MODELLING AND FORECASTING OF COUNTRY TOURISM DEVELOPMENT IN LITHUANIA

Marija Burinskiene[1], Vitalija Rudzkiene[2]

*[1] Dept of Urban Engineering, Vilnius Gediminas Technical University,
Saulėtekio al. 11, LT-10223 Vilnius-40, Lithuania, e-mail: marija.burinskiene@ap.vtu.lt
[2] Dept of Legal Informatics, Law University of Lithuania,
Ateities g. 20, LT-08303 Vilnius, Lithuania, e-mail: vital@mruni.lt*

Abstract. Development trends of the European Union society show that tourism has the highest increase rate. Tourism is an important branch of every country's economy, because tourism consumption spreads over a large number of goods and services. In the last years development of country tourism in Lithuania has had a strong support from the government as well as the EU funds. The effect of economic variables on forecasting accuracy of guest numbers in Aukštaitija National Park is examined using ARMA models. The empirical results show that country tourism has a positive linear trend and the consumer price index can be used as an explanatory variable forecasting country guest number.

Keywords: tourism development, time series, ARMA models, consumer price index.

1. Introduction

Development trends of the European Union society show that tourism has the highest increase rate.

Tourism is an important branch of every country's economy, because tourism consumption spreads over a large number of goods and services. The number of tourists is influenced by prices as well as by changing preferences, and this is the reason for difficulties in obtaining a good forecast.

Since 1994 the State Department of Tourism annually conducts a sample survey of foreigners from aboard. Foreigners who have spent at least one night but not more than free months and who are leaving Lithuania are interviewed. In a similar way, the same – day visitors to Lithuania are interviewed just with fewer questions.

A great majority of tourists came to Lithuania on their own, 4 percent (3 percent in 2002) of tourists came by purchasing travel packages [2].

19 percent (26 percent in 2002) of all the nights tourists spent in hotels and motels, 5 percent (4 percent) – in rest-houses and sanatoria, 12 percent (13 percent) – in a private accommodation sector, and 58 percent (49 percent) stayed at friends or relatives.

In 2003 one tourist spent in Lithuania on average 931 Lt, 16 percent more than in 2002 (800 Lt). The highest average expenditure was made by tourists from Asia (Israel, Japan) – 1660 Lt (1309 Lt). Among the main markets, the highest average expenditure was made by tourists from the United Kingdom – 1324 Lt (1597 Lt), the USA – 1314 Lt (in 2002–2881 Lt), Russia – 822 Lt (989 Lt). The lowest expenditure was made by tourists from neighbour countries: Belarus – 685 Lt (530 Lt), Estonia – 362 Lt (351 Lt). Average expenditure of one tourist per day was 144 Lt [2].
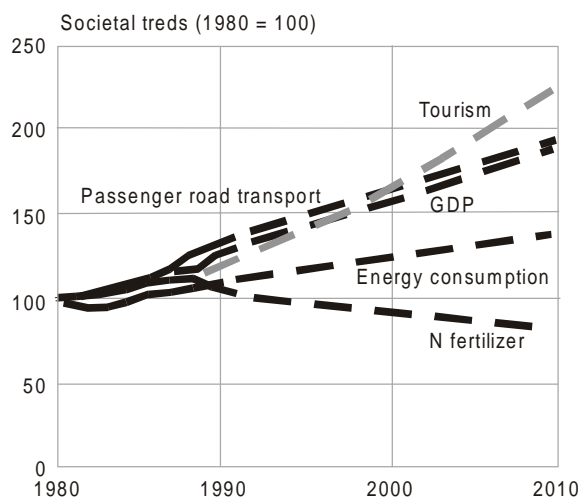
Fig 1. Development trends of the European Union society [1]

Guest nights or guest number is based on pure time series analytic models, on economic demand models or on mixed forms.

Mathematical modeling of tourism suggests that the income level of a country and prices (for domestic tourism) and exchange rate (for inbound tourism) are important determinants (e g [3], [4]). The real gross domestic product (GDP) or consumer price index (CPI) are good proxy variables for income level [5]. As the purchasing power of international tourists the rate of consumer price indices with the exchange rate between countries currency can be specified as a relative price. From a forecasting point of view, pure time-series models often turn out to be as good as the economic models. One reason for a relatively modest forecasting performance of economic models is difficulties in obtaining good forecasts of price variables. Better prediction of tourism demand may be expected by forecasting tourism series jointly with variables in which information on price changes and taste are incorporated.

Time series models depart from micro or macro level. At a micro level day-to-day movements of guest nights are presented. An integer – valued autoregressive model of the order one [INAR(1)] governs these movements. Basic work on the INAR(1) model is due to [6, 7] and others.

Starting from a micro level, i e from a model explaining the daily numbers of guest nights for a specific hotel, the cross-sectional and temporal aggregation is applied. The cross sectional aggregations consider a model for the number of hotel and cottage visitors for a region at a certain day. By aggregation, cross sectionally as well as temporally, the result is a model of a macro level for aggregate guest nights or guest number per month, quarter, etc.

Temporal aggregation can be obtained applying INAR(1) model [8]. For cross-sectional aggregation, AR(1) and INAR(1) are considered, and these models are closely related [5]:

$$y_{it} = a y_{it-1} + \varepsilon_{it}. \qquad (1)$$

In (1) y_{it} is the number of guest nights at day t at the checkout time for hotel (cottage) $i = 1, 2… N$. A random number of guests with check-ins in the period $(t - 1, t)$ is represented by ε_{it}.

For economic-demand models the demand of tourism is evaluated by maximizing the utility function:

$$F(q) = \prod_{j=1}^{m} \left(q_j - k_j \right)^{\phi_j}, \ 1 > \phi_j > 0, \ \sum_{j=1}^{m} \phi_j = 1 \ \text{subject to}$$

$$M = \sum_{j=1}^{m} p_j q_j,$$

where M – *total* expenditures on tourism, q_j number of guest nights in hotels and cottages, p_j – prices for goods and services, k_j – parameter [9].

2. Identifying patterns in time series data

Most social and economic data take the form of time series – e g the Gross National Product (GNP) by years or quarters, indices of consumer and wholesale price, the number of tourists, to name just a few. Usually analysis of time series data focuses on two types of problems:
- Attempting to estimate the factors which produce the patterns of the series;
- Using the estimate in forecasting the future behavior of series.

In reality the social and economic world is a complicated network of interdependencies which is rarely captured well by models involving only two variables (one of which is time). Typically, the value of any variable is influenced not only by one another variable, but also by two, three, or even a multitude of other variables. The number of tourists in a region indeed may depend on the income level, quality of proposed services, prices, preferences and so on. In solving the question whether changes in the values of one variable tend to be associated with changes in the others, we can use several different statistics. The most popular is *Pearson's sample coefficient of correlation*

$$r = \frac{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{S_x S_y}, \qquad (2)$$

where $\bar{x}$, $\bar{y}$ – mean of variables x and y, respectively,

S_x, S_y – standard deviations, n – sample size.

Most time series consist of elements that are serially dependent in the sense that one can estimate a coefficient or a set of coefficients that describe consecutive elements of the series from specific, time-lagged (previous) elements. In addition, each element in the series can also be affected by the former error. This can be summarized in an autoregressive moving average (ARMA) model, introduced by Box and Jenkins [5]:

$$y_t + \sum_{i=1}^{p} a_i y_{t-i} = b_0 \varepsilon_t + \sum_{k=1}^{q} b_k \varepsilon_{t-k}, \qquad (3)$$

where $a_1, a_2,…,a_p$ – autoregressive model parameters, $b_1, b_2,…b_q$ – moving average model parameters. We assume that $\{\varepsilon_t\}$ is random sequence with $E(\varepsilon_t) \geq 0$ and $V(\varepsilon_t) = \delta$.

The input series for ARMA models needs to be stationary, that is, it should have a constant mean, variance and autocorrelation through time. An autoregressive process will be stable, if the parameters are within a certain range, for example, if there is only one autoregressive parameter then it must follow within the interval [–1,1].

In general, time series can be thought of as consisting of different components (a seasonal component, a trend component, a cyclic component and a random component). The trend component produces a steady increase

or decrease in the series. The trend describes the influence of long-term factors the effect of which can be evaluated progressively. Changes in social-economic life are usually slow, so the trend in such cases can be described by a linear equation:

$$y_t = c_0 + c_1 x_t + \varepsilon_t. \qquad (4)$$

The method of least squares is a commonly used approach for calculating coefficients c_0 and c_1.

3. Data and empirical results

We consider the quarter time series of the number of guests in Aukštaitija National Park. The purpose of this investigation is to discover hidden patterns in the series and to predict the process evaluation. Consumer price index and gross domestic product (GDP) [10] (Fig 1) are used as a proxy for the purchasing power of the tourists. The influence of these economical indicators is distinct in different spheres of social life (e g [11]). The figure reveals a strong seasonal pattern in GDP and tourists series with major peaks in summer and troughs in winter period. The GDP is expressed in real prices and based on quarterly observations. Seasonal pattern in Lithuanian CPI is distinct only at the beginning of the investigated period. The series in Fig 2 are presented on two axes: an axis of GDP and tourists is on the left-hand side of the diagram and an axis of CPI is on the right-hand side.

The correlation coefficient between the number of tourists and GDP is 0,42 (*corr* = 0,42) and the correlation coefficient between the number of tourists and CPI is – 0,41 (*corr* = –0,41). Both values are significant with a significance level $p = 0,05$. In Fig 2 we can see that in Lithuania the growth of GDP does not increase CPI; on the contrary, the correlation coefficient between GDP and CPI is – 0,932.

In this study the ARMA models were estimated using a statistical program *Statistica* 5,5. For parameter estimation the *quasi-Newton* method as a function minimization algorithm is used. The most commonly used lack of fit indicator in statistical fitting procedures is the mean square error (MSE). The smallest MSE is applied as one of the criteria for finding the most appropriate models. The value of MSE is computed as an average of
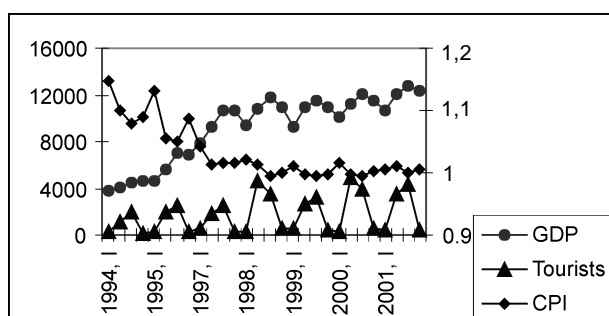

Fig 2. Guest number in Aukštaitija National Park, GDP and CPI per quarter, 1994–2001

square error values. The modelling framework of Box and Jenkins [12] rests on applying the seasonal differencing. In our case the series are rather short ($n = 28$) and the analysis reveals that in most cases there is no substantial improvement in the results applying differencing.

We started with estimating an order of the ARMA model. Analysis of autocorrelation and partial autocorrelation functions together with mean square errors helped us to find the most appropriate orders of models.

Regarding time series of the number of guests, we obtained the following estimation results:

$$y_t = 1655,6 + 0,575 y_{t-1} - 0,89 y_{t-2} + 0,58 y_{t-3} + \varepsilon_t. \quad (5)$$

For all the parameter estimates asymptotic standard errors were computed. These are computed from the matrix of second-order partial derivatives (i e Hessian matrix). The respective t value (t statistics) is calculated as parameters divided by standard errors.

Standard errors are, respectively: $\delta(a_1) = 0,18$; $\delta(a_2) = 0,13$; $\delta(a_3) = 0,20$.

All the parameters are significant at a significance level 0,05, MSE = 903 800.

When a trend is subtracted, the equation acquires a new form:

$$y_t = 865,2 + 58,8t - 0,941 y_{t-2} + \varepsilon_t. \qquad (6)$$

The first-order parameter is insignificant ($a_1 = 0,002$; $\delta(a_1) = 0,101$; $\delta(a_2) = 0,104$; MSE = 548 300.

In the next model the income variable is specified as a real gross domestic product (GDP). The positive value of the correlation coefficient between the number of tourists and GDP shows that increasing GDP increases demand for tourism. A linear trend was subtracted from the data series $y_t = y_t - 0,231 * GDP + 413,71$. Then the series was differenced with lag 1 ($z_t = \nabla y_t = y_t - y_{t-1}$) because the autocorrelation function had no decay, and partial autocorrelation functions had no significant spikes. After differencing, the AR(1) model was indentified:

$$z_t = -0,733 z_{t-1} + \varepsilon_t. \qquad (7)$$

The standard error is $\delta(a_1) = 0,15$; MSE = 2 758 000.

When the consumer price index is used as an income variable, we have the following model:

$$y_t = 16836 - 14650 \cdot CPI - 0,902 y_{t-2} + \varepsilon_t. \qquad (8)$$

The first-order parameter is insignificant ($a_1 = 0,07$; $\delta(a_1) = 0,11$; $\delta(a_2) = 0,117$), MSE = 553 700.

The forecasts are obtained as follows: the identified models are estimated on the data up to 2000-IV quarter. Then, 2000-IV is taken as the forecast origin for forecasting from 1 up to 4 steps ahead. The model is then re-estimated of the data through 2001-I, with the form of the model unchanged. Then 2000-I is taken as the forecast origin and the procedure is repeated. The forecasts and the observed values (data) are presented in Fig 3.

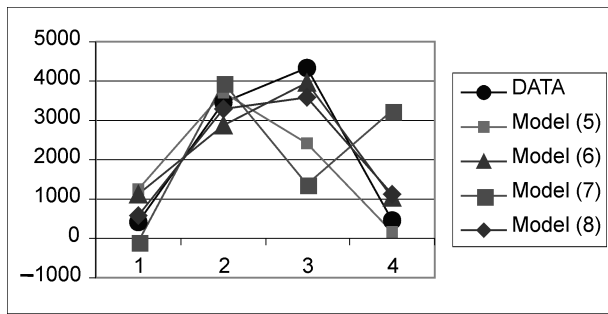The mean square errors of differences between the

Fig 3. Forecasts of guest number for 2001

Fig 4. Forecasts of guest number for 2002

forecast and real values were calculated in order to compare the forecasts. By this criterion, model (8) gives the best forecast (MSE = 272 400). Model (6) with MSE = 332 400 ranks a second one. The forecasts of the fourth and sixth models were less exact (MSE = 1 110 000 and MSE = 4 234 320).

After evaluating the quality of the forecast for 2001 of the investigated ARMA models, we can predict future cases for 2002. For this purpose the use of models (7) and (5) would be reasonable. For model (8), we will assume that the level of CPI in 2002 will be the same as in 2001. Computed forecasts for 2002 are presented in Fig 4.

As we see in Fig 4, the values of forecasts evaluated by model (6) are larger than values predicted by model (8). The accuracy of computed forecasts by model (8) depends on changes in Lithuanian CPI. The forecasts given by model (6) are influenced only by time.

As we see in Fig 4, the values of forecasts evaluated by model (6) are larger than those predicted by model (8). The accuracy of computed forecasts by model (8) depends on changes in Lithuanian CPI. The forecasts given by model (6) are influenced only by time.

4. Conclusions

Development trends of the European Union society show that tourism has the highest increase rate. Tourism is an important branch of every country's economy, because tourism consumption spreads over a large number of goods and services. In the last years development of country tourism in Lithuania has had a strong support from the government as well as the EU funds.

Since 1994, the State Department of Tourism annually conducts a sample survey of the foreigners from aboard. A great majority of tourists came to Lithuania on their own, 19 percent (in the year 2002 – 26 percent) of all the nights the tourists spent in hotels and motels, 5 percent (4 percent) – in rest-houses and sanatoria. In 2003 one tourist spent in Lithuania on average 931 Lt, 16 percent more than in 2002 (800 Lt). Average expenditure of one tourist per day was 144 Lt

An empirical work with guest series and the purchasing power of guests in Aukštaitija National Park investigates hidden patterns in the data series and differ-
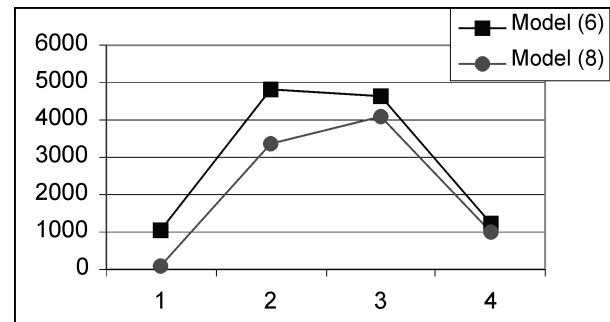
ences in forecasts produced by using different ARMA models. Relatively similar estimates of models (6) and (8) suggest that the modelling approach has some support from the data. The estimation results identify some problems in evaluating the income level of guests and obtaining a reasonable proxy for the purchasing power. The most watched variable in any economy is GDP or GDP per capita. This variable is widely used for an economic analysis. However, GDP is not relevant as an explanatory variable for assessment and forecasting of country tourism in Lithuania. This effect can be explained by lack of a direct dependence between GDP and the purchasing power of guests.

References

1. European Commision. Taking European Environment Policy into the 21st century. Luxemburg. 1996, p. 22.

2. Lithuanian Tourism statistics. 2003. Vilnius, 2004. 44 p.

3. Melenberg, B. and van Soest, A. Parametric and Semi-Parametric Modelling of Vacation Expenditures. *Journal of Applied Econometrics*, 11, 1996, p 59–76.

4. Nordström, J. Unobserved Components in International Tourism Demand. *Umeå Economic Studies*, 502, 1999.

5. Bränäs, K.; Hellström, J. and Nordström, J. A New Approach to Modelling and Forecasting Monthly Guest Nigts in Hotels. *Umeå Economic Studies,* 503, 1999.

6. Alzoid, A. A. and Al-Osh, M. A. First Order Integer Valued Autoregressive (INAR(1)) Process: Distributional and Regression Properties. *Statistica Neerlandica,* 42, 1998, p 53–61.

7. McKenzie, E. Autoregressive Moving – Average Processes with Negative Binomial and Geometric Marginal Distributions. *Advances in Applied Probability*, 18, 1986, p 679–705.

8. Bränäs, K. Explanatory Variables in the AR(1) Count Data Model. *Umeå Economic Studies*, 381, 1995.

9. Anderson, R. A. Perfect Price Aggregation and Empirical Demand Analysis. *Econometrica*, 47, 1979, p 1209–1230.

10. http://www.std.lt/

11. Dzemydienė, D. and Rudzkienė, V. Multiple Regression Analysis in Crime Pattern Warehouse for Decision Support. LNCS 2453: Database and Expert Systems Applications, Springer, 2002, p 249–258.

12. Box, G. E. P. and Jenkins G. M. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisko, 1976.

KAIMO TURIZMO PLĖTROS MODELIAVIMAS IR PROGNOZAVIMAS

M. Burinskienė, V. Rudzkienė

S a n t r a u k a

Europos Sąjungos raidos tendencijos byloja, kad didžiausi yra turizmo plėtros tempai. Turizmas – tai svarbi kiekvienos šalies ekonomikos dalis, nes turizmo poreikiai sukuria daug papildomų prekių ir paslaugų. Per pastarąjį dešimtmetį kaimo turizmui skiriama valstybės ir Europos Sąjungos struktūrinių fondų parama. Siekiant nustatyti ekonominių kintamųjų įtaką Aukštaitijos nacionalinio parko svečių skaičiaus prognozavimo tikslumui, taikytas ARMA modelis. Empiriniai rezultatai rodo, kad kaimo turizmas plėtojasi teigiama augimo kryptimi, o vartojimo kainų indeksas gali būti taikomas kaip pagrindinis kintamasis prognozuojant svečių skaičių šalyje.

Raktažodžiai: turizmo plėtra, laiko eilutės, ARMA modeliai, vartojimo kainų indeksas.

Marija BURINSKIENĖ. Dr, Prof, head of Dept of Urban Engineering, Vilnius Gediminas Technical University (VGTU).
Teaching: Urban Infrastructure, Transportation Planning, Engineering Equipment of Territories. Publications: author of over 70 scientific publications, 2 monographs, co-author of a textbook. Projects: project manager of more than 40 national projects. Conferences: participant of more than 30 international conferences. Research interests: sustainable development, transportation engineering and decision support system.

Vitalija RUDZKIENĖ. Dr, Prof, Dept of Informatics and Statistics, Mykolas Romeris University.
Teaching: Social Statistics, Statistical Technologies in Management, Statistical Technologies in Psychology, Legal Statistics. Publications: author of more than 20 research papers. Conferences: participant of nearly 30 conferences. Projects: expert of 4 international projects. Research interests: application of multidimensional statistical methods for social-economic processes analysis, statistical technologies in management; data mining methods, modeling and simulation of sustainable development, knowledge representation.