10.2478/v10103-012-0024-8

EWA K. DŁUBAKOWSKA-PUZIO[*], MICHAŁ KARPUK[**], KAMIL PUZIO[***]

Analysis of Tourism Service Quality in Kołobrzeg Region by Means of Time Series Models

Abstract

The aim of the given paper is to present the analysis of tourism services by means of time series models and forecasting of evaluation of tourism services. Tourism services are analysed according to various parameters: hotel stay price, hotel services quality (such as cleanness of rooms, check-in, information provision etc.), catering quality and medical service quality.

The research has been undertaken on the basis of responses of foreign guests of 13 hotels in Kołobrzeg region taken from 1400 questionnaires divided according to age and sex of respondents. Various econometric models were used for the analysis of statistic regularities. First, customers evaluated quality of their stay in hotels. These data were examined during the 2006-2009 time period as a stochastic process. It was found that the processes are nonstationary, that is why the ARIMA (1,1,1) model was used in the study. On the basis of the analyses and prognoses one can deduce that models of time series make it possible to estimate a tendency that occurs for an analysed parameter, however misprediction is quite possible to appear (up to 30 %). Similar results were achieved during the analysis of evaluation of hotel stay quality on the basis of sex of respondents. One should point out that an average value of male evaluation rate is higher than an average value of female evaluation rate.

[*] Ph.D., University of Technology in Koszalin

[**] Ph.D., University of Technology in Koszalin

[***] Social School of Entrepreneurship and Management in Lodz, Branch Faculty in Kołobrzeg

Time series for analysed variables were integrated into level I(1). A certain co-integrational connection was found between the evaluation of hotel stay and number of stays in a given hotel, where tourists that have already visited a given hotel for several times presumably give a high evaluation rate of hotel stay quality. The result of the analysis of long-term relation between hotel stay and duration of check-in is presented in high mutual dependence of hotel evaluation rate on evaluation of check-in duration. A hotel stay price influences hotel stay evaluation in a negative way, but catering and medical services are considered as standard and do not have any particular influence on hotel stay evaluation.

The undertaken study shows that methods that take into account time series can be successfully used in analysis of parameters of tourist comfort and in evaluation of hotel services.

1. Introduction

Development of such economic branch as tourism is directly connected with development of hotel market and increase of quality of customer services. Hotel and tourism businesses must be closely related to a client and offer more services in order to achieve better financial results. Enterprise activity of tourism businesses occurs under pressure of strong competition and gets worse because of the world financial crisis (Woźniak 2009, pp. 127-138). Therefore broadening or provision of new services is of great importance for activity of tourism businesses.

Moreover, analysis of customers’ opinions on quality of provided services and their feedback on relations between buyers and producers of tourism services has very significant positive influence on results of financial activity of tourism companies.

The given paper presents the analysis of tourism services according to various parameters: hotel stay price, hotel services quality (such as cleanness of rooms, check-in, information provision etc.), catering quality and medical service quality and quantity (Dłubakowska-Puzio et al. 2009, pp. 181-194). A questionnaire consisted of 16 questions which were grouped according to the following segments: hotel services quality, catering quality, medical services quality, and evaluation of hotel stay. Some of these questions are presented in the Table 1.

Table 1. Several questions from the questionnaire

No.	Question	Scale of evaluation rates (min-max)
X0a	Is it your first time at the hotel?	0 – 1
X1	Check-in duration.	0 – 5
X6	Breakfast offer.	0 – 5
X12a	Range of treatments: number of treatments offered.	0 – 5
X14a	Medical care: nurse attitude towards guests.	0 – 5
X15	Guests guide care	0 – 5
X16	How will you evaluate the given hotel during your conversation with friends and family?	0 – 5

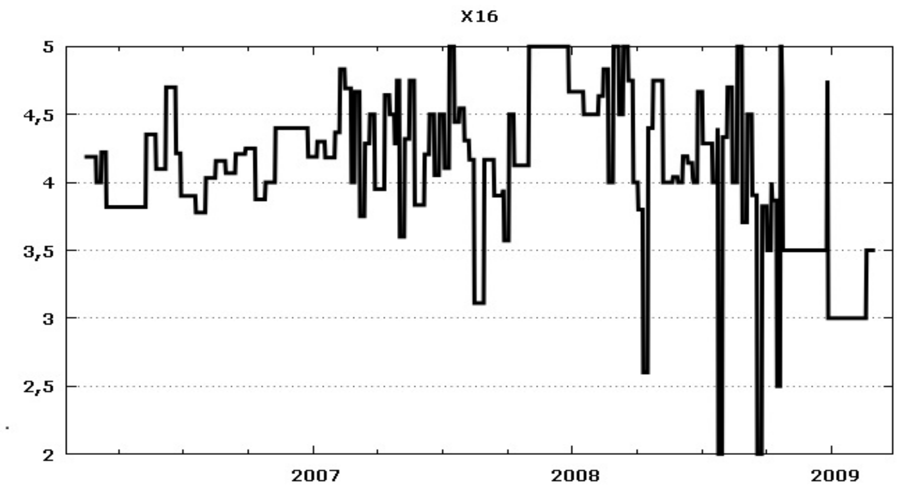
Source: own data.

Moreover, monitoring of hotel stay prices has been launched. Average price depends on time and can be described with dynamic variable SrCena.

The research has been undertaken in time period 2006-2009 on the basis of responses of foreign guests of 13 hotels in Kołobrzeg region taken from 1400 questionnaires divided according to age and sex of respondents.

Traditionally, statistical research uses methods of analysis of means and dispersions, and detection of correlational dependences between various data, on the basis of which conclusions about significant dependences between different phenomena are made (Maddala 2006, pp. 577-607). However, this kind of research does not take into account dynamism of concerned phenomena, their changes in time, which causes substantial problems with forecasting. For example, Figure 1 presents a time series of average evaluation rates of hotel service quality given by responding tourists during the time period 2006-2009. Similar time dependences are typical also for average evaluation rates made by tourists responding to other kinds of questions.

Figure 1. Average evaluation rate of hotel stay quality (question X16)



Source: own data.

Phenomena taking place in the tourism branch, evaluated by tourists in questionnaires, are considered as stochastic processes during the 2006-2009 time period. Therefore, the given paper uses methods of time series research. Hotel quality evaluation is correlated with a range of factors, such as hotel room prices, quality of services provided at the reception desk, catering quality. It is necessary to admit that one of special features of tourism in Kołobrzeg region is its health-improving character and provision of spa and rehabilitation services of high quality for tourists.

The given paper had several goals, which are the following:

1. Research of evaluation data of tourism services quality in Kołobrzeg region in the 2006-2009 time period as a stochastic process and detection of its regularities and features;
2. Forecasting of evaluation of tourism services quality on the basis of time-series methods;
3. Research of influence of other features (hotel room prices, quality of medical services, catering quality etc.) on evaluation of service quality by tourists;
4. Research of influence of sex of tourists on evaluation of hotel stay quality by tourists.

2. Statistical methods used in the article

ARIMA model for forecasting of evaluation of hotel stay quality

In econometric studies, ARIMA models (Autoregressive Integrated Moving Average), introduced by Box and Jenkins (Box, Jenkins, 1976) are usually used for time series analysis.

General view of ARIMA model (p,d,q) is the following:

$$x_t^d = \varphi_1 x_{t-1}^d + \varphi_2 x_{t-2}^d + ... + \varphi_p x_{t-p}^d + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \Theta_2 \varepsilon_{t-2} - ... - \Theta_q \varepsilon_{t-q} \qquad (1)$$

where x_t is the analysed time series (the given study analyses time series on the basis of answers to the question).

$$x_t^d = \Delta^d x_t = x_t - C_d^1 x_{t-1} + C_d^2 x_{t-2} - ... + (-1)^d C_d^d x_{t-d} \qquad (2)$$

ε_t – rest of the model for a time period t

$\varphi_1, \varphi_2, ..., \varphi_p, \Theta_1, \Theta_2, ..., \Theta_q$ - parameters of the model

a series $x_t^d, t = 1,..., T - d$, extracted from x_t, after d-fold method of successive differences was applied, can be described by ARMA model (p,q).

In ARIMA model (p,d,q) three parameters can be defined: autoregressive parameter p, time-series differentiation d, moving average parameter q. An input time-series for ARIMA method should be stationary – it must have a time-constant mean, dispersion and it must lack autocorrelation. Therefore a time-series usually requires time-series differentiation d until stationarity is achieved.

Methodology of ARIMA model modelling for examined time-series consists of the following fundamental stages:

- time-series stationarity test. In case of time-series nonstationarity, time-series differentiation d is necessary;
- identification of exploratory model;
- model parameters evaluations and diagnostic verification of model adequacy;
- use of model for forecasting.

First, we should acquire a stationary series. In order to reach this result, we must analyse an autocorrelation function (ACF) and a partial autocorrelation function (PACF) (see Figure 2. Horizontal axis *lag* presents delay, vertical axis presents values of (auto)correlation variables). Quick decay of ACF values is a simple stationarity test. Moreover, at the given stage statistical tests of the unit root (Dickey-Fuller test or Augmented Dickey-Fuller test) (Dickey, Fuller 1979, pp. 427-431) are used.

If a series is nonstationary according to Dickey-Fuller statistic or rates, then, in order to achieve a stationary series, one should apply the successive differences operator which defines a value of parameter d (time-series differentiation). Thus, we become familiar with a value of one parameter in ARIMA(p,d,q).

After the achievement of a stationary series, behaviour pattern of selective ACF and PACF is investigated and hypotheses about values of parameters *p* (autoregressive scheme) and *q* (moving average scheme) are put forward. In this case several models with different *p* and *q* are used.

It is necessary to evaluate parameters after identification of the model. To this effect maximum likelihood estimation method (MLE) is used.

For every analysed model verification the number of residues is analysed. Number of residues of an adequate model should resemble a "white noise", which means that their PACF should not differ from zero. Moreover, for verification of hypothesis that observable data are a realization of the "white noise" the Q-statistic of Ljung and Box (Ljung, Box, 1978, pp. 297-303) is used. The statistic has an asymptotic distribution χ^2 with degrees of freedom *K*-*p*-*q*, where *K* is a maximum lag during the model examination.

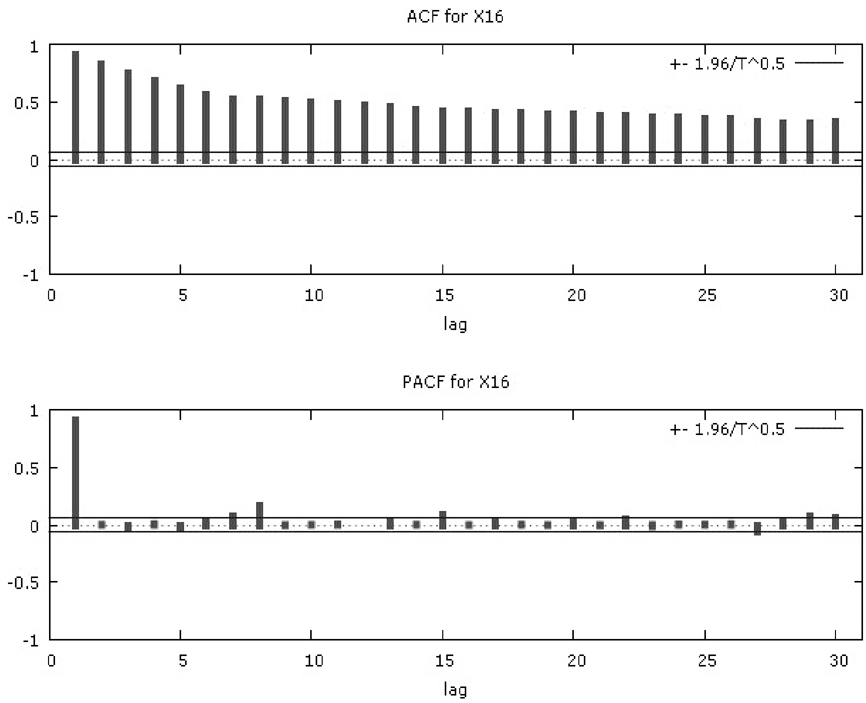
$$Q(K) = T(T+2)\sum_{i=1}^{K}\frac{r_i^2}{T-i} \tag{3}$$

where r_i is an evaluation of variable correlation coefficient x_t, x_{t-i}

In order to compare variants of models and to choose the best model we can use Akaike information criterion (Akaike 1974, pp. 716-723).

For the analysis we have used *gretl* free software (http://gretl.sourceforge.net/), which is actively used for the analysis of time series.

Figure 2. ACF and PACF function for the variable X16



Source: own data.

ACF and PACF functions for the variable X16 allow us to conclude that the variable X16 is a nonstationary variable, which means that an average value of stay quality evaluation in hotels in Kołobrzeg region changes with time under the influence of hotel clients' preferences. Therefore, human factor plays an important role in stay quality.

Similar conclusion can be made during the analysis of tourists' responses on the following questions: X1 Check-in duration., X6 Breakfast offer, X12 Range of treatments, X14 Medical care, X15 Guests guide care, and also an average price SrCena.

In order to create a stationary model for the X16 variable there were various differentiation levels of the X16 variable analysed. It was found out that after an application of ADF test (and additional KPSS test) (Kwiatkowski et al. 1992, pp. 159–178) we can achieve stationarity of time series for series differences d=1.

Analysis of correlograms of ARIMA model (p,1,q) for different values of p, q, and use of Ljung-Box and Akaike tests brings us to conclusion that the best model is the ARIMA model (1,1,1) where p=1 and q=1. Model can be presented as follows

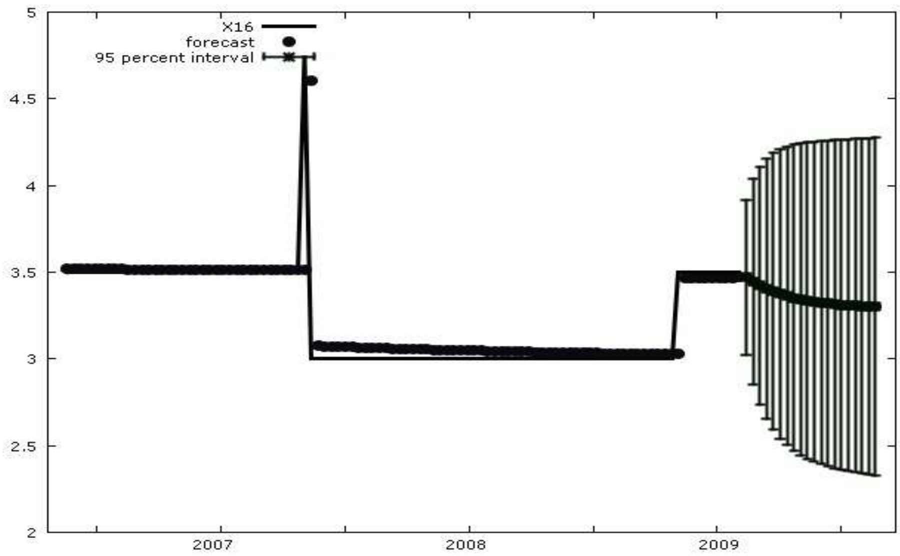
$$x_t^1 = 0.857560 x_{t-1}^1 - 0.980884 \varepsilon_{t-1} \quad (4)$$

where $x_t^1 = \Delta x_t = x_t - x_{t-1}$

Mean absolute percentage error (MAPE) makes 2.22%, coefficient of determination (R-squared) R^2=0,982 shows that the created model (4) explains 98,2% of changeability of the analysed X16 variable.

Similar results (optimal ARIMA model (1,1,1)), high values of R-squared and low values of MAPE) were received for other variables presented in Table 1.

Figure 3. Results of forecast by ARIMA model (1,1,1) for the X16 variable for 20 series



Source: own data.

It is necessary to admit that the ARIMA methodology was also used for analysis of X1, X6, X12, X14, X15 variables depending on sex. As a result, it specified that coefficients of male and female models differ for maximum 10%, even though average male evaluation value for the whole period of analysis is higher than average female evaluation value of the above-mentioned variables. It allows us to conclude that male and female models are very similar, which can be interpreted as a tiny difference in hotel stay evaluation and hotel service evaluation.

Created ARIMA models ARIMA(1,1,1) were used to forecast average evaluation values for the above-mentioned variables. Figure 3 presents forecast results for 20 days for the X16 variable on the basis of the model (4). Results of modelling show that the average value for the X16 variable decreases from 3,5 to 3,3. Relative forecast error (APE) increases from 12 % to 30%.

Similar results are achieved for other variables.

The given study shows that use of ARIMA method allows to forecast data received from tourists. The data is described by nonstationary time series with a small amount of error, which is impossible while using other methods.

3. Cointegration of processes in the tourism branch.

Analysis of economic empirical data often faces problems of nonstationary series presence and/or series with trends. As it was mentioned above, time series that are used to process data received from tourists are also nonstationary. In this case it is necessary to use differences in order to achieve stationarity and analyse the results using Box-Jenkins method.

Moreover, establishing dependences between a few nonstationary processes is lso of great importance.

In a regression model

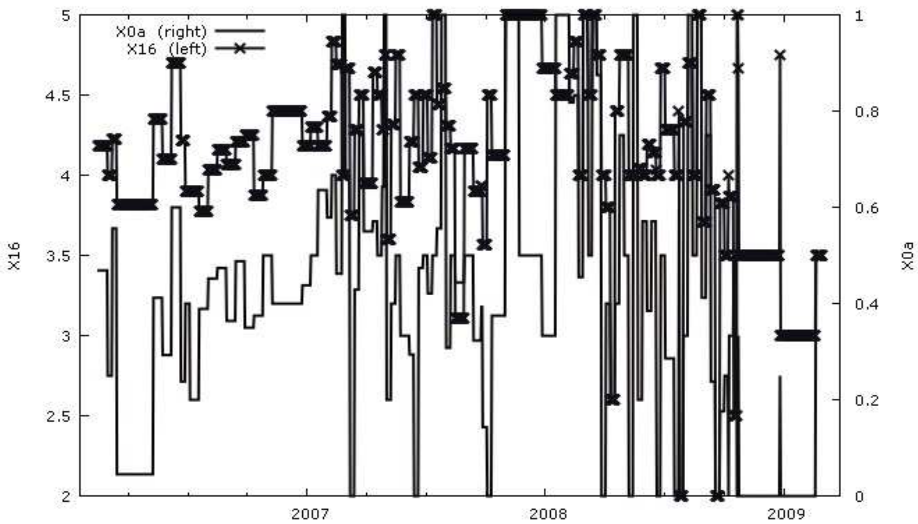
$$y_t = b \cdot x_t + \varepsilon_t \tag{5}$$

we assume that white noise ε_t is stationary. The assumption might not be right if *y* and *x* were integrated. Generally, if two processes are integrated into two different degrees, their linear combination's integration scheme will be equal to the biggest degree. Therefore, if *x* and *y* are integrated into the degree I(1), and both series include trend, then linear combination x and y will be integrate into the degree I(1) regardless of value b. If both *y* and *x* have their own trend of a similar sign, then, in case that there is no dependence between the trends, their difference will also have a trend. On the other hand, if both series are integrated into the degree I(k), then difference between them should vary around the fixed value. It means that series are characterised by similar growth rates. Two series that meet this demand are called cointegrative, and vector [1,-b] is called a cointegration vector. In this case we can notice a difference between a long-term dependence between *y* and *x*, and a short-term dynamism of deviations of *y* and *x* from their long-term tendencies (Charemza, Deadman, 1997, pp.103-144).

Notion of cointegration allows to analyse long-term relations in case of nonstationary variables. This can be connected with a notion of long-term equilibrium with the help of Granger theorem (Engle, Granger 1987, pp.251-276). Granger theorem bounds the notion of cointegration with the error correction mechanism (ECM). Error connection mechanism describes a way mentioned variable is adjusted to a long-term relation. Granger theorem allow us to interpret cointegration vector a a long-term relation between variables.

Gretl programme was also used for the cointegration analysis. Time series for analysed variables had an integration degree I(1).

Figure 4. Integrated time series X16 and X0a



Source: own data.

There was a significant cointegrational relation between variables X16 *How will you evaluate the given hotel during your conversation with friends and family?* and X0a *Is it your first time at the hotel?* (Figure 4) established, which can be interpreted as follows: tourists that have already visited the hotel (X0a→1) generally give higher evaluation rates of the hotel stay quality (X16 → 5).

Table 2 presents a very interesting result of the analysis of long-term hotel stay dependence (X16 variable) on the X1 variable (Check-in duration). Determination coefficient makes 0.990, which shows high interrelation between stay evaluation and check-in duration evaluation.

Table 2. Engle–Granger test results for the variables X16 and X1

Cointegrating regression - OLS, using observations 2006/02/16-2009/02/28 (T = 1109)
Dependent variable: X16

	coefficient	std. error	t-ratio	p-value	
X1	0.955752	0.00286437	333.7	0.0000	***

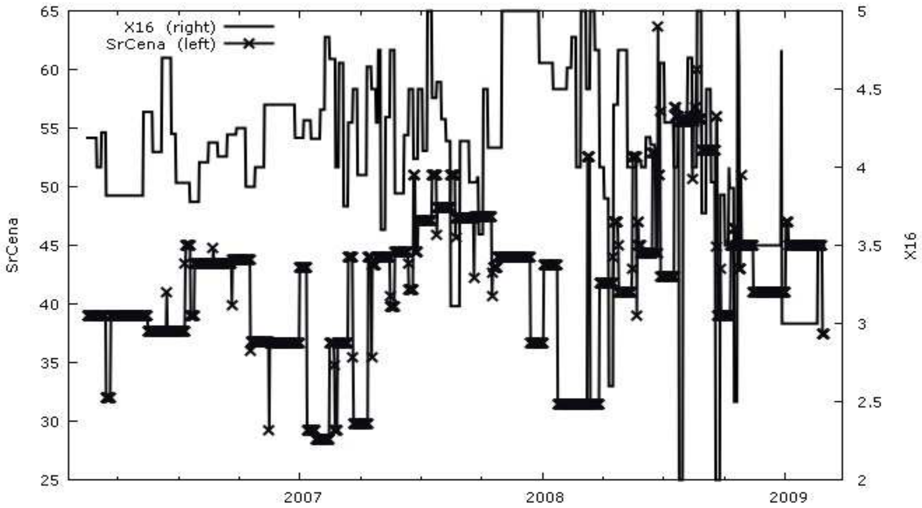
Mean dependent var	4.102426	S.D. dependent var	0.563983
Sum squared resid	187.3889	S.E. of regression	0.411246
R-squared	0.990146	Adjusted R-squared	0.990146
Log-likelihood	-587.6865	Akaike criterion	1177.373
Schwarz criterion	1182.384	Hannan-Quinn	1179.268
Rho	0.888808	Durbin-Watson	0.225046

Source: own research.

Figure 5 presents time series of the X16 and SrCena variables. During the analysis of nonstationary time series for the X16 and SrCena variables there was a very significant long-term negative dependence
$X16_t = 4{,}906 - 0{,}019 \cdot SrCena_t + \varepsilon_t$ determined.

The dependence can be interpreted as an decrease of hotel stay evaluation by 0,02 along with an increase of price by 1 euro. Increase of price will always be connected with dissatisfaction with hotel stay and smaller stay evaluation value.

Figure 5. Integrated time series X16 and SrCena



Source: own data.

4. Conclusions

The undertaken analysis of data received from hotel guests concludes the nonstationarity of time series, which describe indicators from tourism branch. For real analysis and data forecast it is necessary to use methods of analysis of nonstationary time series (Box-Jenkins methodology). Moreover, it is required to pay attention to cointegration of processes which describe data received from tourists.

During our research we were interested in influence of various factors on the evaluation of hotel stay quality by foreign guests. It was found out that check-in duration at the reception desk and hotel stay price is of great importance. Spa and rehabilitation services and catering services have less influence. Therefore we can deduce that services that are considered as standard services by foreign tourists should be offered on a high level. Hotel stay evaluation made by males and females does not have any significant difference, but males' evaluation rates are higher to a little degree, which can be interpreted as smaller demands of males concerning standards of stay in a hotel.

References

Akaike, H. (1974), *A New Look at the Statistical Model Identification*, 'I.E.E.E. Transactions on Automatic Control', 19

Box, G.E.P., Jenkins G.M. (1976), *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco

Charemza W.W., Deadman D.F. (1997), *Nowa ekonometria [New econometrics]*, PWE, Warszawa

Dickey D.A., Fuller W.A. (1979), *Distribution of the Estimators for Autoregressive Time Series with a Unit Root*, 'Journal of the American Statistical Association', 74

Dłubakowska–Puzio E.K., Karpuk M., Puzio K. (2009), *Przyszłość uzdrowiska Kołobrzeg w obliczu ogólnoświatowych zjawisk kryzysowych [The future of Kołobrzeg resort in the face of global crisis phenomena]*, [in:] M.Malicki, S.Bilan (ed.) *Polityka gospodarcza, społeczna i ekologiczna w dobie kryzysu światowego [The economic, social and ecological policy in times of global crisis]*, Wydawnictwo Politechniki Szczecińskiej, Szczecin-Kijów

Engle R. F., Granger C. W. J. (1987), *Co-integration and error correction: Representation, estimation, and testing*, 'Econometrica', 55

http://gretl.sourceforge.net/

Kwiatkowski D., Phillips P. C. B., Schmidt P., Shin Y. (1992), *Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root*, 'Journal of Econometrics', 54

Ljung G.M., Box G.E.P. (1978), *On a Measure of Lack of Fit in Time Series Models*, 'Biometrika', 65

Maddala G.S. (2006), *Ekonometria [Econometrics]*, PWN, Warszawa

Woźniak R. (2009), *Pozycja konkurencyjna Kołobrzegu w świetle najnowszych badań empirycznych [The competitive position of Kołobrzeg in the light of recent empirical research]*, [in:] J. Chotkowski (ed.) *Marketing w rozwoju turystyki [Marketing in tourism development]*, Wydawnictwo Politechniki Koszalińskiej, Koszalin

Streszczenie

ANALIZA JAKOŚCI USŁUG TURYSTYCZNYCH W REGIONIE KOŁOBRZESKIM Z WYKORZYSTANIEM MODELI SZEREGÓW CZASOWYCH

Celem niniejszej pracy jest analiza jakości świadczonych usług w zależności od szeregu czynników: ceny pobytu w hotelu, jakości usług hotelarskich (czystość w pokoju, rejestracja pobytu, informowanie klientów itp.), jakości usług gastronomicznych (oferty gastronomiczne), ilości i jakości usług sanatoryjno-rehabilitacyjnych z wykorzystaniem metod szeregów czasowych oraz prognozowania oceny jakości świadczonych usług.

Badania były przeprowadzone na podstawie ankietowania gości zagranicznych z 13 hoteli regionu kołobrzeskiego, razem powyżej 1400 ankiet w zależności od wieku i płci ankietowanych. Dla ustalenia prawidłowości statystycznych zastosowano różne modele ekonometryczne. Po pierwsze, ocenianie jakości pobytu w hotelach przez klientów było rozpatrywane w przestrzeni czasowej 2006-2009 jako proces stochastyczny. Ustalono, że procesy są niestacjonarne, dlatego badanie bylo przeprowadzone na podstawie modelu ARIMA(1,1,1). Badania i prognozy pozwalają zrobić wniosek, że modele szeregów czasowych mogą oszacować trend, występujący dla badanego wskaźnika, natomiast błąd prognozowania jest dość wysoki (do 30%). Podobne wyniki były otrzymane rownież podczas badania ocen jakości pobytu w hotelach w zależności od płci badanych. Zaznaczymy, że średnia wartość ocen mężczyzn jest wyższa niż średnia ocen kobiet.

Szeregi czasowe dla badanych zmiennych miały stopien zintegrowania I(1). Została ustalona znaczna więź integracyjna między oceną pobytu w hotelu oraz ilością odwiedzin wybranego hotelu, kiedy turyści nie po raz pierwszy odwiedzający hotel przeważnie wystawiają wysokie oceny jakości pobytu w hotelu. Wynikiem badania zależności długookresowej pobytu w hotelu od czasu trwania procedury zameldowania jest bardza wysoka współzależność oceny pobytu od oceny czasu trwania procedury zameldowania. Cena pobytu w hotelu negatywnie wpływa na ocenę pobytu w hotelu, natomiast usługi sanatoryjne i gastronomiczne odbierane są jako standard i nieznacznie wpływają na ocenę pobytu w hotelu.

Przeprowadzone badania wskazują, że metody szeregów czasowych mogą być skutecznie zastosowane w badaniu wskaźników turystycznych i ocenie jakości usług hotelarskich.