CRENoS

Centro Ricerche Economiche Nord Sud
Università di Cagliari
Università di Sassari

QUANTITATIVE FORECASTING FOR TOURISM:NEL OLS AND ARIMAX APPROACHES

Manuela Pulina

CONTRIBUTI DI RICERCA
03/03

Manuela Pulina
University of Sassari and CRENoS
manuela.pulina@crenos.it

QUANTITATIVE FORECASTING FOR TOURISM: OLS AND ARIMAX APPROACHES

Abstract
The paper analyses, estimates and forecasts the demand for international and domestic tourism to Sardinia (Italy). Monthly data are used for the sample period from 1987 to 2002. Concepts such as seasonal and long run unit roots are employed. Two econometric approaches, the OLS and ARIMAX, are used that give satisfactory results in terms of both the estimation and forecasting phases. A full range of diagnostic tests is provided. An *ex-ante* forecasting exercise is run for tourism demand to Sardinia for the period between January and December 2003.

Keywords: monthly data, unit roots, OLS, ARIMAX, forecasting.

May 2003

1. Introduction*

This paper has the objective to analyse, estimate and forecast the demand of tourism to Sardinia. As reported in Crenos (2003), tourism activity has a relevant role within the economic system of Sardinia and indeed Italy. In Sardinia, in 2001 tourism expenditure has reached the 2.8% of the Italian total expenditure for tourism in goods and services. 42% of the total tourism expenditure has been spent by Italians, 38% by the locals and a final 20% has been spent by foreigners (Touring Club, 2003). According to Cao and Usai (2002), tourism expenditure in Sardinia represents 7% of the total regional value added.

In estimating and forecasting Sardinian tourism demand, four time series are used. The first series is the number of nights of stay in the hotel accommodation. The second time series is the number of nights of stay in extra-hotel accommodation. The third is the number of nights of stay in all registered accommodation by Italians and, finally, the fourth is the number of nights of stay in all registered accommodation by foreigners. Over the period 1987 and 2002, the nights of stay in hotel count for 67% against 33% in extra-hotel accommodation; moreover, domestic nights of stay count for 79% against 21% of foreigners. The majority of empirical studies focus on the analysis of international demand for tourism. In general, there has been little attention in analysis, estimating and forecasting the domestic demand for tourism in developed countries. Additionally, only a few studies consider the domestic component (see Raeside *et al.*, 1997; Seddighi and Shearing, 1997) or both the components (Malacarni, 1991; Pulina, 2002) and even more rarely the hotel and extra-hotel components.

Song *et al.* (2000) have shown that more sophisticated econometric approaches have given significant results in analysing, modelling

*I would like to thank Maddalena Chessa for providing the data and discussing ideas. For very helpful comments and ideas, I would like to thank Stefano Usai. Any errors are mine.

and testing economic theory. Kulendran and Witt (2001) have also shown that the forecasts obtained by using more recent econometric methodologies are more accurate than those obtained by least squares regression. Can advanced econometric approaches give more insight in modelling and forecasting tourism demand to Sardinia? In this study, the Ordinary Least Squares (OLS) and the "*causal time-series*"ARIMAX approaches are used. These approaches have given the best results both in the estimation and forecasting exercise.

In the majority of the empirical studies, tourism demand is estimated by using either annual or quarterly time series (Witt and Witt, 1992; Seddighi and Shearing, 1997; Song *et al.*, 2000); whereas only few studies make use of monthly time series (Gonzàles and Moral, 1996; Contu, 1997; Crenos, 2002). In O'Brien and Pulina (2002), evidence is found that the monthly and quarterly models present homogenous results in terms of seasonal and long run unit roots. Annual data show different and perhaps misleading results, in particular, when a relative short sample set is used. Furthermore, it emerges that monthly data models are more likely to reflect consumers' decisions to be taken several months in advance or sometimes at the last minute in response to "special offers". On this basis monthly data are used in this study.

The paper is structured as follows. In the next section, the economic model and methodology adopted are presented. In the third section, a seasonal and long run integration status analysis is undertaken for the economic time series under investigation. In the fourth section, four dynamic econometric models are estimated for each of the time series under study. In the fifth section, a forecast is run for each of the econometric model. A summary and conclusive remarks are given in the last section.

2. Methodology.

In this section, the distinct research steps for this empirical investigation are presented.

2.1. The model.

The demand equation for tourism, as used in this paper, is defined in the following manner:

(1) $D = f(DS, DP, T, ID)$

As in function (1), the variables of interest are defined in the following manner:

D = demand for tourism in Sardinia (Source: EPT[1])[2]. Number of Italians and foreigners' nights of stay in hotel accommodation (*LPRAL*); number of Italians and foreigners' nights of stay in extra-hotel accommodation (*LPREX*); number of Italians' nights of stay in all Sardinian registered accommodation (*LPRIT*); number of foreigners' nights of stay in all registered accommodation (*LPRST*).

DS = seasonal dummies. Such variables have been included to evaluate seasonal factors, cyclical holidays (such as Christmas) and seasonal weather conditions effecting the demand for tourism to Sardinia.

DP = "Easter" dummy. This variable is included into the model in order to capture the Easter holiday effect. This effect, in fact, "cannot be captured by the seasonal components due to its mobility so it has to be modelled separately" (Gonzàles and Moral 1996, p.748). As far as the period under modelling is concerned, Easter falls between the 26th March and the 22nd April. The dummy variable "Easter" has, therefore, been constructed giving the value one in the Easter month and zero otherwise. Note also that the Saturday before Easter has been considered as the first day of the holiday, in the case when the Easter period is split into

---

[1] EPT is the Ente Provinciale per il Turismo. Each time series is obtained as the aggregation of the four Sardinian Provinces (Cagliari, Nuoro, Oristano and Sassari).

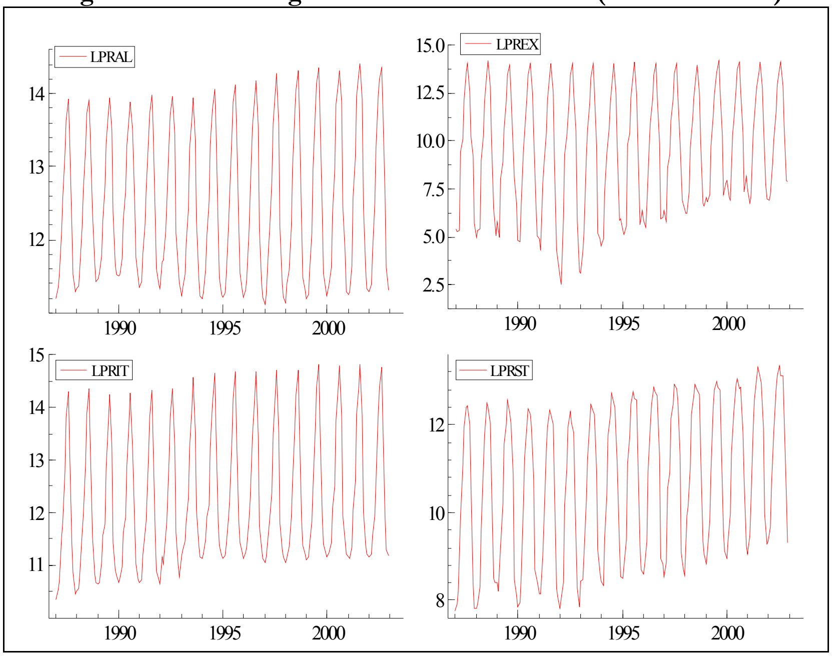
[2] Pulina (2002) has shown that the number of nights of stay is highly correlated with tourism expenditure with a coefficient of 0.83. Moreover, it is argued that the longer the time spent by tourists in a given destination the higher the expenditure in tourism goods and services. Hence, the number of nights of stay can be thought as a valid proxy of tourism expenditure; the latter is not available in the statistical sources.

March and April. This worked better empirically than giving a value 0.5 in each month (see Gonzàles and Moral, 1996).
T = trend. This variable is recognised to pick up possible changes in consumers' tastes for a specific destination over time. In this study, a time trend is included in the final restricted model as having a statistically significant coefficient.
ID = impulse dummies. These qualitative variables are constructed in order to avoid non-normality problems in the residuals. Using monthly data, such dummies are not always easy to interpret. Possible factors for outliers could be related to particular events, such as strikes for boats or planes, or particular discounts for holiday packages in Sardinia. Particular sport events could also be thought to have positively effected the demand for tourism such as rallies, cycle races and so on.
Graphs of each economic series are provided in Figure 1. In each case the natural logarithm[3] of the variables is used. A strong seasonality and a time trend are visible in each of the series. Upon a further inspection, the coefficient of the time trend is found statistically significant at the 1% level of significance by estimating each dependent variable on a time trend and seasonal.

---

[3] In a preliminary investigation, empirical evidence is given in using the logarithmic specification as the correct specification in accordance to the Box and Cox (1974) test. A linear and a logarithmic model have been run. The Box-Cox test suggests that the null hypothesis that two models are empirically equivalent does not hold. Moreover, by comparing the sum of the squared errors (SSE) of the logarithmic and auxiliary regression it emerges that the logarithmic specification form fits the data better than the linear form.

Figure 1: Natural Logarithm of the Time Series (1987:1 - 2002:12)



2.2. Seasonal and long run unit roots tests.

A "pre-modelling" analysis for testing both possible long run and seasonal unit roots is run. The theory suggests that a series can be non-stationary in the level. In particular, a series whose first differences are stationary may be a random walk. To test this, one can use the so-called Augmented Dickey and Fuller (ADF) unit root test. In this paper, Dickey and Fuller's (1981) framework will be used. The ADF test consists in running equation (2):

$$(2)\quad \Delta Y_t = a + \beta T + (\rho - 1)\, Y_{t\text{-}1} + \sum_{i=1}^{p} \lambda_i\, Y_{t\text{-}1} + \eta\, D + \varepsilon t$$

where a constant, the first lag of the series, the lagged difference terms, a time trend (say *T*) and seasonal dummies (say *D*) are included. The augmentation is set to the first statistically significant lag testing downwards. Results for the ADF test are given for each of the following possible combinations: equation (2) with the inclusion of the constant term, the constant and the trend, the constant and seasonals and, finally, the constant, the trend and the seasonals. Given the generic model (2), the ADF test consists by running a *t*-test on the coefficient of the first lag of the dependent variable. Hence, the null hypothesis is $\rho=1$. When failing to reject the null hypothesis, one treats the dependent variable as non-stationary[4].

The seasonal unit roots test is based on Franses (1991a and 1991b), respectively. These tests allow one to study in a systematic manner characteristics and properties of tourism seasonality in Sardinia. Many recent studies have involved the investigation of seasonal variation. This development is due to the realisation that the seasonal components can be the main cause for the variations in many economic time series, and that the seasonal variation in many time series is often irregular. As Hyllerberg points out (see Hargreaves, 1994, pp.153-177), there are many different causes for seasonal variation. As far as tourism is concerned, a change in tourists' preferences or a change in the timing of vacations by institutions and/or employers can cause a shift in the seasonal pattern. The possibility of an irregular seasonal pattern can be tested by means of investigating the possible existence of seasonal unit roots. 'Testing for unit roots in monthly time series is equivalent to testing for the significance of the parameters in the auxiliary regression' (Franses 1991a, p.202) estimated by Ordinary Least Squares (OLS):

(3) $\phi^*(B)y_{8,t} = \pi_1 y_{1,t-1}+\pi_2 y_{2,t-1} + \pi_3 y_{3,t-1} + \pi_4 y_{3,t-2}+$
$+ \pi_5 y_{4,t-1} + \pi_6 y_{4,t-2} + \pi_7 y_{5,t-1} + \pi_8 y_{5,t-2} + \pi_9 y_{6,t-1} +$
$+ \pi_{10} y_{6,t-2} + \pi_{11} y_{7,t-1} + \pi_{12} y_{7,t-2} + \mu_t + \varepsilon_t$

where, μ_t, the deterministic part, consists of a constant, a time trend and 11 seasonal dummies. The null hypothesis of unit roots

4 Results are obtained using PcGive 10.0 modulus (Doornik, 2001).

is tested both by running a t-test of the separate π's, as well as the joint F-test of the pairs, and the π's in the interval $\pi_3 \ldots \pi_{12}$[5]. If the null hypothesis is rejected one can treat the variable of interest as seasonally stationary[6].

2.3. Econometric approaches.

The transformed variables are employed for the estimation of an econometric model. In this empirical study, two main approaches are used. The first is the well-known OLS approach and the "general-to-specific" strategy is adopted. Starting with a very general model, it is possible via a testing down procedure to reach a congruent and encompassing model, via statistical tests[7], information criteria[8] and a set of diagnostic tests[9] (see Mizon, 1996). A second approach is also used to estimate and forecast Sardinian tourism demand, the ARIMAX[10] (AutoRegressive,

---

5 Critical values for the seasonal unit roots test are given in Franses (1991b, pp. 161-165).

6 This test is run in Microfit 4.0 (Pesaran and Pesaran, 1997).

7 In estimating an autoregressive distributed lag model the choice of the lag length is of extreme importance. In choosing the lag length the statistical joint F-test (or Wald test) is adopted. This test allows one to test whether it is statistically significant to reduce the lag length by one.

8 The lag length of a model and the goodness of the fit can also be chosen by making use of information criteria, that is Hannan-Quinn, denoted as HQ criterion; Schwartz, denoted as SC criterion; finally, Akaike, denoted as AIC criterion. The estimated information criteria are chosen so that they are minimised.

9 For the OLS model: DW, Durbin-Watson statistic; AR, autocorrelation test; ARCH, conditional heteroscedasticity; NORM, normality test; HETER, heteroscedasticity test; RESET, functional form test; CHOW, prediction test; WALD, long run coefficients statistical significance test, excluding the constant. For the ARIMAX approach: Chi-Squared test for group of regressors (the null hypothesis is the group of regressors are not statistically significant); the information criteria as defined in Footnote 8; Q test (monitoring and quality assessment statistics are all satisfactory; acceptance region between 0 and 1). See also Doornik X12arima Modulus in GiveWin 2.00 (2001).

10 See Findley *et al.* (1998).

dIfferences, Moving Average and eXogenous variables). This approach consists in several stages: identification of the model; identification of the ARIMA errors of order (p,d,q) (P,D,Q)s[11]; simultaneous estimation of the regression (which also includes exogenous variables on the right hand side of the equation) and the ARIMA errors; finally, backcast (or *ex-post* forecasts) and forecast (or *ex-ante* forecasts) performance and analysis.

3. Seasonal and long run unit roots: results.

In this section, results for long run and seasonal unit roots are reported for each of the time series under study.
Equation (2) is fitted by OLS for each of the four time series mentioned above, for the sample period from 1987:1 to 2002:12. The results are reported in Table 1.

---

[11] (*p,d,q*) for the non-seasonal part, where: *p* the order of the autoregressive part (AR), *d* the number of regular differences used to produce a stationary series (e.g. *d*=1 implies $Y_t = Y_t - Y_{t-1}$); *q* the order of the moving average (MA). (*P,D,Q*)s for the seasonal part, with s=12 as a monthly frequency is used. Fitting an ARMA model requires that the data are stationary.

Table 1 Augmented Dickey-Fuller Unit Root Test (1987:1 - 2002:12) - 192 observations					
Time Series	*ADF*(1)	*LAG*(2)	***Time Series***	*ADF*(1)	*LAG*(2)
LPRAL(c)	- 4,08 **	9	**LPREX**(c)	- 5,35 **	7
LPRAL(c,t)	- 12,97 **	7	LPREX(c,t)	- 4,48 **	9
LPRAL(c,s)	- 5,40 **	6	LPREX(c,s)	- 5,50 **	0
LPRAL(c,s,t)	- 6,25 **	2	LPREX(c,s,t)	- 6,65 **	0
LPRST(c)	- 3,34 **	6	**LPRIT**(c)	- 5,27 **	8
LPRST(c,t)	- 14,21 **	5	LPRIT(c,t)	- 13,61 **	8
LPRST(c,s)	- 3,37 *	0	LPRIT(c,s)	- 3,52 **	0
LPRST(c,s,t)	- 3,94 *	4	LPRIT(c,s,t)	- 6,64 **	0

Notes: * and ** indicate that the unit root null hypothesis is rejected at the 5% and 1% level, respectively. (1) Augmented Dickey-Fuller statistics with constant (*c*) critical values = -2.872 at 5% and -3.455 at 1% level; with constant and trend (*c, t*) c.v.= -3.428 at 5% and -3.995 at 1% level; with constant and seasonals (*i.e. c, s*) c.v. = -2.872 at 5% and -3.456 at 1% level; with constant, trend and seasonals (*i.e. c, t, s*) c.v. = -3.428 at 5% and -3.995 at 1% level. (2) Number of lags set to the first statistically significant lag, testing downwards and upon white residuals. Note that ADF(0) corresponds to the DF test.

The main result is that none of the series denotes a random walk pattern. All the series are stationary in the level and no difference transformation is needed.

Equation (3) is fitted by OLS for each of the four time series mentioned above, for the sample period from 1987:1 to 2002:12. The results are reported in Table 2.

Table 2 Seasonal Unit Root Test (1987:1-2002:12; 192 observations)				
t-statistics	Variable			
	LPRAL	***LPREX***	***LPRST***	***LPRIT***
π1	- 0,642	-1,00	-2.144	-1,804
π2	- 4,950 ***	-4,582***	-4.870 ***	-3,308 *
π3	- 2,729 ***	-1,853 *	-0.851	-1,031
π4	-3,535 **	-3,462 *	-5.120 ***	-5,184 ***
π5	-3,894 ***	-5.172 ***	-4.388 ***	-5,437 ***
π6	-4,903 ***	-5,783 ***	-5.318 ***	-5,822 ***
π7	1,304	1,309	2.125	1.909
π8	-1,226	-2,362	- 3.327 *	-2.200
π9	-2,226	-2,166	- 0.429	-3,463 ***
π10	-4,636 ***	-4,945 ***	- 4.115 ***	-5,485 ***
π11	1,770	0,921	1.101	1,443
π12	-3,851 ***	-3,083 20%	-3.465 **	-4,069 ***
F-statistics	***LPRAL***	***LPREX***	***LPRST***	***LPRIT***
π3, π4	10,000 ***	7,991 **	13.616 ***	14,115 ***
π5, π6	12,364 ***	16,783 ***	14.289 ***	17,250 ***
π7, π8	0,862	4,007	6.913 *	2,421
π9, π10	10,756 ***	12,316 ***	10.786 ***	15,429 ***
π11, π12	7,442 **	5,035	6.309 **	8,523 ***
π3=...= π12	10,130 ***	10,937 ***	12.418 ***	12,048 ***

Notes: ***, ** and * indicate that the seasonal unit root null hypothesis is rejected at the 1%, 5% and 10% level, respectively.

In accordance with the findings in Franses (1991b), it appears that there is no evidence of seasonal unit roots in the international demand for tourism variable (*LPRST*); however, for the other series an exception can be noticed. The null hypothesis cannot be rejected for one pair of π's: π_7, π_8 (*i.e.* at frequency $\pi/3$). In accordance with the findings in Franses (1991b) it appears that there is no strong evidence for the presence of seasonal unit roots. Hence, the four series are treated as having a deterministic seasonal pattern. However, as Webb (1995) notices "….other types of nonstationarity are also possible. An alternative….involves large, infrequent shocks." (p.277). The possibility of the existence of level shifts is investigated further in the next section.

4. Modelling phases and results.

In this section, the demand for tourism to Sardinia is estimated. The pre-modelling phase, so far, allows one to understand the characteristics of the economic variables under study, as well as other specific properties of the seasonal pattern.

The main aim of this study is to identify the most satisfactory models so that to forecasting tourism flows into Sardinia for 2003. A preliminary investigation has led to choosing several models and approaches. Two sets of minima and maxima forecasts are obtained by employing both the OLS and the ARIMAX approaches, together with a normalisation of the dependent variable for the "trading-day" factors, whenever required by the diagnostics and by the goodness of the forecasting. As Baron (1989) points out trading-day factors might be important in the analysis of monthly data; these take into account the effects of four or five Saturdays in a particular month. It is likely that the higher concentration of tourists occurs on the weekends. Moreover, the majority of the charter flights and boat trips to the north of Sardinia occur on a Saturday. Given these assumptions, whenever required, the dependent variable is defined as follows:

ND = *D*/ *N,* where *D* is the total number of tourists' nights of stay and *N* is the number of Saturdays in a month.

An account of the main estimation findings follows for each of the econometric models run for the sample period from 1987:1-2001:12. The last twelve observations are left out for the forecasting *ex-post* exercise. The results are reported in Table 3.

- LPRAL (hotel nights of stay). The maximum value of the *ex-ante* forecast (see Table 4) is obtained by using the following model: ARMAX (0,0,1) (0,0,2)12. This specification is choosen in accordance to the total and partial sample autocorrelation functions. The seasonality of the series is readily apparent in the sample autocorrelation function that shows peaks at k = 6, 18, 30 and troughs at k = 12, 24, 36. In particular, the autocorrelation function exhibits moving average properties that have been considered as first order for the non-seasonal part (q=1, say *MAq1* in Table 3), and of second order for the seasonal part (Q=2, say *MAQ1* and *MAQ2* in Table 3). The final restricted model also includes the following exogenous variables: the seasonal dummies (*DS*), a level shift (*ls92m1*[12]), and a trend (*T*). All these variables are statistically significant at the 1% level. This model has been validated by the minimisation of the information criteria and other satisfactory statistical tests (see Table 3 and Footnote 8). Moreover, the goodness of forecasting is verified by a Mean Squared Error[13] (MSE) equals to 0,011.
- LNPRAL (hotel nights of stay, normalised by the number of Saturdays in a month). The minimum value of the *ex-ante* forecast (see Table 4) is obtained by using the following model:

---

12 A level shift is added in order to pick up ouliers permanently affecting the level of the series after January 1992. The statistically significant parameter (see Table 3) suggests that this series is characterised by a discontinuity or irregularity in line with Franses results (see Table 2).

13 The Mean Squared Error is defined by the following formula: $MSE = 1/n \sum_{t=1}^{n} e_t^2$.

ARMAX (0,0,0) (0,0,1)12. This specification is choosen in accordance to the total and partial sample autocorrelation functions. In particular, the autocorrelation function exhibits moving average properties that have been considered as first order for the seasonal part (Q=1, say *MAQ1* in Table 3). The model also includes the following exogenous variables: the seasonal dummies (*DS*), a level shift (*ls92m12*), a trend (*T*) and the Easter dummy (*DP*). This model has been validated by the minimisation of the information criteria as well by the other statistical tests. The goodness of forecasting is verified by a Mean Square Error (MSE) equals to 0,02.

- LPREX (extra-hotel nights of stay). The maximum value of the *ex-ante* forecast (see Table 4) is obtained by using the OLS approach. The model presents diagnostic problems in the residuals. Nevertheless, this specification is one of the best models obtained given the data available in this study. Note that the impulse dummy (*i89m11*) and the step dummy[14] (*s92m1m2*), though improving the diagnostics, could not eliminate the non-normality issue. The lag coefficients of the extra-hotel demand for tourism, as explanatory variables, present an overall positive sign. This indicates that tourists, who choose extra-hotel resorts, are possibly 'psychocentric' (Sinclair and Stabler, 1997) and that Sardinia is viewed as a desirable destination area. This time series has presented much volatility and a further investigation is needed. The inclusion of further explanatory variables into the equation (such as a weather variable) might reduce or eliminate the problems in the residuals. The MSE in this case is equal to 0.23.
- LNPREX (extra-hotel nights of stay, normalised for the number of Saturdays in a month). The minimum value of the *ex-ante* forecast (see Table 4) is obtained by using the OLS approach. The model presents diagnostic problems in the residuals; heteroschedasticity (at the 5% level), conditional

[14] This variable takes the value one in January and February 1992 and zero otherwise.

heteroschedasticity (at the 1% level) and non-normality (at the 1% level) are detected. The R^2 explains around 96% of the variance of the dependent variable. Moreover, as the relevant *F*-statistic indicates, the overall significance of the regression is satisfactory. The inclusion of either seasonal dummies or dummy variables does not reduce the problems in the residuals. It is worthwhile noting that the specification form is valid in comparison to the specification form without the normalisation of the dependent variable (see Reset test in Table 3 for *LPREX* and *LNPREX*). Again, this is one of the best specification obtainable given the availability of the data used in this study. Once more, this time series has presented much volatility. The inclusion of extra explanatory variables into the equation might reduce or eliminate such problems in the residuals. Further investigation is needed. The goodness of forecasting is given by the Chow F-test (the null hypothesis of parameter costancy is accepted) and by a MSE equals to 0.17.

- LPRIT (Italian nights of stay in all registered accommodation). The maximum value of the *ex-ante* forecast (see Table 4) is obtained by using the following model: ARMAX (0,0,1) (0,0,1)12. This specification is choosen in accordance to the total and partial sample autocorrelation functions. The seasonality of the series is readily apparent in the sample autocorrelation function that shows peaks at k = 6, 18, 30 and troughs at k = 12, 24, 36. In particular, the autocorrelation function exhibits moving average properties that have been considered as first order for the non-seasonal part (q=1, say *MAq1* in Table 3), and of first order for the seasonal part (Q=1, say *MAQ1* in Table 3). The model also includes the following exogenous variables: the seasonal dummies (*DS*), a level shift (*ls92m1*[15]), and a trend (*T*). The Easter dummy is

[15] A level shift is added in order to pick up ouliers permanently affecting the level of the series after January 1992. The statistically significant parameter (see Table 3) suggests that this series is characterised by a discontinuity or irregularity in line with Franses results (see Table 2).

excluded from the final restricted model as the coefficient is not statistically significant. This model has been validated by the minimisation of the information criteria and other satisfactory statistical tests (see Table 3 and Footnote 8). Moreover, the goodness of forecasting is verified by a MSE equals to 0,02.

- LPRIT (Italian nights of stay in all registered accommodation). The minimum value of the *ex-ante* forecast (see Table 4) is obtained by using the OLS approach. The model is well specified and does not present any problems in the diagnostics. The final restricted model includes the first, eleventh and twelveth lag of the dependent variable, 11 seasonal dummies, the Easter dummy, three impulse dummies (*i92m2*, *i93m1* and *i93m10*) which detect outliers causing problems of non-normality in the residuals. Again, the lag coefficients of the domestic demand for tourism, as explanatory variables, present an overall positive sign. This indicates that domestic tourists are possibly 'psychocentric' and that Sardinia is viewed as a desirable destination area. The goodness of forecasting is given by the Chow F-test (the null hypothesis of parameter costancy is accepted) and by a MSE equals to 0.008.
- LPRST (extra-hotel nights of stay). The maximum value of the *ex-ante* forecast (see Table 4) is obtained by using the OLS approach. The first and second lag coefficient of the dependent variable are statistically significant, denoting that foreigners regard Sardinia as a desiderable tourist destination. Eleven seasonal dummies, a time trend, the Easter dummy and two impulse dummies are also included in the final restricted model (*i00m11* and *i98m2*) The model does not present diagnostic problems in the residuals. However, it has to be noted that the null hypothesis of homoscedasticity for the disturbances is marginally rejected at the 5% level. In this case the 'ordinary least-squares parameter estimators are unbiased and consistent, but they are not efficient; *i.e.* the variances of the estimated parameters are not the minimum variances' (Pindyck and Rubinfeld 1991, p.128). A White

correction for heteroscedasticity has been used for the standard errors, and the final results are reported in Table 3. The MSE in this case is equal to 0.23.

- LNPRST (extra-hotel nights of stay, normalised for the number of Saturdays in a month). The minimum value of the *ex-ante* forecast (see Table 4) is obtained by using the OLS approach. Overall, the model is well-specified with the only exception for the mis-specification Reset test. This failure might be due to the exclusion of other important economic variables which might affect the dependent variable. Further research is needed to investigate this issue. The R^2 explains around 98% of the variance of the dependent variable. Moreover, as the relevant *F*-statistic indicates, the overall significance of the regression is satisfactory. The goodness of forecasting is given by the Chow F-test (the null hypothesis of parameter costancy is accepted) and by a MSE equals to 0.12.

So far, the results obtained show some volatility in estimating the international demand for tourism as well as the extra-hotel demand. Notably, a great number of foreigners make use of extra-hotel resorts. The inclusion of other extra variables (such as exchange rate, tourist price index, weather conditions and so on) might have an important role in explaining these components. Further work is needed to assess such possibilities.

Table 3. Estimated models using ARIMAX and OLS approaches

Variable Estimated Model

LPRAL (ARMAX) $LPRAL_t$ = 12.187 - 0.388 LSm1 + 0.139 T- 1.180 Jan - 1.088 Feb
(maximum forecast) (45.18) (-3.49) (6.02) (-30.82) (-28.20)
- 0.810 Mar - 0.245 Apr +1.713 May + 1.030 Jun + 1.425 Jul
(-21.11) (-6.39) (7.42) (26.86) (37.16)
+ 1.650 Aug + 1.141 Sep - 0.207 Oct - 0.90 Nov - 1.142 Dec (derived)
(43.02) (29.75) (-4.59) (-23.17) (-29.81)
- 0.453 MAq1 - 0.3016 MAQ1 - 0.404 MAQ2
(-6.67) (-4.14) (-5.48)

Chi-squared Tests for Groups of Regressors; seasonals Chi^2(11) = 4539.81**; user-defined Chi^2(2) = 54.40**;
Q = 0.72 accepted; MSE = 0.011

LNPRAL (ARMAX) $LPRAL_t$ = 10.682 -0.169LSm12 + 0.003T+ 0.089DP - 1.193 Jan - 0.988 Feb
(minimum forecast) (175.83) (-3.55) (5.50) (1.61) (-22.96) (-19.18)
- 0.819 Mar - 0.301 Apr +0.271 May + 1.048 Jun + 1.411 Jul
(-15.72) (-4.81) (5.21) (20.18) (27.19)
+ 1.641 Aug + 1.142 Sep - 0.201 Oct - 0.859 Nov - 1.143 Dec (derived)
(31.60) (22.00) (-3.88) (-16.53) (-22.00)
- 0.487 MAQ1
(-7.00)

Chi-squared Tests for Groups of Regressors; seasonals Chi^2(11) = 4214.61 **; user-defined Chi^2(2)=32.88 **;
Q = 0.86 conditionally accepted; MSE= 0,02.

LPREX (OLS) $LPREX_t$ = 0.357 + 0.291$LPREX_{t-1}$+ 0.155$LPREX_{t-10}$+ 0.203$LPREX_{t-11}$+0.146$LPREX_{t-12}$
(maximum forecast) (1.04) (7.85) (3.10) (3.52) (2.62)
- 0.143 Jan - 0.911 Feb -0.720 Mar - 0.534 Apr +0.107 May + 0.507 Jun +
(-0.775) (-3.57) (2.31) (1.44) (0.28) (1.20)
+ 1.126Jul + 1.678 Aug + 1.187 Sep + 0.533Oct - 1.476Nov
(2.42) (4.04) (2.92) (1.55) (-5.92)
+ 1.606 DP + 2.167 i89m11 - 1.782 s92m1m2
(7.85) (4.83) (-5.23)

R^2=0.984384 F(18,149) = 521.8** DW = 1.79 AR _F(7,142) = 0.66393 ARCH_ F(7,135) = 1.3196
NORM_Chi^2(2)=28.160 ** HETER_F(22, 126) =2.3323* RESET_F(1,48) = 8.6597** CHOW_F(12,137) = 0.28126;
MSE=0.23.

LNPREX (OLS) $LNPREX_t$=1.533+ 0.436$LNPREX_{t-1}$-0.102$LNPREX_{t-2}$-0.012$LNPREX_{t-3}$+ 0.046$LNPREX_{t-4}$+
(mimimum forecast) (1.48) (4.33) (-1.51) (0.18) (0.70)
- 0.116 $LNPREX_{t-5}$+ 0.068 $LNPREX_{t-6}$ - 0.134 $LNPREX_{t-7}$ + 0.062$LNPREX_{t-8}$
(-1.71) (1.00) (1.96) (0.91)
- 0.038 $LNPREX_{t-9}$ - 0.010 $LNPREX_{t-10}$ + 0.191$LNPREX_{t-11}$+ 0.530 $LNPREX_{t-12}$ +
(-0.59) (-0.16) (2.87) (8.03)
+ 0.002 T
(2.11)

R^2=0.959195 F(13,154) = 278.5** DW = 1.97 AR _F(7,147) = 0.44187 ARCH_ F(7,140) = 2.4887*
NORM_Chi^2(2) = 33.456** HETER_F(26,127)=1.9594 RESET_F(1,153) = 3.8036 CHOW_F(12,131)= 0.47607
MSE=0.17.

LPRIT (ARMAX) $LPRIT_t$ = 11.920 + 0.098LSm1 + 0.003T - 1.278Jan - 1.182Feb
(maximum forecast) (251.55) (2.42) (9.22) (-42.24) (-38.99)
- 0.954 Mar - 0.405Apr + 0.022May + 1.072Jun + 1.859Jul
(-31.49) (-13.38) (0.74) (35.43) (61.42)
+ 2.344Aug + 1.289Sep - 0.465Oct - 1.019Nov - 1.238Dec (derived)
(77.42) (42.59) (-15.35) (-33.64) (-40.89)
- 0.385 MAq1 - 0.193 MAQ1
(-5.43) (-2.54)

Chi-squared Tests for Groups of Regressors; seasonals Chi^2(11) =11544.48**; Q = 0.58 accepted; MSE = 0.02.

Table 3 continuous

LPRIT (OLS) LPRIT$_t$ = 0.295+ 0.563LPRIT$_{t-1}$ + 0.190 LPRIT$_{t-11}$ + 0.212 LPRIT$_{t-12}$ +
(mimimum forecast) (0.91) (10.5) (3.03) (3.08)
+ 0.037 Jan + 0.088 Feb + 0.110 Mar + 0.225 Apr + 0.153 May + 0.648 Jun +
(1.10) (2.21) (2.21) (3.52) (1.62) (5.72)
+ 0.552 Jul + 0.696 Aug - 0.080Sep - 0.779 Oct - 0.161 Nov +
(5.04) (6.38) (-0.55) (-7.44) (-4.06)
+ 0.151DP + 0.368 i92m2 + 0.391 i93m1 + 0.344 i93m10
(4.04) (4.37) (4.52) (4.13)
R^2=0.996496 F(13,154) =2165.0 ** DW = 2.09 AR _F(7,142) = 0.92631 ARCH_ F(7,135) = 0.29221
NORM_Chi^2(2)= 4.9524 HETER_F(21,127)= 1.5127 RESET_F(1,148) = 0.027828 CHOW_F(12,137)= 1.0532
MSE=0.008.

LPRST (OLS) LPREX$_t$ = 2.490+ 0.486LPRST$_{t-1}$+ 0.168LPRST$_{t-11}$+ 0.003 T + 0.496 DP +
(maximum forecast) (3.18) (6.14) (2.72) (3.11) (3.56)
+ 0.316 Jan + 1.023Feb + 1.687 Mar + 2.560 Apr 3.542 May + 3.675Jun +
(2.64) (4.43) (6.32) (8.63) (11.9) (11.8)
+ 3.561 Jul + 3.092 Aug + 2.841 Sep + 1.569 Oct - 0.159 Nov
(12.5) (11.5) (11.4) (6.70) (-1.00)
+ 1.118 i00m11 + 0.917 i98m2
(17.97) (9.87)
R^2=0.984165 F(17,151) = 902** DW = 1.94 AR _F(7,144) = 0.92526 ARCH_ F(7,137) = 1.2198
NORM_Chi^2(2)=5.2366 HETER_F(20,130)=1.6556* F(1,150) =2.0386 CHOW_F(12,139)= 1.2199; MSE=0.08.

LNPRST (OLS) LNPRST$_t$ = 2.179+ 0.284 LNPR$_{t-1}$ + 0.215 LNPRST$_{t-3}$ - 0.004 T+ 0.379 DP +
(mimimum forecast) (2.99) (4.20) (3.23) (5.00) (3.56)
+ 0.076Jan + 0.911 Feb + 0.194 Mar + 0.529 Apr +1.746May + 1.713 Jun +
(1.04) (3.03) (3.91) (3.77) (6.02) (4.98)
+ 1.830Jul + 1.511Aug + 1.642 Sep + 0.844 Oct - 0.632Nov
(5.00) (3.91) (4.77) (2.83) (-2.96)
+ 1.118 i00m11 + 0.917 i98m2
(17.97) (9.87)

R^2=0.984165 F(13,154) = 278.5** DW = 2.00 AR _F(7,144) = 1.8965 ARCH_ F(7,137) = 1.4069
NORM_Chi^2(2) = 0.52381 HETER_F(20,130)= 1.4254 RESET_F(1,150) =9.9287**
CHOW_F(12,139)= 1.2136; MSE=0.12.

Note: t-value in parenthesis

From Table 3, it appears that 5 of the 8 models estimated are well specified and do not present any problems in the diagnostics. The coefficient of the time trend (*T*) is positive and statistically significant at the 5% level in 6 of the 8 models estimated. This finding denotes that both Italians and foreigners regard Sardinia as a good quality destination. In 5 of the 8 models estimated, the coefficient of the Easter dummy is statistically significant at the 5% level. This fact reveals the particular importance of the Easter holiday in explaining the pattern of tourism. Finally, in all the models with the only exception for the *LNPREX* model (extra-hotel nights of stay, with normalisation) the seasonal dummy variables demonstrate that the demand for tourism is rather highly influenced by seasonal factors, including statutory or religious holidays such as Christmas. Notably, in the estimation phase the main differences between Italian and foreigners' seasonal pattern appear. On one hand, foreigners regard Sardinia as an appealing destination in May, June, July and September. On the other hand, Italians regard Sardinia as an appealing destination in the peak months (August and July).

5. Forecasting phases and results.

The aim of this section is to present the forecasting results obtained for the year 2003. There is a vast amount of literature on tourism forecasting and an effort must still be made to consider the models that best predict the future. Quantitative and qualitative approaches can be used for forecasting. Amongst the quantitative approaches two main types are distinguishable: time series "non-causal" and "causal" models. In the present study, the ARIMAX approach is used to mediate between the former approaches. The ARIMAX approach, in fact, has several advantages. It makes use of the more sophisticated time series Box-Jenkins as well as the "causal" analysis. In this study several non-economic variables are included on the right hand side of the equation. This approach has the advantage of avoiding the use of economic predictor variables (such as exchange rate, tourist price index, etc.) needed to forecast the dependent variable.

Data from January 1987 to December 2001 are used to forecast the demand of tourism to Sardinia for the following twelve months (January 2002 to December 2003). As previously stated, this is the *ex-post* forecasting in an effort to track the goodness of forecasting by comparing the actual with the forecast observations. The MSE, as defined above, is used for this purpose and the results presented in Table 3 for each of the estimated models. Figure 2 (maxima forecasting models) and Figure 3 (minima forecasting models) shows the actual and forecast number of nights of stay.

Figure 2 Nights of Stay in Sardinia: actual and *ex-post* forecasting for January to December 2002 (models for maxima)

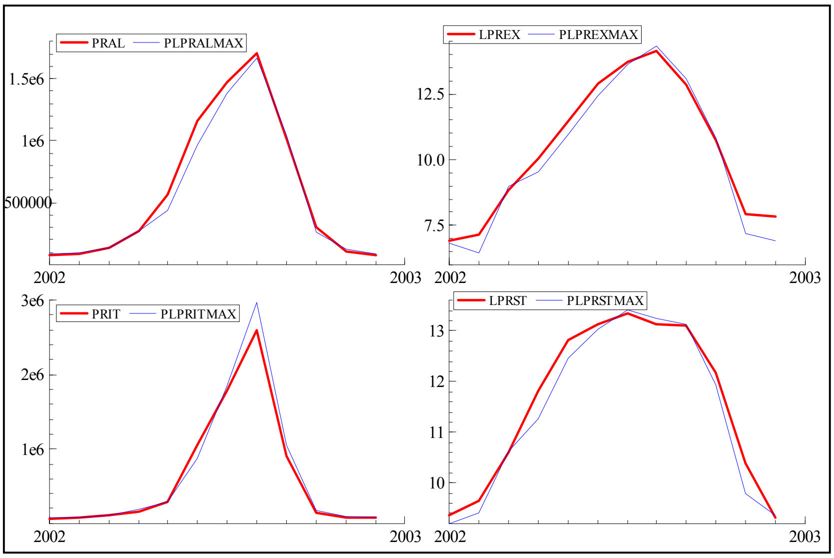
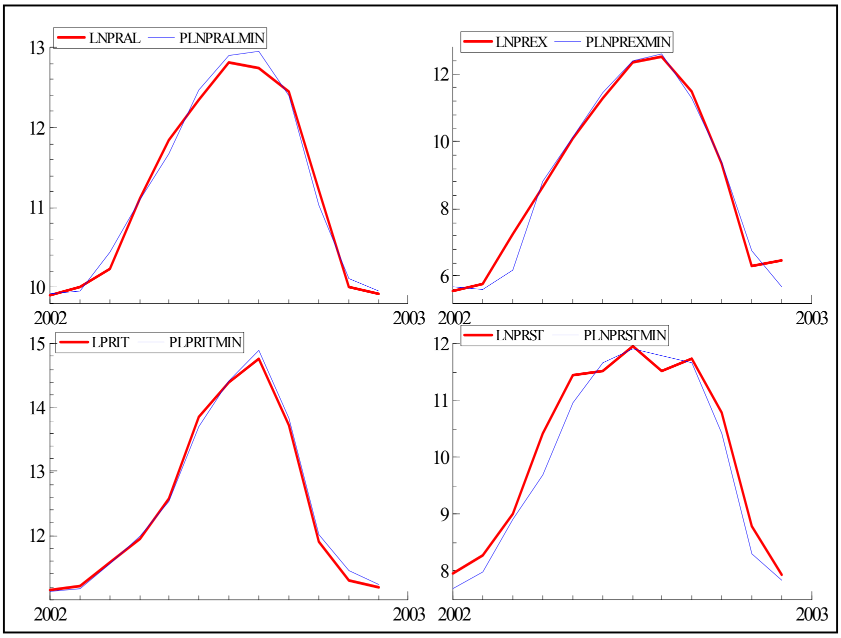


Figure 3 Nights of Stay in Sardinia: actual and *ex-post* forecasting for January to December 2002 (models for minima)



Figures 2 and 3 show forecast values capturing the seasonal distribution as well as the main turning points. The best fit model, is the OLS approach for the Italian demand for tourism which is also confirmed by the lowest MSE value obtained (0.008). The least best fit models are those for the international demand for tourism, as well as for the extra-hotel demand, confirmed both by the econometric analysis and the higher MSE values. These findings confirm those in other empirical studies where it is shown that international demand for tourism is more volatile than domestic demand (Witt and Witt, 1992).

These models and approaches are employed to forecast the *ex-ante* observations from January to December 2003. Table 4 reports the annual percentage variation from 2000 up to 2003. The last percentage variation is calculated using actual value for 2002 and forecast values for 2003. In the last row, the total flow variation to Sardinia is reported. The *ex-ante* forecast is obtained as an average of all the four components forecast of Sardinian tourism demand. One might argue that the major differences obtained in the pre-modelling analysis (see Table 2), denoting a different seasonal distribution for foreigners and Italians, the greater volatility of the international and extra-hotel components, confirmed by the econometric analysis could lead to a further mis-specification using aggregated data. These findings also confirm the result obtained in Pulina (2002).

Table 4 Nights of Stay Growth and *ex-ante* forecasting for 2003.

Models	% variation 2000-01	% variation 2001-02	Forecast % variation 2002-03	
			Minima	Maxima
	var.%00-01	var.% 01-02	var.% 02-03	var.% 02-03
Hotel Flows	10%	-2,6%	-6,1%	-2,4%
Extra-Hotel Flows	0,6%	7,7%	11,9%	15,6%
Italians	2,5%	-3,3%	5,8%	7,2%
Foreigners	22%	10,9%	3,7%	5,5%
Total Flows to Sardinia	7%	0,4%	2,3%	5%

As one can notice the year 2002 has presented stagnation in the Sardinian tourism activity. A reduction of tourism flows with respect to 2001 is remarkable for hotel accommodation and the Italian demand. This fact is argued to be due to the unusual poor weather conditions of the past season characterised by continuous rainfalls during summer (Crenos, 2003).

On one side, the 2003 forecast is more optimistic for Italians who are predicted to return to Sardinia, after the fall in 2002. On the other side, foreigners should register a slow down after high percentages of growth in the two past years. The extra-hotel accommodation should see a further boost and one expects that Italians will make use of this type of accommodation, such as "agriturismo" and bed and breakfast that have developed in Sardinia in recent years. However, a further decrease of hotel users is forecast.

6. Conclusions.

In this paper, an empirical analysis and a forecast of the demand of tourism to Sardinia for the period between 1997 and 2003 has been presented.

More advanced econometric tools have been used in the analysis of tourism demand to investigate the characteristics and properties of the time series under study. Long run and seasonal unit roots tests have led to the investigation of properties of the variables under study; major differences have been detected between the Italian and foreigners' seasonal distribution. The use of the OLS and ARIMAX approaches have given satisfactory results both in the estimation and forecasting phases. A full range of diagnostic statistics has been provided which is much neglected in tourism literature. Further work is needed to improve the estimation for international and extra-hotel accommodation demand.

Seasonality is one of the main features of tourism activity. Hence, the understanding of seasonality is a necessity for both private and public operators. In this paper, the use of monthly time series has given a deeper insight into the characteristics of the seasonal pattern for each of the tourism demand components. There are reasons for believing that the public administration, at a regional level, should adopt promotion policies to encourage a de-seasonalisation process, in particular for the domestic demand. The objective of the local authorities should also be that of promoting Sardinian tourism in off-peak months to Italian clients. Moreover, the importance of the Easter break for tourists could be used by the private and public sectors to adopt price discrimination for tourist consumers, together with higher standard of quality of the goods and services supplied during "second holiday" periods.

The importance of trading-day factors has been confirmed in this study. This finding is a further informative element for tourist operators, who should be more aware of the concentration of tourist flows in the months where the number of Saturdays in a month is higher.

The forecast for the year 2003 predicts the growth of the tourism demand to Sardinia with respect to the previous year. However, a further decrease is predicted for the flows in hotel accommodation. Overall, the demand for tourism to Sardinia should regain the positive trend of the past years, given the international scenario previously described.

As a further research step, the *ex-ante* forecasting results have been employed in a wider methodology context in Crenos (2003). Using both the quantitative approaches described so far and a Delphi qualitative approach an integrative forecasting has been run. The qualitative approach, in fact, based on expert-opinions, has the purpose to mediate the quantitative approach deficiencies derived from not including further relevant quantitative and qualitative variables into the econometric models.

Bibliography

Box, G.E.P. and Cox D.R. (1964), "An Analysis of Transformations", *Journal of the Royal Statistical Society, Series B,* 26(2), pp. 211- 43.

Contu, G. (1997), *Domanda ed Offerta nel Nord Sardegna e Ruolo dei Prezzi e Qualità*, Tesi di Laurea in Scienze Politiche, Università degli Studi di Sassari.

CreNos (2002), *9° Rapporto sull'Economia della Sardegna*, Università degli Studi di Cagliari. Università degli Studi di Sassari.

CreNos (2003), *10° Rapporto sull'Economia della Sardegna*, Università degli Studi di Cagliari. Università degli Studi di Sassari.

Dickey, D.A. and Fuller W.A. (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica,* 49(4), pp. 1061-1072.

Doornik, J.A. (2001), *Give Win 2.00 - An Interface to Empirical Modelling,* International Thomson Business Press, London.

Findley, D., Monsell B.C., Bell W., Otto M.C. and Chen B.C. (1998), New capabilities and methods of the X-12-ARIMA seasonal-adjustment program, *Journal of Business & Economic Statistics,* vol.16, n.2, pp. 27-177.

Franses, P.H. (1991a), "Model Selection and Seasonality in Time Series", *Institute Research Series*, Tinbergen, 10, pp. 94-165.

Franses, P.H. (1991b), "Seasonality, Non-Stationarity and the Forecasting of Monthly Time Series", *International Journal of Forecasting*, 7, pp. 199-208.

Gonzàles, P. and Moral P. (1996), "Analysis of Tourism Trends in Spain", *Annals of Tourism Research*, 23(4), pp. 739-754.

Hargreaves, C.P. (1994), *Nonstationary Time Series Analysis and Cointegration,* New York, Oxford University Press.

Kulendran, N. and Witt S.F. (2001), "Cointegration Versus Least Squares Regression", *Annals of Tourism Research*, 28(2), pp. 291-311.

Malacarni, L. (1991), "Analisi Econometrica della Domanda Turistica in Italia", *Politica del Turismo,* 2, pp. 19-29.

Mizon, G.E. (1996), "Progressive Modelling of Macroeconomic Time Series - The LSE Methodology", Ch. 4 in Hoover, K.D. (ed), *Macroeconometrics,* California, Kluwer Academic Publishers.

Pesaran, M.H. and Pesaran B. (1997), *Working with Microfit 4.0*, O.U.P.

Pindyck, R.S. and Rubinfeld D.L. (1991), *Econometric Models and Economic Forecasts*, Singapore, International Editions.

Pulina, M. (2002), *Modelling demand for tourism in Italy,* PhD Thesis, Southampton University.

Pulina, M. and O'Brien R.J. (2002), "Monthly, Annual and Quarterly Frequencies: a Comparison of Models for Tourism in Sardinia" Discussion Paper, Southampton University.

Raeside, R., Adams J., Rush R. and Haldane K. (1997), "Modelling and Forecasting Domestic Tourism in Scotland", *Paper presented at the 17th International Symposium on Forecasting, Barbados*, pp. 6-11.

Seddighi, H.R. and Shearing D.F. (1997), "The Demand for Tourism in North East England with Special Reference to Northumbria: An Empirical Analysis".

Sinclair, M.T. and Stabler M. (1997), *The Economics of Tourism.* Routledge, London and New York.

Song, H., Romilly P. and Liu X. (2000), "An Empirical Study of Outbound Tourism Demand in the UK", *Applied Economics*, 32, pp. 611-624.

Touring Club Italia (2003), *Annuario del Turismo 2003,* TCI, Milano.

Webb, R.H. (1995), "Forecasts of Inflation from VAR Models", *Journal of Forecasting,* vol. 14, pp. 267-285.

Witt, S. and Witt C.A. (1992), *Modelling and Forecasting Demand in Tourism*, London, Academic Press.