Estudios Turísticos, n.º 146 (2000), pp. 95-104

Instituto de Estudios Turísticos
Secretaría General de Turismo
Secretaría de Estado de Comercio y Turismo

MODELOS DINÁMICOS LINEALES PARA PREDICCIONES TURÍSTICAS

E. Fernández*, J. M. Marín*, I. Olmeda y D. Ríos Insúa***

Resumen: La gestión turística requiere de manera esencial el acceso a sistemas de predicción. El artículo describe una clase de modelos especialmente útiles en este contexto, los modelos dinámicos lineales, implementados en PLAZA, un sistema de gestión del turismo en las Islas Baleares. El proyecto incluye problemas estadísticos complejos, como la predicción del número de visitantes y de las tasas de ocupación hotelera, basados en la existencia de diversas bases de datos.

I. SISTEMAS DE PREDICCIÓN TURÍSTICA

De manera esencial la gestión turística requiere acceso a buenos sistemas de predicción que se integren dentro de sistemas de gestión turística. El ejemplo más claro es el de gestión de la ocupación hotelera, problema en el que debemos evitar situaciones de sobrerreserva *(overbooking),* con los consiguientes costes económicos y pérdida de imagen, y de subreserva, que implican pérdida de recursos por mantener plazas abiertas de forma innecesaria. El disponer de buenos sistemas de predicción permitiría abrir un número más ajustado de plazas hoteleras, una vez que se integre el sistema de predicción dentro de un sistema de ayuda a la decisión, ver French y Ríos Insua (2000). Una revisión de los diferentes enfoques de predicción de la demanda puede verse en Witt y Witt (1995) o Lim (1997).

En este trabajo se describe e ilustra una clase de modelos de predicción, los modelos dinámicos lineales, que hemos empleado con éxito en la predicción de series temporales turísticas, dentro de un proyecto más amplio denominado PLAZA (Olmeda *et al.,* 1999). PLAZA es un sistema de información turístico para las Islas Baleares, que permite, entre otras funciones, hacer predicciones *on line* de la ocupación hotelera para distintos horizontes de tiempo, con posibilidad de considerar diferentes niveles de agregación: las Islas Baleares, cada una de las islas por separado, distintas zonas de cada isla y hoteles de manera individual. Se facilitaría así la gestión hotelera, siendo los principales usuarios del sistema la administración autonómica, los hoteleros, los operadores turísticos e incluso, con un acceso restringido, los propios clientes potenciales que quisieran programar sus viajes. Para ello,

* Universidad Rey Juan Carlos.

** Director del Proyecto. Universidad Rey Juan Carlos.

se integran diversas bases de datos y diversos modelos de predicción bajo un interfaz de acceso vía Internet.

Un rasgo general importante del turismo en las Islas Baleares es la fuerte estacionalidad del mismo. Por otro lado, más del 90% de los visitantes llegan por vía aérea, de manera que el estudio del tráfico y reservas de vuelos es un indicativo bastante fiable del comportamiento del turismo en cada una de las temporadas en estudio. Se presenta también incertidumbre respecto al nivel de ocupación hotelera, lo que origina frecuentes problemas de *overbooking* o exceso de oferta hotelera. En definitiva, la gestión turística en las Islas Baleares se ve sumamente dificultada por problemas de predicción muy complicadas.

En este trabajo, tras realizar una breve introducción a los modelos dinámicos lineales, presentamos dos ilustraciones de los mismos en problemas de predicción de pasajeros en el aeropuerto de Palma y de ocupación hotelera en las Islas.

II. INTRODUCCIÓN A LOS MODELOS DINÁMICOS LINEALES

La modelización matemática de procesos de series temporales se basa en la clase de los *modelos dinámicos*, refiriéndose el término *dinámico* a cambios en tales procesos debidos al paso del tiempo como fuerza motriz fundamental. Esbozamos las ideas principales sobre predicción con modelos dinámicos lineales que nos están resultando de gran utilidad en el análisis de series temporales de carácter turístico, dentro del proyecto PLAZA.

Los modelos lineales clásicos se escriben habitualmente en términos de la ecuación:

$$Y = F'\theta + v,$$

donde Y es el vector respuesta, F' es la matriz de variables regresoras o independientes, θ es un vector de parámetros desconocidos y v es el vector de errores. Si se quiere dar un sentido dinámico o evolutivo al sistema anterior, se han de poder variar los elementos del modelo con el paso del tiempo t y permitir que el conjunto de parámetros θ evolucione a su vez con el tiempo. La evolución toma la forma de un proceso de primer orden de Markov. De este modo, se podría escribir un modelo formado por dos ecuaciones simultáneas:

$$\begin{aligned} Y_t &= F_t'\theta_t + v_t, \\ \theta_t &= G_t\theta_{t-1} + \omega_t, \end{aligned}$$

Introducimos, a continuación, el modelo lineal dinámico MDL normal general para observaciones Y_t unidimensionales. Este modelo vendrá caracterizado por la cuádrupla

$$\{F_t, G_t, V_t, W_t\}$$

donde, para cada instante de tiempo t, F_t es un vector conocido de dimensión $n \times 1$, G_t es una matriz conocida $n \times n$, V_t es una varianza conocida y W_t es una matriz de varianzas conocida $n \times n$.

Definición 1. *El modelo dinámico lineal normal general asociado a la cuádrupla*

$\{F_t, G_t, V_t, W_t\}$ *viene definido por la especificación secuencial de distribuciones*

$$Y_t|\theta_t \sim N(F_t'\theta_t, V_t)$$
$$\theta_t|\theta_{t-1} \sim N(G_t\theta_{t-1}, W_t)$$
$$\theta_0|D_0 \sim N(m_0, C_0)$$

con m_0, vector de medias a priori, y C_0, matriz de covarianzas a priori. Observemos que tal modelo admite la representación alternativa

- *Ecuación de observación*

$$Y_t = F_t'\theta_t + v_t, \quad v_t \sim N(0, V_t)$$

- *Ecuación de sistema*

$$\theta_t = G_t\theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t)$$

- *Información inicial*

$$\theta_0|D_0 \sim N(m_0, C_0)$$

Se supone que las sucesiones de errores de observación y de evolución son independientes, mutuamente independientes e independientes de $\theta_0|D_0$. Los dos resultados siguientes resumen las características básicas del MDL:

Teorema 2. *Para el MDL general univariante las distribuciones a posteriori y las predicciones a un paso, para cada t, son*

- *A posteriori en t – 1*

$$\theta_{t-1}|D_{t-1} \sim N(m_{t-1}, C_{t-1})$$

- *A priori en t*

$$\theta_t|D_{t-1} \sim N(a_t, R_t)$$

- *Predicción a un paso*

$$Y_t|D_{t-1} \sim N(f_t, Q_t)$$

con $a_t = G_t m_{t-1}$, $R_t = G_t C_{t-1} G_t' + W_t$, $f_t = F_t' a_t$, $Q_t = F_t' R_t F_t + V_t$, $e_t = Y_t - f_t$, $A_t = R_t F_t Q_t^{-1}$, $C_t = R_t - A_t A_t' Q_t$ y $m_t = a_t + A_t e_t$.

Este modelo permite, además, hacer predicciones a un horizonte de k unidades de tiempo:

Teorema 3. *Para el MDL general univariante, para cada t y $k \geqslant 1$, las distribuciones para θ_{t+k} e Y_{t+k} dado D_t vienen dadas por:*

- *Distribución del estado*

$$(\theta_{t+k}|D_t) \sim N(a_t(k), R_t(k))$$

- *Distribución predictiva*

$$(Y_{t+k}|D_t) \sim N\big(f_t(k), Q_t(k)\big)$$

con momentos definidos recursivamente por $f_t(k) = F_t' a_t(k)$, $Q_t(k) = F_t' R_t(k) F_t + V_{t+k}$, siendo $a_t(k) = G_{t+k} a_t(k-1)$ y $R_t(k) = G_{t+k} R_t(k-1) G_{t+k}' + W_{t+k}$ y los valores iniciales $a_t(0) = m_t$ y $r_t(0) = C_t$. Además, para cualesquiera enteros j y k con $1 \leqslant j < k$,

$$C(\theta_{t+k}, \theta_{t+j}|D_t) = C_t(k, j)$$

y

$$C(Y_{t+k}, Y_{t+j}|D_t) = F_{t+k}' C_t(k, j) F_{t+j}$$

con $C_t(r, j) = G_{t+r}C_t(r-1, j)$, $r = j + 1, \ldots, k$ *con valor inicial* $C_t(j, j) = R_t(j)$, $\forall t$.

Tales resultados permiten predecir, a uno o más pasos, las observaciones futuras de la(s) variable(s) de interés, de forma eficiente en sentido computacional. West y Harrison (1997) proporcionan abundante información sobre la modelización de los elementos participantes. Desde el punto de vista de las aplicaciones turísticas, el principio más relevante es el de superposición de MDLs, que es un nuevo MDL. A partir de este principio, y teniendo en cuenta la descomposición básica en tendencia, parte estacional, término regresivo (dinámico) y término autorregresivo pueden especificarse numerosos modelos que resultan de enorme utilidad en las aplicaciones turísticas, como ilustramos a continuación.

III. DOS APLICACIONES DE PREDICCIÓN TURÍSTICA

En lo que sigue, consideramos un par de ejemplos de aplicación de los modelos dinámicos lineales en el ámbito turístico. Ambos problemas aparecen en el desarrollo de PLAZA.

III.1. Predicción del número de pasajeros

Como hemos indicado, dado que la gran mayoría de visitantes llega a las Islas Baleares a través del Aeropuerto de Palma, es importante disponer de un sistema de predicción para tal variable. Específicamente, el problema que consideramos es predecir las llegadas al aeropuerto internacional de Palma, con una semana de antelación. Para ello, disponemos del número de pasajeros que llegan cada día y el número de reservas efectuadas con una semana de antelación.

La serie y_t de pasajeros puede modelizarse mediante un *modelo dinámico lineal* con los siguientes componentes:

- Nivel: m_t, que describe un nivel subyacente de visitantes.
- Término de regresión (para las reservas): $\hat{x}$ que intenta explicar, parcialmente, el número de visitantes, basados en el número de reservas.
- Término semanal: ϕ_t^i, que describe la variación debida al día de la semana.
- Término mensual: Z_t^j, que describe la variación debida al mes del año.
- Término autorregresivo: y_{t-1}, que intenta explicar la variabilidad no explicada por los anteriores términos estructurales.

Estos términos se relacionan según la siguiente ecuación:

$$y_t = m_t + \alpha_t + \hat{x}_t + \phi_t^i + \theta_t Z_t^j + \beta_t y_{t-1} + v_t,$$

donde $j = 1, \ldots, 12$ e $i = 1, \ldots, 7$ corresponden a los meses y a los días respectivamente; $\hat{x}_t$ son las series de pasajeros predichos por **Aena,** según las reservas. Los errores v_t se distribuyen según una normal y son independientes entre sí.

Los parámetros del modelo evolucionan según el siguiente sistema de ecuaciones:

$$\begin{cases} m_t = m_{t-1} + \omega_{t,1} \\ \alpha_t = \alpha_{t-1} + \omega_{t,2} \\ \phi_t^1 = \phi_{t-1}^2 + \omega_{t,3} \\ \phi_t^2 = \phi_{t-1}^3 + \omega_{t,4} \\ \phi_t^3 = \phi_{t-1}^4 + \omega_{t,5} \\ \phi_t^4 = \phi_{t-1}^5 + \omega_{t,6} \\ \phi_t^5 = \phi_{t-1}^6 + \omega_{t,7} \\ \phi_t^6 = \phi_{t-1}^7 + \omega_{t,8} \\ \theta_t^7 = \theta_{t-1}^8 + \omega_{t,9} \\ \theta_t = \theta_{t-1} + \omega_{t,10} \\ \beta_t = \beta_{t-1} + \omega_{t,11} \end{cases}$$

donde los errores $\omega_{t,i}$, para $i = 1,\ldots, 11$, son normales e independientes.

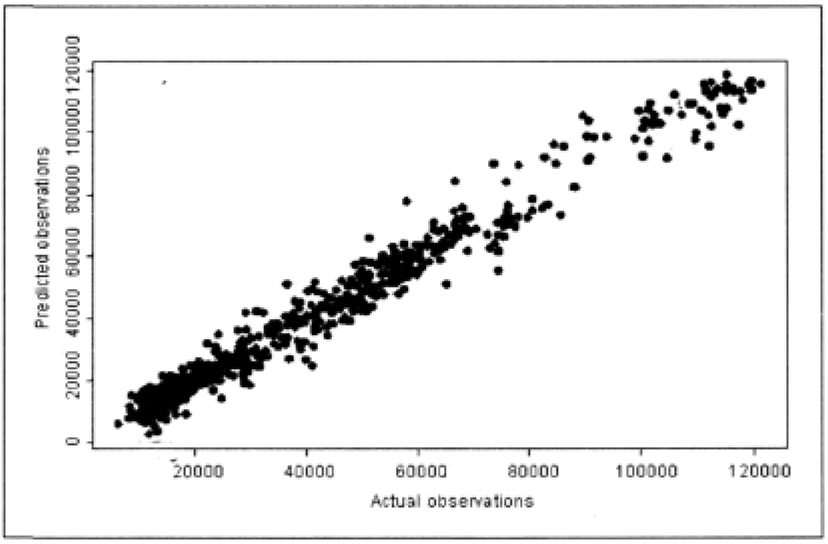
Claramente, el modelo previo puede describirse como un modelo dinámico lineal,

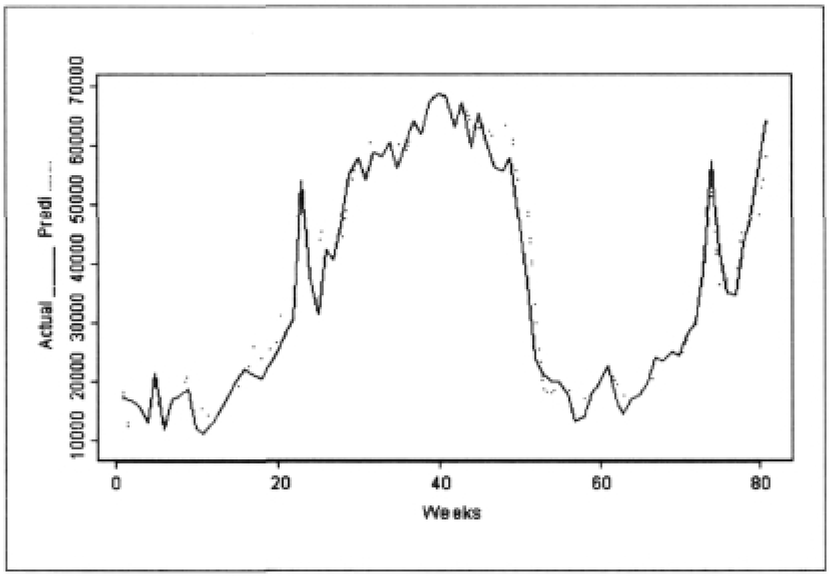
$$\begin{aligned} y_t &= F_t'\theta_t + v_t \\ \theta_t &= G_t\theta_{t-1} + W_t, \end{aligned}$$

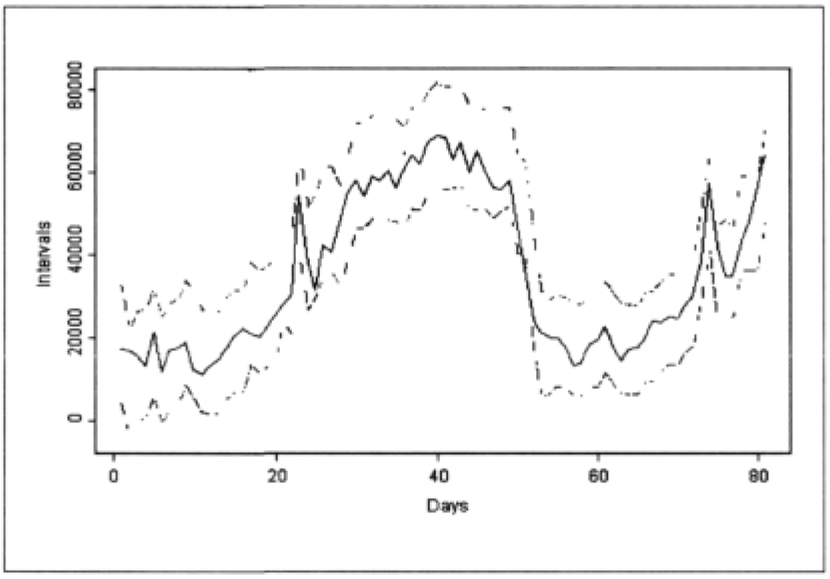
donde se supone normalidad de los errores con varianza desconocida. Los parámetros de la distribución a priori se determinan a partir de los datos del año 1996, estimando el vector de medias y la matriz de covarianzas del vector de parámetros θ sólo con las observaciones de ese año. Posteriormente las predicciones del modelo se realizan a partir del año 1997.

Con ayuda de los resultados de la Sección 2, podemos predecir el número de pasajeros que van a llegar al aeropuerto de Palma con una semana de antelación. Tales predicciones, se evalúan en varios sentidos, como describimos a continuación. Así, si representamos los errores de predicción en los años 1998 y 1999 se observa que no presentan tendencias y que no son, en general, demasiado grandes, lo que sugiere que el modelo es adecuado.

Las siguientes gráficas comparan el número real de pasajeros frente a su predicción, una semana antes, con ayuda del modelo dinámico lineal. Las dos primeras proporcionan observaciones vs predicciones, la segunda preservando el sentido temporal.







Por último, consideramos intervalos de probabilidad predictiva 0,95 (de máxima densidad predictiva).

Todas ellas sugieren un modelo predictivo de suficiente calidad para realizar predicciones a una semana vista del número de pasajeros que lleguen al aeropuerto de Palma.

III.2. PREDICCIÓN DE LA OCUPACIÓN HOTELERA MENSUAL

En el caso de la predicción de ocupación hotelera, podrían considerarse diversos niveles de agregación: el conjunto de las islas, por islas, por zonas costeras dentro de cada isla, o por hoteles dentro de cada zona.

Específicamente, los datos disponibles son las tasas de ocupación hotelera mensuales en las Islas Baleares al completo, en cada isla por separado y en diversas zonas costeras de cada una de las islas. En este ejemplo, se producen ausencias de datos, lo que complica algo más el problema de predicción. Se dispone, igualmente, de datos recogidos por el instituto de estadística local (*IBAE*). El problema que consideraremos es el de predicción de la ocupación en las cuatro islas.

Para ello, empleamos un modelo dinámico lineal con dos términos autorregresivos correspondientes a los datos del mes previo y a los datos de un año antes. El modelo aplicado se describe como:

$$\begin{aligned}
y_t^1 &= \theta_{11}^t y_{t-1}^1 + \theta_{12}^t y_{t-12}^1 + v_{t1}\\
y_t^2 &= \theta_{21}^t y_{t-1}^2 + \theta_{22}^t y_{t-12}^2 + v_{t2}\\
y_t^3 &= \theta_{31}^t y_{t-1}^3 + \theta_{32}^t y_{t-12}^3 + v_{t3}\\
y_t^4 &= \theta_{41}^t y_{t-1}^4 + \theta_{42}^t y_{t-12}^4 + v_{t4},
\end{aligned}$$

donde y_t^1, y_t^2, y_t^3, y_t^4 designan, respectivamente, las ocupaciones el mes t en Mallorca, Menorca, Ibiza y Formentera; cada θ_{ij}^t puede variar aleatoriamente con el tiempo y los errores v_{ti} están distribuidos según una normal independientemente entre sí.

El modelo anterior es un modelo dinámico lineal multivariante:

$$\begin{pmatrix} y_t^1 \\ y_t^2 \\ y_t^3 \\ y_t^4 \end{pmatrix} = \begin{pmatrix} y_{t-1}^1 & y_{t-12}^1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & y_{t-1}^2 & y_{t-12}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & y_{t-1}^3 & y_{t-12}^3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & y_{t-1}^4 & y_{t-12}^4 \end{pmatrix} * \begin{pmatrix} \theta_{11}^t \\ \theta_{12}^t \\ \theta_{21}^t \\ \theta_{22}^t \\ \theta_{31}^t \\ \theta_{32}^t \\ \theta_{41}^t \\ \theta_{42}^t \end{pmatrix} + v_t$$

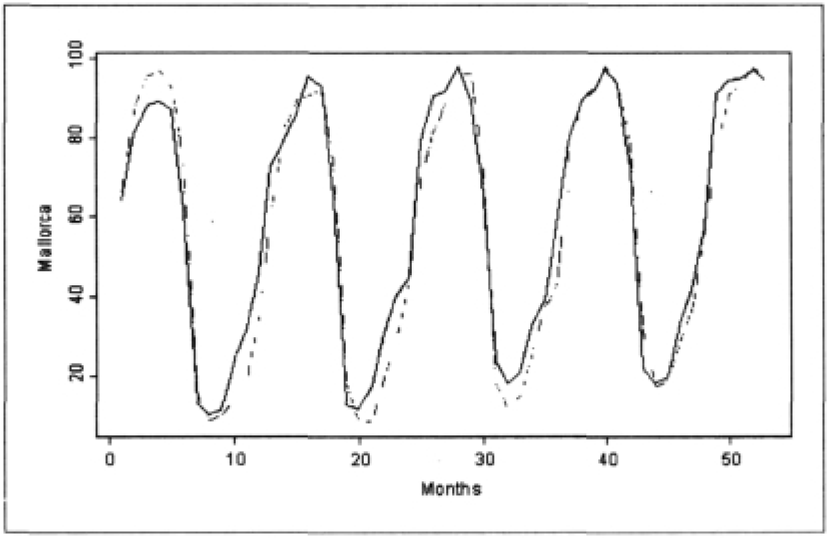
$$\begin{pmatrix} \theta_{11}^t \\ \theta_{12}^t \\ \theta_{21}^t \\ \theta_{22}^t \\ \theta_{31}^t \\ \theta_{32}^t \\ \theta_{41}^t \\ \theta_{42}^t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} \theta_{11}^{t-1} \\ \theta_{12}^{t-1} \\ \theta_{21}^{t-1} \\ \theta_{22}^{t-1} \\ \theta_{31}^{t-1} \\ \theta_{32}^{t-1} \\ \theta_{41}^{t-1} \\ \theta_{42}^{t-1} \end{pmatrix} + W_t$$

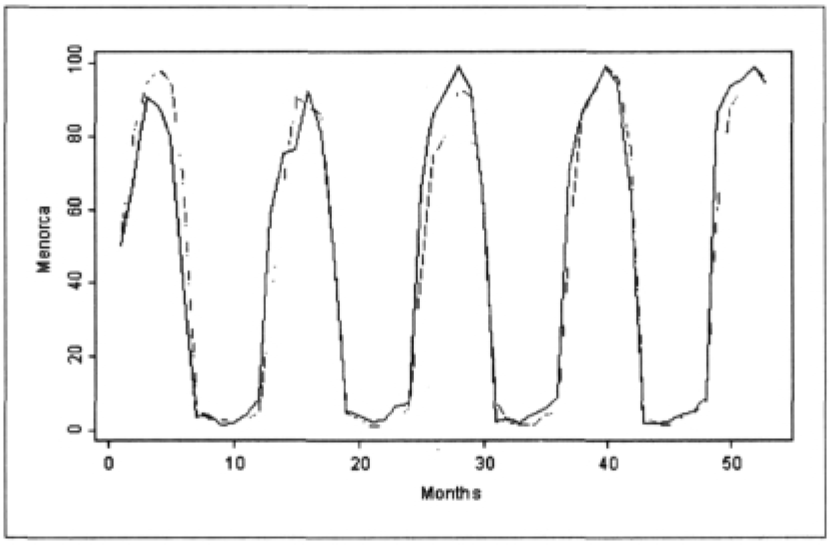
donde v_t y W_t se distribuyen según una normal.

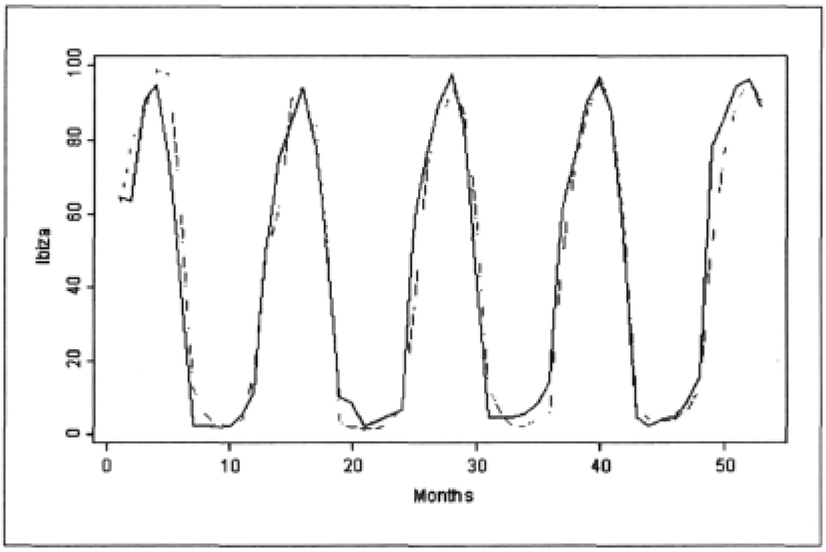
Como los datos se refieren a proporciones, se aplica la transformación $\ln[p/(1-p)]$, donde p es cada uno de los valores observados. Los parámetros de la distribución a priori se determinan de manera heurística dando más relevancia a la parte estacional del modelo, y para estimar los datos ausentes se emplean los valores para el total de las Islas Baleares predichos por el *IBAE*, ponderándolos de manera diferente según el caso de cada isla.

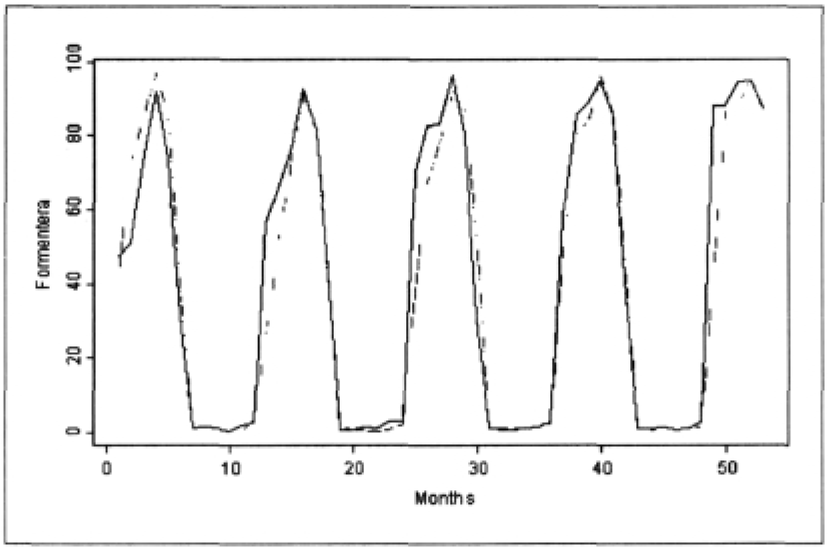
En cualquier caso, dada la escasez de datos y la presencia de datos faltantes, la información previa a priori que se puede introducir en el modelo no es demasiado precisa. Esto provoca que las primeras predicciones no sean muy ajustadas, aunque después el modelo mejora el funcionamiento cuando va incluyendo la información de las observaciones reales.

Las predicciones se hacen, en este modelo, a un mes vista, con ayuda de una extensión inmediata de los resultados de predicción al caso multivariante. Las siguientes gráficas muestran para cada isla las tasas de ocupación reales (continuas) frente a las predicciones (puntuales) procedentes del modelo.









De nuevo, la evaluación del modelo sugiere buena calidad en sus predicciones, salvo en algunas observaciones relativas a mayo, que viene a ser un mes indicativo de lo buena que va a ser la temporada turística. Su modelización no resulta sencilla. Podrían buscarse otros términos de regresión, o intervenciones.

IV. CONCLUSIONES Y LÍNEAS FUTURAS DE DESARROLLO

El sistema PLAZA es un sistema para la predicción turística que actualmente está en desarrollo y en la que aparecen diversos problemas complejos en cuanto a predicción e incluso teoría de la decisión.

Hemos destacado, en este trabajo, la predicción por medio de los modelos lineales dinámicos, aunque se están estudiando modelos alternativos, como son las redes neuronales y que también son utilizadas en dicho proyecto.

Agradecimientos

Trabajo parcialmente apoyado por un proyecto SUP.COM del Eurostat.

BIBLIOGRAFÍA

FRENCH, S., y RIOS-INSUA, D. (2000): *Statistical Decision Theory,* Arnold.

Harrison, P. J., y Stevens, C. F. (1976): «Bayesian Forecasting (with discussion)», *J. R. Statist. Soc. B.,* 38, 205-247.

LIM, C. (1997): «Review of international tourism demand models», *Annals of Tourism Research,* 24, 835-849.

OLMEDA, I.; FERNANDEZ, E., y DE MIGUEL, M. M. (1999): «Sistemas de predicción para la demanda de plazas hoteleras: el proyecto PLAZA», *Estudios Turísticos,* 142, 85-96.

POLE, A.; WEST, M., y HARRISON, J. (1994): *Applied Bayesian Forecasting and Time Series Analysis,* Chapman and Hall, Nueva York.

WEST, M., y HARRISON, J. (1997): *Bayesian Forecasting and Dynamic Models,* Springer, Nueva York.

WITT, S. F., y WITT, C. A. (1995): «Forecasting Tourism demand: A review of empirical research», *International Journal of Forecasting,* 11, 447-475.