ON MODELLING OF DEVELOPMENT OF TOURISM IN SELECTED EUROPEAN COUNTRIES

Eva Litavcová – Petra Vašaničová

Abstract

Understanding tourism development over time requires analysing time series of available data and assessing relevant information and essential tourism indicators. The aim of this paper is modelling of the tourism development of the 24 European Union countries using Box-Jenkins methodology to capture and explain the patterns and the determinants of tourism in the European Union countries. The variable "Total Nights Spent at Tourist Accommodation Establishments" per month was considered, and was recorded for the period from January 2002 to September 2018 by the Eurostat database, since this is one of the variables that best expresses effective demand. Results of founded 24 SARIMA models reported in the overview table can be considered as preliminary analysis for next examining using ARFIMA models with long memory, artificial neural networks, or models based on Engle's methodology. For the best six models, ex-ante analysis of accuracy is done. The best model among them is the model for Germany.

Key words: Tourism, development, European Union, ARIMA models, Box-Jenkins methodology

JEL Code: C32, L83, Z32

Introduction (Times New Roman, 14 pt., bold)

Accurate tourism demand modelling and forecasting can help tourism businesses to establish effective marketing and investment plans and the government to formulate appropriate policies. According to Nor et al. (2018), tourism has become one of the important industries that contributes to the country's economy. Tourism demand forecasting gives valuable information to policy makers, decision makers and organizations related to tourism industry in order to make crucial decision and planning. However, it is challenging to produce an accurate forecast since economic data such as the tourism data is affected by social, economic and environmental factors. Nowadays, tourism is becoming more vulnerable and more responsive to economic challenges, lack of security and safety.

An important characteristic of tourism is seasonality. Forecasting tourism development, respectively tourism demand is realized by the traditional most widely used methods of time series examination, specifically by Exponential Smoothing and Box-Jenkins methodology (Vašaničová, Litavcová, Jenčová, 2017). Box-Jenkins method is an appropriate for a medium to a long (at least 50) time series data observation. When modelling a medium to a long time series, the difficulty arose in choosing the accurate order of model identification level and in discovering the right parameter estimation. Method is widely used in the context of tourism, where the lengths of time series use to be from medium to long. In the recent study of Perles-Ribes et al. (2019), there was analysed the immediate impact that the instability associated to the political situation in Catalonia has had on the arrivals and spending of international tourists in the region using the classical Box-Jenkins method (ARIMA) and the more recent Bayesian Structural Time-Series Models. Petruška (2018) explored arrivals at tourist accommodation establishments from the time series periodicity point of view. Using spectral analysis, he has shown that the most significant wavelengths are 6 and 12 months. According to the dominant frequency, he has divided European countries into four groups, and thus identified four patterns of behaviour.

In this paper, we analyse tourism data, specifically time series of total nights spent at tourist accommodation establishments with a length of 201 observations for 24 countries, and we use Box-Jenkins methodology. Results of founded SARIMA models reported in the overview table can be considered as preliminary analysis for next examining using ARFIMA models with long memory, artificial neural networks, or models based on Engle's methodology (Engle, 2003).

1 Indicators measuring tourism demand

We can argue that the need for and relevance of forecasting tourism demand has recently become a highly discussed topic. Understanding tourism development over time requires analysing time series of available data and assessing relevant information and essential tourism indicators in the context of competition. Providing uniform tourism indicators for several countries is of great importance in international comparisons.

Indicators measuring tourism demand are collected in various ways. For countries of the European Union (EU), Eurostat (2018) provides data on tourism. For example, on monthly period, it gives information about occupancy of tourist accommodation establishments, specifically arrivals and nights spent by residents and non-residents, and about net occupancy rate of bed-places and bedrooms in hotels and similar accommodation. Applying these indicators to international tourism research contributes to effective decision-making by tourism policy makers. From the broader perspective, these indicators provide valuable information in a process of reviewing the tourism “strategic management and also in national and international benchmarking” (Stefko, Gavurova, Korony, 2016, p. 177).

According to Eurostat (2018) “a night spent (or overnight stay) is each night a guest actually spends in a tourist accommodation establishment or non-rented accommodation. An arrival is defined as a tourist who arrives at tourist accommodation establishment and checks in or arrives at non-rented accommodation.” It is important to note that the arrivals of same-day visitors are excluded from accommodation statistics. Moreover, NACE Rev. 2 classification divides accommodation establishments in three groups; specifically, I551 – hotels and similar accommodation, I552 – holiday and other short-stay accommodation, I553 – camping grounds, recreational vehicle parks and trailer parks. Each indicator is specified for residents, non-residents, and in total.

In this paper, we use data of nights spent by residents and non-residents in total in the first group of NACE Rev. 2 classification, i.e., in hotels and similar accommodation.

2 Data and methodology

The aim of the paper is to model the development of total nights spent at tourist accommodation establishments among selected EU countries using SARIMA models. Obtained monthly data from the Eurostat database for 24 countries have monthly periodicity from January 2002 (2003) to September 2018. For six countries, specifically for Croatia, Estonia, Latvia, Malta, Poland, Slovakia, we have only data from 2003. There are no more missing observations for other countries. When analysing data, we apply Box-Jenkins methodology (Box et al., 1994) and model the development of time series for the next 6 months.

Box and Jenkins (1970) proposed methodology, which considers a random component that can be created by correlated random variables as an essential element of the time series model construction. As Taneja et al. (2016) further says, a simple equation to define the autoregressive moving average (ARMA)(p,q) model for a stationary time series is given below:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ... + \varphi_p Y_{t-p} + \varepsilon_t + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} \qquad (1)$$

The first term in ARIMA model represents an autoregressive (AR) term of the order *p* having the form of

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t \quad (2)$$

This (AR) term refers to the current time series values Y_t as a function of past time series values Y_{t-1}, Y_{t-2},..., Y_{t-p}. The φ_1, φ_2, φ_3 are autoregressive coefficients that relates Y_t to Y_{t-1}, …, Y_{t-p}. The moving average MA(q) term of the model is represented as,

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \quad (3)$$

where, $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q}$ are the past random shocks or independent white noise sequence with mean = 0 and variance = σ^2; $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients relating Y_t to $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q}$.

When the (AR) and (MA) specifications are combined together with integration (differencing) term, they constitute an ARIMA (p,d,q) model, where p, d and q indicate orders of autoregression, differencing and moving average. The model is mathematically given as

$$(1-B)^d Y_t = \frac{\theta(B)}{\varphi(B)} \varepsilon_t \quad (4)$$

where, *t* denotes the time indices, *B* is the backshift operator, ie., $BY_t = Y_{t-1}$. $\varphi(B)$ and $\theta(B)$ are the autoregressive and moving average operators respectively and can be written as

$$\varphi(B) = 1 - \varphi_1 B^1 - \varphi_2 B^2 - \ldots - \varphi_p B^p \quad (5)$$

$$\theta(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \ldots - \theta_q B^q \quad (6)$$

Seasonality is a pattern which is repeating itself over a fixed time interval. Here, the monthly dataset is presenting a seasonal period of 12 months. In order to obtain a stationarity, seasonal differencing is performed by taking difference between the present and corresponding observation from the previous year. Taking into consideration the seasonality of our time series, a seasonal ARIMA denoted as SARIMA (p, d, q) x $(P, D, Q)_s$, is used, where P,D,Q represent seasonal autoregressive, differencing and moving average orders respectively and *s* is number of seasons. For the present study, *s* = 12. SARIMA(p, d, q)$(P, D, Q)_s$ without constant built for the time series is defined as:

$$\varphi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (7)$$

where, *B* is the backshift or lag operator, *s* is the seasonal lag (in 'month' for present study); ε_t represents error variables; *d* and *D* are non-seasonal and seasonal differences; φ and Φ are the non-seasonal and seasonal autoregressive parameters; θ and Θ are the non-seasonal and seasonal moving average parameters; $\Phi_P\left(B^s\right)$ and $\Theta_Q\left(B^s\right)$ are seasonal parts of (AR) and (MA) specifications respectively. (more Arlt and Arltova, 2009).

2 Results

Table 1 states optimal SARMA models founded by IBM SPSS 20 software. The requirement for the normality of residuals has been met for eight countries, and they are marked in bold (Belgium, Denmark, Estonia, Germany, Hungary, Latvia, Lichtenstein, and Luxembourg).

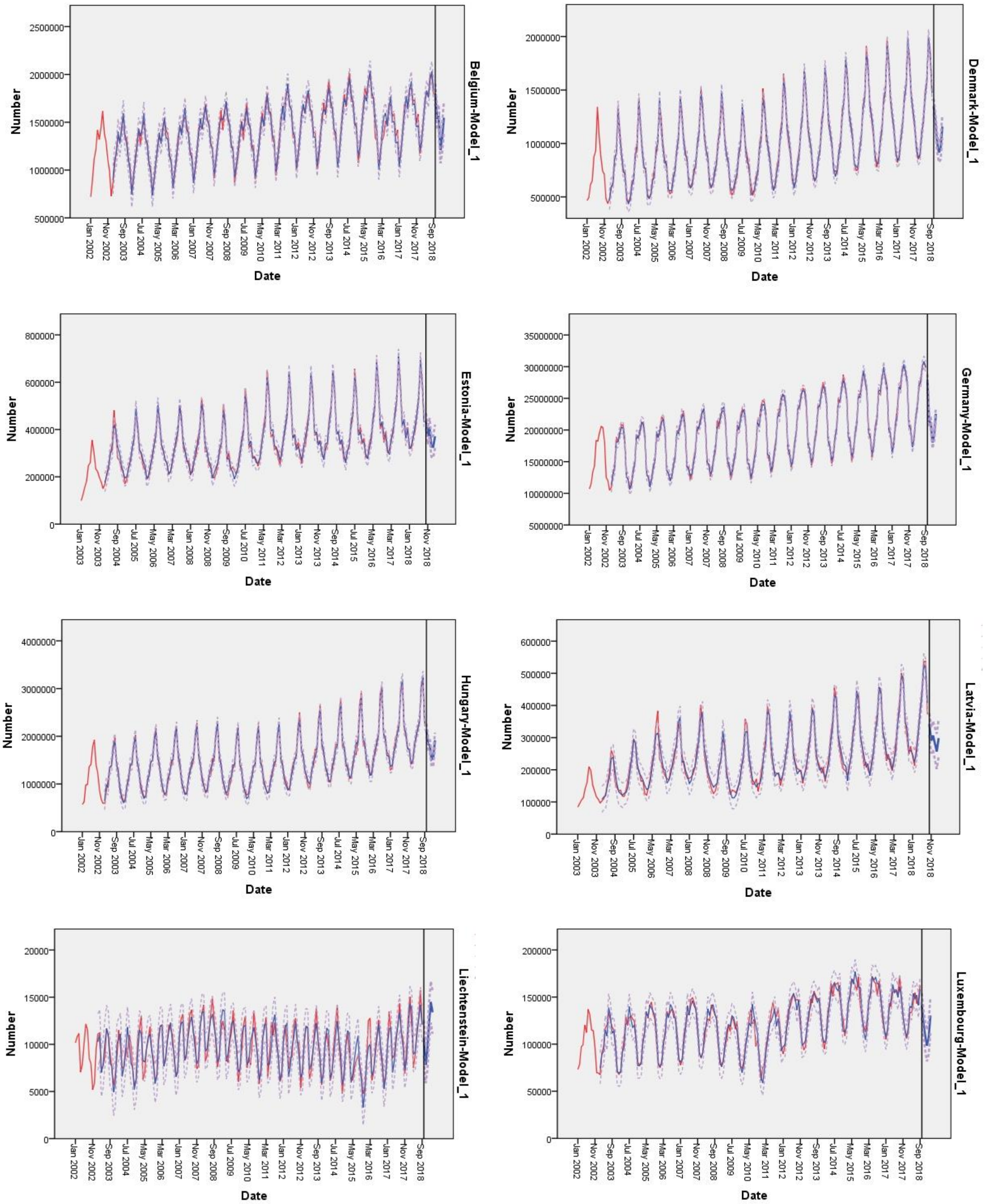
Tab. 1: Optimal SARIMA models for variable Total nights spent among 24 EU countries

Country	**ARIMA model**	**Stat. R^2**	**MAPE**	**Norm. BIC**	**LB Stat.**	**DF**	**LB Sig.**	**SW Sig.**
Austria	(1,0,0)(0,1,1)	0.438	3.991	25.745	31.756	16	*0.011*	*0.000*
Belgium	(0,1,1)(0,1,1)	0.404	2.996	21.837	24.620	16	0.077	0.092
Croatia	(0,0,11)(1,1,1)	0.395	8.254	23.415	18.293	13	0.147	*0.000*
Cyprus	(1,0,11)(0,1,1)	0.588	5.961	22.406	13.950	15	0.529	*0.001*
Czechia	(0,1,1)(0,1,1)	0.356	3.189	23.211	22.440	16	0.130	*0.003*
Denmark	(0,1,6)(0,1,1)	0.375	2.923	21.043	20.084	15	0.169	0.112
Estonia	(2,1,0)(0,1,1)	0.279	3.748	19.485	15.773	15	0.397	0.307
Finland	(0,1,5)(0,1,1)	0.406	2.234	21.269	10.740	15	0.771	*0.002*
France	(0,1,1)(0,1,1)	0.613	2.739	26.600	13.863	16	0.609	*0.008*
Germany	(0,1,1)(0,1,1)	0.469	1.772	26.004	21.769	16	0.151	0.583
Greece	(1,0,0)(2,1,1)	0.541	6.075	25.488	22.433	14	0.070	*0.000*
Hungary	(0,1,1)(0,1,1)	0.525	3.768	22.346	23.505	16	0.101	0.137
Italy	(0,1,1)(0,1,1)	0.557	3.324	27.338	22.778	16	0.120	*0.011*
Latvia	(0,1,1)(0,1,1)	0.341	6.002	19.570	16.838	16	0.396	0.067
Liechtenstein	(0,1,1)(0,1,1)	0.611	7.830	13.828	20.013	16	0.220	0.094
Luxembourg	(0,1,1)(0,1,1)	0.335	4.104	17.629	24.170	16	0.086	0.075
Malta	(1,0,0)(0,1,1)	0.683	3.429	20.587	26.547	16	*0.047*	*0.002*
Netherlands	(0,1,1)(0,1,1)	0.506	2.722	23.147	27.663	16	*0.035*	*0.040*
Poland	(1,0,1)(0,1,0)	0.623	2.686	22.863	13.356	16	0.647	*0.001*
Portugal	(0,1,11)(1,1,0)	0.512	4.181	24.194	26.795	15	*0.030*	*0.000*
Romania	(1,0,2)(0,1,1)	0.536	5.246	23.264	13.095	15	0.595	*0.000*
Slovakia	(0,1,0)(0,1,1)	0.127	3.711	20.675	23.656	17	0.129	*0.022*
Spain	(2,1,11)(0,1,1)	0.583	3.433	27.538	19.264	14	0.155	*0.023*
Sweden	(1,0,1)(0,1,1)	0.266	2.592	22.842	30.180	15	*0.011*	*0.000*

Source: own processing in IBM SPSS 20.

Note: Stat. R^2 denotes stationary R-squared, Norm. BIC denotes normalized Bayesian information criterion, LB Stat. denotes Ljung-Box statistics, DF denotes degrees of freedom, LB Sig. denotes significance of Ljung-Box test, while we consider significance level of 0.05, SW Sig. denotes significance of Shapiro-Wilk test.

Fig. 1: Observed (red) and predicted (blue) time series of 8 chosen countries with 95% confidence limits (blue dashed)



Source: own processing

Other 16 countries have not been met the requirement for the normality of residuals, what we demonstrated by using a Shapiro-Wilk test in Table 1. Four of them have a systematic component in residuals, what we confirmed by using a Ljung-Box test at a significance level of 0.05 (significance is marked in italic).

Tab. 2: Forecasts of 8 selected countries and verification of accuracy of the models

	Oct 2018	Nov 2018	Dec 2018	Jan 2019	Feb 2019	Mar 2019	UI	UII
Belgium o.	1784699	1598986	1640866				0.018	0.035
Forecast	1800499	1546263	1556151	1211379	1329864	1547637		
UCL	1904700	1663658	1685400	1351482	1480038	1707247		
LCL	1696298	1428868	1426903	1071277	1179691	1388027		
Denmark o.	1357430	1225760	986990				0.009	0.017
Forecast	1349673	1198712	1008211	917984	947408	1158145		
UCL	1419456	1277153	1094446	1011365	1047424	1264384		
LCL	1279889	1120270	921976	824604	847392	1051907		
Estonia o.	430937	362447	417085				0.010	0.019
Forecast	436831	372022	409842	327159	326660	371357		
UCL	468624	407116	448056	370127	372930	420734		
LCL	405038	336929	371628	284190	280390	321979		
Germany o.	27748119	21717558	21336784				0.005	0.011
Forecast	27957386	21396466	21109547	18029993	19207855	22551678		
UCL	28802849	22272756	22015615	18964892	20170721	23541722		
LCL	27111923	20520176	20203479	17095095	18244989	21561635		
Hungary o.	2146736	1816378	1764292				0.008	0.016
Forecast	2129506	1836649	1808769	1513345	1544112	1916580		
UCL	2264946	1976811	1953498	1662502	1697569	2074219		
LCL	1994065	1696487	1664040	1364189	1390656	1758940		
Latvia o.	336138	281250	294216				0.015	0.031
Forecast	342347	290914	306114	281144	257393	298585		
UCL	376080	330983	351646	331551	312243	357544		
LCL	308614	250844	260581	230737	202544	239627		
Liechtenstein o.	10365	7733	8794				0.016	0.033
Forecast	10677	7828	9196	12361	14480	13391		
UCL	12542	9782	11235	14480	16679	15665		
LCL	8812	5875	7158	10241	12282	11117		
Luxembourg o.	144388	115954	108896				0.015	0.030
Forecast	141287	121579	108278	98429	106647	130461		
UCL	154110	135791	123755	115075	124385	149227		
LCL	128464	107368	92801	81783	88910	111695		

Source: own processing.

Note: UCL denotes upper, and LCL denotes lower 95% confidence limits of forecast. Abbreviation o. means observed. UI denotes Theil's inequality coefficient and UII is Theil's modified inequality coefficients (both computed ex-ante on the base of three months with known current data).

The time series of the analysed countries indicate a growing trend in most of them with a slight decline or stagnation in the time of crisis. We see four behaviour patterns in the graphs of the examined time series, in accordance to spectral analysis of Petruška (2018). In Figure 1, we present only graphs of SARIMA models of eight countries that we consider the most appropriate compared with those of other countries.

To measure the forecast accuracy we use Theil's inequality coefficient *UI* (in modified form as *UII*), which is given according to Bliemel (1973). These coefficients is from the interval $\langle 0,1 \rangle$, where *UI* = 0 means case of equality and *UI* = 1 means maximum inequality. In Table 2 are forecasts of selected 8 countries with confidence bounds for the next six months and computed Theil's inequality coefficients *UI* and *UII* ex-ante, on the base of three months with known current data. Table 2 shows that no observed value has gone beyond the confidence limits of forecasts. Moreover, the best ex-ante estimation of reality was for Germany, Hungary, Denmark and Estonia (in order).

Conclusion

In this article the variable "Total Nights Spent at Tourist Accommodation Establishments" per month of length 201 months was analysed for 24 countries, since this is one of the variables that best expresses effective tourism demand. Box-Jenkins methodology was used to model development in time series. Of the 24 SARIMA models found, 6 were those that met the required assumptions. For these models, we made the analysis of their accuracy (Table 2), and it turned out that models are convenient. Results of all founded SARIMA models reported in the overview Table 1 can be considered as preliminary analysis for next examining using ARFIMA models with long memory, artificial neural networks, or models based on Engle's methodology.

The tourism sector is considered to be a complex system of relationships (Lyócsa, Vašaničová, Litavcová, 2019). Forecasting is an essential prerequisite and desirable determinant of tourism activities. With a forecasted trend and a concrete model for future tourism demand, the government can create a well-organized strategy, allocate sufficient financial resources and provide visitors with better infrastructure; in the private sector, an appropriate marketing strategy can be developed to gain the benefits from growing tourism, set performance targets, effectively managed offer and pricing. (Hirashima et al., 2017; Liang, 2014; Hadavandi et al., 2011). The conclusions of our research may be useful for a deeper understanding of the tourism industry in the analyzed countries. Accurate forecasting of tourism demand patterns can be valuable also for tourism-related industries in formulating useful and efficient strategies to maintain and promote tourism.

Acknowledgment

This contribution was supported by research grant VEGA No. 1/0470/18 “Economic activity of tourism in the European area” and GAPU 43/2019 “Cultural difference and their impact on tourism”. This work was supported by the Slovak Research and Development Agency under the contract no. APVV-17-0166.

References

Arlt, J., & Arltová, M. (2009). *Ekonomické časové řady*. Praha: Proffesional Publishing.

Bliemel. F. (1973). Theil's forecast accuracy coefficient: A clarification. *Journal of Marketing Research*, 10, 444-446.

Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

Box, G. E. P., Jenkins, G. M., & Reinsel, G.C. 1994. *Time Series Analysis: Forecasting and Control*. Englewood Cliff: Prentice-Hall.

Engle, R.F. (2003). *Risk and volatility: Econometric models and financial practice. Nobel Lecture*. New York: New York University, Department of Finance.

Eurostat. (2018). *Occupancy of tourist accommodation establishments: Reference Metadata*. Retrieved from https://ec.europa.eu/eurostat/cache/metadata/en/tour_occ_esms.htm

Hadavandi, E., Ghanbari, A., Shahanaghi, K., & Abbasian-Naghneh, S. (2011). Tourist arrival forecasting by evolutionary fuzzy systems. *Tourism Management*, 32(5), 1196-1203.

Hirashima, A., Jones, J., Bonham, C. S., & Fuleky, P. (2017). Forecasting in a mixed up world: nowcasting Hawaii tourism. *Annals of Tourism Research*, 63, 191-202.Liang, Y. H. (2014). Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan. *Computers & Industrial Engineering*, 74, 111-119.

Lyócsa, Š., Vašaničová, P., & Litavcová, E. (2019). Interconnectedness of international tourism demand in Europe: A cross-quantilogram network approach. *Physica A: Statistical Mechanics and its Applications*, 526, 1-13.

Nor, M. E., Nurul, A. I. M., & Rusiman, M. S. (2018). A Hybrid Approach on Tourism Demand Forecasting. In *Journal of Physics Conference Series*, 995(1), 1-11.

Perles-Ribes, J. F., Ramon-Rodriguez, A. B., Such-Devesa, M. J., & Moreno-Izquierdo, L. (2019). Effects of political instability in consolidated destinations: The case of Catalonia (Spain). *Tourism Management*, 70, 134-139.

Petruška, I. (2018). Spektrálna analýza ukazovateľov cestovného ruchu európskych krajín (Spectral analysis indicators of tourism in European Countries). In S. Jenčová & E. Litavcová (Eds.), *Ekonomická aktivita turizmu (Economic Activity of Tourism): Conference proceedings* (pp. 120-128). Prešov: Faculty of Management.

Štefko, R., Gavurová, B., & Korony, S. (2016). Efficiency measurement in healthcare work management using Malmquist indices. *Polish Journal of Management Studies*, 13, 168-180.

Taneja, K., Ahmad, S., Ahmad, K., & Attri, S. D. (2016). Time series analysis of aerosol optical depth over New Delhi using Box–Jenkins ARIMA modeling approach. *Atmospheric Pollution Research*, 7, 585-596.

Vašaničová, P., Litavcová, E., & Jenčová, S. (2017). On Modelling of the Development of Turnover in Services in the Slovak Republic: Tourism Approach. In P. Pražák (Eds.), *35th International Conference of Mathematical Methods in Economics MME 2017 Conference Proceedings* (pp. 824-829). Hradec Králové: University of Hradec Králové.

Contact

doc. Mgr. Eva Litavcová, PhD.
University of Presov, Faculty of Management, Department of Mathematical Methods and Managerial Informatics
Konštantínova 16, 080 01 Prešov, Slovakia
eva.litavcova@unipo.sk

Mgr. Petra Vašaničová, PhD.
University of Presov, Faculty of Management, Department of Mathematical Methods and Managerial Informatics
Konštantínova 16, 080 01 Prešov, Slovakia
petra.vasanicova@unipo.sk