THE USE OF ARIMA MODELS FOR FORECASTING THE SUPPLY AND DEMAND INDICATORS FROM TOURISM SECTOR

University Assistant Ph.D. Student **Iulian CONDRATOV**
Faculty of Economics and Public Administration
"Stefan cel Mare" University of Suceava, Romania
iulianc@seap.usv.ro

University Assistant Ph.D. Student **Pavel STANCIU**
Faculty of Economics and Public Administration
"Stefan cel Mare" University of Suceava, Romania
pavels@seap.usv.ro

Abstract:
This paper presents the use of ARIMA models in making prognosis for the tourism industry. Within this analysis were taken variables that capture both the supply and the demand for tourist services. These variable data are related to the tourism sector from the County of Suceava – Romania and have been taken from TEMPO-Online data base-time series, managed by the National Institute of Statistics. The analysis results were able to highlight, even under the severe impact that the financial crisis had the last three years that the tourism sector of Suceava will continue to be on an ascendant trend in terms of supply and demand indicators.

Key words: ARIMA, demand, supply, tourism sector

JEL classification: L83, C10

INTRODUCTION

The tourism sector in Suceava County has benefited from a remarkable development in recent years. The analysis of the statistic data provided by Suceava County Department of Statistics reveals that at least the last ten years ascendant trends in most indicators are found and they quantify Suceava tourism market through two components: demand and supply.

In the analysis that we propose on tourism market in the Suceava County, we chose 3 features that capture both supply and demand for tourist services recorded within the profile market of Suceava County, as follows:

- Overnight stays in tourist structures;
- Number of accommodated tourists;
- Accommodation capacity.

The above features are tracked chronologically, the recording being made annually. Although tourism activity is an activity where a strong seasonal pattern can be found, by taking into account the analysis of annual data has been removed the influence of the climatic factors (changes of seasons) and also the influence it has on various tourist festivals throughout the year.

Thus, we will try to achieve an econometric model that includes the major economic factors that influence the 3 variables analysed (Table no.1).

Table no. 1. Evolution of indicators of interest from 1990 to 2011

Year	Overnight stays in tourist structures -OSTS-	Accommodated tourists -AT-	Accommodation capacity -AC-
1990	1050185	390163	6841
1991	1046696	372710	6478
1992	629000	308469	6511
1993	628308	240719	6300
1994	605590	235032	6235
1995	592808	235507	5654
1996	502705	206315	4979

Year	Overnight stays in tourist structures -OSTS-	Accommodated tourists -AT-	Accommodation capacity -AC-
1997	407446	163398	5455
1998	409859	149395	5277
1999	424337	153416	5416
2000	480298	153515	5269
2001	461095	151370	5034
2002	406651	162423	5192
2003	423208	162473	5577
2004	432448	187412	5755
2005	435199	192120	6526
2006	500302	211003	7012
2007	535078	226277	6831
2008	530110	229068	7029
2009	479402	209725	7554
2010	460637	194365	8033
2011	556249	229519	8843

Source: (The National Institute of Statistics) www.insse.ro

1. DETERMINATION OF ARIMA MODELS FOR PROGNOSIS

Box & Jenkins (1970) have proposed a methodology to forecast a variable using as a database only its past and present. These models are very popular due to:

- The quality of the generated forecasts;
- The flexibility of the models;
- The rigor of the mathematical base of the model;
- The fact that this is an appropriate method for predicting variables with an irregular trend, too.

An autoregressive-moving average ARMA (*p,q*) model has an autoregressive type component or an average moving type component:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + ... + a_p Y_{t-p} - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - ... - b_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

Where p is the order of the autoregressive part, q is the moving average order and ε_t is a white noise type process (a sequence of independent and identically distributed random variables with zero mean).

To elaborate such models the following reasons are advised:

- The evolution of the economic phenomena is fuelled by the existing resources, the already created capacities, the experience and the tradition. Variables in economy have an inertial character, a strong autoregressive component being presented (mainly in macroeconomic indicators);
- Moving average type component is the effect of unpredictable events on the variable, effects gradually assimilated in time. This component is justified by the intervention of sudden unexpected changes among the external factors correlated to the variable. For example, strikes effect in Greece during the last two years led to the deterioration of the tourism activity indicators of the country. The moving average part captures the gradual assimilation of shocks (accidental deviations) from outside the system.

The ARMA models are suitable for stationary series. These were generalized for non-stationary series that become stationary by differentiation; the resulting models are called autoregressive-integrated-moving average ARIMA (*p,d,q*) where d is the order of differentiation required for stationary series.

Stages (the methodology) to develop an ARIMA (*p,d,q*) model are:

- Identifying the model → specifying the appropriate values for p, d or q;
- Estimation of model parameters → estimating the coefficients a_i, b_i;

- Testing the validity of the model. If the model is not valid then the model is once more specified (other plausible values for *p,d,q*) and the previous stages are repeated;
- Using the model to generate predictions (after it passed the validation tests).

The analysis was performed using SPSS v.17 statistical analysis software.
There was an attempt for retrieving an ARIMA (*p,d,q*) model to forecast the values for the future periods for the 3 variables (overnight stays in tourist structures, accommodated tourists, accommodation capacity).
While processing time series through Box-Jenkins strategy, the identification of the 3 parameters represents one of the important stages of the approach. Relevant information to determine p and q parameters are obtained by calculating the autocorrelation function values - ACF and partial autocorrelation function - PACF, and by drawing the corresponding graphs of the two functions.
The first step in building the model is to make graphics for data series.

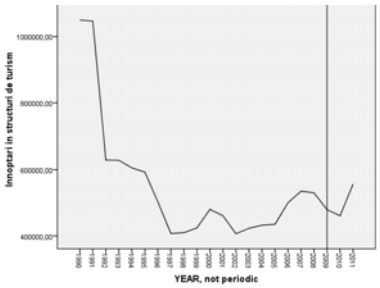

Figure no.1. The evolution of overnight stays in tourist structures indicator during 1990- 2011

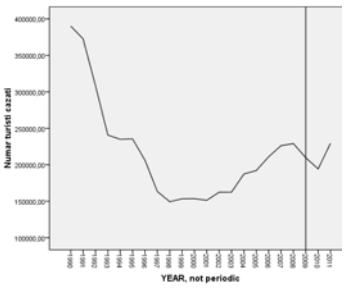

Figure no.2. The evolution of the number of accommodated tourists indicator during 1990-2011

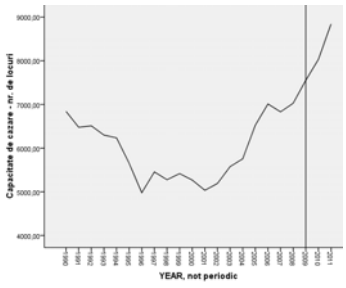

Figure no.3. The evolution of accommodated capacity indicator (no. of places) during 1990-2011

The features 'graphics cannot present a seasonal manifestation.

The decline of Suceava tourism sector that followed after 1990 is obvious, the most severe reduction of the overnight stays in tourist structures variable was during 1991-1997 – due to the reduction in the number of tourists and the capacity of accommodation.

Signs of revival can be found starting from 2001-2002, but, apart from accommodation capacity, the recorded levels at the end of 2011 for the other two features (overnight stays in tourist structures, number of accommodated tourists) are still far from those recorded at the end of 1990.

In the first part of the analysis we will attempt to retrieve candidate models suitable for considered forecasts.

The trend component can be observed in all 3 graphics, indicating the necessity for differentiate the data series. For clearer evidence of the stationary feature and of the auto-regressive component of data series it is essential to do the analysis for both the autocorrelation and the partial autocorrelation of the values of each series.

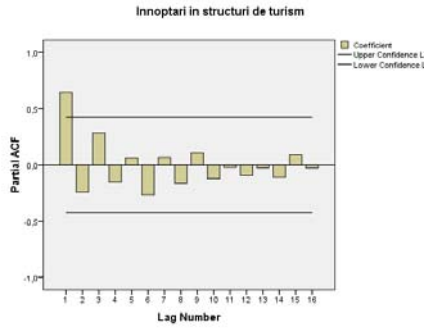

Figure no.4. Partial autocorrelation function chart for the overnight stays in tourism structures variable

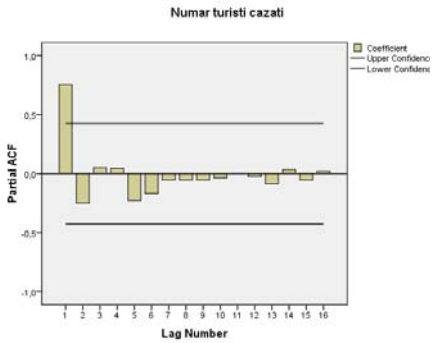

Figure no.5. Partial autocorrelation function chart for the number of accommodated tourists variable

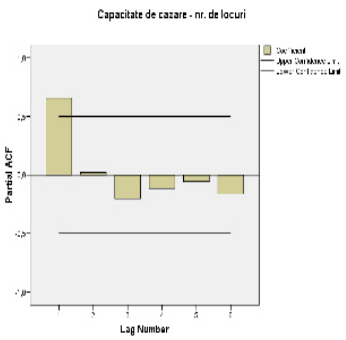

Figure no.6. Partial autocorrelation function chart for the accommodation capacity variable *(no. of places)*

One can easily see within the associated graphics that all 3 variables indicate a non-stationary state of the values as for a *lag* of order 1 the partial autocorrelation function achieves breakthroughs of the statistical significance. Thus, a differentiation of order 1 could resolve the non-stationary state problem.
The differentiation of order 1 leads to the following results regarding partial autocorrelation graphs:

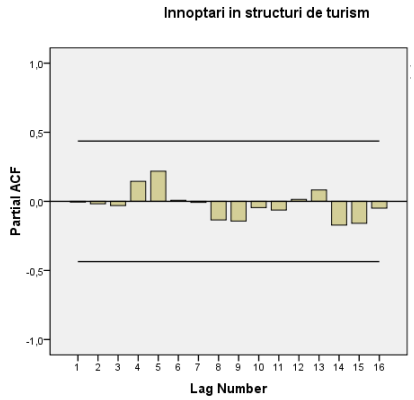

Figure no.7. Partial autocorrelation function chart for the overnight stays in tourism structures variable – *differentiation of order 1*

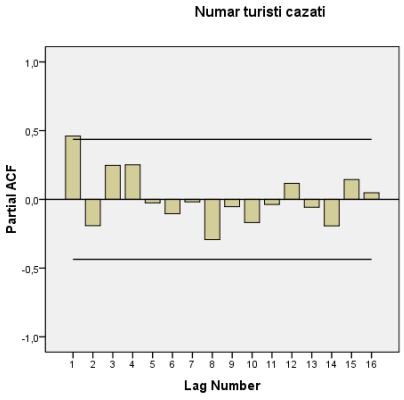

Figure no.8. Partial autocorrelation function chart for the number of accommodated tourists variable – *differentiation of order 1*

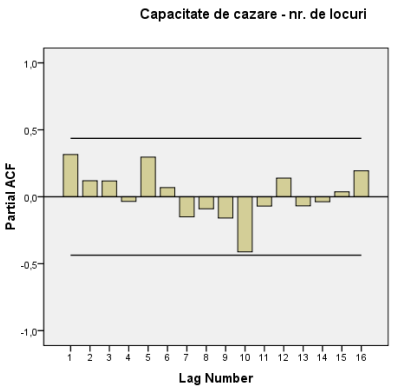

Figure no.9. Partial autocorrelation function chart for the accommodation capacity variable – *differentiation of order 1*

There is no partial autocorrelation coefficient above the critical level of significance; that indicates the absence of non-stationary data series and the necessity of change, implying the differentiation of order 1 for these variables when applying a regression analysis.

However, there is no fixed algorithm in establishing an accurate ARIMA (*p,d,q*) model optimally; it is necessary to complete an iterative process through which the 3 parameters will take different values, then will be applied a criterion for candidate models found .

Thus, by refining different models using the stationary R^2 criterion we chose the following adjustment models for data series related to the 3 variables:

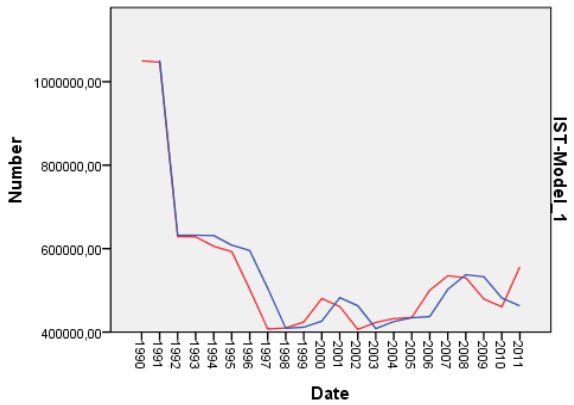

Figure no.10. The ARIMA model chart for the overnight stays in tourist structures variable

For the overnight stays in tourist structures variable the ARIMA (*0,1,0*) model was found. Diagnosing an ARIMA candidate model is a crucial stage of the construction process of the final model that involves the verification of random distribution of residues.

First, by analysing the stationary R^2 indicator (see Table no.2) we observe that the built model explains 57% of the variation in the series. Also, the Ljung-Box test is not statistically significant, so the model can be considered viable.

Table no.2. Diagnosing the ARIMA model for the overnight stays in tourist structures variable

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Overnight stays in tourist structures-Model_1	0	.571	22.664	18	.204	2

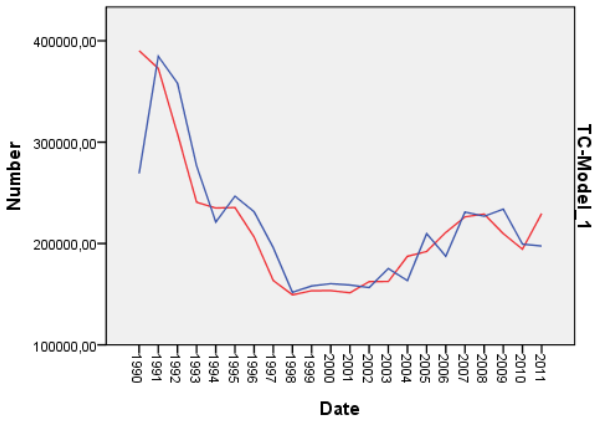

Figure no.11. The ARIMA model for the number of accommodated tourists variable

For the number of accommodated tourists variable the best was the ARIMA (*1,0,1*) model. Analysing the stationary R^2 indicator we find that the built model explains 74% of the variation in the series. Also, the Ljung-Box test is not statistically significant, so the model can be considered viable.

Table no.3. Diagnosing the ARIMA model for the number of accommodated tourists' variable

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Number of accommodated tourists-Model_1	0	.740	9.090	16	.910	0

Figure no.12. The ARIMA model for the accommodation capacity variable

For the accommodation capacity variable the best was the ARIMA (*0,2,0*) model. Regarding the accommodation capacity variable the model explains 48% of the variation in the series.
Even if the Box-Ljung test is not statistically significant confirming the fact that residues are randomly distributed, at the same time it provides evidence of building an accurate model.

Table no.4. Diagnosing the ARIMA model for the accommodation capacity variable

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Accommodation capacity - no. of places-Model_1	0	.483	13.968	18	.731	1

2. INTRODUCTION TO THE PREDICTIVE MODELS OF THE INDEPENDENT VARIABLES

Although, as we determined, the retrieved models succeeded to explain in a relatively large proportion the variation found in the observed data series, their ability to predict could be improved by introducing into the econometric model, of independent variables correlated with the dependant variables of interest.

Therefore, we chose to enter the following variables:

- Civilian employees in the tourism sector;
- Share of the turnover of companies in the tourism sector of total small and medium enterprises;
- Gross domestic product;
- Net investments in tourism sector;
- Real earnings.

Given the need for the calculation of values expressed in comparable prices, we chose the variables: GDP, net investments in tourism sector and real earnings, using the calculated indices with comparable values.

Table no.5. Chosen indicators as independent variables to increase the prediction power of ARIMA models

Year	Civilian employees – hotels and restaurants – thousands of people	Share of the turnover of companies of active small and medium enterprises in the tourism sector/total	Indices of real earnings – 1990=100%	Indices of net investments – 1990=100 %	Indices of GDP – 1990=100%
1990	186	-	100	100	100
1991	213	-	81.5	114.7	74.1
1992	175	65.3	70.8	196.1	68.7
1993	131	69.1	58.9	140.2	54.4
1994	136	77	59.1	172.7	43.5
1995	123	79.9	66.5	308.8	59.2
1996	116	73.3	72.7	221.4	81.3
1997	130	69.9	56.2	170.9	67.9
1998	98	71.9	58.4	207.3	61.6
1999	100	73.4	57	237.8	96.6
2000	93	74.7	59.4	153.9	94.2
2001	79	74	62.4	162.5	95.1
2002	95	76.1	63.9	162	102.3
2003	105	79.1	70.8	197.8	104.4
2004	133	77.8	78.3	265.4	111
2005	133	78.8	89.5	300.7	120.4
2006	136	79.3	97.4	359.6	128.3
2007	142	81.5	111.8	416.1	136.4
2008	162	83.8	130.3	598.2	146.4
2009	125	84.6	128.3	362.2	136
2010	*133*	*84.3*	*123.6*	*403.4*	*134.9*

Source: (The National Institute of Statistics) www.insse.ro

The study of the correlation matrix of variables (Appendix no. 1) indicates statistically significant correlations between them. The share of turnover achieved by active small and medium enterprises/total, Indices of the real earnings - 1990=100%, Indices of net investments in the tourist sector - 1990=100% and Indices of GDP - 1990=100%. Thus, to avoid the phenomenon of multicollinearity, of the 5 proposed variables were retained only *Civilian employees -in the tourism sector – and Indices of GDP - 1990=100%.*

The further analysis builds on the results of univariate ARIMA models found in the previous stage. Even though we have these landmarks, the iterative process of finding the parameters (*p,d,q*) will continue to capture the optimal prediction model.

The model for - ***Overnight stays in tourist structures variable*** was ARIMA (*4,2,1*), in which, after the iterative process of modifying the parameters, we kept only the independent variable *Civilian employees - the tourism sector*. This variable manages to estimate a proportion of 81.7% of variation recorded during 1990-2011 by the dependent variable.

Table no.6. Diagnosing the ARIMA model for the overnight stays in tourist structures variable

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Overnight stays in tourist structures-Model_1	1	.817	13.447	13	.414	1

Table no.7. Description of the ARIMA model for the overnight stays in tourist structures variable

			Estimated	SE	t	Sig.
Overnight stays in tourist structures	*Constant*		41729.370	71891.100	.580	.572
	AR	Lag 1	.554	.731	.758	.463
		Lag 2	-.907	.356	-2.546	.026
		Lag 3	.244	.538	.453	.659
		Lag 4	-.082	.371	-.222	.828
	Difference		2			
	MA	Lag 1	.984	8,175	.120	.906
Civilians employed - the tourist sector	*Numerator*	Lag 0	-299.000	610.842	-.489	.633

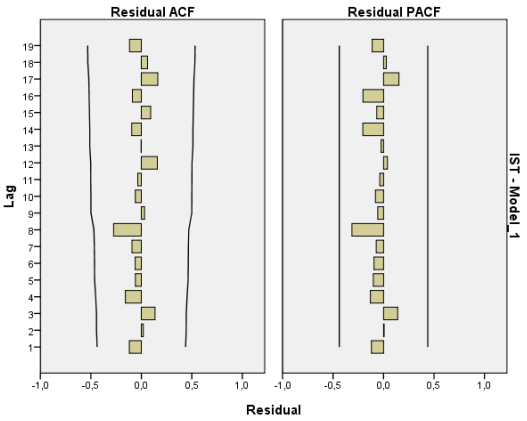

Figure no.13. Autocorrelation function chart and partial autocorrelation for the overnight stays in tourist structures variable

In diagnosing the model we notice the absence of autocorrelation of residues.

Although because of the reduced volume of the analysed sample the estimations of the parameters for econometric models lose their significance, we are able to specify, through the only statistically representative parameter that a one unit increase in the value of the difference between the number of nights spent in tourist structures two years ago, will reduce the current value of this feature by 0.907 units.

Unfortunately for - ***The number of accommodated tourists variable,*** the new introduced variables failed to improve the model, so the initial model named ARIMA (*1,0,1*) will be kept with no predictors.

If the first 2 variables targeted the demand for tourist services, the latter introduced in the analysis concerns the supply in tourist services, namely the evolution of ***the accommodation capacity***.

In order to maximize the stationary R^2 criterion, we chose the ARIMA (*4,0,5*) model and we introduced in the econometric modelling only the independent variable *Civilian employees – in tourist sector*.

Table 8 – Diagnosing the ARIMA model for the accommodation capacity variable

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Accommodation capacity – no. of places-Model_1	1	.920	14.375	9	.110	0

Table no.9. Description of the ARIMA model for the accommodation capacity variable

			Estimate	SE	t	Sig.
Accommo-dation capacity *(no. of places)*	*Constant*		7483.698	1230.696	6.081	.000
	AR	Lag 1	1.229	050	24.595	.000
		Lag 2	.237	.004	53.577	.000
		Lag 3	-.327	.021	-15.881	.000
		Lag 4	-.208	.006	-32.850	.000
	MA	Lag 1	.572	4.417	.130	.899
		Lag 2	.280	2.455	.114	.911
		Lag 3	.580	3.229	.180	.861
		Lag 4	.238	2.432	.098	.924
		Lag 5	-.825	4.132	-.200	.845
Predictor 1	*Numerator*	Lag 0	-3.195	5.054	-.632	.540

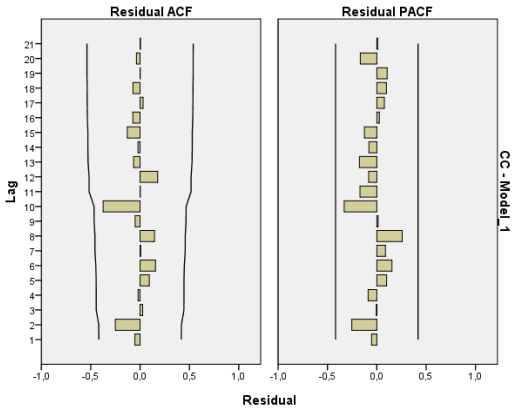

Figure no.14. Autocorrelation function chart and partial autocorrelation for the accommodation capacity variable

The model explains 92% of the total variation of the accommodation capacity during the analysed period. Through the parameters statistically representative, we can specify the following:

- If they could cancel the influence of other factors, the value of the dependent variable would be 7484 accommodation places;
- A one unit up of the value of the difference of accommodation capacity 1 year ago would increase the current value of the feature by 1.229 units. To increase by one unit the next years, the current value of the feature would follow a descendant trend (from 0.237 to -0.208).

Consequently, for two of the analysed variables, one characterizing the market demand, the other characterizing the existing offer, they managed to improve the model by introducing an independent variable.

Before making forecasts on variables of interest, it is recommended to compare the univariate ARIMA models found in the first part of the study and the predictor models identified above to choose the final model and after that to forecast the values of the variables of interest for the next 2 years.

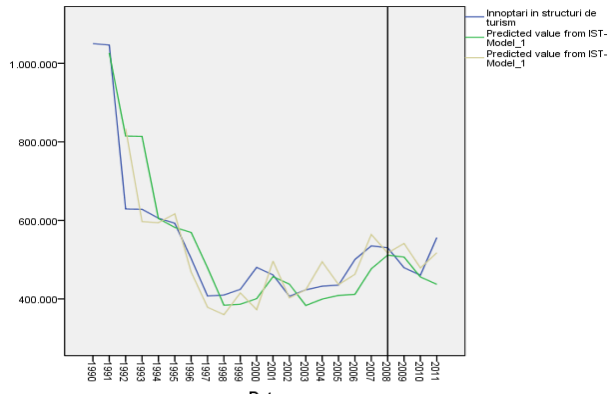

Figure no.15. The chart of ARIMA models built for the overnight stays in tourist structures variable

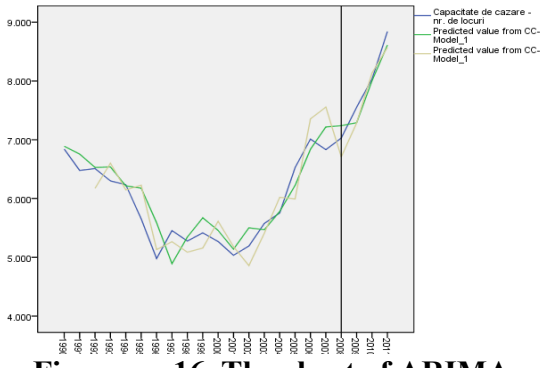

Figure no.16. The chart of ARIMA models built for the accommodation capacity variable

It can be seen clearly the fact that the predictor models are better in estimating the recorded variation in data series during the analysed period. Moreover, if we analyse the models' behaviour over the last three years, when the effects of the economic crisis occurred, we find that the predictor models perform better, the unvariate model related to the overnight stays in tourist structures indicates an inaccurate evolution.

Thus, to determine the expected values of the 3 analysed variables, the following models will be used:

- For the variable *Overnight stays in tourist structures* -the ARIMA (*4,2,1*) model with the predictor civilian employees in tourist sector;
- For the variable *Number of accommodated tourists* - the univariate ARIMA(*1,0,1*) model;
- For the variable *Accommodation capacity* - the ARIMA (*4,0,5*) model with the predictor civilian employees in tourist sector.

After applying these models, the predicted values for the 3 variables of interest, during 2012-2013, are as follows:

Table no.10. Predicted values for the next 2 years for the 3 analysed variables

		2012	2013
Overnight stays in tourist structures	Forecast	642752.74	637531.55
	UCL	767277.65	869867.21
	LCL	518227.83	399195.89
Number of accommodated tourists	Forecast	253331.54	254686.50
	UCL	301133.32	302981.81
	LCL	205529.75	166383.18
Accommodation capacity *(no. of places)*	Forecast	8828.14	8941.22
	UCL	9576.31	9856.84
	LCL	8079.97	8025.60

CONCLUSIONS

All 3 indicators of demand and supply in tourism services present increased predicted values and we assume that the development of Suceava tourist sector will continue during the next years. According to the values determined through the analysis, an increase with 86504 of the number of overnight stays in tourist structures is estimated with a probability of 95%, for 2012, compared to 2011. Regarding the *accommodated tourists,* their number will increase in 2012 compared to 2011 with 23813 tourists.

If the first two variables are expected to increase significantly in 2012, for the variable accommodation capacity *(no. of places)* stagnation is estimated in 2012, an increase with approximately 100 accommodation places is expected to be registered in 2013.

It is necessary however, that this trend be supported by a proper market strategy, by increasing the quality of tourism services and, perhaps, most importantly, by developing the road infrastructure which will allow an increased number of tourists, given that the most used means of transportation for people that arrive in Bucovina zone is the automobile.

REFERENCES

1. Andrei, T.S. (2003), *Statistică. Teorie şi aplicaţii* (ed. Ediţia a II-a), Bucureşti, Editura Economică.

2. Alleyne, D. (2006), *Can Seasonal Unit Root Testing Improve the Forecasting Accuracy of Tourist Arrivals?* Tourism Economics, 12, pp.45-64.

3. Brannas, K., Hellstrom, J., Nordstrom, J. (2002), *A new approach to modelling and forecasting monthly guest nights in hotels.* International Journal of Forecasting, Vol. 18, pp.19-30.
4. Divisekera, S. (2003), *A Model of Demand for International Tourism*. Annals of Tourism Research, 30, pp.31-49.
5. du Preez J., Witt, S.F. (2003), *Univariate versus multivariate time series forecasting: an application to international tourism demand*. International Journal of Forecasting, Vol. 19, pp.435-451.
6. Hapenciuc, C.V. (2003), *Cercetare statistică în turism – Studiul fenomenului turistic în județul Suceava,* Bucureşti, Editura Didactică şi Pedagogică.
7. Hapenciuc, C.V. (2004), *Elemente de analiză şi prognoză în turism, Studiu de caz – județul Suceava.* Iaşi, Editura Junimea.
8. Kulendran, N., Witt S.F. (2003), *Forecasting the demand for international business tourism,* Journal of Travel Research, 41, pp.265–271.
9. Lee, C.C., Chang, C.P. (2008), *Tourism Development and Economic Growth: A Closer Look at Panels*, Tourism Management 29, pp.180-192.
10. Louca, C. (2006), *Income and Expenditure in the Tourism Industry: Time Series Evidence from Cyprus*, Tourism Economics 12, pp.603-617.
11. Li, G., Song H. Witt S.F. (2005), *Recent developments in econometric modelling and forecasting*, Journal of Travel Research, 44, pp.82-99.
12. Oh, C., Morzuch B.J. (2005), *Evaluating Time-Series Models to Forecast the Demand for Tourism in Singapore*. Journal of Travel Research, 43, pp.404-413.
13. Popescu, T. (2000), *Serii de timp. Aplicații în analiza sistemelor.* Bucureşti: Editura Tehnică.
14. Proença, S., Soukiazis, E. (2005), *Tourism as an Alternative Source of Regional Growth in Portugal: A Panel Data Analysis at NUTS II and III Levels*, Portuguese Economic Journal, 6, pp.121-135.
15. Wong, K.K.F., Song, H., Witt, S.F., Wu, D.C. (2007), *Tourism Forecasting: To Combine or Not to Combine?* Tourism Management, 28, pp.1068-1078.
16. *** Institutul National de Statistică, *TEMPO-Online - serii de timp*, 2012, https://statistici.insse.ro/shop/.

APPENDIX 1

Table The Matrix of Correlations

		Civilian employees – tourist sector	Share of the turnover of companies of active small and medium enterprises in the tourism sector/total	Indices of real earnings - 1990=100%	Indices of net investments in tourist sector- 1990=100%	Indices of GDP - 1990=100%
Civilian employees – tourist sector	*Pearson Correlation*	1	-.031	.413	.067	.019
	Sig. (2-tailed)		.895	.063	.772	.935
	N	21	21	21	21	21
Share of the turnover of companies of active small and medium enterprises in the tourism sector/total	*Pearson Correlation*	-.031	1	.747**	.739**	.733**
	Sig. (2-tailed)	.895		.000	.000	.000
	N	21	21	21	21	21
Indices of real earnings - 1990=100%	*Pearson Correlation*	.413	.747**	1	.759**	.816**
	Sig. (2-tailed)	.063	.000		.000	.000
	N	21	21	21	21	21
Indices of net investments in tourist sector - 1990=100%	*Pearson Correlation*	.067	.739**	.759**	1	.714**
	Sig. (2-tailed)	.772	.000	.000		.000
	N	21	21	21	21	21
Indices of gross domestic product - 1990=100%	*Pearson Correlation*	.019	.733**	.816**	.714**	1
	Sig. (2-tailed)	.935	.000	.000	.000	
	N	21	21	21	21	21

** - Correlation is significant at the 0.01 level (2-tailed)