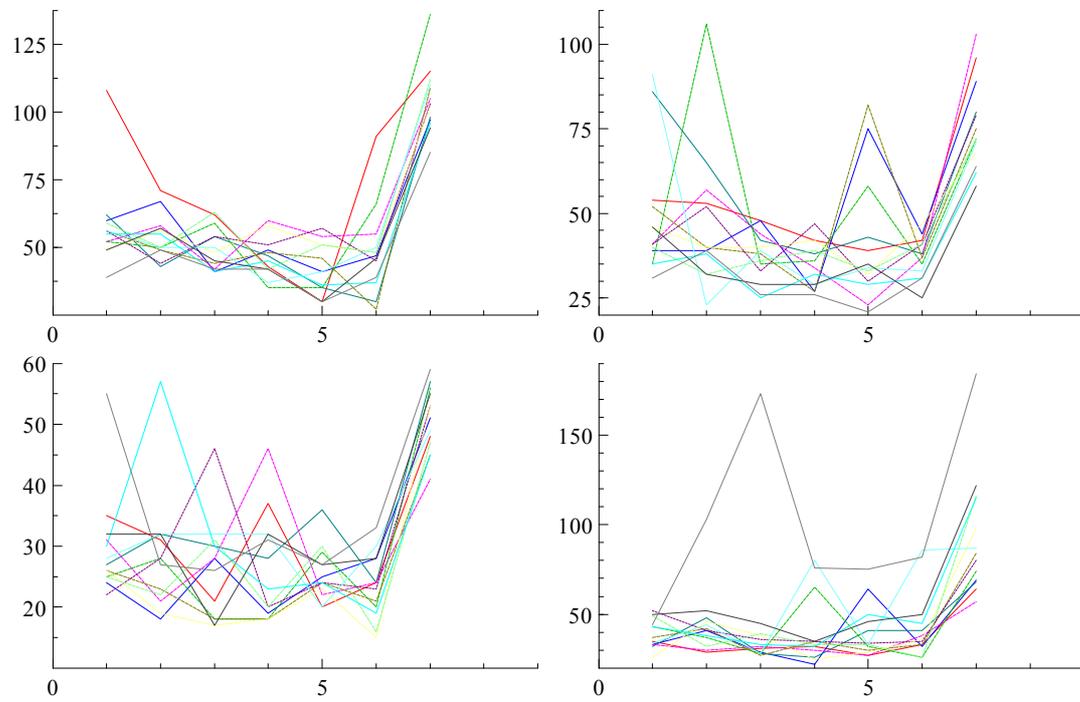
## Año 1999



**Año 2000**

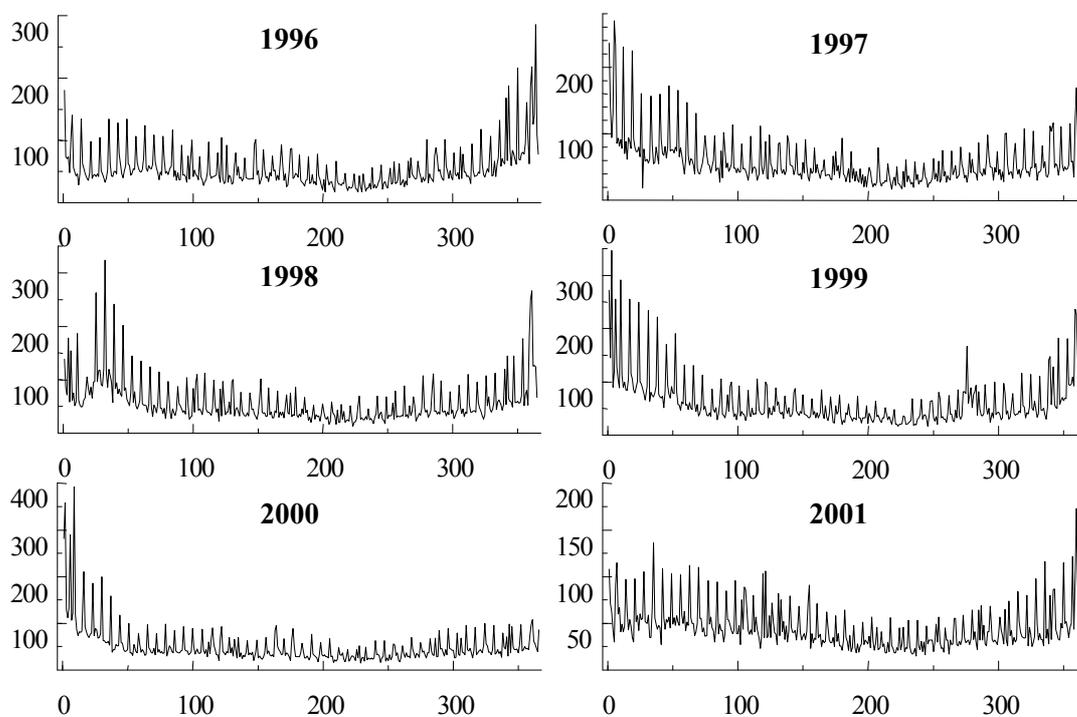


**Año 2001**





## Anexo 4.2. Evolución anual del número diario de visitas registradas en el servicio de urgencias del Barcelonès Nord





## Anexo 4.3. Resultados de la estimación del modelo estructural para la transformación logarítmica de la serie de número diario de visitas

### A. Modelo sin intervenciones

#### Varianzas de los términos de perturbación

	Irregular ( $\hat{s}_e^2$ )	Nivel ( $\hat{s}_h^2$ )
<b>Varianza</b>	0.052505	0.00077640
<b>Ratio señal-ruido</b>	(1.0000)	(0.0148)

En el modelo que incluía estacionalidad semanal estocástica resultó que  $\hat{s}_w^2 = 0$ .

#### Vector de estado final

Componente	Estimación
$\hat{m}_T$	3.6623

#### Parámetros libres en la especificación de la *spline*

Parámetro	Estimación	Estadístico $t$	Parámetro	Estimación	Estadístico $t$
$\hat{g}_{2,0}^*$	0.7784	5.6798	$\hat{g}_{11,0}^*$	-1917.6000	-0.7621(*)
$\hat{g}_{3,0}^*$	1.0636	3.4180	$\hat{g}_{12,0}^*$	-170.6300	-1.7900(*)
$\hat{g}_{4,0}^*$	-0.2319(*)	-1.3894	$\hat{g}_{13,0}^*$	443580	2.1660(*)
$\hat{g}_{5,0}^*$	2.4978	3.2635	$\hat{g}_{14,0}^*$	-436080	-1.7151(*)
$\hat{g}_{6,0}^*$	-29.3560(*)	-1.8318	$\hat{g}_{15,0}^*$	-12774	-1.7311(*)
$\hat{g}_{7,0}^*$	45.8820(*)	1.6897	$\hat{g}_{16,0}^*$	386960	7.0990
$\hat{g}_{8,0}^*$	-1082.400(*)	-1.3089	$\hat{g}_{17,0}^*$	-8.4943x10 <sup>6</sup>	-7.6453
$\hat{g}_{9,0}^*$	-11.1650(*)	-0.9614	$\hat{g}_{18,0}^*$	5.7485x10 <sup>6</sup>	6.8944
$\hat{g}_{10,0}^*$	69.1390	3.8550			

(\*) Parámetros no significativos al 95 % de confianza.

#### Componente estacional de periodo semanal (estimación para la semana del 25 al 31 de diciembre de 2001)

Día	Estimación	Día	Estimación
Martes, 25-12-01	-0.0738	Sábado, 29-12-01	-0.1100
Miércoles, 26-12-01	-0.1374	Domingo, 30-12-01	0.6595
Jueves, 27-12-01	-0.1690	Lunes, 31-12-01	0.0011
Viernes, 28-12-01	-0.1702		

**B. Modelo con intervenciones**

**Varianzas de los términos de perturbación**

	Irregular ( $\hat{s}_e^2$ )	Nivel ( $\hat{s}_h^2$ )
<b>Varianza</b>	0.046426	0.00052916
<b>Ratio señal-ruido</b>	(1.0000)	(0.0114)

En el modelo que incluía estacionalidad semanal estocástica resultó que  $\hat{s}_w^2 = 0$ .

**Vector de estado final**

Componente	Estimación
$\hat{m}_r$	3.6279

**Párametros libres en la especificación de la *spline***

Parámetro	Estimación	Parámetro	Estimación	Parámetro	Estimación
$\hat{g}_{2,0}^*$	0.8054	$\hat{g}_{8,0}^*$	-1207.4000(*)	$\hat{g}_{14,0}^*$	-372010(*)
$\hat{g}_{3,0}^*$	0.7805	$\hat{g}_{9,0}^*$	-7.1154(*)	$\hat{g}_{15,0}^*$	-13195(*)
$\hat{g}_{4,0}^*$	-0.1642(*)	$\hat{g}_{10,0}^*$	55.6600	$\hat{g}_{16,0}^*$	370880
$\hat{g}_{5,0}^*$	2.0549	$\hat{g}_{11,0}^*$	-940.2900(*)	$\hat{g}_{17,0}^*$	-8.2317x10 <sup>6</sup>
$\hat{g}_{6,0}^*$	-28.3040	$\hat{g}_{12,0}^*$	-170.9300	$\hat{g}_{18,0}^*$	5.7071x10 <sup>6</sup>
$\hat{g}_{7,0}^*$	47.3410	$\hat{g}_{13,0}^*$	386730		

(\*) Parámetros no significativos al 95 % de confianza.

**Párametros de las variables de intervención**

Parámetro	Estimación	Parámetro	Estimación	Parámetro	Estimación
<b>En el componente irregular</b>					
$\hat{I}_{01-05-96(f)}$	0.9793	$\hat{I}_{12-10-98(f)}$	0.8409	$\hat{I}_{06-01-00(f)}$	0.9867
$\hat{I}_{06-12-96(f)}$	0.9430	$\hat{I}_{13-04-98(f)}$	0.9404	$\hat{I}_{12-06-00(f)}$	0.9867
$\hat{I}_{27-01-97}$	-1.4812	$\hat{I}_{01-06-98(f)}$	0.9207	$\hat{I}_{01-05-01(f)}$	0.9113
$\hat{I}_{01-11-97(f)}$	0.8808	$\hat{I}_{01-05-99(f)}$	0.8614	$\hat{I}_{04-06-01(f)}$	0.9216
$\hat{I}_{31-12-97(f)}$	-0.8489	$\hat{I}_{11-09-99(f)}$	0.9515		
<b>En el componente de nivel</b>					
$\hat{I}_{19-01-98(v)}$	0.4975	$\hat{I}_{30-09-99}$	0.6593	$\hat{I}_{11-10-99(v)}$	-0.5394

En las intervenciones se ha indicado si ese día fue festivo (f), domingo (d) o víspera de festivo (v).

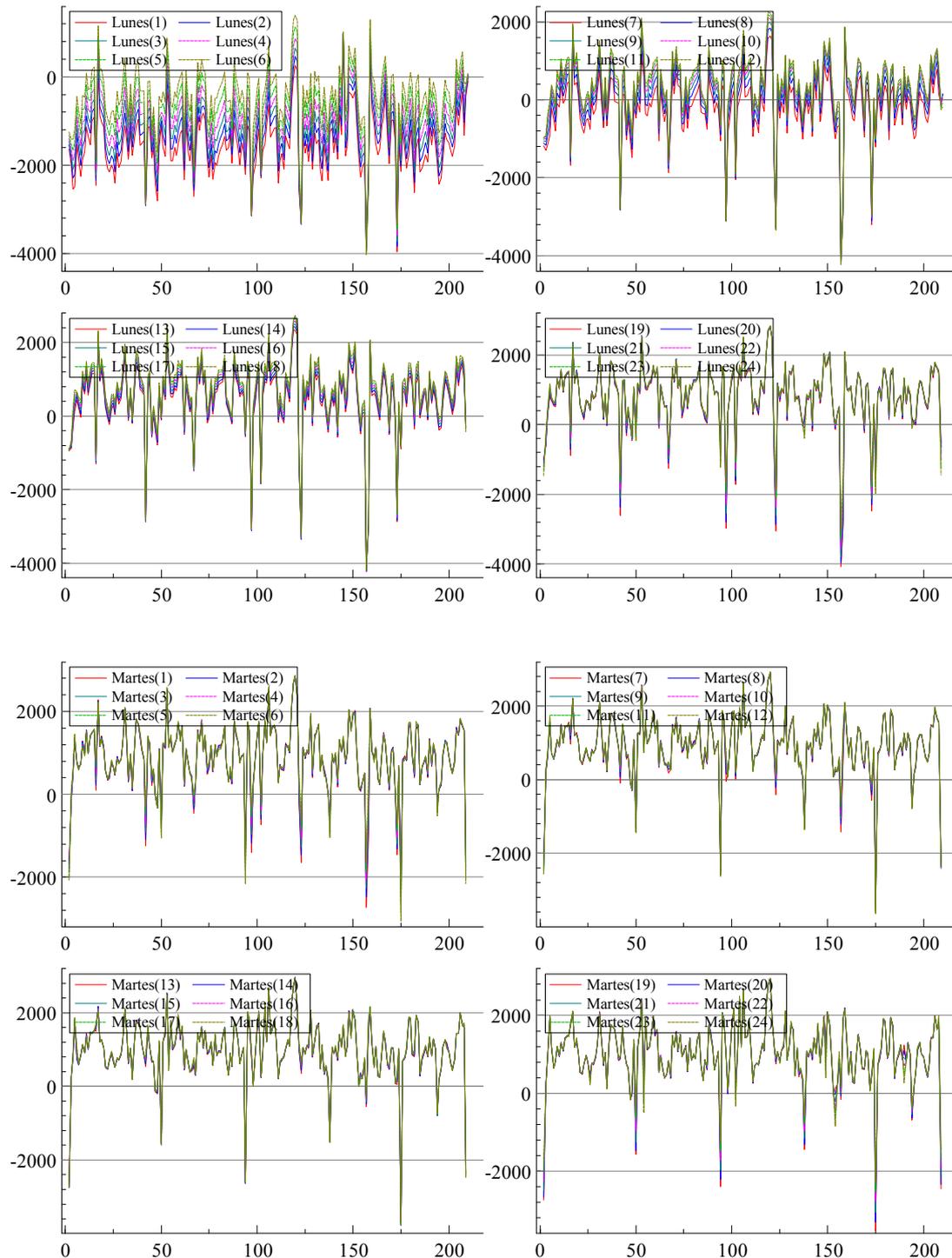
Para la especificación de una intervención en el componente de nivel, véase capítulo 1:21.

**Componente estacional de periodo semanal (estimación para la semana del 25 al 31 de diciembre de 2001)**

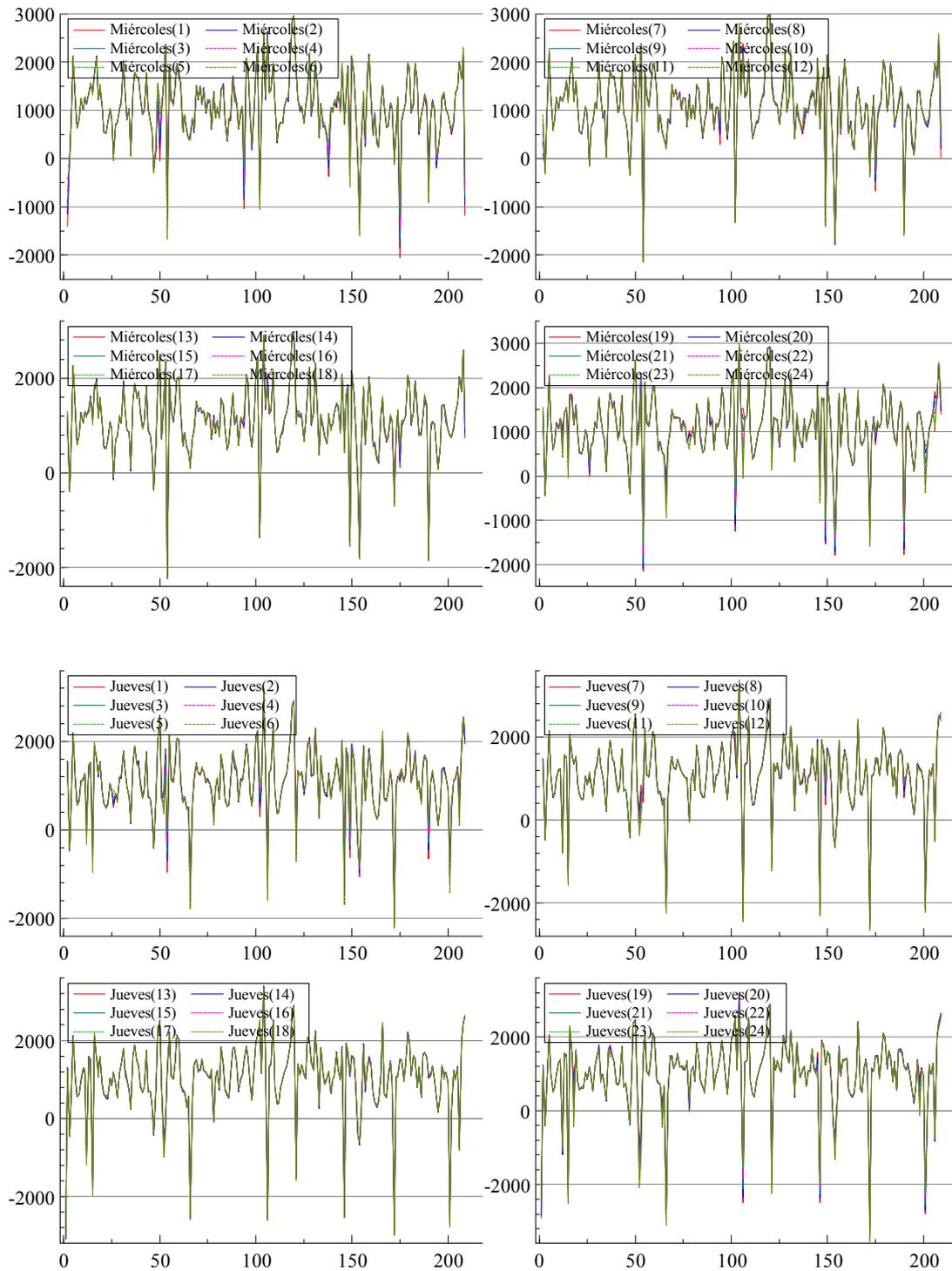
<b>Día</b>	<b>Estimación</b>	<b>Día</b>	<b>Estimación</b>
Martes, 25-12-01	-0.0721	Sábado, 29-12-01	-0.1149
Miércoles, 26-12-01	-0.1330	Domingo, 30-12-01	0.6634
Jueves, 27-12-01	-0.1691	Lunes, 31-12-01	-0.0043
Viernes, 28-12-01	-0.1699		

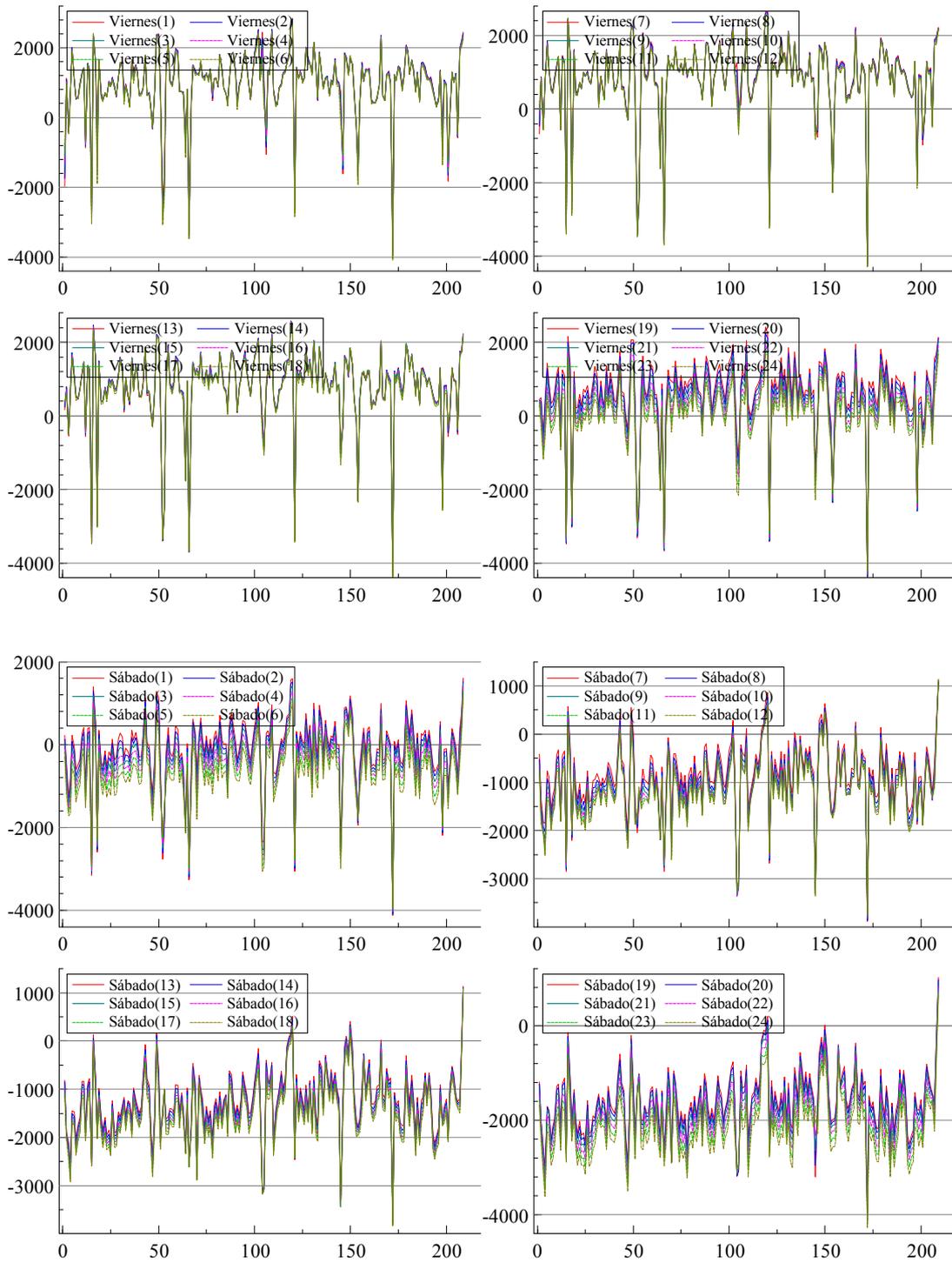


# Anexo 5.1. Evolución por hora de la semana de las medias móviles de periodo 24 sobre la serie de demanda horaria de energía eléctrica filtrada de componente tendencial determinístico y de fluctuación estacional de periodo anual, $\hat{y}_{2,h^a}^{\text{anual}}$

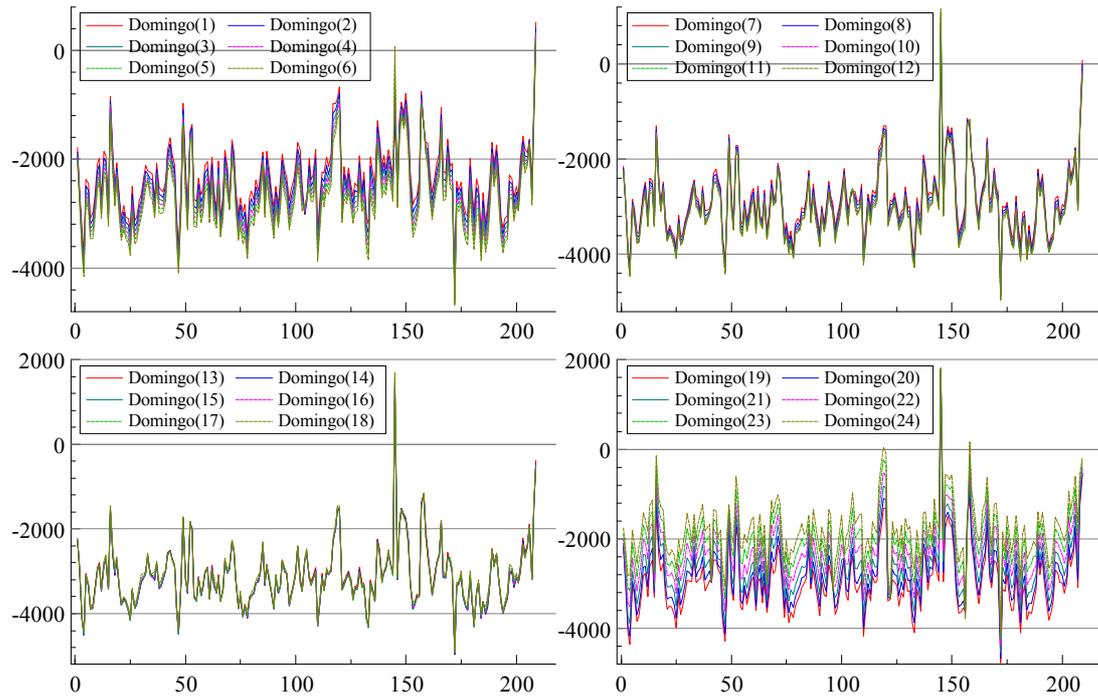


# Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

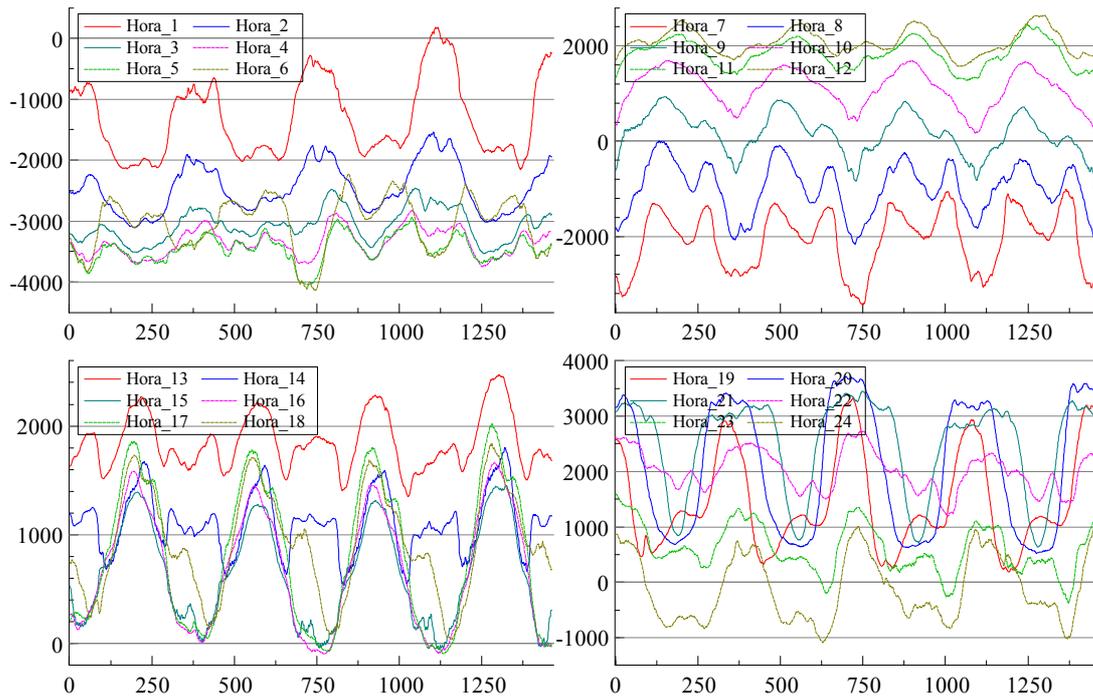




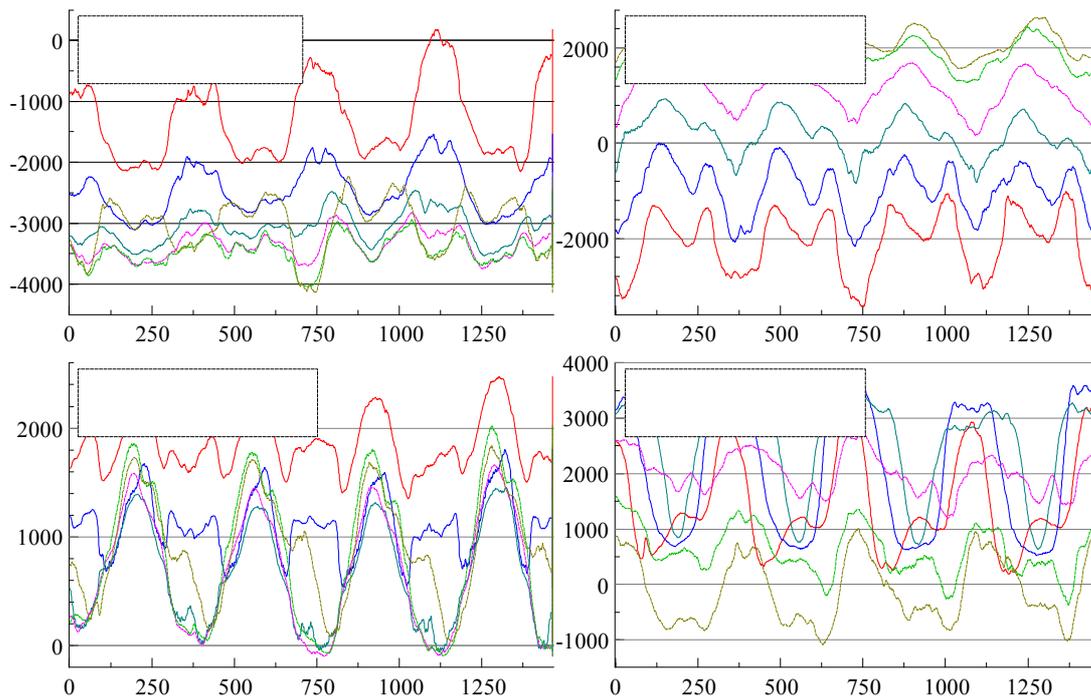
## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia



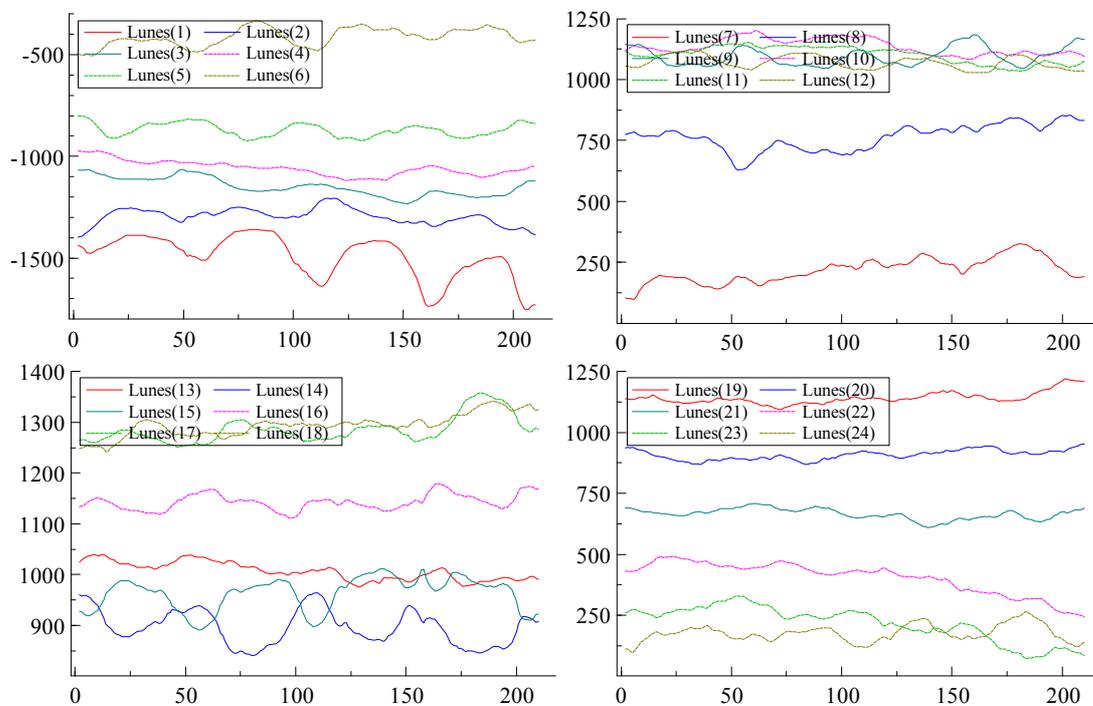
**Anexo 5.2. Evolución por hora del día de la serie de demanda horaria de energía eléctrica filtrada de tendencia lineal determinística y de las fluctuaciones estacionales de periodo anual,  $\{\hat{v}_{2,h^a}^{anual}\}$ , y semanal,  $\{\hat{v}_{2,h^s}^{semanal}\}$**



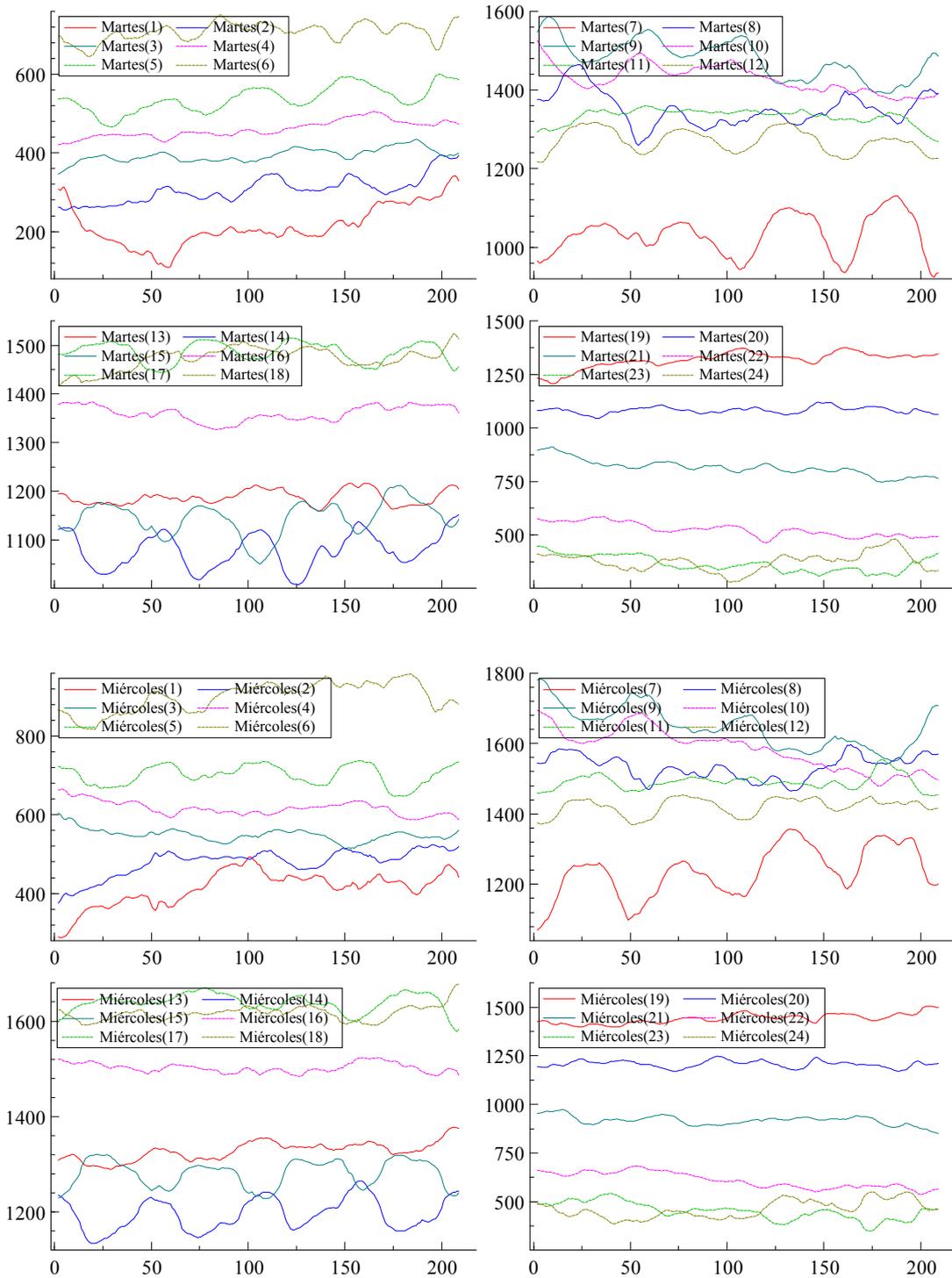
**Anexo 5.3. Evolución por hora del día de la fluctuación estacional diaria estimada en el modelo  $y_t = \mu_t + \gamma_t^{\text{diaria}} + \gamma_t^{\text{semanal}} + \gamma_t^{\text{anual}} + \delta\text{CH}_t + \varepsilon_t$ , con nivel local y fluctuación diaria estocástica y fluctuaciones estacionales anual y semanal determinísticas (*splines*)**

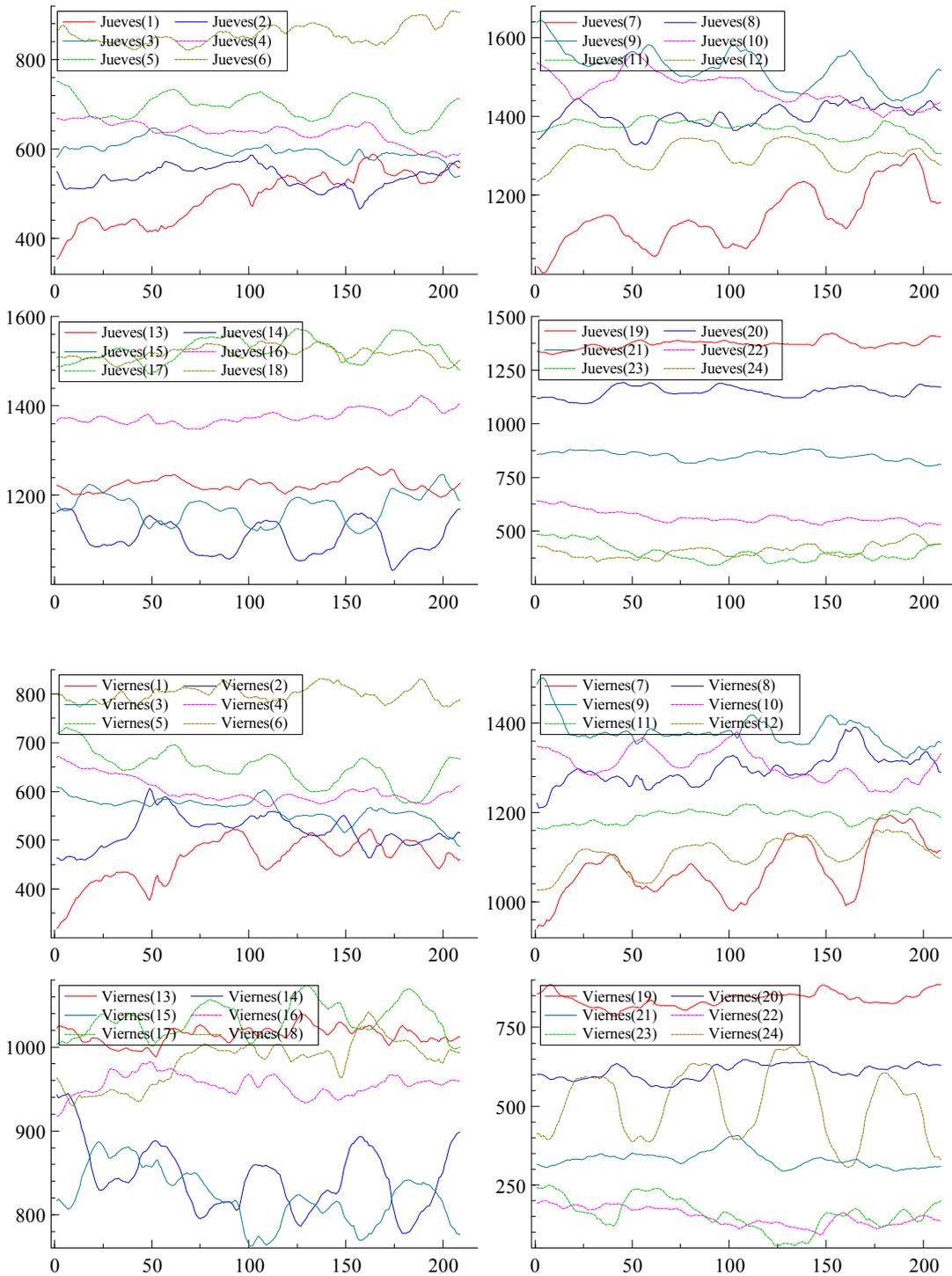


**Anexo 5.4. Evolución por hora de la semana de la fluctuación estacional semanal estimada en el modelo  $y_t^s = \mu_t + \gamma_t^{\text{semanal}} + \varepsilon_t$ , con nivel local y fluctuación semanal estocástica (*dummies*), donde  $y_t^s$  es la serie de demanda horaria de energía eléctrica filtrada de las estimaciones de la fluctuación estacional estocástica de periodo diario y del resto de componentes de naturaleza determinística del modelo  $y_t = \mu_t + \gamma_t^{\text{diaria}} + \gamma_t^{\text{semanal}} + \gamma_t^{\text{anual}} + \delta\text{CH}_t + \varepsilon_t$**

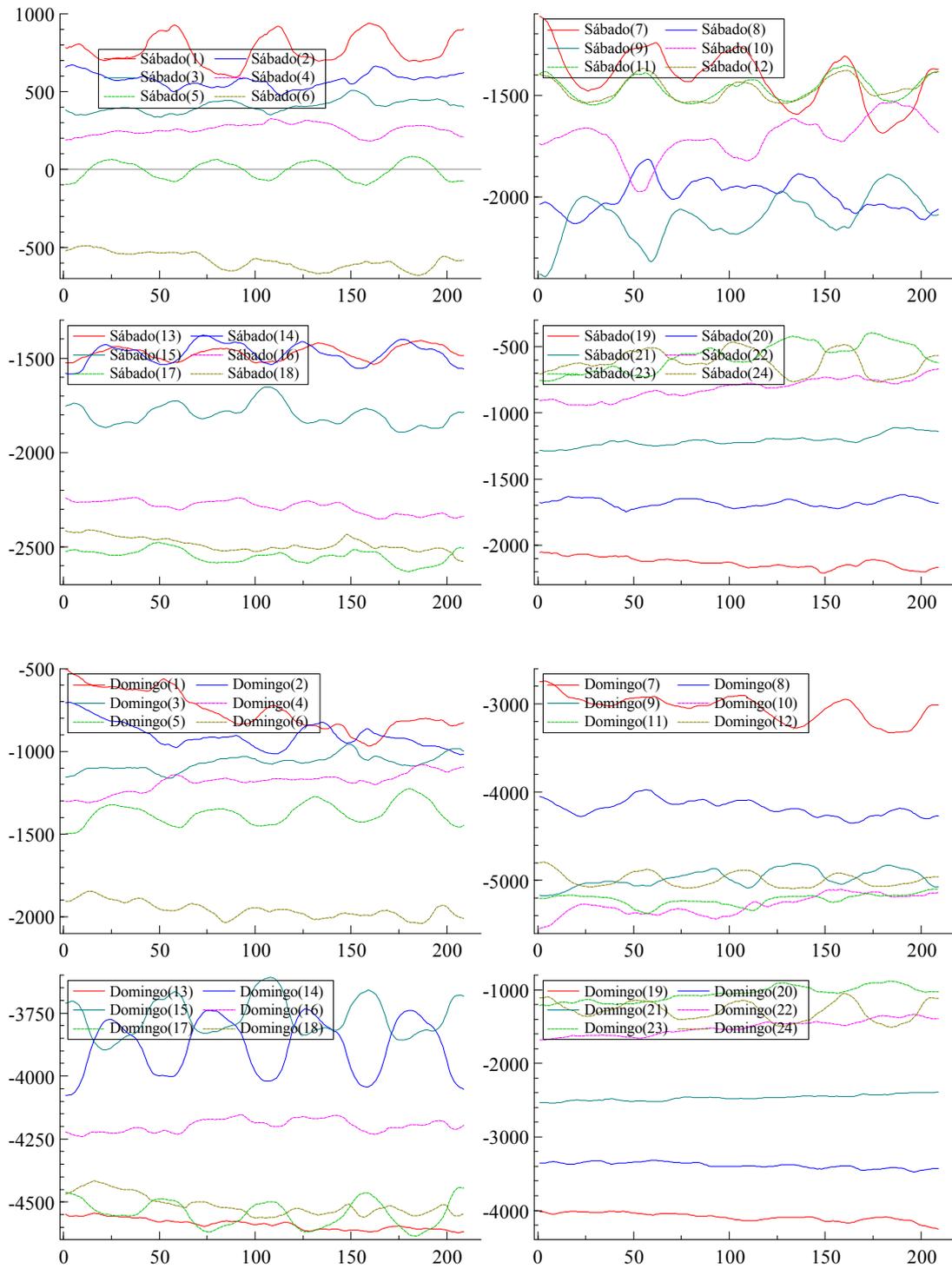


# Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia





# Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia



## Anexo 5.5. Resultados de la estimación del modelo estructural para la transformación logarítmica de la serie de demanda horaria de energía eléctrica

### A. Modelo sin intervenciones

#### Varianzas de los términos de perturbación

	Irregular ( $\hat{s}_e^2$ )	Nivel ( $\hat{s}_h^2$ )	Estacionalidad diaria ( $\hat{s}_w^2$ )
Varianza	0.0000	0.00049696	$2.4659 \times 10^{-8}$
q-ratios	(0.0000)	(1.0000)	(0.0000)

#### Vector de estado final

Componente	Estimación	Componente	Estimación
$\hat{m}_T$	9.8430	$\hat{g}_{6,T}^*$	-0.0013(*)
$\hat{g}_{1,T}$	-0.0051(*)	$\hat{g}_{7,T}$	0.0013(*)
$\hat{g}_{1,T}^*$	-0.1133	$\hat{g}_{7,T}^*$	0.0026(*)
$\hat{g}_{2,T}$	0.0547	$\hat{g}_{8,T}$	-0.0025(*)
$\hat{g}_{2,T}^*$	-0.0488	$\hat{g}_{8,T}^*$	0.0040
$\hat{g}_{3,T}$	-0.0249	$\hat{g}_{9,T}$	0.0040
$\hat{g}_{3,T}^*$	-0.0028(*)	$\hat{g}_{9,T}^*$	0.0052
$\hat{g}_{4,T}$	0.0062	$\hat{g}_{10,T}$	0.0047
$\hat{g}_{4,T}^*$	0.0122	$\hat{g}_{10,T}^*$	-0.0016(*)
$\hat{g}_{5,T}$	0.0095	$\hat{g}_{11,T}$	0.0030
$\hat{g}_{5,T}^*$	0.0085	$\hat{g}_{11,T}^*$	-0.0003(*)
$\hat{g}_{6,T}$	0.0080	$\hat{g}_{12,T}$	-0.0006(*)

(\*) Parámetros no significativos al 90 % de confianza. El valor del estadístico del contraste de significación conjunta del componente estacional de periodo diario fue  $c_{23}^2 = 1470.73$ .

**Parámetros de las variables explicativas**

Parámetro	Estimación	Estadístico t	Parámetro	Estimación	Estadístico t
$\hat{g}_{2,0}^{a*}$	$-5.9406 \times 10^{-1}$	-0.7895	$\hat{g}_{12,0}^{a*}$	$-10604 \times 10^{-1}$	-0.2865
$\hat{g}_{3,0}^{a*}$	$2.1011 \times 10^{-1}$	0.6797	$\hat{g}_{13,0}^{a*}$	$1.6105 \times 10^{-2}$	0.0128
$\hat{g}_{4,0}^{a*}$	2.9166	0.5777	$\hat{g}_0^{29F}$	-0.0001	-1.0811
$\hat{g}_{5,0}^{a*}$	$-6.7300 \times 10^3$	-0.9420	$\hat{g}_1^{29F}$	$2.1540 \times 10^{-2}$	1.0785
$\hat{g}_{6,0}^{a*}$	$2.3838 \times 10^3$	0.9601	$\hat{g}_{2,0}^{s*}$	$-5.4881 \times 10^{-3}$	-36.5100
$\hat{g}_{7,0}^{a*}$	$-5.4118 \times 10^1$	-0.9429	$\hat{g}_{3,0}^{s*}$	$3.1758 \times 10^{-1}$	27.4790
$\hat{g}_{8,0}^{a*}$	$6.5287 \times 10^2$	1.3061	$\hat{g}_{4,0}^{s*}$	$4.4741 \times 10^1$	10.5090
$\hat{g}_{9,0}^{a*}$	-0.0006	-1.6953	$\hat{g}_{5,0}^{s*}$	$2.8779 \times 10^3$	11.2670
$\hat{g}_{10,0}^{a*}$	$6.7925 \times 10^4$	1.4633	$\hat{g}_{6,0}^{s*}$	$-5.3911 \times 10^1$	-22.4430
$\hat{g}_{11,0}^{a*}$	$-4.1076 \times 10^3$	-0.3057	$\hat{d}$	-0.0208	-2.5386

**B. Modelo con intervenciones**

**Varianzas de los términos de perturbación**

	Irregular ( $\hat{s}_e^2$ )	Nivel ( $\hat{s}_h^2$ )	Estacionalidad diaria ( $\hat{s}_w^2$ )
<b>Varianza</b>	0.0000	0.00049399	$2.4693 \times 10^{-8}$
<b>q-ratios</b>	(0.0000)	(1.0000)	(0.0000)

**Vector de estado final**

Componente	Estimación	Componente	Estimación
$\hat{m}_T$	9.8794	$\hat{g}_{6,T}^*$	-0.0013(*)
$\hat{g}_{1,T}$	-0.0050(*)	$\hat{g}_{7,T}$	0.0013(*)
$\hat{g}_{1,T}^*$	-0.1132	$\hat{g}_{7,T}^*$	0.0026(*)
$\hat{g}_{2,T}$	0.0547	$\hat{g}_{8,T}$	-0.0025(*)
$\hat{g}_{2,T}^*$	-0.0488	$\hat{g}_{8,T}^*$	0.0040
$\hat{g}_{3,T}$	-0.0249	$\hat{g}_{9,T}$	0.0040
$\hat{g}_{3,T}^*$	-0.0028(*)	$\hat{g}_{9,T}^*$	0.0052
$\hat{g}_{4,T}$	0.0062	$\hat{g}_{10,T}$	0.0047
$\hat{g}_{4,T}^*$	0.0122	$\hat{g}_{10,T}^*$	-0.0016(*)
$\hat{g}_{5,T}$	0.0095	$\hat{g}_{11,T}$	0.0030
$\hat{g}_{5,T}^*$	0.0085	$\hat{g}_{11,T}^*$	-0.0003(*)
$\hat{g}_{6,T}$	0.0080	$\hat{g}_{12,T}$	-0.0006(*)

(\*) Parámetros no significativos al 90 % de confianza.

El valor del estadístico del contraste de significación conjunta del componente estacional de periodo diario fue  $\chi_{23}^2 = 1471.9$ .

**Parámetros de las variables explicativas**

Parámetro	Estimación	Estadístico t	Parámetro	Estimación	Estadístico t
$\hat{g}_{2,0}^{a*}$	$-5.6362 \times 10^{-1}$	-0.7512	$\hat{g}_{12,0}^{a*}$	$-6.8320 \times 10^{-4}$	-0.1851
$\hat{g}_{3,0}^{a*}$	$1.5984 \times 10^{-1}$	0.5184	$\hat{g}_{13,0}^{a*}$	$-1.4965 \times 10^{-3}$	-0.1192
$\hat{g}_{4,0}^{a*}$	3.1220	0.6202	$\hat{g}_0^{29F}$	-0.0001	-1.0869
$\hat{g}_{5,0}^{a*}$	$-6.7826 \times 10^3$	-0.9523	$\hat{g}_1^{29F}$	$2.1593 \times 10^{-2}$	1.0843
$\hat{g}_{6,0}^{a*}$	$2.4028 \times 10^3$	0.9707	$\hat{g}_{2,0}^{s*}$	$-5.5219 \times 10^{-3}$	-36.8240
$\hat{g}_{7,0}^{a*}$	$-5.5381 \times 10^1$	-0.9678	$\hat{g}_{3,0}^{s*}$	$3.2026 \times 10^{-1}$	27.7850
$\hat{g}_{8,0}^{a*}$	$6.8599 \times 10^2$	1.3764	$\hat{g}_{4,0}^{s*}$	$4.5514 \times 10^1$	10.6330
$\hat{g}_{9,0}^{a*}$	-0.0006	-1.7400	$\hat{g}_{5,0}^{s*}$	$2.8998 \times 10^3$	11.3860
$\hat{g}_{10,0}^{a*}$	$6.9569 \times 10^4$	1.5032	$\hat{g}_{6,0}^{s*}$	$-5.4083 \times 10^1$	-22.5760
$\hat{g}_{11,0}^{a*}$	$-4.6962 \times 10^3$	-0.3506	$\hat{d}$	-0.0208	-2.5460

**Parámetros de las variables de intervención**

Parámetro	Estimación	Estadístico t	Parámetro	Estimación	Estadístico t
$\hat{I}_{81-8}$	-0.1188	-5.1300	$\hat{I}_{1097-9}$	-0.1073	-4.6352
$\hat{I}_{366-1}$	0.1036	4.4723	$\hat{I}_{1111-1}$	-0.1097	-4.7381
$\hat{I}_{410-9}$	-0.1188	-5.1312	$\hat{I}_{1118-1}$	-0.1243	-5.3682
$\hat{I}_{731-1}$	0.1224	5.2853	$\hat{I}_{1405-1}$	-0.1199	-5.1805

**Componente estacional de periodo diario (estimación correspondiente al lunes 31 de diciembre de 2001)**

Hora del día	Estimación	Hora del día	Estimación	Hora del día	Estimación
1	-0.0072	9	-0.0544	17	0.0034
2	-0.0872	10	0.0024	18	0.0348
3	-0.1345	11	0.0699	19	0.1368
4	-0.1498	12	0.0825	20	0.1517
5	-0.1609	13	0.0789	21	0.1364
6	-0.1632	14	0.0582	22	0.1033
7	-0.1324	15	0.0202	23	0.0515
8	-0.1043	16	0.0053	24	0.0584

**C. Modelo con estacionalidad semanal estocástica**

**Varianzas de los términos de perturbación**

	Irregular ( $\hat{s}_e^2$ )	Nivel ( $\hat{s}_h^2$ )	Estacionalidad semanal ( $\hat{s}_w^2$ )
Varianza	$1.1915 \times 10^{-5}$	0.00020442	0.0000
q-ratios	0.0583	1.0000	0.0000

## Bibliografía

- Abraham, B. y G.E.P. Box (1978) «Deterministic and forecast-adaptive time dependent models», *Applied Statistics* **27**:120-130.
- Albertos, J., A. Martínez y J.A. Sans (1987) *La Agricultura Canaria en la Comunidad Europea*, Consejería de Agricultura, Ganadería y Pesca. Gobierno de Canarias.
- Aldanondo, A. (1995) «Cultivo y producción de tomate en la Unión Europea y en España», en F. Nuez (ed.) *El Cultivo del Tomate*, Mundi Prensa, 695-740.
- Anderson, B.D.O. y J.B. Moore (1979) *Optimal Filtering*. Prentice-Hall.
- Anderson, T.W. (1971) *The Statistical Analysis of Time Series*. Wiley.
- Andrén, K.G. y U. Rosenqvist (1985) «Heavy users of an emergency department: psychosocial and medical characteristics, other health care contacts and the effect of a hospital social worker intervention», *Social Science & Medicine* **21**:761-770.
- Andrén, K.G. y U. Rosenqvist (1987) «Heavy users of an emergency department: a two year follow-up study», *Social Science & Medicine* **25(7)**:825-831.
- Ansley, C.F. y R. Kohn (1982) «A geometrical derivation of the fixed interval smoothing algorithm», *Biometrika* **69**:486-7.
- Ansley, C.F. y R. Kohn (1985) «Estimation, filtering and smoothing in state space models with incompletely specified initial conditions», *Annals of Statistics* **13**:1286-1316.
- Ansley, C.F. y R. Kohn (1986) «Prediction mean square error for state space models with estimated parameters», *Biometrika* **73**:467-74.
- Ansley, C.F. y R. Kohn (1990) «Filtering and smoothing in state space models with partially diffuse initial conditions», *Journal of Time Series Analysis* **11**:275-293.
- Aoki, M. (1990) *State Space Modelling of Time Series*. Springer-Verlag.
- Atkinson, A.C., S.J. Koopman y N. Shephard (1997) «Detecting shocks: outliers and breaks in time series», *Journal of Econometrics* **80**:387-422.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

- Bajo-Rubio, O., S. Sosvilla-Rivero y F. Fernández-Rodríguez (2002) «Non-Linear Forecasting Methods: Some applications to the Analysis of Financial Series», *Documento de Trabajo de la Fundación de Estudios de Economía Aplicada* **2002-01**.
- Banerjee, A., J. Dolado, J.W. Galbraith y D.F. Hendry (1993) *Cointegration, Error Correction and the Econometric Analysis of Non-stationary data*. Oxford University Press.
- Beaulieu, J.J. y J.A. Miron (1992) «A cross country comparison of seasonal cycles and business cycles», *The Economic Journal* **102**:772-788.
- Bell, W.R. y S.C. Hillmer (1983) «Modelling time series with calendar variation», *Journal of the American Statistical Association* **78**(383):526-534.
- Bell, W.R. y S.C. Hillmer (1991) «Initializing the Kalman filter for nonstationary time series models», *Journal of the Time Series Analysis* **12**:283-300.
- Benz, J.R. y J.C. Shank (1982) «Alteration of emergency room usage in a family practice residency programme», *Journal of Family Practice* **15**:1135-1139.
- Billón, M. (1995) *La Exportación Hortofrutícola. El Caso del Albaricoque en Fresco y la Lechuga Iceberg*. MAPA-SGT. Serie Estudios.
- Birnbaum, A.E., J. Gallagher, M. Utkewicz, P. Gennis y W. Carter (1994) «Failure to validate a predictive model for refusal of care to emergency-department patients», *Academic Emergency Medicine* **1**:213-217.
- Bohara, A.K. y C. Sauer (1992) «Competing Macro-Hyphotesis in the United States: A Kalman filter approach», *Applied Economics* **24**:389-399.
- Bolíbar, I., X. Balanzó, A. Armada, J.L. Fernández, G. Foz, E. Sanz y M. De la Torre (1996) «El impacto de la reforma de la atención primaria en la utilización de servicios de urgencias hospitalarios», *Medicina Clínica* **107**:289-295.
- Bolland, P.J., J.T. Connor y A.P.N. Refenes (1998) «Application of neural network to forecast high frequency data: foreign exchange», en C. Dunis y B. Zhou (eds.) *Nonlinear Modelling of High Frequency Financial Time Series*, John Wiley & Sons, 225-246.
- De Boor, C. (1978) *A Practical Guide to Splines*. Springer-Verlag.
- Box, G.E.P. y G.C. Tiao (1975) «Intervention analysis with applications to economic and environmental problems», *Journal of the American Statistical Association* **70**:70-79.

- Box, G.E.P. y G.M. Jenkins (1976) *Time Series Analysis: forecasting and control*. Holden Day.
- Bricio, C. (2001): *Algunos factores no médicos asociados a las variaciones de la demanda asistencial en atención continuada. Utilidad de un modelo de predicción basado en estos factores*. Tesina del Máster de Economía de la Salud y Gestión Sanitaria. UPF. Barcelona. Mimeo.
- Briz, J. (1990) «La comercialización de productos hortofrutícolas en la CEE», en *III Jornadas de Transferencia Tecnológica. El cultivo del tomate*. Universidad de La Laguna, Consejería de Agricultura, Ganadería y Pesca, Cabildo Insular de Tenerife, Cooperativa Agrícola de Guía de Isora. Ediciones y Promociones LAV, S.L., 289-324.
- Brockwell, P.J. y R.A. Davis (1991) *Time Series: theory and methods*. Springer.
- Bull, C. y R. Frydman (1983) «The derivation and interpretation of the Lucas supply function», *Journal of Money, Credit and Banking* **15**:82-95.
- Burmeister, E., K.D. Wall y J.D. Hamilton (1986) «Estimation of unobserved expected monthly inflation usign Kalman filtering», *Journal of Business and Economic Statistics* **4**:147-160.
- Burrige, P. y K.F. Wallis (1985) «Calculating the variance of seasonally adjusted series», *Journal of the American Statistical Association*, **80**:541-552.
- Buse, A. y L. Lim (1977) «Cubic splines as a special case of restricted least squares», *Journal of the American Statistical Association* **72**:64-68.
- Cabrero, A., G. Camba-Méndez, A. Hirsch y F. Nieto (2002) «Modelling the daily banknotes in circulation in the context of the liquidity management of the European Central Bank», Documento de trabajo del *Servicio de Estudios del Banco de España* **0211**.
- Cáceres, J.J. (1997) *El Tomate Canario de Exportación. Estacionalidad de las series semanales de oferta y precios*. Universidad de La Laguna (CD-ROM).
- Cáceres, J.J. (2000). *La Exportación de Tomate en Canarias. Elementos para una estrategia competitiva*. Ediciones Canarias.
- Cáceres, J.J. (2001) «Optimalidad del patrón estacional de las exportaciones canarias de tomate», *Estudios de Economía Aplicada* **18**:41-66.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

- Canova, F. (1992) «An alternative approach to modelling and forecasting seasonal time series», *Journal of Business and Economic Statistics* **10(1)**:97-108.
- Chan, L.S., M.A. Galaif y C.L. Kushi (1985) «Referrals from hospital emergency departments to primary care centers for nonurgent care», *Journal of Ambulatory Care Management* **8**:57-69.
- Chande, V.T., N. Wyss y V. Exum (1996) «Educational interventions to alter pediatric emergency department utilization patterns», *Archives of Pediatrics & Adolescent Medicine* **150**:525-528.
- CIES (1968) «Economía Canaria 67», *Centro de Investigación Económica y Social de la Caja Insular de Ahorros de Gran Canaria* **2**.
- Cortés, R. (1989) *El Comercio Exterior del Tomate*. FEPEX.
- Dale, J., J. Green, F. Reid y E. Glucksman (1995) «Primary care in the accident and emergency department: I prospective identification of patients», *British Medical Journal* **311**:423-426.
- Dale, J., J. Green, F. Reid, E. Glucksman y R. Higgs (1995) «Primary care in the accident and emergency department: II comparison of general practitioners and hospital doctors», *British Medical Journal* **311**:427-430.
- Dale, J., H. Lang, J.A. Roberts, J. Green y E. Glucksman (1996) «Cost effectiveness of treating primary care patients in accident and emergency: a comparison between general practitioners, senior house officers and registrars», *British Medical Journal* **312**:1340-1344.
- Derlet, R.W. (2002) «Overcrowding in emergency department: increased demand and decreased capacity», *Annals of Emergency Medicine* **39(4)**:430-432.
- Derlet, R.W., D. Kinser y L. Ray (1995) «Prospective identification and triage of nonemergency patients out of an emergency department: A 5-year study», *Annals of Emergency Medicine* **25**:215-244.
- Derlet, R.W. y D.A. Nishio (1990) «Refusing care to patients who present to an emergency department», *Annals of Emergency Medicine* **19**:262-267.
- Derlet, R.W., D.A. Nishio, L.M. Cole y J. Silva (1992) «Triage of patients out of the emergency department: three-year experience», *American Journal of Emergency Medicine* **10**:195-199.

- Derlet, R.W. y J.R. Richards (2000) «Overcrowding in the nation's emergency departments: complex causes and disturbing effects», *Annals of Emergency Medicine* **35(1)**:63-68.
- Driscoll, P.A., C.A. Vincent y M. Wilkinson (1987) «The use of the accident and emergency department», *Archives of Pediatrics & Adolescent Medicine* **4**:77-82.
- Dunis, C. y B. Zhou (1998) *Nonlinear Modelling of High Frequency Financial Time Series*, John Wiley & Sons.
- Durbin, J. y S.J. Koopman (2001) *Time Series Analysis by State Space Models*. Oxford University Press.
- Engle, R.F. (1978) «Estimating structural models of seasonality», en A. Zellner (ed.) *Seasonal Analysis of Economic Time Series*, 281-308, Bureau of the Census.
- Engle, R. F. (2000) «The Econometrics of Ultra-High-Frequency Data», *Econometrica* **68(1)**:1-22.
- Engle, R.F. y M.W. Watson (1981) «A one-factor multivariate time series model of metropolitan wage rates», *Journal of the American Statistical Association* **76**:744-781.
- Espasa, A. y J.R. Cancelo (1993) *Métodos cuantitativos para el análisis de la coyuntura económica*. Alianza Editorial.
- FEDEX (1994) «Islas Canarias, más de un siglo de exportación», *Fruits & Legumes Distribution* **8**:25-27.
- Fernández, F.J. y A.C. Harvey (1990) «Seemingly unrelated time series equations and a test for homogeneity», *Journal of Business and Economic Statistics* **8**:71-82.
- Fleming, G.V. y R.M. Andersen (1986) «The municipal health services program: improving access to primary care without increasing expenditures», *Medical Care* **24**:565-579.
- Foldes, S.S., L.R. Fischer y K. Kaminsky (1994) «What is an emergency? The judgments of two physicians», *Annals of Emergency Medicine* **23**:833-840.
- Franses, P.H. (1996a) «Recent Advances in Modelling Seasonality», *Journal of Economic Surveys* **10**:299-345.
- Franses, P.H. (1996b) *Periodicity and Stochastic Trends in Economic Time Series*. Oxford University Press.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

- Franses, P.H. (1997) «Seasonality in economic time series», en A. Ullah y D.E.A. Giles (eds.) *Handbook of Applied Economic Statistics* (Documento de Trabajo del Econometric Institute Report).
- Fuller, W.A. (1969) «Grafted polynomials as approximating functions», *Australian Journal of Agricultural Economics* **13**:35-46.
- Gallant, A.R. y W.A. Fuller (1973) «Fitting segmented polynomial regression model whose join points have to be estimated», *Journal of the American Statistical Association* **68**:144-147.
- Ghysels, E. (1994) «On the economics and econometrics of seasonality», en C.A. Sims (ed.) *Advances in Econometrics*. Sixth World Congress, Cambridge University Press. vol **1**, cap. 7, 257-316.
- Ghysels, E. y D.R. Osborn (2001) *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press.
- Gil, J.M. y A. Gil (1988) *Informe sobre Servicios de Urgencia del Sistema Sanitario Público*. Madrid. Defensor del Pueblo.
- Gill, J.M. (1999) «Use of hospital emergency departments for nonurgent care: a persistent problem with no easy solutions», *American Journal of Managed Care* **5(12)**:1565-1568.
- Gill, J.M y J.J. Diamond (1996) «Effect of primary care referral on emergency department use: evaluation of a statewide Medicaid program», *Family Medicine* **28**:178-182.
- Gill, J.M., A.G. Mainous y M. Nsereko (2000) «The effect of continuity of care on emergency department use», *Archives of Family Medicine* **9(4)**:333-338.
- Gill, J.M., C.L. Reese y J.J. Diamond (1996) «Disagreement among health care professional about the urgent care needs of emergency department patients», *Annals of Emergency Medicine* **28(5)**:474-479.
- Gómez, V y A. Maravall (1994) «Estimation, prediction and interpolation for non-stationary time series with the Kalman filter», *Journal of the American Statistical Association* **89(426)**:611-624.
- Granger, C.W.J. (1978) «Seasonality: causation, interpretation and implications», en A. Zellner (ed.) *Seasonal Analysis of Economic Time Series*, Bureau of the Census.
- Grossman, L.K., L.N. Rich y C. Johnson (1998) «Decreasing nonurgent emergency department utilization by Medicaid children», *Pediatrics* **102**:20-24.

- Hamilton, D. (1994) *Time Series Analysis*. Princeton University Press.
- Hannan, E.J. y M. Deistler (1988) *The Statistical Theory of Linear System*. Wiley.
- Hannan, E.J., R.D. Terrell y N. Tuckwell (1970) «The seasonal adjustment of economic time series», *International Economic Review* **11**:24-52.
- Hansagi, H., P. Allebeck y O. Edhag (1989) «Health care utilization after referral from a hospital emergency department», *Scandinavian Journal of Social Medicine* **17**:291-299.
- Hansagi, H., B. Carlsson, M. Olsson y O. Edhag (1987) «Trial of a method of reducing inappropriate demands on a hospital emergency department», *Public Health* **101**:99-105.
- Harrison, P.J. y C.F. Stevens (1976) «Bayesian forecasting», *Journal of the Royal Statistical Society series B* **38**:205-247.
- Hartley, H.O. (1961) «The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares», *Technometrics* **3**:269-280.
- Harvey, A.C. (1985) «Trends and cycles in macroeconomic time series», *Journal of Business and Economic Statistics* **3**:216-227.
- Harvey, A.C. (1989) *Forecasting, Structural Time Series and the Kalman Filter*. Cambridge University Press.
- Harvey, A.C. (1993) *Time Series Model*. Harvester Wheatsheaf.
- Harvey, A.C. (1997) «Trend, cycles and autoregressions», *The Economic Journal* **107**:192-201.
- Harvey, A.C. (2000) «Trend Analysis». Mimeo.
- Harvey, A.C., B. Henry, S. Peters y S. Wren-Lewin (1986) «Stochastic trends in dynamic regression models: an application to the employment-output equation», *Economic Journal* **96**:975-985.
- Harvey, A.C. y A. Jaeger (1993) «Detrending, stylized facts and the business cycles», *Journal of Applied Econometrics* **8**:231-247.
- Harvey, A.C. y S.J. Koopman (1992) «Diagnostic checking of unobserved components time series models», *Journal of Business and Economic Statistics* **10**:377-389.
- Harvey, A.C. y S.J. Koopman (1993) «Forecasting hourly electricity demand using time varying splines», *Journal of the American Statistical Association* **88**:1228-1236.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

- Harvey, A.C. y S.J. Koopman (1996) «Multivariate structural time series models», en C. Heij, H. Schumacher, B. Hanzon y C. Praagman (eds.) *Systematic Dynamics in Economic and Financial Models*. John Wiley and Sons.
- Harvey, A.C., S.J. Koopman y J. Penzer (1997) «Messy Time Series: A Unified Approach», *LSE Discussion Paper EM/97/327*.
- Harvey, A.C., S.J. Koopman y M. Riani (1997) «The modelling and seasonal adjustment of weekly observations», *Journal of Business and Economic Statistics* **15**(3):354-368.
- Harvey, A.C. y P. Marshall (1991) «Inter-fuel substitution, technical change and the demand for energy in the UK economy», *Applied Economics* **23**:1077-1086.
- Harvey, A.C. y S. Peters (1990) «Estimation procedures for structural time-series models», *Journal of Forecasting* **9**:89-108.
- Harvey, A.C. y G.D.A. Phillips (1979) «The estimation of regression models with autoregressive-moving average disturbance», *Biometrika* **66**:49-58.
- Harvey, A.C. y R.G. Pierse (1984) «Estimating missing observations in economic time series», *Journal of the American Statistical Association* **79**:125-131.
- Harvey, A.C. y A. Scott (1994) «Seasonality in dynamic regression models», *Economic Journal* **104**:1324-1345.
- Harvey, A.C. y N. Shephard (1993) «Structural time series models», en G.S. Maddala, C.R. Rao y H.D. Vinod (eds.) *Handbook of Statistics* **11**. Elsevier Science Publishers.
- Harvey, A.C. y M. Streibel (1996) «Test for deterministic versus indeterministic cycles», *LSE Statistics Research Report, LSERR29*.
- Harvey, A.C. y P.H. Todd (1983) «Forecasting economic time series with structural and Box-Jenkins models» (with discussion), *Journal of Business & Economic Statistics* **1**:299-315.
- Hilditch, J. R. (1980) «Changes in hospital emergency department use associated with increased family physician availability», *Journal of Family Practice* **11**:91-96.
- Hudson, D.J. (1966) «Fitting segmented curves whose join points have to be estimated», *Journal of the American Statistical Association* **61**:1097-1129.
- Hylleberg, S. (1986) *Seasonality in Regression*. Academic Press Inc.
- Hylleberg, S. (1992) *Modelling Seasonality*. Oxford University Press.

- Hylleberg, S. (1994a) «The economics of seasonal cycles», en C.A. Sims (ed.) *Advances in Econometrics*. Sixth World Congress, Cambridge University Press. Vol 1, cap. 6, 252-255.
- Hylleberg, S. (1994b) «Modelling seasonal variation», en C.P. Hargreaves (ed.) *Nonstationary Time Series Analysis and Cointegration*, 153-178. Oxford University Press.
- IDAE (2001) *Ahorro de Energía en el Sector Hotelero: Recomendaciones y Soluciones de Bajo Riesgo*. Instituto para la Diversificación y Ahorro de la Energía.
- Informe sobre la Explotación del Sistema Eléctrico Nacional, *Delegación del Gobierno en la Explotación del Sistema Eléctrico*. Varios años.
- Jazwinski, A.H. (1970) *Stochastic Processes and Filtering Theory*. Academic Press.
- Johnston, J. (1987) *Métodos de Econometría*. Vicens Universidad.
- Jones, R.H. (1993) *Longitudinal Data with Serial Correlation: A state-space approach*. Chapman & Hall.
- De Jong, P. (1991) «The diffuse Kalman filter», *Annals of Statistics* **19**:1073-1083.
- De Jong, P. y Chu-Chun-Lin, S. (1994) «Stationary and non-stationary state space models», *Journal of Time Series Analysis* **15(2)**:151-166.
- Kaiser, R. y A. Maravall (1999) «Seasonal outliers in time series», *Working Paper Statistics and Econometrics Series 15 de la Universidad Carlos III de Madrid* **99-49**.
- Kalman, R.E. (1960) «A new approach to linear filtering and prediction problems», *Transactions ASME, Series D, Journal of Basic Engineering* **82**:35-45.
- Kalman, R.E. y R.S. Bucy (1961) «New results in linear filtering and prediction theory», *Transactions ASME, Series D, Journal of Basic Engineering* **83**:95-108.
- Kaplan, D.W., B.N. Calonge, B.P. Guernsey y M.B. Hanrahan (1998) «Managed care and school-based health centres. Use of health services», *Archives of Pediatrics & Adolescent Medicine* **152**:25-33.
- Kim, S. (1993) «Sources of monetary growth, uncertainty and economic activity: the time varying parameter model with heteroskedastic disturbances», *The Review of Economics and Statistics* **74**:483-492.
- Kitagawa, G. (1984) «State space modelling of nonstationary time series and smoothing of unequally spaced data», en E. Parzen (ed) *Time Series Analysis of Irregularly Observed Data*, Lectures Notes in Statistics **25**:189-210. Springer-Verlag, Nueva York.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

- Kitagawa, G. (1987) «Non-gaussian state space modelling of nonstationary time series», *Journal of the American Statistical Association* **82**:1032-1063.
- Kitagawa, G. y W. Gersch (1984) «A smoothness prior-state space modelling of time series with trend and seasonality», *Journal of the American Statistical Association* **79**:378-389.
- Kohn, R. y C.F. Ansley (1983) «Fixed interval estimation in state space models when some of the data are missing or aggregated», *Biometrika* **70**:683-8.
- Koopman, S.J. (1992) *Diagnostic checking and intra-daily effects in time series models*. Tinbergen Institute research series **27**.
- Koopman, S.J. (1993) «Disturbance smoother for state space models», *Biometrika* **80**:117-126.
- Koopman, S.J. (1996) «Kalman filtering and smoothing». Mimeo.
- Koopman, S.J. (1997) «Exact initial Kalman filtering and smoothing for non-stationary time series models», *Journal of the American Statistical Association* **92**:1630-8.
- Koopman, S.J., A.C. Harvey, J.A. Doornik y N. Shephard (2000) *Stamp 6.0: Structural Time Series Analyser, Modeller and Predictor*. Timberlake Consultants.
- Liu, L.M. (1980) «Analysis of time series with calendar effects», *Management Science* **26**(1):106-112.
- Liu, T.C. (1963) «An exploratory quarterly econometric model of effective demand in the postwar U.S. economy», *Econometrica* **31**:301-348.
- Liu, T.C. (1969) «A monthly recursive model of the U.S. economy», *The Review of Economics and Statistics* **51**:1-13.
- De los Llanos, M<sup>a</sup> y A. Valentina (1994) «Métodos para la extracción de señales y para la trimestralización. Una aplicación: trimestralización del deflactor del consumo privado nacional». *Servicio de estudios del Banco de España* **9415**.
- Long, S.H. y F.F. Settle (1988) «An evaluation of Utah's primary care case management program for Medicaid recipients», *Medical Care*, **26**:1021-1032.
- López, G., V. Ortun y C. Murillo (1997) «El sistema sanitario español: informe de una década», *Documento de trabajo de la fundación BBV*.
- Lotric, U. y A. Dobnikar (1999) «Using smoothing splines in time series prediction with neural networks», en A. Dobnikar, N.C. Steele, D.W. Pearson y R.F. Albrecht (eds.)

- Artificial Neural Nets and Genetic Algorithms*. Springer Computer Science. Springer Verlag Wien, 121-126.
- Lowe, R.A. y A.B. Bindman (1994) «Refusing care to emergency department of patients: evaluation of published triage guidelines», *Annals of Emergency Medicine* **23**:377-379.
- Lucas, R.E. (1976) «Understanding business cycles», en K. Brunner y A. Meltzer (eds.) *Stabilization of the Domestic and International Economy*, North-Holland.
- MacKoul, D., M. Feldman, J. Savageau y A. Krumholz (1995) «Emergency department utilization in a large pediatric group practice», *American Journal of Medical Quality* **10**:88-92.
- Maravall, A. y A. del Río (2001) «Time aggregation and The Hodrick-Prescott filter», Documento de trabajo del *Sevicio de Estudios del Banco de España* **0108**.
- Marsh, L. (1983) «On estimating spline regression. Proceedings of SAS», *User's Group International* **8**:723-728.
- Marsh, L. (1986) «Estimating the number and location of knots in spline regression», *Journal of Applied Business Research* **3**:60-70.
- Marsh, L. y D. Cormier (2001) *Spline regresión models*, SAGE university papers series on quantitative applications in the Social Sciences, **07-137**. Thousand Oaks, CA:Sage.
- Marsh, L., M. Maudgal y J. Raman (1990) «Alternative methods of estimating piecewise linear and higher order regression models using SAS software», *User's Group International* **15**:523-527.
- Martín, G., V.J. Cano y J.J. Cáceres (2002) «Exportación de tomate en Canarias: ¿un patrón estacional estable?», *Economía Agraria y Recursos Naturales* (en prensa).
- Martínez, A. (1989) «Exportaciones agrícolas: campañas 1988/1989», *Canarias Agraria y Pesquera* **7**:40-43.
- Maybeck, P. (1979) *Stochastic Models, Estimation and Control*. Academic Press.
- Maynard, E.J. y J.S. Dodge (1983) «Introducing a community health centre at Mosgiel, New Zealand: effects on use of hospital accident and emergency (A&E) department», *Medical Care* **21**:379-389.
- Meinhold, R. y N. Singpurwalla (1983) «Understanding the Kalman filter», *American Statistician* **37**:123-127.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

- Meinhold, R. y N. Singpurwalla (1989) «Robustification of Kalman filter models», *Journal of the American Statistical Association* **84**:479-486.
- Miró, O., M. Sánchez, B. Coll-Vinent y J. Millá (2000) «Estimación del efecto relativo que ejercen los determinantes externos e internos sobre la eficacia de un servicio de urgencias de medicina», *Medicina Clínica* **115(8)**:294-296.
- Miró, O., M. Sánchez, B. Coll-Vinent y J. Millá (2001) «Indicadores de calidad en urgencias: comportamiento en relación con la presión asistencial», *Medicina Clínica* **116(3)**:92-97.
- Miron, J.A. (1994) «The economics of seasonal cycles», en C.A. Sims (ed.) *Advances in Econometrics*. Sixth World Congress, Cambridge University Press. Vol **1**, cap. **6**.
- Miron, J.A. (1996) *The Economics of Seasonal Cycles*, MIT Press.
- Miron, J.A. y J.J. Beaulieu (1996) «What have macroeconomists learned about business cycles from the study of seasonal cycles», *The Review of Economics and Statistics* **78**:54-66.
- Moody, J. y L. Wu (1998) «High frequency foreign exchange rates: prices behavior analysis and *true prices* models», en C. Dunis y B. Zhou (eds.) *Nonlinear Modelling of High Frequency Financial Time Series*, John Wiley & Sons, 23-47.
- Murphy, A.W., G. Bury, P.K. Plunkett, D. Gibney, M. Smith, E. Mullan y otros (1996) «Randomised controlled trial of general practitioner versus usual medical care in an urban accident and emergency department: process, outcome and comparative cost», *British Medical Journal* **312**:1135-1142.
- Muth, J.F. (1960) «Optimal properties of exponentially weighted forecasts», *Journal of the American Statistical Association* **55(290)**:299-306.
- Nerlove, M. (1964) «Spectral analysis of seasonal adjustment procedures», *Econometrica* **34**:241-285.
- Nielsen, H.B. (1998) *Cubic Splines*, IMM Department of Mathematical Modelling. Technical University of Denmark.
- Noguera, P. (1996) *Estacionalidad y Transmisión Vertical de Precios en el Sector Hortofrutícola Español*. Tesis Doctoral. Universidad de Murcia.

## Bibliografía

- O'Brien, G.M., M.D. Stein, M.J. Fagan, M.J. Shapiro y A. Nasta (1999) «Enhanced emergency department referral improves primary care access», *American Journal of managed care* **5**(10):1265-1269.
- O'Brien, G.M., M.J. Shapiro, R.W. Woolard, P.S. O'Sullivan y M.D. Stein (1996) «"Inappropriate" emergency department use: a comparison of three methodologies for identification», *Academic Emergency Medicine* **3**:252-257.
- O'Grady K.F., W.G. Manning, J.P. Newhouse y R.H. Brook (1985) «The impact of cost-sharing on emergency department use», *New England Journal of Medicine* **313**:484-490.
- O'Shea, J.S., E.W. Collins y J.C. Pezzullo (1984) «An attempt to influence health care visits of frequent hospital emergency facility users», *Clinical Pediatrics* **23**:559-562.
- Olson, M. y H. Hansagi (2001) «Repeated use of the emergency department: qualitative study of the patient's perspectives», *Emergency Medicine Journal* **18**(6):430-434.
- Osborn, D.R. (1991) «The implications of periodically varying coefficients for seasonal time series processes», *Journal of Econometrics* **48**:373-384.
- Oterino, D., S. Peiró, R. Calvo, P. Sutil, O. Fernández, G. Pérez, P. Torres, M. López y T. Sempere (1999) «Utilización inadecuada de un servicio de urgencias hospitalario», *Gaceta Sanitaria* **13**(5):361-370.
- Otter, P. y S. Mittnik (1989) *Parameter identifiability in Kalman filtering models*. Technical report, University of Groningen.
- Ovens, H.J. y B.T.B. Chan (2001) «Heavy users of emergency services: a population-based review», *Canadian Medical Association Journal* **165**(8):1049-1050.
- Padgett, D.K. y B. Brodsky (1992) «Psychosocial factors influencing non-urgent use of the emergency room: a review of the literature and recommendations for research and improved service delivery», *Social Science & Medicine* **35**:1189-1197.
- Padgett, D.K., E.L. Struening, H. Andrews y J. Pittman (1995) «Predictors of emergency room use by homeless adults in New York City: the influence of predisposing, enabling and need factors», *Social science & medicine* **41**(4):547-556.
- Pagan, A. (1975) «A note on the extraction of components from time series», *Econometrica* **43**(1):163-168.
- Parreño, J.M. y M.C. Domínguez (1997) «La competencia entre los productores canarios y marroquíes en el mercado comunitario del tomate», *Agricultura y Sociedad* **83**: 117-142.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

- Pascual, P.M. (1996a) «Informe sobre la campaña de exportación 94/95 de tomate canario», *Canarias Agraria y Pesquera* **31**:43-47.
- Pascual, P.M. (1996b) *Tomate. Memoria Campaña 95/96*. SOIVRE. Tenerife.
- Pascual, P.M., R. Marrero y E.V. Hernández (1994) *Tomate. Memoria Campaña 93/94*. SOIVRE. Tenerife.
- Peiró, S., T. Sempere y D. Oterino (1999): *Efectividad de las intervenciones para reducir la utilización inapropiada de los servicios hospitalarios de urgencias. Revisando la literatura 10 años después del Informe del Defensor del Pueblo*. Informe Técnico nº 2. Economía y Salud, Boletín Informativo de la Asociación de Economía de la Salud. **12/abril/33**.
- Pierce, D.A., M.R. Gruppe y W.P. Cleveland (1984) «Seasonal adjustment of the weekly monetary aggregate: a model-based approach», *Journal of Business and Economic Statistics* **2**:260-270.
- Poirier, D.J. (1973) «Piecewise regression using cubic splines», *Journal of the American Statistical Association* **68**:515-524.
- Poirier, D.J. (1974) *The econometric of structural change*. North Holland.
- Poirier, D.J. (1975) «On the use of bilinear splines in economics», *Journal of Econometrics* **3**:23-24.
- Poirier, D.J. (1976) *The econometric of structural change with special emphasis on spline functions*, Amsterdam, North Holland Publishing Company.
- Poirier, D.J. y S.G. Garber (1974) «The determinants of aerospace profit rates 1951-1971», *Southern Economic Journal* **41**:1541-1546.
- Quilis, E.M. (2001) «Notas sobre desagregación temporal de series económicas», *Papeles de Trabajo del Instituto de Estudios Fiscales* **1/01**.
- Quintana, J.M., I. Aróstegui, A. Arcelay, M. García, M. Ortega y S. Candelas (1999) «Encuesta de satisfacción a pacientes de servicios de urgencias de hospitales agudos», *Gaceta Sanitaria*, **13(1)**:38-45.
- Ravn, M. O. y H. Uhlig (2002) «On adjusting the Hodrick-Prescott filter for the frequency of observations», *The Review of Economics and Statistics* **84(2)**:371-380.

## Bibliografía

- Redelmeier, D.A., J.P. Molin y R.J. Tibshirani (1995) «A randomized trial of compassionate care for the homeless in an emergency department», *Lancet* **345**:1131-1134.
- Reinsch, C.H. (1967) «Smoothing by spline functions», *Numerische Mathematik* **10**:177-183.
- Richards, J.R., M.L. Navarro y R.W. Derlet (2000) «Survey of directors of emergency departments in California on overcrowding», *Western Journal of Medicine* **172(6)**:385-388.
- Rizos, J., P. Anglin, I. Grava-Gubins y C. Lazar (1990) «Walk-in clinics: implications for family practice», *Canadian Medical Association Journal* **143**:740-745.
- Robison, D.E. (1964) «Estimates for the points of intersection of two polynomial regressions», *Journal of the American Statistical Association* **59**:214-224.
- Rodríguez, L y F. Castro (1994) «Aspectos Económicos de la Configuración del Sector Eléctrico en España», *Cuadernos Económicos del ICE* **57**:161-183.
- Rodríguez, R., J.M. Tabares y J.A. Medina (1989) *Cultivo Moderno del Tomate*, Ediciones Mundi Prensa.
- Rodríguez, W. (1986) *La Agricultura de Exportación en Canarias*. Consejería de Agricultura, Ganadería y Pesca. Gobierno de Canarias.
- Rosenberg, B. (1973) «Random coefficient models: the analysis of a cross section of time series by stochastically convergent parameter regression», *Annals of Economic and Social Measurement* **2**:399-428.
- Sansó, A. (1996) *Anàlisi de l'Estacionalitat no Estacionària*. Tesis Doctoral. Universidad de Barcelona.
- Sanz, I. y C. Carlos (1997) «La energía como actividad económica en España. El caso particular del sector eléctrico», *Economía Industrial* **316**:199-212.
- Schoenberg, I.J. (1946) «Contributions to the problem of approximation of equidistant data by analytic functions », *Quarterly Journal of Applied Mathematics* **4**:45-99, 112-141.
- Schoenberg, I.J. (1964) «Spline functions and the problem of graduation», *Proceedings of the Natural Academy of Sciences* **52**:947-950.
- Schweppe, C.F. (1965) «Evaluation of likelihood function for gaussian signals», *IEEE Transactions on Information Theory* **11**:61-70.

## Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia

- Scott, A. (1995) «Why is consumption so seasonal?», *Oxford University Applied Economics Discussion Paper*, **172**.
- Selby, J.V., B.H. Fireman y B.E. Swain (1996) «Effect of copayment on use of the emergency department in a health maintenance organization», *New England Journal of Medicine* **334**:635-641.
- Shapiro, M.F., J.E.J. Ware y C.D. Sherbourne (1986) «Effects of cost sharing on seeking care for serious and minor symptoms: results of a randomized control trial», *Annals of Internal Medicine* **104**:246-251.
- Sims, C.A. (1993) «Rational expectations modelling with seasonally adjusted data», *Journal of Econometrics* **55**:9-20.
- Slade, M.E. (1989) «Modelling stochastic and cyclical components of structural change: an application of the Kalman filter», *Journal of Econometrics* **41**:363-383.
- Slemrod, J., S. Yitzhaki, J. Mayshar y M. Lundholm (1994) «The optimal two-bracket linear income tax», *Journal of Public Economics* **53**:269-290.
- Smith, P.L. (1979) «Splines as a useful and convenient statistical tool», *The American Statistician* **33**:57-32.
- Steyn, I.J. (1996) *State Space Models in Econometrics: a field guide*. Vrije Universiteit.
- Stock, J.H. y M.W. Watson (1988) «Testing for common trends», *Journal of the American Statistical Association* **83**:1097-1107.
- Straus J.H., S.T. Orr y E. Charney (1983) «Referrals from an emergency room to primary care practices at an urban hospital», *American Journal of Public Health* **73**:57-61.
- Strawczynski, M. (1998) «Social insurance and the optimal piecewise linear income tax», *Journal of Public Economics* **69**:371-388.
- Suits, D.B., A. Mason y L. Chan (1978) «Spline functions fitted by standard regression methods», *Review of Economics and Statistics* **60**:132-139.
- Tiao, G. y M. Grupe (1980) «Autoregressive moving average models in time series data», *Biometrika* **67**:365-373.
- Vargas, M. (1999) «Modelización de Series Temporales Estacionarias en Espacio de Estados», *Documento de Trabajo de la Facultad de CCEE de la Universidad de Castilla la Mancha* **2/1999/4**.

## Bibliografía

- Vargas, M. (2000) «Análisis de Tendencias Comunes y Cointegración en Espacio de Estados», *Documento de Trabajo 2/2000/3 de la Facultad de CCEE de la Universidad de Castilla la Mancha 2/1999/4*.
- Vargas, M. y R. Mínguez (2000) «A State-Space Modelization of Economic Growth Among Spanish Regions», *Documento de Trabajo de la Facultad de CCEE de la Universidad de Castilla la Mancha 2/2000/1*.
- Villalba, E. (1978) *Estudio del cultivo del tomate en Tenerife y Gran Canaria*. Cámara Oficial de Comercio, Industria y Navegación de Santa Cruz de Tenerife. Instituto Tinerfeño de Expansión Económica.
- Wecker, W.E. y C.F. Ansley (1983) «The signal extraction approach to nonlinear regression and spline smoothing», *Journal of the American Statistical Association* **78**:81-89.
- Wegman, E.J. (1982) «Kalman filtering», en N. Johnson y S. Kotz (eds.) *Encyclopedia of Statistics*. John Wiley.
- Welch, G. y G. Bishop (2001) «An introduction to the Kalman Filter», [http://www.cs.unc.edu/~welch/kalman/kalman\\_filter/kalman.html](http://www.cs.unc.edu/~welch/kalman/kalman_filter/kalman.html).
- Whittaker, E.T. (1923) «On a new method of graduation», *Proceedings of the Edinburgh Mathematical Society* **41**:63-75.
- Zoroa, P. y N. Zoroa (1991) *Introducción a la probabilidad y la medida*. PPU.

## **Modelos estructurales y estacionalidad en series temporales económicas de alta frecuencia**