

EXTRACCIÓN DE SEÑAL Y PREDICCIÓN EN SERIES TURÍSTICAS

*José Mondéjar Jiménez, Manuel Vargas Vargas, Juan Antonio Mondéjar
Jiménez y Carlota Lorenzo Romero*
Universidad de Castilla-La Mancha

RESUMEN

En los últimos años existe un creciente interés en materia de investigación turística, en un sector dinámico y de comportamiento especialmente complejo. En este escenario resulta de especial dificultad el manejo de las series temporales disponibles. La extracción de señal relevante en las series así como la realización de predicciones del comportamiento futuro de las cifras turísticas, cobran una especial importancia. En este trabajo se pretende realizar una comparación entre los métodos tradicionales y bayesianos de extracción de señal, de forma que el análisis empírico se aplica a las series turísticas más representativas de todas las provincias de la comunidad autónoma de Castilla-La Mancha. Los trabajos más frecuentes en la materia han abordado un análisis de estas series temporales a nivel internacional, nacional o regional; por ello, al descender a un nivel de desagregación provincial, los datos muestran mayores fluctuaciones y patrones de comportamiento y, por consiguiente, su modelización y predicción a futuro se hace más compleja.

Palabras clave: turismo, series temporales, extracción de señal.

Sign extraction and prediction in tourist series

ABSTRACT

In the last several years, there has been an increasing interest in the field of tourism research, a dynamic, particularly complex sector where the use of available temporary series proves to be especially difficult. Relevant sign extractions, as well as predictions of future behavior of tourism figures, take on special significance. The purpose of this project is to carry out a comparison between traditional and bayesian sign extraction methods, thus

Fecha de recepción: 5 de febrero de 2007

Fecha de aceptación: 23 de octubre de 2007

Facultad de Ciencias Sociales. Universidad de Castilla-La Mancha. Avda de los Alfares, 44. 16071 CUENCA (España). E-mail: Jose.Mondejar@uclm.es; Manuel.Vargas@uclm.es; Juan.Antonio.Mondejar@uclm.es; Carlota.Lorenzo@uclm.es

applying empirical analysis to the tourist series which best represent all provinces in the region of Castilla-La Mancha. Other studies on this subject have commonly tackled the analysis of these temporary series at international, national and regional levels. It is for this reason that, when dealing with separate provinces, the data show bigger fluctuations and different behavioral patterns, making modelling and future prediction more complex.

Key words: tourism, time series, signal extraction.

1. INTRODUCCIÓN

Los investigadores han mostrado en los últimos años su reiterado interés por analizar el sector turístico. Concretamente, la relación entre la investigación sobre turismo ha ido aumentando como muestra del creciente interés de los investigadores y permitiendo, a su vez, mejorar su gestión (Sánchez y Marín, 2003). La principal razón de esta proliferación de trabajos está directamente relacionada con la introducción de los estudios de turismo en la Universidad española (Esteban, 2000).

La propia naturaleza del turismo como conjunto de actividades relativas a alojamiento, transporte, animación, promoción, comercialización, entre otras, así como la amplia gama de dimensiones que confluyen en la configuración del producto argumentan sobradamente su interpretación como hecho multiforme. De esta forma, su estudio se aborda desde distintos ámbitos disciplinares que tratan de analizar el turismo con sus correspondientes metodologías y herramientas y desde sus propios fundamentos epistemológicos (Vera e Ivars, 2001).

El sector turístico se ha convertido en uno de los principales sectores a nivel nacional y, particularmente, en Castilla-La Mancha, donde día tras día se consolida como uno de los sectores con más peso relativo en su economía, desbancando a otros sectores históricamente más fuertes como son la industria y la agricultura. A partir de este creciente interés por analizar el sector, surge la necesidad de establecer herramientas que permitan canalizar la gran cantidad de información que, a nivel institucional, se deriva de las estadísticas oficiales.

En este sentido, cabe destacar el interés suscitado en la aplicación de técnicas estadísticas al sector turístico y, más específicamente, la modelización y predicción de cualquiera de las estadísticas oficiales existentes. Así, el Instituto Nacional de Estadística (INE) elabora numerosas estadísticas oficiales relacionadas con el sector, a saber: encuesta de ocupación hotelera, encuesta de ocupación en acampamentos turísticos, encuesta de ocupación en apartamentos turísticos, encuesta de ocupación en alojamientos de turismo rural, índice de precios hoteleros e índice de ingresos hoteleros, índice de precios de acampamentos y de apartamentos turísticos, encuesta sobre la estructura de las empresas hoteleras, encuesta sobre la estructura de empresas de agencias de viajes, etc... Además, el Instituto de Estudios Turísticos (IET) analiza turismo receptor y turismo nacional en base a sus encuestas FRONTUR y FAMILITUR junto a otras muchas estadísticas oficiales, de gran interés para los investigadores de la materia.

Es por ello que en los últimos años han proliferado los trabajos publicados en este ámbito, en todos los niveles de desagregación: regional, nacional e internacional (por

ejemplo, Daniel y Ramos, 2002; Durán y Flores, 1998; Esteban, 1996; Ledesma y Navarro, 2000; Lim, 1997; Otero, 1996, 1999; Sorensen, 2003 y Zou y Yang, 2004). Dentro de los diversos métodos de previsión, las series temporales constituyen una herramienta sencilla y precisa como punto de partida en investigaciones con técnicas más complejas (modelos econométricos). Estos métodos de previsión univariantes son empleados fundamentalmente para la predicción de algunas variables a corto y medio plazo, que puedan conservar las características de evolución del pasado en momentos futuros, donde estos métodos suelen obtener muy buenos resultados.

Por tanto, el análisis de series temporales proporciona una herramienta ideal para el tratamiento de series turísticas, demostrándose su capacidad predicativa y su utilidad en la toma de decisiones (Zou y Yang, 2004).

Según Song, Witt y Jensen (2003), el mejor rendimiento de las series temporales en detrimento de otros métodos puede deberse, fundamentalmente, a tres motivos:

- El rendimiento de los modelos econométricos es muy sensible a las diferentes metodologías utilizadas.
- La existencia de diferentes frecuencias en los datos puede dar lugar a diferentes interpretaciones.
- Los modelos econométricos trabajan usualmente con hipótesis de partida tales como permanencia de la estructura del modelo en el tiempo, hecho este que no es aplicable a sectores como el turismo, debido a la existencia de una constante evolución en el mismo (Lise y Tol, 1999).

2. EXTRACCIÓN DE SEÑAL EN SERIES TEMPORALES

El concepto de filtrado incluye una serie de procedimientos derivados del hecho de que la señal que transportan los datos está contaminada con componentes que no son deseables y que se pueden considerar ruido en este contexto. Se trata, por tanto, de limpiar esta señal, es decir, de depurar la serie de los componentes irregulares y estacionales.

Lo que se entiende por ruido, sin embargo, no es concebido del mismo modo por todos los autores. Así, por ejemplo, Gómez y Maravall (1998) o Espasa y Cancelo (1993) comienzan el planteamiento depurando la estacionalidad y los componentes irregulares para obtener el componente ciclo tendencia. Sin embargo, otros autores como Baxter y King (1995) no tratan el problema de la estacionalidad y se preocupan de obtener el componente cíclico a través de un filtro, imponiendo restricciones de estacionalidad y eliminando componentes irregulares. En un tercer enfoque, Canova (1998) plantea el problema de una forma diferente: se trata de extraer el componente tendencial de la propia serie ciclo-tendencia, obteniendo como residuo el componente puramente cíclico.

Estos distintos enfoques pueden parecer complementarios y no excluyentes desde el momento en que la extracción de ambos componentes, irregular y estacional, proporciona una serie de ciclo-tendencia sobre la que aplicar una segunda fase de extracción de ciclo y tendencia. Sin embargo, considerar la aplicación de los distintos filtros de forma independiente puede llevar a dos tipos de problemas: el primero es la intervención sobre los componentes estacionales que, sistemáticamente, han recibido un tratamiento

aislado y diferente del filtrado y la extracción del resto de los componentes. Algunos de los filtros de desestacionalización, típicamente el procedimiento X-11, suponen aplicar tratamientos previos para depurar la serie original de los componentes estacionales. Esto conduce al segundo problema: el derivado de las consecuencias de aplicar filtros en cascada.

Podemos, por tanto, caracterizar los diferentes procedimientos de filtrado, describiendo sus efectos desde una perspectiva de análisis espectral, es decir, caracterizándolos como filtros de frecuencias. En la propia noción de filtro subyace la idea de dejar pasar determinadas frecuencias y eliminar otras. Un filtro ideal es aquel que está diseñado para permitir el paso de información en una banda de frecuencias y eliminar o inhibir la información contenida en otras bandas de frecuencia consideradas no deseables. Se trata de analizar qué tipo de frecuencia deja pasar cada filtro y cuáles de ellas son eliminadas, así como la eficacia relativa de cada filtro respecto a las frecuencias filtradas.

Cualquier filtro tiene un coste informativo que se traduce en las observaciones que se pierden en el tramo final e inicial que hacen necesario sustituir los datos por predicciones provisionales sujetas a revisión. No es nada nuevo en el trabajo de los analistas de coyuntura tener que recalcular las series filtradas cuando llega un nuevo dato que, a veces, llega a alterar de forma dramática los resultados previamente obtenidos. Este desfase temporal de las series filtradas será analizado mediante el análisis de fase. Se trata, pues, de clarificar las consecuencias de la aplicación de diferentes filtros ya que como señala Melis (1991): *«El desarrollo de las técnicas ARIMA ha contribuido indirectamente, al facilitar la predicción a corto, a oscurecer la frontera entre señal y predicción»*.

Pasaremos, a continuación, a exponer brevemente los métodos más importantes de filtrado y sus características, clasificando los filtros en distintos tipos (tan solo se hará un desarrollo más profundo del filtro bayesiano):

2.1. Filtros autoregresivos (AR)

El filtro univariante más sencillo es el filtro AR, es decir, el operador diferencia. Dentro de los filtros AR el más simple es AR(1), es decir, la primera diferencia ordinaria. Este tipo de filtros tiene unas propiedades características: cuando se aplica sobre series que son estacionarias es un filtro de paso alto, es decir, elimina las frecuencias bajas pero acentúa las frecuencias altas. En series suaves permite obtener estimaciones de los componentes estacional e irregular y aproximar la tasa de crecimiento de la serie original cuando se filtra el logaritmo y, por lo tanto, ha sido utilizado como indicador del ciclo cuando las series tienen comportamientos suaves. Sin embargo, la vasta utilización que se ha hecho de este filtro se debe a que convierte en estacionarias series I(1), es decir, series integradas de orden 1, permitiendo la utilización del análisis ARMA para la predicción y estimación. La clave está en analizar correctamente el comportamiento de la serie original. Para cualquier comportamiento de la serie original, el componente cíclico debe ser en todo caso estacionario. Sin embargo, el componente tendencial puede ser estacionario o no. Según cuál sea el comportamiento de la tendencia estaremos ante un ciclo clásico o de crecimiento.

2.2. Filtros aditivos o de medias móviles (MA)

Son uno de los filtros más usados para extraer componentes en análisis del ciclo. En términos frecuenciales son filtros pasabanda, es decir, filtros que dejan pasar intacta la información contenida en determinada banda de frecuencias mientras que eliminan o acentúan las restantes. Se trata de eliminar movimientos tendenciales correspondientes a las frecuencias más bajas y los movimientos irregulares o vinculados a las frecuencias más altas. En cualquier proceso MA(m), la función de ganancia cuando los pesos son iguales es la siguiente:

$$G(w) = \frac{\text{sen}(mw/2)}{m \text{sen}(w/2)}$$

Esta función presenta ceros en $\frac{2\pi k}{m}$, $k=1,2,\dots$ y máximos locales (lóbulos laterales) en $\frac{\pi(2k+1)}{m}$, $k=1,2,\dots$. Precisamente estas bandas típicas de las medias móviles simples suponen el principal inconveniente para su uso como filtros, ya que parte de la varianza de la serie original pasa a la serie filtrada contaminándola de oscilaciones cortas de alta frecuencia. Una de las medias más usadas es la media móvil de periodo estacional. Se trata de uno de los filtros más sencillos para desestacionalizar extrayendo la componente ciclo-tendencia. La función de transferencia tiene ceros en los armónicos estacionales, aunque los lóbulos laterales que induce hacen que esté muy alejado del filtro ideal.

2.3. Filtro bayesiano en espacio de estados

El planteamiento bayesiano del filtrado de indicadores está íntimamente ligado a los modelos dinámicos en espacio de estados. El aprendizaje bayesiano combina, mediante el teorema de Bayes, la información de las observaciones, expresada por la función de verosimilitud, con el estado de conocimiento del analista antes de disponer de las observaciones; se dispone así de un mecanismo de actualización del conocimiento sobre la serie que se puede esquematizar como sigue:

1.- Dado el indicador Y_t y su señal de ciclo-tendencia, θ_t , se considera que su relación está perturbada por un ruido, y que la señal presenta una evolución markoviana como se recoge en la siguiente ecuación:

$$Y_t = F \theta_t + V_t$$

$$\theta_{t-1} = G \theta_t + W_t$$

2.- Denotemos por D_{t-1} el conocimiento existente en el instante (t-1). Podemos expresar la información a priori sobre la señal en el instante t mediante la distribución de θ_t/D_{t-1} .

3.- La predicción a partir de la información a priori se generará a través de la ecuación de observación

$$Y_t/D_{t-1} = (F_t \theta_t / D_{t-1}) + (V_t / D_{t-1}) = F_t (\theta_t / D_{t-1}) + (V_t / D_{t-1})$$

4.- Ahora, considerando $D_t = D_{t-1} \cup Y_t$, podemos obtener la distribución a posteriori, que será proporcional al producto de la distribución a priori por la verosimilitud.

$$(\theta_t / D_t) \propto (Y_t / \{\theta_t, V_t\}) \cdot (\theta_t / D_{t-1})$$

5.- Por último, utilizando la ecuación de transición, podemos obtener la distribución a priori para la señal en el instante (t+1) como

$$(\theta_{t+1} / D_t) = G_t (\theta_t / D_t) + (W_t / D_t)$$

cerrándose así el ciclo «a priori - verosimilitud - a posteriori - a priori», fundamento del aprendizaje bayesiano.

Esta formulación permite una gran flexibilidad, tanto de análisis como de intervención, propiedad deseable dentro de un contexto dinámico como son los indicadores de coyuntura.

Asumiendo la normalidad para la distribución a priori y mediante el algoritmo expuesto, podemos actualizar el conocimiento sobre la serie de ciclo-tendencia, quedándonos las distribuciones a posteriori

$$(\theta_t / D_t) \rightarrow t_n [m_t, C_t]$$

$$(\varphi_t / D_t) \rightarrow G \left[\frac{n_t}{2}, \frac{d_t}{2} \right]$$

donde las actualizaciones son

$$m_t = a_t + A_t e_t \quad C_t = \left(\frac{S_t}{S_{t-1}} \right) [R_t - A_t A_t' Q_t]$$

$$e_t = Y_t - f_t \quad A_t = \frac{R_t F_t}{Q_t}$$

$$n_t = \delta_t n_{t-1} + 1 \quad d_t = \delta_t d_t + \frac{S_{t-1} e_t^2}{Q_t}$$

completándose así el proceso de aprendizaje de los datos.

Hemos introducido toda la información disponible en la serie dentro de nuestro modelo de una forma coherente y bastante versátil. En particular, la formulación de un modelo lineal dinámico permite la introducción de información extramuestral en cualquier instante del período con sólo alterar la distribución a priori en ese período.

Otra ventaja interesante es la detección y tratamiento de anómalos. Supongamos que llamamos M al modelo propuesto por el algoritmo anterior y M_A a un modelo alternativo. Si definimos ahora

$$H_t = \frac{p(Y_t/D_{t-1}, M)}{p(Y_t/D_{t-1}, M_A)}$$

llamado factor de Bayes, este cociente nos permite cuantificar la evidencia a favor o en contra del modelo propuesto frente al alternativo. Con este mecanismo, Jeffreys (1961) sugiere que existe evidencia a favor del modelo alternativo.

$$H_{t\wedge}(k) = \frac{p(Y_t, Y_{t-1}, \dots, Y_{t-k+1}/D_{t-k}, M)}{p(Y_t, Y_{t-1}, \dots, Y_{t-k+1}/D_{t-k}, M_A)}$$

De esta forma, podemos detectar la presencia de outlier cuando el factor de Bayes indique evidencia contraria al modelo propuesto, siendo el modelo alternativo el resultante de la eliminación de la información contenida en el dato presuntamente anómalo. El hecho de tomar k períodos permite la detección de sólo aquellos anómalos que presentan influencia relativamente importante.

En el caso de estar interesados en un horizonte de predicción más amplio, $k \geq 1$, las distribuciones predictivas proporcionadas por el modelo lineal dinámico serían

$$\begin{aligned} (\theta_{t+k}/D_t) &\rightarrow t_{\delta m} [a_t(k), R_t(k)] \\ (Y_{t+k}/D_t) &\rightarrow t_{\delta m} [f_t(k), Q_t(k)] \end{aligned}$$

donde las actualizaciones futuras se hacen en función de

$$\begin{aligned} a_t(k) &= G_{t+k} a_t(k-1) & a_t(0) &= m_t \\ R_t(k) &= G_{t+k} R_t(k-1) G'_{t+k} + W_{t+k} & R_t(0) &= C_t \\ f_t(k) &= F'_{t+k} a_t(k) & Q_t(k) &= F'_{t+k} R_t(k) F_{t+k} + S_t \end{aligned}$$

Por último, para obtener la distribución del ciclo-tendencia histórica, basta con retrotraer la información desde el último periodo disponible, utilizando las distribuciones:

$$\begin{aligned} (\theta_{t-k}/D_t) &\rightarrow t_{m(-k)} [a_t(-k), R_t(-k)] \\ (\varphi_{t-k}/D_t) &\rightarrow G \left[\frac{n_t(-k)}{2}, \frac{d_t(-k)}{2} \right] \end{aligned}$$

donde se tiene que

$$\begin{aligned}
 a_t(-k) &= m_t - k - B_{t-k} [a_{t-k+1}(-k) - a_t(-k+1)] \\
 R_t(-k) &= C_{t-k} - B_{t-k} [R_{t-k+1} - R_t(-k+1)] B'_{t-k} \\
 B_t &= C_t G'_{t+1} + R_{t+1}^{-1} \\
 n_t(-k) &= n_{t-k} + \delta_{t-k+1} [n_t(-k+1) - \delta_{t-k+1} n_{t-k}] \\
 S_t^{-1}(-k) &= S_{t+1}^{-1} + \delta_{t-k+1} (S_t^{-1}(-k+1) - R_{t-k}^{-1}) \\
 d_t(-k) &= n_t(-k) S_t(-k)
 \end{aligned}$$

Así, obtenemos una estimación suavizada del ciclo-tendencia dado el conjunto de información disponible¹.

2.4. Extracción de la señal relevante con Seats

Por todo lo comentado antes, nos basaremos en el enfoque basado en modelos ARIMA. Para la extracción de señal utilizaremos el programa SEATS (Gómez y Maravall, 1997). El nombre de este programa se corresponde con las letras iniciales de la traducción al inglés de 'extracción de señal en series de tiempo ARIMA' («Signal Extraction in ARIMA Time Series»). Vamos a describir de manera breve las utilidades que proporciona este programa con el objeto de explicar qué hace el programa. Para ello, aportamos unos extractos de la descripción del programa que Gómez y Maravall (1997) hacen²: «SEATS es un programa para la identificación de componentes no observados en series temporales siguiendo el enfoque llamado 'basado en modelos ARIMA'. Los componentes tendenciales, estacionales, irregulares, y cíclicos son estimados y predichos con técnicas de extracción de señal aplicadas a los modelos ARIMA. Son obtenidos los errores estándar de las estimaciones y predicciones y la estructura basada en modelo se explota para contestar a cuestiones de interés en el análisis en el corto plazo de los datos. (...) Cuando se usan [los programas TRAMO y SEATS] para ajuste estacional, TRAMO previamente adapta la serie que va a ser ajustada por SEATS.».

2.5. Otros filtros

En el proceso de filtrado es posible que estemos interesados en aplicar una combinación de varios métodos de filtrado. Un filtro compuesto es la aplicación de varios filtros simples sucesivamente sobre la serie original. El módulo de la función de transferencia resultado de la combinación de filtros tendrá un módulo igual al producto de los módulos de los filtros componentes y una fase igual a la suma de las fases de los mismos. La aplicación sucesiva de un filtro MA y un filtro AR tendrá como consecuencia respectiva la atenuación de altas frecuencias y la atenuación de bajas frecuencias, resultando un pico en la función de transferencia de la serie filtrada y una atenuación de determinadas frecuen-

1 Algunas aplicaciones de esta metodología pueden encontrarse en Mondéjar (2003 y 2006).

2 Una exposición adecuada requiere la lectura íntegra de dicha referencia y otras que allí se indican.

cias intermedias. Los filtros más ampliamente usados, sin embargo, son filtros ad-hoc, de expresión variable. Concretamente, filtros de la familia Butterworth a la cual pertenece el filtro de Hodrick y Prescott (1980), Prescott (1986) y los de tipo Henderson entre los que destaca el utilizado en el método de desestacionalización X-11.

El filtro de Hodrick y Prescott (HP) adopta la expresión:

$$y^c = \left(\frac{\lambda(1-L)^2(1-L^{-1})^2}{1 + \lambda(1-L)^2(1-L^{-1})^2} \right) y$$

Donde λ es el parámetro que penaliza la variación en el crecimiento y para el cual los autores recomiendan un valor arbitrario de $\lambda=1600$ cuando se utilizan datos trimestrales. Este filtro elimina el componente no estacionario, es simétrico y no causa desfase en la serie filtrada. Este filtro es parecido a un filtro de paso alto, es decir, asigna valor cero a la frecuencia 0 y valores cercanos a uno para valores de frecuencias altas. Precisamente para este valor de $\lambda=1600$, se aproxima a un filtro lineal tipo Butterworth que permite el paso de la información para frecuencias de corte $w = \frac{\pi}{16}$, es decir, permite el paso de información cíclica por encima de 32 trimestres. El filtro X-11, ampliamente usado para desestacionalizar las series económicas, es un filtro tipo Henderson cuya expresión inicial en su primera iteración en datos mensuales sería del tipo 1-MA(12). Se trata de otro filtro de paso alto centrado en el procedimiento, además, se combina con el análisis ARIMA utilizando estimaciones para los extremos de la media móvil que se pierden en el centrado.

3. SITUACIÓN ACTUAL DEL SECTOR TURÍSTICO EN CASTILLA-LA MANCHA

Dado que el trabajo empírico se centrará en las provincias de Castilla-La Mancha (Albacete, Ciudad Real, Cuenca, Guadalajara y Toledo), resulta conveniente matizar el crecimiento y el enorme interés de este sector en nuestra región. En este sentido, es preciso destacar que la única fuente de información sobre la demanda turística extrapolable al conjunto de provincias de la Comunidad es, a priori, la proporcionada por el INE y por los estudios del IET. En el ámbito provincial existen estadísticas de ciertas tipologías de alojamiento, pero ninguna de ellas es común para todas las provincias y, por tanto, no se pueden estimar datos para la Comunidad.

3.1. La demanda en Castilla-La Mancha

Castilla-La Mancha ha concluido el año 2006 con más de dos millones de viajeros y por encima de tres millones seiscientos mil pernoctaciones, según datos de la Encuesta de Ocupación Hotelera (EOH) del INE, lo que supone una variación interanual de más del 11% respecto al ejercicio 2005.

En este sentido, la región ocupa el noveno lugar como comunidad autónoma de destino de viajeros nacionales alojados en establecimientos hoteleros, con aproximadamente un

2,6% del total nacional, en cifras similares a las de Aragón y País Vasco. La procedencia de estos viajeros corresponden, mayoritariamente, a la Comunidad de Madrid (22,9%), Comunidad Valenciana (16,6%), la propia Castilla-La Mancha (15,5%) y Andalucía (12%). Sin embargo, los viajeros extranjeros suponen algo menos del 15% del total de viajeros recibido en la comunidad autónoma, y suponen casi el 1% del total recibido a nivel nacional.

Si tenemos en cuenta el número de pernoctaciones realizadas por los visitantes, Castilla-La Mancha se sitúa en undécimo lugar, representando el 1,46% del total nacional.

A tenor de estos datos, la estancia media en nuestra región es de 1,73 días, según datos del INE, frente a 3,25 días de la media para el total de España, lo que convierte a Castilla-La Mancha en un destino turístico de corta duración. Este aspecto nos lleva a considerar el importante número de turistas considerados como excursionistas, es decir, aquellos que visitan el destino turístico sin pernoctar y que, por consiguiente, no son recogidos en la estadística oficial del INE, de tal forma que podríamos estimar la cifra de visitantes que utilizan esta fórmula en nuestra región en el 25%.

Desde el punto de vista del análisis según el tipo de alojamiento utilizado, en los últimos años se observa claramente un retroceso en establecimientos hoteleros tradicionales (por ejemplo, hoteles y pensiones), en cambio aumentan nuevas fórmulas: concretamente, el turismo rural ha sufrido un importante incremento en Castilla-La Mancha del 16,7% en términos de la tasa de variación interanual del número de viajeros, y del 14,8% para el total de pernoctaciones.

En la curva de estacionalidad se pueden observar dos picos, uno menor en abril, coincidiendo con el periodo vacacional de Semana Santa, y otro más acusado en los meses de Julio a Septiembre. En los últimos años no ha habido grandes variaciones en la distribución de los viajeros por meses.

Tabla 1
NÚMERO DE ESTABLECIMIENTOS, PLAZAS Y PERSONAL
EMPLEADO. 2006

Provincias	Número establecimientos abiertos	Número de plazas estimadas	Total personal empleado
Albacete	125	4.834	520
Ciudad Real	159	6.395	715
Cuenca	136	4.713	518
Guadalajara	140	4.564	505
Toledo	191	8.382	1.171
Total Castilla-La Mancha	751	28.887	3.429

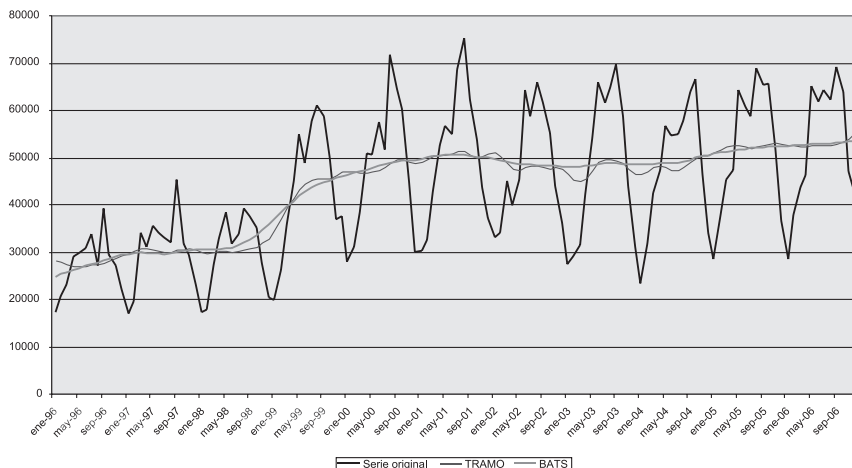
Fuente: Elaboración propia a partir de datos del INE (2007).

4. PREDICCIÓN DE SERIES TURÍSTICAS EN CASTILLA-LA MANCHA

En esta fase se pretende llevar a cabo una aplicación práctica con el fin de eliminar de cada serie de turismo el ruido inherente propio de cada una de ellas; dejando, por tanto, la componente ciclo-tendencia. Para realizar dicha extracción se ha considerado la posibilidad de utilización de un método clásico (ARIMA) frente a la metodología Bayesiana, más concretamente, utilizar el software *TRAMO-SEATS*, basado en un modelo ARIMA univariante para la serie, a partir del cual se obtienen modelos univariantes para los componentes ciclo-tendencia, estacional y ruido (metodología UCARIMA), frente a la extracción de señal proporcionada por *BATS*, modelización dinámico bayesiana para obtener el componente ciclo-tendencia de cada indicador mediante un modelo en espacio de estados.

Debido a las particularidades de las series del turismo, éstas se ven fuertemente afectadas por la estacionalidad (vacaciones de verano, puentes, festivos, etc...) y por un importante «efecto calendario» (de especial importancia para Castilla-La Mancha resulta, por ejemplo, el efecto de Pascua móvil). Es por lo que, para realizar una análisis más en profundidad, deberán extraerse las series ciclo-tendencia, tal y como podemos observar en las siguientes figuras:

Figura 1
EVOLUCIÓN MENSUAL DE LAS PernoctACIONES EN LA PROVINCIA DE ALBACETE (SERIE ORIGINAL Y CICLO-TENDENCIA)

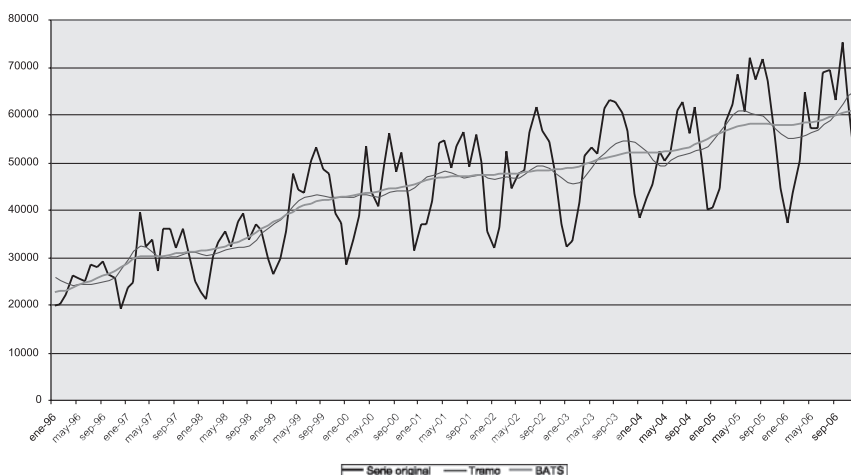


Fuente: Elaboración propia a partir de datos del INE (2007).

Como se observa en los gráficos, la serie original muestra una elevada componente estacional, derivada de las características del sector turístico que antes hemos descrito, y que, junto a la provincia de Cuenca, tiene los efectos más acusados. En este sentido, la componente ciclo-tendencia de las series en base a nuestra metodología supone obtener unas líneas más adecuadas a una evolución que pueda ser comparable desde el punto de vista analítico.

Así, considerando la componente ciclo-tendencia de la provincia de Albacete, vemos que se produce un crecimiento muy leve en los últimos años; concretamente, la tasa de variación interanual de 2006 respecto al año anterior ha sido de $-0,4\%$.

Figura 2
EVOLUCIÓN MENSUAL DE LAS PERNOCTACIONES EN LA PROVINCIA DE CIUDAD REAL (SERIE ORIGINAL Y CICLO-TENDENCIA)



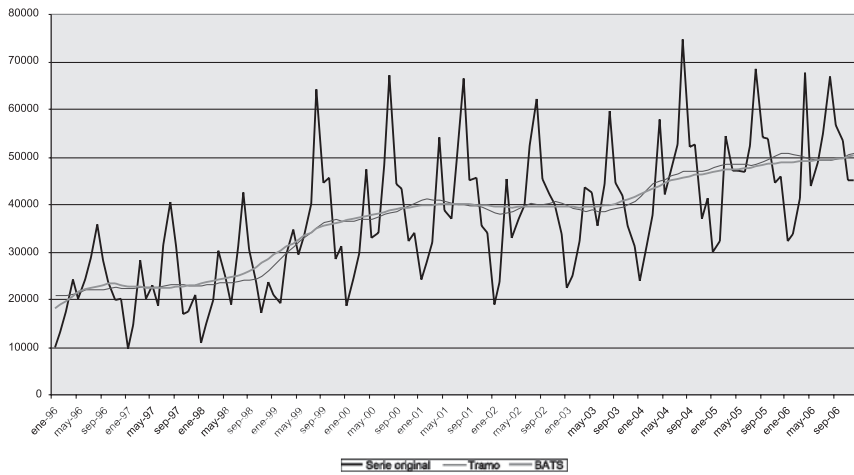
Fuente: Elaboración propia a partir de datos del INE (2007).

La provincia de Ciudad Real ha sufrido el retroceso más importante respecto al número de pernoctaciones en Castilla-La Mancha, con un $-2,4\%$, pero ha mantenido, sin embargo, una evolución constante y positiva en los últimos años.

Como puede apreciarse en la figura 2, el perfil obtenido con la metodología bayesiana es mucho más suave, lo que permite al investigador prever escenarios a más largo plazo.

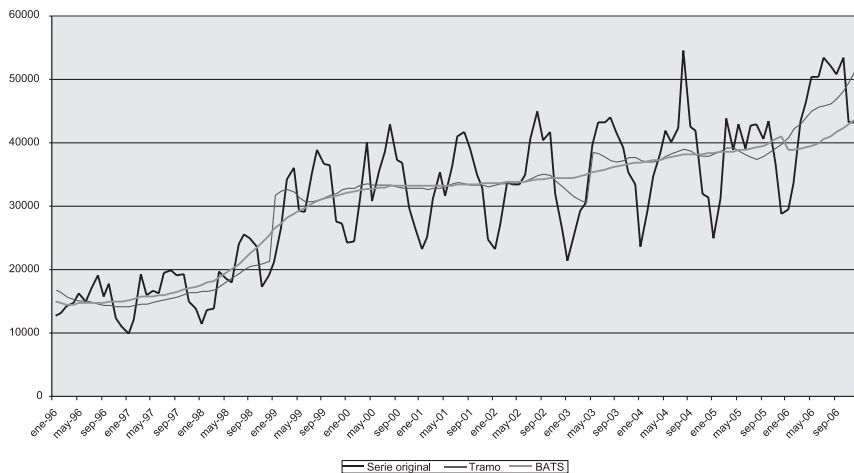
La figura 3 muestra, por su parte, la evolución correspondiente a la provincia de Cuenca. La serie original pone de manifiesto los picos más bajos en los meses invernales, debido a las temperaturas extremas de la provincia en esas épocas. Asimismo, la Semana Santa de Cuenca —declarada de interés turístico internacional— conlleva un grado de ocupación del 100% en la ciudad y prácticamente toda la provincia, año tras año, tal y como muestra la figura anterior.

Figura 3
EVOLUCIÓN MENSUAL DE LAS PERNOCTACIONES EN LA PROVINCIA DE CUENCA (SERIE ORIGINAL Y CICLO-TENDENCIA)



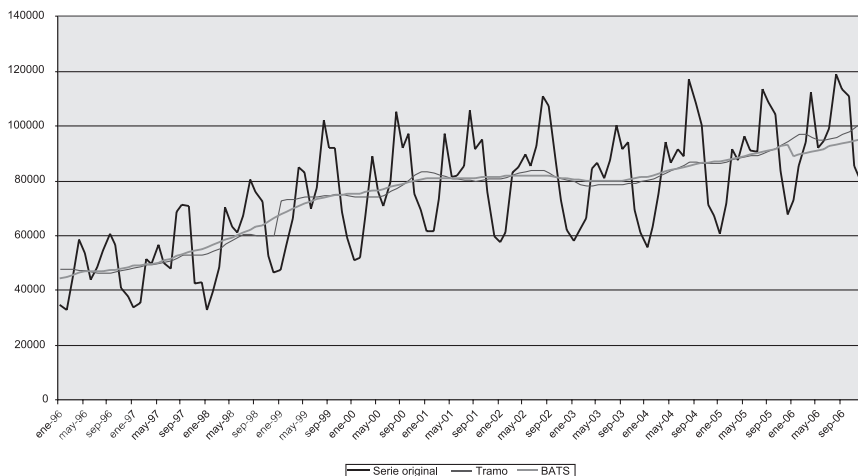
Fuente: Elaboración propia a partir de datos del INE (2007).

Figura 4
EVOLUCIÓN MENSUAL DE LAS PERNOCTACIONES EN LA PROVINCIA DE GUADALAJARA (SERIE ORIGINAL Y CICLO-TENDENCIA)



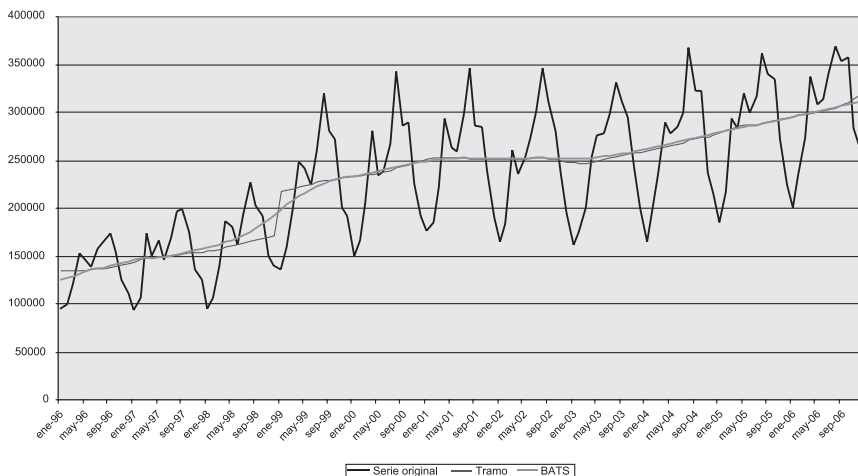
Fuente: Elaboración propia a partir de datos del INE (2007).

Figura 5
EVOLUCIÓN MENSUAL DE LAS PERNOCTACIONES EN LA PROVINCIA DE TOLEDO (SERIE ORIGINAL Y CICLO-TENDENCIA)



Fuente: Elaboración propia a partir de Datos INE (2007).

Figura 6
EVOLUCIÓN MENSUAL DE LAS PERNOCTACIONES EN CASTILLA-LA MANCHA (SERIE ORIGINAL Y CICLO-TENDENCIA)



Fuente: Elaboración propia a partir de datos del INE (2007).

Guadalajara presenta el perfil más alcista de toda Castilla-La Mancha, con unos excelentes resultados para el último año 2006. Concretamente, ha obtenido una tasa de variación anual tres veces superior a la media nacional (18,6%), lo que confirma sus buenas perspectivas de futuro.

Asimismo, Toledo presenta una evolución constante al alza (Fig. 5), y se consolida como el primer destino turístico de nuestra región. Observamos, sin embargo, la importante corrección que reciben los datos tras la desestacionalización, evitando las grandes fluctuaciones derivadas de esta componente.

Si, además del análisis anterior, realizamos una estimación de las previsiones de evolución de los datos correspondientes a las cinco provincias y al total agregado regional, el resultado aparece en la tabla 2.

Tabla 2
PREVISIÓN DE PERNOCACIONES POR PROVINCIAS

Pernoc.	AB Tramo	AB BATS	CR Tramo	CR BATS	CU Tramo	CU BATS	GU Tramo	GU BATS	TO Tramo	TO BATS	CLM Tramo	CLM BATS
ene-07	30.862	30.752	47.391	43.749	30.551	33.368	34.672	35.631	74.419	75.699	212.448	219.199
feb-07	39.270	38.621	51.171	48.605	34.497	36.091	40.727	40.403	84.777	85.839	248.809	249.559
mar-07	46.244	45.723	61.021	57.026	42.846	47.523	49.877	49.691	94.860	98.189	290.346	298.152
abr-07	53.984	49.047	76.289	66.513	67.773	61.631	55.298	50.623	118.816	110.506	365.137	338.320
may-07	67.907	64.415	71.639	64.546	47.562	48.078	55.966	54.451	105.954	102.060	341.196	333.549
jun-07	68.556	63.476	67.951	62.602	51.105	50.480	55.598	53.865	104.294	101.859	341.222	332.281
jul-07	68.529	63.830	80.153	73.148	57.543	57.037	60.373	56.902	108.384	104.547	366.278	355.465
ago-07	72.475	66.602	82.282	73.118	74.041	71.714	63.442	58.265	133.036	125.746	408.401	395.445
sep-07	76.352	69.658	77.025	69.867	60.107	57.787	58.078	55.051	124.183	119.688	386.529	372.052
oct-07	71.092	66.269	79.200	74.917	55.679	56.059	59.604	56.693	121.701	115.823	383.389	369.760
nov-07	53.867	50.414	69.537	64.378	46.249	46.758	48.550	48.212	92.843	91.718	305.973	301.479
dic-07	43.382	41.515	55.484	53.078	47.400	47.221	44.022	45.589	82.403	83.589	273.796	270.992
ene-08	33.150	32.125	51.536	46.868	32.017	35.340	37.524	40.566	79.317	81.067	227.680	235.966
feb-08	42.089	39.994	55.569	51.724	35.988	38.063	44.115	45.337	90.526	91.207	266.466	266.325
mar-08	55.936	47.096	76.936	60.145	63.921	49.495	60.039	54.625	115.664	103.557	367.200	314.918
abr-08	51.336	50.419	71.555	69.633	49.894	63.603	53.808	55.558	110.694	115.874	331.601	355.086
may-08	73.167	65.788	78.282	67.665	50.647	50.050	60.389	59.386	112.192	107.427	366.542	350.316
jun-08	73.435	64.848	73.894	65.722	53.559	52.452	60.171	58.799	111.158	107.227	365.688	349.048
jul-08	72.978	65.203	86.743	76.268	59.352	59.009	65.534	61.837	116.275	109.915	391.596	372.231
ago-08	78.409	67.975	90.347	76.237	80.109	73.686	68.252	63.200	139.950	131.113	439.798	412.211
sep-08	80.808	71.031	82.956	72.986	61.015	59.760	63.232	59.986	134.098	125.056	412.252	388.819
oct-08	76.032	67.641	86.127	78.037	58.352	58.031	64.508	61.628	129.711	121.191	410.878	386.527
nov-08	58.216	51.787	76.353	67.497	50.039	48.730	52.231	53.146	97.669	97.085	329.496	318.245
dic-08	45.869	42.888	59.756	56.197	48.116	49.194	47.929	50.523	88.982	88.957	292.016	287.759

Fuente: Elaboración propia.

Según los modelos univariantes de previsión de series temporales utilizados en nuestro análisis, se prevé un incremento en la tasa de variación interanual del 7% y 6%, respectivamente, para los años 2007 y 2008. Estos modelos de series temporales ponen de manifiesto la rápida incorporación de las posibles fluctuaciones periódicas recogidas durante el periodo previsto.

Por último, es destacable la capacidad de los modelos para incorporar, de forma coherente y natural, toda aquella información —no necesariamente cuantitativa— que puede afectar a la evolución de la serie, todo ello con un coste operacional muy bajo.

5. CONCLUSIONES, LIMITACIONES Y OPORTUNIDADES DE INVESTIGACIÓN

El filtrado de los datos proporciona al investigador unas series más adecuadas para el análisis y estudio del sector turístico debido a la fuerte estacionalidad del mismo. El uso de un determinado modelo de filtrado condiciona los resultados obtenidos: mientras los modelos tradicionales (metodología ARIMA) conforman unos mejores resultados a corto plazo, los modelos bayesianos son más recomendables para el estudio de la evolución a largo plazo.

Este filtrado de series supone el punto inicial de un estudio más ambicioso donde se realizará un análisis de cointegración de las variables estudiadas, para establecer las relaciones de cointegración que pueden presentar las cinco provincias de la comunidad autónoma de Castilla-La Mancha y su comportamiento a nivel agregado para el conjunto de la región.

Ambos procedimientos de filtrado representan una herramienta útil que permite ayudar en la toma de decisiones, la metodología ARIMA, representada por el software *TRAMO-SEATS*, tiene una aplicación más inmediata cuando se trata de realizar predicciones en el corto plazo, por su parte la metodología bayesiana (*BATS*), presenta mejores resultados comparativos en el medio y largo plazo, en las variables estudiadas en nuestro trabajo empírico sujetas a una fuerte estacionalidad.

La utilización de métodos más complejos de predicción (modelos econométricos, series temporales múltiples, redes neuronales, etc.) suponen unas estimaciones más precisas, pero también un coste tanto a nivel informativo (mayor número de variables), como a nivel operacional.

Desde el punto de vista de las limitaciones de nuestro trabajo, debemos destacar el hecho de haber considerado exclusivamente datos de una Comunidad Autónoma, lo que deriva inmediatamente la primera línea de investigación futura en que trabajamos actualmente: ampliar el estudio al resto de comunidades autónomas, de tal forma que podamos efectuar una comparación a nivel nacional.

A este mismo nivel se refiere el hecho de trabajar exclusivamente con un reducido número de variables de las que forman parte de la EOH. Es posible ampliar el análisis a nuevos ámbitos de estudio que permiten obtener una mayor riqueza informativa desde el punto de vista de la investigación turística.

6. BIBLIOGRAFÍA

- BAXTER, M. y KING, M. (1995): «Measuring business cycles approximate band-pass filters for economic time series». *NBER Working Paper n. 5022*, Cambridge (Massachusetts).
- CANOVA, F. (1998): «Detrending and business cycle facts». *Journal of Monetary Economics*, vol. 41, pp. 475-512.
- CONSEJERÍA DE INDUSTRIA Y TECNOLOGÍA (2006): *Plan de Ordenación y Promoción del Turismo en Castilla-La Mancha*. Dirección General de Turismo y Artesanía, Junta de Comunidades de Castilla-La Mancha, Toledo.
- DANIEL, A. y RAMOS, F. (2002): «Modelling Inbound International Tourism Demand to Portugal», *International Journal of Tourism Research*, vol. 4, pp. 193-209.
- DURÁN, J. y FLORES, B. (1998): «Forecasting Practices in Mexican Companies». *Interfaces*, vol. 28, nº 6, pp. 56-62.
- ESPASA, A. y CANCELO, J. R. (1993): *Métodos cuantitativos para el análisis de la coyuntura económica*. Alianza Economía, Madrid.
- ESTEBAN, A. (1996): «Previsiones de demanda turística», *Boletín Económico de Información Comercial Española*, nº 749, pp. 89-97.
- ESTEBAN, A. (2000): «La investigación turística en la Universidad Española», *Estudios Turísticos*, nº 144-145, pp. 155-180.
- GALLOSO, A. (2004): «El análisis de coyuntura regional en España y la evolución comparada de la coyuntura económica de las distintas comunidades autónomas: estado de la cuestión, deficiencias y lagunas», *CLM. Economía*, vol. 4, pp. 299-320.
- GÓMEZ, V. y MARAVALL, A. (1997): «Programs TRAMO (Times Series Regression with ARIMA Noise, Missing Observations, and Outliers) and SEATS (Signal Extraction in ARIMA Time Series). Instructions for the User.
- GÓMEZ, V. y MARAVALL, A. (1998): «Seasonal adjustment and signal extraction in economic time series». *Documento de trabajo 9809*. Servicio de Estudios del Banco de España. Madrid.
- GÓMEZ, M. A., MONDÉJAR, J.A. y SEVILLA, C. (coords.) (2005): *Gestión del turismo cultural y de ciudad*. Ediciones de la Universidad de Castilla-La Mancha, Cuenca.
- HERNÁNDEZ, E. y GIL, N. (1998): «Estadísticas de demanda turística en la región de Murcia». *Cuadernos de Turismo*, nº 1, pp. 55-71
- HODRICK, R. y PRESCOTT E. C. (1980): «Post-ward U.S. Business cycle: an empirical investigation», mimeo, Carnegie-Mellon University, Pittsburgh P.A.
- INSTITUTO NACIONAL DE ESTADÍSTICA (2007): Encuesta de Ocupación Hotelera (Datos 2006 y anteriores). Disponible en www.ine.es.
- JEFFREYS, H. (1961): *Theory of Probability*. Oxford University Press, London.
- JIMÉNEZ, J.F.; GÁZQUEZ, J.C. y SÁNCHEZ, R. (2006): «La capacidad predictiva en los modelos Box-Jenkins y Holt-Winters: una aplicación al sector turístico». *Revista Europea de Dirección y Economía de la Empresa*, vol. 15, nº 3, pp. 185-198.
- LEDESMA, F y NAVARRO, M. (2000): *Datos de panel y demanda de turismo: el caso de Tenerife*, Documento de trabajo, Universidad de La Laguna.

- LIM, C. (1997): «An econometric classification and review of international tourism demand models». *Tourism Economics*, vol. 3, pp. 69-81.
- LISE, W. y TOL, R. (1999): «On the impact of climate on tourist destination choice». *Working paper, W-99/30*. Institute for environmental studies. Amsterdam.
- MELIS, F. (1991): «La estimación del ritmo de variación en series económicas». *Estadística Española*, vol. 33, n.º 126, pp. 7-56.
- MONDÉJAR, J. (2003): Evaluation methods for complex indicators. *56th International Atlantic Economic Conference*. Québec City (Canadá), 16 a 19 de octubre.
- MONDÉJAR, J. (2006): *Análisis cuantitativo de la coyuntura económica. Una aplicación de la representación en espacio de estados de series temporales múltiples*. Tesis Doctoral, Ediciones de la Universidad de Castilla-La Mancha, Cuenca.
- OTERO, J. (1996): «Principales determinantes del flujo de pasajeros extranjeros desembarcados en el aeropuerto de Málaga». *Estudios de Economía Aplicada*, vol. 5, pp. 369-380.
- OTERO, J. (1999): «Las fluctuaciones cíclicas del turismo en Andalucía». *Revista de Estudios de Economía Aplicada*, vol. 13, pp. 105-120.
- PRESCOTT, E. (1986): «Theory ahead of business cycle measurement». *Carnegie-Rochester Conference series on Public Policy*, n.º 25, pp. 11-66.
- SÁNCHEZ, M. y MARÍN, M.B. (2003): «La investigación en turismo y economía de la empresa publicada en revistas especializadas españolas: 1996-2001». *Papers de Turisme*, n.º 33, pp. 6-39.
- SONG, H.; WITT, S. y JENSEN, T. (2003): «Tourism Forecasting: Accuracy of Alternative Econometric Models», *International Journal of Forecasting*, vol. 19, pp. 123-141.
- SORENSEN, N. (2003): «Modelling and Monthly Seasonal Forecasting of Hotel Nights in Denmark». En N. Kaergaard [ed.]: *Symposium for adventt statistik*, pp. 35-50, KVL Press.
- VERA, J.F. e IVARS, J.A. (2001): La formación y la investigación turística en España: una visión de síntesis. *Papers de Turisme*, n.º 29, pp. 6-27.
- ZOU, H. y YANG, Y. (2004): «Combining Time Series Models for Forecasting», *International Journal of Forecasting*, vol. 20, n.º 1, pp. 69-84.