A LEADING INDICATOR FOR THE FOREIGN TOURISM DEMAND IN PORTUGAL

Luis Delfim Santos*

and Margarida Macedo**

Paper presented to the *Fourth International Forum on Tourism Statistics*

Copenhagen, 17-19 june 1998

* Faculdade de Economia do Porto and CEMPRE
** Universidade Católica Portuguesa (Centro Regional do Porto)

Introduction

Tourism is one of the main economic activities at world level. In 1996, receipts attributed to tourism were of about 423 billions of US dollars and the number of tourists was of 592 millions. This importance will tend to be reinforced in years to come: the forecasts of the World Tourism Organisation point to an annual growth rate of receipts of 6 to 7% (reaching 2 000 billions of US dollars by 2020), the double of the growth rate predicted for the whole world economy.

As of 1996, Portugal occupied the 16th position as a tourist destination, with a share of just 1.7% of the world total. But, if the relative weight of Portugal in the global context is reduced, the importance of tourism in the Portuguese economy is unquestionable. The number of tourists was, in 1996, of 9,9 millions, a value similar to that of the resident population. In terms of Balance of Payments, receipts attributed to tourism were, also in that year, of about 805 billions of Portuguese escudos and the payments of about 380 billions, leaving a net balance of 425 billions of Portuguese escudos. The weight of tourist gross value added in Portuguese GDP is, according to OCDE estimates, of 8%, one of the highest values among European countries.

Forecasting tourism activity, and tourism demand, in particular, is, in this context, especially relevant. A considerable number of studies has been published on the field, with the use of the most varied forecasting methods. Martin and Witt (1989) compared the accuracy of several quantitative methods in forecasting tourism demand in an international context; Johnson and Ashworth (1990) made a survey of articles published after 1980 on the determinants of tourism demand; Witt and Witt (1995) presented a review of 65 empirical studies about tourism demand and, more recently, Sinclair and Stabler (1997) made a synthesis of several tourism analyses, discussing, particularly, the use of different kinds of models, with explicit consideration of the single-equation and systems of equations approaches, and their respective advantages and disadvantages.

It is the purpose of this study to present a leading indicator for the foreign tourist demand in Portugal, defined by the mean of a transfer function, and to analyse its forecasting ability. Transfer function models make a useful combination between causal and non causal methods; based in ARIMA models (Box-Jenkins models), they allow the use of one or more series related to the one which is being forecasted, thus permitting the explicit consideration of explanatory variables in the model.

Univariate ARIMA models deal with a single time series, forecasted on the basis of its own past values (and a white noise); transfer function models extend this analysis to multiple time series, therefore the forecast of one variable being also affected by past values of the other (explanatory) variables.

Transfer function models

Consider the case of only one explanatory series, X_t, designated as the input series which establishes a causal relationship with the series to be forecasted, Y_t, the output series.

Let X_t and Y_t be two joint stationary processes. The transfer function model is written as follows:

$$\delta(B)\, Y_t = \omega(B)\, X_{t-b} + \eta_t$$

where

$\delta(B) = 1 - \delta_1 B - \ldots - \delta_r B^r \quad \omega(B) = \omega_0 - \omega_1 B - \ldots - \omega_s B^s$

where B is the lag operator ($B^k Y_t \equiv Y_{t-k}$)

and η_t is a stationary and invertibility ARMA(p,q) residual series (“noise”), such that $E(\eta_t) = 0$, $\forall\, t$, $Cov(\eta_s , X_t) = 0$, $\forall\, s,t$

and

$$\phi_t(B)\, \eta_t = \theta_t(B)\, \varepsilon_t$$

where ε_t is a *white noise* term.

Output is described in function of its own past values, of the input and of a residual term that combines the effects of other factors influencing Y_t. This is a dynamic model in the sense that the effect of the input in the output is delayed b periods. That is to say, an input change at moment t produces a response in the output only at moment t+b. Therefore, the input is an advanced indicator (leading indicator) of the output.

The characterisation of the transfer function model is done through 5 parameters (r,s,b) and (p,q), and the modelisation of the function implies the estimation of the coefficients of the polynomials δ (B), ω(B), $\phi_t(B)$, $\theta_t(B)$.

The data

The first question to be dealt with is the choice of the variable to be used as an indicator of tourism demand. In the literature, the variables usually modelised are the number of tourists, the number of nights spent by tourists in hotels or in other types of tourist accommodation and the receipts attributed to tourism.

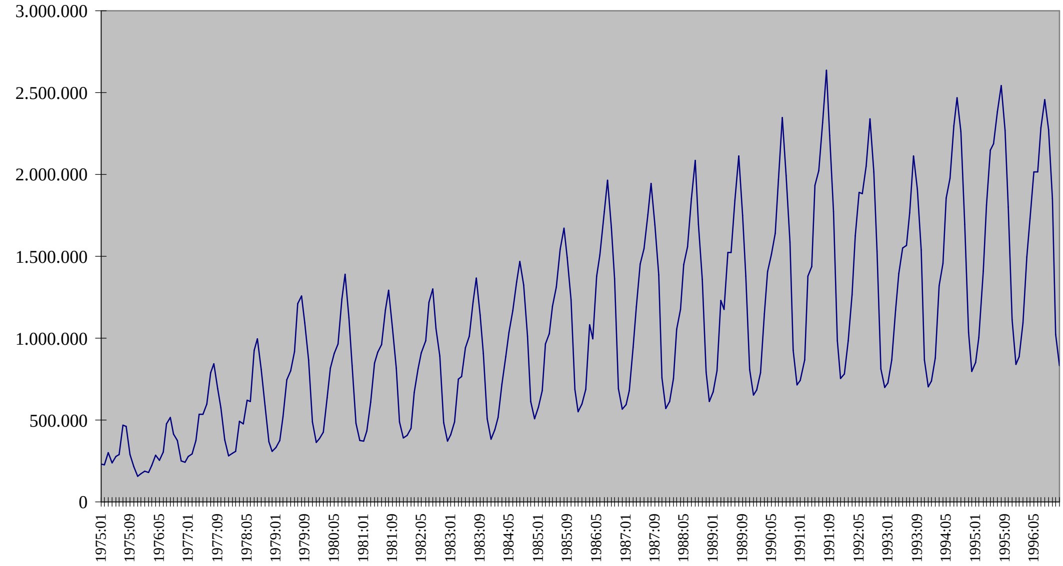
Receipts data on an infra-annual basis being unavailable, we chose to discard the last variable. Consideration of tourism as an economic activity led us to the option for the number of nights spent by tourists as an indicator of demand, given its great economic importance than if we considered only the number of tourist's entrances.

The variable chosen, the "output" in terms of the terminology above mentioned, is then the number of nights spent by foreign tourists in hotels and similar establishments in Portugal. Monthly data for this variable for the period 1975 to 1996 were collected from the *Tourism Statistics* of the "Instituto Nacional de Estatística" (the Portuguese national statistical office).

Although we had observations for those 22 years, the model was estimated only with the data from 1975 to 1994 (240 observations), the values of the last 2 years being kept aside to allow for ex-ante evaluation of the predictive ability of the model.

The following graph displays the main characteristics of the output series:

Nights spent by foreign tourists in hotels and similar establishments in Portugal



It is clearly a nonstationary series, both in the mean as in the variance, as can be detected from the analysis of the autocorrelation function (ACF) and of the partial autocorrelation function (PACF), and by the detection of a unit root with the augmented Dickey-Fuller (ADF) test. The series exhibits a trend and evident signs of seasonality, with peaks in summer months, and low values in winter, characteristics to be expected from the "sun and beach" tourism typical of most Portuguese resorts.

In order to remove nonstationarity of this series, a linearizing transformation of the form Y_t^{λ} [redacted] with $\lambda = 0.5$ [1] and a seasonal differentiation of order 12 were used. In the ADF test to the transformed series the null hypothesis of the existence of a unit root was rejected, what suggests that the transformed series is effectively stationary.

For the input series we employed one series resulting from the simple average of two indicators: one, relative to the real exchange rate (indicator of the relative prices between foreign countries and

---

[1] This linearizing transformation is a sort of the Box-Cox transformation, which is defined by $\frac{Y_t^{\lambda} - 1}{\lambda}$. The package used to estimate the model is AUTOBOX, which uses a simple version of this transformation considering only Y_t^{λ} [redacted] and choosing the optimal value of λ, in this case 0.5 .

Portugal) and another one relative to the economic activity level of the issuing countries of tourism to Portugal (indicator of revenue).

“INPUT” = 0,5 * ITCER + 0,5 * IPI

where $ITCER = \sum_i \alpha_i \, ITCER_i$

ITCER = index (1990=100) of the real exchange rate (TCE)

$$TCE = TC_{(PTE/M_i)} \cdot \frac{IPC_i}{IPC_P}$$

$TC_{(PTE/Mi)}$ = exchange rate (number of Portuguese escudos by one unit of currency of the country i)

IPC = consumer’s price index (1990=100) (i = country ;P = Portugal)

Data were collected for the following countries: United Kingdom, Germany, Spain, the Netherlands and France, that is to say, the 5 main issuing countries of tourism to Portugal. In 1995, the hotel nights of residents of those five countries amounted to 73% of the total. The weights, α_i, used in the aggregation were proportional to the importance of each country in the total of nights spent in hotels (simple average of several years in the period 1975 to 1994).

For the indicator of the economic activity level we used the index of industrial production (IPI), with the same countries and the same weights.

$$IPI = \sum_i \alpha_i \, IPI_i$$

The utilisation of the industrial production to indicate the level of economic activity is related to its availability in monthly data for the 22 considered years and for the countries envolved. That was not the case with other possible indicators, even more global ones, such as GDP.

The source for those variables was the *International Financial Statistics* of the International Monetary Fund.

The estimation of the model

Estimation of the transfer model function was done according to the following stages:

1. Preparation of the input and output time series

The first stage deals with the stationarity of the input and output series. As mentioned above, the output series Y_t was nonstationary, and to become stationary it was necessary a linearizing transformation and a seasonal differentiation of order 12. The new stationary series, denoted by y_t, is defined as:

$$y_t = (1 - B^{12}) \sqrt{Y_t}$$

The input series, X_t, was nonstationary, but the package AUTOBOX did not detect the need for a linearizing transformation. No seasonal pattern was apparent in X_t. Stationarity of the data was achieved by means of a first-order differentiation :

$$x_t = (1-B)\, X_t$$

2. Prewhitening the input series

The input series was prewhitened by means of an ARMA(0,10) model (t ratio in brackets):

$$x_t = \left(1 + 0.19045\, B^{10}\right) \alpha_t$$

[- 2,98]

The evaluation of this model indicates a high statistical quality, particularly the ACF and PACF for the residuals showing no autocorrelation coefficient and partial autocorrelation coefficient significantly different from zero.

3. “Prewhitening” the output series

In this 3rd stage we proceeded to the transformation of the output series applying the same prewhitening transformation used to the input series:

$$Y_t = \left(1 + 0.19045\, B^{10}\right) b_t$$

We obtain a transformed series, β_t , which is not, generally, a white noise term, but having with α_t the some functional relation existing between X_t e Y_t, and so preserving the integrity of the transfer function model.

4. Computing the cross-correlations and the specification of (r,s,b)

The specification of (r,s,b) is made by the mean of the analysis of the cross-correlations between the two series, α_t and β_t . The first significant lag is the lag 12, which corresponds to the value identified for the parameter b. As was pointed out previously, this parameter refers to the lag between an input change and the response; so, the input series (prices and revenue) is assumed to lead the output (hotel nights) with a lag of 12 periods. From an economic point of view, this value means, that a change in the level of the effective real exchange rate or in the resident's revenue in the issuing countries would have effects in the number of nights spent in hotels in Portugal 12 months later, that is to say, one year latter (the following holidays).

Identification of the parameters r and s is a little more complex. We have to analyse the cross-correlations after lag b, and decide on the pattern of these coefficients. The consideration of a fixed pattern would lead to the consideration of a parameter s different from zero (a fixed pattern is expected between lags b and b+s) and a decaying pattern would allow us to consider a parameter r different from zero.

Analysis of the cross correlations, although not in a conclusive way, suggests a fixed pattern between lags 12 and 16 (in which case s=4) and, after lag 16, a decaying pattern (in which case $r \neq 0$). With those guidelines, we have applied the common practice of trying a few different models (r,s,12) and made a choice amongst them at the diagnostic stage. Doing so, and with due allowance to the parsimony principle, we took the final choice of a (0,0,12) specification.

5. The noise series and the specification of (p,q)

The values to the parameters p and q result of the identification of the ARMA model for the noise series:

$$\phi_t(B)\, \eta_t = \theta_t(B)\, \varepsilon_t$$

where ε_t is a *white noise* term.

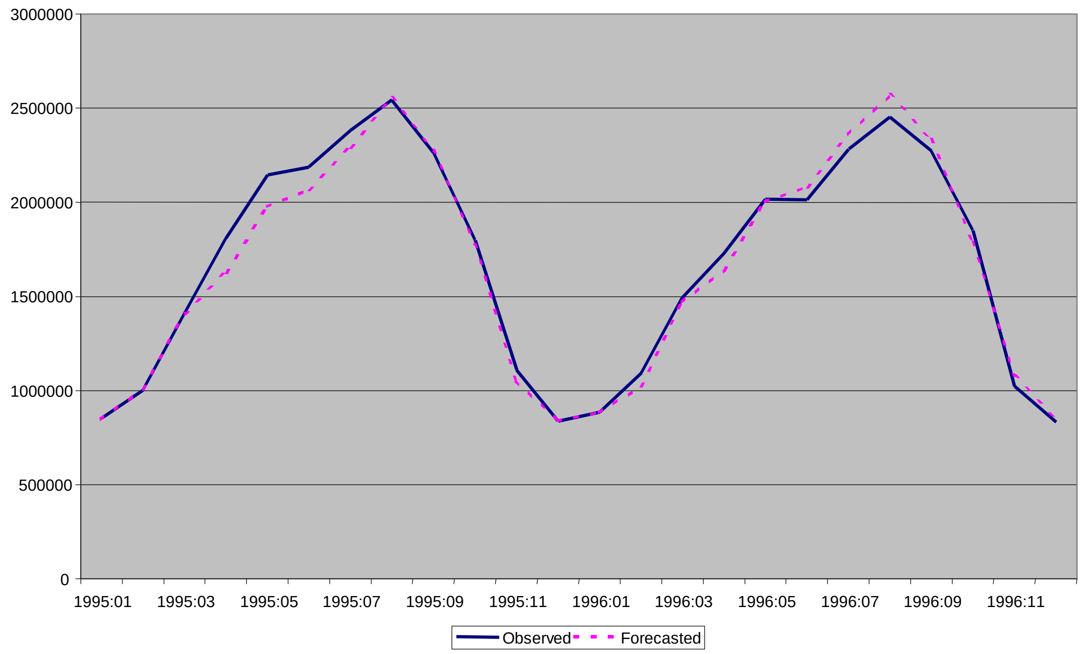
The analysis of the ACF and PACF for the noise series confirms that η_t is not a white noise (a sizeable significantly different from zero number of autocorrelations coefficients and partial autocorrelations coefficients was detected, as well as the existence of seasonal components).
With regard to the non-seasonal component, the ACF displays an exponential decay towards zero, whereas the PACF decays abruptly to zero. This is suggestive of an autoregressive model. The seasonal component reveals the same characteristics. So, the noise was identified by the following model:

$$\underset{[\,9{,}52\,]}{(1 - 0{,}61307\, B} - \underset{[\,4{,}43\,]}{0{,}28914\, B^2})(1 - \underset{[\,4{,}19\,]}{0{,}26595\, B^{11}} + \underset{[\,-5{,}80\,]}{0{,}36476\, B^{12}})\, \eta_t = \varepsilon_t$$

Final identification of the transfer function model is:

Forecasting with the transfer function model

With this estimated model, we forecasted the foreign hotel nights in Portugal for the years 1995 and 1996. The following graph presents the observed and forecasted values:



Forecasts are quite close to the real values; particularly, the turning points caused by the seasonality were satisfactorily taken in account. Observed values are always between the limits of the 95% confidence interval for the output.

In a more formal way, we have, for the 24 forecasted values, a mean absolute percentage error (MAPE) of 3,20% and a root mean squared error (RMSE) of 75,890 (observed values for these two years are in the range 800,000 – 2,500,000). Those two measures of forecasting accuracy show a good behaviour of the model in ex-ante forecasting terms.

References

Auerbach, A. (1982), "The index of leading indicators: Measurement without theory, thirty-five years later", *The Review of Economics and Statistics*, 64, 589-595.

Box, G.E.P. and Jenkins, G.M. (1976), *Times Series Analysis, Forecasting and Control*, 2nd edition, Holden-Day.

De Leeuw, F. (1991), Toward a Theory of Leading Indicators in Kajal Lahiri and Geoffrey Moore (ed.), *Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge University Press.

Diebold, G. and Rudebuch, G. (1989), "Scoring the Leading Indicators", *Journal of Business*, 62,