



UNIVERSITAT DE VALÈNCIA
Facultat de Matemàtiques
Departament d'Estadística i Investigació Operativa



**UN ANÁLISIS BAYESIANO DE MODELOS
MULTIVARIANTES DE SUAVIZADO EXPONENCIAL**

Tesis doctoral
Realizada por: Ana Corberán Vallet
Dirigida por: José D. Bermúdez Edo y
Enriqueta Vercher González

UMI Number: U603105

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U603105

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

D. José Domingo Bermúdez Edo y Dña. M^a Enriqueta Vercher González,
profesores titulares del Departamento de Estadística e Investigación
Operativa de la Universidad de Valencia

CERTIFICAN que la presente memoria de investigación

**”Un Análisis Bayesiano de Modelos Multivariantes
de Suavizado Exponencial”**

ha sido realizada bajo su dirección por Ana Corberán Vallet, y consti-
tuye su tesis para optar al grado de Doctor en Ciencias Matemáticas.

Y para que así conste, en cumplimiento con la normativa vigente,
autorizan su presentación ante la Facultad de Matemáticas de la Uni-
versidad de Valencia para que pueda ser tramitada su lectura y defensa
pública.

Valencia, 24 de febrero de 2009.



José D. Bermúdez Edo.



M^a Enriqueta Vercher González.

A Javi

Agradecimientos

La realización de esta tesis ha sido posible gracias al apoyo de muchas personas a quienes estaré siempre agradecida.

A José Bermúdez y Enriqueta Vercher, mis directores de tesis, porque sin su tiempo, dedicación y motivación esta tesis nunca habría existido, pero sobre todo por su amistad y porque gracias a ellos he disfrutado mucho haciendo este trabajo.

A M^a Jesús Bayarri, Francisco Javier Girón y José Vicente Segura, por su trabajo de revisión y apreciados comentarios.

Quiero agradecer especialmente a M^a Jesús Bayarri toda su ayuda en relación a mi estancia en Estados Unidos, ya que gracias a ella he tenido una experiencia muy enriquecedora, tanto en lo profesional como en lo personal.

Al profesor Jim Berger, por facilitar mi estancia en SAMSI durante 3 meses, por su amabilidad y sus valiosas sugerencias.

A toda la gente que allí conocí, en especial a Edsel Peña, Jaya Bishwal y Megan Owen, porque sin ellos SAMSI no hubiera sido lo mismo.

A todo el departamento de Estadística e Investigación Operativa, por su acogida y los buenos momentos pasados.

A toda mi familia, y muy especialmente a mis padres, Paco y Amparo, a mis hermanos, Carlos y Marta, y a mi marido Javi, por su ayuda y ánimos en los momentos necesitados, por estar siempre a mi lado, y porque sin su amor y confianza nunca lo hubiera conseguido. Gracias!

Prólogo

El objetivo principal de esta tesis es la obtención de predicciones precisas, tanto puntuales como intervalos de predicción, de series temporales con errores correlacionados para cada instante temporal mediante el modelo de Holt-Winters multivariante. Con ello hemos pretendido realizar una aportación al análisis multivariante de series temporales mediante la metodología de suavizado exponencial.

Los modelos de suavizado exponencial son reconocidos como una de las herramientas más empleadas en la predicción a corto plazo de series temporales univariantes (Gardner, 1985; 2006). Su popularidad se debe principalmente a su sencilla formulación y a los buenos resultados obtenidos en competiciones de predicción (Makridakis and Hibon, 2000).

En un principio, los métodos de suavizado exponencial, con origen en el trabajo de Brown y Holt realizado en la década de los años cincuenta, no estaban basados explícitamente en ningún modelo probabilístico, sino en un análisis previo de los datos que conformaban las series temporales para determinar, por ejemplo, la existencia de tendencia o comportamientos cíclicos. A partir de sencillas ecuaciones de actualización, de las que se derivan posibles cambios en el nivel local de la serie, tendencia o estacionalidad, los métodos obtienen predicciones para los valores futuros de las series. La elección adecuada de los valores de los parámetros (parámetros de suavizado y condiciones iniciales del nivel, tendencia y estacionalidad) es crucial en la obtención de predicciones; sin embargo, el mejor procedimiento para su obtención es una cuestión que todavía está por determinar. Existe una gran variedad de procedimientos heurísticos para estimar el valor de las condiciones iniciales, sin evidencia empírica a favor de ninguno de ellos, obteniendo como resultado predicciones que pueden ser substancialmente diferentes (Chatfield and Yar, 1988; Segura and Vercher, 2001). Una vez

II

fijado el valor de las condiciones iniciales, los parámetros de suavizado son obtenidos, generalmente, minimizando los errores de predicción de un paso de los datos históricos de la serie.

Posteriormente, se introduce una clase de modelos de espacio de estado para los cuales los métodos de suavizado exponencial son óptimos (Ord et al., 1997; Hyndman et al., 2002). Dicha equivalencia posibilita el estudio de las propiedades estadísticas de los modelos de suavizado exponencial, el cálculo de los estimadores máximo-verosímiles de los parámetros y la obtención de predicciones tanto puntuales como intervalos de predicción. Las predicciones puntuales basadas en la estimación máximo-verosímil (Hyndman et al., 2002; Bermúdez et al., 2007) u otros métodos de optimización (Bermúdez et al., 2006a;b) han demostrado ser muy precisas. En cambio, los intervalos de predicción tienden a ser demasiado estrechos, pues las diferentes ecuaciones propuestas para su cálculo no tienen en cuenta el error de estimación de los parámetros del modelo.

Recientemente, se ha analizado el modelo de Holt-Winters univariante aditivo desde el enfoque Bayesiano (Bermúdez et al., 2008b). Ésto les ha permitido a los autores acomodar la incertidumbre propia del modelo y obtener así intervalos de predicción con cobertura empírica próxima a la nominal.

El enfoque Bayesiano para la inferencia, al igual que para la predicción, supone condicionar a lo conocido (datos observados, estructuras, etc.) para obtener inferencias acerca de lo desconocido, dando lugar a problemas de integración que en la mayoría de los casos no tienen solución analítica. Los avances en integración numérica y métodos de simulación han incrementado el número de problemas de predicción que pueden ser tratados desde el punto de vista Bayesiano (Geweke and Whiteman, 2006).

Por otro lado, en la práctica es común encontrar conjuntos de series temporales sujetas a componentes aleatorias similares o donde las observaciones de una serie dependen, no sólo de los valores pasados de la serie, sino también de los valores pasados y presentes de otras series. Algunos ejemplos son las series de ventas de diferentes productos de una misma compañía, series de demanda eléctrica en diferentes regiones de un país, series de tipo de cambio, etc. En estas ocasiones, con el uso de modelos de series temporales multivariantes, los cuales incorporan la relación existente entre las series, es posible mejorar tanto el ajuste como la predicción respecto a los análisis univariantes. Consecuentemente, varios autores (Jones, 1966; Enns et al., 1982;

Harvey, 1986; Pfeffermann and Allon, 1989; Fernández and Harvey, 1990; de Silva et al., 2007) han tratado de extender los modelos de suavizado exponencial al caso multivariante. Una vez estimados los parámetros del modelo, generalmente mediante estimación máximo-verosímil, predicciones puntuales para los valores futuros de las series pueden ser calculadas a partir del filtro de Kalman. No obstante, en el análisis conjunto de series temporales correlacionadas mediante modelos de suavizado exponencial multivariante quedan todavía cuestiones por determinar. Prueba de ello es que, en la literatura existente, no encontramos referencias para el cálculo de intervalos de predicción.

En esta memoria, introducimos una nueva formulación para el modelo de Holt-Winters multivariante con estacionalidad aditiva y errores aditivos, el cual puede ser expresado como un modelo de regresión aparentemente no relacionado (modelo SUR, Zellner, 1962). El modelo es analizado desde la perspectiva Bayesiana. Dados los datos observados que conforman las series temporales, la distribución a posteriori para los parámetros del modelo es obtenida. Dicha distribución, aunque no es analíticamente tratable, puede ser estimada mediante métodos de simulación MCMC. Finalmente, la distribución predictiva es estimada mediante integración por Monte Carlo.

Todo lo mencionado hasta aquí lo exponemos detalladamente en los distintos capítulos de la tesis, seis en total, cerrando el trabajo con la indicación de la bibliografía utilizada. En el primer capítulo introducimos los modelos de suavizado exponencial univariante y las generalizaciones propuestas en la literatura para el caso multivariante. Posteriormente, revisamos brevemente las principales herramientas Bayesianas propuestas para tratar el problema de selección de modelos. En el segundo capítulo presentamos la formulación para el modelo de Holt-Winters multivariante empleada en esta memoria y mostramos su análisis Bayesiano. El tercer capítulo está dedicado al estudio de dos casos particulares del modelo de Holt-Winters multivariante: el modelo de Holt multivariante, adecuado para el análisis de series temporales con tendencia pero sin componente estacional, y el modelo de Holt-Winters multivariante con parámetros de suavizado comunes para los modelos univariantes. En el cuarto capítulo aplicamos los diferentes procedimientos de selección de modelos para seleccionar el modelo de Holt-Winters multivariante que más se adecua a los datos que conforman las series temporales de estudio. En el quinto capítulo mostramos los resultados obtenidos en la predicción de dos conjuntos reales de series temporales correlacionadas: las series diarias del valor de cierre del Dow Jones y del Ibex 35 y las series

IV

mensuales del número de pasajeros en cinco aeropuertos de Londres. Finalizando, el sexto capítulo está dedicado a conclusiones y futuras líneas de investigación.

Índice general

1. Introducción	1
1.1. Modelos de suavizado exponencial	2
1.1.1. Análisis Bayesiano del modelo de Holt-Winters aditivo univariante	8
1.2. Series temporales de ocupación hotelera	11
1.3. Modelos de suavizado exponencial multivariante	14
1.4. Principales técnicas de selección de modelos	18
1.4.1. Factores Bayes	18
1.4.2. Selección de modelos como un problema de decisión .	25
1.4.3. Criterios de selección de modelos	26
2. Holt-Winters multivariante	29
2.1. Formulación del modelo	31
2.2. Análisis Bayesiano	34
2.2.1. Análisis Bayesiano partiendo de la priori de Jeffreys .	35
2.2.2. Algoritmo de Metropolis-within-Gibbs para simular de la distribución a posteriori	39
2.2.3. Análisis Bayesiano de referencia	41
2.2.4. Ejemplo de ocupación hotelera (continuación)	46
2.3. Distribución predictiva	49
2.3.1. Ejemplo de ocupación hotelera (continuación)	57
2.4. Notas computacionales	58
3. Holt-Winters multivariante: casos particulares	61
3.1. Holt multivariante	62
3.1.1. Predicción de un paso cuando los valores futuros son observados secuencialmente	65
3.2. Holt-Winters multivariante con parámetros de suavizado iguales	68
3.2.1. Análisis Bayesiano	72

3.2.2. Método de aceptación-rechazo para simular de $f(\theta y)$	75
3.2.3. Distribución predictiva	76
3.2.4. Ejemplo de ocupación hotelera (continuación)	79
3.2.5. Análisis Bayesiano de referencia	81
3.3. Notas computacionales	86
4. Selección de modelos	89
4.1. Factores Bayes para la selección del modelo de Holt-Winters multivariante	91
4.2. Propiedades predictivas para la selección del modelo de Holt-Winters multivariante	98
4.3. Criterios de selección para el modelo de Holt-Winters multivariante	100
4.4. Ejemplo de ocupación hotelera(resumen)	102
5. Resultados numéricos	107
5.1. Índices bursátiles Dow Jones e Ibex 35	108
5.2. Pasajeros de líneas aéreas	114
6. Conclusiones y futuras líneas de investigación	123
6.1. Conclusiones	123
6.2. Futuras líneas de investigación	125
Referencias	129

Índice de figuras

1.1.	Gráfico temporal de las series mensuales de ocupación hotelera en Castellón, Valencia y Alicante desde Enero de 2001 hasta Diciembre de 2006. Unidades: miles de viajeros	12
1.2.	Histogramas de los parámetros de suavizado simulados de su distribución a posteriori en los análisis univariantes de las series temporales de ocupación hotelera	12
1.3.	Línea negra: ajuste y predicciones obtenidos para cada una de las series temporales de ocupación hotelera con el modelo de Holt-Winters univariante. Líneas discontinuas: intervalos de predicción del 95 %	13
3.1.	Histogramas de los parámetros de suavizado simulados de su distribución a posteriori cuando las series de ocupación hotelera son analizadas mediante el modelo de Holt-Winters multivariante con parámetros de suavizado iguales. Líneas rojas verticales: medias muestrales	80
3.2.	Densidades estimadas de los coeficientes de correlación de las series temporales de ocupación hotelera. Las líneas verticales representan las medias y los intervalos de probabilidad a posteriori del 95 %	80
5.1.	Gráfico temporal de las series del valor de cierre del Dow Jones (desde el 1 de Marzo hasta el 12 de Julio de 2007) y del Ibex 35 (desde el 2 de Marzo hasta el 13 de Julio de 2007)	109
5.2.	Histogramas de los parámetros de suavizado simulados de su distribución a posteriori cuando las series del Dow Jones y del Ibex 35 son analizadas mediante el modelo de Holt multivariante con parámetros de suavizado iguales. Líneas rojas verticales: medias muestrales	111

- 5.3. Densidad estimada del coeficiente de correlación del modelo de Holt multivariante con parámetros de suavizado iguales resultante del análisis conjunto de las series del Dow Jones y del Ibex 35 111
- 5.4. Gráficos temporales univariantes y multivariante de las series mensuales del número de pasajeros, en millones, de cinco aeropuertos de Londres 115
- 5.5. Ajuste y predicciones puntuales obtenidas, con el modelo de Holt-Winters multivariante con parámetros de suavizado iguales, para las series temporales del número de pasajeros en cinco aeropuertos de Londres. Líneas discontinuas: intervalos de predicción del 95% 118

Índice de tablas

1.1. Principales métodos de suavizado exponencial	3
1.2. Error SMAPE de ajuste y de predicción resultantes de los análisis univariantes de las series temporales de ocupación hotelera	14
2.1. Estimaciones de la media y del intervalo de probabilidad a posteriori del 95 % para las condiciones iniciales del modelo de Holt-Winters multivariante	48
2.2. Estimaciones de la media y del intervalo de probabilidad a posteriori del 95 % para los parámetros de suavizado del modelo de Holt-Winters multivariante	49
2.3. Estimaciones de la media y del intervalo de probabilidad a posteriori del 95 % para los distintos elementos de la matriz de covarianzas del modelo de Holt-Winters multivariante	49
2.4. Estimaciones de la media y del intervalo de probabilidad a posteriori del 95 % para los coeficientes de correlación del modelo de Holt-Winters multivariante	49
2.5. Error SMAPE de ajuste y de predicción para cada una de las series de ocupación hotelera resultantes del análisis Bayesiano multivariante	58
3.1. Error SMAPE de ajuste y de predicción cuando las series temporales de ocupación hotelera son analizadas mediante el modelo de Holt-Winters multivariante con parámetros de suavizado iguales	81
4.1. Criterios de selección para el modelo de Holt-Winters multivariante en el análisis conjunto de las series de ocupación hotelera	105

4.2.	Errores de ajuste obtenidos, para cada una de las series temporales de ocupación hotelera, en los análisis univariantes y multivariantes con el modelo de Holt-Winters general y el modelo con parámetros de suavizado iguales	106
4.3.	Errores de predicción obtenidos, para cada una de las series temporales de ocupación hotelera, en los análisis univariantes y multivariantes con el modelo de Holt-Winters general y el modelo con parámetros de suavizado iguales	106
5.1.	Estimaciones de las medias y de los intervalos de probabilidad a posteriori del 95 % para los parámetros del modelo de Holt multivariante resultante del análisis conjunto de las series del Dow Jones y del Ibex 35	110
5.2.	Criterios de selección para el modelo de Holt multivariante en el análisis conjunto de las series temporales del Dow Jones y del Ibex 35	112
5.3.	Predicciones de un paso para la serie del Dow Jones cuando es analizada mediante el modelo \mathcal{M}_1	113
5.4.	Predicciones de un paso para la serie del Ibex 35 cuando es analizada mediante el modelo \mathcal{M}_1	113
5.5.	Predicciones de un paso para la serie del Ibex 35 resultantes de modificar las predicciones originales en función del error de predicción de un paso asociado al Dow Jones	113
5.6.	Estimaciones de la media y del intervalo de probabilidad a posteriori del 95 % para los coeficientes de correlación del modelo de Holt-Winters multivariante en el análisis conjunto de las series temporales del número de pasajeros de cinco aeropuertos de Londres	116
5.7.	Criterios de selección para el modelo de Holt-Winters multivariante en el análisis conjunto de las series temporales del número de pasajeros en cinco aeropuertos de Londres	117
5.8.	Errores de ajuste y de predicción resultantes de los análisis multivariantes y univariantes de las series temporales del número de pasajeros en cinco aeropuertos de Londres	119
5.9.	Criterios de selección para el modelo de Holt-Winters multivariante en el análisis conjunto de las series temporales del número de pasajeros en Gatwick, Heathrow, Luton y Stansted	120

5.10. Errores de ajuste y de predicción resultantes del análisis conjunto de las series temporales del número de pasajeros de Gatwick, Heathrow, Luton y Stansted mediante el modelo de Holt-Winters multivariante y del análisis univariante de la serie de London City 120

Capítulo 1

Introducción

En la práctica es común encontrar conjuntos de series temporales sujetas a componentes aleatorias similares o donde las observaciones de una serie dependen, no sólo de los valores pasados de la serie, sino también de los valores pasados y presentes de otras series. En estas ocasiones, el análisis conjunto de las series mediante modelos multivariantes, que incorporan la relación existente entre éstas, permite obtener predicciones más precisas que las resultantes de los análisis univariantes.

La sencilla formulación de los modelos de suavizado exponencial y los buenos resultados obtenidos, con un mínimo esfuerzo, en competiciones de predicción (Makridakis and Hibon, 2000) hacen de estos modelos una de las herramientas más empleadas en la predicción a corto plazo de series temporales univariantes (Gardner, 2006; Bermúdez et al., 2008a). Es por ello que, en este trabajo, desarrollamos un procedimiento Bayesiano de predicción basado en el modelo de Holt-Winters aditivo multivariante que nos permite obtener predicciones precisas, tanto puntuales como intervalos de predicción, de series temporales con errores correlacionados para cada instante temporal.

En la siguiente sección de este primer capítulo introducimos los modelos de suavizado exponencial univariante, mostrando el análisis Bayesiano del modelo de Holt-Winters aditivo. En la segunda sección mostramos los resultados obtenidos en los análisis univariantes de las series de ocupación hotelera en tres provincias de España, ejemplo de referencia a lo largo de la memoria. En la sección tercera presentamos las generalizaciones existentes en la literatura para el caso multivariante. En la última sección revisamos las principales herramientas Bayesianas propuestas para tratar el problema de selección de modelos, parte fundamental de cualquier análisis estadístico.

1.1. Modelos de suavizado exponencial

A lo largo de los años, el ‘método de Holt’ se ha convertido en un término estándar de referencia para el método de predicción basado en el uso de medias móviles ponderadas exponencialmente para obtener predicciones de series temporales con cierta tendencia. Según Ord (2004), el término ha llegado a ser tan conocido que a menudo es utilizado sin referencias. Cuando se cita a Holt (1957), the Office of Naval Research report [ONR Research Memorandum Number 52], hay una sospecha de que el autor esté siguiendo la costumbre y no haya consultado el informe original, puesto que no fue publicado por Holt hasta el año 2004 (Holt, 2004). No obstante, el método ganó popularidad en el año 1960 al ser estudiado por Winters, pasando a ser conocido como el método de Holt-Winters (Winters, 1960).

La tabla 1.1 contiene las ecuaciones para los principales métodos de suavizado exponencial, todos ellos extensiones del trabajo original de Holt y Winters (1960). En la notación empleada, y_t es el valor observado de la serie temporal en el instante t ; a_t , b_t y c_t representan el nivel, la tendencia y la componente estacional respectivamente en el instante t ; s es la longitud del ciclo estacional e \hat{y}_{t+h} la predicción de h pasos calculada en t . Los métodos son clasificados empleando la nomenclatura introducida por Hyndman et al. (2002), de manera que el método N-N denota el método de suavizado exponencial simple, sin tendencia ni componente estacional, A-N es el método de Holt, con tendencia aditiva pero sin componente estacional, A-A es el método de Holt-Winters aditivo, con tendencia y estacionalidad aditivas, A-M representa al método de Holt-Winters multiplicativo, es decir, tendencia aditiva y estacionalidad multiplicativa, etc.

Para emplear la metodología de suavizado exponencial, el analista debe proporcionar los valores del nivel, tendencia y estacionalidad al comienzo de la serie ($a_0, b_0, c_{1-s}, c_{2-s}, \dots, c_0$) así como el valor de los parámetros de suavizado (α, β, γ). Para la estimación de las condiciones iniciales encontramos una gran variedad de procedimientos heurísticos, sin evidencia empírica a favor de ninguno de ellos, obteniendo como resultado predicciones que pueden ser substancialmente diferentes (Chatfield and Yar, 1988; Segura and Vercher, 2001). Además, estos procedimientos utilizan parte de los datos que conforman la serie temporal, al menos los del primer ciclo estacional, para estimar las condiciones iniciales, lo que supone un problema si se dispone de un número limitado de observaciones. Una vez fijado su valor, los parámetros de suavizado suelen obtenerse minimizando los errores de predicción de un paso de los datos históricos de la serie.

Tendencia	Estacionalidad		
	N	A	M
N	$a_t = \alpha y_t + (1 - \alpha)a_{t-1}$ $\hat{y}_{t+h} = a_t$	$a_t = \alpha(y_t - c_{t-s}) + (1 - \alpha)a_{t-1}$ $c_t = \gamma(y_t - a_t) + (1 - \gamma)c_{t-s}$ $\hat{y}_{t+h} = a_t + c_{t+h-s}$	$a_t = \alpha(y_t/c_{t-s}) + (1 - \alpha)a_{t-1}$ $c_t = \gamma(y_t/a_t) + (1 - \gamma)c_{t-s}$ $\hat{y}_{t+h} = a_t c_{t+h-s}$
A	$a_t = \alpha y_t + (1 - \alpha)(a_{t-1} + b_{t-1})$ $b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$ $\hat{y}_{t+h} = a_t + hb_t$	$a_t = \alpha(y_t - c_{t-s}) + (1 - \alpha)(a_{t-1} + b_{t-1})$ $b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$ $c_t = \gamma(y_t - a_t) + (1 - \gamma)c_{t-s}$ $\hat{y}_{t+h} = a_t + hb_t + c_{t+h-s}$	$a_t = \alpha(y_t/c_{t-s}) + (1 - \alpha)(a_{t-1} + b_{t-1})$ $b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$ $c_t = \gamma(y_t/a_t) + (1 - \gamma)c_{t-s}$ $\hat{y}_{t+h} = (a_t + hb_t)c_{t+h-s}$
M	$a_t = \alpha y_t + (1 - \alpha)(a_{t-1}b_{t-1})$ $b_t = \beta(a_t/a_{t-1}) + (1 - \beta)b_{t-1}$ $\hat{y}_{t+h} = a_t b_t^h$	$a_t = \alpha(y_t - c_{t-s}) + (1 - \alpha)(a_{t-1}b_{t-1})$ $b_t = \beta(a_t/a_{t-1}) + (1 - \beta)b_{t-1}$ $c_t = \gamma(y_t - a_t) + (1 - \gamma)c_{t-s}$ $\hat{y}_{t+h} = a_t b_t^h + c_{t+h-s}$	$a_t = \alpha(y_t/c_{t-s}) + (1 - \alpha)(a_{t-1}b_{t-1})$ $b_t = \beta(a_t/a_{t-1}) + (1 - \beta)b_{t-1}$ $c_t = \gamma(y_t/a_t) + (1 - \gamma)c_{t-s}$ $\hat{y}_{t+h} = (a_t b_t^h)c_{t+h-s}$

Tabla 1.1: Principales métodos de suavizado exponencial

Otro enfoque, el cual simplifica el análisis y produce resultados más satisfactorios que los anteriores, consiste en considerar las condiciones iniciales como parámetros del modelo y determinar su valor conjuntamente con el de los parámetros de suavizado mediante algún esquema de optimización adecuado (Bermúdez et al., 2006a;b).

La introducción de una clase de modelos de espacio de estado no lineales con una única fuente de error subyacentes a los métodos de suavizado exponencial supuso un gran avance para la metodología de suavizado exponencial (Ord et al., 1997; Hyndman et al., 2002). De este modo, los métodos pueden disfrutar de las ventajas propias de los procedimientos de predicción basados en modelos estadísticos como, por ejemplo, el cálculo de la función de verosimilitud y de intervalos de predicción para los valores futuros de las series. Para cada método existen dos posibles modelos de espacio de estado subyacentes, según si se suponen errores aditivos o multiplicativos. Las predicciones puntuales obtenidas con ambos modelos son equivalentes, aunque no ocurre lo mismo con los intervalos de predicción.

Los modelos de espacio de estado no lineales con una única fuente de error, definidos como:

$$\begin{aligned} y_t &= h(x_{t-1}, \varphi) + k(x_{t-1}, \varphi)\epsilon_t \\ x_t &= f(x_{t-1}, \varphi) + g(x_{t-1}, \varphi)\epsilon_t \end{aligned}$$

son un caso particular de los tradicionales modelos de espacio de estado, con una única componente aleatoria común para las ecuaciones de observación y de transición. Esta particularidad permite la estimación máximo-verosímil del modelo, no siendo necesario recurrir al filtro de Kalman extendido como ocurre en el caso general. En las ecuaciones anteriores, y_t es la observación en t ; x_{t-1} es el vector que representa el estado del proceso subyacente al comienzo del instante t ; φ es un vector de parámetros y $\{\epsilon_t\}_{t=1}^n$ son los errores, independientes entre sí e idénticamente distribuidos con media 0 y varianza σ^2 .

La versión lineal de estos modelos viene dada por:

$$\begin{aligned} y_t &= h'x_{t-1} + k_t(x_{t-1})\epsilon_t \\ x_t &= Fx_{t-1} + \delta k_t(x_{t-1})\epsilon_t \end{aligned}$$

siendo $k_t(x_{t-1}) = 1$ para el caso homocedástico.

Entonces, los modelos de suavizado exponencial con errores aditivos son obtenidos al suponer que la observación de la serie temporal en el instante t , y_t , viene de la variable aleatoria Y_t , suma del nivel local de la serie, posibles componentes de tendencia y estacionalidad y de la componente aleatoria ϵ_t , distribuida según la Normal $N(0, \sigma^2)$. Cada vez que se dispone de una nueva observación, los términos de nivel, tendencia y estacionalidad son actualizados conforme a las correspondientes ecuaciones de actualización mostradas en la tabla 1.1. Así por ejemplo, el modelo de Holt-Winters aditivo es obtenido al suponer que la observación en el instante t viene de la variable aleatoria Y_t definida como:

$$Y_t = a_{t-1} + b_{t-1} + c_{t-s} + \epsilon_t \quad (1.1)$$

donde a_t , b_t y c_t son el nivel, la tendencia y la componente estacional de la serie en el instante t y $\{\epsilon_t\}_{t=1}^n$ son las componentes aleatorias del modelo, supuestas independientes e idénticamente distribuidas $N(0, \sigma^2)$.

Definiendo $x_t = (a_t, b_t, c_t, c_{t-1}, \dots, c_{t-s+2}, c_{t-s+1})'$ el vector $(s+2) \times 1$ de estado y $\delta = (\alpha, \alpha\beta, \gamma, 0, \dots, 0, 0)'$ el vector de parámetros, el modelo de Holt-Winters aditivo con errores aditivos puede expresarse como un modelo de espacio de estado lineal homocedástico con una única fuente de error de la siguiente manera:

$$y_t = (1, 1, 0, 0, \dots, 0, 1)x_{t-1} + \epsilon_t$$

$$\begin{pmatrix} a_t \\ b_t \\ c_t \\ c_{t-1} \\ \vdots \\ c_{t-s+2} \\ c_{t-s+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} a_{t-1} \\ b_{t-1} \\ c_{t-1} \\ c_{t-2} \\ \vdots \\ c_{t-s+1} \\ c_{t-s} \end{pmatrix} + \begin{pmatrix} \alpha \\ \alpha\beta \\ \gamma \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \epsilon_t \quad (1.2)$$

donde las ecuaciones de actualización son expresadas en la forma equivalente de corrección de error:

$$a_t = \alpha(y_t - c_{t-s}) + (1 - \alpha)(a_{t-1} + b_{t-1}) = a_{t-1} + b_{t-1} + \alpha\epsilon_t \quad (1.3)$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} = b_{t-1} + \alpha\beta\epsilon_t \quad (1.4)$$

$$c_t = \gamma(y_t - a_{t-1} - b_{t-1}) + (1 - \gamma)c_{t-s} = c_{t-s} + \gamma\epsilon_t \quad (1.5)$$

Notar que la ecuación de actualización de la componente estacional no es la original de Winters (1960), sino la propuesta por Ord et al. (1997). Para más referencias consultar Gardner (2006).

La versión del modelo con errores multiplicativos sería equivalente a la formulación anterior sustituyendo ϵ_t por $\epsilon_t = (a_{t-1} + b_{t-1} + c_{t-s})\epsilon_t$. El resto de modelos de suavizado exponencial, tanto con errores aditivos como con errores multiplicativos, se definen de manera análoga.

Posteriormente, Bermúdez et al. (2007) formulan el modelo de Holt-Winters aditivo con errores aditivos como un modelo lineal heterocedástico, lo que simplifica el cálculo de los estimadores máximo-verosímiles de los parámetros del modelo, parámetros de suavizado, condiciones iniciales y varianza de los errores, así como la obtención de intervalos de predicción.

Utilizando recursivamente la ecuación de observación (1.1) junto con las ecuaciones de actualización (1.3), (1.4) y (1.5), los datos pueden ser expresados en función de las condiciones iniciales y de los parámetros de suavizado de la siguiente manera:

$$\begin{aligned}
y_1 &= a_0 + b_0 + c_{1-s} + \epsilon_1 \\
y_2 &= a_0 + 2b_0 + c_{2-s} + \alpha(1 + \beta)\epsilon_1 + \epsilon_2 \\
y_3 &= a_0 + 3b_0 + c_{3-s} + \alpha(1 + 2\beta)\epsilon_1 + \alpha(1 + \beta)\epsilon_2 + \epsilon_3 \\
&\vdots \\
y_s &= a_0 + sb_0 + c_0 + \alpha \sum_{r=1}^{s-1} (1 + (s-r)\beta)\epsilon_r + \epsilon_s \\
y_{s+1} &= a_0 + (s+1)b_0 + c_{1-s} + \gamma\epsilon_1 + \alpha \sum_{r=1}^s (1 + (s+1-r)\beta)\epsilon_r + \epsilon_{s+1} \\
y_{s+2} &= a_0 + (s+2)b_0 + c_{2-s} + \gamma\epsilon_2 + \alpha \sum_{r=1}^{s+1} (1 + (s+2-r)\beta)\epsilon_r + \epsilon_{s+2} \\
&\vdots
\end{aligned} \tag{1.6}$$

Entonces, el vector de datos $y = (y_1, y_2, y_3, \dots, y_n)'$ puede ser expresado matricialmente como:

$$y = A\omega + L\epsilon \tag{1.7}$$

donde $\omega = (a_0, b_0, c_{1-s}, c_{2-s}, \dots, c_0)'$ es el vector $(s+2) \times 1$ de las condiciones iniciales; $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)'$ es el vector de errores, distribuido según

la Normal multivariante $N_n(0, \sigma^2 I_n)$; A es la matriz de orden $n \times (s + 2)$ cuya primera columna es el vector $(1, 1, \dots, 1)'$, su segunda columna el vector $(1, 2, \dots, n)'$ y las últimas s columnas están compuestas por matrices identidad de orden s una a continuación de otra hasta completar las n filas; L es la matriz $n \times n$ triangular inferior definida como:

$$L = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ l_2 & 1 & 0 & \dots & 0 & 0 \\ l_3 & l_2 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ l_{n-1} & l_{n-2} & l_{n-3} & \dots & 1 & 0 \\ l_n & l_{n-1} & l_{n-2} & \dots & l_2 & 1 \end{pmatrix} \quad (1.8)$$

con $l_i = \alpha(1 + (i - 1)\beta) + \gamma(i \equiv 1(\text{mod } s))$. Entonces, la distribución conjunta del vector de datos y es la Normal multivariante con vector de medias $E(y) = A\omega$ y matriz de covarianzas $V(y) = \sigma^2 LL'$, que depende del vector $\theta = (\alpha, \beta, \gamma)'$ de los parámetros de suavizado y es siempre definida positiva, pues $|L| = 1$ para todo valor de θ .

La matriz de diseño A es una matriz constante y conocida, pues no depende ni de las condiciones iniciales ni de los parámetros de suavizado, de rango no completo; de hecho, $rg(A) = s + 1$, pues la primera columna es la suma de las últimas s columnas. Consecuentemente, las condiciones iniciales no son funciones estimables y para que el modelo sea identificable es necesario imponer una restricción lineal sobre ω . Bermúdez et al. (2007) utilizan, sin pérdida de generalidad, la restricción $a_0 + b_0 = 0$, en lugar de la generalmente asumida $c_{1-s} + c_{2-s} + \dots + c_0 = 0$. La expresión matricial del vector de datos resultante de considerar esta restricción es de la forma:

$$y = M\psi + L\varepsilon \quad (1.9)$$

donde $\psi = (b_0, c_{1-s}, c_{2-s}, \dots, c_0)'$ es el nuevo vector $(s + 1) \times 1$ identificable de las condiciones iniciales; M es la matriz $n \times (s + 1)$ cuya primera columna es el vector $(0, 1, 2, \dots, n - 1)'$ y las últimas s columnas están compuestas por matrices identidad de orden s una a continuación de otra hasta completar las n filas; L y ε están definidos como en (1.7).

La función de verosimilitud asociada al modelo de Holt-Winters aditivo univariante, expresado como un modelo lineal heterocedástico, es de la forma:

$$f(y|\psi, \theta, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (y - M\psi)' (LL')^{-1} (y - M\psi) \right\} \quad (1.10)$$

Para un análisis detallado del modelo basado en la estimación máximo-verosímil consultar Bermúdez et al. (2007). En la siguiente sección se muestra su análisis Bayesiano.

1.1.1. Análisis Bayesiano del modelo de Holt-Winters aditivo univariante

La finalidad principal de los modelos de suavizado exponencial es la obtención de predicciones para los valores futuros de las series temporales de estudio. A diferencia de las predicciones puntuales basadas en la estimación máximo-verosímil (Hyndman et al., 2002; Bermúdez et al., 2007) u otros métodos de optimización (Bermúdez et al., 2006a;b), que han demostrado ser muy precisas, los intervalos de predicción tienden a ser demasiado estrechos, pues en su cálculo no se tiene en cuenta el error de estimación de los parámetros del modelo.

Bermúdez et al. (2008b) analizan el modelo de Holt-Winters univariante aditivo desde la perspectiva Bayesiana, incorporando así el error de estimación en el cálculo de predicciones y obteniendo intervalos de predicción con cobertura empírica próxima a la nominal.

Sea $X = L^{-1}M$, $P_X = X(X'X)^{-1}X'$ la matriz de proyección ortogonal sobre el espacio generado por las columnas de X y $\tilde{\psi} = (X'X)^{-1}X'L^{-1}y$ el estimador mínimos cuadrados de ψ cuando el vector θ de parámetros de suavizado y, por tanto, las matrices L y X son conocidos. Entonces, la forma cuadrática en (1.10) puede descomponerse como:

$$\begin{aligned} (y - M\psi)'(LL')^{-1}(y - M\psi) &= \\ &= (L^{-1}y - X\tilde{\psi} + X(\tilde{\psi} - \psi))'(L^{-1}y - X\tilde{\psi} + X(\tilde{\psi} - \psi)) \\ &= (\psi - \tilde{\psi})'X'X(\psi - \tilde{\psi}) + (L^{-1}y)'(I_n - P_X)L^{-1}y \end{aligned}$$

y la función de verosimilitud es proporcional a

$$(\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\psi - \tilde{\psi})'X'X(\psi - \tilde{\psi}) \right\} \exp \left\{ -\frac{1}{2\sigma^2} (L^{-1}y)'(I_n - P_X)L^{-1}y \right\} \quad (1.11)$$

Si el vector $\theta = (\alpha, \beta, \gamma)'$ de los parámetros de suavizado fuera conocido, la matriz L sería conocida y el análisis Bayesiano del modelo (1.9) es conocido (ver, por ejemplo, O'Hagan and Foster, 2004, capítulo 11). A partir de la distribución a priori no informativa convencional $f(\psi, \sigma|\theta) \propto \sigma^{-1}$, la



distribución a posteriori para los parámetros del modelo es de la forma:

$$f(\psi|y, \sigma, \theta) = N_{s+1}(\tilde{\psi}, \sigma^2(X'X)^{-1}) \quad (1.12)$$

$$f(\sigma^2|y, \theta) = Ga^{-1}\left(\frac{n-s-1}{2}, \frac{1}{2}(L^{-1}y)'(I_n - P_X)L^{-1}y\right) \quad (1.13)$$

En la práctica, el vector θ de los parámetros de suavizado es siempre desconocido. Tomando $f(\psi, \sigma, \theta) = f(\psi, \sigma|\theta)f(\theta)$ donde $f(\psi, \sigma|\theta) \propto \sigma^{-1}$ y $f(\theta) \propto 1$, la distribución a posteriori condicional $f(\psi, \sigma|y, \theta)$ viene dada por las expresiones (1.12) y (1.13), mientras que la distribución a posteriori marginal $f(\theta|y)$ es proporcional a:

$$f(\theta|y) \propto |X'X|^{-\frac{1}{2}}((L^{-1}y)'(I_n - P_X)L^{-1}y)^{-\frac{n-s-1}{2}} \quad (1.14)$$

Generalmente se asume que el valor de los parámetros de suavizado, las tres componentes del vector $\theta = (\alpha, \beta, \gamma)'$, está incluido en el intervalo $(0, 1)$. Esta restricción puede ser introducida fácilmente en el modelo a través de la distribución a priori $f(\theta)$, considerando $f(\theta) \propto 1$ si las componentes de $\theta \in (0, 1)$ y $f(\theta) = 0$ en otro caso. En este caso, la distribución a posteriori $f(\theta|y)$ viene dada por (1.14) si todos los parámetros de suavizado pertenecen al intervalo unidad, mientras que toma el valor cero en caso contrario. En lo que resta de la memoria, salvo que se indique lo contrario, trabajaremos con esta restricción.

Dado que $f(\theta|y)$ no es analíticamente tratable, Bermúdez et al. (2008b) proponen utilizar un mecanismo de aceptación-rechazo para obtener una muestra de dicha distribución y estimar así las características de interés.

Sea $\{R_i\}_{i=1}^T$ una partición del cubo unidad, por ejemplo, la obtenida al dividir cada uno de los intervalos $(0,1)$ en $K = 4$ subintervalos, y sea $M_i = \sup_{\theta \in R_i} f(\theta|y)$. Entonces, la función importante $s(\theta|y)$ se define proporcional a M_i si $\theta \in R_i$, para $i = 1, 2, \dots, T$, y cero en otro caso; es decir, la función importante es la uniforme a bloques en el cubo unidad.

Una vez definida la función importante, en cada iteración del algoritmo se selecciona una región R_j con probabilidad proporcional a M_j y se simula $\theta^{(j)}$ uniforme en R_j . Si $uM_j < f(\theta^{(j)}|y)$, con $u \sim Un(0, 1)$, el valor $\theta^{(j)}$ es aceptado; en otro caso, el valor es rechazado y el proceso es repetido. Tras repetir el algoritmo de simulación N veces, se obtiene una muestra $\{\theta^{(j)}\}_{j=1}^N$ de tamaño N de la distribución a posteriori $f(\theta|y)$.

Una vez obtenida la distribución a posteriori para los parámetros del modelo, la distribución predictiva, la cual contiene toda la información acerca de los valores futuros de la serie, puede ser calculada. Sea P el vector $h \times 1$ de los datos futuros de la serie temporal. Entonces, la distribución predictiva $f(P|y)$ es obtenida al resolver la siguiente integral:

$$f(P|y) = \int \int \int f(P|y, \psi, \theta, \sigma) f(\psi, \theta, \sigma|y) d\psi d\theta d\sigma$$

Para el cálculo de $f(P|y, \psi, \theta, \sigma)$, Bermúdez et al. (2008b) suponen que el vector $(n + h) \times 1$ de los datos observados y futuros de la serie temporal, $(y', P)'$, se ajusta al modelo de Holt-Winters aditivo univariante y, por tanto, satisface la ecuación (1.9):

$$\begin{bmatrix} y \\ P \end{bmatrix} = \begin{bmatrix} M \\ M_p \end{bmatrix} \psi + \begin{bmatrix} L^1 & 0 \\ L^{21} & L^2 \end{bmatrix} \begin{bmatrix} \varepsilon \\ v \end{bmatrix} \quad (1.15)$$

donde las matrices M y L en (1.9) han sido completadas hasta orden $n + h$ y particionadas de manera análoga al vector $(y', P)'$; v es el vector de errores asociado a los datos futuros.

Si el vector θ de los parámetros de suavizado es conocido, la distribución predictiva resultante $f(P|y, \theta)$ es la t-Student multivariante con $n - s - 1$ grados de libertad y vector de medias y matriz de covarianzas dados por:

$$E(P|y, \theta) = M_p \tilde{\psi} + L^{21} (L^1)^{-1} (y - M \tilde{\psi}) \quad (1.16)$$

$$V(P|y, \theta) = \frac{n - s - 1}{n - s - 3} \hat{\sigma}^2 (L^2 (L^2)' + (M_p - L^{21} X) (X' X)^{-1} (M_p - L^{21} X)') \quad (1.17)$$

siendo $X = (L^1)^{-1} M$, $\tilde{\psi} = (X' X)^{-1} X' (L^1)^{-1} y$ el estimador mínimos cuadrados de ψ y $\hat{\sigma}^2 = \frac{1}{n - s - 1} ((L^1)^{-1} y)' (I_n - P_X) (L^1)^{-1} y$ el estimador insesgado de σ^2 .

En el caso general, θ desconocido, la distribución predictiva, así como sus principales momentos, pueden ser estimados mediante integración por Monte Carlo. Sea $\{\theta^{(j)}\}_{j=1}^N$ una muestra de la distribución a posteriori $f(\theta|y)$,

1.2. SERIES TEMPORALES DE OCUPACIÓN HOTELERA 11

entonces:

$$f(P|y) \approx \frac{1}{N} \sum_{j=1}^N MSt(\mu_P^{(j)}, V_P^{(j)}, n-s-1) \quad (1.18)$$

$$E(P|y) \approx \frac{1}{N} \sum_{j=1}^N \mu_P^{(j)} \quad (1.19)$$

$$V(P|y) \approx \frac{1}{N} \sum_{j=1}^N V_P^{(j)} + \frac{1}{N} \sum_{j=1}^N \mu_P^{(j)} (\mu_P^{(j)})' - E(P|y)E'(P|y) \quad (1.20)$$

donde $MSt(\mu_P^{(j)}, V_P^{(j)}, n-s-1)$ es la t-Student multivariante con $n-s-1$ grados de libertad y vector de medias $\mu_P^{(j)}$ y matriz de covarianzas $V_P^{(j)}$ resultantes de sustituir θ por $\theta^{(j)}$ en (1.16) y (1.17) respectivamente.

1.2. Series temporales de ocupación hotelera

A lo largo de la memoria, a fin de clarificar los procedimientos de análisis que introducimos, haremos referencia al siguiente ejemplo en el que se estudian las series mensuales de ocupación hotelera en Castellón, Valencia y Alicante desde Enero de 2001 hasta Diciembre de 2006. Los datos están disponibles en la página web del Instituto Nacional de Estadística (<http://www.ine.es>).

La actividad turística en la Comunidad Valenciana tiene una gran trascendencia económica y social, siendo una de las principales fuentes de ingresos. Dado que la demanda turística depende de factores tales como la inversión llevada a cabo en la zona, la calidad de los servicios ofertados, las preferencias de los turistas, etc., la predicción de la ocupación hotelera puede ayudar a las autoridades locales y a los profesionales del sector a adoptar las medidas necesarias de acuerdo a la situación prevista.

A continuación mostramos los principales resultados obtenidos en los análisis univariantes de las series temporales. Notar que el uso del modelo de Holt-Winters está justificado, pues las series presentan una tendencia creciente y estacionalidad aditiva, ver figura 1.1. Consideramos como datos históricos las observaciones correspondientes a los cinco primeros años (2001 - 2005), mientras que el año 2006 es utilizado para valorar la precisión de nuestra predicción fuera de la muestra.

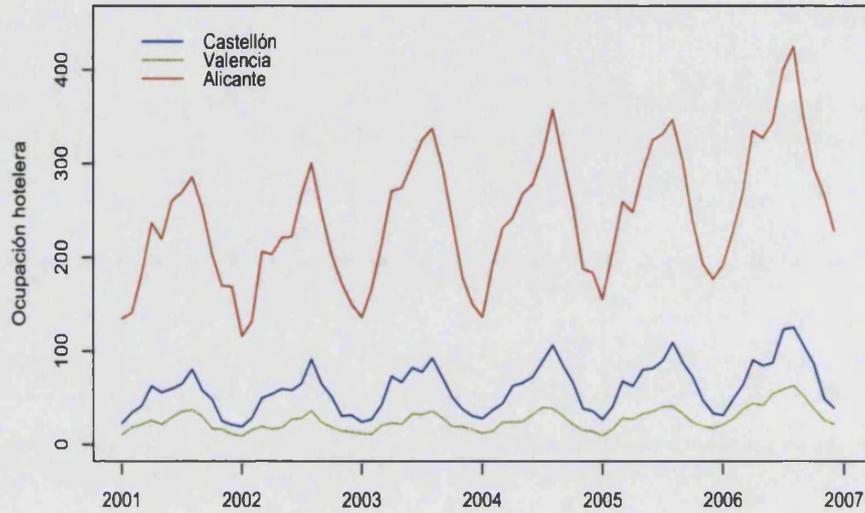


Figura 1.1: Gráfico temporal de las series mensuales de ocupación hotelera en Castellón, Valencia y Alicante desde Enero de 2001 hasta Diciembre de 2006. Unidades: miles de viajeros

La figura 1.2 muestra los histogramas de los parámetros de suavizado simulados de las distribuciones a posteriori $f(\theta_i|y_i)$ para $i = 1, 2, 3$, ver ecuación (1.14). En concreto simulamos $N = 10000$ valores utilizando el método de aceptación-rechazo descrito anteriormente. Las líneas rojas verticales representan las medias muestrales.

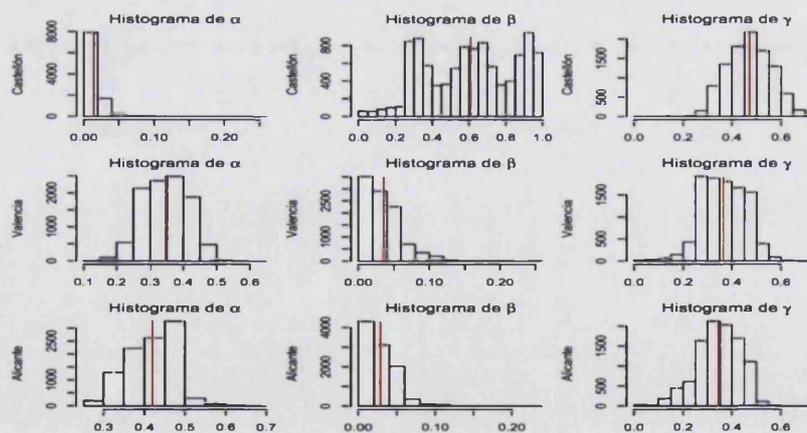


Figura 1.2: Histogramas de los parámetros de suavizado simulados de su distribución a posteriori en los análisis univariantes de las series temporales de ocupación hotelera

1.2. SERIES TEMPORALES DE OCUPACIÓN HOTELERA 13

El hecho de que el parámetro β correspondiente a Castellón presente tanta variabilidad se debe a que el valor de α es prácticamente cero y, por tanto, el parámetro β influye poco en el análisis de la serie.

En la figura 1.3 encontramos, en color negro, el ajuste obtenido para cada una de las series temporales en la zona de estimación (2001-2005) así como las predicciones para el año 2006. Las líneas negras discontinuas representan los intervalos de predicción del 95 %. Una vez obtenida la muestra $\{\theta_i^{(j)}\}_{j=1}^N$, las estimaciones de las observaciones que conforman la serie temporal i -ésima son calculadas utilizando (1.9), sustituyendo los parámetros de suavizado y las condiciones iniciales por sus medias muestrales. Análogamente, las predicciones puntuales son obtenidas a partir de (1.19). Como podemos ver, el año 2006 es un año atípico, con un incremento en la ocupación hotelera mayor que el esperado. En el caso de Valencia, este aumento puede ser debido a la celebración de las conocidas regatas de la America's Cup.

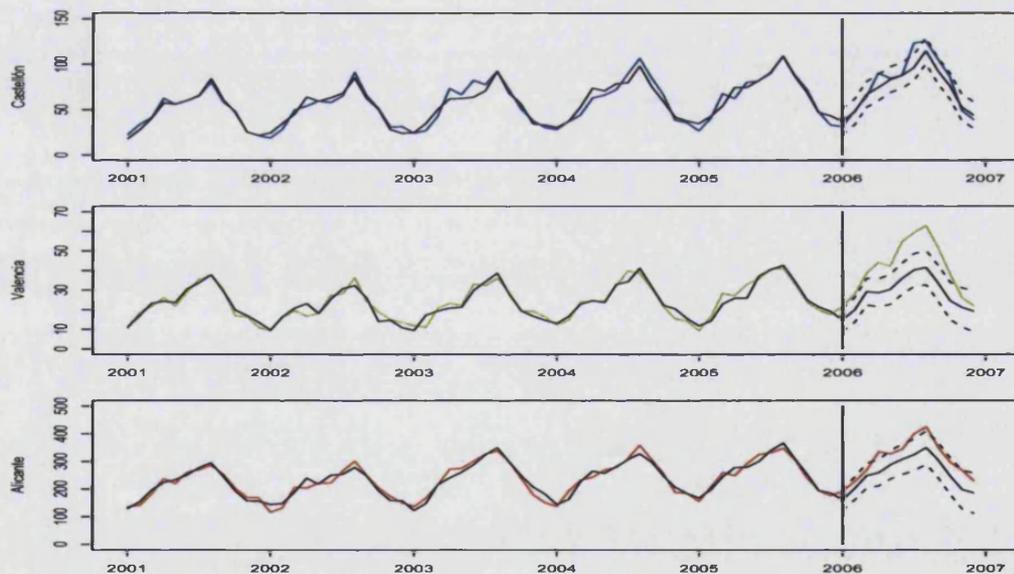


Figura 1.3: Línea negra: ajuste y predicciones obtenidos para cada una de las series temporales de ocupación hotelera con el modelo de Holt-Winters univariante. Líneas discontinuas: intervalos de predicción del 95 %

Por último, en la tabla 1.2 se presentan los errores de ajuste y de predicción correspondientes a los análisis univariantes de las series temporales. Como medida de precisión utilizamos el error SMAPE (symmetric mean

absolute percentage error), definido como:

$$\text{SMAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t + \hat{y}_t} * 200$$

donde y_t es la observación en el instante t e \hat{y}_t su estimación. Elegimos este error porque es independiente de la escala, simétrico y acotado: fluctúa entre -200 % y 200 %. Además, es una de las medidas de precisión más empleadas para comparar los modelos de suavizado exponencial.

	Castellón	Valencia	Alicante	Media
Error ajuste	9.29	9.42	6.23	8.31
Error predicción	10.13	35.00	17.58	20.90

Tabla 1.2: Error SMAPE de ajuste y de predicción resultantes de los análisis univariantes de las series temporales de ocupación hotelera

1.3. Modelos de suavizado exponencial multivariante

Jones (1966) es el primer autor que generalizó los modelos de suavizado exponencial simple al caso multivariante, expresándolos como modelos de espacio de estado con múltiples fuentes de error de la siguiente manera:

$$y_t = \mu_t + \xi_t \quad (1.21)$$

$$\mu_t = \mu_{t-1} + \eta_t \quad (1.22)$$

donde $y_t = (y_{t1}, y_{t2}, \dots, y_{tm})'$ es el vector $m \times 1$ de las observaciones en t ; μ_t es el vector $m \times 1$ de estado; $\xi_t = (\epsilon_{t1}, \epsilon_{t2}, \dots, \epsilon_{tm})'$ es el vector de los errores observados en t , distribuido según la Normal multivariante $N_m(0, \Sigma)$ y η_t es el vector de errores asociado al vector de estado, con distribución $N_m(0, Q)$.

En su artículo, Jones propone un método para la estimación de la matriz de peso óptima, equivalente matricial al parámetro de suavizado del modelo univariante, sin considerar la estimación máximo-verosímil de los parámetros del modelo.

Posteriormente, Enns et al. (1982) introducen una nueva clase de modelos de suavizado exponencial simple multivariantes, cuya formulación como

modelos de espacio de estado coincide con la propuesta previamente por Jones (1966), ecuaciones (1.21) y (1.22), con la propiedad de que las matrices de covarianzas de los vectores de errores asociados al vector de las observaciones y al vector de estado son proporcionales entre sí, es decir,

$$Q = q\Sigma \quad (1.23)$$

Harvey (1986) demuestra que dicha propiedad, equivalente a suponer que las series temporales univariantes tienen la misma estructura estocástica, permite obtener los estimadores máximo-verosímiles de los parámetros del modelo mediante técnicas univariantes, pues la función de verosimilitud del modelo multivariante puede ser expresada en términos de cantidades que aparecen únicamente en las funciones de verosimilitud univariantes. Del mismo modo, predicciones para los valores futuros de las series pueden ser calculadas mediante la aplicación del filtro de Kalman univariante, que será común para todas las series.

Así mismo, Harvey (1986) generaliza los modelos multivariantes propuestos de manera que las series temporales puedan presentar tendencia y estacionalidad. Su formulación como modelos de espacio de estado viene dada por:

$$y_t = (I_m \otimes z_t')\mu_t + \xi_t \quad (1.24)$$

$$\mu_t = (I_m \otimes T_t)\mu_{t-1} + \eta_t \quad (1.25)$$

donde z_t es un vector conocido de orden $k \times 1$; μ_t es el vector $km \times 1$ de estado; T_t es la matriz $k \times k$ de transición, no necesariamente conocida; η_t es el vector $km \times 1$ de errores asociado al vector de estado, distribuido según la Normal multivariante $N_{km}(0, \Sigma \otimes Q_t)$ y \otimes representa el producto de Kronecker. El hecho de que la matriz de covarianzas del vector de errores η_t sea de la forma $\Sigma \otimes Q_t$ permite, nuevamente, obtener los estimadores máximo-verosímiles de los parámetros del modelo y las predicciones para los valores futuros de las series a partir de técnicas univariantes.

Fernández and Harvey (1990) presentan una clase de modelos estructurales de series temporales multivariantes, compuestos de nivel, tendencia y estacionalidad, cuya formulación es similar a la propuesta por Harvey (1986), ecuaciones (1.24) y (1.25), pero donde las matrices de covarianzas no están sujetas, en principio, a ninguna restricción. La estimación máximo-verosímil de los parámetros del modelo es llevada a cabo mediante la aplicación del filtro de Kalman o, alternativamente, utilizando técnicas espectrales si la

verosimilitud es expresada en el dominio de las frecuencias (para un estudio detallado de estas técnicas consultar Fernández, 1990). Una vez los parámetros del modelo han sido estimados, predicciones para los valores futuros de las series pueden ser obtenidas a partir del filtro de Kalman. Posteriormente, los autores proponen un test para comprobar la hipótesis de homogeneidad supuesta en trabajos previos (Enns et al., 1982; Harvey, 1986).

Un enfoque distinto lo encontramos en el trabajo de Pfeffermann and Allon (1989), donde se introduce una nueva generalización del modelo de Holt-Winters para el caso multivariante. Sea $\{y_t\}_{t=1}^n$ la serie temporal multivariante de estudio, donde $y_t = (y_{t1}, y_{t2}, \dots, y_{tm})'$ es el vector $m \times 1$ de las observaciones en t . Entonces, el modelo supone que los vectores de las observaciones en cada instante temporal admiten la descomposición:

$$y_t = \tilde{L}_t + \tilde{S}_t + \xi_t \quad (1.26)$$

siendo \tilde{L}_t y \tilde{S}_t el nivel de la serie y la componente estacional respectivamente en t . ξ_t es el vector de los errores en t , distribuido según la Normal multivariante $N_m(0, \Sigma)$. Cada vez que un vector y_t es observado, los términos de nivel, tendencia y estacionalidad de cada una de las series temporales univariantes son actualizados mediante una media ponderada de las estimaciones derivadas del modelo univariante y de factores de corrección que recogen la información derivada de las demás series temporales. Es decir,

$$L_t = A(y_t - S_{t-1}^t) + (I_m - A)(L_{t-1} + T_{t-1}) \quad (1.27)$$

$$T_t = \Gamma(L_t - L_{t-1}) + (I_m - \Gamma)T_{t-1} \quad (1.28)$$

$$S_t^{*t} = \Delta(y_t - L_t) + (I_m - \Delta)S_{t-1}^t \quad (1.29)$$

$$\sum_{i=0}^{s-1} S_t^{t+i} = 0 \quad (1.30)$$

donde L_{t-1} representa la estimación del nivel de la serie en $t - 1$, T_{t-1} la última estimación de la tendencia, s la longitud del ciclo estacional y S_{t-1}^t la estimación obtenida en $t - 1$ para la componente estacional en t . Notar que la actualización de la componente estacional es llevada a cabo en dos pasos: en el primero de ellos se calcula la nueva estimación para t , S_t^{*t} , mientras que en el segundo, dicha estimación, junto con las obtenidas en $t - 1$ para los otros $s - 1$ periodos, son normalizadas de manera que su suma es cero.

La aplicación del modelo requiere de la especificación de los valores de inicio L_0 , T_0 y $\{S_0^i\}_{i=1}^s$, y de las matrices de suavizado A , Γ y Δ . La estimación de las condiciones iniciales es llevada a cabo mediante un algoritmo

heurístico, independiente del valor de las matrices de suavizado. Las matrices de los parámetros de suavizado pueden ser estimadas de manera subjetiva o minimizando alguna función de los errores de predicción de un paso. Predicciones puntuales para los valores futuros de las series temporales pueden ser calculadas como:

$$\hat{y}_{n+h} = L_n + hT_n + S_n^{n+h} \quad (1.31)$$

Recientemente, de Silva et al. (2007) han presentado un nuevo modelo multivariante, el modelo VISTS (vector innovation structural time series model), para el análisis conjunto de series temporales correlacionadas. Dicho modelo, aunque similar a los modelos estructurales de series temporales multivariantes (Harvey, 1986; Fernández and Harvey, 1990), se caracteriza por la incorporación de una única fuente de error. Errores contemporáneos en las series temporales son asumidos independientes entre sí, y la correlación entre éstas es introducida a través de las ecuaciones de actualización. Es decir, la evolución de las componentes no observables de las series univariantes, nivel, tendencia y estacionalidad, depende, no sólo del propio error de la serie, sino también de los errores correspondientes a las demás series. En concreto, el modelo puede ser descrito mediante las siguientes ecuaciones:

$$y_t = H\mu_{t-1} + \xi_t \quad (1.32)$$

$$\mu_t = F\mu_{t-1} + G\xi_t \quad (1.33)$$

donde $y_t = (y_{t1}, y_{t2}, \dots, y_{tm})'$ es el vector $m \times 1$ de las observaciones en t ; H es la matriz $m \times km$ de diseño, conocida y cuyos elementos toman el valor cero o uno; μ_{t-1} es el vector $km \times 1$ de estado; F es la matriz $km \times km$ de transición; G es una matriz $km \times m$ de parámetros generalmente desconocida y ξ_t es el vector $m \times 1$ de los errores observados en t , con distribución $N_m(0, \Sigma)$, siendo Σ una matriz diagonal.

Notar que, en realidad, este modelo es un caso particular del modelo propuesto previamente por Pfeffermann and Allon (1989), pues la ecuación de observación (1.26) y las ecuaciones de actualización (1.27), (1.28) y (1.29), expresadas en la forma de corrección de error, son equivalentes a (1.32) y (1.33) respectivamente, con la diferencia de que, en el modelo VISTS, Σ es supuesta diagonal; es decir, se supone independencia entre errores contemporáneos en las distintas series univariantes.

Una vez los estimadores máximo-verosímiles de los parámetros del modelo, μ_0 , G y Σ , son obtenidos, predicciones puntuales para los valores futuros

de las series temporales pueden ser calculadas a partir de las ecuaciones del modelo.

1.4. Principales técnicas de selección de modelos

En esta sección revisamos brevemente las principales técnicas Bayesianas que han sido propuestas para la resolución del problema de selección de modelos. Una de las herramientas más populares es el factor Bayes (Kass and Raftery, 1995), mediante el cual es posible comparar las probabilidades a posteriori de los modelos competidores. Otra propuesta Bayesiana, alternativa al factor Bayes, consiste en tratar el problema de selección de modelos como un problema de decisión, donde el modelo más complejo es asumido cierto y, basándose en las propiedades predictivas de los modelos, se valora la pérdida que supone sustituir el modelo verdadero por otro modelo más sencillo (Goutis and Robert, 1998; Bernardo and Rueda, 2002). Finalmente, el problema de selección de modelos puede ser resuelto mediante la utilización de un criterio de selección, como puede ser el AIC (Akaike Information Criterion, Akaike, 1974), el BIC (Bayesian Information Criterion, Schwarz, 1978) o el DIC (Deviance Information Criterion, Spiegelhalter et al., 2002).

1.4.1. Factores Bayes

Sean $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_K$ los modelos competidores para representar el comportamiento de los datos de estudio, donde cada modelo representa una familia de distribuciones $\{f(y|\varphi_k, \mathcal{M}_k)\}$ con φ_k el vector de parámetros del modelo k .

Una solución Bayesiana común al problema de selección de modelos (Robert, 2007) consiste en tomar el modelo con mayor probabilidad a posteriori. Para ello, además de asignar probabilidades a priori sobre los parámetros de cada modelo, es necesario asignar probabilidades a priori sobre los posibles modelos, es decir, para $k = 1, 2, \dots, K$ se determina:

1. π_k , probabilidad a priori del modelo \mathcal{M}_k
2. $f_k(\varphi_k)$, probabilidad a priori del vector $\varphi_k \in \Theta_k$ de los parámetros del modelo \mathcal{M}_k

Entonces, si $f_k(y|\varphi_k)$ es la función de verosimilitud correspondiente al modelo k , las probabilidades a posteriori de los modelos pueden calcularse

como:

$$\pi(\mathcal{M}_k|y) = \frac{f(y|\mathcal{M}_k)\pi_k}{\sum_{i=1}^K f(y|\mathcal{M}_i)\pi_i}$$

Comparaciones de modelos dos a dos pueden realizarse a partir del cociente de las probabilidades a posteriori:

$$\frac{\pi(\mathcal{M}_k|y)}{\pi(\mathcal{M}_j|y)} = \frac{f(y|\mathcal{M}_k)}{f(y|\mathcal{M}_j)} \times \frac{\pi_k}{\pi_j}$$

donde el cociente entre las verosimilitudes marginales, conocido como el factor Bayes B_{kj} , puede ser calculado como:

$$B_{kj} = \frac{f(y|\mathcal{M}_k)}{f(y|\mathcal{M}_j)} = \frac{\int_{\Theta_k} f_k(y|\varphi_k) f_k(\varphi_k) d\varphi_k}{\int_{\Theta_j} f_j(y|\varphi_j) f_j(\varphi_j) d\varphi_j} \quad (1.34)$$

El factor Bayes, factor de transición entre el cociente de las probabilidades a priori y el cociente de las probabilidades a posteriori, recoge la evidencia a favor del modelo k proporcionada por los datos. Los factores Bayes son los análogos Bayesianos de los tests de razón de verosimilitudes, donde los parámetros son eliminados mediante integración en lugar de maximización.

A pesar de la popularidad de los factores Bayes como solución al problema de selección de modelos (Kass and Raftery, 1995), pues son muy flexibles y permiten comparar simultáneamente hipótesis múltiples, en la práctica son difíciles de calcular. A diferencia de lo que ocurre en estimación, la elección de las distribuciones a priori en selección de modelos es fundamental. Las distribuciones a priori $f_k(\varphi_k)$ deben ser propias. En caso contrario, el factor Bayes dependerá de una constante arbitraria y, por tanto, no tendrá validez. Dicha restricción puede ser relajada para el caso de parámetros comunes en los distintos modelos pues, al ser la constante de normalización la misma, es posible utilizar distribuciones impropias. Del mismo modo, distribuciones a priori vagas, es decir, distribuciones propias con una varianza muy grande, tampoco resultan adecuadas.

i) Factores Bayes parciales

Varias alternativas han sido propuestas para resolver el problema de arbitrariedad del factor Bayes como consecuencia de utilizar distribuciones a priori impropias. Entre ellas cabe destacar los factores Bayes parciales que utilizan una parte de los datos, muestra de entrenamiento, para convertir la

distribución a priori impropia de los parámetros del modelo en una distribución propia. El resto de datos son utilizados para calcular el factor Bayes, utilizando como priori para los parámetros la distribución a posteriori obtenida con la muestra de entrenamiento. Así pues, si $y = (y'_1, y'_2)'$, donde y_1 es el vector de datos que conforman la muestra de entrenamiento e y_2 son los datos restantes empleados para la comparación de modelos, el factor Bayes parcial se define como:

$$B_{kj}(y_2|y_1) = \frac{f(y_2|\mathcal{M}_k)}{f(y_2|\mathcal{M}_j)} = \frac{\int_{\Theta_k} f_k(y_2|\varphi_k) f_k(\varphi_k|y_1) d\varphi_k}{\int_{\Theta_j} f_j(y_2|\varphi_j) f_j(\varphi_j|y_1) d\varphi_j} = \frac{B_{kj}}{B_{kj}(y_1)} \quad (1.35)$$

siendo B_{kj} el factor Bayes calculado a partir de y y $B_{kj}(y_1)$ el factor Bayes calculado a partir de y_1 , ver ecuación (1.34). El problema que presenta este enfoque es la determinación de la muestra de entrenamiento, pues para n observaciones existen $\binom{n}{r}$ muestras de entrenamiento de tamaño r .

O'Hagan (1995) presenta el factor Bayes fraccional como una expresión alternativa al factor Bayes parcial que evita la elección de una muestra de entrenamiento. Sea $b = r/n$, por analogía con (1.35) el factor Bayes fraccional se define como:

$$B_{kj}^F(b) = \frac{\int_{\Theta_k} f_k(y|\varphi_k) f_k(\varphi_k) d\varphi_k \int_{\Theta_j} [f_j(y|\varphi_j)]^b f_j(\varphi_j) d\varphi_j}{\int_{\Theta_j} f_j(y|\varphi_j) f_j(\varphi_j) d\varphi_j \int_{\Theta_k} [f_k(y|\varphi_k)]^b f_k(\varphi_k) d\varphi_k} \quad (1.36)$$

Para su cálculo es necesario definir previamente el valor de $b \in [\frac{m_0}{n}, 1]$, siendo m_0 el tamaño de una muestra de entrenamiento mínima. Dicho valor influye tanto en el poder discriminatorio del método, que aumenta conforme b disminuye, como en la robustez frente a la definición de las distribuciones a priori, que aumenta con b . O'Hagan propone tomar $b = m_0/n$ cuando la robustez no sea fundamental. Posteriormente, De Santis and Spezzaferri (1997), basándose en el trabajo de Berger and Pericchi (1996), demuestran que los factores Bayes fraccionales pueden ser asintóticamente equivalentes a factores Bayes calculados a partir de distribuciones a priori razonables.

Berger and Pericchi (1996) introducen un nuevo criterio, el factor Bayes intrínseco, que es totalmente automático en el sentido de que para su cálculo sólo se requieren distribuciones a priori convencionales no informativas. En su artículo demuestran que este nuevo criterio se corresponde, al menos asintóticamente, con factores Bayes reales obtenidos a partir de distribuciones a priori razonables conocidas como prioris intrínsecas. Sea

$y_{\mathcal{T}} = \{y_1^{(1)}, y_1^{(2)}, \dots, y_1^{(L)}\}$ el conjunto de todas las muestras de entrenamiento mínimas. En lugar de escoger una muestra de entrenamiento para el cálculo del factor Bayes parcial, la idea detrás de los factores Bayes intrínsecos es calcular el promedio de los factores Bayes parciales sobre todas las muestras de $y_{\mathcal{T}}$. Dependiendo de si la media es aritmética o geométrica se definen respectivamente:

$$B_{kj}^{AI} = \frac{1}{L} \sum_{l=1}^L B_{kj}(y_2^{(l)} | y_1^{(l)}) = B_{kj} \frac{1}{L} \sum_{l=1}^L B_{jk}(y_1^{(l)}) \quad (1.37)$$

$$B_{kj}^{GI} = \left(\prod_{l=1}^L B_{kj}(y_2^{(l)} | y_1^{(l)}) \right)^{\frac{1}{L}} = B_{kj} \left(\prod_{l=1}^L B_{jk}(y_1^{(l)}) \right)^{\frac{1}{L}} \quad (1.38)$$

donde $y_1^{(l)}$ representa una muestra de entrenamiento mínima e $y_2^{(l)} = y \setminus y_1^{(l)}$ el resto de datos utilizados para la comparación de modelos. Para modelos no encajados, y con el fin de evitar posibles problemas de inestabilidad, se recomienda construir un modelo general que abarque a todos los modelos propuestos, \mathcal{M}_0 , y calcular las comparaciones dos a dos a partir de $\{B_{0i}^{AI}\}_{i=1}^K$ utilizando la definición para modelos encajados. Para un estudio detallado de posibles modificaciones de estos factores se puede consultar Berger and Pericchi (1996).

ii) Distribuciones a priori esperanzas de posteriores

Posteriormente, Pérez and Berger (2002) desarrollan un método para calcular distribuciones a priori para los parámetros del modelo, a partir de prioris convencionales no informativas, de manera que los factores Bayes resultantes estén bien definidos. La clave está en utilizar una distribución predictiva subyacente común para todos los modelos en la definición de estas distribuciones a priori, conocidas como distribuciones a priori esperanzas de posteriores (PEP).

Se supone que se dispone de una muestra de entrenamiento imaginaria, y^* , que puede ser utilizada para convertir una distribución a priori convencional no informativa $f_i^N(\varphi_i)$ en propia, y sea $m^*(y^*)$ una distribución predictiva adecuada para y^* . Entonces, la PEP es calculada como:

$$f_i^*(\varphi_i) = \int f_i^N(\varphi_i | y^*) m^*(y^*) dy^* \quad (1.39)$$

mientras que el factor Bayes es de la forma:

$$B_{kj}^* = \frac{\int_{\Theta_k} f_k(y|\varphi_k) f_k^*(\varphi_k) d\varphi_k}{\int_{\Theta_j} f_j(y|\varphi_j) f_j^*(\varphi_j) d\varphi_j}$$

De este modo se evita el problema de indeterminación de los factores Bayes. $f_i^N(\varphi_i|y^*)$ no depende de constantes arbitrarias y el que la predictiva $m^*(y^*)$ no sea propia no repercute en la arbitrariedad del factor Bayes, pues aparece en todos los modelos. Como posible distribución predictiva puede tomarse la distribución empírica de los datos reales y :

$$m^*(y^*) = \frac{1}{L} \sum_{l=1}^L I_{y(l)}(y^*)$$

donde I_A representa la función indicatriz de un conjunto A e $\{y(l)\}_{l=1}^L$ son submuestras de y de tamaño m_0 tales que $f_i^N(\varphi_i|y(l))$ existe para todos los modelos. La PEP correspondiente a esta predictiva, PEP empírica, es de la forma:

$$f_i^*(\varphi_i) = \frac{1}{L} \sum_{l=1}^L f_i^N(\varphi_i|y(l))$$

Notar que éste es un posible uso de muestras de entrenamiento mínimas alternativo al descrito anteriormente para el cálculo de los factores Bayes intrínsecos, ecuaciones (1.37) y (1.38). En lugar de calcular el promedio de los factores Bayes parciales correspondientes a las distintas muestras de entrenamiento, primero se calcula el promedio de las distribuciones a posteriori propias obtenidas con las diferentes muestras de entrenamiento y, posteriormente, se calculan los factores Bayes con las distribuciones obtenidas.

iii) Estimación del factor Bayes mediante técnicas de simulación MCMC

Un segundo problema relacionado con los factores Bayes es que, con frecuencia, las verosimilitudes marginales $f(y|\mathcal{M}_k) = \int f_k(y|\varphi_k) f_k(\varphi_k) d\varphi_k$ no pueden ser calculadas de manera analítica. Varios autores han recurrido a la simulación con el fin de obtener estimaciones adecuadas para la selección.

Carlin and Chib (1995) aplican la metodología de Gibbs sampling al problema de selección de modelos. Además de las probabilidades a priori para los posibles modelos, $\{\pi_k\}_{k=1}^K$, y de las distribuciones a priori para los parámetros de cada modelo, $f_k(\varphi_k)$, introducen un nuevo parámetro M ,

indicador del modelo, y las distribuciones a priori $\{f_k(\varphi_i)\}_{i \neq k}$, llamadas pseudo-prioris. De este modo, la distribución conjunta del vector de datos y y de los parámetros de todos los modelos, $\varphi_1, \varphi_2, \dots, \varphi_K$, cuando $M = j$ es de la forma:

$$f(y, \varphi_1, \varphi_2, \dots, \varphi_K, M = j) = \pi_j \left(\prod_{i=1}^K f_j(\varphi_i) \right) f_j(y|\varphi_j)$$

y el método de Gibbs sampling es implementado a partir de las siguientes condicionales completas:

$$f(\varphi_k|y, \varphi_{i \neq k}, M = j) \propto \begin{cases} f_j(y|\varphi_j) f_j(\varphi_j) & \text{si } k = j \\ f_j(\varphi_k) & \text{si } k \neq j \end{cases}$$

$$\pi(M = j|y, \varphi_1, \varphi_2, \dots, \varphi_K) = \frac{\pi_j \left(\prod_{i=1}^K f_j(\varphi_i) \right) f_j(y|\varphi_j)}{\sum_{k=1}^K \pi_k \left(\prod_{i=1}^K f_k(\varphi_i) \right) f_k(y|\varphi_k)}$$

Finalmente, la selección de modelos se lleva a cabo a partir de las estimaciones:

$$\hat{\pi}(M = j|y) = \frac{\text{número de } M^{(g)} = j}{\text{número total de } M^{(g)}}, \quad j = 1, 2, \dots, K$$

Green (1995), argumentando que este método puede resultar aparatoso e ineficiente debido a la necesidad de definir y simular de las pseudo-prioris $\{f_k(\varphi_i)\}_{i \neq k}$, propone un nuevo procedimiento que recurre a los métodos MCMC para generar observaciones de la distribución a posteriori conjunta $f(\mathcal{M}_k, \varphi_k|y)$. Dicho método, conocido como método MCMC de salto reversible, está basado en la creación de una cadena de Markov que puede saltar entre los distintos modelos. Si $(\mathcal{M}_k, \varphi_k)$ es el estado actual de la cadena de Markov, se propone un nuevo modelo \mathcal{M}_j con probabilidad $j(\mathcal{M}_k, \mathcal{M}_j)$ y se obtiene una simulación de los parámetros de este modelo. Para ello se introducen parámetros auxiliares de manera que los dos modelos tengan la misma dimensión y biyecciones entre los nuevos vectores de parámetros de los modelos. El salto a $(\mathcal{M}_j, \varphi_j)$ dependerá de una cierta probabilidad $\alpha_{k,j}$.

En el trabajo de Raftery (1996) podemos encontrar distintos métodos de estimación de las verosimilitudes marginales de los modelos propuestos, $f(y|\mathcal{M}_k)$ para $k = 1, 2, \dots, K$. La primera clase de métodos, conocidos como estimadores por muestreo importante, requieren de una densidad conocida

$cg(\varphi_k)$ de la que es posible obtener una muestra $\{\varphi_k^{(j)}\}_{j=1}^N$. Entonces, la verosimilitud marginal expresada en la forma:

$$f(y|\mathcal{M}_k) = \int_{\Theta_k} f_k(y|\varphi_k) \frac{f_k(\varphi_k)}{cg(\varphi_k)} cg(\varphi_k) d\varphi_k$$

puede ser estimada mediante integración por Monte Carlo como:

$$\widehat{f}(y|\mathcal{M}_k) \approx \frac{1}{N} \sum_{j=1}^N \frac{f_k(y|\varphi_k^{(j)}) f_k(\varphi_k^{(j)})}{cg(\varphi_k^{(j)})} \quad (1.40)$$

En el caso en el que la constante $c = \left[\int_{\Theta_k} g(\varphi_k) d\varphi_k \right]^{-1}$ no es conocida, la ecuación (1.40) adopta la forma:

$$\widehat{f}_{IS}(y|\mathcal{M}_k) \approx \frac{\frac{1}{N} \sum_{j=1}^N \frac{f_k(y|\varphi_k^{(j)}) f_k(\varphi_k^{(j)})}{g(\varphi_k^{(j)})}}{\frac{1}{N} \sum_{j=1}^N \frac{f_k(\varphi_k^{(j)})}{g(\varphi_k^{(j)})}} \quad (1.41)$$

El estimador más sencillo se obtiene al tomar $g(\varphi_k) = f_k(\varphi_k)$. En este caso, la estimación de la verosimilitud marginal resultante es el promedio de las verosimilitudes calculadas en la muestra obtenida a partir de la distribución a priori de los parámetros del modelo. Esta estimación puede resultar muy ineficiente si la distribución a posteriori está mucho más concentrada que la priori, pues la mayoría de las simulaciones tendrán un valor de la verosimilitud muy pequeño y la estimación estará dominada por unos pocos valores. Como posible alternativa puede tomarse $g(\varphi_k) = f_k(y|\varphi_k) f_k(\varphi_k)$, es decir, la distribución a posteriori. La estimación obtenida, conocida como media armónica de la verosimilitud, es muy inestable, con varianza generalmente infinita. Una posible solución, que da lugar a estimaciones más precisas que las anteriores, consiste en tomar como función importante una mixtura de la priori y la posteriori, $g(\varphi_k) = \delta f_k(\varphi_k) + (1-\delta) f_k(y|\varphi_k) f_k(\varphi_k)$ con $0 < \delta < 1$. DiCiccio et al. (1997) utilizan la muestra obtenida de la distribución a posteriori de los parámetros del modelo para definir la función importante. En particular, proponen tomar $cg(\varphi_k) = N_p(\widehat{\varphi}_k, \widehat{\Sigma})$, es decir, la Normal multivariante de media $\widehat{\varphi}_k$ la moda a posteriori y matriz de covarianzas $\widehat{\Sigma}$ igual a menos la inversa del Hessiano del logaritmo de la posteriori evaluada en $\widehat{\varphi}_k$.

El siguiente método que encontramos en el trabajo de Raftery (1996) es el estimador Laplace-Metropolis cuyo objetivo es estimar los términos

que aparecen en el estimador de Laplace, pues en muchas ocasiones no pueden ser calculados analíticamente, mediante simulación. Sea $h(\varphi_k) = f_k(y|\varphi_k)f_k(\varphi_k)$, el estimador de Laplace de la verosimilitud marginal es de la forma:

$$\hat{f}_L(y|\mathcal{M}_k) = (2\pi)^{\frac{p}{2}} |\hat{\Sigma}|^{\frac{1}{2}} h(\hat{\varphi}_k) \quad (1.42)$$

donde p es el número de parámetros, $\hat{\varphi}_k$ la moda a posteriori y $\hat{\Sigma}$ menos la inversa del Hessiano del logaritmo de $h(\varphi_k)$ evaluada en $\hat{\varphi}_k$. Entonces, el estimador de Laplace-Metropolis se obtiene al sustituir en (1.42) $\hat{\varphi}_k$ y $\hat{\Sigma}$ por sus estimaciones obtenidas mediante simulación: $\hat{\varphi}_k$ puede ser estimada por la media o mediana muestral y $\hat{\Sigma}$ por la matriz de covarianzas muestral. Para un estudio detallado de estimadores más robustos, ver Raftery (1996), DiCiccio et al. (1997) y las referencias que ahí aparecen.

1.4.2. Selección de modelos como un problema de decisión

Un enfoque distinto al presentado para la selección de modelos, basado en la comparación de las propiedades predictivas de los modelos, lo podemos encontrar en los trabajos de Goutis and Robert (1998) y Bernardo and Rueda (2002). Dado un modelo general $\mathcal{M} \equiv \{f(y|\varphi), \varphi \in \Theta\}$ asumido cierto, se desea evaluar la compatibilidad de los datos con un modelo más sencillo $\mathcal{M}_0 \equiv \{f(y|\varphi_0), \varphi_0 \in \Theta_0 \subset \Theta\}$, obtenido al considerar ciertas restricciones sobre el vector de parámetros. Formalmente, este problema de selección puede ser formulado como un problema de decisión donde el espacio de acciones está constituido por dos elementos: A_0 , aceptar el modelo simplificado \mathcal{M}_0 , y A_1 , rechazar \mathcal{M}_0 . Para resolverlo es necesario introducir unas funciones de pérdida, $l(A_i|\varphi)$ para $i = 0, 1$, que valoren las consecuencias de aceptar o rechazar \mathcal{M}_0 . De este modo, \mathcal{M}_0 será rechazado si la pérdida esperada a posteriori de aceptarlo es mayor que la de rechazarlo; es decir,

$$\text{Rechazar } \mathcal{M}_0 \quad \text{si y sólo si} \quad \int_{\Theta} \underbrace{(l(A_0|\varphi) - l(A_1|\varphi))}_{\Delta l(\mathcal{M}_0|\varphi)} f(\varphi|y) d\varphi > 0$$

Tomando $\Delta l(\mathcal{M}_0|\varphi) = \delta(\varphi_0, \varphi) - d^*$ donde $\delta(\varphi_0, \varphi)$ mide la discrepancia entre los dos modelos y d^* es un valor de utilidad que valora la ventaja que supone trabajar con el modelo más sencillo cuando éste es válido, la solución al problema de selección viene dada por:

$$\text{Rechazar } \mathcal{M}_0 \quad \text{si y sólo si} \quad \int_{\Theta} \delta(\varphi_0, \varphi) f(\varphi|y) d\varphi > d^*$$

Goutis and Robert (1998) proponen tomar como medida de discrepancia entre los dos modelos la función:

$$\delta(\varphi_0, \varphi) = \inf_{\varphi_0 \in \Theta_0} \underbrace{\int f(y|\varphi) \log \frac{f(y|\varphi)}{f(y|\varphi_0)} dy}_{d\{f(\cdot|\varphi), f(\cdot|\varphi_0)\}} \quad (1.43)$$

donde $d\{f(\cdot|\varphi), f(\cdot|\varphi_0)\}$, distancia generalizada de Kullback-Leibler, valora el efecto de sustituir $f(y|\varphi)$ por $f(y|\varphi_0)$ cuando $f(y|\varphi)$ es cierta.

De manera similar, Bernardo and Rueda (2002) proponen como medida de discrepancia la siguiente función:

$$\delta(\varphi_0, \varphi) = \min_{\varphi_0 \in \Theta_0} \left\{ \min \left\{ \int f(y|\varphi) \log \frac{f(y|\varphi)}{f(y|\varphi_0)} dy, \int f(y|\varphi_0) \log \frac{f(y|\varphi_0)}{f(y|\varphi)} dy \right\} \right\} \quad (1.44)$$

donde la diferencia respecto a (1.43) es que, mientras Goutis and Robert (1998) suponen que el modelo general \mathcal{M} es el modelo verdadero, Bernardo and Rueda (2002) suponen que \mathcal{M} es aceptado válido para representar el comportamiento de los datos, pero consideran la posibilidad de que \mathcal{M}_0 sea el modelo verdadero.

Notar que cuando se considera como función de pérdida la función 0-1, la cual no tiene en cuenta la distancia al modelo verdadero en caso de que éste no sea seleccionado, la solución al problema de selección de modelos basada en la teoría de la decisión coincide con la solución previamente propuesta de tomar aquel modelo con mayor probabilidad a posteriori (factores Bayes).

1.4.3. Criterios de selección de modelos

Finalmente, el problema de selección de modelos puede ser resuelto utilizando distintos criterios de selección. Entre los más conocidos podemos destacar el AIC (Akaike Information Criterion, Akaike, 1974) y el BIC (Bayesian Information Criterion, Schwarz, 1978), definidos como:

$$AIC = -2 \log f(y|\hat{\varphi}) + 2p \quad (1.45)$$

$$BIC = -2 \log f(y|\hat{\varphi}) + \log(n)p \quad (1.46)$$

donde $\hat{\varphi}$ es el estimador máximo verosímil, p el número de parámetros del modelo y n el número de observaciones. Entonces, dado un conjunto de posibles modelos, el modelo seleccionado para un análisis posterior de los

datos es aquel que minimiza el criterio utilizado. El primer término en la definición valora el ajuste obtenido por el modelo, mientras que el segundo término, función del número de parámetros incluidos en el modelo, penaliza la sobreparametrización. Como podemos observar, los modelos seleccionados mediante el criterio BIC serán, en general, más sencillos que los escogidos mediante el AIC.

El criterio AIC es un estimador asintóticamente insesgado de la información de Kullback-Leibler esperada, que mide la información perdida cuando el modelo cierto es sustituido por un modelo aproximado. Una justificación Bayesiana a favor de su uso es que, cuando la precisión de la distribución a priori es comparable a la de la verosimilitud, las comparaciones de modelos basadas en los factores Bayes y en el criterio AIC son equivalentes asintóticamente, aunque esta propiedad no es común en la práctica. Para un estudio más detallado de las propiedades de este criterio consultar Burnham and Anderson (2004).

Cuando el número de parámetros es grande en comparación con el número de observaciones, o cuando el número de observaciones es pequeño, existe una modificación del criterio AIC, conocido como el criterio AICc, definido de la siguiente manera (Hurvich and Tsai, 1989):

$$AIC_c = -2 \log f(y|\hat{\varphi}) + 2p + \frac{2p(p+1)}{n-p-1} \quad (1.47)$$

Este criterio debería ser utilizado siempre y cuando no se verifique la condición $n/p > 40$ para el modelo con mayor número de parámetros.

El criterio BIC es una aproximación asintótica de $-2 \log (\pi(\mathcal{M}|y))$, siendo $\pi(\mathcal{M}|y)$ la probabilidad a posteriori del modelo \mathcal{M} . De este modo, el modelo con un valor del BIC menor es el modelo con mayor probabilidad a posteriori. Neath and Cavanaugh (1997) investigan la derivación de este criterio con el objetivo de determinar términos descartados en su definición por ser asintóticamente insignificantes pero que podrían ser relevantes cuando el número de observaciones es pequeño, proponiendo varias modificaciones resultantes de incluir algunos de los términos descartados.

Numerosos criterios han sido propuestos desde la aparición del AIC y del BIC, de los cuales cabe destacar el DIC (Deviance Information Criterion,

Spiegelhalter et al., 2002):

$$DIC = E(D(\varphi)|y) + \underbrace{E(D(\varphi)|y) - D(E(\varphi|y))}_{P_D} \quad (1.48)$$

siendo $D(\varphi) = -2 \log f(y|\varphi) + 2 \log f'(y)$ la desviación Bayesiana, con $f'(y)$ un término conocido dependiente únicamente de los datos. Notar que para la comparación de modelos, dado que $f'(y)$ depende de los datos y no del modelo, puede tomarse $D(\varphi) = -2 \log f(y|\varphi)$.

El primer término de la definición, esperanza a posteriori de la desviación, es una medida Bayesiana de la bondad del ajuste, mientras que el término P_D , conocido como número efectivo de parámetros en el modelo, es una medida de la complejidad, la cual es razonable que dependa de la información a priori acerca de los parámetros de interés y de las observaciones. Su justificación está basada en la teoría de la información.

Expresando el criterio como

$$DIC = D(E(\varphi|y)) + 2P_D$$

puede verse su analogía con el AIC o el BIC. Sin embargo, este criterio puede resultar más satisfactorio que los anteriores en el sentido de que tiene en cuenta la información a priori. Además, pueden utilizarse distribuciones a priori impropias, pues cada modelo es considerado por separado.

Capítulo 2

Holt-Winters multivariante

Para el caso univariante, los métodos de suavizado exponencial, tanto con tendencia como con estacionalidad aditiva o multiplicativa, han sido analizados desde la metodología de espacio de estado (Hyndman et al., 2002; Bermúdez et al., 2006b). Recientemente, Bermúdez et al. (2007) han formulado el modelo de Holt-Winters como un modelo lineal heterocedástico, cuyos coeficientes vienen dados por las condiciones iniciales y la matriz de covarianzas depende de los parámetros de suavizado. Esta formulación simplifica el cálculo de los estimadores máximo-verosímiles de los parámetros del modelo y la obtención de predicciones, tanto puntuales como intervalos de predicción, para los valores futuros de la serie temporal.

Basándonos en ese trabajo, introducimos una nueva formulación para el modelo de Holt-Winters multivariante con estacionalidad aditiva y errores aditivos. El modelo asume que cada una de las series temporales se ajusta al modelo de Holt-Winters univariante y que existe una correlación contemporánea entre errores correspondientes en los modelos univariantes. Entonces, utilizando la formulación para el modelo de Holt-Winters univariante como un modelo lineal heterocedástico, el modelo de Holt-Winters multivariante, resultante de formular conjuntamente los modelos univariantes, puede ser expresado como un modelo de regresión aparentemente no relacionado (modelo SUR, Zellner, 1962).

La representación de los métodos de suavizado exponencial como modelos de espacio de estado permite estudiar sus propiedades estadísticas, calcular la función de verosimilitud y obtener predicciones puntuales e intervalos de predicción. A diferencia de las predicciones puntuales, que han demos-

trado ser muy precisas, los intervalos de predicción tienden a ser demasiado estrechos. Ésto es debido principalmente a que en su cálculo no se tiene en cuenta el error de estimación de los parámetros del modelo. Bermúdez et al. (2008b) analizan el modelo de Holt-Winters univariante desde el punto de vista Bayesiano, resolviendo de este modo el problema de error de estimación y obteniendo intervalos de predicción con cobertura empírica próxima a la nominal.

El enfoque Bayesiano para el análisis de un modelo probabilístico que describe el comportamiento de una serie temporal implica condicionar a lo conocido, datos observados, estructuras, etc., para obtener inferencias acerca de lo desconocido: parámetros del modelo y valores futuros de la serie. Ésto implica la resolución de problemas de integración que en la mayoría de los casos no tienen solución analítica. Los avances en integración numérica y métodos de simulación han incrementado el número de problemas de predicción que pueden ser tratados desde el punto de vista Bayesiano (Geweke and Whiteman, 2006).

En este trabajo analizamos el modelo de Holt-Winters multivariante desde una perspectiva Bayesiana, obteniendo la distribución a posteriori de los parámetros del modelo (condiciones iniciales, parámetros de suavizado y matriz de covarianzas) dados los datos observados que conforman las series temporales. En el análisis consideramos distintas distribuciones a priori. La primera de ellas es la distribución a priori convencional no informativa de Jeffreys, mientras que las otras son distribuciones a priori de referencia propuestas por Chang and Eaves (1990) y Yang and Berger (1994). Las distribuciones a posteriori obtenidas no son analíticamente tratables pero pueden ser estimadas mediante métodos MCMC. El procedimiento de simulación empleado nos permite obtener una muestra de la distribución a posteriori de los coeficientes de correlación y, por tanto, estimar la correlación entre las series temporales. Finalmente, la distribución predictiva de los valores futuros de las series temporales, la cual nos permite obtener tanto estimaciones puntuales como intervalos de predicción, es estimada mediante integración por Monte Carlo.

En la siguiente sección se presenta la formulación del modelo de Holt-Winters multivariante empleada en la memoria. La segunda sección muestra el análisis Bayesiano del modelo y los métodos MCMC empleados en la simulación de muestras de las distribuciones a posteriori obtenidas para los parámetros del modelo. La sección tercera está dedicada al cálculo de la

distribución predictiva y a la obtención de predicciones puntuales e intervalos de predicción.

2.1. Formulación del modelo

Supongamos que tenemos n observaciones de m series temporales, $y_i = (y_{1i}, y_{2i}, \dots, y_{ni})'$ para $i = 1, 2, \dots, m$, cada una de las cuales se ajusta al modelo de Holt-Winters univariante y, por tanto, satisfacen (1.9):

$$y_i = M\psi_i + L_i\varepsilon_i \quad (2.1)$$

donde $\psi_i = (b_{0i}, c_{1-s,i}, c_{2-s,i}, \dots, c_{0,i})'$ es el vector $(s+1) \times 1$ de las condiciones iniciales, $\theta_i = (\alpha_i, \beta_i, \gamma_i)'$ es el vector de los parámetros de suavizado, L_i es la matriz L calculada en θ_i y $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i}, \dots, \varepsilon_{ni})'$ es el vector $n \times 1$ de errores de la i -ésima serie temporal, independientes entre sí e idénticamente distribuidos $N(0, \sigma_i^2)$.

Suponiendo además que existe una correlación contemporánea entre errores correspondientes en las diferentes ecuaciones, es decir, la distribución del vector $m \times 1$ de los errores observados en el instante t , $\xi_t = (\varepsilon_{t1}, \varepsilon_{t2}, \dots, \varepsilon_{tm})'$, es $N_m(0, \Sigma)$, el modelo de Holt-Winters multivariante resultante de formular conjuntamente los m modelos univariantes es de la forma:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_y = \underbrace{\begin{bmatrix} M & 0 & \dots & 0 \\ 0 & M & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M \end{bmatrix}}_{M_D} \underbrace{\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix}}_\psi + \underbrace{\begin{bmatrix} L_1 & 0 & \dots & 0 \\ 0 & L_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L_m \end{bmatrix}}_{L_B} \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}}_\varepsilon \quad (2.2)$$

donde el vector de errores ε está distribuido según una Normal multivariante de media el vector nulo y matriz de covarianzas $\Sigma \otimes I_n$, siendo \otimes el producto de Kronecker. Por tanto, la distribución del vector de datos es $N_{nm}(M_D\psi, L_B(\Sigma \otimes I_n)L_B')$. El subíndice D que aparece en M_D indica que M_D es una matriz diagonal a bloques, siendo todos sus bloques iguales a M . En cambio, el subíndice B que aparece en L_B indica que es una matriz diagonal a bloques, con bloques (L_1, L_2, \dots, L_m) distintos entre sí.

Multiplicando (2.2) por la matriz inversa de L_B se obtiene la expresión

matricial equivalente:

$$\underbrace{\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}}_z = \underbrace{\begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_m \end{bmatrix}}_{X_B} \underbrace{\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix}}_\psi + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}}_\varepsilon \quad (2.3)$$

siendo $z_i = (L_i)^{-1}y_i$ y $X_i = (L_i)^{-1}M$, para $i = 1, 2, \dots, m$. Entonces, el modelo de Holt-Winters multivariante puede formularse como un modelo de regresión aparentemente no relacionado (Zellner, 1962). En este caso, la matriz de regresión X_B depende de los parámetros de suavizado y por tanto es desconocida.

La función de verosimilitud asociada al modelo de Holt-Winters multivariante viene dada por la distribución Normal multivariante del vector de datos:

$$\begin{aligned} f(y|\psi, \Sigma, \theta) &\propto |L_B(\Sigma \otimes I_n)L'_B|^{-\frac{1}{2}} \\ &\quad \exp \left\{ -\frac{1}{2}(y - M_D\psi)'(L'_B)^{-1}(\Sigma^{-1} \otimes I_n)L_B^{-1}(y - M_D\psi) \right\} \\ &\propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2}(z - X_B\psi)'(\Sigma^{-1} \otimes I_n)(z - X_B\psi) \right\} \end{aligned} \quad (2.4)$$

$$\propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2}\text{tr} [\Sigma^{-1}H'H] \right\} \quad (2.5)$$

donde $\theta = (\theta'_1, \theta'_2, \dots, \theta'_m)'$ es el vector de los parámetros de suavizado de los m modelos univariantes y $H = [z_1 - X_1\psi_1, z_2 - X_2\psi_2, \dots, z_m - X_m\psi_m]$.

Con el objetivo de comparar nuestra formulación para el modelo de Holt-Winters multivariante con las propuestas realizadas anteriormente por otros autores, apartado 1.3, lo expresamos como un modelo de espacio de estado. Para $i = 1, 2, \dots, m$, la observación en el instante t de la serie temporal i -ésima puede expresarse (ver ecuación (1.2)) como:

$$y_{ti} = f'\varphi_{ti} + \epsilon_{ti} \quad (2.6)$$

siendo $f' = (1, 1, 0, 0, \dots, 0, 1)$ y $\varphi_{ti} = (a_{t-1,i}, b_{t-1,i}, c_{t-1,i}, c_{t-2,i}, \dots, c_{t-s,i})'$ el vector $(s+2) \times 1$ de estado tal que:

$$\varphi_{ti} = G\varphi_{t-1,i} + \eta_{ti} \quad (2.7)$$

donde G es la matriz de transición $(s+2) \times (s+2)$ definida como en (1.2) y $\eta_{ti} = (\alpha_i \epsilon_{t-1,i}, \alpha_i \beta_i \epsilon_{t-1,i}, \gamma_i \epsilon_{t-1,i}, 0, \dots, 0, 0)'$ es el vector de errores asociado al vector de estado, distribuido según la Normal multivariante de media 0 y matriz de covarianzas dada por:

$$V(\eta_{ti}) = \sigma_i^2 \begin{pmatrix} W_{ii} & 0_{3 \times (s-1)} \\ 0_{(s-1) \times 3} & 0_{(s-1) \times (s-1)} \end{pmatrix} \quad (2.8)$$

$$\text{con } W_{ii} = \begin{pmatrix} \alpha_i^2 & \alpha_i^2 \beta_i & \alpha_i \gamma_i \\ \alpha_i^2 \beta_i & \alpha_i^2 \beta_i^2 & \alpha_i \beta_i \gamma_i \\ \alpha_i \gamma_i & \alpha_i \beta_i \gamma_i & \gamma_i^2 \end{pmatrix}$$

Entonces, el vector $y_t = (y_{t1}, y_{t2}, \dots, y_{tm})'$ de las m observaciones en t puede ser expresado como:

$$y_t = (I_m \otimes f') \varphi_t + \xi_t \quad (2.9)$$

donde

$$\varphi_t = (I_m \otimes G) \varphi_{t-1} + \eta_t \quad (2.10)$$

siendo $\varphi_t = (\varphi'_{t1}, \varphi'_{t2}, \dots, \varphi'_{tm})'$ el vector $(s+2)m \times 1$ de estado; $\xi_t = (\epsilon_{t1}, \epsilon_{t2}, \dots, \epsilon_{tm})'$ el vector de los errores en t , cuya distribución es la Normal multivariante $N_m(0, \Sigma)$ y $\eta_t = (\eta'_{t1}, \eta'_{t2}, \dots, \eta'_{tm})'$ el vector de errores $(s+2)m \times 1$ asociado al vector de estado, con media 0 y donde el bloque (i, j) de la matriz de covarianzas es de la forma:

$$\text{Cov}(\eta_{ti}, \eta_{tj}) = \sigma_{ij} \begin{pmatrix} \alpha_i \alpha_j & \alpha_i \alpha_j \beta_j & \alpha_i \gamma_j & 0_{1 \times (s-1)} \\ \alpha_i \beta_i \alpha_j & \alpha_i \beta_i \alpha_j \beta_j & \alpha_i \beta_i \gamma_j & 0_{1 \times (s-1)} \\ \gamma_i \alpha_j & \gamma_i \alpha_j \beta_j & \gamma_i \gamma_j & 0_{1 \times (s-1)} \\ 0_{(s-1) \times 1} & 0_{(s-1) \times 1} & 0_{(s-1) \times 1} & 0_{(s-1) \times (s-1)} \end{pmatrix}$$

Si comparamos nuestra formulación con la propuesta por Harvey (1986), ecuaciones (1.24) y (1.25), la primera diferencia que encontramos es que nosotros trabajamos con una única fuente de error. No obstante, para poder expresar la matriz de covarianzas del vector η_t como $\Sigma \otimes W$, es necesario que los parámetros de suavizado de las distintas series temporales univariantes sean iguales. En este sentido, nuestro modelo es más general.

Por otro lado, el modelo de Holt-Winters multivariante propuesto en esta memoria podría verse como un caso particular del modelo introducido por Pfeffermann and Allon (1989), ecuaciones (1.26) - (1.29). En nuestra formulación, la correlación entre las series es introducida únicamente a través de

la matriz de covarianzas del vector ξ_t de los m errores observados en t . La actualización de los términos de nivel, tendencia y estacionalidad de cada una de las series temporales univariantes coincide con la derivada del modelo univariante, pues no depende de los errores asociados a las demás series. Así pues, si comparamos las ecuaciones (1.3), (1.4) y (1.5) con (1.27), (1.28) y (1.29) respectivamente, nuestro modelo sería equivalente al resultante de considerar las matrices A , Γ y Δ diagonales, con elementos en la diagonal principal iguales a $\{\alpha_i\}_{i=1}^m$, $\{\beta_i\}_{i=1}^m$ y $\{\gamma_i\}_{i=1}^m$ respectivamente.

Consecuentemente, nuestra formulación y la propuesta recientemente por de Silva et al. (2007) no están relacionadas. Aunque ambos modelos introducen una única fuente de error y pueden considerarse casos particulares del modelo propuesto por Pfeffermann and Allon (1989), la correlación entre las series es introducida de manera distinta.

2.2. Análisis Bayesiano

En esta sección analizamos el modelo de Holt-Winters multivariante propuesto en la sección anterior empleando un enfoque Bayesiano. Dados los datos observados que conforman las series temporales obtenemos la distribución a posteriori de los parámetros del modelo: condiciones iniciales, parámetros de suavizado y matriz de covarianzas.

La especificación de la distribución a priori es una de las cuestiones más importantes de todo análisis Bayesiano. En problemas multiparamétricos, como es el caso de modelos de regresión, dicha especificación puede resultar difícil. En tales situaciones, así como cuando se dispone de poca información a priori, es común el uso de prioris convencionales no informativas. La distribución a priori no informativa de Jeffreys (ver, por ejemplo, Zellner, 1971, pag. 225) es la priori que generalmente se utiliza en el análisis de modelos Normales multivariantes cuando se desea realizar un análisis Bayesiano objetivo. Dicha distribución, proporcional a la raíz cuadrada del determinante de la matriz de información de Fisher, tiene la propiedad de ser invariante frente a transformaciones y tiende a reproducir resultados clásicos. Bernardo (1979) muestra que, bajo ciertas condiciones, la distribución a priori de Jeffreys es una distribución de referencia óptima en el sentido de que maximiza la información perdida, aunque en problemas multiparamétricos puede resultar inadecuada en la obtención de distribuciones a priori marginales poco informativas. Por otro lado, puede presentar problemas a la hora de

acotar el rango ('shrinkage') de los valores propios de la matriz de covarianzas adecuadamente (Yang and Berger, 1994).

En el análisis que presentamos suponemos independencia entre los distintos elementos del vector de las condiciones iniciales, del vector de los parámetros de suavizado y de la matriz de covarianzas. Como distribuciones a priori para las condiciones iniciales y parámetros de suavizado tomamos la uniforme, mientras que para la matriz de covarianzas consideramos distintas distribuciones a priori: la priori de Jeffreys y, a fin de solventar las posibles deficiencias de dicha distribución, las prioris de referencia propuestas por Chang and Eaves (1990) y Yang and Berger (1994). Las distribuciones a posteriori obtenidas son analíticamente intratables y por tanto, para poder estimarlas, es necesario recurrir a la simulación.

2.2.1. Análisis Bayesiano partiendo de la priori de Jeffreys

Zellner (1971) estudió el modelo de regresión aparentemente no relacionado (modelo SUR) en el contexto Bayesiano, popularizando así la inferencia Bayesiana en Econometría. Sea

$$z = X_B \psi + \varepsilon \quad (2.11)$$

el modelo de regresión aparentemente no relacionado, donde z es el vector $nm \times 1$ de datos observados, X_B es la matriz de regresión, de orden $nm \times km$ y supuesta conocida, ψ es el vector $km \times 1$ de parámetros de regresión y ε el vector $nm \times 1$ de errores, distribuido $N_{nm}(0, \Sigma \otimes I_n)$. Suponiendo poca información a priori acerca de los parámetros del modelo, independencia entre los elementos de ψ y Σ y utilizando la teoría de la invarianza de Jeffreys, Zellner considera la siguiente distribución a priori:

$$\begin{aligned} f(\psi, \Sigma) &= f(\psi)f(\Sigma) \\ &\propto |\Sigma|^{-\frac{m+1}{2}} \end{aligned} \quad (2.12)$$

Entonces, la distribución a posteriori, proporcional al producto de la priori por la verosimilitud, es de la forma:

$$f(\psi, \Sigma|z) \propto |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} (z - X_B \psi)' (\Sigma^{-1} \otimes I_n) (z - X_B \psi) \right\} \quad (2.13)$$

$$\propto |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} H' H] \right\} \quad (2.14)$$

donde $H = [z_1 - X_1 \psi_1, z_2 - X_2 \psi_2, \dots, z_m - X_m \psi_m]$.

La forma cuadrática en (2.13), $FC = (z - X_B\psi)'(\Sigma^{-1} \otimes I_n)(z - X_B\psi)$, puede descomponerse como:

$$\begin{aligned} FC &= z'(\Sigma^{-1} \otimes I_n)z - z'(\Sigma^{-1} \otimes I_n)X_B\psi \\ &\quad - \psi'X_B'(\Sigma^{-1} \otimes I_n)z + \psi'X_B'(\Sigma^{-1} \otimes I_n)X_B\psi \\ &= (z - X_B\hat{\psi})'(\Sigma^{-1} \otimes I_n)(z - X_B\hat{\psi}) + (\psi - \hat{\psi})'X_B'(\Sigma^{-1} \otimes I_n)X_B(\psi - \hat{\psi}) \end{aligned}$$

donde $\hat{\psi} = (X_B'(\Sigma^{-1} \otimes I_n)X_B)^{-1}X_B'(\Sigma^{-1} \otimes I_n)z$ es el estimador mínimos cuadrados generalizado de ψ (Zellner, 1962). Entonces, la distribución a posteriori de los parámetros del modelo dada por (2.13) puede expresarse como:

$$\begin{aligned} f(\psi, \Sigma|z) &\propto |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2}(z - X_B\hat{\psi})'(\Sigma^{-1} \otimes I_n)(z - X_B\hat{\psi}) \right\} \\ &\quad \exp \left\{ -\frac{1}{2}(\psi - \hat{\psi})'X_B'(\Sigma^{-1} \otimes I_n)X_B(\psi - \hat{\psi}) \right\} \end{aligned} \quad (2.15)$$

de donde se deduce que la distribución a posteriori de ψ , dada la matriz Σ , es Normal multivariante, en particular,

$$f(\psi|z, \Sigma) = N_{km}(\hat{\psi}, (X_B'(\Sigma^{-1} \otimes I_n)X_B)^{-1}) \quad (2.16)$$

De (2.14) se deduce que la distribución a posteriori de Σ condicionada a ψ es una Wishart invertida, en concreto,

$$f(\Sigma|z, \psi) = IW(n, H'H) \quad (2.17)$$

En la notación empleada en este trabajo, una matriz G de orden $m \times m$, simétrica y definida positiva, sigue una distribución Wishart invertida si su función de densidad es de la forma:

$$f(G|\nu, A) \propto \frac{|A|^{\frac{\nu}{2}}}{|G|^{\frac{\nu+m+1}{2}}} \exp \left\{ -\frac{1}{2} \text{tr}[G^{-1}A] \right\}$$

donde $\nu \geq m$ y A es una matriz simétrica definida positiva. Una propiedad que se deriva de esta definición es que si la matriz G sigue una distribución Wishart invertida, entonces su matriz inversa G^{-1} sigue la distribución Wishart $W(\nu, A^{-1})$, de manera que $E(G^{-1}) = \nu A^{-1}$.

Por tanto, (2.17) es equivalente a

$$f(\Sigma^{-1}|z, \psi) = W(n, (H'H)^{-1}) \quad (2.18)$$

Integrando la distribución a posteriori respecto a Σ o ψ se obtienen, respectivamente, las marginales a posteriori para ψ y Σ :

$$\begin{aligned} f(\psi|z) &= \int f(\psi, \Sigma|z) d\Sigma \\ &\propto \int |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} H' H] \right\} d\Sigma \\ &\propto |H' H|^{-\frac{n}{2}} \end{aligned} \quad (2.19)$$

$$\begin{aligned} f(\Sigma|z) &= \int f(\psi, \Sigma|z) d\psi \\ &\propto \int |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} (z - X_B \hat{\psi})' (\Sigma^{-1} \otimes I_n) (z - X_B \hat{\psi}) \right\} \\ &\quad \exp \left\{ -\frac{1}{2} (\psi - \hat{\psi})' X_B' (\Sigma^{-1} \otimes I_n) X_B (\psi - \hat{\psi}) \right\} d\psi \\ &\propto |\Sigma|^{-\frac{n+m+1}{2}} |X_B' (\Sigma^{-1} \otimes I_n) X_B|^{-\frac{1}{2}} \\ &\quad \exp \left\{ -\frac{1}{2} (z - X_B \hat{\psi})' (\Sigma^{-1} \otimes I_n) (z - X_B \hat{\psi}) \right\} \end{aligned} \quad (2.20)$$

que no se corresponden con ninguna de las distribuciones conocidas.

Con el fin de obtener inferencias acerca del vector ψ de los parámetros de regresión, Zellner (1971) propuso sustituir Σ por un estimador consistente $\hat{\Sigma}$, en concreto el formado a partir de los residuos obtenidos al estimar las ecuaciones de manera individual mediante mínimos cuadrados, es decir,

$$(n - k) \hat{\Sigma}[i, j] = (z_i - X_i \hat{\psi}_i)' (z_j - X_j \hat{\psi}_j)$$

donde $\hat{\psi}_i = (X_i' X_i)^{-1} X_i' z_i$, y continuar el análisis bajo la suposición $\Sigma = \hat{\Sigma}$.

Posteriormente, la aplicación de técnicas MCMC en la inferencia Bayesiana facilitó el análisis del modelo SUR desde el punto de vista Bayesiano. Percy (1992) calcula la distribución a posteriori para los parámetros del modelo y muestra como, en general, la distribución predictiva de un paso no puede ser evaluada analíticamente, por lo que propone dos posibles aproximaciones. La primera de ellas utiliza el método de Gibbs sampling para generar recursivamente z_{n+1} , ψ y Σ a partir de sus condicionales completas. La otra utiliza una aproximación de primer orden basada en una estimación Bayesiana de la matriz de precisión.

De manera similar, Chib and Greenberg (1995) utilizan el algoritmo de Gibbs para analizar distintas versiones jerárquicas del modelo de regresión aparentemente no relacionado, donde el vector de regresión es proyectado sucesivamente en subespacios de menor dimensión. Esta generalización permite incorporar restricciones cruzadas en las ecuaciones, evaluar la validez de un modelo reducido y analizar el caso en el que los parámetros varían con el tiempo.

Griffiths (2001) muestra el análisis Bayesiano del modelo SUR partiendo de la priori de Jeffreys. En su trabajo utiliza el método de Gibbs sampling para obtener una muestra de la distribución a posteriori de los parámetros del modelo a partir de sus condicionales completas $f(\psi|z, \Sigma)$ y $f(\Sigma|z, \psi)$. Si el número de ecuaciones es grande y sólo se desea obtener inferencias acerca del vector de regresión, Griffiths propone utilizar el método de Gibbs sampling con las condicionales completas $f(\psi_i|z, \psi_1, \dots, \psi_{i-1}, \psi_{i+1}, \dots, \psi_m)$, que se corresponden con t-Student multivariantes, o bien utilizar el algoritmo de Metropolis-Hastings para obtener una muestra de la distribución marginal $f(\psi|z)$.

En nuestra formulación del modelo de Holt-Winters multivariante como un modelo de regresión aparentemente no relacionado, la matriz de regresión X_B depende de los parámetros de suavizado y, por tanto, es desconocida. Si el vector θ de parámetros de suavizado es conocido, entonces X_B es conocida y podemos aplicar los resultados anteriores. Es decir, partiendo de la distribución a priori de Jeffreys

$$f(\psi, \Sigma|\theta) \propto |\Sigma|^{-\frac{m+1}{2}} \quad (2.21)$$

obtenemos las siguientes distribuciones a posteriori condicionales

$$f(\psi|y, \Sigma, \theta) = N_{(s+1)m}(\hat{\psi}, (X_B'(\Sigma^{-1} \otimes I_n)X_B)^{-1}) \quad (2.22)$$

$$f(\Sigma|y, \psi, \theta) = IW(n, H'H) \quad (2.23)$$

o, equivalentemente

$$f(\Sigma^{-1}|y, \psi, \theta) = W(n, (H'H)^{-1}) \quad (2.24)$$

En la práctica, el vector θ de parámetros de suavizado es siempre desconocido. En tal caso, necesitamos también una distribución a priori para θ . Nosotros proponemos tomar la uniforme en la región $[0, 1]^{3m}$ como priori para θ , pues trabajamos con la hipótesis generalmente asumida de que el

valor de los parámetros de suavizado pertenece al intervalo $[0, 1]$, y la priori (2.21) para el resto de parámetros. Es decir,

$$\begin{aligned} f(\psi, \Sigma, \theta) &= f(\psi, \Sigma|\theta)f(\theta) \\ &\propto |\Sigma|^{-\frac{m+1}{2}} \end{aligned} \quad (2.25)$$

Entonces, la distribución a posteriori de los parámetros del modelo es:

$$f(\psi, \Sigma, \theta|y) \propto |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2}(z - X_B\psi)'(\Sigma^{-1} \otimes I_n)(z - X_B\psi) \right\} \quad (2.26)$$

siendo $z = L_B^{-1}y$ el vector de los datos transformados. Dicha distribución a posteriori puede ser expresada alternativamente en la forma (2.14) o (2.15). Las condicionales completas a posteriori que se deducen para ψ y Σ vienen dadas por (2.22) y (2.23) respectivamente, mientras que la condicional completa para θ es de la forma

$$f(\theta|y, \psi, \Sigma) \propto \exp \left\{ -\frac{1}{2}(z - X_B\psi)'(\Sigma^{-1} \otimes I_n)(z - X_B\psi) \right\} \quad (2.27)$$

que no se corresponde con ninguna distribución conocida.

Para poder estimar tanto la distribución a posteriori de los parámetros del modelo como cualquiera de sus características en las que estemos interesados, es necesario recurrir a la simulación y obtener así una muestra $\{(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)})\}_{j=1}^N$ de dicha distribución. Siguiendo la metodología de Griffiths (2001), nos planteamos utilizar el algoritmo de Metropolis-within-Gibbs, que nos permite aplicar el algoritmo de Metropolis-Hastings para simular de la condicional completa $f(\theta|y, \psi, \Sigma)$ que no es analíticamente tratable. En la siguiente sección se describe el algoritmo empleado.

2.2.2. Algoritmo de Metropolis-within-Gibbs para simular de la distribución a posteriori

El algoritmo de Metropolis-within-Gibbs, uno de los procedimientos de simulación más importantes en la modelización econométrica, combina los algoritmos de Gibbs sampling y Metropolis-Hastings para simular de una distribución desconocida mediante el procedimiento de Gibbs aunque algunas de las distribuciones condicionales completas no sean conocidas. Sea $\varphi = (\varphi'_1, \varphi'_2, \dots, \varphi'_s)'$ el vector de parámetros del modelo tal que $f(\varphi_j|\varphi_i (i \neq j))$

es intratable, entonces el algoritmo de Metropolis-within-Gibbs consiste en emplear el mecanismo de Metropolis-Hasting en la simulación del bloque intratable, φ_j , mientras que los otros bloques son tratados en la manera habitual. Para un estudio detallado del algoritmo así como de las condiciones de convergencia, consultar la obra de Robert and Casella (1999) o Geweke and Whiteman (2006).

Dados unos valores iniciales $\psi^{(0)}$ y $\theta^{(0)}$, la iteración i -ésima del algoritmo de Metropolis-within-Gibbs que proponemos se compone de los siguientes pasos:

1. Simular $\Sigma^{(i)}$ de la condicional completa $f(\Sigma|y, \psi^{(i-1)}, \theta^{(i-1)})$, dada por (2.23)
2. Simular $\psi^{(i)}$ a partir de $f(\psi|y, \Sigma^{(i)}, \theta^{(i-1)})$, dada por (2.22)
3. Dado $\theta^{(i-1)} = (\theta_1^{(i-1)}, \theta_2^{(i-1)}, \dots, \theta_m^{(i-1)})$, donde $\theta_j^{(i-1)}$ es el vector simulado de los parámetros de suavizado asociados a la serie temporal j -ésima, simular $\theta^{(i)}$ mediante el siguiente algoritmo de Metropolis:
 - a) Para $j = 1, 2, \dots, m$, simular θ_j^* utilizando como función generadora la uniforme en el cubo centrado en $\theta_j^{(i-1)}$ y lado $2l$. Construir $\theta^* = (\theta_1^*, \theta_2^*, \dots, \theta_m^*)$
 - b) Calcular $\alpha(\theta^{(i-1)}, \theta^*) = \min \left\{ 1, \frac{f(\theta^*|y, \psi^{(i)}, \Sigma^{(i)})}{f(\theta^{(i-1)}|y, \psi^{(i)}, \Sigma^{(i)})} \right\}$ utilizando (2.27)
 - c) Obtener $r \sim Un(0, 1)$
 - d) Si $r \leq \alpha(\theta^{(i-1)}, \theta^*)$, fijar $\theta^{(i)} = \theta^*$; en otro caso, $\theta^{(i)} = \theta^{(i-1)}$

Como valores de inicio $\psi^{(0)}$ y $\theta^{(0)}$, necesarios para la implementación del algoritmo, tomamos:

$$\begin{aligned} \psi_k^{(0)} &= (b_{0k}, c_{1-s,k}, c_{2-s,k}, \dots, c_{0,k})^{(0)} = (0, y_{1k}, y_{2k}, \dots, y_{sk})' \\ \theta_k^{(0)} &= (\alpha_k, \beta_k, \gamma_k)^{(0)} = (0.5, 0.5, 0.5)' \end{aligned}$$

para $k = 1, 2, \dots, m$. Por otro lado, proponemos tomar $l = 0.1$ ya que, como el valor de los parámetros de suavizado pertenece al intervalo $[0, 1]$, éste valor nos permite que la cadena visite todo el soporte de la distribución final.

Tras repetir este proceso $N_0 + N$ veces y eliminar las N_0 primeras iteraciones de calentamiento, obtenemos una muestra $\{(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)})\}_{j=1}^N$ de la distribución a posteriori de los parámetros del modelo que nos permite

estimar cualquier característica de interés.

Notar que, a partir de la muestra obtenida, es posible hallar una muestra de la distribución a posteriori de los coeficientes de correlación y, por tanto, comprobar la hipótesis de correlación contemporánea entre errores correspondientes en los modelos univariantes. Sea $\{\Sigma^{(j)}\}_{j=1}^N$ una muestra de la distribución marginal a posteriori $f(\Sigma|y)$. Para cada $\Sigma^{(j)}$ calcular $\rho_{kl}^{(j)}$, para $k = 1, 2, \dots, m-1$ y $l = k+1, \dots, m$, como:

$$\rho_{kl}^{(j)} = \frac{\Sigma^{(j)}[k, l]}{\sqrt{\Sigma^{(j)}[k, k]\Sigma^{(j)}[l, l]}} \quad (2.28)$$

$\{\rho_{kl}^{(j)}\}_{j=1}^N$ es una muestra de $f(\rho_{kl}|y)$ que nos permite aproximar cualquier característica de interés de la distribución. En particular, podemos estimar su media e intervalos de probabilidad a posteriori.

2.2.3. Análisis Bayesiano de referencia

Dados los problemas que puede presentar la priori de Jeffreys, Chang and Eaves (1990) proponen utilizar, para el análisis de modelos Normales multivariantes donde el parámetro de interés es la matriz de correlaciones, la siguiente distribución a priori de referencia:

$$f(\mu, \Sigma) = \frac{1}{|\Sigma|^{\frac{m+1}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{\frac{1}{2}}} \quad (2.29)$$

donde \circ representa el producto de Hadamard (elemento a elemento).

Cuando los parámetros de interés son los valores propios de la matriz de covarianzas, $d_1 > d_2 > \dots > d_m$, la priori de referencia propuesta por Chang and Eaves adopta la forma:

$$f(\mu, \Sigma) = \frac{1}{|\Sigma| \prod_{i < j} (d_i - d_j)}$$

Posteriormente, y con el objetivo de estimar adecuadamente la estructura de valores y vectores propios de la matriz de covarianzas, Yang and Berger (1994), basándose en el trabajo de Berger and Bernardo (1992), derivan la priori de referencia no informativa para la matriz de covarianzas dada por:

$$f(\Sigma) \propto \frac{1}{|\Sigma| \prod_{i < j} (d_i - d_j)} \quad (2.30)$$

la cual, suponiendo independencia entre μ y Σ y tomando como distribución a priori para el vector de medias la uniforme, coincide con la distribución a priori conjunta propuesta por Chang and Eaves (1990). Comparada con la priori de Jeffreys, la distribución de referencia da considerablemente más peso a las matrices de covarianzas con valores propios iguales y, por tanto, intuitivamente es posible que produzca estimaciones con mejor acotación del rango ('shrinkage') de los valores propios.

Para el análisis Bayesiano de referencia del modelo de Holt-Winters multivariante suponemos, nuevamente, que el vector θ de los parámetros de suavizado, el vector ψ de las condiciones iniciales y la matriz Σ de covarianzas están distribuidos de manera independiente. Como distribución a priori para θ y ψ consideramos la uniforme, mientras que para la matriz Σ , siguiendo el enfoque de Chang and Eaves (1990), proponemos tomar la siguiente distribución:

$$f(\Sigma) = \frac{1}{|\Sigma|^{\frac{m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{\frac{1}{2}}} \quad (2.31)$$

cuya diferencia respecto a (2.29) se debe a que sólo estamos buscando la distribución de referencia para la matriz de covarianzas Σ , pues suponemos independencia entre los parámetros. Por ello, en su cálculo partimos de la distribución a priori $f(\Sigma) = |\Sigma|^{-\frac{m+1}{2}}$ en lugar de considerar la priori conjunta de Jeffreys $f(\psi, \Sigma) = |\Sigma|^{-\frac{m+2}{2}}$. La distribución a priori para los parámetros del modelo, $f(\psi, \Sigma, \theta)$, viene dada por (2.31).

En este caso, la distribución a posteriori obtenida es de la forma:

$$\begin{aligned} f(\psi, \Sigma, \theta|y) &\propto |\Sigma|^{-\frac{n+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} \\ &\quad \exp \left\{ -\frac{1}{2} (z - X_B \psi)' (\Sigma^{-1} \otimes I_n) (z - X_B \psi) \right\} \\ &\propto |\Sigma|^{-\frac{n+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} \\ &\quad \exp \left\{ -\frac{1}{2} (z - X_B \hat{\psi})' (\Sigma^{-1} \otimes I_n) (z - X_B \hat{\psi}) \right\} \\ &\quad \exp \left\{ -\frac{1}{2} (\psi - \hat{\psi})' X_B' (\Sigma^{-1} \otimes I_n) X_B (\psi - \hat{\psi}) \right\} \\ &\propto |\Sigma|^{-\frac{n+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} H' H] \right\} \end{aligned} \quad (2.32)$$

siendo $\hat{\psi} = (X'_B(\Sigma^{-1} \otimes I_n)X_B)^{-1}X'_B(\Sigma^{-1} \otimes I_n)z$ el estimador mínimos cuadrados generalizado de ψ y $H = [z_1 - X_1\psi_1, z_2 - X_2\psi_2, \dots, z_m - X_m\psi_m]$. A partir de (2.32) se concluye que sólo la distribución a posteriori condicional completa para ψ se corresponde con una densidad conocida, la Normal multivariante dada por:

$$f(\psi|y, \Sigma, \theta) = N_{(s+1)m}(\hat{\psi}, (X'_B(\Sigma^{-1} \otimes I_n)X_B)^{-1}) \quad (2.33)$$

mientras que para Σ y θ se obtienen las siguientes condicionales completas:

$$f(\Sigma|y, \psi, \theta) \propto |\Sigma|^{-\frac{n+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} H' H] \right\} \quad (2.34)$$

$$f(\theta|y, \psi, \Sigma) \propto \exp \left\{ -\frac{1}{2} (z - X_B\psi)' (\Sigma^{-1} \otimes I_n) (z - X_B\psi) \right\} \quad (2.35)$$

Como puede observarse, las condicionales para ψ y θ , ecuaciones (2.33) y (2.35), coinciden con las obtenidas en el análisis Bayesiano cuando se parte de la priori de Jeffreys, ecuaciones (2.22) y (2.27) respectivamente.

Con el fin de obtener inferencias acerca de los parámetros del modelo es necesario nuevamente recurrir a la simulación. El mecanismo de simulación que proponemos en este caso para obtener una muestra de la distribución a posteriori, ecuación (2.32), es una modificación del algoritmo de Metropolis-within-Gibbs desarrollado anteriormente (sección 2.2.2). En concreto, en el paso 1 empleamos un método de aceptación-rechazo para simular de la matriz de covarianzas Σ , pues su distribución condicional completa, ecuación (2.34), deja de ser conocida. El método de aceptación-rechazo que utilizamos para simular de $f(\Sigma|y, \psi, \theta)$ es similar al propuesto por Sun and Berger (2006):

1. Simular $\Sigma^* \sim IW(n-1, H'H) = g(\Sigma)$
2. Simular $u \sim Un(0, 1)$. Si $u \leq 2^{\frac{m}{2}} |I_m + \Sigma^* \circ (\Sigma^*)^{-1}|^{-\frac{1}{2}}$, fijar $\Sigma^{(i)} = \Sigma^*$. En otro caso, volver al paso 1

Para la demostración de que éste es un método de aceptación-rechazo válido para la simulación de $f(\Sigma|y, \psi, \theta)$ podemos hacer uso del siguiente resultado (Horn and Johnson, 1994, pag. 313):

Resultado 2.1 Sean A y B dos matrices simétricas y semidefinidas positivas, entonces

$$\lambda_{\min}(A \circ B) \geq \lambda_{\min}(AB)$$

donde $\lambda_{\min}(A)$ representa el menor valor propio de la matriz A

En nuestro caso, tenemos

$$\lambda_{\min}(\Sigma \circ \Sigma^{-1}) \geq \lambda_{\min}(\Sigma \Sigma^{-1}) = 1$$

por lo que, si $\lambda_1 > \lambda_2 > \dots > \lambda_m$ representan los valores propios de la matriz $\Sigma \circ \Sigma^{-1}$, se deduce que:

$$|I_m + \Sigma \circ \Sigma^{-1}| = \prod_{i=1}^m (1 + \lambda_i) \geq \prod_{i=1}^m (1 + \lambda_m) = (1 + \lambda_m)^m \geq 2^m$$

y por tanto

$$\frac{f(\Sigma|y, \psi, \theta)}{g(\Sigma)} = \frac{1}{|I_m + \Sigma \circ \Sigma^{-1}|^{\frac{1}{2}}} \leq \frac{1}{2^{\frac{m}{2}}} < \frac{1}{M}$$

Entonces, tomando $M = 2^{-\frac{m}{2}}$ se cumple que $f(\Sigma|y, \psi, \theta) \leq M g(\Sigma)$ y la condición de aceptación del método de aceptación-rechazo, $u M g(\Sigma) \leq f(\Sigma|y, \psi, \theta)$, es la utilizada en el algoritmo.

Si en lugar de la priori de referencia para la matriz de covarianzas dada por (2.31) partimos de la priori propuesta por Yang and Berger (1994), ecuación (2.30), la distribución a posteriori de los parámetros del modelo es de la forma:

$$\begin{aligned} f(\psi, \Sigma, \theta|y) &\propto |\Sigma|^{-\frac{n+2}{2}} \left(\prod_{i < j} (d_i - d_j) \right)^{-1} \\ &\quad \exp \left\{ -\frac{1}{2} (z - X_B \psi)' (\Sigma^{-1} \otimes I_n) (z - X_B \psi) \right\} \\ &\propto |\Sigma|^{-\frac{n+2}{2}} \left(\prod_{i < j} (d_i - d_j) \right)^{-1} \\ &\quad \exp \left\{ -\frac{1}{2} (z - X_B \hat{\psi})' (\Sigma^{-1} \otimes I_n) (z - X_B \hat{\psi}) \right\} \\ &\quad \exp \left\{ -\frac{1}{2} (\psi - \hat{\psi})' X_B' (\Sigma^{-1} \otimes I_n) X_B (\psi - \hat{\psi}) \right\} \\ &\propto |\Sigma|^{-\frac{n+2}{2}} \left(\prod_{i < j} (d_i - d_j) \right)^{-1} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} H' H] \right\} \end{aligned} \tag{2.36}$$

Las condicionales completas para el vector ψ de las condiciones iniciales y el vector θ de los parámetros de suavizado coinciden con (2.33) y (2.35) respectivamente, mientras que la condicional completa para la matriz de covarianzas es de la forma:

$$f(\Sigma|y, \psi, \theta) \propto |\Sigma|^{-\frac{n+2}{2}} \left(\prod_{i < j} (d_i - d_j) \right)^{-1} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} H' H] \right\} \quad (2.37)$$

El mecanismo de Metropolis-within-Gibbs que proponemos para simular de la distribución a posteriori de los parámetros de modelo en este caso, ecuación (2.36), coincide nuevamente con el descrito en la sección 2.2.2 excepto el primer paso, donde para simular de $f(\Sigma|y, \psi^{(i-1)}, \theta^{(i-1)})$ hacemos uso del siguiente algoritmo de Metropolis-Hastings:

1. Simular $\Sigma^* \sim IW(n - m + 1, H' H) = q(\Sigma|\Sigma^{(i-1)}, \psi^{(i-1)}, \theta^{(i-1)})$

2. Calcular

$$\begin{aligned} \alpha(\Sigma^{(i-1)}, \Sigma^*) &= \min \left\{ 1, \frac{q(\Sigma^{(i-1)}|\Sigma^*, \psi^{(i-1)}, \theta^{(i-1)}) f(\Sigma^*|y, \psi^{(i-1)}, \theta^{(i-1)})}{q(\Sigma^*|\Sigma^{(i-1)}, \psi^{(i-1)}, \theta^{(i-1)}) f(\Sigma^{(i-1)}|y, \psi^{(i-1)}, \theta^{(i-1)})} \right\} \\ &= \min \left\{ 1, \frac{\prod_{i < j} (d_i - d_j)}{\prod_{i < j} (d_i^* - d_j^*)} \right\} \end{aligned}$$

siendo $\{d_i\}_{i=1}^m$ los valores propios de $\Sigma^{(i-1)}$ y $\{d_i^*\}_{i=1}^m$ los de Σ^*

3. Obtener $r \sim Un(0, 1)$

4. Si $r \leq \alpha(\Sigma^{(i-1)}, \Sigma^*)$, fijar $\Sigma^{(i)} = \Sigma^*$; en otro caso, $\Sigma^{(i)} = \Sigma^{(i-1)}$

Otros análisis Bayesianos del modelo SUR donde la matriz de regresión no es conocida los encontramos en los trabajos de Smith and Kohn (2000) y Holmes et al. (2002). En ambos casos, las funciones de regresión que relacionan los datos que conforman cada una de las series temporales univariantes con las variables explicativas no son necesariamente lineales. Expresando cada una de estas funciones de regresión como una combinación lineal de funciones base, se obtiene un modelo SUR donde la matriz de regresión X_B es considerada a su vez como una variable aleatoria. Smith and Kohn (2000), en su análisis Bayesiano, simulan variables binarias indicadoras asociadas a los coeficientes de regresión con el fin de determinar las funciones base redundantes. Holmes et al. (2002) parten de una distribución a priori sobre la matriz de regresión y utilizan algoritmos MCMC de dimensión variable, que permiten variar la dimensión de la matriz de regresión, para estimar

finalmente las cantidades de interés. Notar que en nuestra formulación del modelo de Holt-Winters multivariante como un modelo de regresión aparentemente no relacionado, la matriz de regresión es desconocida pero de dimensión fija, por lo que estos métodos no son válidos para su análisis.

2.2.4. Ejemplo de ocupación hotelera (continuación)

En la sección 1.2 mostramos los resultados obtenidos en el análisis de las series temporales de ocupación hotelera en Castellón, Valencia y Alicante cuando éstas son analizadas de manera independiente mediante el modelo de Holt-Winters univariante. A continuación analizamos las series conjuntamente mediante el modelo de Holt-Winters multivariante.

Como distribuciones a priori para los parámetros del modelo consideramos la priori de Jeffreys, ecuación (2.25), y la priori de referencia dada por (2.31). Para la simulación de muestras de las distribuciones a posteriori resultantes, ecuaciones (2.26) y (2.32) respectivamente, empleamos los algoritmos de Metropolis-within-Gibbs descritos en este capítulo. Con el fin de acelerar su convergencia repetimos, en cada iteración del algoritmo, 20 veces el paso 3, correspondiente al algoritmo de Metropolis para simular θ de su condicional completa $f(\theta|y, \psi, \Sigma)$. En concreto simulamos $N_0 + N = 55000$ valores. Entonces, tras eliminar las 5000 primeras simulaciones de calentamiento, obtenemos muestras de tamaño $N = 50000$ de las distribuciones a posteriori que nos permiten estimar cualquiera de sus características. Cuando partimos de la priori de referencia observamos que la muestra obtenida no converge, siendo necesario considerar las 40000 primeras simulaciones de calentamiento. Por ello, para esta distribución, simulamos $N_0 + N = 90000$ valores. La media e intervalos de confianza o de probabilidad a posteriori del 95 % estimados, a partir de las correspondientes muestras obtenidas, para los parámetros del modelo pueden verse en las tablas 2.1, 2.2 y 2.3. La tabla 2.4 muestra las estimaciones de la media y del intervalo de confianza del 95 % de las distribuciones a posteriori de los coeficientes de correlación.

Dado que las condiciones iniciales en $\psi_i = (b_{0i}, c_{1-s,i}, c_{2-s,i}, \dots, c_{0i})'$, para $i = 1, 2, \dots, m$, pueden resultar difíciles de interpretar, los resultados mostrados en la tabla 2.1 hacen referencia a la parametrización habitual $\omega_i = (\tilde{a}_{0i}, \tilde{b}_{0i}, \tilde{c}_{1-s,i}, \tilde{c}_{2-s,i}, \dots, \tilde{c}_{0i})'$ donde $\tilde{c}_{1-s,i} + \tilde{c}_{2-s,i} + \dots + \tilde{c}_{0i} = 0$. Notar que, dado $\psi_i^{(j)}$, un valor simulado de la distribución a posteriori $f(\psi_i|y)$, el

vector $\omega_i^{(j)}$ calculado como:

$$\begin{aligned}\tilde{a}_{0i}^{(j)} &= \frac{1}{s} \sum_{k=1}^s c_{k-s,i}^{(j)} - b_{0i}^{(j)} \\ \tilde{b}_{0i}^{(j)} &= b_{0i}^{(j)} \\ \tilde{c}_{r-s,i}^{(j)} &= c_{r-s,i}^{(j)} - \frac{1}{s} \sum_{k=1}^s c_{k-s,i}^{(j)} \text{ para } r = 1, 2, \dots, s\end{aligned}$$

se distribuye según $f(\omega_i|y)$.

Como puede observarse, las medias y los intervalos de probabilidad a posteriori del 95 % obtenidos para los parámetros del modelo son prácticamente iguales en ambos casos. Ésto lleva a pensar que, para el análisis multivariante de un conjunto moderado de series temporales correlacionadas, el uso de la distribución a priori de Jeffreys o de la priori de referencia no influye en los resultados obtenidos. Por otro lado, el tiempo de cómputo de la posteriori de referencia es, aproximadamente, un 80 % mayor que el de la posteriori de Jeffreys. Notar que, en general, el tiempo de computación necesario para obtener una muestra de tamaño N de la distribución a posteriori de referencia será mayor que el requerido para obtener una muestra del mismo tamaño de la distribución a posteriori de Jeffreys pues, como se muestra en los algoritmos de Metropolis-within-Gibbs propuestos, en el caso de la posteriori de referencia la matriz de covarianzas Σ es simulada mediante un procedimiento de aceptación-rechazo que, en general, será más lento que cuando la matriz es simulada a partir de su distribución a posteriori condicional completa cuando ésta es conocida. Además, la priori de referencia requiere de un mayor calentamiento. Consecuentemente, el uso de la distribución a priori de Jeffreys puede ser aconsejado en problemas multivariantes de dimensión reducida.

Del mismo modo, parece razonable no emplear la distribución a priori de referencia dada por (2.30). Notar que para obtener una muestra de la distribución a posteriori resultante se utiliza un algoritmo de Metropolis-within-Gibbs donde sólo se conoce la condicional completa para el vector de las condiciones iniciales ψ , siendo necesario recurrir al procedimiento de Metropolis-Hastings en la simulación del vector θ de los parámetros de suavizado y de la matriz de covarianzas Σ . Ésto implica que el algoritmo, además de requerir de un tiempo de computación mayor, presente mayores problemas de convergencia.

	Priori Jeffreys			Priori referencia		
	$q_{0.025}$	media	$q_{0.975}$	$q_{0.025}$	media	$q_{0.975}$
\tilde{a}_{01}	41.15	48.73	61.37	40.99	48.08	60.34
\tilde{b}_{01}	-1.50	0.12	1.28	-1.38	0.15	1.23
$\tilde{c}_{-11,1}$	-38.91	-28.22	-14.69	-39.05	-27.81	-14.15
$\tilde{c}_{-10,1}$	-26.83	-15.98	-0.02	-26.99	-15.44	0.35
$\tilde{c}_{-9,1}$	-15.11	-4.54	9.81	-15.39	-4.18	10.23
$\tilde{c}_{-8,1}$	0.38	10.03	22.15	0.13	10.51	22.99
$\tilde{c}_{-7,1}$	-0.02	10.35	23.76	-0.47	10.62	24.31
$\tilde{c}_{-6,1}$	0.13	12.08	21.99	-0.74	11.82	22.43
$\tilde{c}_{-5,1}$	5.99	18.49	28.51	5.46	18.25	29.04
$\tilde{c}_{-4,1}$	23.67	36.89	46.72	23.09	36.43	46.87
$\tilde{c}_{-3,1}$	-1.86	13.17	23.67	-2.56	12.59	23.87
$\tilde{c}_{-2,1}$	-12.32	-1.31	9.11	-12.80	-1.41	9.65
$\tilde{c}_{-1,1}$	-37.51	-23.76	-13.72	-37.77	-24.06	-13.39
$\tilde{c}_{0,1}$	-39.50	-27.19	-16.05	-39.98	-27.30	-15.69
\tilde{a}_{02}	19.18	24.16	27.82	19.07	23.99	27.72
\tilde{b}_{02}	-0.93	-0.06	0.92	-0.97	-0.07	0.93
$\tilde{c}_{-11,2}$	-19.39	-13.42	-8.26	-19.57	-13.34	-8.09
$\tilde{c}_{-10,2}$	-12.04	-6.56	-0.97	-12.05	-6.39	-0.91
$\tilde{c}_{-9,2}$	-7.33	-1.66	3.35	-7.43	-1.56	3.43
$\tilde{c}_{-8,2}$	-4.49	0.51	5.34	-4.34	0.70	5.62
$\tilde{c}_{-7,2}$	-6.10	-0.94	3.77	-6.19	-0.85	3.92
$\tilde{c}_{-6,2}$	1.56	6.73	11.37	1.32	6.61	11.23
$\tilde{c}_{-5,2}$	6.83	11.81	17.25	6.70	11.71	17.39
$\tilde{c}_{-4,2}$	10.31	15.31	20.86	10.11	15.10	20.84
$\tilde{c}_{-3,2}$	2.18	7.61	13.74	2.01	7.44	13.80
$\tilde{c}_{-2,2}$	-9.43	-4.51	0.14	-9.38	-4.45	0.28
$\tilde{c}_{-1,2}$	-10.14	-5.39	-0.13	-10.21	-5.47	-0.11
$\tilde{c}_{0,2}$	-14.49	-9.49	-4.62	-14.47	-9.50	-4.60
\tilde{a}_{03}	178.46	206.12	236.11	177.86	205.18	234.66
\tilde{b}_{03}	-5.35	0.46	5.94	-5.46	0.46	5.96
$\tilde{c}_{-11,3}$	-103.78	-81.94	-51.74	-103.83	-80.98	-50.38
$\tilde{c}_{-10,3}$	-81.37	-57.11	-33.15	-81.36	-56.67	-32.11
$\tilde{c}_{-9,3}$	-35.64	-10.62	12.49	-35.62	-10.17	13.85
$\tilde{c}_{-8,3}$	-3.49	18.49	44.39	-3.34	19.53	45.99
$\tilde{c}_{-7,3}$	-4.66	23.29	46.17	-5.11	23.51	47.03
$\tilde{c}_{-6,3}$	20.75	44.26	65.54	19.77	44.06	66.31
$\tilde{c}_{-5,3}$	42.09	65.82	87.21	40.65	65.49	87.55
$\tilde{c}_{-4,3}$	59.74	87.24	108.08	58.53	86.34	108.37
$\tilde{c}_{-3,3}$	16.15	42.24	65.44	14.39	41.17	65.27
$\tilde{c}_{-2,3}$	-37.19	-12.56	8.89	-38.64	-13.14	9.32
$\tilde{c}_{-1,3}$	-78.69	-55.35	-30.21	-79.37	-55.68	-29.65
$\tilde{c}_{0,3}$	-85.59	-63.76	-33.89	-86.40	-63.48	-33.35

Tabla 2.1: Estimaciones de la media y del intervalo de probabilidad a posteriori del 95% para las condiciones iniciales del modelo de Holt-Winters multivariante

	Priori Jeffreys			Priori referencia		
	$q_{0.025}$	media	$q_{0.975}$	$q_{0.025}$	media	$q_{0.975}$
α_1	0.01	0.17	0.55	0.01	0.15	0.53
β_1	0.01	0.23	0.92	0.01	0.21	0.84
γ_1	0.02	0.43	0.86	0.03	0.47	0.87
α_2	0.06	0.21	0.50	0.06	0.23	0.51
β_2	0.00	0.28	0.94	0.01	0.25	0.91
γ_2	0.07	0.58	0.96	0.08	0.58	0.97
α_3	0.18	0.40	0.66	0.18	0.40	0.65
β_3	0.00	0.05	0.18	0.00	0.05	0.19
γ_3	0.01	0.28	0.64	0.02	0.31	0.64

Tabla 2.2: Estimaciones de la media y del intervalo de probabilidad a posteriori del 95 % para los parámetros de suavizado del modelo de Holt-Winters multivariante

	Priori Jeffreys			Priori referencia		
	$q_{0.025}$	media	$q_{0.975}$	$q_{0.025}$	media	$q_{0.975}$
σ_{11}	33.93	56.19	89.82	35.34	57.78	91.78
σ_{12}	5.08	13.51	24.84	4.74	13.21	24.34
σ_{13}	30.50	78.01	145.76	31.55	79.70	147.29
σ_{22}	6.39	10.61	16.70	6.48	10.74	16.90
σ_{23}	-7.64	13.18	37.72	-8.59	12.75	37.86
σ_{33}	223.55	362.66	576.38	230.32	373.15	590.37

Tabla 2.3: Estimaciones de la media y del intervalo de probabilidad a posteriori del 95 % para los distintos elementos de la matriz de covarianzas del modelo de Holt-Winters multivariante

	Priori Jeffreys			Priori referencia		
	$q_{0.025}$	media	$q_{0.975}$	$q_{0.025}$	media	$q_{0.975}$
ρ_{12}	0.25	0.55	0.78	0.23	0.53	0.76
ρ_{13}	0.27	0.54	0.76	0.26	0.54	0.75
ρ_{23}	-0.13	0.21	0.52	-0.14	0.20	0.51

Tabla 2.4: Estimaciones de la media y del intervalo de probabilidad a posteriori del 95 % para los coeficientes de correlación del modelo de Holt-Winters multivariante

2.3. Distribución predictiva

Una vez calculada la distribución a posteriori de los parámetros del modelo nos planteamos obtener la distribución predictiva, la cual, desde el

punto de vista Bayesiano, contiene toda la información acerca de los valores futuros de las series temporales. En particular, nos permite obtener estimaciones puntuales, como por ejemplo la media o la mediana de la distribución, e intervalos de predicción, calculados a partir de los correspondientes cuantiles de la distribución.

Sea P_i el vector $h \times 1$ de los datos futuros de la serie temporal i -ésima. Suponemos que el vector $(n + h) \times 1$ de los datos observados y futuros de la serie temporal i -ésima, $(y'_i, P'_i)'$, se ajusta al modelo de Holt-Winters univariante y por tanto satisface la ecuación (2.1)

$$\begin{bmatrix} y_i \\ P_i \end{bmatrix} = \begin{bmatrix} M \\ M_p \end{bmatrix} \psi_i + \begin{bmatrix} L_i^1 & 0 \\ L_i^{21} & L_i^2 \end{bmatrix} \begin{bmatrix} \varepsilon_i \\ v_i \end{bmatrix} \quad (2.38)$$

donde las matrices M y L_i en (2.1) han sido completadas hasta orden $n + h$ y particionadas de manera análoga al vector $(y'_i, P'_i)'$. Es decir, L_i^1 es la submatriz de L_i de orden $n \times n$ correspondiente a las n primeras filas y columnas, L_i^{21} es la submatriz $h \times n$ asociada a las últimas h filas y primeras n columnas y L_i^2 está formada por las últimas h filas y columnas de L_i . De nuestra definición de L_i como matriz triangular inferior se deriva que el bloque L_i^{12} es la submatriz $n \times h$ nula. v_i es el vector de los errores asociados a los datos futuros, independientes entre sí y distribuidos según $N(0, \sigma_i^2)$.

Entonces, el modelo multivariante resultante de formular conjuntamente las m series temporales es de la forma:

$$\begin{bmatrix} y_1 \\ P_1 \\ y_2 \\ P_2 \\ \vdots \\ y_m \\ P_m \end{bmatrix} = \begin{bmatrix} M & 0 & \dots & 0 \\ M_p & 0 & \dots & 0 \\ 0 & M & \dots & 0 \\ 0 & M_p & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M \\ 0 & 0 & \dots & M_p \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix} + \begin{bmatrix} L_1^1 & 0 & 0 & 0 & \dots & 0 & 0 \\ L_1^{21} & L_1^2 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & L_2^1 & 0 & \dots & 0 & 0 \\ 0 & 0 & L_2^{21} & L_2^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & L_m^1 & 0 \\ 0 & 0 & 0 & 0 & \dots & L_m^{21} & L_m^2 \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ v_1 \\ \varepsilon_2 \\ v_2 \\ \vdots \\ \varepsilon_m \\ v_m \end{bmatrix} \quad (2.39)$$

donde el vector de errores está distribuido según $N_{(n+h)m}(0, \Sigma \otimes I_{n+h})$.

Sea $y = (y'_1, y'_2, \dots, y'_m)'$ el vector $nm \times 1$ de los datos observados y $P = (P'_1, P'_2, \dots, P'_m)'$ el vector $hm \times 1$ de los valores futuros de las m series temporales. Entonces, la distribución predictiva a posteriori, $f(P|y)$, se obtiene al resolver la integral:

$$f(P|y) = \int \int \int f(P|y, \psi, \Sigma, \theta) f(\psi, \Sigma, \theta|y) d\psi d\Sigma d\theta$$

donde $f(\psi, \Sigma, \theta|y)$ es la distribución a posteriori de los parámetros del modelo, obtenida en el apartado anterior. Por tanto, para obtener la distribución predictiva es necesario el cálculo de $f(P|y, \psi, \Sigma, \theta)$, es decir, la distribución del vector de los datos futuros dados los datos observados y los parámetros del modelo.

Considerando en (2.39) únicamente el vector formado por los datos observados se obtiene la expresión matricial:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} M & 0 & \dots & 0 \\ 0 & M & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix} + \begin{bmatrix} L_1^1 & 0 & \dots & 0 \\ 0 & L_2^1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L_m^1 \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

de donde se deduce:

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} - \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_m \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix} \quad (2.40)$$

siendo $z_i = (L_i^1)^{-1}y_i$ y $X_i = (L_i^1)^{-1}M$, para $i = 1, 2, \dots, m$.

Por otro lado, si en (2.39) consideramos el vector de las observaciones futuras se tiene:

$$\begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{bmatrix} = \begin{bmatrix} M_p & 0 & \dots & 0 \\ 0 & M_p & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_p \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix} + \begin{bmatrix} L_1^{21} & 0 & \dots & 0 \\ 0 & L_2^{21} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L_m^{21} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} \\ + \begin{bmatrix} L_1^2 & 0 & \dots & 0 \\ 0 & L_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L_m^2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix} \quad (2.41)$$

Sustituyendo (2.40) en (2.41), se obtiene finalmente que el vector de los datos futuros puede ser representado matricialmente como:

$$\begin{aligned}
 \underbrace{\begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{bmatrix}}_P &= \underbrace{\begin{bmatrix} L_1^{21} z_1 \\ L_2^{21} z_2 \\ \vdots \\ L_m^{21} z_m \end{bmatrix}}_{\tilde{z}} + \underbrace{\begin{bmatrix} M_p - L_1^{21} X_1 & 0 & \cdots & 0 \\ 0 & M_p - L_2^{21} X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_p - L_m^{21} X_m \end{bmatrix}}_D \underbrace{\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix}}_\psi \\
 &+ \underbrace{\begin{bmatrix} L_1^2 & 0 & \cdots & 0 \\ 0 & L_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & L_m^2 \end{bmatrix}}_{L_B^2} \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}}_v
 \end{aligned} \tag{2.42}$$

donde el vector de errores v está distribuido según $N_{hm}(0, \Sigma \otimes I_h)$. Por tanto, dados los datos observados que conforman las series temporales y los parámetros del modelo, la distribución del vector P de los datos futuros es Normal multivariante, en particular,

$$f(P|y, \psi, \Sigma, \theta) = N_{hm}(\tilde{z} + D\psi, L_B^2(\Sigma \otimes I_h)(L_B^2)')$$

Entonces, la distribución predictiva resultante de considerar la distribución a posteriori de los parámetros del modelo obtenida a partir de la priori de Jeffreys, ecuación (2.26), viene dada por:

$$\begin{aligned}
 f(P|y) &\propto \int \int \int |\Sigma|^{-\frac{n+h+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} H' H] \right\} \\
 &\exp \left\{ -\frac{1}{2} (P - \tilde{z} - D\psi)' (L_B^2)'^{-1} (\Sigma^{-1} \otimes I_h) (L_B^2)^{-1} (P - \tilde{z} - D\psi) \right\} d\psi d\Sigma d\theta \\
 &\propto \int \int \int |\Sigma|^{-\frac{n+h+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} H' H] \right\} \\
 &\exp \left\{ -\frac{1}{2} \text{tr} \left[\Sigma^{-1} (\tilde{P} - \tilde{\mu}_P)' (\tilde{P} - \tilde{\mu}_P) \right] \right\} d\psi d\Sigma d\theta
 \end{aligned}$$

donde $H = [z_1 - X_1\psi_1, z_2 - X_2\psi_2, \dots, z_m - X_m\psi_m]$ y \tilde{P} y $\tilde{\mu}_P$ son las matrices $h \times m$ definidas como:

$$\begin{aligned}
 \tilde{P} &= [(L_1^2)^{-1} P_1, (L_2^2)^{-1} P_2, \dots, (L_m^2)^{-1} P_m] \\
 \tilde{\mu}_P &= [(L_1^2)^{-1} (\tilde{z}_1 + (M_p - L_1^{21} X_1)\psi_1), \dots, (L_m^2)^{-1} (\tilde{z}_m + (M_p - L_m^{21} X_m)\psi_m)]
 \end{aligned}$$

De este modo,

$$\begin{aligned} f(P|y) &\propto \int \int \int |\Sigma|^{-\frac{n+h+m+1}{2}} \\ &\quad \exp \left\{ -\frac{1}{2} \text{tr} \left[\Sigma^{-1} \left(H'H + (\tilde{P} - \tilde{\mu}_P)'(\tilde{P} - \tilde{\mu}_P) \right) \right] \right\} d\Sigma d\psi d\theta \\ &\propto \int \int |H'H + (\tilde{P} - \tilde{\mu}_P)'(\tilde{P} - \tilde{\mu}_P)|^{-\frac{n+h}{2}} d\psi d\theta \end{aligned}$$

que no tiene solución analítica. Entonces,

$$f(P|y, \psi, \theta) \propto |H'H + (\tilde{P} - \tilde{\mu}_P)'(\tilde{P} - \tilde{\mu}_P)|^{-\frac{n+h}{2}} \quad (2.43)$$

de donde se deduce que la matriz \tilde{P} , dados los datos observados, ψ y θ , sigue una distribución t-Student generalizada, en concreto

$$f(\tilde{P}|y, \psi, \theta) = GSt(I_h, H'H, \tilde{\mu}_P, n+h) \quad (2.44)$$

En la notación utilizada en este trabajo, decimos que la matriz T de orden $n \times m$ está distribuida según la t-Student generalizada $GSt(A, B, t, \nu)$ si y sólo si su función de densidad es de la forma:

$$f(T|A, B, t, \nu) \propto \frac{|B|^{\frac{\nu-n}{2}} |A|^{\frac{m}{2}}}{|B + (T-t)'A(T-t)|^{\frac{\nu}{2}}}$$

donde t representa la media de la distribución. De su definición se deriva la siguiente propiedad (ver, por ejemplo, Zellner, 1971, pag. 397):

Si la matriz T está distribuida según $GSt(A, B, t, \nu)$, entonces la distribución marginal de cada una de sus columnas es de la forma:

$$f(T[, i]) \propto \left[1 + (T[, i] - t[, i])' \frac{A}{B[i, i]} (T[, i] - t[, i]) \right]^{-\frac{\nu-(m-1)}{2}}$$

es decir, la t-Student multivariante $MSt(\mu = t[, i], H = \frac{A}{B[i, i]}, d = \nu - n - m + 1)$ donde μ representa el vector de medias. Si $d > 2$, entonces la matriz de covarianzas es $\frac{H^{-1}}{d-2}$.

La matriz $(\tilde{P} - \tilde{\mu}_P)'(\tilde{P} - \tilde{\mu}_P)$ es de la forma:

$$\begin{bmatrix} \hat{P}'_1(L_1^2)^{-1}(L_1^2)^{-1}\hat{P}_1 & \hat{P}'_1(L_1^2)^{-1}(L_2^2)^{-1}\hat{P}_2 & \dots & \hat{P}'_1(L_1^2)^{-1}(L_m^2)^{-1}\hat{P}_m \\ \hat{P}'_2(L_2^2)^{-1}(L_1^2)^{-1}\hat{P}_1 & \hat{P}'_2(L_2^2)^{-1}(L_2^2)^{-1}\hat{P}_2 & \dots & \hat{P}'_2(L_2^2)^{-1}(L_m^2)^{-1}\hat{P}_m \\ \vdots & \vdots & \ddots & \vdots \\ \hat{P}'_m(L_m^2)^{-1}(L_1^2)^{-1}\hat{P}_1 & \hat{P}'_m(L_m^2)^{-1}(L_2^2)^{-1}\hat{P}_2 & \dots & \hat{P}'_m(L_m^2)^{-1}(L_m^2)^{-1}\hat{P}_m \end{bmatrix}$$

donde $\widehat{P}_i = P_i - \widetilde{z}_i - (M_p - L_i^{21} X_i)\psi_i$, para $i = 1, 2, \dots, m$. Debido a que las matrices L_i^2 asociadas a cada una de las series temporales univariantes son distintas, a partir de la distribución t-Student generalizada obtenida para la matriz \widetilde{P} no es posible obtener la distribución análoga para la matriz de los datos futuros $[P_1, P_2, \dots, P_m]$. Sin embargo, aplicando la propiedad de la t-Student generalizada vista anteriormente, se tiene que la distribución marginal de cada una de las columnas de la matriz \widetilde{P} es de la forma:

$$\begin{aligned} f(\widetilde{P}[i]|y, \psi, \theta) &\propto \left[1 + (\widetilde{P}[i] - \widetilde{\mu}_P[i])' \frac{I_h}{H'H[i, i]} (\widetilde{P}[i] - \widetilde{\mu}_P[i]) \right]^{-\frac{n+h-m+1}{2}} \\ &\propto \left[1 + (P_i - \mu_{P_i})' \frac{(L_i^2)'^{-1}(L_i^2)^{-1}}{H'H[i, i]} (P_i - \mu_{P_i}) \right]^{-\frac{n+h-m+1}{2}} \end{aligned}$$

siendo $\mu_{P_i} = \widetilde{z}_i + (M_p - L_i^{21} X_i)\psi_i$. Entonces, la distribución del vector de los datos futuros de la serie temporal i -ésima, dados los datos observados, ψ y θ , es la t-Student multivariante:

$$f(P_i|y, \psi, \theta) = MSt \left(\mu_{P_i}, \frac{(L_i^2)'^{-1}(L_i^2)^{-1}}{H'H[i, i]}, \nu = n - m + 1 \right) \quad (2.45)$$

con media y varianza:

$$E(P_i|y, \psi, \theta) = \mu_{P_i} = L_i^{21} z_i + (M_p - L_i^{21} X_i)\psi_i \quad (2.46)$$

$$V(P_i|y, \psi, \theta) = V_{P_i} = \frac{H'H[i, i]}{\nu - 2} L_i^2 (L_i^2)' \quad (2.47)$$

Con el objetivo de estimar la distribución predictiva, pues no puede ser calculada de manera analítica, recurrimos a la integración por Monte Carlo. Sea $\{(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)})\}_{j=1}^N$ una muestra de la distribución a posteriori de los parámetros del modelo $f(\psi, \Sigma, \theta|y)$, entonces la distribución predictiva puede ser estimada como:

$$\begin{aligned} f(P_i|y) &\approx \frac{1}{N} \sum_{j=1}^N f(P_i|y, \psi^{(j)}, \theta^{(j)}) \\ &\approx \frac{1}{N} \sum_{j=1}^N MSt^{(j)} \left(\mu_{P_i}, \frac{(L_i^2)'^{-1}(L_i^2)^{-1}}{H'H[i, i]}, n - m + 1 \right) \end{aligned} \quad (2.48)$$

para $i = 1, 2, \dots, m$. El superíndice (j) en $MSt^{(j)}$ indica que los parámetros de las distribuciones han sido calculados en el punto $(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)})$. Del

mismo modo, los principales momentos de la predictiva son estimados como:

$$E(P_i|y) \approx \frac{1}{N} \sum_{j=1}^N \mu_{P_i}^{(j)} \quad (2.49)$$

$$V(P_i|y) \approx \frac{1}{N} \sum_{j=1}^N V_{P_i}^{(j)} + \frac{1}{N} \sum_{j=1}^N \mu_{P_i}^{(j)} (\mu_{P_i}^{(j)})' - E(P_i|y)E'(P_i|y) \quad (2.50)$$

Análogamente, los intervalos de predicción pueden ser estimados a partir de los correspondientes cuantiles de la mixtura de distribuciones t-Student. Sin embargo, estos cuantiles no tienen expresión analítica y para poder calcularlos tenemos que recurrir a métodos numéricos, como por ejemplo el de prueba y error, teniendo en cuenta que el cuantil de orden ω , q_ω , es el punto que satisface:

$$\frac{1}{N} \sum_{j=1}^N F^{(j)}(q_\omega) \approx \omega$$

donde $F^{(j)}$ representa la función de distribución de la t-Student multivariante dada por (2.45).

Alternativamente se podría haber calculado la distribución predictiva integrando, en primer lugar, con respecto a ψ en lugar de Σ . Para ello interesa expresar la distribución a posteriori de los parámetros del modelo como en (2.15), de manera que:

$$\begin{aligned} f(P|y) &\propto \int \int \int |\Sigma|^{-\frac{n+h+m+1}{2}} \\ &\quad \exp \left\{ -\frac{1}{2} (P - \tilde{z} - D\psi)' (L_B^2)^{-1} (\Sigma^{-1} \otimes I_h) (L_B^2)^{-1} (P - \tilde{z} - D\psi) \right\} \\ &\quad \exp \left\{ -\frac{1}{2} (z - X_B \hat{\psi})' (\Sigma^{-1} \otimes I_n) (z - X_B \hat{\psi}) \right\} \\ &\quad \exp \left\{ -\frac{1}{2} (\psi - \hat{\psi})' X_B' (\Sigma^{-1} \otimes I_n) X_B (\psi - \hat{\psi}) \right\} d\psi d\Sigma d\theta \\ &\propto \int \int \int |\Sigma|^{-\frac{n+h+m+1}{2}} \exp \left\{ -\frac{1}{2} (\tilde{P} - \tilde{D}\psi)' (\Sigma^{-1} \otimes I_h) (\tilde{P} - \tilde{D}\psi) \right\} \\ &\quad \exp \left\{ -\frac{1}{2} (z - X_B \hat{\psi})' (\Sigma^{-1} \otimes I_n) (z - X_B \hat{\psi}) \right\} \\ &\quad \exp \left\{ -\frac{1}{2} (\psi - \hat{\psi})' X_B' (\Sigma^{-1} \otimes I_n) X_B (\psi - \hat{\psi}) \right\} d\psi d\Sigma d\theta \quad (2.51) \end{aligned}$$

siendo $\tilde{P} = (L_B^2)^{-1} (P - \tilde{z})$ y $\tilde{D} = (L_B^2)^{-1} D$. Teniendo en cuenta que las

dos formas cuadráticas que involucran a ψ pueden agruparse como

$$\begin{aligned} & (\tilde{P} - \tilde{D}\psi)'(\Sigma^{-1} \otimes I_h)(\tilde{P} - \tilde{D}\psi) + (\psi - \hat{\psi})'X'_B(\Sigma^{-1} \otimes I_n)X_B(\psi - \hat{\psi}) = \\ & = (\psi - \mu_\psi)'V_\psi(\psi - \mu_\psi) + \tilde{P}'(\Sigma^{-1} \otimes I_h)\tilde{P} + \hat{\psi}'X'_B(\Sigma^{-1} \otimes I_n)X_B\hat{\psi} - \mu'_\psi V_\psi \mu_\psi \end{aligned}$$

donde

$$\begin{aligned} V_\psi &= X'_B(\Sigma^{-1} \otimes I_n)X_B + \tilde{D}'(\Sigma^{-1} \otimes I_h)\tilde{D} \\ \mu_\psi &= V_\psi^{-1}(X'_B(\Sigma^{-1} \otimes I_n)X_B\hat{\psi} + \tilde{D}'(\Sigma^{-1} \otimes I_h)\tilde{P}) \end{aligned}$$

la distribución predictiva puede expresarse como:

$$\begin{aligned} f(P|y) &\propto \int \int \int |\Sigma|^{-\frac{n+h+m+1}{2}} \exp \left\{ -\frac{1}{2}(\psi - \mu_\psi)'V_\psi(\psi - \mu_\psi) \right\} \\ &\quad \exp \left\{ -\frac{1}{2} \left(\tilde{P}'(\Sigma^{-1} \otimes I_h)\tilde{P} + \hat{\psi}'X'_B(\Sigma^{-1} \otimes I_n)X_B\hat{\psi} - \mu'_\psi V_\psi \mu_\psi \right) \right\} \\ &\quad \exp \left\{ -\frac{1}{2}(z - X_B\hat{\psi})'(\Sigma^{-1} \otimes I_n)(z - X_B\hat{\psi}) \right\} d\psi d\Sigma d\theta \\ &\propto \int \int |\Sigma|^{-\frac{n+h+m+1}{2}} |V_\psi|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(z - X_B\hat{\psi})'(\Sigma^{-1} \otimes I_n)(z - X_B\hat{\psi}) \right\} \\ &\quad \exp \left\{ -\frac{1}{2} \left(\tilde{P}'(\Sigma^{-1} \otimes I_h)\tilde{P} + \hat{\psi}'X'_B(\Sigma^{-1} \otimes I_n)X_B\hat{\psi} - \mu'_\psi V_\psi \mu_\psi \right) \right\} d\Sigma d\theta \end{aligned}$$

que no tiene solución analítica. Por lo tanto, tras eliminar los términos que no dependen de P se obtiene:

$$f(P|y, \Sigma, \theta) \propto \exp \left\{ -\frac{1}{2} \left(\tilde{P}'(\Sigma^{-1} \otimes I_h)\tilde{P} - \mu'_\psi V_\psi \mu_\psi \right) \right\} \quad (2.52)$$

Por otro lado,

$$\begin{aligned} \tilde{P}'(\Sigma^{-1} \otimes I_h)\tilde{P} - \mu'_\psi V_\psi \mu_\psi &= \tilde{P}'(\Sigma^{-1} \otimes I_h)\tilde{P} \\ &\quad - \tilde{P}'(\Sigma^{-1} \otimes I_h)\tilde{D}V_\psi^{-1}\tilde{D}'(\Sigma^{-1} \otimes I_h)\tilde{P} \\ &\quad - \tilde{P}'(\Sigma^{-1} \otimes I_h)\tilde{D}V_\psi^{-1}X'_B(\Sigma^{-1} \otimes I_n)X_B\hat{\psi} \\ &\quad - \hat{\psi}'X'_B(\Sigma^{-1} \otimes I_n)X_BV_\psi^{-1}\tilde{D}'(\Sigma^{-1} \otimes I_h)\tilde{P} \\ &\quad - \hat{\psi}'X'_B(\Sigma^{-1} \otimes I_n)X_BV_\psi^{-1}X'_B(\Sigma^{-1} \otimes I_n)X_B\hat{\psi} \end{aligned}$$

que, completando cuadrados en \tilde{P} , puede expresarse como:

$$\begin{aligned} \tilde{P}'(\Sigma^{-1} \otimes I_h)\tilde{P} - \mu'_\psi V_\psi \mu_\psi &= (\tilde{P} - c_p)'V_p(\tilde{P} - c_p) - c'_p V_p c_p \\ &\quad - \hat{\psi}'X'_B(\Sigma^{-1} \otimes I_n)X_BV_\psi^{-1}X'_B(\Sigma^{-1} \otimes I_n)X_B\hat{\psi} \end{aligned} \quad (2.53)$$

siendo V_p la matriz $(\Sigma^{-1} \otimes I_h) - (\Sigma^{-1} \otimes I_h) \tilde{D} V_\psi^{-1} \tilde{D}' (\Sigma^{-1} \otimes I_h)$ y $c_p = V_p^{-1} (\Sigma^{-1} \otimes I_h) \tilde{D} V_\psi^{-1} X_B' (\Sigma^{-1} \otimes I_n) X_B \hat{\psi}$.

Finalmente, sustituyendo (2.53) en (2.52) y eliminando los términos que no dependen de P , la distribución predictiva condicionada a Σ y θ puede expresarse como

$$f(P|y, \Sigma, \theta) \propto \exp \left\{ -\frac{1}{2} (\tilde{P} - c_p)' V_p (\tilde{P} - c_p) \right\}$$

de donde se deduce que el vector \tilde{P} de los datos futuros transformados sigue una distribución Normal multivariante con vector de medias c_p y matriz de covarianzas V_p^{-1} . Por tanto, la distribución del vector P de los datos futuros es la Normal multivariante:

$$f(P|y, \Sigma, \theta) = N_{hm}(L_B^2 c_p + \tilde{z}, L_B^2 V_p^{-1} (L_B^2)') \quad (2.54)$$

Mediante integración por Monte Carlo, de manera análoga a la mostrada en las ecuaciones (2.48), (2.49) y (2.50), es posible estimar tanto la distribución predictiva como sus principales momentos.

Si en lugar de partir de la distribución a priori de Jeffreys partimos de cualquiera de las distribuciones a priori de referencia, ecuaciones (2.30) y (2.31), en el cálculo de la distribución predictiva tenemos que integrar respecto a ψ , pues la distribución para la matriz Σ deja de ser conocida. En este caso obtenemos la distribución predictiva condicional $f(P|y, \Sigma, \theta)$, ecuación (2.54), y la distribución predictiva es estimada mediante integración por Monte Carlo.

2.3.1. Ejemplo de ocupación hotelera (continuación)

Una vez estimada la distribución a posteriori de los parámetros del modelo de Holt-Winters multivariante resultante de analizar conjuntamente las series temporales de ocupación hotelera, sección 2.2.4, nos planteamos el cálculo de predicciones para los valores futuros de las series. Recordar que consideramos como datos históricos las observaciones correspondientes a los cinco primeros años (2001 - 2005), mientras que el año 2006 es utilizado para valorar la precisión de nuestra predicción fuera de la muestra.

La tabla 2.5 muestra el error SMAPE de ajuste y de predicción para cada una de las series de ocupación hotelera. Como posibles distribuciones

a priori para los parámetros del modelo consideramos la priori de Jeffreys y la priori de referencia (2.31). Dada la muestra $\{(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)})\}_{j=1}^N$ de la distribución a posteriori de los parámetros del modelo, las estimaciones para las observaciones que conforman la serie temporal i -ésima son obtenidas a partir de (2.1) sustituyendo los parámetros de suavizado y las condiciones iniciales del modelo por sus medias muestrales. Las predicciones puntuales coinciden con la media de la distribución predictiva, que es estimada mediante integración por Monte-Carlo como se muestra en (2.49).

	Priori Jeffreys		Priori referencia	
	Error ajuste	Error predicción	Error ajuste	Error predicción
Castellón	9.70	10.57	9.75	10.33
Valencia	11.14	28.58	10.92	29.13
Alicante	6.37	17.78	6.38	17.52
Media	9.07	18.98	9.02	18.99

Tabla 2.5: Error SMAPE de ajuste y de predicción para cada una de las series de ocupación hotelera resultantes del análisis Bayesiano multivariante

A partir de los errores mostrados, observamos que el ajuste y la predicción obtenidos en el análisis multivariante de las series temporales no están afectados por el uso de la distribución a priori de Jeffreys o la priori de referencia. Así pues, tal y como sugeríamos en la sección 2.2.4, el uso de la priori de Jeffreys puede ser aconsejado.

Por otro lado, mientras que los errores de predicción resultantes de los análisis univariantes de las series de Castellón y Alicante son del mismo orden (ver tabla 1.2), el error de predicción correspondiente a Valencia, que es la serie que presenta menor precisión en la predicción, se reduce considerablemente con el uso del modelo multivariante.

2.4. Notas computacionales

La matriz L_i , para $i = 1, 2, \dots, m$, es una matriz triangular inferior que, dada su estructura, puede ser representada a través de su primera columna; es decir, para trabajar con esta matriz sólo es necesario calcular y almacenar su primera columna, pues la matriz puede ser reconstruida a partir de ésta. Dicha propiedad se conserva también en su inversa, por lo que sólo será necesario el cálculo de su primera columna $(l^1, l^2, \dots, l^n)'$, donde $l^1 = 1$ y los demás elementos pueden ser calculados recursivamente como

$$\sum_{i=1}^k l^i l_{k-i+1} = 0, \text{ para } k = 2, 3, \dots, n.$$

Para la obtención del estimador mínimos cuadrados generalizado de ψ es necesario el cálculo de $(X'_B(\Sigma^{-1} \otimes I_n)X_B)^{-1}$, el cual puede originar problemas numéricos. Para solventar este problema hacemos uso, en primer lugar, de la descomposición de Cholesky de Σ^{-1} .

Definición 2.1 Sea A una matriz $n \times n$ simétrica y definida positiva. La descomposición de Cholesky de la matriz A es:

$$A = R'R$$

donde R es una matriz triangular superior.

Aplicando dicha descomposición a Σ^{-1} y haciendo uso de las propiedades del producto de Kronecker, obtenemos que $\Sigma^{-1} \otimes I_n = (R \otimes I_n)'(R \otimes I_n)$. A continuación, calculamos la descomposición en valores singulares de la matriz $(R \otimes I_n)X_B$ que, numéricamente, es mucho más estable que el cálculo de la matriz inversa.

Definición 2.2 Sea A una matriz $n \times m$. La descomposición en valores singulares de la matriz A es:

$$A = UDV'$$

donde D es una matriz diagonal $m \times m$, con los elementos de la diagonal la raíz cuadrada de los valores propios de $A'A$, que es una matriz simétrica y definida no negativa, por lo que sus valores propios serán todos no negativos. V' es la matriz $m \times m$ ortogonal formada por los vectores propios de $A'A$. U es una matriz $n \times m$ de columnas ortogonales.

Aplicando dicha descomposición a la matriz $(R \otimes I_n)X_B$, obtenemos que ésta puede ser expresada como $(R \otimes I_n)X_B = UDV'$, por lo que:

$$X'_B(\Sigma^{-1} \otimes I_n)X_B = X'_B(R \otimes I_n)'(R \otimes I_n)X_B = VDDV'$$

y

$$(X'_B(\Sigma^{-1} \otimes I_n)X_B)^{-1} = VD^{-1}D^{-1}V' \quad (2.55)$$

La aplicación informática que empleamos en esta memoria es R (R Development Core Team, 2008), donde la descomposición de Cholesky de una matriz A se obtiene mediante la instrucción $chol(A)$, que origina la matriz R de la definición. De manera similar, la descomposición en valores singulares

se obtiene mediante la función $svd(A)$, la cual da como salida una lista con tres elementos: u , v y d , siendo u y v las matrices U y V de la definición de descomposición en valores singulares y d el vector cuyas componentes son los elementos de la diagonal de la matriz D .

Para la simulación de la distribución Normal multivariante, requerida en la ejecución del algoritmo de Metropolis-within-Gibbs, hacemos uso de la siguiente función de R:

$$rmnorm(n, mean, varcov)$$

donde n representa el número de simulaciones que se desean obtener y $mean$ y $varcov$ el vector de medias y la matriz de covarianzas de la Normal multivariante. Esta función origina una matriz de n vectores aleatorios.

Analogamente, para obtener simulaciones de una Wishart utilizamos la instrucción:

$$rwishart(nu, V)$$

siendo nu los grados de libertad y V la matriz de localización, tales que si $f(\Sigma) = W(nu, V)$, $E[\Sigma] = nuV$. Si se desea simular de una Wishart invertida, es necesario invertir previamente la matriz de localización, pues si $f(\Sigma) = IW(nu, V)$, entonces $f(\Sigma^{-1}) = W(nu, V^{-1})$. El resultado de esta instrucción es una lista con dos elementos: W , matriz simulada a partir de la Wishart e IW , matriz simulada de la Wishart invertida e inversa de la matriz W .

Para el cálculo de la matriz V_p^{-1} , la cual aparece tanto en el vector de medias como en la matriz de covarianzas de la distribución predictiva condicionada a Σ y θ , ecuación (2.54), puede emplearse el siguiente resultado (Mardia et al., 1979, pag. 459):

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

De esta forma,

$$\begin{aligned} V_p^{-1} &= (I_{mh} - \tilde{D}V_\psi^{-1}\tilde{D}'(\Sigma^{-1} \otimes I_h))^{-1}(\Sigma \otimes I_h) \\ &= (I_{mh} + \tilde{D}(V_\psi - \tilde{D}'(\Sigma^{-1} \otimes I_h)\tilde{D})^{-1}\tilde{D}'(\Sigma^{-1} \otimes I_h))(\Sigma \otimes I_h) \\ &= (\Sigma \otimes I_h) + \tilde{D}(V_\psi - \tilde{D}'(\Sigma^{-1} \otimes I_h)\tilde{D})^{-1}\tilde{D}' \\ &= (\Sigma \otimes I_h) + \tilde{D}(X_B'(\Sigma^{-1} \otimes I_n)X_B)^{-1}\tilde{D}' \end{aligned}$$

donde $(X_B'(\Sigma^{-1} \otimes I_n)X_B)^{-1}$ se calcula utilizando la descomposición en valores singulares, ecuación (2.55).

Capítulo 3

Holt-Winters multivariante: casos particulares

El modelo de Holt-Winters es el más general dentro de la clase de modelos de suavizado exponencial. Dicho modelo resulta adecuado para el análisis de series temporales que evolucionan en el tiempo y presentan oscilaciones que se producen a corto plazo y se repiten periódicamente. Sin embargo, es frecuente en la práctica observar series temporales con tendencia pero sin componente estacional. La mayoría de las series que cumplen estas características surgen en el ámbito económico-financiero: series de evolución de los tipos de interés, del Producto Interior Bruto (PIB), del Índice de Producción Industrial (IPI), del Índice de Precios al Consumo (IPC), series de tipo de cambio, de cotización en bolsa, etc. En estas ocasiones, el uso del modelo de Holt, al incorporar únicamente la tendencia de la serie, resulta más adecuado que el modelo de Holt-Winters, el cual fuerza a estimar la componente estacional.

Dada la importancia que la economía financiera tiene en la vida económica actual, dedicamos la primera parte de este capítulo al estudio del modelo de Holt multivariante, el cual nos permite analizar de manera conjunta series temporales sin componente estacional sujetas a errores correlacionados. La formulación multivariante que proponemos es equivalente a la propuesta para el modelo de Holt-Winters multivariante, por lo que su análisis se corresponde con el descrito en el capítulo anterior.

Por otro lado, cuando se analizan series económico-financieras con el objetivo de obtener predicciones precisas para los datos futuros de las series,

hay que tener en cuenta que dichos valores pueden ser observados de manera secuencial. En estas situaciones, las predicciones correspondientes a las series temporales observadas con posterioridad podrían ser mejoradas si se incorpora la información disponible en la distribución predictiva calculada inicialmente.

La segunda parte del capítulo está dedicada al estudio del modelo de Holt-Winters multivariante bajo el supuesto de parámetros de suavizado iguales para todos los modelos univariantes. Esta condición, equivalente a suponer que los modelos univariantes comparten una misma estructura, ha sido supuesta en trabajos previos (Enns et al., 1982; Harvey, 1986 West and Harrison, 1989). En este caso, el modelo multivariante puede formularse como un modelo de regresión multivariante tradicional (Tiao and Zellner, 1964), lo que simplifica considerablemente su análisis.

3.1. Holt multivariante

Supongamos que tenemos n observaciones de m series temporales, $y_i = (y_{1i}, y_{2i}, \dots, y_{ni})'$ para $i = 1, 2, \dots, m$, cada una de las cuales se ajusta al modelo de Holt univariante con errores aditivos. Entonces, utilizando la formulación del modelo de Holt como modelo de espacio de estado, se tiene:

Ecuación de observación

$$y_{ti} = a_{t-1,i} + b_{t-1,i} + \epsilon_{ti}$$

Ecuaciones de transición

$$\begin{aligned} a_{ti} &= a_{t-1,i} + b_{t-1,i} + \alpha_i \epsilon_{ti} \\ b_{ti} &= b_{t-1,i} + \alpha_i \beta_i \epsilon_{ti} \end{aligned}$$

donde a_{ti} y b_{ti} son el nivel y la tendencia en el instante t ; $\epsilon_i = (\epsilon_{1i}, \epsilon_{2i}, \dots, \epsilon_{ni})'$ es el vector de los errores, independientes entre sí e idénticamente distribuidos $N(0, \sigma_i^2)$; $\omega_i = (a_{0i}, b_{0i})'$ es el vector de las condiciones iniciales y $\theta_i = (\alpha_i, \beta_i)'$ es el vector de los parámetros de suavizado asociados a una serie temporal i -ésima.

Utilizando la ecuación de observación recursivamente junto con las ecuaciones de transición, la observación en el instante t de la serie temporal

i -ésima puede ser descrita (ver (1.6)) como:

$$y_{ti} = a_{0i} + tb_{0i} + \alpha_i \sum_{r=1}^{t-1} (1 + (t-r)\beta_i)\epsilon_{ri} + \epsilon_{ti}$$

Definiendo las matrices

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{pmatrix} \quad y \quad L_i = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{2,i} & 1 & 0 & \dots & 0 \\ l_{3,i} & l_{2,i} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n,i} & l_{n-1,i} & l_{n-2,i} & \dots & 1 \end{pmatrix}$$

donde $l_{j,i} = \alpha_i(1 + (j-1)\beta_i)$ para $j = 2, 3, \dots, n$, los datos que conforman la serie temporal i -ésima pueden ser descritos matricialmente como:

$$y_i = A\omega_i + L_i\epsilon_i \quad (3.1)$$

por lo que el vector de datos y_i se distribuye según la Normal multivariante con vector de medias $E(y_i) = A\omega_i$ y matriz de covarianzas $V(y_i) = \sigma_i^2 L_i L_i'$.

La ecuación (3.1) es similar a la expresión obtenida para el modelo de Holt-Winters univariante, ecuación (1.7). La principal diferencia es que, en el modelo de Holt, el vector ω_i de las condiciones iniciales está compuesto únicamente por dos componentes, el nivel y la tendencia en el instante inicial. La matriz A es una matriz $n \times 2$ de rango completo y ω_i es identificable, no siendo necesario, por tanto, imponer restricciones de identificabilidad sobre las condiciones iniciales. Respecto a la matriz L_i , la única diferencia es que no aparece sumando el parámetro γ_i en aquellos elementos $l_{j,i}$ tales que $j \equiv 1(\text{mod } s)$.

Proposición 3.1 *La expresión matricial dada por (3.1) es equivalente a (1.9) fijando s , periodo del ciclo estacional, igual a 1 y $\psi_i = (b_{0i}, a_{0i} + b_{0i})'$.*

Demostración:

Si en (1.9) fijamos $s = 1$, los elementos $\{l_{j,i}\}_{j=2}^n$ de la matriz L_i son todos de la forma $l_{j,i} = \alpha_i(1 + (j-1)\beta_i)$, pues no hay ningún j tal que $j \equiv 1(\text{mod } 1)$. Por otro lado,

$$\psi_i = \begin{pmatrix} b_{0i} \\ a_{0i} + b_{0i} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a_{0i} \\ b_{0i} \end{pmatrix} = T\omega_i \quad y,$$

$$M\psi_i = MT\omega_i = A\omega_i$$

siendo M la matriz $n \times 2$ cuya primera columna es el vector $(0, 1, \dots, n-1)'$ y la segunda el vector $(1, 1, \dots, 1)'$ ■

Considerando conjuntamente las m series temporales y haciendo uso de la proposición 3.1, el modelo de Holt multivariante puede ser expresado matricialmente como:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_y = \underbrace{\begin{bmatrix} M & 0 & \dots & 0 \\ 0 & M & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M \end{bmatrix}}_{M_D} \underbrace{\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix}}_\psi + \underbrace{\begin{bmatrix} L_1 & 0 & \dots & 0 \\ 0 & L_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L_m \end{bmatrix}}_{L_B} \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}}_\varepsilon \quad (3.2)$$

o, equivalentemente, como:

$$\underbrace{\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}}_z = \underbrace{\begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_m \end{bmatrix}}_{X_B} \underbrace{\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix}}_\psi + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}}_\varepsilon \quad (3.3)$$

siendo $z_i = (L_i)^{-1}y_i$ y $X_i = (L_i)^{-1}M$, para $i = 1, 2, \dots, m$. La distribución del vector de errores ε es la Normal multivariante de media el vector nulo y matriz de covarianzas $\Sigma \otimes I_n$.

De este modo, la formulación propuesta para el modelo de Holt multivariante, ecuación (3.2), coincide con la obtenida para el modelo de Holt-Winters multivariante, ecuación (2.2), siendo $s = 1$ y $\psi_i = (b_{0i}, c_{0i})'$ con $c_{0i} = a_{0i} + b_{0i}$, para $i = 1, 2, \dots, m$. Entonces, la distribución a posteriori para los parámetros del modelo, $f(\psi, \Sigma, \theta|y)$, coincide con la obtenida en el capítulo 2, ecuación (2.26) si se parte de la priori de Jeffreys, ecuación (2.32) si se toma la priori dada por (2.31), o ecuación (2.36) si se considera la priori de referencia de Yang and Berger (1994), ecuación (2.30). Respecto al algoritmo de Metropolis-within-Gibbs propuesto para obtener una muestra de la distribución a posteriori de los parámetros del modelo, la única diferencia del modelo de Holt multivariante en relación al modelo de Holt-Winters es que los vectores θ_i son ahora de dimensión 2 y, por tanto, utilizaremos

como función generadora para su simulación la uniforme en el cuadrado en lugar de la uniforme en el cubo. En cuanto a la distribución predictiva, los resultados obtenidos coinciden con los vistos en el capítulo 2, sección 2.3.

Dados los valores simulados $\{\psi_i^{(j)} = (b_{0i}^{(j)}, c_{0i}^{(j)})'\}_{j=1}^N$ y aplicando la transformación $a_{0i}^{(j)} = c_{0i}^{(j)} - b_{0i}^{(j)}$, obtenemos una muestra $\{(a_{0i}^{(j)}, b_{0i}^{(j)})'\}_{j=1}^N$ de la distribución $f(\omega_i|y)$, para $i = 1, 2, \dots, m$, que nos permite obtener inferencias acerca de las condiciones iniciales originales.

3.1.1. Predicción de un paso cuando los valores futuros son observados secuencialmente

La aplicación de diferentes técnicas estadísticas de análisis de series temporales en economía ha sido extensa, especialmente en el estudio de la evolución de las bolsas y de los índices que las sintetizan, pues son reconocidos como los principales indicadores de la economía de un país. La introducción de modelos multivariantes ha supuesto un avance importante en su estudio, pues permiten incorporar información proporcionada por series temporales correlacionadas y obtener así análisis más robustos. En estos casos, hay que considerar la posibilidad de valores observados de manera secuencial para las distintas series univariantes. Cuando el objetivo final del análisis es obtener predicciones de los valores futuros, puede resultar conveniente modificar las predicciones obtenidas originalmente conforme se dispone de nuevos datos.

En este apartado mostramos como modificar la distribución predictiva de un único paso en cada serie para incorporar la información proporcionada por los valores futuros de las series temporales cuando éstos son observados de manera secuencial. Es decir, si $y = (y'_1, y'_2, \dots, y'_m)'$ es el vector $nm \times 1$ de los datos que conforman las m series temporales, ordenadas de manera que para el instante t , y_{t1} se observa en primer lugar, seguido de y_{t2} y así sucesivamente hasta y_{tm} , y $p = (p_1, p_2, \dots, p_m)'$ es el vector $m \times 1$ de los datos futuros de las m series temporales, en esta sección nos planteamos obtener $f(p_i, p_{i+1}, \dots, p_m | y, p_1, p_2, \dots, p_{i-1})$.

Fijando $h = 1$ en los resultados obtenidos en el capítulo 2, pues trabajamos con la distribución predictiva de un paso, se tiene que la distribución del vector p de los datos futuros, dados los datos observados, el vector ψ de las condiciones iniciales y el vector θ de los parámetros de suavizado, es de

la forma (ver ecuación (2.43)):

$$f(p|y, \psi, \theta) \propto |H'H + (\tilde{P} - \tilde{\mu}_P)'(\tilde{P} - \tilde{\mu}_P)|^{-\frac{n+1}{2}} \quad (3.4)$$

donde $\tilde{P} = [p_1, p_2, \dots, p_m]$, $H = [z_1 - X_1\psi_1, z_2 - X_2\psi_2, \dots, z_m - X_m\psi_m]$ y $\tilde{\mu}_P = [\tilde{z}_1 + (M_p - L_1^{21}X_1)\psi_1, \tilde{z}_2 + (M_p - L_2^{21}X_2)\psi_2, \dots, \tilde{z}_m + (M_p - L_m^{21}X_m)\psi_m]$, con $\tilde{z}_i = L_i^{21}z_i$. Notar que cuando el horizonte de predicción es $h = 1$, de la definición de L_i como matriz triangular inferior con unos en la diagonal principal, se tiene que $L_i^2 = 1$ para $i = 1, 2, \dots, m$, de ahí que no aparezcan los términos $(L_i^2)^{-1}$ en las definiciones anteriores. Por tanto, la matriz \tilde{P} de los datos futuros, dados los datos observados, ψ y θ , se distribuye según la t-Student generalizada

$$f(\tilde{P}|y, \psi, \theta) = GSt(1, H'H, \tilde{\mu}_P, n + 1) \quad (3.5)$$

Sea $P_1 = (p_1, p_2, \dots, p_{i-1})'$ el vector $(i-1) \times 1$ de los valores de las series en el instante $n + 1$ que ya han sido observados y $P_2 = (p_i, p_{i+1}, \dots, p_m)'$ el vector $(m - i + 1) \times 1$ de los valores futuros de las series que todavía no han sido observados y que se quieren predecir. Entonces, particionando $\tilde{\mu}_P$ y $H'H$ de manera análoga a $\tilde{P} = [P_1', P_2']$, es decir:

$$\begin{aligned} \tilde{\mu}_P &= [\tilde{\mu}_{P_1}, \tilde{\mu}_{P_2}] \\ H'H &= \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \end{aligned}$$

y haciendo uso de las propiedades de la t-Student generalizada (Zellner, 1971, pag. 397), la distribución predictiva del vector P_2' , dados los datos observados, ψ , θ y P_1 , es t-Student generalizada, en particular:

$$f(P_2'|y, \psi, \theta, P_1) = GSt(A_{2.1}, B_{2.1}, \mu_{2.1}, n + 1) \quad (3.6)$$

donde $A_{2.1} = (1 + (P_1' - \tilde{\mu}_{P_1})Q_{11}^{-1}(P_1' - \tilde{\mu}_{P_1}))^{-1}$, $B_{2.1} = Q_{22} - Q_{21}Q_{11}^{-1}Q_{12}$ y $\mu_{2.1} = \tilde{\mu}_{P_2} + (P_1' - \tilde{\mu}_{P_1})Q_{11}^{-1}Q_{12}$.

Finalmente, de (3.6) se deduce que la distribución de los elementos de P_2 , p_k para $k = i, i + 1, \dots, m$, dados los datos observados, ψ , θ y P_1 , es de la forma:

$$f(p_k|y, \psi, \theta, P_1) \propto [B_{2.1}[r, r] + A_{2.1}(p_k - \mu_{2.1}[r])^2]^{-\frac{n+1-m+i}{2}} \quad (3.7)$$

donde $r = k - i + 1$. Es decir, la t-Student univariante con $\nu = n - m + i$ grados de libertad y media y varianza dadas por:

$$\begin{aligned} E(p_k|y, \psi, \theta, P_1) = \mu_{p_k} &= \mu_{2.1}[k - i + 1] \\ &= \tilde{\mu}_{P_2}[k - i + 1] + (P_1' - \tilde{\mu}_{P_1})Q_{11}^{-1}Q_{12}[k - i + 1] \\ V(p_k|y, \psi, \theta, P_1) = V_{p_k} &= \frac{1}{\nu - 2} \frac{B_{2.1}[k - i + 1, k - i + 1]}{A_{2.1}} \end{aligned}$$

para $k = i, i + 1, \dots, m$. Como podemos observar, tanto la media como la varianza son modificadas en función del error de predicción asociado a P_1 .

Aunque no es posible obtener de manera analítica la distribución predictiva de las variables p_i, p_{i+1}, \dots, p_m , dada $\{(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)})\}_{j=1}^N$ una muestra de la distribución a posteriori de los parámetros del modelo y aplicando integración por Monte Carlo, obtenemos las siguientes estimaciones:

$$f(p_k|y, P_1) \approx \frac{1}{N} \sum_{j=1}^N USt(\mu_{p_k}^{(j)}, V_{p_k}^{(j)}, \nu = n - m + i) \quad (3.8)$$

$$E(p_k|y, P_1) \approx \frac{1}{N} \sum_{j=1}^N \mu_{p_k}^{(j)} \quad (3.9)$$

$$V(p_k|y, P_1) \approx \frac{1}{N} \sum_{j=1}^N V_{p_k}^{(j)} + \frac{1}{N} \sum_{j=1}^N (\mu_{p_k}^{(j)})^2 - E^2(p_k|y, P_1) \quad (3.10)$$

donde USt hace referencia a la distribución t-Student univariante.

Si en lugar de considerar la distribución predictiva condicionada a ψ y θ , $f(p|y, \psi, \theta)$, consideramos la predictiva condicionada a Σ y θ , ecuación (2.54), se tiene que, cuando el horizonte de predicción es $h = 1$, el vector de los datos futuros se distribuye según la Normal multivariante:

$$f(p|y, \Sigma, \theta) = N_m(c_p + \tilde{z}, V_p^{-1})$$

siendo

$$\begin{aligned} \tilde{z} &= (L_1^{21}z_1, L_2^{21}z_2, \dots, L_m^{21}z_m)' \\ c_p &= V_p^{-1}\Sigma^{-1}DV_\psi^{-1}X_B'(\Sigma^{-1} \otimes I_n)X_B\hat{\psi} \\ V_p &= \Sigma^{-1} - \Sigma^{-1}DV_\psi^{-1}D'\Sigma^{-1} \\ V_\psi &= X_B'(\Sigma^{-1} \otimes I_n)X_B + D'\Sigma^{-1}D \end{aligned}$$

y D la matriz $m \times (s + 1)m$ diagonal a bloques, con bloques iguales a $M_p - L_i^{21} X_i$ para $i = 1, 2, \dots, m$ y siendo $s = 1$ para el modelo de Holt. En las definiciones anteriores se ha tenido en cuenta que $L_B^2 = I_m$ cuando $h = 1$.

Definiendo nuevamente $P_1 = (p_1, p_2, \dots, p_{i-1})'$ como el vector $(i - 1) \times 1$ de los valores de las series en $n + 1$ que ya han sido observados y $P_2 = (p_i, p_{i+1}, \dots, p_m)'$ como el vector $(m - i + 1) \times 1$ de los datos futuros que todavía no han sido observados, y particionando el vector de medias y la matriz de covarianzas de manera análoga a $p = (P_1', P_2')'$, es decir,

$$\begin{aligned} c_p + \tilde{z} &= (\mu'_{P_1}, \mu'_{P_2})' \\ V_P^{-1} &= \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \end{aligned}$$

la distribución de P_2 condicionada a los datos observados, Σ , θ y P_1 es (ver, por ejemplo, Mardia et al., 1979, pag. 63) la Normal multivariante:

$$f(P_2|y, \Sigma, \theta, P_1) = N_{m-i+1}(\mu_{P_{2.1}}, V_{P_{2.1}}) \quad (3.11)$$

con $\mu_{P_{2.1}} = \mu_{P_2} + V_{21} V_{11}^{-1} (P_1 - \mu_{P_1})$ y $V_{P_{2.1}} = V_{22} - V_{21} V_{11}^{-1} V_{12}$.

Para obtener una estimación tanto de la distribución predictiva de P_2 como de sus principales momentos es necesario nuevamente recurrir a la simulación.

Notar que, aunque hemos introducido este apartado en el marco de series temporales económico-financieras sin estacionalidad, los resultados obtenidos son aplicables a cualquier situación práctica en la que las series temporales se observen secuencialmente, y pueden ser generalizados de manera directa al caso de series con estacionalidad.

3.2. Holt-Winters multivariante con parámetros de suavizado iguales

Varios autores (Enns et al., 1982; Harvey, 1986) han trabajado bajo el supuesto de homogeneidad del modelo multivariante, equivalente a suponer que los modelos univariantes que lo constituyen comparten una misma estructura, pues de este modo su análisis se simplifica notablemente. Para los modelos de suavizado exponencial, expresados como modelos de espacio de

estado con una única fuente de error, esta suposición implica parámetros de suavizado iguales para los modelos univariantes. Esta propiedad resulta adecuada en muchas situaciones prácticas como, por ejemplo, en la modelización de indicadores financieros similares, precios de acciones o series de tipo de cambio. Por ello, en esta sección mostramos el análisis Bayesiano del modelo de Holt-Winters propuesto en el capítulo 2 bajo esta hipótesis.

Si en (2.2) suponemos que los parámetros de suavizado asociados a las series temporales univariantes son iguales, es decir, $\theta_1 = \theta_2 = \dots = \theta_m = \theta$, entonces $L_1 = L_2 = \dots = L_m = L$, pues todas ellas tienen la misma estructura y dependen únicamente de los parámetros de suavizado, y el modelo de Holt-Winters multivariante es de la forma:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_y = \underbrace{\begin{bmatrix} M & 0 & \dots & 0 \\ 0 & M & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M \end{bmatrix}}_{M_D} \underbrace{\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix}}_\psi + \underbrace{\begin{bmatrix} L & 0 & \dots & 0 \\ 0 & L & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L \end{bmatrix}}_{L_D} \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}}_\varepsilon \quad (3.12)$$

o, equivalentemente,

$$\underbrace{[y_1, y_2, \dots, y_m]}_{y_{mat}} = M \underbrace{[\psi_1, \psi_2, \dots, \psi_m]}_{\psi_{mat}} + L \underbrace{[\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m]}_{\varepsilon_{mat}} \quad (3.13)$$

siendo ψ_{mat} la matriz $(s+1) \times m$ de las condiciones iniciales y ε_{mat} la matriz $n \times m$ de errores, con distribución matriz Normal $N_{n,m}(0, \Sigma \otimes I_n)$.

Siguiendo la notación de Timm (2002), decimos que la distribución de una matriz aleatoria E de orden $n \times m$ es la matriz Normal $N_{n,m}(M, V \otimes W)$ si su densidad es de la forma:

$$f(E|M, V, W) = (2\pi)^{-\frac{nm}{2}} |V|^{-\frac{n}{2}} |W|^{-\frac{m}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [V^{-1}(E - M)'W^{-1}(E - M)] \right\}$$

donde V y W son matrices definidas positivas de orden $m \times m$ y $n \times n$ respectivamente. M representa la matriz de medias, V la matriz de covarianzas entre las columnas de E y W la matriz de covarianzas entre las filas de E . De la definición se sigue que la distribución del vector $vec(E)$, obtenido al colocar las columnas de la matriz E una a continuación de otra, es la Normal multivariante $N_{nm}(vec(M), V \otimes W)$.

Multiplicando (3.13) por L^{-1} se tiene que, cuando los parámetros de suavizado de los modelos univariantes son iguales, el modelo de Holt-Winters multivariante puede ser expresado como un modelo de regresión multivariante tradicional (Tiao and Zellner, 1964):

$$\underbrace{[z_1, z_2, \dots, z_m]}_{z_{mat}} = X [\psi_1, \psi_2, \dots, \psi_m] + [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m] \quad (3.14)$$

donde z_{mat} es la matriz $n \times m$ de datos transformados, con $z_i = L^{-1}y_i$ para $i = 1, 2, \dots, m$, y $X = L^{-1}M$ es la matriz $n \times (s+1)$ de diseño, dependiente de los parámetros de suavizado y, por consiguiente, desconocida.

La función de verosimilitud asociada al modelo de Holt-Winters multivariante en este caso viene dada por la distribución matriz Normal de la matriz de datos:

$$\begin{aligned} f(y|\psi_{mat}, \Sigma, \theta) &\propto |\Sigma|^{-\frac{n}{2}} |LL'|^{-\frac{m}{2}} \\ &\exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (y_{mat} - M\psi_{mat})' (LL')^{-1} (y_{mat} - M\psi_{mat})] \right\} \\ &\propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat})] \right\} \end{aligned} \quad (3.15)$$

La forma cuadrática en (3.15) puede descomponerse como:

$$\begin{aligned} &(z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat}) = \\ &(z_{mat} - X\tilde{\psi}_{mat} + X(\tilde{\psi}_{mat} - \psi_{mat}))' (z_{mat} - X\tilde{\psi}_{mat} + X(\tilde{\psi}_{mat} - \psi_{mat})) = \\ &((I - P_X)z_{mat} + X(\tilde{\psi}_{mat} - \psi_{mat}))' ((I - P_X)z_{mat} + X(\tilde{\psi}_{mat} - \psi_{mat})) = \\ &z_{mat}' (I_n - P_X) z_{mat} + (\psi_{mat} - \tilde{\psi}_{mat})' X' X (\psi_{mat} - \tilde{\psi}_{mat}) \end{aligned}$$

donde $\tilde{\psi}_{mat} = (X'X)^{-1}X'z_{mat}$ es el estimador mínimos cuadrados de ψ_{mat} cuando el valor de θ es conocido y $P_X = X(X'X)^{-1}X'$ es la matriz de proyección ortogonal sobre el espacio generado por las columnas de la matriz X , simétrica, idempotente y tal que $P_X X = X$. Entonces, la función de verosimilitud puede expresarse como:

$$\begin{aligned} f(y|\psi_{mat}, \Sigma, \theta) &\propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} z_{mat}' (I_n - P_X) z_{mat}] \right\} \\ &\exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (\psi_{mat} - \tilde{\psi}_{mat})' X' X (\psi_{mat} - \tilde{\psi}_{mat})] \right\} \end{aligned} \quad (3.16)$$

Proposición 3.2 *Bajo el supuesto de parámetros de suavizado iguales y conocidos, el modelo de Holt-Winters multivariante puede ser expresado como*

un modelo lineal dinámico Matriz Normal.

Demostración:

Para $i = 1, 2, \dots, m$, la observación en el instante t de la serie temporal i -ésima puede expresarse, ver ecuaciones (2.6) y (2.7), en la forma:

$$\begin{aligned} y_{ti} &= a_{t-1,i} + b_{t-1,i} + c_{t-s,i} + \epsilon_{ti} \\ &= \underbrace{(1, 1, 0, 0, \dots, 0, 1)}_{f'} \varphi_{ti} + \epsilon_{ti} \end{aligned}$$

donde

$$\underbrace{\begin{pmatrix} a_{t-1,i} \\ b_{t-1,i} \\ c_{t-1,i} \\ c_{t-2,i} \\ c_{t-3,i} \\ \vdots \\ c_{t-s+1,i} \\ c_{t-s,i} \end{pmatrix}}_{\varphi_{ti}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}}_G \underbrace{\begin{pmatrix} a_{t-2,i} \\ b_{t-2,i} \\ c_{t-2,i} \\ c_{t-3,i} \\ c_{t-4,i} \\ \vdots \\ c_{t-s,i} \\ c_{t-s-1,i} \end{pmatrix}}_{\varphi_{t-1,i}} + \underbrace{\begin{pmatrix} \alpha_i \epsilon_{t-1,i} \\ \alpha_i \beta_i \epsilon_{t-1,i} \\ \gamma_i \epsilon_{t-1,i} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}}_{\eta_{ti}}$$

siendo la distribución del vector de errores η_{ti} la Normal multivariante de media el vector nulo y matriz de covarianzas dada por:

$$V(\eta_{ti}) = \sigma_i^2 \underbrace{\begin{pmatrix} W_{ii} & 0_{3 \times (s-1)} \\ 0_{(s-1) \times 3} & 0_{(s-1) \times (s-1)} \end{pmatrix}}_{W_i}, \quad \text{con } W_{ii} = \begin{pmatrix} \alpha_i^2 & \alpha_i^2 \beta_i & \alpha_i \gamma_i \\ \alpha_i^2 \beta_i & \alpha_i^2 \beta_i^2 & \alpha_i \beta_i \gamma_i \\ \alpha_i \gamma_i & \alpha_i \beta_i \gamma_i & \gamma_i^2 \end{pmatrix}$$

Entonces, si los parámetros de suavizado de las series univariantes son iguales, $W_1 = W_2 = \dots = W_m = W$, y el vector de las m observaciones en el instante t , $y_t = (y_{t1}, y_{t2}, \dots, y_{tm})'$, puede ser expresado como:

$$\begin{aligned} \underbrace{[y_{t1}, y_{t2}, \dots, y_{tm}]}_{y'_t} &= f' \underbrace{[\varphi_{t1}, \varphi_{t2}, \dots, \varphi_{tm}]}_{\Theta_t} + \underbrace{[\epsilon_{t1}, \epsilon_{t2}, \dots, \epsilon_{tm}]}_{\xi'_t} \\ \underbrace{[\varphi_{t1}, \varphi_{t2}, \dots, \varphi_{tm}]}_{\Theta_t} &= G \underbrace{[\varphi_{t-1,1}, \varphi_{t-1,2}, \dots, \varphi_{t-1,m}]}_{\Theta_{t-1}} + \underbrace{[\eta_{t1}, \eta_{t2}, \dots, \eta_{tm}]}_{\Upsilon_t} \end{aligned}$$

donde el vector de errores ξ_t está distribuido según la Normal multivariante $N_m(0, \Sigma)$ y la matriz Υ_t según la matriz Normal $N_{(s+2),m}(0, \Sigma \otimes W)$ ■

Esta clase de modelos (West and Harrison, 1989, capítulo 15) permite un análisis conjugado secuencial que generaliza el análisis estándar de series univariantes con varianza desconocida mediante el filtro de Kalman. El análisis multivariante esta basado en el uso de distribuciones a priori y posteriori Matriz Normal/Wishart invertida para la matriz de estado Θ_t y la matriz de covarianzas Σ respectivamente, suponiendo que la matriz W es conocida. En el caso del modelo de Holt-Winters multivariante, este análisis es posible sólo si los parámetros de suavizado son conocidos, lo cual no ocurre en la práctica. Por otro lado, la generalización de este enfoque al caso de parámetros de suavizado desconocidos no es sencilla. Por ello, y con el fin de obtener inferencias acerca de los parámetros del modelo, calculamos, al igual que hemos hecho en el modelo general, su distribución final como el producto de la priori por la verosimilitud.

3.2.1. Análisis Bayesiano

Si el vector θ de parámetros de suavizado es conocido, lo cual implica que las matrices L y X sean conocidas, el análisis Bayesiano del modelo de regresión multivariante es conocido (Zellner, 1971, capítulo 8). Asumiendo nuevamente poca información a priori, independencia entre los elementos de ψ_{mat} y Σ y utilizando la teoría de la invarianza de Jeffreys, la distribución a priori propuesta es:

$$\begin{aligned} f(\psi_{mat}, \Sigma | \theta) &\propto f(\psi_{mat} | \theta) f(\Sigma | \theta) \\ &\propto |\Sigma|^{-\frac{m+1}{2}} \end{aligned} \quad (3.17)$$

Entonces, la distribución a posteriori, proporcional al producto de la priori por la verosimilitud, es de la forma:

$$\begin{aligned} f(\psi_{mat}, \Sigma | y, \theta) &\propto |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat})] \right\} \\ &\propto |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} z'_{mat} (I_n - P_X) z_{mat}] \right\} \\ &\quad \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (\psi_{mat} - \tilde{\psi}_{mat})' X' X (\psi_{mat} - \tilde{\psi}_{mat})] \right\} \end{aligned} \quad (3.18)$$

de donde se deducen las siguientes distribuciones:

$$f(\psi_{mat} | y, \Sigma, \theta) = N_{(s+1),m}(\tilde{\psi}_{mat}, \Sigma \otimes (X'X)^{-1}) \quad (3.19)$$

$$f(\Sigma | y, \theta) = IW(n - s - 1, z'_{mat} (I_n - P_X) z_{mat}) \quad (3.20)$$

siendo (3.19) equivalente a:

$$f(\psi | y, \Sigma, \theta) = N_{(s+1)m}(\text{vec}(\tilde{\psi}_{mat}), \Sigma \otimes (X'X)^{-1}) \quad (3.21)$$

Por otro lado, integrando respecto a Σ la distribución a posteriori conjunta, ecuación (3.18), se obtiene la marginal para ψ_{mat} :

$$\begin{aligned} f(\psi_{mat}|y, \theta) &= \int f(\psi_{mat}, \Sigma|y, \theta) d\Sigma \\ &\propto \int |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (z_{mat} - X\psi_{mat})'(z_{mat} - X\psi_{mat})] \right\} d\Sigma \\ &\propto |(z_{mat} - X\psi_{mat})'(z_{mat} - X\psi_{mat})|^{-\frac{n}{2}} \\ &\propto |z'_{mat}(I_n - P_X)z_{mat} + (\psi_{mat} - \tilde{\psi}_{mat})'X'X(\psi_{mat} - \tilde{\psi}_{mat})|^{-\frac{n}{2}} \end{aligned}$$

de donde se deduce que la distribución a posteriori de ψ_{mat} , supuesto conocido el valor de los parámetros de suavizado, es una t-Student generalizada, en concreto,

$$f(\psi_{mat}|y, \theta) = GSt(X'X, z'_{mat}(I_n - P_X)z_{mat}, \tilde{\psi}_{mat}, n) \quad (3.22)$$

Entonces, la distribución marginal a posteriori de cada una de las columnas de la matriz ψ_{mat} es una t-Student multivariante, en particular:

$$f(\psi_i|y, \theta) = MSt \left(\tilde{\psi}_{mat}[i], \frac{X'X}{S[i, i]}, \nu = n - m - s \right) \quad (3.23)$$

siendo $S = z'_{mat}(I_n - P_X)z_{mat}$.

Sin embargo, dado que el vector θ de parámetros de suavizado es siempre desconocido, es necesario el uso de una distribución a priori para θ . Siguiendo el análisis del modelo general, tomamos $f(\theta)$ la uniforme en el cubo unidad, pues en el caso de parámetros de suavizado comunes para los modelos univariantes $\theta = (\alpha, \beta, \gamma)$. Para el resto de parámetros tomamos la distribución (3.17). De este modo,

$$\begin{aligned} f(\psi_{mat}, \Sigma, \theta) &\propto f(\psi_{mat}, \Sigma|\theta)f(\theta) \\ &\propto |\Sigma|^{-\frac{m+1}{2}} \end{aligned} \quad (3.24)$$

y la distribución a posteriori de los parámetros del modelo, producto de la priori por la verosimilitud, es de la forma (3.18). Las distribuciones a posteriori para las matrices ψ_{mat} y Σ condicionadas al vector θ coinciden con (3.22) y (3.20) respectivamente, mientras que la distribución marginal

del vector de parámetros de suavizado viene dada por:

$$\begin{aligned}
 f(\theta|y) &= \int \int f(\psi_{mat}, \Sigma, \theta|y) d\psi_{mat} d\Sigma \\
 &\propto \int \int |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} tr [\Sigma^{-1} z'_{mat} (I_n - P_X) z_{mat}] \right\} \\
 &\quad \exp \left\{ -\frac{1}{2} tr [\Sigma^{-1} (\psi_{mat} - \tilde{\psi}_{mat})' X' X (\psi_{mat} - \tilde{\psi}_{mat})] \right\} d\psi_{mat} d\Sigma \\
 &\propto |X' X|^{-\frac{m}{2}} \int |\Sigma|^{-\frac{n-(s+1)+m+1}{2}} \exp \left\{ -\frac{1}{2} tr [\Sigma^{-1} z'_{mat} (I_n - P_X) z_{mat}] \right\} d\Sigma \\
 &\propto |X' X|^{-\frac{m}{2}} |z'_{mat} (I_n - P_X) z_{mat}|^{-\frac{n-s-1}{2}} \tag{3.25}
 \end{aligned}$$

que no es analíticamente tratable, pero puede ser estimada mediante simulación.

A diferencia del modelo general, cuando los parámetros de suavizado asociados a los modelos univariantes son iguales, basta con simular de la distribución a posteriori $f(\theta|y)$ pues, una vez obtenida una muestra $\{\theta^{(j)}\}_{j=1}^N$ de dicha distribución, las distribuciones a posteriori marginales para ψ_{mat} y Σ pueden ser estimadas mediante integración por Monte Carlo como:

$$f(\psi_i|y) \approx \frac{1}{N} \sum_{j=1}^N MSt^{(j)} \left(\tilde{\psi}_{mat}[i], \frac{X' X}{S[i, i]}, \nu = n - m - s \right)$$

para $i = 1, 2, \dots, m$, y

$$f(\Sigma|y) \approx \frac{1}{N} \sum_{j=1}^N IW^{(j)}(n - s - 1, S)$$

donde $S = z'_{mat} (I_n - P_X) z_{mat}$ y el superíndice (j) indica que los cálculos han sido realizados para el valor $\theta = \theta^{(j)}$.

A partir de la muestra $\{\theta^{(j)}\}_{j=1}^N$ obtenida, también es posible estimar los coeficientes de correlación y comprobar de este modo que las series temporales univariantes realmente están correlacionadas. Así pues, para cada $\theta^{(j)}$ simulamos $\Sigma^{(j)}$ a partir de su distribución a posteriori $f(\Sigma|y, \theta)$, ecuación (3.20), y calculamos el coeficiente de correlación ρ_{kl} , para $k = 1, 2, \dots, m-1$ y $l = k+1, \dots, m$, como

$$\rho_{kl}^{(j)} = \frac{\Sigma^{(j)}[k, l]}{\sqrt{\Sigma^{(j)}[k, k] \Sigma^{(j)}[l, l]}} \tag{3.26}$$

Por otro lado, dado que la distribución a posteriori del vector θ de parámetros de suavizado, ecuación (3.25), es una función tridimensional con soporte acotado, el cubo unidad, proponemos emplear para su simulación el algoritmo de aceptación-rechazo, el cual nos permite obtener una secuencia de valores independientes e idénticamente distribuidos según $f(\theta|y)$. En particular, el mecanismo que proponemos es el que detallamos a continuación.

3.2.2. Método de aceptación-rechazo para simular de $f(\theta|y)$

Sea $\{R_i\}_{i=1}^T$ una partición del cubo unidad, por ejemplo la obtenida al dividir cada uno de los intervalos $[0, 1]$ en K subintervalos. Para cada región R_i , calcular M_i como

$$M_i = \sup_{\theta \in R_i} k(\theta|y)$$

donde $k(\theta|y) = |X'X|^{-\frac{m}{2}} |z'_{mat}(I_n - P_X)z_{mat}|^{-\frac{n-s-1}{2}}$ es el núcleo de $f(\theta|y)$. Entonces, definimos la función importante $s(\theta|y)$ como

$$s(\theta|y) \propto \begin{cases} M_i & \text{si } \theta \in R_i \\ 0 & \text{en otro caso} \end{cases}$$

es decir, la uniforme en cada una de las regiones R_i .

Una vez calculada la función de distribución importante, cada iteración del algoritmo de aceptación-rechazo que proponemos consta de los siguientes pasos:

1. Seleccionar una región R_j con probabilidad proporcional a M_j
2. Simular $\theta^{(j)}$ uniforme en R_j
3. Obtener $u \sim Un(0, 1)$
4. Si $uM_j > k(\theta^{(j)}|y)$, rechazar $\theta^{(j)}$ y volver al paso 1; en otro caso, aceptar $\theta^{(j)}$

En ocasiones, la distribución $f(\theta|y)$ puede ser muy puntiaguda en algunas regiones R_i , obteniendo así un valor M_i muy grande y como consecuencia un rechazo alto en dicha región. En estas situaciones es recomendable utilizar un algoritmo de aceptación-rechazo dinámico que acelere el proceso de simulación. En particular, cuando el rechazo en una región R_i es grande, proponemos particionar dicha región en T subregiones $\{R_{ij}\}_{j=1}^T$ y repetir el procedimiento con la nueva partición.

Notar que para calcular las constantes M_i necesitamos utilizar procedimientos de optimización no lineales que permitan incorporar restricciones de caja, pues restringimos el valor de los parámetros de suavizado a una región del cubo unidad. En concreto, el método que empleamos en este trabajo es el 'L-BFGS-B' (Byrd et al., 1995).

Por otro lado, en cada paso del algoritmo es necesario seleccionar una región R_j con probabilidad proporcional a M_j . Cuando el número de regiones T es grande, es conveniente utilizar el método alias (Devroye, 1986). Dicho método permite obtener simulaciones de una distribución discreta cuando el número de posibles valores es grande, evitando así el uso de la transformación inversa de su función de distribución que, al exigir demasiadas comparaciones, resulta ineficiente.

3.2.3. Distribución predictiva

Una vez obtenida la distribución a posteriori de los parámetros del modelo, calculamos la distribución predictiva $f(P|y)$, siendo $P = (P'_1, P'_2, \dots, P'_m)'$ el vector $hm \times 1$ de los datos futuros de las m series temporales. Siguiendo el procedimiento mostrado para el modelo general, el vector de los datos futuros puede ser expresado matricialmente, ver ecuación (2.42), como:

$$\underbrace{\begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{bmatrix}}_P = \underbrace{\begin{bmatrix} L^{21}z_1 \\ L^{21}z_2 \\ \vdots \\ L^{21}z_m \end{bmatrix}}_{\tilde{z}} + \underbrace{\begin{bmatrix} M_p - L^{21}X & 0 & \dots & 0 \\ 0 & M_p - L^{21}X & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_p - L^{21}X \end{bmatrix}}_D \underbrace{\begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix}}_{\psi} + \underbrace{\begin{bmatrix} L^2 & 0 & \dots & 0 \\ 0 & L^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L^2 \end{bmatrix}}_{L_D^2} \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}}_v \tag{3.27}$$

donde el vector de errores v está distribuido según $N_{hm}(0, \Sigma \otimes I_h)$. El cambio en la notación que observamos respecto a (2.42) es debido a que $L_1 = L_2 = \dots = L_m = L$, pues los parámetros de suavizado asociados a las series temporales univariantes son iguales. Consecuentemente, tanto los bloques en los que dichas matrices son particionadas como las matrices X_i

son iguales. En este caso, los datos futuros admiten la expresión alternativa:

$$\underbrace{[P_1, P_2, \dots, P_m]}_{P_{mat}} = L^{21} \underbrace{[z_1, z_2, \dots, z_m]}_{z_{mat}} + (M_p - L^{21} X) \underbrace{[\psi_1, \psi_2, \dots, \psi_m]}_{\psi_{mat}} + L^2 \underbrace{[v_1, v_2, \dots, v_m]}_{v_{mat}} \quad (3.28)$$

Definiendo $F = (L^2)^{-1}(P_{mat} - L^{21} z_{mat})$ y $B = (L^2)^{-1}(M_p - L^{21} X)$, (3.28) es equivalente a:

$$F = B\psi_{mat} + v_{mat} \quad (3.29)$$

de donde se deduce que la matriz F de los valores futuros transformados se distribuye según la matriz Normal:

$$f(F|y, \psi_{mat}, \Sigma, \theta) = N_{h,m}(B\psi_{mat}, \Sigma \otimes I_h)$$

Notar que las matrices F y B dependen del vector θ de los parámetros de suavizado y, por tanto, son desconocidas. Si θ fuera conocido, Zellner (1971, capítulo 8) calcula la distribución predictiva para la matriz F , obtenida a partir de la resolución de la siguiente integral:

$$\begin{aligned} f(F|y, \theta) &= \int \int f(F|y, \psi_{mat}, \Sigma, \theta) f(\psi_{mat}, \Sigma|y, \theta) d\psi_{mat} d\Sigma \\ &\propto \int \int |\Sigma|^{-\frac{h}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (F - B\psi_{mat})' (F - B\psi_{mat})] \right\} \\ &\quad |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat})] \right\} d\Sigma d\psi_{mat} \\ &\propto \int |(z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat}) + (F - B\psi_{mat})' (F - B\psi_{mat})|^{-\frac{n+h}{2}} d\psi_{mat} \end{aligned}$$

Completando cuadrados en ψ_{mat} , la suma de las dos formas cuadráticas puede expresarse como:

$$\begin{aligned} (z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat}) + (F - B\psi_{mat})' (F - B\psi_{mat}) &= \\ &= z'_{mat} z_{mat} - z'_{mat} X \psi_{mat} - \psi'_{mat} X' z_{mat} + \psi'_{mat} X' X \psi_{mat} \\ &\quad + F' F - F' B \psi_{mat} - \psi'_{mat} B' F + \psi'_{mat} B' B \psi_{mat} \\ &= (\psi_{mat} - \psi^*)' W (\psi_{mat} - \psi^*) + z'_{mat} z_{mat} + F' F - (\psi^*)' W \psi^* \end{aligned}$$

donde $W = X'X + B'B$ y $\psi^* = W^{-1}(X'z_{mat} + B'F)$. Entonces, la distribución predictiva $f(F|y, \theta)$ es proporcional a

$$\int |z'_{mat} z_{mat} + F' F - (\psi^*)' W \psi^* + (\psi_{mat} - \psi^*)' W (\psi_{mat} - \psi^*)|^{-\frac{n+h}{2}} d\psi_{mat}$$

$$= |z'_{mat}z_{mat} + F'F - (\psi^*)'W\psi^*|^{-\frac{n+h-(s+1)}{2}}$$

Con el fin de simplificar la expresión obtenida, Zellner propone completar cuadrados en F como:

$$\begin{aligned} & z'_{mat}z_{mat} + F'F - (\psi^*)'W\psi^* = \\ & = z'_{mat}z_{mat} + F'F - (z'_{mat}X + F'B)W^{-1}(X'z_{mat} + B'F) \\ & = (F - C^{-1}BW^{-1}X'z_{mat})'C(F - C^{-1}BW^{-1}X'z_{mat}) \\ & \quad + z'_{mat}(I_n - XW^{-1}X' - XW^{-1}B'C^{-1}BW^{-1}X')z_{mat} \end{aligned} \quad (3.30)$$

siendo $C = I_h - BW^{-1}B'$. De este modo, y teniendo en cuenta las siguientes igualdades:

$$\begin{aligned} C^{-1} &= I_h + B(X'X)^{-1}B' \\ C^{-1}BW^{-1} &= BW^{-1} + B(X'X)^{-1}B'BW^{-1} = \\ &= B(X'X)^{-1}(X'X + B'B)W^{-1} = B(X'X)^{-1} \\ XW^{-1}X' + XW^{-1}B'C^{-1}BW^{-1}X' &= XW^{-1}X' + XW^{-1}B'B(X'X)^{-1}X' = \\ &= XW^{-1}(X'X + B'B)(X'X)^{-1}X' = P_X \end{aligned}$$

la distribución predictiva para la matriz F puede ser expresada finalmente como:

$$f(F|y, \theta) \propto |S + (F - B\tilde{\psi}_{mat})'C(F - B\tilde{\psi}_{mat})|^{-\frac{n+h-s-1}{2}} \quad (3.31)$$

siendo $S = z'_{mat}(I_n - P_X)z_{mat}$ y $\tilde{\psi}_{mat} = (X'X)^{-1}X'z_{mat}$ el estimador mínimos cuadrados de ψ_{mat} supuesto conocido el valor de θ .

Premultiplicando por $(L^2)^{-1}L^2$ la matriz $F - B\tilde{\psi}_{mat}$ en (3.31), obtenemos la distribución predictiva para la matriz P_{mat} de los datos futuros:

$$f(P_{mat}|y, \theta) \propto |S + (P_{mat} - \mu_{P_{mat}})'(L^2)'^{-1}C(L^2)^{-1}(P_{mat} - \mu_{P_{mat}})|^{-\frac{n+h-s-1}{2}}$$

que se trata de una t-Student generalizada, en concreto,

$$f(P_{mat}|y, \theta) = GSt((L^2)'^{-1}C(L^2)^{-1}, S, \mu_{P_{mat}}, n + h - s - 1) \quad (3.32)$$

donde $\mu_{P_{mat}} = L^{21}z_{mat} + (M_p - L^{21}X)\tilde{\psi}_{mat}$.

De (3.32) se deduce que la distribución predictiva marginal de cada una de las columnas de la matriz P_{mat} es t-Student multivariante, en particular:

$$f(P_i|y, \theta) = MSt\left(\mu_{P_{mat}}[i], \frac{(L^2)'^{-1}C(L^2)^{-1}}{S[i, i]}, \nu = n - m - s\right) \quad (3.33)$$

con vector de medias y matriz de covarianzas:

$$E(P_i|y, \theta) = \mu_{P_i} = L^{21} z_i + (M_p - L^{21} X)(X' X)^{-1} X' z_i \quad (3.34)$$

$$V(P_i|y, \theta) = V_{P_i} = \frac{S[i, i]}{\nu - 2} L^2 C^{-1} (L^2)' \quad (3.35)$$

En el caso general, θ desconocido, la distribución predictiva puede ser estimada mediante integración por Monte Carlo. Sea $\{\theta^{(j)}\}_{j=1}^N$ una muestra de la distribución a posteriori $f(\theta|y)$, ecuación (3.25). Entonces

$$f(P_i|y) \approx \frac{1}{N} \sum_{j=1}^N MSt^{(j)} \left(\mu_{P_i}, \frac{(L^2)'^{-1} C (L^2)^{-1}}{S[i, i]}, n - m - s \right) \quad (3.36)$$

donde el superíndice (j) indica que los cálculos han sido realizados para el valor $\theta = \theta^{(j)}$. Del mismo modo, los principales momentos de la distribución predictiva pueden ser estimados como

$$E(P_i|y) \approx \frac{1}{N} \sum_{j=1}^N \mu_{P_i}^{(j)} \quad (3.37)$$

$$V(P_i|y) \approx \frac{1}{N} \sum_{j=1}^N V_{P_i}^{(j)} + \frac{1}{N} \sum_{j=1}^N \mu_{P_i}^{(j)} \left(\mu_{P_i}^{(j)} \right)' - E(P_i|y) E'(P_i|y) \quad (3.38)$$

Los intervalos de predicción pueden ser calculados a partir de los correspondientes cuantiles de la mixtura de distribuciones t-Student, siguiendo el procedimiento descrito para el modelo general.

3.2.4. Ejemplo de ocupación hotelera (continuación)

En esta sección mostramos los resultados obtenidos cuando las series temporales de ocupación hotelera en Castellón, Valencia y Alicante son analizadas conjuntamente mediante el modelo de Holt-Winters multivariante con parámetros de suavizado iguales. Los resultados son comparados con los obtenidos en los análisis univariantes de las series, sección 1.2, y en el análisis multivariante mediante el modelo de Holt-Winters general, secciones 2.2.4 y 2.3.1.

La figura 3.1 muestra los histogramas de los $N = 10000$ parámetros de suavizado simulados de su distribución a posteriori, $f(\theta|y)$, utilizando el método de aceptación-rechazo descrito previamente tomando $K = 4$. Las líneas rojas verticales representan las medias muestrales.

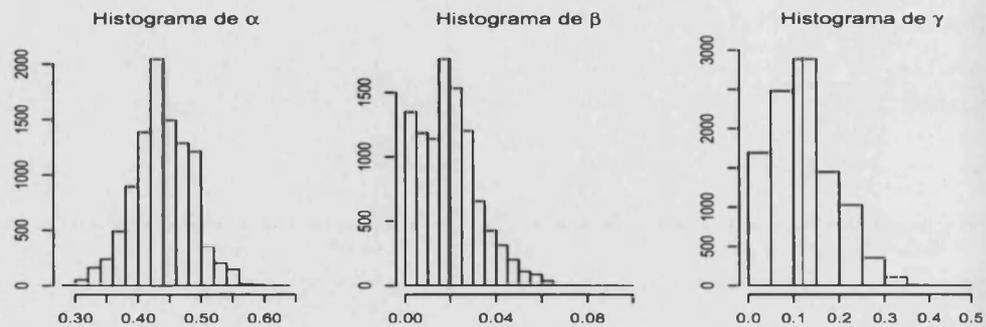


Figura 3.1: Histogramas de los parámetros de suavizado simulados de su distribución a posteriori cuando las series de ocupación hotelera son analizadas mediante el modelo de Holt-Winters multivariante con parámetros de suavizado iguales. Líneas rojas verticales: medias muestrales

Las densidades estimadas de los coeficientes de correlación, simulados como (3.26) utilizando la muestra $\{\theta^{(j)}\}_{j=1}^{10000}$ obtenida, pueden observarse en la figura 3.2. Las líneas verticales representan las medias y los intervalos de probabilidad a posteriori del 95 % de las distribuciones. Como puede observarse, las medias muestrales son ligeramente menores que las obtenidas en el análisis del modelo multivariante general (ver tabla 2.4).

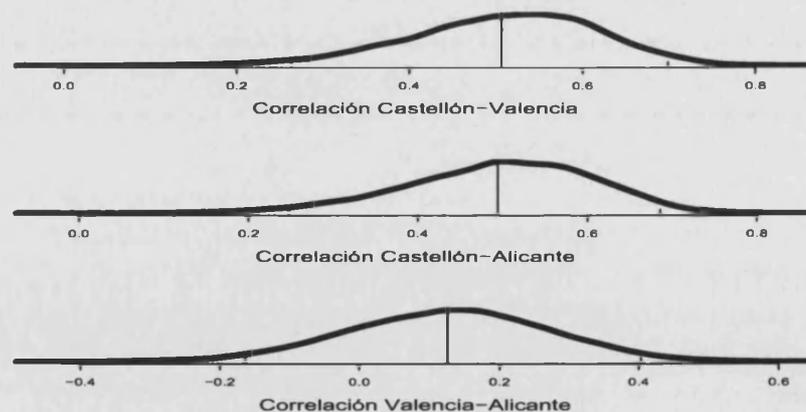


Figura 3.2: Densidades estimadas de los coeficientes de correlación de las series temporales de ocupación hotelera. Las líneas verticales representan las medias y los intervalos de probabilidad a posteriori del 95 %

Finalmente, en la tabla 3.1 se presentan los errores de ajuste y de predicción resultantes del análisis multivariante. Las estimaciones de los datos

que conforman cada una de las series temporales univariantes son obtenidas, nuevamente, a partir de las ecuaciones que definen los modelos univariantes, $y_i = M\psi_i + L\varepsilon_i$ para $i = 1, 2, 3$, sustituyendo los parámetros por sus medias muestrales a posteriori. Las predicciones puntuales coinciden con la media estimada de la distribución predictiva marginal, ver ecuación (3.37).

	Castellón	Valencia	Alicante	Media
Error ajuste	8.78	8.89	6.06	7.91
Error predicción	12.95	37.04	18.77	22.92

Tabla 3.1: Error SMAPE de ajuste y de predicción cuando las series temporales de ocupación hotelera son analizadas mediante el modelo de Holt-Winters multivariante con parámetros de suavizado iguales

A la vista de los resultados obtenidos concluimos que, al analizar las series temporales con el modelo de Holt-Winters multivariante con parámetros de suavizado iguales, mejoramos el ajuste obtenido pero empeoramos la predicción con respecto a los análisis univariantes y al análisis del modelo multivariante general (ver tablas 1.2 y 2.5).

Por otro lado, con el uso del modelo de Holt-Winters multivariante con parámetros de suavizado iguales, el tiempo requerido para el análisis de las series es reducido, aproximadamente, en un 73 % respecto a los análisis univariantes. Ésto es debido a que en los análisis univariantes es necesario obtener una muestra de la distribución a posteriori de cada uno de los vectores de parámetros de suavizado, $f(\theta_i|y_i)$ para $i = 1, 2, 3$, mientras que en el análisis multivariante sólo se obtiene una muestra de la distribución a posteriori del vector de los parámetros de suavizado comunes, $f(\theta|y)$. Por lo tanto, cuando estudiamos series temporales con errores correlacionados y se cumple que los parámetros de suavizado de los modelos univariantes son iguales, el análisis multivariante es más rápido que los análisis univariantes, siendo posible además mejorar el ajuste y/o la predicción.

3.2.5. Análisis Bayesiano de referencia

En los apartados anteriores, sección 3.2.1 - sección 3.2.3, mostramos el análisis Bayesiano del modelo de Holt-Winters multivariante bajo el supuesto de parámetros de suavizado iguales cuando partimos de la distribución a priori de Jeffreys. Sin embargo, al igual que hemos hecho para el modelo general, es posible partir de una distribución a priori de referencia. En esta

sección obtenemos tanto la distribución a posteriori de los parámetros del modelo como la distribución predictiva cuando partimos de una distribución a priori de referencia para la matriz de covarianzas. Para el vector ψ de las condiciones iniciales y el vector θ de los parámetros de suavizado consideramos nuevamente como distribución a priori la uniforme.

Tomando como priori para Σ la distribución dada por (2.31), la distribución a priori conjunta es de la forma:

$$f(\psi_{mat}, \Sigma, \theta) \propto \frac{1}{|\Sigma|^{\frac{m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{\frac{1}{2}}}$$

Entonces, la distribución a posteriori de los parámetros del modelo, proporcional al producto de la priori por la verosimilitud, viene dada por:

$$\begin{aligned} f(\psi_{mat}, \Sigma, \theta|y) &\propto |\Sigma|^{-\frac{n+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} \\ &\quad \exp \left\{ -\frac{1}{2} \text{tr} \left[\Sigma^{-1} (z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat}) \right] \right\} \\ &\propto |\Sigma|^{-\frac{n+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} \\ &\quad \exp \left\{ -\frac{1}{2} \text{tr} \left[\Sigma^{-1} z'_{mat} (I_n - P_X) z_{mat} \right] \right\} \\ &\quad \exp \left\{ -\frac{1}{2} \text{tr} \left[\Sigma^{-1} (\psi_{mat} - \tilde{\psi}_{mat})' X' X (\psi_{mat} - \tilde{\psi}_{mat}) \right] \right\} \end{aligned} \quad (3.39)$$

de donde se deducen las siguientes distribuciones condicionales a posteriori:

$$f(\psi_{mat}|y, \Sigma, \theta) = N_{(s+1),m}(\tilde{\psi}_{mat}, \Sigma \otimes (X'X)^{-1}) \quad (3.40)$$

$$\begin{aligned} f(\Sigma|y, \psi_{mat}, \theta) &\propto |\Sigma|^{-\frac{n+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} \\ &\quad \exp \left\{ -\frac{1}{2} \text{tr} \left[\Sigma^{-1} (z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat}) \right] \right\} \end{aligned} \quad (3.41)$$

$$f(\theta|y, \psi_{mat}, \Sigma) \propto \exp \left\{ -\frac{1}{2} \text{tr} \left[\Sigma^{-1} (z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat}) \right] \right\} \quad (3.42)$$

siendo (3.40) equivalente a

$$f(\psi|y, \Sigma, \theta) = N_{(s+1)m}(\text{vec}(\tilde{\psi}_{mat}), \Sigma \otimes (X'X)^{-1}) \quad (3.43)$$

Notar que, cuando partimos de una distribución a priori de referencia, no es posible obtener las distribuciones a posteriori $f(\psi_{mat}|y, \theta)$ y $f(\Sigma|y, \theta)$

y simular únicamente de la marginal de θ , $f(\theta|y)$, para obtener inferencias acerca de los parámetros del modelo. En este caso, para estimar la distribución a posteriori es necesario utilizar un mecanismo de Metropolis-within-Gibbs similar al utilizado en el análisis del modelo general. En concreto, cada iteración del algoritmo que proponemos consta de los siguientes pasos:

1. Simular $\Sigma^* \sim IW(n-1, (z_{mat} - X\psi_{mat})'(z_{mat} - X\psi_{mat})) = g(\Sigma)$
2. Simular $r_1 \sim Un(0, 1)$. Si $r_1 \leq 2^{\frac{m}{2}} |I_m + \Sigma^* \circ (\Sigma^*)^{-1}|^{-\frac{1}{2}}$, fijar $\Sigma^{(i)} = \Sigma^*$. En caso contrario, volver al paso 1.
3. Simular $\psi^{(i)} \sim f(\psi|y, \Sigma^{(i)}, \theta^{(i-1)})$, dada por (3.43)
4. Dado $\theta^{(i-1)} = (\alpha^{(i-1)}, \beta^{(i-1)}, \gamma^{(i-1)})$, simular $\theta^{(i)}$ mediante el siguiente algoritmo de Metropolis:
 - a) Simular θ^* utilizando como función generadora la uniforme en el cubo centrado en $\theta^{(i-1)}$ y lado $2l$.
 - b) Calcular $\alpha(\theta^{(i-1)}, \theta^*) = \min \left\{ 1, \frac{f(\theta^*|y, \psi^{(i)}, \Sigma^{(i)})}{f(\theta^{(i-1)}|y, \psi^{(i)}, \Sigma^{(i)})} \right\}$ utilizando (3.42)
 - c) Obtener $r_2 \sim Un(0, 1)$
 - d) Si $r_2 \leq \alpha(\theta^{(i-1)}, \theta^*)$, fijar $\theta^{(i)} = \theta^*$. En otro caso, $\theta^{(i)} = \theta^{(i-1)}$

Como valores de inicio $\psi^{(0)}$ y $\theta^{(0)}$ necesarios para la implementación del algoritmo utilizamos $\psi_k^{(0)} = (0, y_{1k}, y_{2k}, \dots, y_{sk})'$, para $k = 1, 2, \dots, m$, y $\theta^{(0)} = (0.5, 0.5, 0.5)'$. Como valor de l tomamos 0.1. Repitiendo el proceso $N_0 + N$ veces y eliminando las N_0 primeras iteraciones de calentamiento, obtenemos una muestra $\{(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)})\}_{j=1}^N$ de la distribución a posteriori de los parámetros del modelo, ecuación (3.39).

Una vez calculada la distribución a posteriori de los parámetros del modelo, podemos obtener la distribución predictiva resolviendo la siguiente integral (ver ecuaciones (3.28) y (3.29)):

$$\begin{aligned}
 f(P_{mat}|y) &= \int \int \int f(P_{mat}|y, \psi_{mat}, \Sigma, \theta) f(\psi_{mat}, \Sigma, \theta|y) d\psi_{mat} d\Sigma d\theta \\
 &\propto \int \int \int |\Sigma|^{-\frac{n+h+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} \\
 &\quad \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (F - B\psi_{mat})' (F - B\psi_{mat})] \right\} \\
 &\quad \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat})] \right\} d\psi_{mat} d\Sigma d\theta
 \end{aligned}$$

El término $|I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}}$ que aparece en la integral hace que no tengamos una distribución Wishart invertida para la matriz Σ y, por tanto, a diferencia de lo que ocurre cuando partimos de la priori de Jeffreys, no podemos integrar respecto a Σ . Para poder integrar respecto a ψ consideramos la expresión alternativa de la distribución a posteriori de los parámetros del modelo (ver ecuación (3.39)). De este modo:

$$f(P_{mat}|y) \propto \int \int \int |\Sigma|^{-\frac{n+h+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}(F - B\psi_{mat})'(F - B\psi_{mat})] \right\} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}z'_{mat}(I_n - P_X)z_{mat}] \right\} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}(\psi_{mat} - \tilde{\psi}_{mat})'X'X(\psi_{mat} - \tilde{\psi}_{mat})] \right\} d\psi_{mat} d\Sigma d\theta$$

Agrupando a continuación las dos formas cuadráticas que involucran a ψ_{mat} como:

$$\begin{aligned} & (F - B\psi_{mat})'(F - B\psi_{mat}) + (\psi_{mat} - \tilde{\psi}_{mat})'X'X(\psi_{mat} - \tilde{\psi}_{mat}) \\ &= F'F - F'B\psi_{mat} - \psi'_{mat}B'F + \psi'_{mat}B'B\psi_{mat} \\ &+ \psi'_{mat}X'X\psi_{mat} - \psi'_{mat}X'X\tilde{\psi}_{mat} - \tilde{\psi}'_{mat}X'X\psi_{mat} + \tilde{\psi}'_{mat}X'X\tilde{\psi}_{mat} \\ &= (\psi_{mat} - \psi^*)'W(\psi_{mat} - \psi^*) + F'F + \psi'_{mat}X'X\tilde{\psi}_{mat} - (\psi^*)'W\psi^* \end{aligned}$$

donde $W = X'X + B'B$ y $\psi^* = W^{-1}(X'X\tilde{\psi}_{mat} + B'F)$, la distribución predictiva viene dada por:

$$\begin{aligned} f(P_{mat}|y) &\propto \int \int \int |\Sigma|^{-\frac{n+h+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} \\ &\exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}z'_{mat}(I_n - P_X)z_{mat}] \right\} \\ &\exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}(\psi_{mat} - \psi^*)'W(\psi_{mat} - \psi^*)] \right\} \\ &\exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}(F'F + \tilde{\psi}'_{mat}X'X\tilde{\psi}_{mat} - (\psi^*)'W\psi^*)] \right\} d\psi_{mat} d\Sigma d\theta \\ &\propto \int \int |\Sigma|^{-\frac{n-s-1+h+m}{2}} |I_m + \Sigma \circ \Sigma^{-1}|^{-\frac{1}{2}} |W|^{-\frac{m}{2}} \\ &\exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}z'_{mat}(I_n - P_X)z_{mat}] \right\} \\ &\exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}(F'F + \tilde{\psi}'_{mat}X'X\tilde{\psi}_{mat} - (\psi^*)'W\psi^*)] \right\} d\Sigma d\theta \end{aligned}$$

que no tiene solución analítica. Por lo tanto, eliminando aquellos términos que no dependen de P , se tiene:

$$f(P_{mat}|y, \Sigma, \theta) \propto \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}(F'F - (\psi^*)'W\psi^*)] \right\}$$

Completando cuadrados en F como:

$$\begin{aligned} F'F - (\psi^*)'W\psi^* &= F'F - (\tilde{\psi}'_{mat}X'X + F'B)W^{-1}(X'X\tilde{\psi}_{mat} + B'F) \\ &= (F - \mu_F)'V_F(F - \mu_F) - \tilde{\psi}'_{mat}X'XW^{-1}X'X\tilde{\psi}_{mat} - \mu'_F V_F \mu_F \end{aligned}$$

siendo $V_F = I_h - BW^{-1}B'$ y $\mu_F = V_F^{-1}BW^{-1}X'X\tilde{\psi}_{mat}$, y eliminando nuevamente los términos que no dependen de P , la distribución predictiva condicional es de la forma:

$$\begin{aligned} f(P_{mat}|y, \Sigma, \theta) &\propto \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}(F - \mu_F)'V_F(F - \mu_F)] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}(F - \mu_F)'(L^2)'(L^2)^{-1}V_F(L^2)^{-1}L^2(F - \mu_F)] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1}(P_{mat} - \mu_{P_{mat}})'(L^2)'(L^2)^{-1}V_F(L^2)^{-1}(P_{mat} - \mu_{P_{mat}})] \right\} \end{aligned}$$

donde

$$\begin{aligned} \mu_{P_{mat}} &= L^{21}z_{mat} + L^2\mu_F \\ &= L^{21}z_{mat} + L^2V_F^{-1}BW^{-1}X'X\tilde{\psi}_{mat} \\ &= L^{21}z_{mat} + L^2V_F^{-1}BW^{-1}X'z_{mat} \end{aligned}$$

y, por tanto, se deduce que la matriz P_{mat} de los datos futuros sigue una distribución matriz Normal, en concreto,

$$f(P_{mat}|y, \Sigma, \theta) = N_{h,m}((L^{21} + L^2V_F^{-1}BW^{-1}X')z_{mat}, \Sigma \otimes L^2V_F^{-1}(L^2)')$$

o, equivalentemente,

$$f(P||y, \Sigma, \theta) = N_{hm}((I_m \otimes (L^{21} + L^2V_F^{-1}BW^{-1}X'))z, \Sigma \otimes L^2V_F^{-1}(L^2)') \quad (3.44)$$

Para poder calcular tanto estimaciones puntuales como intervalos de predicción para los datos futuros es necesario estimar la distribución predictiva marginal. Para ello, dada $\{(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)})\}_{j=1}^N$ una muestra de la distribución a posteriori de los parámetros del modelo, obtenida mediante el mecanismo de Metropolis-within-Gibbs descrito anteriormente, aplicamos

integración por Monte Carlo de manera similar a como hemos mostrado en secciones anteriores.

Como puede verse, el uso de la priori de referencia complica el análisis del modelo de Holt-Winters multivariante en el caso de parámetros de suavizado iguales. Para poder estimar la distribución a posteriori de los parámetros del modelo es necesario emplear un mecanismo de Metropolis-within-Gibbs que nos permita obtener una muestra de la posteriori conjunta $f(\psi, \Sigma, \theta|y)$. Dicho mecanismo es más lento que el método de aceptación-rechazo que empleamos para simular de $f(\theta|y)$ cuando partimos de la priori de Jeffreys y, además, puede tener problemas de convergencia. Por otro lado, la distribución predictiva depende no sólo de θ sino también de Σ , lo cual puede provocar una mayor variabilidad en las estimaciones obtenidas. Con todo ésto, cuando el objetivo principal del análisis no sea obtener inferencias acerca de los parámetros del modelo sino el cálculo de predicciones, y bajo el supuesto de parámetros de suavizado comunes, la priori de Jeffreys puede resultar más conveniente. Además, del análisis de las series temporales de ocupación hotelera mediante el modelo general podemos suponer que, para el análisis multivariante de un conjunto moderado de series temporales correlacionadas, el uso de la priori de Jeffreys o la priori de referencia no ocasiona cambios significativos en los resultados obtenidos.

Fijando s , periodo del ciclo estacional, igual a 1 y $\psi_i = (b_{0i}, c_{0i})'$, con $c_{0i} = a_{0i} + b_{0i}$ para $i = 1, 2, \dots, m$, los resultados obtenidos son aplicables al modelo de Holt multivariante en el caso de parámetros de suavizado iguales. Notar que en este caso $\theta = (\alpha, \beta)$ es de dimensión 2 y, por tanto, en su simulación trabajaremos con el cuadrado unidad en lugar del cubo unidad.

3.3. Notas computacionales

Siguiendo la estructura del capítulo 2, dedicamos esta última sección al desarrollo de resultados que han sido utilizados en la implementación de los algoritmos descritos en este capítulo, así como las rutinas de R necesarias.

Cuando analizamos el modelo de Holt-Winters multivariante en el caso de parámetros de suavizado iguales tenemos que calcular la matriz $(X'X)^{-1}$, que aparece en las distribuciones a posteriori $f(\psi_{mat}|y, \Sigma, \theta)$ y $f(\Sigma|y, \theta)$, ecuaciones (3.19) y (3.20) respectivamente. Para evitar los posibles problemas numéricos que dicho cálculo puede ocasionar, puede hacerse uso de la

descomposición en valores singulares de la matriz X . De esta manera, si $X = UDV'$, la matriz $(X'X)^{-1} = VD^{-1}D^{-1}V'$. La matriz $(X'X)^{-1}X'$ que aparece en el cálculo del estimador mínimos cuadrados de ψ puede ser calculada como $VD^{-1}U'$ y la matrix P_X como UU' .

Por otro lado, en el método de aceptación-rechazo propuesto para simular de la distribución a posteriori marginal del vector θ de parámetros de suavizado, sección 3.2.2, hacemos uso del método 'L-BFGS-B'. La rutina empleada en R para utilizar este método es:

`optim(par, f, method = 'L - BFGS - B', lower, upper)`

donde *par* es el vector con los valores de inicio de los parámetros a optimizar, *f* es la función a optimizar y *lower* y *upper* son, respectivamente, los vectores de cotas inferiores y superiores de los parámetros a optimizar. La salida que se obtiene es una lista con los siguientes elementos: *par*, el mejor valor encontrado para los parámetros; *value*, el valor de la función objetivo correspondiente a *par* y *convergence*, que toma el valor 0 si se ha alcanzado la convergencia y otro valor en caso contrario.

Finalmente, para el cálculo de la matriz V_F^{-1} que aparece en la distribución predictiva condicional cuando se parte de la priori de referencia, ecuación (3.44), puede hacerse uso del resultado de Mardia et al. (1979, pag. 459) visto anteriormente. De este modo,

$$V_F^{-1} = (I_h - BW^{-1}B')^{-1} = I_h + B(W - B'B)^{-1}B' = I_h + B(X'X)^{-1}B'$$

**CAPÍTULO 3. HOLT-WINTERS MULTIVARIANTE: CASOS
88 PARTICULARES**

Capítulo 4

Selección de modelos

Los avances computacionales en las últimas décadas han permitido el desarrollo de nuevos modelos estadísticos, en algunas ocasiones de gran complejidad, capaces de adaptarse a casi cualquier situación práctica. Dados unos datos de estudio, en la actualidad es común considerar un conjunto de posibles modelos, $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_K$, donde cada modelo representa una familia de distribuciones $\{f(y|\varphi_k, \mathcal{M}_k)\}$, lo que permite seleccionar aquel modelo que mejor se ajusta a los datos para su posterior análisis. Así pues, el problema de selección de modelos, al ser una parte fundamental de cualquier análisis estadístico, ha sido objeto de estudio tanto desde la perspectiva frecuentista como Bayesiana. En los últimos años, debido al avance de los métodos de simulación y dado que el enfoque Bayesiano permite tratar de manera conjunta la incertidumbre del modelo así como la de los parámetros de éste, los métodos Bayesianos para el problema de selección de modelos han experimentado un gran desarrollo (Clyde and George, 2003).

La principal situación en la que surge el problema de selección es en la construcción de un modelo estadístico adecuado para representar el comportamiento de los datos de estudio. Sin embargo, hay otras razones por las que se plantea la selección de modelos. Por ejemplo cuando el modelo actual es demasiado complicado y, por razones de parsimonia, se examinan modelos más sencillos para ver si se ajustan a los datos suficientemente, o cuando unos pocos modelos son propuestos porque se ajustaban adecuadamente a muestras previas y se desea ver cual es el que mejor se ajusta a la muestra actual, etc. No obstante, algunos autores prefieren desestimar aquellos modelos que claramente no se ajustan a los datos y mantener un subconjunto para futuras consideraciones.

En nuestro caso, dadas las observaciones que conforman las series temporales, suponemos que cada una de las series univariantes se ajusta al modelo de Holt-Winters univariante y que existe una correlación contemporánea entre errores correspondientes en los modelos univariantes. Considerando las series temporales conjuntamente obtenemos el modelo de Holt-Winters multivariante, ecuación (2.3). A partir de distribuciones a priori convencionales no informativas obtenemos la distribución a posteriori de los parámetros del modelo. Dicha distribución, aunque no es analíticamente tratable, puede ser estimada mediante simulación. En particular, proponemos utilizar un algoritmo de Metropolis-within-Gibbs, sección 2.2.2, que nos permite simular de las condicionales completas a posteriori. Este algoritmo, además de poder presentar problemas de convergencia, requiere de un tiempo de computación elevado.

En el caso de parámetros de suavizado iguales para los modelos univariantes, el modelo de Holt-Winters multivariante puede formularse como un modelo de regresión multivariante tradicional, ecuación (3.14), lo que simplifica su análisis considerablemente. La distribución a posteriori para los parámetros del modelo sigue siendo analíticamente intratable pero, en este caso, sólo es necesario obtener una muestra de la distribución a posteriori marginal del vector θ de los parámetros de suavizado comunes, ecuación (3.25). Las distribuciones a posteriori marginales para el vector de las condiciones iniciales ψ y la matriz de covarianzas Σ son estimadas entonces mediante integración por Monte Carlo. Por otro lado, para obtener una muestra de $f(\theta|y)$, al tratarse de una función tridimensional con soporte acotado, el cubo unidad, es posible emplear un algoritmo de aceptación-rechazo que, además de no presentar problemas de convergencia, es fácil de implementar y reduce el tiempo de computación considerablemente. Por ello, dados los valores que conforman las series temporales, el primer paso del análisis es comprobar si el modelo con parámetros de suavizado iguales se adecua a los datos; es decir, seleccionar entre el modelo de Holt-Winters multivariante general y el modelo con parámetros de suavizado iguales.

Dado que no existe ningún método de selección de modelos que haya demostrado ser superior ante cualquier situación práctica, la elección del método a utilizar depende de la naturaleza del problema. En las distintas secciones de este capítulo intentamos aplicar las principales técnicas Bayesianas de selección de modelos: factores Bayes, técnicas basadas en las propiedades predictivas de los modelos y criterios de selección (ver sección

1.4), para resolver nuestro problema de selección de modelos. Como veremos a continuación, la mayoría de estos métodos resultan difíciles de implementar o requieren de un tiempo de computación excesivo.

Denotamos por \mathcal{M}_0 el modelo de Holt-Winters multivariante general y por \mathcal{M}_1 al modelo con parámetros de suavizado iguales. Suponemos que los dos modelos son a priori igualmente probables, es decir, $\pi_0 = \pi_1 = 1/2$. Como distribuciones a priori para los parámetros del modelo consideramos:

$$f_i(\psi, \Sigma, \theta) \propto |\Sigma|^{-\frac{m+1}{2}} \quad \text{para } i = 0, 1 \quad (4.1)$$

donde la única diferencia entre los dos modelos es que bajo \mathcal{M}_0 , el vector θ de parámetros de suavizado es de la forma $(\theta'_1, \theta'_2, \dots, \theta'_m)'$, con $\theta_i = (\alpha_i, \beta_i, \gamma_i)'$ el vector de parámetros de suavizado asociado a la serie temporal i -ésima, mientras que bajo \mathcal{M}_1 , $\theta = (\alpha, \beta, \gamma)'$ es el vector de los parámetros de suavizado comunes para todos los modelos univariantes. Con el fin de evitar posibles confusiones denotaremos por θ_0 el vector $3m \times 1$ de parámetros de suavizado bajo \mathcal{M}_0 . Notar que el vector ψ de las condiciones iniciales y la matriz de covarianzas Σ están definidos de la misma manera en ambos modelos.

4.1. Factores Bayes para la selección del modelo de Holt-Winters multivariante

En esta sección nos planteamos resolver el problema de selección de modelos asociado al modelo de Holt-Winters multivariante haciendo uso del factor Bayes B_{01} (sección 1.4.1). Dado que suponemos que los dos modelos tienen la misma probabilidad a priori, el factor Bayes coincide con el cociente de las probabilidades a posteriori de los modelos. El primer problema con el que nos encontramos en su cálculo es que las verosimilitudes marginales, $f(y|\mathcal{M}_i)$ para $i = 0, 1$, no tienen solución analítica. Notar que el uso de prioris convencionales no informativas está justificado, pues el vector ψ de las condiciones iniciales y la matriz Σ de covarianzas están definidos de la misma manera en \mathcal{M}_0 y \mathcal{M}_1 , mientras que las distribuciones a priori para los vectores de los parámetros de suavizado son propias.

Bajo \mathcal{M}_0 , a partir de (2.4), se tiene:

$$\begin{aligned}
f(y|\mathcal{M}_0) &= \int \int \int f_0(y|\psi, \Sigma, \theta_0) f_0(\psi, \Sigma, \theta_0) d\psi d\Sigma d\theta_0 \\
&\propto \int \int \int |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2}(z - X_B\psi)'(\Sigma^{-1} \otimes I_n)(z - X_B\psi) \right\} d\psi d\Sigma d\theta_0 \\
&\propto \int \int \int |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2}(z - X_B\hat{\psi})'(\Sigma^{-1} \otimes I_n)(z - X_B\hat{\psi}) \right\} \\
&\quad \exp \left\{ -\frac{1}{2}(\psi - \hat{\psi})'X_B'(\Sigma^{-1} \otimes I_n)X_B(\psi - \hat{\psi}) \right\} d\psi d\Sigma d\theta_0 \\
&\propto (2\pi)^{\frac{m(s+1)}{2}} \int \int |\Sigma|^{-\frac{n+m+1}{2}} |X_B'(\Sigma^{-1} \otimes I_n)X_B|^{-\frac{1}{2}} \\
&\quad \exp \left\{ -\frac{1}{2}(z - X_B\hat{\psi})'(\Sigma^{-1} \otimes I_n)(z - X_B\hat{\psi}) \right\} d\Sigma d\theta_0 \tag{4.2}
\end{aligned}$$

donde $\hat{\psi} = (X_B'(\Sigma^{-1} \otimes I_n)X_B)^{-1}X_B'(\Sigma^{-1} \otimes I_n)z$ es el estimador mínimos cuadrados generalizado de ψ .

Expresando la función de verosimilitud como en (2.5) e integrando en primer lugar respecto a Σ , obtenemos:

$$\begin{aligned}
f(y|\mathcal{M}_0) &= \int \int \int f_0(y|\psi, \Sigma, \theta_0) f_0(\psi, \Sigma, \theta_0) d\psi d\Sigma d\theta_0 \\
&\propto \int \int \int |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} H' H] \right\} d\psi d\Sigma d\theta_0 \\
&\propto k_1^{-1} \int \int |H' H|^{-\frac{n}{2}} d\psi d\theta_0 \tag{4.3}
\end{aligned}$$

siendo $H = [z_1 - X_1\psi_1, z_2 - X_2\psi_2, \dots, z_m - X_m\psi_m]$ y k_1^{-1} la constante definida como:

$$k_1^{-1} = 2^{\frac{nm}{2}} \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^m \Gamma \left(\frac{n+1-i}{2} \right) \tag{4.4}$$

luego la verosimilitud marginal para el modelo \mathcal{M}_0 no puede ser calculada analíticamente.

Cuando intentamos calcular la verosimilitud marginal bajo \mathcal{M}_1 , a partir

de (3.16), se tiene:

$$\begin{aligned}
 f(y|\mathcal{M}_1) &= \int \int \int f_1(y|\psi, \Sigma, \theta) f_1(\psi, \Sigma, \theta) d\psi d\Sigma d\theta \\
 &\propto \int \int \int |\Sigma|^{-\frac{n+m+1}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} z'_{mat} (I_n - P_X) z_{mat}] \right\} \\
 &\quad \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (\psi_{mat} - \tilde{\psi}_{mat})' X' X (\psi_{mat} - \tilde{\psi}_{mat})] \right\} d\psi_{mat} d\Sigma d\theta \\
 &\propto (2\pi)^{\frac{m(s+1)}{2}} \int \int |\Sigma|^{-\frac{n-s-1+m+1}{2}} |X' X|^{-\frac{m}{2}} \\
 &\quad \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} z'_{mat} (I_n - P_X) z_{mat}] \right\} d\Sigma d\theta \\
 &\propto (2\pi)^{\frac{m(s+1)}{2}} k_2^{-1} \int |X' X|^{-\frac{m}{2}} |z'_{mat} (I_n - P_X) z_{mat}|^{-\frac{n-s-1}{2}} d\theta \quad (4.5)
 \end{aligned}$$

con k_2^{-1} definida como (4.4) sustituyendo n por $n - s - 1$. Luego la verosimilitud marginal bajo \mathcal{M}_1 tampoco tiene solución analítica y, por tanto, es necesario recurrir a la simulación para obtener estimaciones de las verosimilitudes marginales o de las probabilidades a posteriori de los modelos (ver sección 1.4.1, apartado (iii)).

Si intentamos aplicar la metodología de Carlin and Chib (1995) hay que definir las pseudo-prioris $f_0(\psi, \Sigma, \theta)$ y $f_1(\psi, \Sigma, \theta_0)$. Aunque en principio pueda parecer que la forma de dichas distribuciones sea irrelevante, pues únicamente son necesarias para que la distribución a posteriori conjunta de todos los parámetros esté bien definida, posteriormente se destaca su influencia en la convergencia del algoritmo de simulación, aconsejando definir las de modo similar a las distribuciones a posteriori obtenidas a partir de los modelos correspondientes. Para nuestra formulación del modelo de Holt-Winters multivariante, el tiempo requerido para obtener una muestra de la distribución a posteriori de los parámetros del modelo puede ser muy grande. Por tanto, si tomamos las pseudo-prioris similares a las distribuciones a posteriori, además de no tener la certeza de que las distribuciones propuestas sean adecuadas para la convergencia del algoritmo, en cada iteración, independientemente del modelo en el que nos encontremos, el tiempo necesario para la simulación de los parámetros de los modelos es muy grande. Por ello, el uso de este procedimiento para la resolución de nuestro problema de selección de modelos no nos parece aconsejable.

Una solución alternativa es obtener los estimadores por muestreo impor-

tante de las verosimilitudes marginales y obtener así un estimador del factor Bayes. Bajo \mathcal{M}_0 es posible trabajar con (4.2) o con (4.3) indistintamente, definiendo la función importante $cg(\Sigma, \theta_0)$ o $cg(\psi, \theta_0)$ respectivamente. Notar que la distribución importante no puede coincidir con la distribución a priori de los parámetros del modelo pues, al tratarse de una distribución no informativa, excepto para el vector θ_0 de parámetros de suavizado, no es posible obtener muestras de esta distribución. Por este motivo, tampoco es posible tomar como función importante una mixtura de la priori y la posteriori. Por otro lado, descartamos la media armónica pues resulta muy inestable en muchas ocasiones.

A sugerencia de J.O. Berger (comunicación personal) definimos las distribuciones importantes asociadas a \mathcal{M}_0 y \mathcal{M}_1 como sigue. Para estimar $f(y|\mathcal{M}_0)$ consideramos la ecuación (4.2). En lugar de trabajar con la matriz Σ , simétrica y con elementos en la diagonal principal positivos, trabajamos con una matriz \tilde{T} triangular cuyos elementos están definidos en la recta real. Entonces, tomamos como función importante definida sobre los elementos no nulos de \tilde{T} una t-Student multivariante con vector de medias y matriz de covarianzas definidos a partir de una muestra simulada de la distribución a posteriori.

Sea $\Sigma^{-1} = T'T$ la descomposición de Cholesky de la matrix Σ^{-1} , donde T es una matriz triangular superior cuyos elementos en la diagonal principal son positivos, $t_{ii} > 0$ para $i = 1, 2, \dots, m$, y sea \tilde{T} la matriz definida como:

$$\tilde{T} = \begin{pmatrix} \log t_{11} & t_{12} & t_{13} & \dots & t_{1m} \\ 0 & \log t_{22} & t_{23} & \dots & t_{2m} \\ 0 & 0 & \log t_{33} & \dots & t_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \log t_{mm} \end{pmatrix}$$

de manera que \tilde{T} es una matriz triangular superior cuyos elementos no están restringidos, es decir, pertenecen a la recta real. Entonces, considerando (4.2) y teniendo en cuenta que los parámetros del modelo están independientemente distribuidos a priori, la verosimilitud marginal $f(y|\mathcal{M}_0)$ puede expresarse como:

$$\begin{aligned} f(y|\mathcal{M}_0) &\propto (2\pi)^{\frac{m(s+1)}{2}} \int \int f_0(y|\Sigma, \theta_0) f_0(\Sigma) f_0(\theta_0) d\Sigma d\theta_0 \\ &\propto (2\pi)^{\frac{m(s+1)}{2}} \int \int f_0(y|\tilde{T}, \theta_0) \tilde{f}_0(\tilde{T}) f_0(\theta_0) d\tilde{T} d\theta_0 \quad (4.6) \end{aligned}$$

donde:

$$f_0(y|\Sigma, \theta_0) \propto |\Sigma|^{-\frac{n}{2}} |X'_B(\Sigma^{-1} \otimes I_n)X_B|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(z - X_B\hat{\psi})'(\Sigma^{-1} \otimes I_n)(z - X_B\hat{\psi}) \right\} \quad (4.7)$$

$$f_0(\Sigma) = |\Sigma|^{-\frac{m+1}{2}} \quad (4.8)$$

Teniendo en cuenta la serie de transformaciones aplicadas a la matriz Σ hasta obtener \tilde{T} :

$$\Sigma \underset{\sim}{\circlearrowright}^{r_1} \Sigma^{-1} \underset{\sim}{\circlearrowright}^{r_2} T \underset{\sim}{\circlearrowright}^{r_3} \tilde{T}$$

la distribución a priori $\tilde{f}_0(\tilde{T})$ puede calcularse como:

$$\begin{aligned} \tilde{f}_0(\tilde{T}) &= f_0(s(\tilde{T})) \left| \frac{\partial s(\tilde{T})}{\partial \tilde{T}} \right| \\ &= f_0(s(\tilde{T})) \left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right| \left| \frac{\partial \Sigma^{-1}}{\partial T} \right| \left| \frac{\partial T}{\partial \tilde{T}} \right| \end{aligned} \quad (4.9)$$

siendo $r = r_1 \circ r_2 \circ r_3$, donde \circ representa la composición de funciones, y $s = r^{-1}$ la función inversa tal que $s(\tilde{T}) = \Sigma$. Veamos a continuación el valor de los Jacobianos que aparecen en la definición de la priori. Es conocido el siguiente resultado (ver, por ejemplo, Mardia et al., 1979, pag. 35):

$$\left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right| = |\Sigma|^{m+1} \quad (4.10)$$

Por otro lado, $\Sigma^{-1} = T'T$ es de la forma:

$$\begin{bmatrix} t_{11}^2 & t_{11}t_{12} & t_{11}t_{13} & \dots & t_{11}t_{1m} \\ t_{11}t_{12} & t_{12}^2 + t_{22}^2 & t_{12}t_{13} + t_{22}t_{23} & \dots & t_{12}t_{1m} + t_{22}t_{2m} \\ t_{11}t_{13} & t_{12}t_{13} + t_{22}t_{23} & t_{13}^2 + t_{23}^2 + t_{33}^2 & \dots & t_{13}t_{1m} + t_{23}t_{2m} + t_{33}t_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{11}t_{1m} & t_{12}t_{1m} + t_{22}t_{2m} & t_{13}t_{1m} + t_{23}t_{2m} + t_{33}t_{3m} & \dots & t_{1m}^2 + t_{2m}^2 + \dots + t_{mm}^2 \end{bmatrix}$$

Sean $vech(\Sigma^{-1})$ y $vech(T)$ los vectores formados por los elementos no repetidos de Σ^{-1} y los elementos no nulos de T , es decir,

$$\begin{aligned} vech(\Sigma^{-1}) &= (\sigma^{11}, \sigma^{12}, \dots, \sigma^{1m}, \sigma^{22}, \dots, \sigma^{2m}, \dots, \sigma^{mm})' \\ vech(T) &= (t_{11}, t_{12}, \dots, t_{1m}, t_{22}, \dots, t_{2m}, \dots, t_{mm})' \end{aligned}$$

entonces, el Jacobiano $|\partial \Sigma^{-1} / \partial T|$ puede ser calculado como (Henderson and Searle, 1979):

$$\left| \frac{\partial \Sigma^{-1}}{\partial T} \right| = \left| \frac{\partial vech(\Sigma^{-1})}{\partial vech(T)} \right|$$

donde el elemento en la fila i y columna j de la matriz Jacobiana es $\frac{d\text{vech}(\Sigma^{-1})[i]}{d\text{vech}(T)[j]}$.

Notar que dicha matriz es una matriz triangular inferior, pues elemento $\text{vech}(\Sigma^{-1})[i]$ sólo involucra términos hasta $\text{vech}(T)[i]$, siendo la diagonal principal el vector

$$\begin{aligned} & \left(\frac{d\sigma^{11}}{dt_{11}}, \frac{d\sigma^{12}}{dt_{12}}, \dots, \frac{d\sigma^{1m}}{dt_{1m}}, \frac{d\sigma^{22}}{dt_{22}}, \dots, \frac{d\sigma^{2m}}{dt_{2m}}, \dots, \frac{d\sigma^{mm}}{dt_{mm}} \right)' \\ & = (2t_{11}, t_{11}, \dots, t_{11}, 2t_{22}, \dots, t_{22}, \dots, 2t_{mm})' \end{aligned}$$

de donde se deduce que

$$\left| \frac{\partial \Sigma^{-1}}{\partial T} \right| = \left| \frac{\partial \text{vech}(\Sigma^{-1})}{\partial \text{vech}(T)} \right| = 2^m \prod_{i=1}^m t_{ii}^{m-i+1} \quad (4.11)$$

De manera análoga, el Jacobiano $|\partial T / \partial \tilde{T}|$ puede ser calculado a partir de los vectores $\text{vech}(T)$ y $\text{vech}(\tilde{T})$. En este caso, la matriz Jacobiana es una matriz diagonal, cuya diagonal es el vector

$$\begin{aligned} & \left(\frac{dt_{11}}{d\tilde{t}_{11}}, \frac{dt_{12}}{d\tilde{t}_{12}}, \dots, \frac{dt_{1m}}{d\tilde{t}_{1m}}, \frac{dt_{22}}{d\tilde{t}_{22}}, \dots, \frac{dt_{2m}}{d\tilde{t}_{2m}}, \dots, \frac{dt_{mm}}{d\tilde{t}_{mm}} \right)' \\ & = (\exp\{\tilde{t}_{11}\}, 1, \dots, 1, \exp\{\tilde{t}_{22}\}, \dots, 1, \dots, \exp\{\tilde{t}_{mm}\})' \end{aligned}$$

por lo que

$$\left| \frac{\partial T}{\partial \tilde{T}} \right| = \exp \left\{ \sum_{i=1}^m \tilde{t}_{ii} \right\} \quad (4.12)$$

Sustituyendo (4.10), (4.11) y (4.12) en (4.9), y teniendo en cuenta que $|\Sigma| = (\prod_{i=1}^m t_{ii})^{-2} = \exp\{-2 \sum_{i=1}^m \tilde{t}_{ii}\}$, se tiene finalmente que la distribución a priori para \tilde{T} es de la forma:

$$\tilde{f}_0(\tilde{T}) = 2^m \exp \left\{ \sum_{i=1}^m (1-i)\tilde{t}_{ii} \right\} \quad (4.13)$$

por lo que $f(y|\mathcal{M}_0)$ puede expresarse como (ver (4.6)):

$$f(y|\mathcal{M}_0) \propto (2\pi)^{\frac{m(s+1)}{2}} \int \int f_0(y|\tilde{T}, \theta_0) \frac{\tilde{f}_0(\tilde{T}) f_0(\theta_0)}{cg(\tilde{T}, \theta_0)} cg(\tilde{T}, \theta_0) d\tilde{T} d\theta_0 \quad (4.14)$$

donde $f_0(y|\tilde{T}, \theta_0)$ es proporcional a:

$$\exp \left\{ n \sum_{i=1}^m \tilde{t}_{ii} \right\} |X'_B(T'T \otimes I_n)X_B|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(z - X_B\hat{\psi})'(T'T \otimes I_n)(z - X_B\hat{\psi}) \right\} \quad (4.15)$$

con T la matriz triangular superior tal que $t_{ii} = \exp\{\tilde{t}_{ii}\}$ para $i = 1, 2, \dots, m$ y $t_{ij} = \tilde{t}_{ij}$ para $i \neq j$; $f_0(\tilde{T})$ es de la forma (4.13); $f_0(\theta_0) \propto 1$ y $cg(\tilde{T}, \theta_0)$ está definida de la siguiente manera:

$$cg(\tilde{T}, \theta_0) = g_1(\tilde{T})g_2(\theta_0) \\ \propto \left[1 + (vech(\tilde{T}) - T^*)' \frac{\bar{V}^{-1}}{\nu - 2} (vech(\tilde{T}) - T^*) \right]^{-\frac{m(m+1)+\nu}{2}}$$

es decir, para los elementos no nulos de la matriz \tilde{T} , $vech(\tilde{T})$, la función importante es una t-Student multivariante con vector de medias T^* y matriz de covarianzas \bar{V} , siendo T^* y \bar{V} el vector de medias y la matriz de covarianzas muestrales calculados a partir de una muestra de la distribución a posteriori de \tilde{T} . Como función importante para θ_0 consideramos la distribución a priori, es decir, la uniforme en $[0, 1]^{3m}$ pues, en nuestra experiencia (ver tabla 2.2 para el ejemplo de ocupación hotelera), la distribución a posteriori del vector θ_0 de parámetros de suavizado no es muy informativa. Entonces, si $\{(\tilde{T}^{(j)}, \theta_0^{(j)})\}_{j=1}^N$ es una muestra obtenida de la función importante $cg(\tilde{T}, \theta_0)$, la verosimilitud marginal puede ser estimada como (ver ecuación (4.14)):

$$\hat{f}_{IS}(y|\mathcal{M}_0) \propto (2\pi)^{\frac{m(s+1)}{2}} \frac{1}{N} \sum_{j=1}^N f_0(y|\tilde{T}^{(j)}, \theta_0^{(j)}) \frac{\tilde{f}_0(\tilde{T}^{(j)})}{g_1(\tilde{T}^{(j)})} \quad (4.16)$$

Bajo \mathcal{M}_1 es necesario definir una nueva función importante $cg(\theta)$. Por analogía con \mathcal{M}_0 , proponemos tomar $cg(\theta) = g_2(\theta) \propto 1$. De este modo:

$$\hat{f}_{IS}(y|\mathcal{M}_1) \propto (2\pi)^{\frac{m(s+1)}{2}} k_2^{-1} \frac{1}{N} \sum_{j=1}^N f_1(y|\theta^{(j)}) \quad (4.17)$$

siendo k_2^{-1} la constante definida como (4.4) donde el valor de n es sustituido por $n - s - 1$; $\{\theta^{(j)}\}_{j=1}^N$ es una muestra obtenida a partir de la distribución uniforme en el cubo unidad y $f_1(y|\theta)$ es de la forma (ver(4.5)):

$$f_1(y|\theta) = |X'X|^{-\frac{m}{2}} |z'_{mat}(I_n - P_X)z_{mat}|^{-\frac{n-s-1}{2}}$$

Tanto para el modelo de Holt-Winters multivariante general como para el modelo con parámetros de suavizado iguales, el estimador de Laplace, al requerir las derivadas parciales del logaritmo de la verosimilitud, es difícil de implementar. Por otro lado, el estimador de Laplace-Metropolis obtenido al sustituir menos la inversa del Hessiano del logaritmo de la posteriori evaluada en la moda a posteriori, $\hat{\Sigma}$, por la matriz de covarianzas muestral puede ser poco preciso. Por ello, preferimos emplear el estimador por muestreo importante, donde la matriz de covarianzas muestral es utilizada para definir la función importante en lugar de estimar directamente las cantidades de interés.

4.2. Propiedades predictivas para la selección del modelo de Holt-Winters multivariante

Una herramienta alternativa para la resolución del problema de selección de modelos asociado al modelo de Holt-Winters multivariante es el procedimiento de selección basado en la comparación de las propiedades predictivas de los modelos (sección 1.4.2). Como mostramos en esta sección, aunque su objetivo coincide con el nuestro, evaluar la compatibilidad de los datos con un modelo más sencillo, resulta difícil de implementar.

Sean $\varphi_0 = (\psi_0, \Sigma_0, \theta_0 = (\theta'_1, \theta'_2, \dots, \theta'_m)')$ y $\varphi_1 = (\psi_1, \Sigma_1, \theta)$ los vectores de parámetros del modelo de Holt-Winters multivariante general, \mathcal{M}_0 , y del modelo con parámetros de suavizado iguales, \mathcal{M}_1 . Bajo \mathcal{M}_1 , $\theta = (\alpha, \beta, \gamma)'$ es el vector de los parámetros de suavizado comunes para los modelos univariantes, resultante de imponer la restricción $\theta_1 = \theta_2 = \dots = \theta_m = \theta$ sobre el vector θ_0 de los parámetros de suavizado del modelo \mathcal{M}_0 . El subíndice 1 es eliminado para evitar posibles confusiones con el vector θ_1 asociado a la primera serie temporal. El vector de las condiciones iniciales y la matriz de covarianzas están igualmente definidos en ambos modelos.

Sea $f_0(y|\varphi_0) = N_{nm}(y|M_D\psi_0, L_B(\Sigma_0 \otimes I_n)L'_B)$ la función de verosimilitud supuesta cierta. Entonces, la distancia generalizada de Kullback-Leibler, utilizada para definir la discrepancia entre los dos modelos, es de la forma

4.2. PROPIEDADES PREDICTIVAS PARA LA SELECCIÓN DEL MODELO DE HOLT-WINTERS MULTIVARIANTE 99

(ver ecuaciones (2.2) y (3.12)):

$$\begin{aligned}
 d\{f_0(y|\varphi_0), f_1(y|\varphi_1)\} &= \int f_0(y|\varphi_0) \log \frac{f_0(y|\varphi_0)}{f_1(y|\varphi_1)} dy \\
 &= \int f_0(y|\varphi_0) \log \frac{|\Sigma_0|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}(y - M_D\psi_0)'(L'_B)^{-1}(\Sigma_0^{-1} \otimes I_n)L_B^{-1}(y - M_D\psi_0)\right\}}{|\Sigma_1|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}(y - M_D\psi_1)'(L'_D)^{-1}(\Sigma_1^{-1} \otimes I_n)L_D^{-1}(y - M_D\psi_1)\right\}} dy \\
 &= -\frac{n}{2} \log \frac{|\Sigma_0|}{|\Sigma_1|} - \frac{1}{2} E_y[(y - M_D\psi_0)'(L'_B)^{-1}(\Sigma_0^{-1} \otimes I_n)L_B^{-1}(y - M_D\psi_0)] \\
 &\quad + \frac{1}{2} E_y[(y - M_D\psi_1)'(L'_D)^{-1}(\Sigma_1^{-1} \otimes I_n)L_D^{-1}(y - M_D\psi_1)]
 \end{aligned}$$

Por otro lado,

$$\begin{aligned}
 &E_y[(y - M_D\psi_0)'(L'_B)^{-1}(\Sigma_0^{-1} \otimes I_n)L_B^{-1}(y - M_D\psi_0)] \\
 &= E_y \left[\text{tr} \left[(y - M_D\psi_0)'(L'_B)^{-1}(\Sigma_0^{-1} \otimes I_n)L_B^{-1}(y - M_D\psi_0) \right] \right] \\
 &= \text{tr} \left[(L'_B)^{-1}(\Sigma_0^{-1} \otimes I_n)L_B^{-1} E_y \left[(y - M_D\psi_0)(y - M_D\psi_0)'\right] \right] \\
 &= nm
 \end{aligned}$$

y, análogamente,

$$\begin{aligned}
 &E_y[(y - M_D\psi_1)'(L'_D)^{-1}(\Sigma_1^{-1} \otimes I_n)L_D^{-1}(y - M_D\psi_1)] \\
 &= E_y \left[\text{tr} \left[(y - M_D\psi_1)'(L'_D)^{-1}(\Sigma_1^{-1} \otimes I_n)L_D^{-1}(y - M_D\psi_1) \right] \right] \\
 &= \text{tr} \left[(L'_D)^{-1}(\Sigma_1^{-1} \otimes I_n)L_D^{-1} E_y \left[(y - M_D\psi_1)(y - M_D\psi_1)'\right] \right] \\
 &= \text{tr} \left[(L'_D)^{-1}(\Sigma_1^{-1} \otimes I_n)L_D^{-1}(L_B(\Sigma_0 \otimes I_n)L'_B + M_D(\psi_0 - \psi_1)(\psi_0 - \psi_1)'M'_D) \right]
 \end{aligned}$$

por lo que la distancia generalizada de Kullback-Leibler viene dada por:

$$\begin{aligned}
 d\{f_0(y|\varphi_0), f_1(y|\varphi_1)\} &= -\frac{n}{2} \log \frac{|\Sigma_0|}{|\Sigma_1|} - \frac{nm}{2} \\
 &\quad + \frac{1}{2} \text{tr} \left[(L'_D)^{-1}(\Sigma_1^{-1} \otimes I_n)L_D^{-1}L_B(\Sigma_0 \otimes I_n)L'_B \right] \\
 &\quad + \frac{1}{2} \text{tr} \left[(L'_D)^{-1}(\Sigma_1^{-1} \otimes I_n)L_D^{-1}M_D(\psi_0 - \psi_1)(\psi_0 - \psi_1)'M'_D \right]
 \end{aligned}$$

y la discrepancia entre los dos modelos, considerando la definición de Goutis and Robert (1998), ecuación (1.43), es de la forma:

$$\delta(\varphi_0, \varphi_1) = \inf_{\psi_1, \Sigma_1, \theta} d\{f_0(y|\varphi_0), f_1(y|\varphi_1)\}$$

El último término de la distancia de Kullback-Leibler puede anularse tomando $\psi_1 = \psi_0$. Sin embargo, para calcular el ínfimo respecto a Σ_1 y

θ nos encontramos ante un problema de optimización con cierta dificultad de resolución. La función objetivo no es una función convexa, por lo que la convergencia global no está garantizada; el procedimiento de optimización puede converger a mínimos locales e incluso no converger. Además, en la optimización respecto a Σ_1 podemos tener problemas de convergencia. El primer término disminuye con Σ_1 , con límite $-\infty$ cuando Σ_1 es la matriz nula, por lo que, aunque el segundo término disminuye conforme aumenta el valor de Σ_1 , con valor 0 si Σ_1^{-1} es la matriz nula, la convergencia no está garantizada. Por último, la optimización respecto al vector θ no es sencilla, pues los parámetros de suavizado del modelo \mathcal{M}_1 aparecen en la matriz L_D en la forma (3.12). Es por ello que hemos decidido no utilizar ese procedimiento en este trabajo.

4.3. Criterios de selección para el modelo de Holt-Winters multivariante

Finalmente, en esta sección empleamos los criterios de selección (ver sección 1.4.3) para resolver el problema de selección de modelos asociado al modelo de Holt-Winters multivariante.

Para el modelo general, \mathcal{M}_0 , la desviación $D_0(\psi, \Sigma, \theta_0) = -2 \log f_0(y|\psi, \Sigma, \theta_0)$ es de la forma (ver (2.4)):

$$D_0(\psi, \Sigma, \theta_0) \propto n \log |\Sigma| + (z - X_B \psi)' (\Sigma^{-1} \otimes I_n) (z - X_B \psi) \quad (4.18)$$

siendo $z = L_B^{-1} y$ y $X_B = L_B^{-1} M_D$.

Sea $\{(\psi^{(j)}, \Sigma^{(j)}, \theta_0^{(j)})\}_{j=1}^N$ una muestra obtenida de la distribución a posteriori de los parámetros del modelo, ecuación (2.26). Entonces, el criterio AIC y el criterio BIC pueden ser calculados, ver ecuaciones (1.45) y (1.46), como:

$$AIC_0 = D_0(\bar{\psi}, \bar{\Sigma}, \bar{\theta}_0) + 2p_0 \quad (4.19)$$

$$BIC_0 = D_0(\bar{\psi}, \bar{\Sigma}, \bar{\theta}_0) + \log(nm)p_0 \quad (4.20)$$

donde $\bar{\psi} = \frac{1}{N} \sum_{j=1}^N \psi^{(j)}$, $\bar{\Sigma} = \frac{1}{N} \sum_{j=1}^N \Sigma^{(j)}$ y $\bar{\theta}_0 = \frac{1}{N} \sum_{j=1}^N \theta_0^{(j)}$ representan las medias muestrales de los parámetros y $p_0 = m(s+1) + m(m+1)/2 + 3m$ es el número de parámetros del modelo.

Cuando el número de observaciones sea pequeño en comparación con el número de parámetros, es decir, cuando $nm/p_0 \leq 40$, proponemos emplear el criterio AICc (ver ecuación (1.47)):

$$AIC_c^0 = D_0(\bar{\psi}, \bar{\Sigma}, \bar{\theta}_0) + 2p_0 + \frac{2p_0(p_0 + 1)}{nm - p_0 - 1} \quad (4.21)$$

Para el cálculo del criterio DIC necesitamos obtener el valor de la esperanza a posteriori de la desviación, resultante de resolver la siguiente integral:

$$E(D_0(\psi, \Sigma, \theta_0)|y) = \int \int \int D_0(\psi, \Sigma, \theta_0) f_0(\psi, \Sigma, \theta_0|y) d\psi d\Sigma d\theta_0$$

Dicha integral no puede ser resuelta analíticamente, pero puede ser estimada mediante integración por Monte Carlo a partir de la muestra simulada de la distribución a posteriori de los parámetros de \mathcal{M}_0 . Así pues, el criterio DIC puede ser calculado, ver ecuación (1.48), como:

$$DIC_0 = \frac{2}{N} \sum_{j=1}^N D_0(\psi^{(j)}, \Sigma^{(j)}, \theta_0^{(j)}) - D_0(\bar{\psi}, \bar{\Sigma}, \bar{\theta}_0) \quad (4.22)$$

donde $D_0(\psi, \Sigma, \theta_0)$ es de la forma (4.18).

Para el modelo de Holt-Winters multivariante con parámetros de suavizado iguales, \mathcal{M}_1 , se tienen resultados parecidos. La desviación $D_1(\psi, \Sigma, \theta) = -2 \log f_1(y|\psi, \Sigma, \theta)$ es de la forma (ver (3.15)):

$$D_1(\psi, \Sigma, \theta) \propto n \log |\Sigma| + \text{tr} [\Sigma^{-1} (z_{mat} - X\psi_{mat})' (z_{mat} - X\psi_{mat})] \quad (4.23)$$

siendo $z_{mat} = L^{-1}y_{mat}$ y $X = L^{-1}M$.

Sea $\{(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)})\}_{j=1}^N$ una muestra de la distribución a posteriori de los parámetros de \mathcal{M}_1 , ecuación (3.18). Los criterios AIC y BIC son calculados como:

$$AIC_1 = D_1(\bar{\psi}, \bar{\Sigma}, \bar{\theta}) + 2p_1 \quad (4.24)$$

$$BIC_1 = D_1(\bar{\psi}, \bar{\Sigma}, \bar{\theta}) + \log(nm)p_1 \quad (4.25)$$

donde $\bar{\psi} = \frac{1}{N} \sum_{j=1}^N \psi^{(j)}$, $\bar{\Sigma} = \frac{1}{N} \sum_{j=1}^N \Sigma^{(j)}$ y $\bar{\theta} = \frac{1}{N} \sum_{j=1}^N \theta^{(j)}$ son las medias muestrales de los parámetros y $p_1 = m(s + 1) + m(m + 1)/2 + 3$ es el número de parámetros del modelo. Cuando el número de observaciones sea pequeño en comparación con el número de parámetros consideraremos

el criterio AICc.

La esperanza a posteriori de la desviación:

$$E(D_1(\psi, \Sigma, \theta)|y) = \int \int \int D_1(\psi, \Sigma, \theta) f_1(\psi, \Sigma, \theta|y) d\psi d\Sigma d\theta$$

puede ser estimada nuevamente mediante integración por Monte Carlo. Entonces, el criterio DIC asociado a \mathcal{M}_1 puede ser calculado como:

$$DIC_1 = \frac{2}{N} \sum_{j=1}^N D_1(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)}) - D_1(\bar{\psi}, \bar{\Sigma}, \bar{\theta}) \quad (4.26)$$

Notar que para obtener una muestra de la distribución a posteriori de los parámetros del modelo \mathcal{M}_1 , en primer lugar obtenemos una muestra $\{\theta^{(j)}\}_{j=1}^N$ de la distribución a posteriori $f_1(\theta|y)$, ecuación (3.25), y, a continuación, para cada $\theta^{(j)}$ simulamos $\psi^{(j)}$ y $\Sigma^{(j)}$ a partir de las condicionales $f_1(\Sigma|y, \theta)$ y $f_1(\psi|y, \Sigma, \theta)$, ecuaciones (3.20) y (3.21) respectivamente.

4.4. Ejemplo de ocupación hotelera (resumen)

A lo largo de la memoria hemos analizado las series temporales de ocupación hotelera en Castellón, Valencia y Alicante con los distintos procedimientos presentados. Así pues, dedicamos esta última sección a la resolución del problema de selección de modelos asociado al análisis conjunto de las series mediante el modelo de Holt-Winters multivariante. Posteriormente, recopilamos los principales resultados obtenidos en su análisis.

Representamos por \mathcal{M}_0 al modelo general, mientras que el modelo con parámetros de suavizado iguales para los modelos univariantes es denotado por \mathcal{M}_1 . En primer lugar calculamos los estimadores por muestreo importante de las verosimilitudes marginales, lo que nos permite obtener una estimación del factor Bayes B_{01} .

Notar que para la comparación de los estimadores por muestreo importante de las verosimilitudes marginales es posible considerar (ver ecuaciones (4.16) y (4.17)) las expresiones:

$$\hat{f}_{IS}(y|\mathcal{M}_0) \approx \frac{1}{N} \sum_{j=1}^N f_0(y|\tilde{T}^{(j)}, \theta_0^{(j)}) \frac{\tilde{f}_0(\tilde{T}^{(j)})}{g_1(\tilde{T}^{(j)})} \quad (4.27)$$

$$\hat{f}_{IS}(y|\mathcal{M}_1) \approx k_2^{-1} \frac{1}{N} \sum_{j=1}^N f_1(y|\theta^{(j)}) \quad (4.28)$$

4.4. EJEMPLO DE OCUPACIÓN HOTELERA (RESUMEN) 103

donde eliminamos las constantes que aparecen en ambos modelos.

Para el modelo \mathcal{M}_0 tomamos como función importante $g_1(\tilde{T})$ la t-Student multivariante con $\nu = 10$ grados de libertad y vector de medias y matriz de covarianzas definidos a partir de la muestra obtenida de la distribución a posteriori de los parámetros del modelo como sigue: Para cada matriz de covarianzas $\Sigma^{(j)}$ simulada calculamos su matriz inversa. Posteriormente, calculamos la descomposición de Cholesky de $(\Sigma^{-1})^{(j)}$ y definimos la matriz $\tilde{T}^{(j)}$ correspondiente. De este modo, obtenemos una muestra de la distribución a posteriori $f(\tilde{T}|y)$ que nos permite estimar la media y la matriz de covarianzas a posteriori de los elementos no nulos de \tilde{T} . Recordar que, en ambos modelos, los parámetros de suavizado son simulados a partir de la distribución uniforme.

Los estimadores por muestreo importante de las verosimilitudes marginales, calculados como en (4.27) y (4.28) respectivamente a partir de una muestra de tamaño $N = 25000$ de las correspondientes distribuciones importantes, son $\hat{f}_{IS}(y|\mathcal{M}_0) \approx 1.17\text{E-}160$ y $\hat{f}_{IS}(y|\mathcal{M}_1) \approx 1.54\text{E-}158$.

Dependiendo de la forma de la función de verosimilitud, en la práctica es posible que la estimación de la verosimilitud marginal sea del orden obtenido. No obstante, valores tan pequeños ponen en duda la validez de las estimaciones resultantes, pues en su cálculo pueden haberse producido errores numéricos que alteren el valor real.

Para valorar la eficiencia del método de muestreo importante propuesto calculamos el tamaño de muestra efectivo (ESS, Liu, 2001, pag. 34), que se define como:

$$ESS = \frac{1}{\sum_{j=1}^N (w_n^{(j)})^2}$$

donde el peso $w^{(j)}$, para $j = 1, 2, \dots, N$, es el cociente entre la distribución objetivo y la distribución importante evaluadas en $\varphi = \varphi^{(j)}$, valor simulado para los parámetros a partir de la densidad importante, y $\{w_n^{(j)} = w^{(j)} / \sum_{k=1}^N w^{(k)}\}_{j=1}^N$ son los pesos normalizados. El tamaño de muestra efectivo es una aproximación del número de muestras independientes, simuladas de la densidad objetivo, que serían necesarias para obtener una estimación con la misma varianza Monte Carlo que la resultante de las N simulaciones de la densidad importante. El método por muestreo importante se considera adecuado si $ESS \geq N/2$ (notar que $1 \leq ESS \leq N$).

En el ejemplo de ocupación hotelera obtenemos los valores $ESS_0 = 4.89$ y $ESS_1 = 187.09$ para \mathcal{M}_0 y \mathcal{M}_1 respectivamente, indicando que las funciones importantes propuestas no son eficientes.

Otra posible función importante para estimar la verosimilitud marginal bajo \mathcal{M}_0 , alternativa a la desarrollada en la sección 4.1, es $cg(\Sigma, \theta_0) = g_1(\Sigma)g_2(\theta_0)$, siendo $g_1(\Sigma)$ una Wishart invertida y $g_2(\theta_0)$ el producto de distribuciones Beta, con parámetros definidos de manera que la media y la varianza de cada Beta coincidan con la media y la varianza muestrales a posteriori del correspondiente parámetro de suavizado. Análogamente, como función importante para estimar la verosimilitud marginal bajo \mathcal{M}_1 proponemos $cg(\theta) = g_2(\theta)$, con $g_2(\theta)$ el producto de distribuciones Beta. Estas distribuciones importantes, aunque producen resultados más satisfactorios que los anteriores, siguen sin ser apropiadas. Notar que éste es un problema frecuente cuando se utiliza el método de muestreo importante para estimar funciones n -dimensionales, pues es difícil obtener distribuciones importantes eficientes. Por ello, esta herramienta de selección de modelos no nos parece aconsejable para la resolución del problema de selección presentado en este trabajo.

Veamos a continuación el modelo de Holt-Winters multivariante seleccionado para el análisis conjunto de las series temporales de ocupación hotelera a partir de los distintos criterios de selección de modelos: AICc, BIC y DIC. En la tabla 4.1 se muestra el valor de estos criterios para el modelo de Holt-Winters multivariante general, \mathcal{M}_0 , y el modelo con parámetros de suavizado iguales, \mathcal{M}_1 , así como los términos de penalización de cada criterio.

Si empleamos los criterios AICc o BIC para la resolución de nuestro problema de selección de modelos, el modelo escogido para el análisis multivariante de las series temporales es el modelo de Holt-Winters multivariante con parámetros de suavizado iguales para los modelos univariantes. En cambio, si el problema de selección es resuelto mediante la utilización del criterio DIC, el modelo seleccionado es el modelo de Holt-Winters general.

4.4. EJEMPLO DE OCUPACIÓN HOTELERA (RESUMEN) 105

	\mathcal{M}_0	\mathcal{M}_1
número parámetros	54	48
$D(\bar{\psi}, \bar{\Sigma}, \bar{\theta})$	852.23	790.88
término penalización AICc	155.52	131.91
AICc	1007.75	922.79
término penalización BIC	280.42	249.26
BIC	1132.65	1040.14
$E(D(\psi, \Sigma, \theta) y)$	860.17	837.48
P_D	7.94	46.60
DIC	868.11	884.08



Tabla 4.1: Criterios de selección para el modelo de Holt-Winters multivariante en el análisis conjunto de las series de ocupación hotelera

Para entender mejor el motivo de esta diferencia en el modelo seleccionado, expresamos el criterio DIC en su forma alternativa

$$DIC = D(E(\psi, \Sigma, \theta|y)) + 2P_D$$

de manera que el primer término de la definición, la desviación Bayesiana calculada en la media a posteriori de los parámetros del modelo, es común en los tres criterios y la única diferencia radica en el segundo término o término de penalización. Dado que el término $D(\bar{\psi}, \bar{\Sigma}, \bar{\theta})$ asociado al modelo \mathcal{M}_1 es menor, los criterios AICc o BIC, cuya penalización al modelo \mathcal{M}_0 es siempre mayor, escogen el modelo \mathcal{M}_1 . En cambio, el término $2P_D$, número eficiente de parámetros incluidos en el modelo, es menor para el modelo \mathcal{M}_0 , siendo finalmente menor el valor del criterio DIC asociado al modelo general \mathcal{M}_0 .

A fin de valorar el comportamiento de los distintos criterios, recopilamos los errores de ajuste y de predicción, tablas 4.2 y 4.3 respectivamente, resultantes del análisis conjunto de las series temporales con el modelo de Holt-Winters multivariante general, \mathcal{M}_0 , y el modelo de Holt-Winters multivariante con parámetros de suavizado iguales, \mathcal{M}_1 . A modo comparativo, incluimos en las tablas los errores derivados de los análisis univariantes.

Como podemos observar, el modelo seleccionado mediante los criterios AICc y BIC, \mathcal{M}_1 , es el que da lugar a un mejor ajuste de los datos históricos que conforman las series temporales. Además, el ajuste obtenido en el análisis multivariante de las series es mejor que el resultante de analizar las series de manera independiente a partir del modelo de Holt-Winters univariante. No obstante, tal y como ocurre en este ejemplo, el modelo que mejor se ajusta a los datos no es necesariamente el modelo con mayor precisión en

Error ajuste	Castellón	Valencia	Alicante	Media
H-W univariante	9.29	9.42	6.23	8.31
H-W multivariante \mathcal{M}_0	9.70	11.14	6.37	9.07
H-W multivariante \mathcal{M}_1	8.78	8.89	6.06	7.91

Tabla 4.2: Errores de ajuste obtenidos, para cada una de las series temporales de ocupación hotelera, en los análisis univariantes y multivariantes con el modelo de Holt-Winters general y el modelo con parámetros de suavizado iguales

Error predicción	Castellón	Valencia	Alicante	Media
H-W univariante	10.13	35.00	17.58	20.90
H-W multivariante \mathcal{M}_0	10.57	28.58	17.78	18.98
H-W multivariante \mathcal{M}_1	12.95	37.04	18.77	22.92

Tabla 4.3: Errores de predicción obtenidos, para cada una de las series temporales de ocupación hotelera, en los análisis univariantes y multivariantes con el modelo de Holt-Winters general y el modelo con parámetros de suavizado iguales

la predicción. Así pues, el modelo escogido a partir del criterio DIC, \mathcal{M}_0 , es el que proporciona predicciones más precisas, reduciendo incluso el error de predicción con respecto a los análisis univariantes.

Como conclusión final podemos destacar que el uso del modelo de Holt-Winters multivariante para el análisis conjunto de las series temporales de ocupación hotelera está justificado, pues las medias e intervalos de confianza a posteriori para los coeficientes de correlación (ver tabla 2.4) verifican la correlación existente entre las series. A partir del modelo multivariante es posible mejorar el ajuste o la predicción con respecto a los análisis univariantes. Por otro lado, mientras que los criterios AICc y BIC seleccionan el modelo multivariante que proporciona un mejor ajuste para los datos históricos de las series, \mathcal{M}_1 , el criterio DIC selecciona el modelo con mayor precisión de predicción, \mathcal{M}_0 . No obstante, si tenemos en cuenta la dificultad de análisis que supone el uso del modelo de Holt-Winters multivariante general, \mathcal{M}_0 , en relación al modelo con parámetros de suavizado iguales, \mathcal{M}_1 , podemos concluir que el criterio DIC penaliza muy poco el exceso de parámetros. Considerando además la reducción en el error de predicción resultante de analizar las series a partir del modelo \mathcal{M}_0 , podemos decir que los criterios AICc y BIC producen resultados más satisfactorios en este caso.

Capítulo 5

Resultados numéricos

El modelo de Holt-Winters multivariante general, al considerar un mayor número de parámetros, es mucho más complejo que el modelo univariante. Por tanto, su uso para el análisis de un conjunto de series temporales resulta adecuado si éstas presentan una correlación alta entre sí. En caso contrario, no existe una compensación entre el número de parámetros incorporados en el modelo y la información adicional proporcionada por las series incluidas. Consecuentemente, los resultados obtenidos tanto en el ajuste como en la predicción de las series pueden ser menos satisfactorios que los resultantes de los análisis univariantes.

Así pues, dado un conjunto de series temporales con errores posiblemente correlacionados en cada instante temporal, debemos decidir entre realizar un análisis multivariante de las mismas o analizarlas de manera independiente mediante el modelo de Holt-Winters univariante. Dicha decisión puede basarse en las estimaciones obtenidas para las distribuciones a posteriori de los coeficientes de correlación.

Si se verifica que errores contemporáneos en las series temporales univariantes están correlacionados, el siguiente paso del análisis es seleccionar entre el modelo de Holt-Winters multivariante general, \mathcal{M}_0 , y el modelo con parámetros de suavizado idénticos para los modelos univariantes, \mathcal{M}_1 . Dicho problema de selección de modelos es resuelto mediante la utilización de los distintos criterios de selección. Una vez seleccionado el modelo, su análisis y el cálculo de predicciones, tanto puntuales como intervalos de predicción, son los descritos en los capítulos anteriores (capítulos 2 y 3 respectivamente para \mathcal{M}_0 y \mathcal{M}_1).

En este capítulo mostramos el análisis completo, mediante el procedimiento Bayesiano de predicción propuesto en esta memoria, de dos conjuntos de series temporales reales. El primero de ellos se corresponde con las series del Ibex 35 y del Dow Jones, principales índices de referencia de las bolsas española y estadounidense. Dado que las series temporales no presentan estacionalidad, utilizamos el modelo de Holt multivariante para su análisis. Además, este ejemplo nos permite mostrar la predicción de un paso cuando los valores de las series temporales son observados secuencialmente. En el segundo ejemplo analizamos las series temporales del número de pasajeros de cinco aeropuertos de Londres: Gatwick, Heathrow, London City, Luton y Stansted.

5.1. Índices bursátiles Dow Jones e Ibex 35

En esta sección se analizan las series temporales correspondientes al valor de cierre del Dow Jones y del Ibex 35. El Ibex 35, elaborado por Bolsas y Mercados Españoles (BME) en 1989, es el principal índice de referencia de la bolsa española. Está formado por las 35 compañías de mayor liquidez que cotizan en el Sistema de Interconexión Bursátil Español (SIBE). Es un índice ponderado por capitalización bursátil, lo que significa que las empresas con mayor liquidez tienen más peso sobre el índice. El Dow Jones, creado en 1883 por Dow Jones & Company, refleja el comportamiento del precio de la acción de las 30 compañías industriales más importantes de Estados Unidos. Es el índice más antiguo y de mayor influencia del mundo.

En concreto, las series temporales que estudiamos están formadas por los valores de cierre de los índices en 91 días consecutivos: desde el 1 de Marzo hasta el 12 de Julio de 2007 para el Dow Jones y desde el 2 de Marzo hasta el 13 de Julio de 2007 para el Ibex 35. El desfase temporal que observamos entre ambas series temporales permite suponer errores correlacionados para cada instante temporal, pues el valor de cierre del Dow Jones en un día particular influye en el valor de cierre del Ibex 35 del día siguiente.

Como puede observarse, las series temporales presentan una tendencia creciente pero no muestran componente estacional, por lo que utilizaremos el modelo de Holt para su análisis. Además, parece razonable suponer que los errores de las series en cada instante temporal están correlacionados, lo que justificaría el uso del modelo de Holt multivariante. Recordar (Proposición

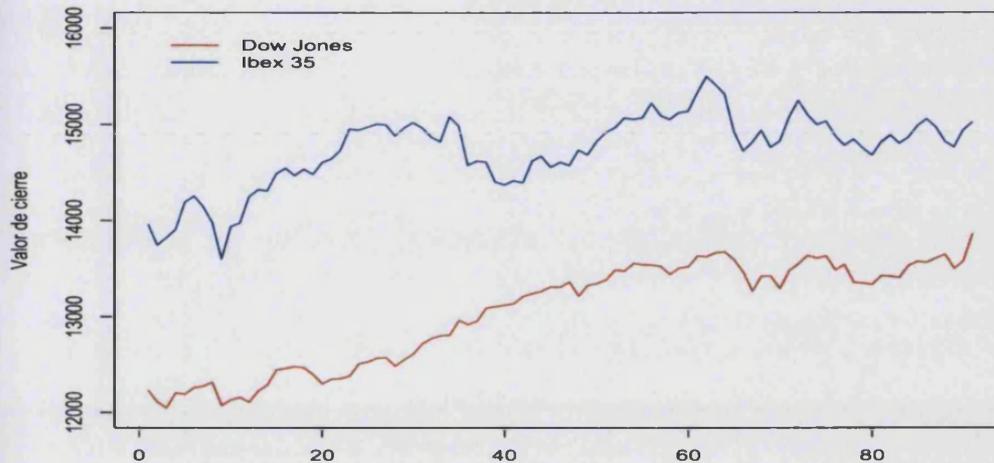


Figura 5.1: Gráfico temporal de las series del valor de cierre del Dow Jones (desde el 1 de Marzo hasta el 12 de Julio de 2007) y del Ibex 35 (desde el 2 de Marzo hasta el 13 de Julio de 2007)

3.1) que el análisis del modelo de Holt multivariante es equivalente al análisis del modelo de Holt-Winters multivariante fijando $s = 1$ y $\psi_i = (b_{0i}, c_{0i})'$ con $c_{0i} = a_{0i} + b_{0i}$ para $i = 1, 2, \dots, m$.

Con el objetivo de mostrar el comportamiento de nuestro procedimiento de predicción, consideramos como datos históricos las 87 primeras observaciones, utilizando las 4 últimas observaciones, correspondientes a la última semana, para valorar la precisión de nuestra predicción fuera de la muestra.

Para estimar la correlación existente entre las series simulamos $N_0 + N = 55000$ valores de la distribución a posteriori de los parámetros del modelo de Holt multivariante, ecuación (2.26). Para ello empleamos el algoritmo de Metropolis-within-Gibbs propuesto en la sección 2.2.2 con la particularidad de que, en cada iteración del algoritmo, repetimos 20 veces el paso 3, correspondiente a la simulación del vector de parámetros de suavizado, pues de este modo aceleramos la convergencia del algoritmo. Tras eliminar las $N_0 = 5000$ primeras simulaciones de calentamiento, obtenemos una muestra $\{(\psi^{(j)}, \Sigma^{(j)}, \theta^{(j)})\}_{j=1}^N$ de la distribución a posteriori de los parámetros del modelo que satisface los criterios de convergencia. En la tabla 5.1 se recogen las estimaciones de la media y del intervalo de probabilidad a posteriori del 95% para los parámetros del modelo de Holt multivariante general. Los subíndices 1 y 2 hacen referencia a las series del Dow Jones y del Ibex 35 respectivamente.

	$q_{0.025}$	media	$q_{0.975}$
a_{01}	12040.41	12211.78	12380.82
b_{01}	-73.17	5.64	69.43
a_{02}	13624.39	13916.14	14207.82
b_{02}	-54.71	17.57	100.03
α_1	0.74	0.89	0.99
β_1	0.00	0.10	0.32
α_2	0.67	0.87	0.99
β_2	0.00	0.04	0.14
σ_1	74.10	86.89	101.85
σ_2	127.96	150.03	175.73
ρ_{12}	0.12	0.34	0.53

Tabla 5.1: Estimaciones de las medias y de los intervalos de probabilidad a posteriori del 95 % para los parámetros del modelo de Holt multivariante resultante del análisis conjunto de las series del Dow Jones y del Ibex 35

A partir de la media y del intervalo de probabilidad a posteriori para el coeficiente de correlación ρ_{12} , podemos concluir que las series temporales están correlacionadas y, por tanto, el uso del modelo multivariante para su análisis está justificado. El siguiente paso del análisis es seleccionar entre el modelo de Holt multivariante general y el modelo con parámetros de suavizado iguales.

Para la resolución del problema de selección de modelos a partir de los distintos criterios de selección (ver sección 4.3) es necesario obtener una muestra de la distribución a posteriori de los parámetros del modelo de Holt multivariante con parámetros de suavizado iguales, \mathcal{M}_1 . Para ello simulamos, en primer lugar, una muestra de la distribución a posteriori de los parámetros de suavizado comunes $f(\theta|y)$, ecuación (3.25), mediante el método de aceptación-rechazo propuesto en la sección 3.2.2. Posteriormente, para cada valor $\theta^{(j)}$ simulamos un valor de la matriz ψ_{mat} de las condiciones iniciales y de la matriz de covarianzas Σ a partir de las distribuciones a posteriori condicionales $f(\Sigma|y, \theta)$ y $f(\psi_{mat}|y, \Sigma, \theta)$, ecuaciones (3.20) y (3.19) respectivamente.

En particular simulamos $N = 10000$ valores de la distribución $f(\theta|y)$. La figura 5.2 muestra los histogramas de los parámetros de suavizado simulados

así como las medias muestrales (líneas rojas verticales).

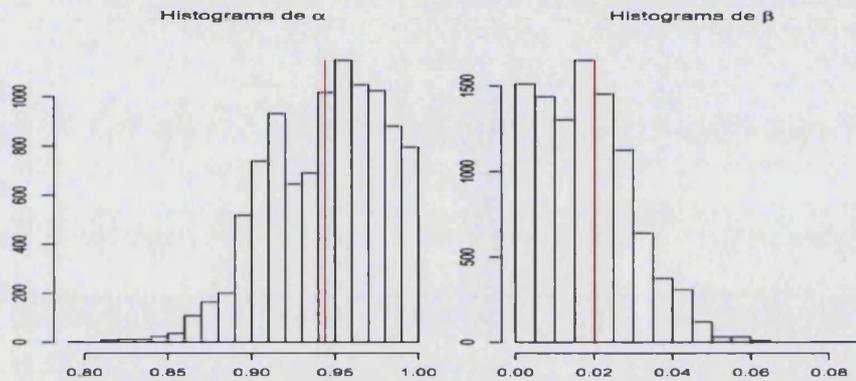


Figura 5.2: Histogramas de los parámetros de suavizado simulados de su distribución a posteriori cuando las series del Dow Jones y del Ibex 35 son analizadas mediante el modelo de Holt multivariante con parámetros de suavizado iguales. Líneas rojas verticales: medias muestrales

La densidad estimada del coeficiente de correlación ρ_{12} junto con su media e intervalo de probabilidad a posteriori del 95 % se muestra en la figura 5.3. Como puede observarse, la media y el intervalo de probabilidad son similares a los obtenidos para el modelo de Holt multivariante general (ver tabla 5.1).

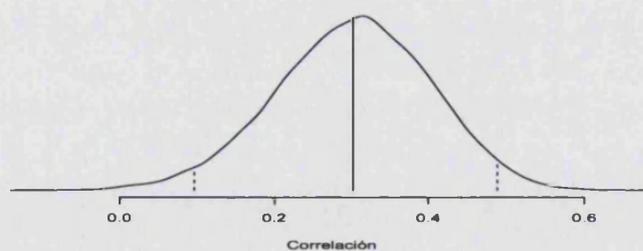


Figura 5.3: Densidad estimada del coeficiente de correlación del modelo de Holt multivariante con parámetros de suavizado iguales resultante del análisis conjunto de las series del Dow Jones y del Ibex 35

El valor de los criterios AICc, BIC y DIC, correspondientes a las muestras simuladas de las distribuciones a posteriori de los parámetros de \mathcal{M}_0 , modelo de Holt multivariante general, y \mathcal{M}_1 , modelo con parámetros de suavizado iguales, se incluyen en tabla 5.2.

	\mathcal{M}_0	\mathcal{M}_1
número parámetros	11	9
$D(\bar{\psi}, \bar{\Sigma}, \bar{\theta})$	1800.32	1796.00
término penalización AICc	23.63	19.09
AICc	1823.95	1815.09
término penalización BIC	56.75	46.43
BIC	1857.07	1842.43
$E(D(\psi, \Sigma, \theta) y)$	1808.36	1803.20
P_D	8.04	7.20
DIC	1816.40	1810.40

Tabla 5.2: Criterios de selección para el modelo de Holt multivariante en el análisis conjunto de las series temporales del Dow Jones y del Ibex 35

El modelo seleccionado para la obtención de predicciones para los valores futuros de las series temporales es, independientemente del criterio utilizado, el modelo de Holt multivariante con parámetros de suavizado iguales, \mathcal{M}_1 .

Por otro lado, este ejemplo nos permite ilustrar el comportamiento de nuestro procedimiento de predicción de un paso cuando los valores de las series son observados de manera secuencial, pues el valor de cierre del Dow Jones en un día concreto es observado con anterioridad al valor de cierre del Ibex 35 del día siguiente. Así pues, es posible modificar las predicciones de un paso obtenidas para el Ibex 35 en función de los valores observados del Dow Jones y obtener, de este modo, predicciones más precisas. Para ello, en lugar de obtener directamente predicciones para los 4 últimos valores de las series (consideramos como datos históricos las $n = 87$ primeras observaciones) fijando el horizonte de predicción $h = 4$, consideramos 4 escenarios diferentes de manera que, en cada escenario, calculamos la predicción de un paso para los valores futuros de las series. Es decir, en el escenario i , para $i = 1, 2, 3, 4$, suponemos conocidas las $n + i - 1$ primeras observaciones y calculamos la predicción de un paso para los valores futuros de las series. Entonces, una vez observado el valor del Dow Jones en el instante $n + i$, la predicción calculada para el Ibex 35 es modificada en función del error de predicción de un paso asociado al Dow Jones, ver ecuaciones (3.8) - (3.10).

Las predicciones de un paso obtenidas, en cada uno de los 4 escenarios, para las series temporales del Dow Jones y del Ibex 35 cuando son analizadas mediante el modelo de Holt multivariante con parámetros de suavizado iguales, \mathcal{M}_1 , aparecen en las tablas 5.3 y 5.4 respectivamente. La tabla 5.5

muestra las predicciones de un paso obtenidas para la serie del Ibex 35 cuando se incorpora la información proporcionada por el correspondiente valor del Dow Jones. Para ilustrar como las predicciones originales del Ibex 35 son modificadas en función de los errores de predicción de un paso asociados al Dow Jones, los valores reales de las series temporales son incluidos en las tablas. Finalmente, las tablas incluyen el error SMAPE de predicción correspondiente a las predicciones calculadas en los cuatro escenarios.

DJ	Predicción de un paso			Valor finalmente observado
	$q_{0.025}$	media	$q_{0.975}$	
E-1	13457.62	13624.32	13791.01	13649.97
E-2	13497.26	13663.07	13828.87	13501.70
E-3	13351.80	13520.80	13689.73	13577.87
E-4	13419.54	13587.33	13755.11	13861.73
error de predicción			0.95	

Tabla 5.3: Predicciones de un paso para la serie del Dow Jones cuando es analizada mediante el modelo \mathcal{M}_1

I35	Predicción de un paso			Valor finalmente observado
	$q_{0.025}$	media	$q_{0.975}$	
E-1	14699.67	14988.84	15278.05	14823.90
E-2	14549.15	14839.25	15129.46	14766.20
E-3	14483.96	14772.71	15061.46	14941.80
E-4	14649.69	14939.10	15228.29	15023.50
error de predicción			0.83	

Tabla 5.4: Predicciones de un paso para la serie del Ibex 35 cuando es analizada mediante el modelo \mathcal{M}_1

I35 DJ	Predicción de un paso			Valor finalmente observado
	$q_{0.025}$	media	$q_{0.975}$	
E-1	14727.99	15002.26	15276.53	14823.90
E-2	14474.55	14755.89	15037.21	14766.20
E-3	14527.15	14801.76	15076.32	14941.80
E-4	14794.51	15084.15	15373.76	15023.50
error de predicción			0.65	

Tabla 5.5: Predicciones de un paso para la serie del Ibex 35 resultantes de modificar las predicciones originales en función del error de predicción de un paso asociado al Dow Jones

Los errores de predicción obtenidos para las series del Dow Jones y del Ibex 35 con el modelo univariante son 0.95 y 0.80 respectivamente, similares a los resultantes del análisis multivariante. La ventaja que presenta el análisis multivariante respecto al univariante es que las predicciones de un paso obtenidas para la serie del Ibex 35 pueden ser modificadas en función del error de predicción asociado al Dow Jones, obteniendo así predicciones más precisas que en el análisis univariante. Por otro lado, el tiempo de computación requerido para el análisis de las series mediante el modelo multivariante se ha reducido en un 45 % aproximadamente. Ésto es debido a que, en cada escenario, sólo es necesario obtener una muestra de la distribución a posteriori del vector de los parámetros de suavizado comunes, mientras que en los análisis univariantes es necesario obtener una muestra de la distribución a posteriori de cada uno de los vectores de parámetros de suavizado.

Finalmente, los errores de predicción resultantes del análisis de las series temporales mediante el modelo de Holt multivariante general son 0.96 y 0.86 respectivamente para las series del Dow Jones y del Ibex 35, similares a los obtenidos en el análisis mediante el modelo \mathcal{M}_1 . Así pues, el modelo seleccionado en este ejemplo a partir de los distintos criterios de selección es el modelo que proporciona ligeramente mayor precisión en la predicción además de requerir de un tiempo de computación menor.

5.2. Pasajeros de líneas aéreas

En esta sección analizamos las series temporales mensuales correspondientes al número de pasajeros de cinco aeropuertos de Londres: Gatwick, Heathrow, London City, Luton y Stansted, desde Enero de 2001 hasta Diciembre de 2007. Como podemos observar en los gráficos temporales, ver figura 5.4, las series presentan tendencia creciente y componente estacional, por lo que su análisis mediante el modelo de Holt-Winters está justificado. Además, parece razonable suponer que errores contemporáneos en las series temporales estén correlacionados, por lo que el uso del modelo multivariante podría resultar conveniente.

Con el objetivo de mostrar el comportamiento del procedimiento Bayesiano de predicción propuesto en esta memoria, consideramos como datos históricos las observaciones correspondientes a los seis primeros años (2001 - 2006), mientras que las observaciones del último año (2007) son utilizadas para valorar la precisión de nuestra predicción fuera de la muestra

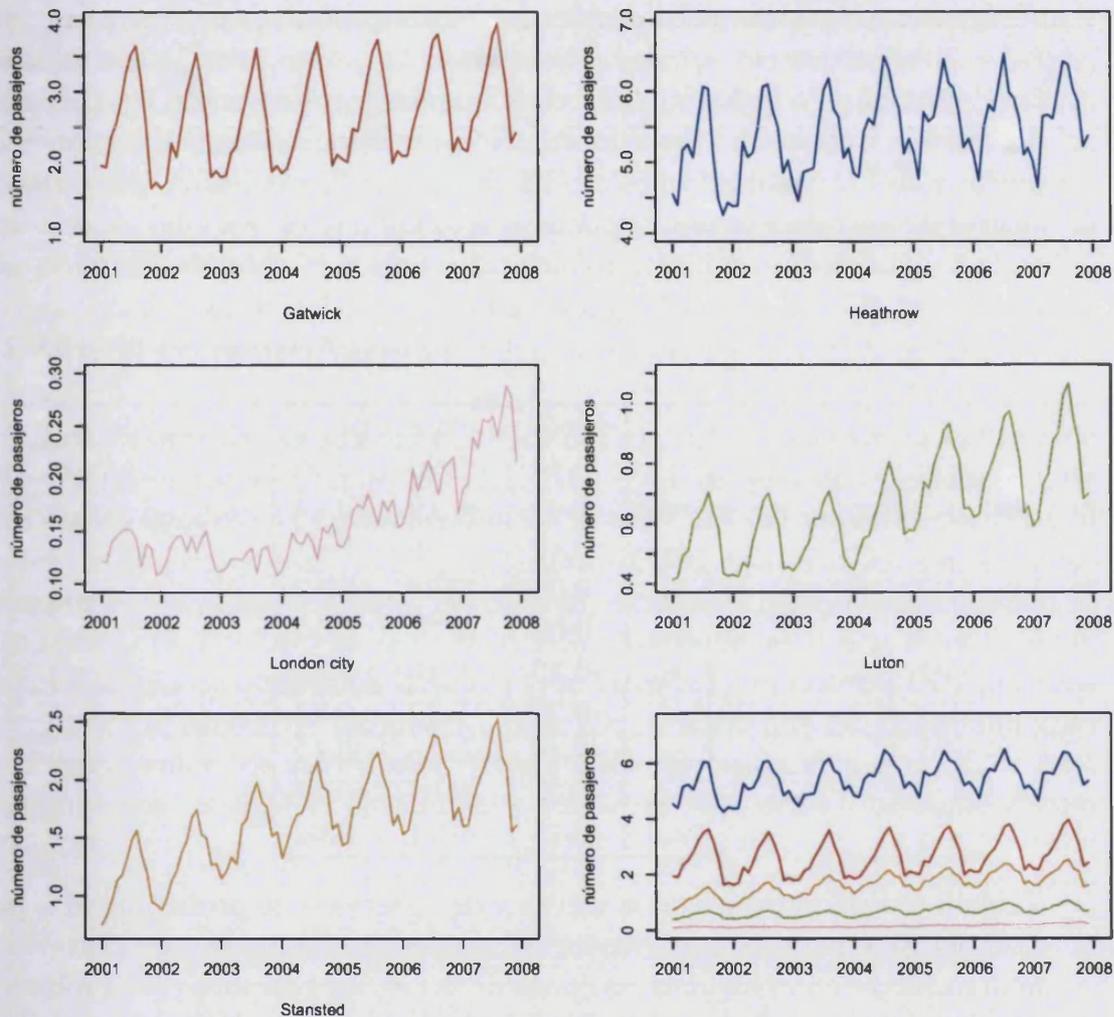


Figura 5.4: Gráficos temporales univariantes y multivariante de las series mensuales del número de pasajeros, en millones, de cinco aeropuertos de Londres

El primer paso del análisis es comprobar si los errores de las series temporales en cada instante temporal realmente están correlacionados. Para ello simulamos, mediante el algoritmo de Metropolis-within-Gibbs propuesto en la sección 2.2.2, $N_0 + N = 100000$ valores de la distribución a posteriori de los parámetros del modelo de Holt-Winters multivariante general, ecuación (2.26). Para acelerar la convergencia del algoritmo de simulación repetimos, en cada iteración del algoritmo, 20 veces el paso 3, es decir, consideramos el valor de los parámetros de suavizado obtenido después de repetir 20 veces

el algoritmo de Metropolis empleado para su simulación. Tras eliminar las $N_0 = 50000$ primeras simulaciones de calentamiento, obtenemos una muestra de tamaño $N = 50000$ de la distribución a posteriori de los parámetros del modelo que verifica los criterios de convergencia. Las estimaciones de la media y del intervalo de probabilidad a posteriori del 95 % para los coeficientes de correlación ρ_{kl} , con $k = 1, 2, 3, 4$ y $l = (k + 1), \dots, 5$, se recogen en la tabla 5.6. El subíndice 1 hace referencia a la serie de Gatwick, el 2 a Heathrow, el 3 a London City, el 4 a Luton y el 5 a Stansted.

	$q_{0.025}$	media	$q_{0.975}$
ρ_{12}	0.40	0.60	0.76
ρ_{13}	-0.46	-0.19	0.11
ρ_{14}	0.12	0.38	0.61
ρ_{15}	0.08	0.34	0.57
ρ_{23}	0.02	0.30	0.55
ρ_{24}	-0.15	0.15	0.42
ρ_{25}	0.23	0.47	0.68
ρ_{34}	-0.46	-0.18	0.12
ρ_{35}	-0.31	-0.02	0.27
ρ_{45}	0.15	0.41	0.62

Tabla 5.6: Estimaciones de la media y del intervalo de probabilidad a posteriori del 95 % para los coeficientes de correlación del modelo de Holt-Winters multivariante en el análisis conjunto de las series temporales del número de pasajeros de cinco aeropuertos de Londres

A la vista de los resultados, podemos concluir que las series temporales de Gatwick, Heathrow, Luton y Stansted presentan una correlación alta entre sí. Por tanto, se recomienda el modelo multivariante para su análisis pues, además de estar justificado, incorpora la correlación existente entre las series. La duda aparece en la inclusión de la serie temporal correspondiente a London City, representada por el subíndice 3, pues sólo está claramente correlacionada con la serie temporal de Heathrow. No obstante, dado que la serie guarda cierta relación con las demás, proponemos analizar las cinco series conjuntamente mediante el modelo de Holt-Winters multivariante.

El siguiente paso del análisis es seleccionar entre el modelo de Holt-Winters multivariante general, \mathcal{M}_0 , y el modelo con parámetros de suavizado comunes para los modelos univariantes, \mathcal{M}_1 . Para ello, es necesario

recurrir nuevamente a la simulación y obtener así una muestra de la distribución a posteriori de los parámetros de \mathcal{M}_1 . En primer lugar simulamos, mediante el algoritmo de aceptación-rechazo propuesto en la sección 3.2.2, una muestra de tamaño $N = 10000$ de la distribución a posteriori de los parámetros de suavizado comunes $f(\theta|y)$, ecuación (3.25). Posteriormente, para cada valor $\theta^{(j)}$ de la muestra simulamos un valor de la matriz ψ_{mat} de las condiciones iniciales y de la matriz de covarianzas Σ a partir de las distribuciones a posteriori condicionales $f(\Sigma|y, \theta)$ y $f(\psi_{mat}|y, \Sigma, \theta)$, ecuaciones (3.20) y (3.19) respectivamente.

Los valores de los criterios AICc, BIC y DIC asociados al modelo de Holt-Winters multivariante general, \mathcal{M}_0 , y al modelo con parámetros de suavizado iguales, \mathcal{M}_1 , se muestran en la tabla 5.7. El modelo seleccionado a partir de los criterios AICc y BIC para el posterior análisis de las series temporales es el modelo \mathcal{M}_1 . En cambio, el modelo seleccionado cuando se utiliza el criterio DIC es el modelo \mathcal{M}_0 . Nuevamente, tal y como ocurría en el ejemplo de ocupación hotelera (ver tabla 4.1), el criterio DIC penaliza muy poco el exceso de parámetros. Por ello, y dado que no existe una gran diferencia entre los valores del criterio DIC correspondientes a \mathcal{M}_0 y \mathcal{M}_1 , proponemos analizar conjuntamente las series temporales mediante el modelo \mathcal{M}_1 .

	\mathcal{M}_0	\mathcal{M}_1
número parámetros	95	83
$D(\psi, \bar{\Sigma}, \bar{\theta})$	7970.87	8000.42
término penalización AICc	259.09	216.52
AICc	8229.96	8216.94
término penalización BIC	559.18	488.55
BIC	8530.05	8488.97
$E(D(\psi, \Sigma, \theta) y)$	8059.18	8084.43
P_D	88.31	84.01
DIC	8147.49	8168.44

Tabla 5.7: Criterios de selección para el modelo de Holt-Winters multivariante en el análisis conjunto de las series temporales del número de pasajeros en cinco aeropuertos de Londres

En la figura 5.5 encontramos, en color negro, el ajuste obtenido para cada una de las series temporales en la zona de estimación (2001 - 2006)

así como las predicciones puntuales para el año 2007 con el modelo \mathcal{M}_1 . Las líneas negras discontinuas representan los intervalos de predicción del 95%.

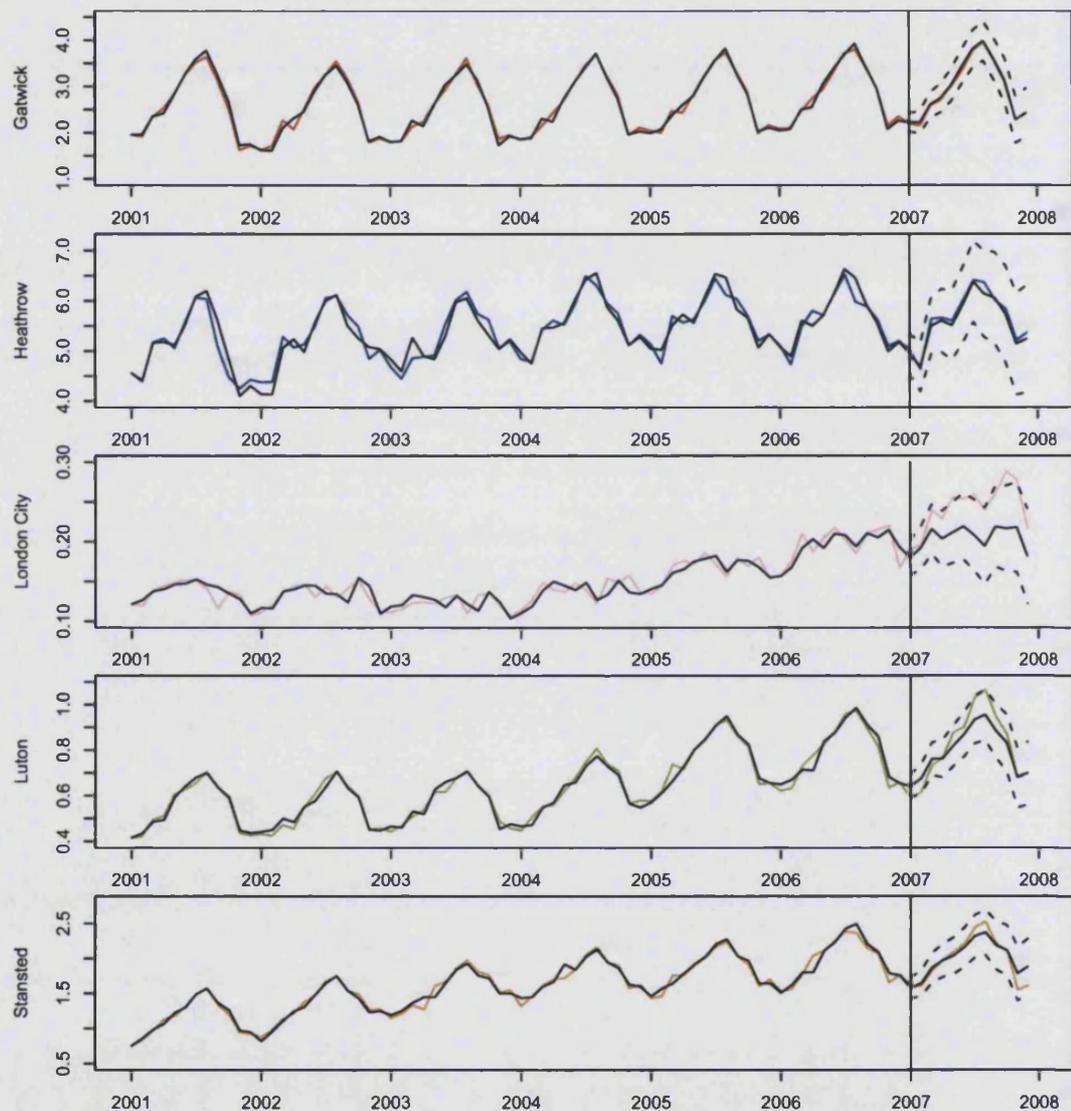


Figura 5.5: Ajuste y predicciones puntuales obtenidas, con el modelo de Holt-Winters multivariante con parámetros de suavizado iguales, para las series temporales del número de pasajeros en cinco aeropuertos de Londres. Líneas discontinuas: intervalos de predicción del 95 %

Los errores SMAPE de ajuste y de predicción resultantes del análisis multivariante de las series temporales mediante el modelo \mathcal{M}_1 se recogen en la tabla 5.8. A modo comparativo, los errores obtenidos en el análisis

multivariante con el modelo general, \mathcal{M}_0 , y en los análisis univariantes de las series son incluidos en la tabla.

Error SMAPE ajuste						
	Gatwick	Heathrow	London City	Luton	Stansted	Media
H-W multivariante \mathcal{M}_1	2.65	2.43	5.09	2.74	2.87	3.16
H-W multivariante \mathcal{M}_0	2.74	2.59	5.07	2.69	2.81	3.18
H-W univariante	2.35	2.48	5.03	2.51	2.66	3.01
Error SMAPE predicción						
H-W multivariante \mathcal{M}_1	0.88	1.65	16.32	5.48	4.40	5.75
H-W multivariante \mathcal{M}_0	2.09	2.44	16.36	7.79	4.37	6.61
H-W univariante	2.05	1.54	17.46	5.22	4.24	6.10

Tabla 5.8: Errores de ajuste y de predicción resultantes de los análisis multivariantes y univariantes de las series temporales del número de pasajeros en cinco aeropuertos de Londres

A partir de estos errores podemos concluir que, mediante el modelo de Holt-Winters multivariante con parámetros de suavizado iguales, \mathcal{M}_1 , se obtienen las predicciones más precisas, mejorando incluso las resultantes de los análisis univariantes, mientras que el ajuste es del mismo orden. Por otro lado, el tiempo de cómputo requerido para el análisis multivariante se ha reducido, aproximadamente, en un 62 % respecto a los análisis univariantes. Así pues, el análisis multivariante, además de ser más rápido, mejora la predicción respecto a los análisis univariantes.

Una vez más, los criterios AICc y BIC producen resultados más satisfactorios que el criterio DIC, que selecciona al modelo más complejo, \mathcal{M}_0 . El ajuste y las predicciones obtenidos con este modelo son similares a los de los análisis univariantes, mientras que el tiempo empleado en el análisis es mucho mayor.

Finalmente, mostramos los resultados obtenidos cuando la serie temporal de London City es analizada de manera independiente con el modelo de Holt-Winters univariante y las demás series conjuntamente a partir del modelo multivariante pues, como hemos comentado anteriormente, sólo podemos asegurar la correlación de esta serie con la de Heathrow y, por tanto, su inclusión en el modelo multivariante podría no resultar adecuada. En primer lugar seleccionamos entre el modelo de Holt-Winters multivariante general, \mathcal{M}_0 , y el modelo con parámetros de suavizado iguales, \mathcal{M}_1 , resultantes de la modelización conjunta de las series temporales de Gatwick, Heathrow, Luton y Stansted. Repitiendo los procedimientos de simulación descritos anteriormente, obtenemos una muestra de las distribuciones a pos-

teriori de los parámetros de cada modelo que nos permiten calcular el valor de los criterios de selección correspondientes a \mathcal{M}_0 y \mathcal{M}_1 , ver tabla 5.9.

	\mathcal{M}_0	\mathcal{M}_1
número parámetros	74	65
$D(\bar{\psi}, \bar{\Sigma}, \bar{\theta})$	6606.21	6569.97
término penalización AICc	200.11	168.65
AICc	6806.32	6738.62
término penalización BIC	419.06	368.09
BIC	7025.27	6938.06
$E(D(\psi, \Sigma, \theta) y)$	6666.07	6634.46
P_D	59.86	64.49
DIC	6725.93	6698.95

Tabla 5.9: Criterios de selección para el modelo de Holt-Winters multivariante en el análisis conjunto de las series temporales del número de pasajeros en Gatwick, Heathrow, Luton y Stansted

El modelo seleccionado para el posterior análisis de las series es, independientemente del criterio empleado, el modelo \mathcal{M}_1 . Los errores de ajuste y de predicción obtenidos en el análisis se muestran en la tabla 5.10. A modo comparativo, los errores resultantes de analizar las series con el modelo de Holt-Winters multivariante general \mathcal{M}_0 son incluidos en la tabla. Notar que para la serie temporal de London City, los resultados son los obtenidos en el análisis univariante.

Error SMAPE ajuste						
	Gatwick	Heathrow	London City	Luton	Stansted	Media
H-W multivariante \mathcal{M}_1	2.28	2.19	5.03	2.53	2.68	2.94
H-W multivariante \mathcal{M}_0	2.67	2.39	5.03	2.68	2.76	3.11
Error SMAPE predicción						
H-W multivariante \mathcal{M}_1	2.02	1.73	17.46	5.22	4.29	6.14
H-W multivariante \mathcal{M}_0	0.96	2.28	17.46	8.22	4.43	6.67

Tabla 5.10: Errores de ajuste y de predicción resultantes del análisis conjunto de las series temporales del número de pasajeros de Gatwick, Heathrow, Luton y Stansted mediante el modelo de Holt-Winters multivariante y del análisis univariante de la serie de London City

Como podemos ver, los resultados obtenidos en la predicción son ligeramente peores que los obtenidos cuando las cinco series temporales son

analizadas de manera conjunta mediante el modelo de Holt-Winters multivariante (ver tabla 5.8). Así pues, tal y como proponíamos en un principio, el análisis univariante de alguna de las series temporales se recomienda solamente cuando ésta, claramente, no presenta correlación con ninguna de las demás series. En otro caso, el modelo multivariante, al incorporar la correlación entre las series, puede dar lugar a un mejor ajuste y predicción.

Capítulo 6

Conclusiones y futuras líneas de investigación

6.1. Conclusiones

En esta memoria presentamos una nueva formulación para el modelo de Holt-Winters multivariante con estacionalidad aditiva y errores aditivos que nos permite obtener predicciones precisas, tanto puntuales como intervalos de predicción, de un conjunto de series temporales con errores correlacionados para cada instante temporal.

Trabajando con alguna transformación de los datos, como por ejemplo la logarítmica, el modelo de Holt-Winters aditivo resulta adecuado para el análisis de series temporales donde la componente estacional o la varianza del error dependen del nivel de la serie, es decir, series que se ajustan al modelo de Holt-Winters multiplicativo.

La formulación del modelo de Holt-Winters multivariante aditivo como un modelo de regresión aparentemente no relacionado, modelo SUR, facilita su análisis Bayesiano. A partir de distribuciones a priori convencionales no informativas obtenemos la distribución a posteriori para los parámetros del modelo que, aunque no es analíticamente tratable, puede ser estimada mediante métodos de simulación. En concreto, proponemos un algoritmo de Metropolis-within-Gibbs que nos permite simular de las condicionales completas a posteriori. Para el análisis conjunto de un número moderado de series temporales se recomienda el uso de la distribución a priori de Jeffreys, pues produce resultados similares a los obtenidos con la distribución de re-

ferencia propuesta por Yang and Berger (1994), mientras que el tiempo de computación requerido para obtener una muestra de la distribución a posteriori correspondiente disminuye ligeramente. Finalmente, la distribución predictiva, que contiene toda la información acerca de los valores futuros de las series temporales, es estimada mediante integración por Monte Carlo.

En el caso de parámetros de suavizado iguales para los modelos univariantes, el modelo de Holt-Winters multivariante puede ser formulado como un modelo de regresión multivariante tradicional, lo que simplifica considerablemente su análisis. La distribución a posteriori de los parámetros del modelo sigue siendo analíticamente intratable pero, en este caso, sólo es necesario obtener una muestra de la distribución a posteriori marginal del vector θ de los parámetros de suavizado comunes. Las distribuciones a posteriori marginales para el vector de las condiciones iniciales ψ y la matriz de covarianzas Σ pueden ser estimadas mediante integración por Monte Carlo. Además, para obtener una muestra de $f(\theta|y)$, al tratarse de una función tridimensional con soporte acotado, el cubo unidad, es posible emplear un algoritmo de aceptación-rechazo que, además de no presentar problemas de convergencia, es fácil de implementar y reduce el tiempo de computación considerablemente.

Así pues, una vez se ha decidido analizar las series temporales conjuntamente mediante el modelo multivariante, es necesario seleccionar entre el modelo de Holt-Winters multivariante general y el modelo con parámetros de suavizado iguales. Tras un estudio de los principales procedimientos Bayesianos de selección de modelos, proponemos emplear los criterios AICc o BIC, pues resultan sencillos de implementar y satisfactorios en los ejemplos numéricos presentados en la tesis.

Los resultados obtenidos en la predicción de los tres conjuntos de series temporales analizados en la memoria son alentadores. Mediante el modelo multivariante podemos obtener predicciones más precisas que las resultantes de los análisis univariantes, obteniendo errores de predicción del mismo orden en el peor de los casos. Además, cuando las series temporales son analizadas conjuntamente a partir del modelo de Holt-Winters multivariante con parámetros de suavizado iguales, el tiempo empleado para su análisis es considerablemente menor que el requerido cuando las series son analizadas de manera independiente. El motivo radica en la necesidad de obtener, en los análisis univariantes, una muestra de la distribución a posteriori de cada uno de los vectores de parámetros de suavizado, $f(\theta_i|y_i)$ para $i = 1, 2, \dots, m$,

mientras que en el análisis multivariante sólo se obtiene una muestra de la distribución a posteriori del vector de los parámetros de suavizado comunes, $f(\theta|y)$. Por lo tanto, cuando estudiamos series temporales con errores correlacionados y se cumple que los parámetros de suavizado de los modelos univariantes son iguales, el análisis multivariante es más rápido que los análisis univariantes, siendo posible además mejorar el ajuste y la predicción.

Por último, cuando las series temporales son observadas secuencialmente, como ocurre en el ejemplo de las series temporales del valor de cierre del Dow Jones y del Ibex 35, el empleo del modelo multivariante permite modificar las predicciones de un paso obtenidas originalmente para los valores futuros de las series en función de los errores de predicción de un paso disponibles. En consecuencia, las predicciones resultantes son mucho más precisas que las derivadas de los análisis univariantes, pues éstos no incorporan información acerca del comportamiento de otras series.

6.2. Futuras líneas de investigación

La realización de este trabajo nos ha abierto nuevos horizontes en los que seguir investigando. En primer lugar nos planteamos como futura línea de investigación el desarrollo de procedimientos de simulación más sofisticados que nos permitan superar las limitaciones de cálculo existentes. Como hemos mencionado anteriormente, el procedimiento de simulación propuesto para obtener una muestra de la distribución a posteriori, especialmente con el modelo de Holt-Winters multivariante general, requiere demasiado tiempo de computación para su utilización rutinaria.

Otro campo de actuación se centraría en el desarrollo de una herramienta de selección de modelos, fácil de implementar y alternativa a los criterios de selección, que nos permita resolver el problema de selección de modelos asociado al modelo de Holt-Winters multivariante. Aunque los criterios AICc y BIC seleccionan al modelo adecuado en los ejemplos numéricos analizados en la memoria, constatamos el problema que presentan estos procedimientos de selección. A pesar de que el modelo seleccionado para el posterior análisis de las series temporales sea el modelo de Holt-Winters multivariante con parámetros de suavizado iguales, el cual reduce considerablemente el tiempo de análisis, es necesario obtener una muestra de la distribución a posteriori para los parámetros de cada modelo, con lo que realmente no se disminuye el tiempo de computación.

**CAPÍTULO 6. CONCLUSIONES Y FUTURAS LÍNEAS DE
128 INVESTIGACIÓN**



Referencias

- H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.
- J.O. Berger and J.M. Bernardo. On the development of reference priors. *Bayesian Statistics*, 4:35–60, 1992.
- J.O. Berger and L.R. Pericchi. The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91:109–122, 1996.
- J.D. Bermúdez, J.V. Segura, and E. Vercher. Improving demand forecasting accuracy using non-linear programming software. *Journal of the Operational Research Society*, 57:94–100, 2006a.
- J.D. Bermúdez, J.V. Segura, and E. Vercher. A decision support system methodology for forecasting of time series based on soft computing. *Computational Statistics and Data Analysis*, 51:177–191, 2006b.
- J.D. Bermúdez, J.V. Segura, and E. Vercher. Holt-Winters forecasting: an alternative formulation applied to UK air passenger data. *Journal of Applied Statistics*, 34:1075–1090, 2007.
- J.D. Bermúdez, J.V. Segura, and E. Vercher. SIOPRED: a prediction and optimisation integrated system for demand. *TOP*, 16:258–271, 2008a.
- J.D. Bermúdez, J.V. Segura, and E. Vercher. Bayesian forecasting with the Holt-Winters model. *Journal of the Operational Research Society*, DOI: 10.1057/jors.2008.152, 2008b.
- J.M. Bernardo. Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B*, 41:113–147, 1979.
- J.M. Bernardo and R. Rueda. Bayesian hypothesis testing: A reference approach. *International Statistical Review*, 70:351–372, 2002.

- K.P. Burnham and D.R. Anderson. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33: 261–304, 2004.
- R.H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16:1190–1208, 1995.
- B.P. Carlin and S. Chib. Bayesian model choice via Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B*, 57: 473–484, 1995.
- T. Chang and D. Eaves. Reference priors for the orbit in a group model. *The Annals of Statistics*, 18:1595–1614, 1990.
- C. Chatfield and M. Yar. Holt-Winters forecasting: Some practical issues. *The Statistician*, 37:129–140, 1988.
- S. Chib and E. Greenberg. Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models. *Journal of Econometrics*, 68:339–360, 1995.
- M. Clyde and E.I. George. Model uncertainty. Technical report, Statistical and Applied Mathematical Sciences Institute, December 2003.
- F. De Santis and F. Spezzaferri. Alternative Bayes factors for model selection. *The Canadian Journal of Statistics*, 25:503–515, 1997.
- A. de Silva, R.J. Hyndman, and R.D. Snyder. The vector innovation structural time series framework: a simple approach to multivariate forecasting. Technical report, Department of Econometrics and Business Statistics, Monash University, Australia, May 2007.
- L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.
- T.J. DiCiccio, R.E. Kass, A. Raftery, and L. Wasserman. Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, 92:903–915, 1997.
- P.G. Enns, J.A. Machak, W.A. Spivey, and W.J. Wroblewski. Forecasting applications of an adaptive multiple exponential smoothing model. *Management Science*, 28:1035–1044, 1982.

- F.J. Fernández. Estimation and testing of a multivariate exponential smoothing model. *Journal of Time Series Analysis*, 11:89–105, 1990.
- F.J. Fernández and A.C. Harvey. Seemingly unrelated time series equations and a test for homogeneity. *Journal of Business & Economic Statistics*, 8:71–81, 1990.
- E.S. Gardner Jr. Exponential Smoothing: The State of the Art. *Journal of Forecasting*, 4:1–28, 1985.
- E.S. Gardner Jr. Exponential Smoothing: The State of the Art - Part II. *International Journal of Forecasting*, 22:637–666, 2006.
- J.F. Geweke and C.H. Whiteman. *Bayesian Forecasting*, en: G. Elliot, C.W.J. Granger and A. Timmermann (Eds.), *The Handbook of Economic Forecasting*. North-Holland, Amsterdam, 2006.
- C. Goutis and C.P. Robert. Model choice in generalised linear models: A Bayesian approach via Kullback-Leibler projections. *Biometrika*, 85:29–37, 1998.
- P. Green. Reversible Jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.
- W.E. Griffiths. Bayesian inference in the Seemingly Unrelated Regression Model. Technical report, Department of Economics, The University of Melbourne, June 2001.
- A.C. Harvey. Analysis and generalisation of a multivariate exponential smoothing model. *Management Science*, 32:374–380, 1986.
- H.V. Henderson and S.R. Searle. Vec and Vech operators for matrices, with some uses in Jacobians and multivariate Statistics. *The Canadian Journal of Statistics*, 7:65–81, 1979.
- C.C. Holmes, D.G.T. Denison, and B.K. Mallick. Accounting for model uncertainty in Seemingly Unrelated Regressions. *Journal of Computational and Graphical Statistics*, 11:533–551, 2002.
- C.C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20:5–10, 2004.
- R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1994.

- C.M. Hurvich and C.L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76:297–307, 1989.
- R.J. Hyndman, A.B. Koehler, R.D. Snyder, and S. Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18:439–454, 2002.
- R.H. Jones. Exponential smoothing for multivariate time series. *Journal of the Royal Statistical Society. Series B*, 28:241–251, 1966.
- R.E. Kass and A.E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995.
- J.S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 2001.
- S. Makridakis and M. Hibon. The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16:451–476, 2000.
- K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.
- A.A. Neath and J.E. Cavanaugh. Regression and time series model selection using variants of the Schwarz information criterion. *Communications in Statistics - Theory and Methods*, 26:559–580, 1997.
- A. O’Hagan. Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society. Series B*, 57:99–138, 1995.
- A. O’Hagan and J. Foster. *Kendall’s Advanced Theory of Statistics, Volume 2B: Bayesian Inference*. Arnold, London, 2nd ed., 2004.
- J.K. Ord, A.B. Koehler, and R.D. Snyder. Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. *Journal of the American Statistical Association*, 92:1621–1629, 1997.
- K. Ord. Charles Holt’s report on exponentially weighted moving averages: an introduction and appreciation. *International Journal of Forecasting*, 20:1–3, 2004.
- D.F. Percy. Prediction for Seemingly Unrelated Regressions. *Journal of the Royal Statistical Society. Series B*, 54:243–252, 1992.
- D. Pfeiffermann and J. Allon. Multivariate exponential smoothing: Method and practice. *International Journal of Forecasting*, 5:83–98, 1989.

- J.M. Pérez and J.O. Berger. Expected-posterior prior distributions for model selection. *Biometrika*, 89:491–511, 2002.
- R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL <http://www.R-project.org>. 3-900051-07-0.
- A.E. Raftery. *Hypothesis Testing and Model Selection*, en: W.R. Gilks, S. Richardson and D. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, 1996.
- C.P. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer, New York, 2007.
- C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, 1999.
- G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.
- J.V. Segura and E. Vercher. A spreadsheet modeling approach to the Holt-Winters optimal forecasting. *European Journal of Operational Research*, 131:375–388, 2001.
- M. Smith and R. Kohn. Nonparametric seemingly unrelated regression. *Journal of Econometrics*, 98:257–281, 2000.
- D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B*, 64:583–639, 2002.
- D. Sun and J.O. Berger. Objective priors for the multivariate Normal model. *Proceedings Valencia / ISBA 8th World Meeting on Bayesian Statistics*, 2006.
- G.C. Tiao and A. Zellner. On the Bayesian estimation of multivariate regression. *Journal of the Royal Statistical Society. Series B*, 26:277–285, 1964.
- N.H. Timm. *Applied Multivariate Analysis*. Springer-Verlag, New York, 2002.
- M. West and J. Harrison. *Bayesian forecasting and Dynamic models*. Springer-Verlag, New York, 1989.

-
- P.R. Winters. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6:324–342, 1960.
- R. Yang and J.O. Berger. Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, 22:1195–1211, 1994.
- A. Zellner. *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons, New York, 1971.
- A. Zellner. An efficient method of estimating Seemingly Unrelated Regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57:348–368, 1962.

UNIVERSITAT DE VALÈNCIA
FACULTAT DE MATEMÀTIQUES

reunido el tribunal que suscribe, en el día de la fecha,
acordó otorgar a esta Tesis Doctoral de D^o ANA
COBBERÁN VALLET
la calificación de EXCELENTE CUM LAURE
Valencia, a 27 de MARZO 2009



El Presidente/a,

El Secretario/a,

