ARTIFICIAL NEURAL NETWORKS IN MODELLING SEASONAL TOURISM DEMAND – CASE STUDY OF CROATIA

Maja Gregorić
PhD, Assistant Professor, University of Rijeka, Faculty of Tourism and Hospitality Management,
Primorska 42, 51 410 Opatija, Croatia; e-mail: majam@fthm.hr

Tea Baldigara
PhD, Full Professor, University of Rijeka, Faculty of Tourism and Hospitality Management,
Primorska 42, 51 410 Opatija, Croatia; e-mail: teab@fthm.hr

ABSTRACT

The purpose of this paper is to design an artificial neural network in the attempt to define the data generating process of the number of German tourist arrivals in Croatia considering the strong seasonal character of empirical data. The presence of seasonal unit roots in tourism demand determinants is analysed using the approach developed by Hylleberg, Engle, Granger and Yoo – Hegy test. The study is based on seasonality analysis and Artificial Neural Networks approach in building a model which intend to describe the behaviour of the German tourist flows to Croatia. Different neural network architectures were trained and tested, and after the modelling phase, the forecasting accuracy and model performances were analysed. Model performance and forecasting accuracy evaluation was tested using the mean absolute percentage error. Based on the augmented HEGY test procedure it can be concluded the German tourist arrivals to the Republic of Croatia have nonstationary behaviour associated with the zero frequency and seasonal frequency. Taking this into consideration, in the analysis of the phenomenon it is necessary to consider its seasonal character. Given the importance of the tourism for Croatian economic development, the research results could be useful, for both, researchers and practitioners, in the process of planning and routing the future Croatian hotel industry development and improvement of business performances.

Key words: *Croatian tourism, seasonality, Hegy seasonal unit root test, artificial neural network model, MAPE*

1.   INTRODUCTION

The seasonality of tourism has been subject of many researches for many years worldwide, as well in Croatia, where tourism is the main driver of economic development. In general, tourism demand shows an increase, but various fluctuations influenced by numerous factors and external influences also. It is no surprise that in a numerous studies of tourism demand attention has been devoted to international tourist flows, due to their significance, but also due the fact that international tourism is statistically better captured than domestic tourism. According to data of Ministry of Tourism (2018), the highest number of foreign tourist arrivals in 2017 was realized by German tourists, followed by tourists from Italy, Slovenia and Austria. Considering the fact that the largest number of international tourist arrivals is realized by German tourists, in this paper German tourist arrivals and basic variables affecting those arrivals were analysed. This kind of analysis should be an integral part of decision making processes regarding tourism policy in the Republic of Croatia.

2.   LITERATURE REVIEW

Considering the great importance of tourism for national economy, and the complexity and variability of tourism market, researchers and practitioners as well have to pay great attention to tourism demand modelling and forecasting. The most recent paper by Song, Qui and Park (2019) reviewed 211 key paper with the main topic of tourism demand modelling and forecasting. The authors concluded that tourism demand forecasting methods have been evolving over past five decades, forecasting models have grown more diversified and there is no single method that performs well for all situations…The development of AI-based models is notable in the most recent tourism demand forecasting studies, despite the theoretical and technological limitations of these methods. In the paper by Wu et.al. (2017) the review of new developments in tourism demand forecasting that consider numerous data measurements of tourism demand and their determinants have also been made. Before this research, a more complex literature review was elaborated by Goh and Law (2011), in time when the adoption of the AI- based models in tourism demand studies were just recent and limited. In the mentioned study 15 AI- based tourism demand forecast papers were identified, and divided into two major categories; AI- based causal methods and AI- based time series models. Still, researchers are pointing out that time series models, econometric approach and artificial intelligence models are the three main categories of quantitative forecasting models (Peng et al. 2014, 182).

Since more researchers show their interest for the issue of more accurate forecasting results, this methodology has been considered interesting in economics and business areas since it is viewed as a valid alternative to the classical forecast approaches event in complex situations (Fernandes et al. 2011).

In Cang's paper (Cang, 2011) the complexity of tourism demand as social phenomenon was modelled by non-linear tourism demand forecast combination model. In the paper nine individual models were proposed; three linear combination models and one non-linear to model UK inbound tourism quarterly arrivals. Comparing mentioned models, the author concluded that the non-linear combination forecasting model, multilayer perceptron neural network showed the best performance compared to individual forecasting models and to all linear combination models.

Lin et al. (2011) modelled the data set of monthly visitors to Taiwan using ARIMA, ANN and MARS. According to this research, ARIMA showed the best forecasting performance. Taking into account a stochastic trend in data on the tourist arrivals to Catalonia (Spain) from 2001 to 2012 from different markets, Claveria and Torra (2014) applied a multivariate neural networks (MLP, RBF and Elman neural network) to obtain its forecasts. The modelling procedure was as usual in neural network modelling; training, validation and testing. In comparison of each neural network, it was concluded that MLP and RBF networks showed better forecasting performance than Elman network. The most significant contribution of those research is the innovative approach to tourism demand modelling and forecasting. Neural networks, as an artificial intelligence methodology for tourism demand modelling have been regarded by many theoretical and practitioners as promising technology, and consequently, in the last few decades more than 2000 articles on neural networks forecasting have been published covering a wide range of application (Yu et al., 2017)." Uysal and El Roubi (1999) demonstrated the usefulness of artificial neural networks as an alternative approach to the use of multiple regression in tourism demand using Canadian tourism expenditures in US as a measure of demand. Those authors identified that the application of ANN in tourism demand studied is useful in identifying existing patterns that may not be relieved by some other forecasting method, and in some cases their implementation result in better estimates in terms of prediction.

When considering studies dealing with implementation of artificial neural in Croatian tourism, the interest of researchers has been growing lately. In the study of Krešić et al. (2013) the particular advantages and shortcomings of ANN-based applications using an empirical case example were demonstrated. The authors used the data from a survey on attitudes and expenditures of tourists in Sarajevo, Bosnia and Herzegovina, and performed a multilayer perceptron-based key-driver analysis to obtain insight into those destination attributes that have a predominant influence on the overall tourist experience in Sarajevo. The results from the ANN-based analysis are compared to results from a regression-based key-driver analysis in order to identify possible significant differences between the approaches. Folgieri et al. (2017) modelled a backpropagation Artificial Neural Network to forecast tourist arrivals in Croatia and compared the results with those obtained with the linear regression methods. The results showed that using the neural network model to predict tourists' arrivals outperforms linear regression techniques. In the second step of their research, the same authors modelled and applied a backpropagation artificial neural network (ANN) to forecast tourists' arrivals in Croatia (Folgieri et al., 2018). Applying a machine learning method for decision support and pattern discovery such as ANN, represents an occasion to achieve a greater accuracy if compared to results usually obtained by other methods, such as linear regression. The authors extended the model of the previously used backpropagation Artificial Neural Network, with the data from sentiment analysis collected through social networks on the Internet. The accuracy of the neural network has been measured by the Mean Squared Error (MSE) and compared to results obtained applying the ANN without data coming from the sentiment analysis. The results showed that including data from sentiment analysis, outperforms the previous obtained results.

Since the seasonality is a one of a prominent feature of tourism, there is a large number of papers dealing with seasonality research and determination. Although numerous studies on seasonality in tourism have been conducted, formal tests of seasonality can be found in relatively few papers. Chang

and Liao (2010) presented the HEGY for the seasonal unit root that was used for the departures of destinations from Taiwan to Hong Kong, Japan and the USA in order to make an accurate inference about data seasonality. Vergori (2012) suggests that it is important to test the stochastic nature of the underlying process by using unit root tests. In order to test for seasonal unit roots, it is necessary to factorize the seasonal difference polynomial that for monthly data is defined as $\Delta\_12=(1-L^{12})$ where L is the backshift operator that is $L^j X_t=X_{(t-j)}$. The research results showed that the seasonal backshift operator should be applied to the sample of monthly arrivals in the Province of Lecce in order to achieve a stationary time series. Chaitipa and Chaiboonsri (2014) provided a non-linear forecasting models for prediction of the international tourist arrival to Thailand by using data from period 1998-2014, and the extent version of seasonal unit root test was performed out to test the data. The main conclusion was that the movement of international tourist arrivals to Thailand was affected by seasonal unit root process during observed period.

As the literature review indicates, the topic of tourism seasonality is analysed in more detail in the foreign literature, while in the domestic literature it is somewhat less. The authors of this paper, therefore, attempted to examine the character of international tourist arrivals in Croatia and to generate an artificial neural network model that represents the observed data. Given the importance of accurate forecasts for the dynamic and complex tourism market, the aim of this research is to identify the possibilities of applying an artificial neural network approach in analysing phenomena that a strong seasonal component.

3. DATA AND METHODOLOGY

In the chapters below the selection of variables and methodology will be presented.

3. 1 The selection of variables

The selection of tourism demand measure, as dependent variable, and independent variables as its determinants, was established and confirmed by relevant literature on tourism demand studies (Uysal and El Roubi, 1999; Fretchling, 2001; Song, et.al., 2012; Shafiullah et.al., 2019). In this study the following variables were used:

$TA_{GER,t}$ - the total German tourism arrivals in Croatia

$GDP_{GER,t}$ - the German gross domestic product at market prices

$RCPI_{GER,t}$ - the tourism price variable for German tourists, which is calculated from the following equation:

$$RCPI_{GER,t} = \frac{CPI_{CRO,t}/EX_{CRO/GER,t}}{CPI_{GER,T}}$$

This variable takes into account the effect of both relative inflation and the exchange rate on the demand for tourism to Croatia.

$D_{1,t} \dots D_{11,t}$ - set of dummy variables for each month

t - some period time

At any time of period t, one of the seasonal dummies $D_{1,t}$, $D_{2,t}$, $D_{3,t}$,…, $D_{s,t}$ will equal 1, while all the others will equal 0. The rule is to use one fewer dummy variables than categories. So, for monthly data, in this research eleven dummy variables were used.

The main hypothesis is that the German tourist arrivals can be explained as a function of the variables selected in the study. Monthly data, from January 2001 to December 2018 (216 observations), were collected from Croatian Bureau of Statistics (First Releases) and Eurostat databases. Since gross domestic product measures the total output of an entire economy by adding up total consumption, investment, government expenditure, and net exports, in this research German gross domestic product at market prices is considered a quality approximation of income for an entire economy in the observed period. Due to the lack of monthly data on gross domestic product, in the research data for this variable were collected on a quarterly basis and were converted to monthly data using linear-match last low to high frequency method. This method inserts the low observation value into the last period of the high frequency data, then performs linear interpolation on the missing values.

The analysis was performed using Eviews 10.0., SPSS 22.0., and Stata 14.0. data analysis software packages.

3. 2 Methodology

An investigation of seasonal unit roots presence in this paper is conducted by the approach developed by Hylleberg et al. (1990); general procedure which allows tests for unit root at some seasonal frequencies without maintaining that unit roots are present at all seasonal frequencies. The augmentation lag selection was suggested by Rodriques and Taylor (2007). After seasonality identification, an artificial neural network model was established and evaluated using the mean absolute percentage error.

3. 2. 1 Testing for seasonal unit roots in monthly data – HEGY test

Univariate seasonal time series data generating process is assumed to be as follows:

$$Y_{St+s} = \mu_{St+s} + x_{St+s}$$
$$\alpha(L)x_{St+s} = u_{St+s}$$
$$s = 1 - S, \dots, 0 \qquad t = 1, 2, \dots, N \tag{1}$$

Where:

S - number of seasons

μ_{St+s} - deterministic part of time series

x_{St+s} - stochastic part of time series

$\alpha(L)$ - an AR(S) polynomial α(L)=(1-α_1^* L-α_2^* L^2-...-α_S^* L^S)

L - lag operator

General HEGY procedure explains how to test for seasonal unit roots in processes that may also exhibit deterministic or stationary stochastic seasonality. HEGY procedure, simplified by Beaulieu and Miron (1993) is presented below.

The approach of Hylleberg et. al. (1990) is based on expanding α(L) around zero and seasonal frequency unit roots {exp (±i2πj/S); j=0,…,S/2}. Generally, the testing equation of the Hegy approach can be written as:

$$\Delta_{SY_{St+s}^{\xi}=\pi_0 y_{0,St+s}^{\xi}} + \pi_{S/2 Y_{S/2,St+s}^{\xi}} + \sum_{j=1}^{S/2-1}(\pi_{1j}Y_{1j,St+s}^{\xi} + \pi_{2j}Y_{2j,St+s}^{\xi}) + \sum_{j=1}^{k}d_j \Delta_{SY_{St+s-j}^{\xi}} + e_{St+s,k}^{\xi} \qquad (2)$$

where:

$$y_{0,St+s}^{\xi} = \Delta_0^0(L)Y_{St+s-j}^{\xi} = \sum_{i=0}^{S-1}Y_{St+s,k}^{\xi}$$

$$Y_{S/2,St+s}^{\xi} = \Delta_{S/2}^0(L)Y_{Sn+s-j}^{\xi} = \sum_{i=0}^{S-1}\cos\{(i+1)\pi\}Y_{St+s-i-1}^{\xi}$$

$$Y_{1j,St+s}^{\xi} = -\{cos(\omega_j) - L\}\Delta_j^0 Y_{Sn+s-i-1}^{\xi} = \sum_{q=0}^{S-1}cos\{(q+1)\omega_j\}Y_{St+s-q-1}^{\xi} \qquad (3)$$

$$Y_{2j,St+s}^{\xi} = sin(\omega_j)\Delta_j^0(L)Y_{St+s-1}^{\xi} = -\sum_{q=0}^{S-1}sin\{(q+1)\omega_j\}Y_{St+s-q-1}^{\xi}$$

$$j = 1, \ldots, s/2 - 1$$

As this study considers monthly data, equations (2) and (3) become:

$$\Delta_{12Y_{12t+s}^{\xi}=\pi_0 y_{0,12t+s}^{\xi}} + \pi_{6Y_{6,12t+s}^{\xi}} + \sum_{j=1}^{5}(\pi_{1j}Y_{1j,12t+s}^{\xi} + \pi_{2j}Y_{2j,12t+s}^{\xi}) + \sum_{j=1}^{k}d_j \Delta_{SY_{12t+s-j}^{\xi}} + e_{12t+s,k}^{\xi} \qquad (4)$$

Where:

$$y_{0,12t+s}^{\xi} = Y_{12t+s-1}^{\xi} + Y_{12t+s-2}^{\xi} + Y_{12t+s-4}^{\xi} + Y_{12t+s-5}^{\xi} + Y_{12t+s-6}^{\xi} + Y_{12t+s-7}^{\xi} + Y_{12t+s-8}^{\xi}$$
$$+ Y_{12t+s-9}^{\xi} + Y_{12t+s-10}^{\xi} + Y_{12t+s-11}^{\xi} + Y_{12(t-1)+s}^{\xi}$$
$$Y_{6,12t+s}^{\xi} = Y_{12t+s-1}^{\xi} + Y_{12t+s-2}^{\xi} + Y_{12t+s-4}^{\xi} + Y_{12t+s-5}^{\xi} + Y_{12t+s-6}^{\xi} + Y_{12t+s-7}^{\xi} + Y_{12t+s-8}^{\xi}$$
$$+ Y_{12t+s-9}^{\xi} + Y_{12t+s-10}^{\xi} + Y_{12t+s-11}^{\xi} + Y_{12(t-1)+s}^{\xi}$$

$$X_{1j,12t+s}^{\xi} = \sum_{q=0}^{11} cos\left\{(q+1)\frac{2\pi j}{s}\right\} Y_{12t+s-q-1}^{\xi} \qquad j = 1,2,3,4,5$$
$$X_{2j,12t+s}^{\xi} = -\sum_{q=0}^{11} sin\left\{(q+1)\frac{2\pi j}{s}\right\} Y_{12t+s-q-1}^{\xi} \qquad j = 1,2,3,4,5$$

$$t = 1,2,\dots,N \qquad s = -11,-10,-9,-8,\dots-1,0$$

(5)

Hegy procedure testing reported the following tests:

H_0	Roots	Frequncy ω_j	$\frac{2\pi^*}{\omega_j}$	test	Test tail
$\pi_0 = 0$	One real root	0	∞	t[0]	left tail
$\pi_6 = 0$	One real root	π	2	t[pi]	left tail
$\pi_{11} = \pi_{21} = 0$	Two complex conjugate roots	$\frac{\pi}{6}$	12	t[pi/6]	upper tail
$\pi_{12} = \pi_{22} = 0$	Two complex conjugate roots	$\frac{\pi}{3}$	6	t[pi/3]	upper tail
$\pi_{13} = \pi_{23} = 0$	Two complex conjugate roots	$\frac{\pi}{2}$	4	t[pi/2]	upper tail
$\pi_{14} = \pi_{24} = 0$	Two complex conjugate roots	$\frac{2\pi}{3}$	3	t[2pi/3]	upper tail
$\pi_{15} = \pi_{25} = 0$	Two complex conjugate roots	$\frac{5\pi}{6}$	$\frac{12}{5}$	t[5pi/6]	upper tail

*number of periods to complete a full cycle

The following testes are reported also:

H_0	Roots	Frequencies ω_j	Test	Test tail
$\pi_{11} = \pi_{21} = \pi_{12}$ $= \pi_{22} = \pi_{13} = \pi_{23}$ $= \pi_{14} = \pi_{24} = \pi_{15}$ $= \pi_{25} = \pi_6 = 0$	All seasonal roots	$\frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}, \pi$	F[seas]	upper tail

$\pi_0 = \pi_{11} = \pi_{21} = \pi_{12}$ $= \pi_{22} = \pi_{13} = \pi_{23}$ $= \pi_{14} = \pi_{24} = \pi_{15}$ $= \pi_{25} = \pi_6 = 0$	All roots in $(1 - L^{12})$	$0, \dfrac{\pi}{6}, \dfrac{\pi}{3}, \dfrac{\pi}{2}, \dfrac{2\pi}{3}, \dfrac{5\pi}{6}, \pi$	F[All]	upper tail

3. 2. 2 Lag – length selection

According to literature (Hall, 1994; Ng and Perron, 1995) different methods can be used in order to determine the order of augmentation for the Hegy regression (e.g. criteria such as AIC, BIC or MAIC, or sequential method). In this paper the sequential method has been used. The maximum lag length (k_{max}) was determined using following equation:

$$k\_max = Int[12(T/100)^{\frac{1}{4}}] \tag{6}$$

Where Int[.] is the integer part and T is total sample size. In the case when sequential method is used, the lag selection procedure starts by fitting the model with $k = k_{max}$ and tests sequentially for the significance of the coefficient associated with ΔY^{ξ}_{St+s-k} until the null is rejected. Following the suggestion of Ng and Perron (1995), to determine the individual significance of the lags of ΔY^{ξ}_{St+s-k}, the researcher have to use the standard normal distribution critical values and select a 10% level of significance the default value.

3. 2. 3 Artificial neural networks

Due to the importance of tourism, complexity and variability of the numerous elements that should be considered in the implementation of an effective model, tourism demand modelling and forecasting is one of the most challenging fields in tourism research. In this context, the implementation of artificial neural networks has proved to be particularly advisable, thanks to their ability to approximate any kind of function within a desirable range (Thiesing and Vornberger, 1997; Crone, 2002).

The artificial neural network models consist of set of nodes (neurons) for processing input data and a set of connections for "memorizing" information. Through an activation transfer function, each node processes the input signal received from other nodes and produces a transformed output signal to other nodes (Claveira et al., 2015, 493).

Figure 1. A neural network model
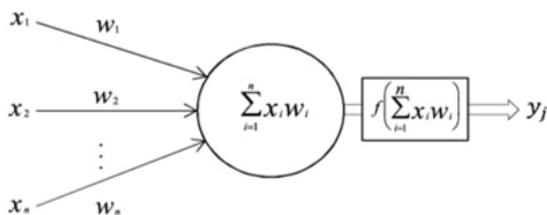
Source: Authors adoption according to Fernandes and Texeira (1999)

The inputs in a multilayer artificial neural network are connected to the output via a middle layer. Generally, the network model can be specified as:

$$Y_t = f(X_t, \alpha, \beta) + \varepsilon_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j F(\sum_{i=1}^{12} \beta_{ij} x_{ij} + \beta_{0j}) + \varepsilon_t \qquad (7)$$

where:

Y_t - dependent variable (regressand variable)

X - a vector of regressors which include eleven monthly dummy variables

n - the number of units in the middle layer

F - Sigmoid transfer function

α_t - a vector of coefficients (or weights) from the middle to output layer unit

β_t - a matrix of coefficients from the input to middle-layer units at time t

t - some period time

Depending on the connecting patterns of different layers, two major types of architecture can be distinguished; feed-forward networks and recurrent (back propagation) network. Main difference between them is that in feed-forward networks the information runs only in one direction, since recurrent networks are feedback connections from outer layers of neuron to lower layer of neurons, which takes into account the temporal structure of the data. The data-driven nature of artificial neural networks makes those models overcome the limitation of traditional forecasting methods, including misspecification, biased outliers, assumption of linearity, and re-estimation. The modelling approach used in this paper consists of four main phases, as follow: data pre-processing, architecture selecting, training process and model performance evaluating. Data pre-processing implies that data were inspected and selected so there were no missing values. Also, all data were normalised by subtracting the minimum and dividing by the range, (*x-min*)/(*max-min*), so all normalized values were in [0,1] interval. Since the range of values of observed data varies widely, in some machine learning algorithms, objective functions will not work properly without normalization.

Constructing an efficient neural network involves choosing an appropriate architecture – the number of layers, units in each layer and the connections among units – selecting the transfer function of the middle and output units, designing a training algorithm, choosing initial weights and specifying rule. For this research, a common artificial neural network architecture is used, which can be described as: Continuous-valued input layer containing thirteen nodes (processing units) corresponding to the number of variables proposed to estimate the number of German tourist arrivals to Croatia, one hidden layer containing eight nodes to handle nonlinearities inherent in data set, and one output layer containing one node corresponding to the one dependent variable subject to the analysis.

After architecture selection it is very important to select which parts of the data set will be used in training, validation and holdout. The training set is the subsection of an observed data from which

the machine learning algorithm "learns," relationships between the input and the target variable. The validation set is another subset of the input data to which the machine learning algorithm has to be applied in order to see how accurately it identifies relationships between the known outcomes for the target variable and the dataset's other features. The holdout subset, called also "testing" set, provides a final estimate of the machine learning model's performance after it has been trained and validated.

In this paper the standard three-layer feedback artificial neural network model is used. One output node is employed. The activation function for hidden nodes is sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{8}$$

The sigmoid function takes real-valued arguments and transforms them to the range 0 to 1. The activation function for the output node is the identity function (pure linear function):

$$f(x) = x \tag{9}$$

This function takes real-valued arguments and returns them unchanged. Bias terms were used in both hidden and output nodes.

After selecting the activation functions, the batch type of training was selected. This kind of training updates synaptic weights only after passing all training data records which means that this kind of training uses information from all records in training data set. The advantage of this kind of training is that it directly minimizes the total error, but for that purpose it need to update weights many times until one of stopping rules is met. Hence, the batch training may need many data passes.

3. 2. 4 Mean Absolute Percentage Error

After modelling German tourist arrivals to Croatia, it was necessary to examine the forecast accuracy of the chosen artificial neural network. Due to its advantages of scale-independency, and interpretability as well, in this research as a measure of modelling accuracy mean absolute percentage error (MAPE) has been used.

The Mean Absolute Percentage Error is expressed in generic percentage terms and it is computed by the following formula:

$$MAPE = \frac{1}{N} \sum_{t}^{n} \frac{|(A_t - F_t)|}{A_t} \cdot 100 \tag{10}$$

Where:

A_t - the actual value

F_t - the forecasted value

N - the number of observation

t - some period time

MAPE is measure which permits the comparison of different models, with different time periods and number of observations. Measures based on percentage errors, such this, have the disadvantage of being infinite or undefined if $y_t=0$ for any observation in the test set, and having extreme values when any y_t is close to zero.

4.   EMPIRICAL RESULTS AND DISCUSSION

In this section, the results obtained from empirical research are presented and discussed in detail. As previously mentioned, the results are based on a sample size that covers the period from January 2001 to December 2018, which includes 216 observations.

Figures 2 to 4 depict the evaluation of the analysed series in total [figure (a)] and in natural logarithm [figure (b)]. Primary goal is to determine whether the seasonality observed in time series show only a determinist nature or whether it is caused by the presence of unit roots.

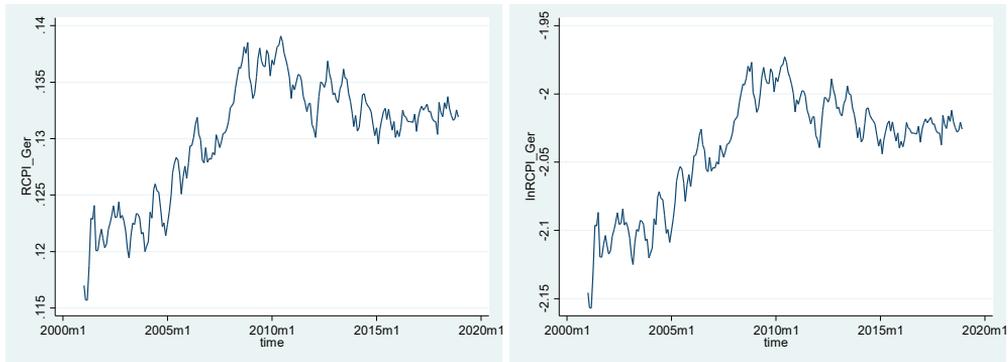Figure 2. Total and natural logarithm of German tourist arrivals to Croatia

a) total   b) natural logarithm
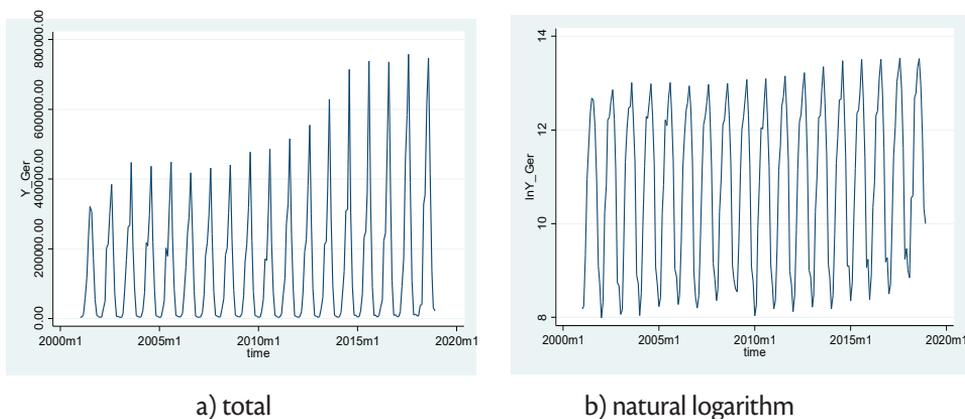
Source: Authors

Figure 3. Total and natural logarithm of gross domestic product at market prices

a) total   b) natural logarithm
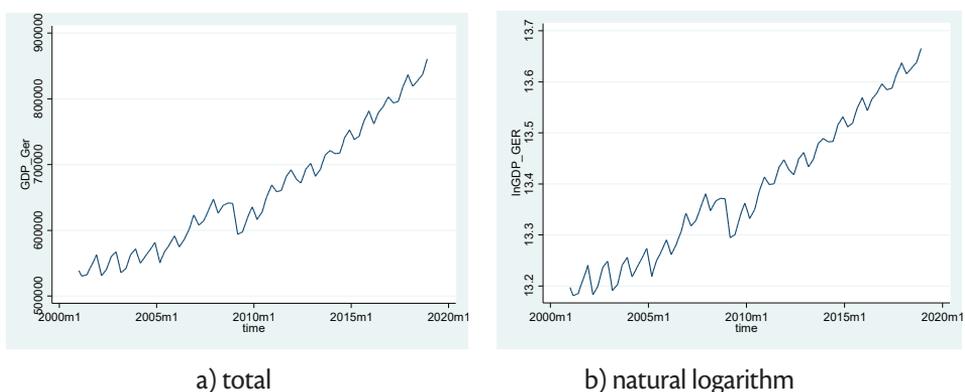
Source: Authors

Figure 4. Total and natural logarithm of tourism price variable

a) total	b) natural logarithm

Source: Authors

Based on the figures 2 - 4 it is decided to perform the further analysis on natural logarithm values of the data. Namely, to avoid the negative effects on the results of the seasonal unit-root testing which can be caused by changing variance (which is case in observed sample), it is advisable to analyse natural logs of the observed data.

The results of the performed seasonal unit-root testing by HEGY approach are shown in tables 1 – to 3. The results will be interpreted at the 5% level significance.

Table 1. HEGY seasonal unit root test for lnY_Ger

```
HEGY Monthly seasonal unit root test for lnY_Ger

Number of observations  :  193
Deterministic variables : Seasonal dummies and seasonal trends
Optimal lag selection method: Sequential at 10% level
Lags tested: 14
Augmented by lags : 11
```

	Stat	1% critical	5% critical	10% critical
t[0]	0.752	-3.843	-3.284	-3.001
t[Pi]	-3.891	-3.843	-3.284	-3.001
F[Pi/6]	4.680	11.422	8.969	7.832
F[Pi/3]	7.444	11.422	8.969	7.832
F[Pi/2]	10.511	11.422	8.969	7.832
F[2*Pi/3]	11.473	11.422	8.969	7.832
F[5*Pi/6]	14.706	11.422	8.969	7.832
F[All seas]	18.316	8.393	7.304	6.771
F[All]	16.827	8.332	7.274	6.755

Source: Authors

Based on table 1, the null hypothesis of the presence of a unit root at frequency zero cannot be rejected. Note that t[0] statistics is a left-tail test; the reported statistics is less negative than the critical value associated with 5%l level. Therefore, the statistics do not fall in the rejection region of the null hypothesis. In the case of t[Pi] test (the other left-tail test), situation is opposite. Since the test statistic is smaller than the critical value, the null hypothesis of the presence of a unit root at the frequency π can be rejected. All the remaining tests (F[Pi/6] to F[All]) are upper-tail test, and as such, the null hypothesis can be rejected when the test statistics is bigger than the critical value. In this case, the null hypothesis is rejected in all tests, except in the case of the two complex roots associated with frequency π/6 (test F[Pi/6]) and π/3 (test F[Pi/3]). It can be concluded that one unit root at the frequency zero is found, as well as the two complex conjugate unit roots at π/6 and π/3 frequency. Based on HEGY test, when the lag length is determined by sequential method, the German tourist arrivals to Croatia have nonstationary behaviour associated with the zero frequency and the frequency associated with seasonal oscillations that complete a full cycle after 6 or 12 months.

Table 2. HEGY seasonal unit root test for lnGDP_Ger

```
HEGY Monthly seasonal unit root test for lnGDP_GER_MON

Number of observations  :  199
Deterministic variables : Seasonal dummies and seasonal trends
Optimal lag selection method: Sequential at 10% level
Lags tested: 14
Augmented by lags : 5


              Stat   1% critical  5% critical  10% critical
       ----------------------------------------------------------------
t[0]        -2.292    -3.846      -3.288      -3.005
t[Pi]       -2.058    -3.846      -3.288      -3.005

F[Pi/6]     13.208    11.442       8.991       7.853
F[Pi/3]      4.024    11.442       8.991       7.853
F[Pi/2]     14.167    11.442       8.991       7.853
F[2*Pi/3]   26.563    11.442       8.991       7.853
F[5*Pi/6]   14.411    11.442       8.991       7.853
F[All seas] 15.805     8.379       7.298       6.770
F[All]      15.296     8.316       7.267       6.752
```

Source: Authors

According to data in table 2, it can be concluded that logarithm of German gross domestic product has one unit root at frequency zero and π, but also two complex conjugate unit roots at the frequency π/3. That means that logarithm of German gross domestic product at market prices has nonstationary behaviour associated with the zero frequency, and with the frequency associated with seasonal oscillations that complete a full cycle after six months also.

Table 3. HEGY seasonal unit root test for lnRCPI_Ger

```
HEGY Monthly seasonal unit root test for lnRCPI_Ger

Number of observations  :  192
Deterministic variables : Seasonal dummies and seasonal trends
Optimal lag selection method: Sequential at 10% level
Lags tested: 14
Augmented by lags : 12


              Stat    1% critical  5% critical  10% critical
---------------------------------------------------------------
t[0]         -0.941    -3.842        -3.284       -3.000
t[Pi]        -4.075    -3.842        -3.284       -3.000

F[Pi/6]       8.283    11.418         8.966        7.828
F[Pi/3]       5.920    11.418         8.966        7.828
F[Pi/2]       7.476    11.418         8.966        7.828
F[2*Pi/3]     8.497    11.418         8.966        7.828
F[5*Pi/6]    15.565    11.418         8.966        7.828
F[All seas]  15.416     8.396         7.305        6.772
F[All]       14.242     8.335         7.275        6.756
```

Source: Authors

Based on the table 3, it can be concluded that the logarithm of the tourism price variable has nonstationary behaviour also. This series have one unit root at the frequency zero, and two complex conjugate roots at all other frequencies, except π and $5\pi/6$. Since the dummy variables are seasonal categorical variables, the seasonal unit root test does not need to be performed on those variables.

After the presence of seasonal nonstationary behaviour has been confirmed, the creation of an artificial neural network was initiated. It should be stressed that a pre-processing was undertaken on the input and output data, corresponding only to a normalization in the range –1 to 1, for all the variables.

Table 4. Case Processing Summary

|		N	Percent
Sample	Training	125	57,9%
	Testing	68	31,5%
	Holdout	23	10,6%
Valid		216	100,0%
Excluded		0	
Total		216	

Source: Authors

As shown in table 4, the analysed time series data were divided into three sets: the training set (125 observations), validation set (68 observations) and holdout set (23 observations).

As already said, the selected artificial neural network model is a multilayer backpropagation model type 13:8:1. As activation function in hidden layers the sigmoid function was used, and for the output node the identity (linear).
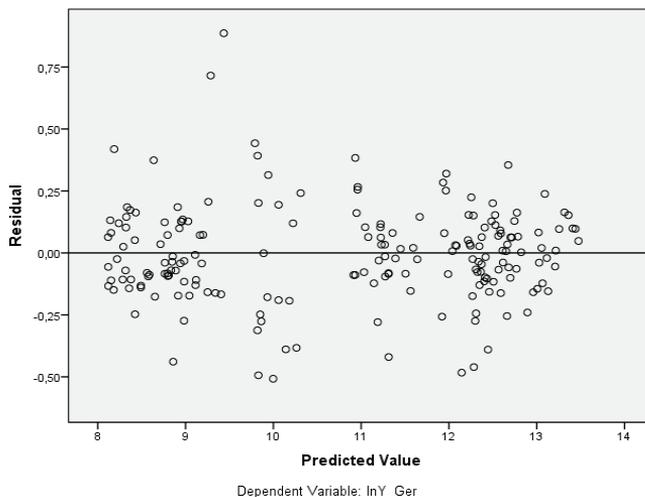
The final artificial neural network is depicted in figure 5.

Figure 5. The artificial neural network model
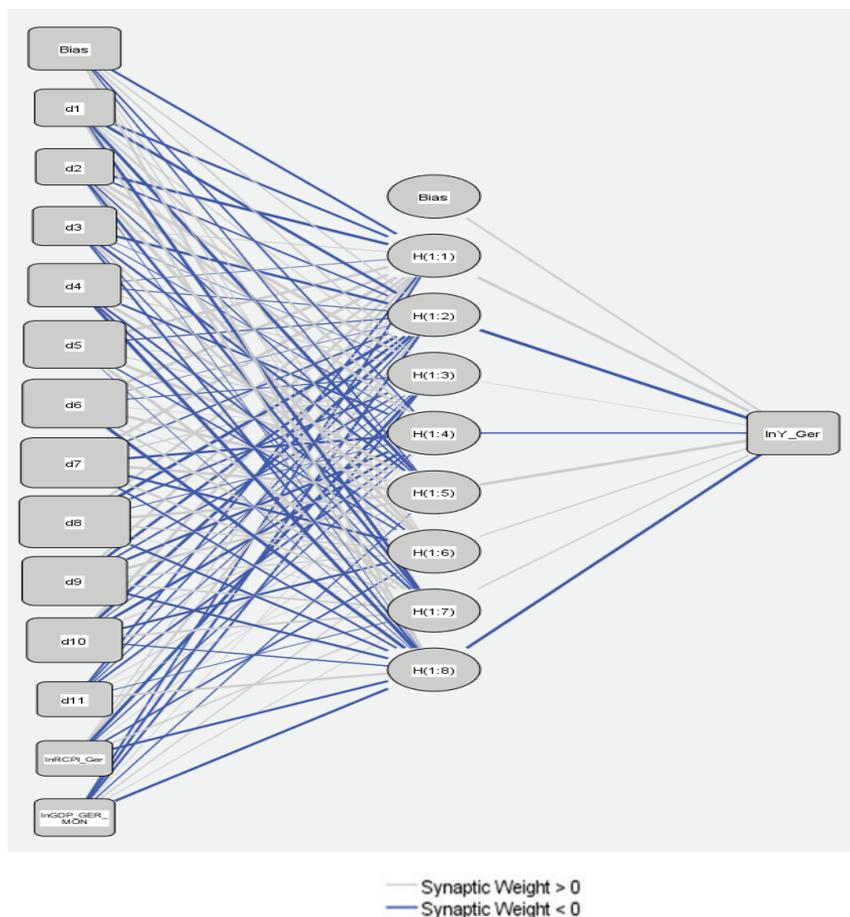
The information from software output shows a couple of positive signs, which are shown below.

Figure 6. Residual by predicted chart


Dependent Variable: lnY_Ger

Source: Authors

The figure 6 presents a residual-by-predicted-value chart for scale-dependent variable. As evident, there are no visible patterns between residuals and predicted values of dependent variable.

The importance of an independent variable is a measure of how much the network's model-predicted value changes for different values of the independent variable. The normalized importance is the importance values divided by the largest importance values and expressed as percentages.

Figure 7. Independent variable importance chart
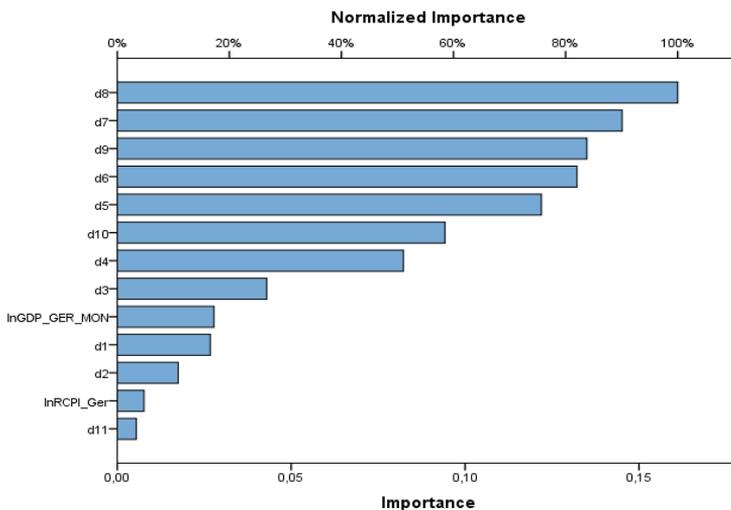
Source: Authors

The figure 7 shows a bar chart of the values of importance, sorted in descending value of importance. As expected, the seasonal dummy variables for August and July have the greatest importance. Those variables have the greatest effect on the German tourist arrivals to Croatia, which was one of the assumptions of this research. As shown in figure, the weakest effect on the dependent variable is the effect of dummy variable for November. The following table presents the calculated values of independent variable importance.

Table 5. Independent variable importance

	Importance	Normalized Importance
d1	,027	16,6%
d2	,018	10,9%
d3	,043	26,7%
d4	,082	51,1%
d5	,122	75,7%
d6	,132	82,0%
d7	,145	90,1%
d8	,161	100,0%
d9	,135	83,8%
d10	,094	58,4%
d11	,005	3,4%
lnRCPI_Ger	,008	4,8%
lnGDP_GER_MON	,028	17,2%

Source: Authors

The table above confirms the interpretation of the figure 7. The most important variables are dummy variables for August (a contribution of 100% to the outputs) and July (a contribution of 90,1% to the outputs). The price variable has surprisingly low importance (4,8%), while the gross domestic product variable has more importance (17,2%) to the target variable. Such results can be explained by the assumption that the German tourists are primary motivated for traveling to Croatia not because of the service price or their purchasing power. They are probably motivated by some other variables, which are, in this research, included in the dummy variables.

The obtained artificial neural network model has been evaluated according to some accuracy criteria. According to Baggio and Klobas (2011,151) a MAPE value of 1,601% indicates a high model forecasting accuracy.

Figure 8. Total and natural logarithm of actual and fitted values of German tourist arrivals
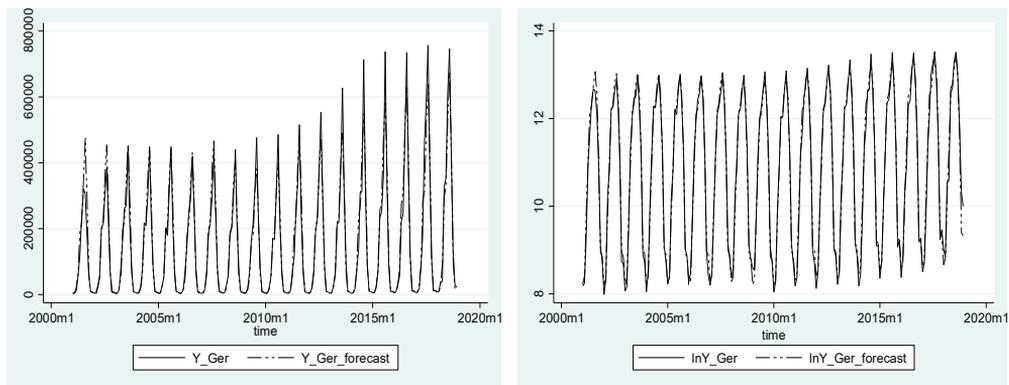
a) total   b) natural logarithm

Source: Authors

The figure above confirms that backpropagation artificial neural model fits well the actual data through the sample period. Selected model shows that forecasted values of analysed series follow the behaviour of the empirical data set. The assumption that German tourist arrivals to Croatia are, besides tourism price variable and gross domestic product at market prices variable, strongly affected by seasonal impact is confirmed, which confirms credibility of Hegy seasonal unit root test conducted in the first part of the research.

5. CONCLUSION

Due to the importance of tourism for the Croatian economy, and the fact that tourists from German markets make the main number of international tourist arrivals to Croatia, the main purpose of this research was to examine the seasonality of the number of German tourist arrivals in the Republic of Croatia. The presence of seasonal unit roots is analysed using the approach developed by Hylleberg, Engle, Granger and Yoo – Hegy test. Based on the augmented HEGY test procedure the main conclusion is made; the German tourist arrivals to the Republic of Croatia have nonstationary behaviour associated with the zero frequency and seasonal frequency. After testing the seasonality occurrence, different neural network architectures were trained and tested. The proposed artificial neural network model confirmed the main assumption of this research - the German tourists are not primary motivated for traveling to Croatia because of the service price or their purchasing power. Analysis showed that the price and gross domestic product variables has much lower importance to the target variable in comparison to seasonal dummy variables. The calculated mean absolute percentage error – MAPE value of 1,601% indicates a high model forecasting accuracy. This kind of analysis forms an integral part of tourism policy and decision making regarding investment in public and private sector, as well as in marketing decisions in the Republic of Croatia. Therefore, this research is considered to be a good starting point for future research of the possibility of applying artificial intelligence methods in analysing

tourism phenomenon. Future tourism research should be certainly addressed in investigating the possibility of using some other concepts of artificial intelligence. The ability of artificial intelligence to perform tasks which, traditionally, require human cognitive function, has made those methods useful for research in the tourism industry. As some previous research showed, as well as this one, the implementation of artificial intelligence approach primarily can save time and money and, at the same time, potentially eliminate human error and allow tasks to be performed quickly.

REFERENCES

Baggio, R. & Klobas, J. (2011) ˝Quantitative Methods in Tourism˝, Bristol: Channel View Publications

Beaulieu, J.J., Miron, J.A. (1993) "Seasonal unit roots in aggregate U.S. data", *Journal of Econometrics*, 55(1-2). pp. 305-328. https://doi.org/10.1016/0304-4076(93)90018-Z

Del Barrio Castro, T., Osborn, D.R., Taylor, A.M.R. (2016) "The Performance of Lag Selection and Detrending Methods for HEGY Seasonal Unit Root Tests", *Econometric Reviews*, 35(1), pp. 122-168. https://doi.org/10.1080/07474938.2013.807710

Del Barrio Castro, T., Osborn, D.R. i Taylor, A.M.R. (2011) "On Augmented HEGY Tests for Seasonal Unit Roots", The School of Economics Discussion Paper Series EDP 1121, https://doi.org/10.1017/S0266466612000060

Cang, S. (2011) "A Non- Linear Tourism Demand Forecast Combination Model", *Tourism Economics*, 17(1), pp. 5- 20. 10.5367/te.2011.0031

Chang, Y.W. and Liao, M.Y. (2010) "A Seasonal ARIMA Model of Tourism Forecasting: The Case of Taiwan", *Journal Asia Pacific Journal of Tourism Research*, 15(2), pp. 215-221.https://doi.org/10.1080/10941661003630001

Chaitipa, P. and Chaiboonsrib, Ch. (2014) "International Tourists Arrival to Thailand: Forecasting by Non-linear Model", *Procedia Economics and Finance*, 14, pp. 100-109. https://doi.org/10.1016/S2212-5671(14)00691-1

Claveria, O., Torra, S. (2014) "Forecasting tourism demand to Catalonia: Neural networks vs. time series models", *Economic Modelling*, 36, pp. 220- 228. doi:10.1016/j.econmod.2013.09.024

Crone, S. F. (2002) "Training artificial neural networks for time series prediction using asymmetric cost functions" In: *Proceedings of the 9th International Conference on Neural Information Processing*, 18-22 November, Singapore, pp. 2374-2380. doi:10.1109/iconip.2002.1201919

Fernandes, P.O, Texeira, J. (2008) "Applying the artificial neural netwoek methodology for forecasting the tourism time seris", In: *Proceeding of 5ᵗʰ International Scintific Conference Business and Management 2008*, 16-17 May 2008, Vilnius, Latvia: Faculty of Business Management, pp. 653-658.

Fernandes, P.O., et.al. (2011) "Forecasting Tourism Demand with the Artificial Neural Networks", In: *Book of proceedings Vol. II– International conference on tourism & management studies*, 26-29 October, Algarve, pp. 1017-1019.

Folgieri, R., Baldigara, T., Mamula M. (2018) "Sentiment analysis and artificial neural networks-based econometric models for tourism demand forecasting", In: *Proceedings of the 25th International Congress Tourism and Hospitality Industry 2018*, April 26 – 27, Opatija, Hrvatska: Fakultet za menadžment u turizmu i ugostiteljstvu, pp. 88-97.

Folgieri, R., Baldigara, T., Mamula M. (2017) "Artificial neural networks-based econometric models for tourism demand forecasting" In: *Proceedings of the 4th International Scientific Conference Tourism and Creative Industries: Trends and Challenges*, May 4 – 6, 2017, Opatija, Hrvatska: Fakultet za menadžment u turizmu i ugostiteljstvu, pp. 169-182.

Frechtling, D. C. (2001) *Forecasting tourism demand: methods and strategies*, Oxford: Elsevier Ltd.

Goh, C., Law, R. (2011) "The Methodological Progress of Tourism Demand Forecasting: A Review of Related Literature", *Journal of Travel & Tourism Marketing*, 28(3), pp. 296- 317. https://doi.org/10.1080/10548408.2011.562856

Hall, A. (1994) "Testing for a unit root in time series with pretest data-based model selection", *Journal of Business and Economic Statistics*, 12(4), pp. 461–470. 10.2307/1392214

Hylleberg R. F. et al. (1990) "Seasonal integration and cointegration", *Journal of Econometrics*, 4(1-2), pp.215-238. https://doi.org/10.1016/0304-4076(90)90080-D

Krešić, D., Mikulić, J., Kožić, I. (2013) "Artificial Neural Network-Based Applications in Travel and Tourism Research: A Review and Case Study" In: *International critical tourism studies conference*, Sarajevo, BIH, http://cts.som.surrey.ac.uk/publication/artificial-neural-network-based-applications-in-travel-and-tourism-research-a-review-and-case-study/wppa_open/ (15.6.2019.)

Lin, C.J., Chen, H.F., Lee, T.S.. (2011) "Forecasting Tourism Demand Using Time Series, Artificial Neural Networks and Multivariate Adaptive Regression Splines: Evidence from Taiwan", *International Journal of Business Administration*, Vol. 2, No. 2, pp. 14- 25. doi: 10.5430/ijba.v2n2p14

Ministry of Tourism (2018.), *Turizam u brojkama*, 1-43. https://mint.gov.hr/UserDocsImages//AA_2018_cdokumenti//190614_HTZ%20TUB_2018_hrv.pdf

Ng, S., P. Perron. (1995) "Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag", *Journal of the American Statistical Association* 90: 268–281. 10.2307/2291151

Peng, B., Song, H., Crouch, G.I. (2014) "A meta-analysis of international tourism demand forecasting and implications for practice", *Tourism Management*, 45, pp. 181-193. https://doi.org/10.1016/j.tourman.2014.04.005

Rodrigues, P. M. M., Taylor A.M.R. (2004) "Asymptotic distributions for regressionbased seasonal unit root test statistics in a near-integrated model", *Econometric Theory*, 20(4), pp. 645–670. https://doi.org/10.1017/S0266466604204029

Song, H., Witt, S. F., Li, G. (2012) "*The Advanced Econometrics of Tourism Demand*", New York: Taylor & Francis Ltd

Song, H., Qui, R. T. R., Park, J. (2019) "A review of research on tourism demand forecasting", *Annals of Tourism Research*, 75: 338-362. https://doi.org/10.1016/j.annals.2018.12.001

Shafiullah, M., Okafor, L. E., & Khalid, U. (2019) "Determinants of international tourism demand: Evidence from Australian states and territories", *Tourism Economics* 25(2), pp. 274-296. 135481661880064.doi:10.1177/1354816618800642

Thiesing, F. M., Vornberger, O. (1997) "Forecasting sales using neural networks", Lecture Notes in Computer Science, pp.321–328, 10.1007/3-540-62868-1_125

Uysal, M. El Roubi, S. (1999) "Artificial Neural Networks versus Multiple Regression in Tourism Demand Analysis", *Journal of Travel Research*, pp. 111-118, https://doi.org/10.1177/004728759903800203

Yu, Y., Wang, Y., Gao, S., Tang, Z. (2017) "Modelling inbound tourism data", Computational Intelligence and Neuroscience, 2017. pp.1-9. https://doi.org/10.1155/2017/7436948

Vergori, A. S. (2012) "Forecasting tourism demand: the role of seasonality", *Tourism Economics*, 8(5), pp. 915–930. 10.5367/te.2012.0153

Wu C. D., Song, H., Shen, S. (2017) "New developments in tourism and hotel demand modeling and forecasting", *International Journal of Contemporary Hospitality Management*", 29(1), pp. 507-529. https://doi.org/10.1108/IJCHM-05-2015-0249

UMJETNE NEURONSKE MREŽE U MODELIRANJU SEZONSKE TURISTIČKE POTRAŽNJE – STUDIJA SLUČAJA HRVATSKE

Maja Gregorić
Dr. sc., docentica, Sveučilište u Rijeci, Fakultet za menadžment u turizmu i ugostiteljstvu, Primorska 42, 51 410 Opatija, Hrvatska. *e-mail*: majam@fthm.hr

Tea Baldigara
Dr. sc., redovita profesorica, Sveučilište u Rijeci, Fakultet za menadžment u turizmu i ugostiteljstvu, Primorska 42, 51 410 Opatija, Hrvatska. *e-mail*: teab@fthm.hr

SAŽETAK

Svrha je ovog rada dizajnirati umjetnu neuronsku mrežu sa ciljem definiranja procesa generiranja podataka o broju dolazaka njemačkih turista u Hrvatsku, uzimajući u obzir sezonski karakter empirijskih podataka. Prisutnost sezonskoga jediničnog korijena u odrednicama turističke potražnje analizirana je pristupom koji su razvili Hylleberg, Engle, Granger and Yoo – Hegyijevim testom. Istraživanje je temeljeno na analizi sezonalnosti i umjetnim neuronskim mrežama u stvaranju modela koji za cilj ima opisati ponašanje njemačkih turističkih tokova u Hrvatskoj. Trenirane su i testirane različite arhitekture neuronskih mreža, te je nakon faze modeliranja ocijenjena pouzdanost i analizirane su performanse modela. Prognostička pouzdanost modela testirana je srednje apsolutnom postotnom pogreškom. Temeljem rezultata Hegyijevog testa može se zaključiti kako je serija broja dolazaka njemačkih turista nestacionarna, na razini nulte i sezonske frekvencije. Uzimajući to u obzir, u analizi navedene pojave potrebno je uvažiti njezin sezonski karakter. S obzirom na važnost turizma za ekonomski razvoj Hrvatske, ovo bi istraživanje trebalo koristiti, kako istraživačima tako i praktičarima, u procesu planiranja razvoja hrvatske turističke industrije i u poboljšanja poslovnih performansi.

Ključne riječi: *hrvatski turizam, sezonalnost, Hegy-jev test sezonskog jediničnog korijena, model umjetne neuronske mreže, MAPE*