

## A COMPARATIVE STUDY ON MODELLING AND FORECASTING TOURISM REVENUES: THE CASE OF TURKEY

Murat ÇUHADAR <sup>1</sup>

*Department of Tourism Management, Süleyman Demirel University, Turkey*

ORCID: 0000-0003-0434-1550

### ABSTRACT

Tourism revenues have important implications for tourism countries in terms of management of tourism-related policies. In order to accurately direct production planning, pricing, promotion and strategic marketing programs, labor and capital resources, accurate and reliable forecasts are needed. Forecasting the developments in tourism with scientific basis methods is an important guide for central and local public administration programs and tourism operators. When reviewing the literature, comparative studies on modeling and forecasting tourism revenues using Artificial Neural Networks (ANNs) are limited and this paper aims to fill this gap. Based on the gap seen in the literature, the purpose of this study is to develop the optimal forecasting model that yields the highest accuracy when comparing the performances of three different methods namely Exponential Smoothing, Box-Jenkins and ANNs for forecasting Turkey's tourism revenues. Forecasting performances of the models were measured by MAPE statistics. As a result of the analyses performed, it was found that ANN Model with [4:5:1] architecture was the best one among the all models applied in this study.

### Article History

Received 07 July 2020

Revised 14 October 2020

Accepted 14 October 2020

### Keywords

Tourism revenues

Modelling

Forecasting

ANN

---

<sup>1</sup> Address correspondence to Murat ÇUHADAR (PhD), Department of Tourism Management, Faculty of Economics and Administrative Sciences, Süleyman Demirel University, Isparta, TURKEY. E-mail: [muratcuhadar@sdu.edu.tr](mailto:muratcuhadar@sdu.edu.tr)

## INTRODUCTION

Due to the globalization, increasing welfare level, increasingly safe, fast and comfortable airline transportation, increase in the number of airline companies and rapid developments in information technologies, tourism sector has been among the fastest growing and revenue generating sectors in the world. Every year, millions of people travel to different parts of the world as tourists, spending billions of dollars and causing huge economic activity throughout the world. When the global tourism flows are analysed, while the number of people participating in international tourism in 2005 was 807 million, in 2019, it reached 1.5 billion people with an increase of 3.8% compared to the previous year (UNWTO, 2020a). In a study published by the World Travel and Tourism Council, it is stated that in 2019, tourism sector provided 10.3% of global gross product with a contribution of 8.9 trillion dollars and 10% of total employment with 330 million jobs worldwide. In the same study, it is predicted that the number of people to be employed by the tourism sector will reach 425 million in 2030, and will constitute 11.3% of the global gross product. Looking at global tourism revenues, those from international tourism flows in 2019 amounted to 1.7 trillion dollars, which means 6.8% of the total exports worldwide; 28.3% of global services sector exports (WTTC, 2020). Although tourism is a concept with historical, cultural, social and economic perspectives, it is an activity that has been handled mostly with the economic aspects. The tourism sector has an important place for the economies of both developing and developed countries. Many countries see the income from international tourism as a source of foreign exchange used to cover the deficits in the balance of payments. For this reason, it can be said that many tourism countries compete in the international tourism market in order to provide the foreign currency they need for development and growth and to increase their national income.

Tourism revenues are based on tourism expenditures. Tourism expenditure refers to the amount paid for the acquisition of consumer goods and services as well as valuables for own use or gift during and during tourist trips. It includes expenses paid or reimbursed by others as well as expenses incurred by the visitors themselves (United Nations, 2010). These revenues have changed hands many times within the country, reaching a total effect far above the original amount, thus contributing to the national income and the development of the country's economy. In other words, the expenditures made by tourists in the countries they visit provide income not only to businesses and individuals operating in the tourism sector, but also indirectly to various sectors. Throughout the years, tourism in Turkey

has also caused direct and indirect economic contributions, which made it one of the most significant sectors for country's economy. Among the positive economic impact of tourism regarding Turkey, its contribution to income and employment, as well as generating foreign exchange effect, which is important in reducing deficit in the balance of payments, can be observed. According to statistics published by Culture and Tourism Ministry, Turkey's tourism revenues reached 34.522 billion dollars in 2019, an increase of 16.9% compared to the previous year. 83.2 percent of this income was obtained from foreign tourists, 16.5 percent from citizens residing abroad and came to the country as tourists, and the remaining 0.4% from GSM roaming and marina services. Total revenue from GSM roaming services and marina services amounted to \$ 127.1 million (Ministry of Culture and Tourism, 2020) The tourism sector is an important economic activity due to the results it has created worldwide. Given the economic results of tourism, it is a sector where many countries strive to benefit from these results. For this reason, many countries, regions and touristic destinations in the world desire to analyse and improve their tourism dynamics and increase their economic contributions. One of the most effective options for achieving this goal is to analyse tourism from past to present and make predictions about the future. When studies on modelling tourism revenues are reviewed, it is observed that several issues were in focus. Among these issues, studies where the relationship between tourism revenues and economic growth is tested (Gunduz & Hatemi-J, 2005; Öztürk & Acaravcı, 2009; Gökovalı, 2010; Ertugrul & Mangir, 2015; Hüseyini et al., 2017; Qin et al., 2018; Wu & Wu, 2018) stand out. Another of the study topics that the authors focus on is the determinants of tourism revenues, such as foreign visitor statistics, total investment amount, employment, number of beds, real exchange rates (Payne & Mervar, 2002; Kara et al., 2003; Aktaş et al., 2014; Kaplan & Aktas, 2016; Akay et al., 2017; Ongan et al., 2017; Çalışkan et al., 2019). As an alternative to the statistical forecasting methods commonly used in the field of tourism, use of machine learning methods for forecasting studies is evident in recent years. However, comparative studies on modelling and forecasting tourism revenues using ANNs and statistical methods are limited and this paper aims to fill this gap. According to many authors (e.g. Law & Pine, 2004; Kon & Turner, 2005; Song & Li, 2008; Peng et al., 2014; Önder, 2017; Höpken et al., 2020; Zhang et al., 2020) in the tourism forecasts, there is no single method that produces the best results in all cases. Many researchers working in this field have agreed that the results will vary according to the method and model to be used, the characteristics of the data set, the number of observations and the estimated period. In other words, there is no magic model or method that can produce

the best result in every situation. Based on this approach, the purpose of this study is to develop the optimal forecasting model that yields the highest accuracy when compared the forecast performances of three different methods, namely, Exponential Smoothing, Box-Jenkins and Artificial Neural Networks for forecasting Turkey's tourism revenues. Having introduced the emphasis of forecasting in tourism and research objectives, the rest of this paper is structured as follows. Next section describes the research methodology. Section 3 describes the data employed in the study as well as research design. Thereafter, forecasting performances of alternative methods and models are compared and the empirical findings are presented. In the last section, the implications and contributions of the current study and suggestions for future ones are included.

## METHODS USED IN THE STUDY

In the following sub-sections, explanations about forecasting methods used in predicting and modelling of data are included.

### Exponential Smoothing

The exponential smoothing method consists of different application options in which the estimates are constantly updated, considering the recent changes in the data. In the Exponential smoothing method, the weighted averages of past period values are calculated and included in the models as the estimated value of the future periods. The working principle of this method is that more recent data and observations are more important than the effect of very old observations or data. Holt-Winter's seasonal exponential smoothing method is suitable for modelling and estimating data containing trend and seasonal effects. This method uses three separate sets of equations, one used to determine the level of the series in the  $t$  period, the other to trend, that is, the third, to model the seasonal component. The equations used in the calculation of the Multiplicative-Seasonal Holt-Winter's method are given below (Makridakis et al., 1998);

$$\text{Level:} \quad L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1})$$

$$\text{Seasonal:} \quad S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s}$$

$$\text{Forecast:} \quad F_{t+m} = L_t + S_{t-s+m}$$

In the given equations;

$S$  = Season length,  
 $L_t$  = The general level of the series in the  $t$  period,  
 $b_t$  = Trend component,  
 $S_t$  = Seasonal component,  
 $F_{t+m}$  =  $m$  forecasting value for the next period,  
 $\alpha$  (level),  $\beta$  (trend) and  $\gamma$  (season) smoothing constants.

### Box-Jenkins Method

The Box-Jenkins approach is a well-established and commonly used method, especially in short and medium-term time series forecasting. It is an assumption of the method that the data set applied in this method, which provides successful results in short and medium term modelling and forecasting studies, is a discrete and stationary data set consisting of observation values with equal time intervals. The basic principle of the Box-Jenkins methodology is based on its value in any period of time series, a combination of past observations and error terms. Since Box-Jenkins models can only be used in stationary series, stationary condition is important in determining the appropriate model group (Anvari et al., 2016). Data used in practice, especially financial and economic data, often do not meet stationarity conditions. The stationarity of this type of data is disrupted by trend, season and cyclical fluctuations and random factors. In general, the expression of the ARIMA (p,d,q) model is as follows:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

The tendency of time series, which consist of values observed in weekly, monthly or quarterly time periods, to reach the highest and lowest level in the same periods of successive years or months, indicates the presence of the seasonal component in the data. The representation of the seasonal ARIMA model is ARIMA (p,d,q)(P,D,Q)<sub>s</sub> or simply SARIMA. Here, P represents the degree of seasonal autoregression (SAR), D represents the number of seasonal differences, Q is the order of the seasonal moving average (SMA) model, and S is the seasonal period. Seasonal ARIMA model in P, D and Q orders with back shift operator is expressed as:

$$\Phi_p(B^S)\Delta_s^D y_t = \Theta_q(B^S)\varepsilon_t$$

In this model,  $\Delta_s$  represents the seasonal difference operator and S is the seasonal period, and S= 12 for monthly data and S= 4 for quarterly data. In

the model, the operator  $\Delta^D$  specifies the degree (D) of the seasonal difference of the data. Non-stationary data is transformed into stationary data with the differencing process indicated by  $\Delta_S^D$ . The term  $\Phi_P$  in the model refers to the seasonal auto regression (SAR) parameter;  $\Theta_Q$  is the seasonal moving average (SMA) parameter, and  $y_t$  is the non-stationary series. The Box-Jenkins method uses a four-step, and iterative process to determine a suitable model among the candidate model options. These stages are briefly; model identification, parameter estimation, diagnostic control and forecasting. The method, combines autoregressive, moving average differencing/integration procedures and finally tests the model for statistical validity. In the model selection, factors such as the fact that the data meet the conditions of stationarity and whether they are affected by the seasonal component are determinants. The basic strategy in the model building process is based on the principle of "parsimony" (Moeeni & Bonakdari, 2016). Therefore, it is essential to determine the time series characteristics of the data used first in the modelling and prediction process with the Box-Jenkins method, and to create candidate models suitable for the characteristics of the data and the most appropriate and statistically significant model among them.

### **Artificial Neural Networks (ANNs)**

Experiments and researches in the field of science, natural and engineering sciences regarding the brain structure have opened a horizon in modelling the information processing processes of the human brain and have led to the development of systems that mimic its working principles. The smallest element of the human brain is nerve cells. Nerve cells perform their information processing activities in a group, not alone. Recently, ANN-based models have been one of the most important non-parametric methods proposed for time series forecasting. ANN is defined as a structure consisting of several interconnected units or artificial nerve cells that mimic biological neural networks (Chen et al., 2007). According to another definition, ANNs are complex systems that are created in the form of artificial nerve cells (neurons), which are compared to biological neurons in the human brain, to form connections with each other at different effect levels. Artificial neurons come together to form ANN. Neurons are composed of five main elements: input connections, weights, addition function, activation function and output connections. The basic element of an ANN is the neuron as shown in Figure 1.

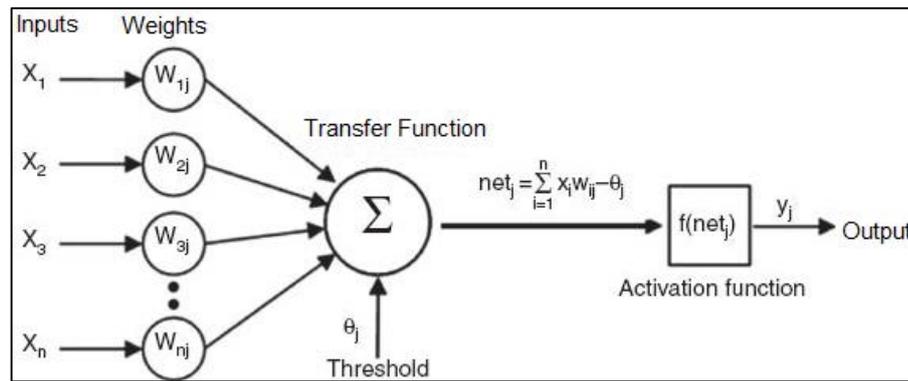


Figure 1. *Basic Elements of an Artificial Neuron*

In Figure 1,  $(x_i)$  represents inputs,  $(w_{ij})$  weights,  $(\theta_j)$  threshold value,  $(f)$  activation function and  $(y_j)$  output. In the information processing continuum, information input from the outside world to the neuron takes place first through the input connections  $(x_1, x_2, \dots, x_i)$ . Information entry to a neuron from other neurons is done by multiplying by a weight  $(w_{1j}, w_{2j}, \dots, w_{ij})$  with a value between  $(-1)$  and  $(+1)$ . In the next step, the information multiplied by weights is passed through an activation function  $(f)$  to obtain the output  $(y_j)$  (Golam & Hasin, 2013). The activation function is briefly the mechanism for converting input signals to output signals for each process element. There are generally three layers in an ANN: input layer with interactive neurons, output layer and hidden layer (s). The first layer in the ANN is the input (also called the input) layer, which functions to receive external data from the artificial neural network. The input layer is the first layer in ANN, and serves to network the data from the outside world. The input layer consists of data from variables that are assumed to be related to the problem being solved. Therefore, the number and quality of neurons in the input layer are shaped according to the variables that affect the problem. The last layer in the ANN is the output layer that performs the function of exporting the processed information. The layer between the input and output layers is called the hidden layer. Neurons in the hidden layer do not interact with the outside world, they take the task of receiving the signals from the input layer and transmit them to the output layer. In Figure 2, an exemplary multi-layered ANN model is presented.

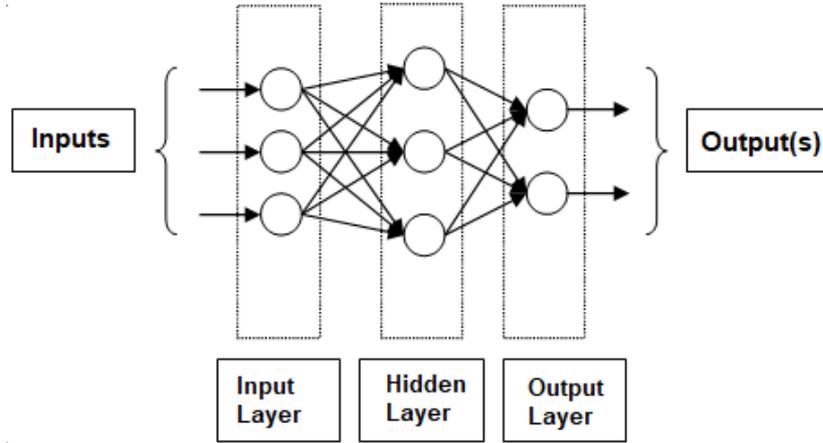


Figure 2. A Multi-layered ANN model

ANNs are trained by learning a problem through the datasets shown to it. Training of ANNs, in other words the learning of the network is the process of updating the connection weights to fulfil a function expected by the user. The most used method in the literature for updating ANN parameters is the "Back Propagation" algorithm. Over the years, ANN architectures and models such as MLP, RBF, Hopfield, Jordan, Elman, Kohonen SOM, which are suitable for use in different fields, have been developed. The most widely used ANN architecture for predictive purposes in the social sciences and tourism forecasting field is the back propagation-feed forward multi-layered perceptron (MLP) model (Wong et al., 2000; Zhang & Qi, 2005; Song & Li, 2008; Moreno et al., 2011; Teixeira & Fernandes, 2012; Bayramoğlu & Başarir, 2018). MLP model is very popular because it is widely applicable in solving business related problems such as prediction, classification and modelling (Smith, 2002). In the use of MLP models for forecasting, determining the network architecture is an especially important issue. In the literature, it is stated that network architectures with a single hidden layer are sufficient in estimating time series. The inputs to be presented to the network are among the lagged observations ( $y_{t-1}, y_{t-3}, y_{t-12} \dots y_{t-n}$ ) of the data set used; output consists of original ( $y_t$ ) observations. The relationship between the output value and the inputs is as follows:

$$y_t = w_0 + \sum_{j=1}^p w_j f \left( w_{0j} + \sum_{i=1}^N v_{ij} y_{t-i} \right) + e_t$$

In this equation,  $w_j, v_{ij}$  represents the weight values between neurons,  $p$  represents the number of neurons in the hidden layer and  $f$  is the activation function used in the hidden layer. The most used activation functions are sigmoid  $f(x) = \frac{1}{1+e^{-x}}$  and hyperbolic tangent  $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$  functions (Egrioglu et al., 2017).

## DATA ANALYSIS

In modelling and analysis of the data, Turkey's quarterly inbound tourism revenues statistics in US dollars realized in 2003-Q1 and 2019-Q4 periods were used. Data were obtained from the TurkStat. First, the time series characteristics of the data used were analysed and the time series components that were effective on the data were determined. In the following stage, analyses were carried out to create forecasting models suitable for the structure of the data from the Exponential Smoothing, Box-Jenkins and ANN methods. In the process of determining the appropriate model, the smoothing constants that minimize the mean of error squares for exponential smoothing models, Autocorrelation (ACF) and Partial Autocorrelation functions (PACF) and Bayes Information Criteria (BIC) for Box-Jenkins models were taken into consideration. When constructing appropriate ANN models, different time delays in the input layer and the instantaneous data values are used in the output layer. For ANN models, parameters and parameter values that will minimize the prediction error produced by the model were used. Forecasting values generated by different models are compared with real tourism income values, and it is determined which method produces more successful forecasting results. Forecasting performances of the applied methods were measured by "Mean Absolute Percentage Error (MAPE)" statistic. Various statistics are available to measure the estimation performance of alternative models. Examples of these are Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), etc.. However, MAPE statistics have some advantages over others. First of all, it eliminates the possible disadvantages in comparing models with different unit values since it expresses the estimation errors proportionally. Secondly, it is not affected by the size or smallness of the observation values, as it gives relative measurement results. In addition, producing interpretable results can be counted among its advantages over other statistics. The mathematical expression of MAPE is as follows:

$$MAPE = \frac{\sum_{t=1}^n |e_t|}{n \cdot y_t} 100(\%)$$

where including ( $e_t = y_t - \hat{y}_t$ )

$y_t$  = Value of the observation at time  $t$ ,

$\hat{y}_t$  = Fitted value for the observation at time  $t$ ,

$n$  is the length of forecasting horizon.

In the evaluation of alternative models, if the MAPE statistic value is below 10%, it is accepted as a high accuracy, good accuracy between 10-20%, reasonable accuracy between 20-50% and an unsuccessful estimate over 50% (Lewis, 1982).

### Analysis of Time Series Properties of Data

The time graph of the quarterly tourism revenues data from January 2003-Q1 to December 2019-Q4 used in the study is given in Figure 3.

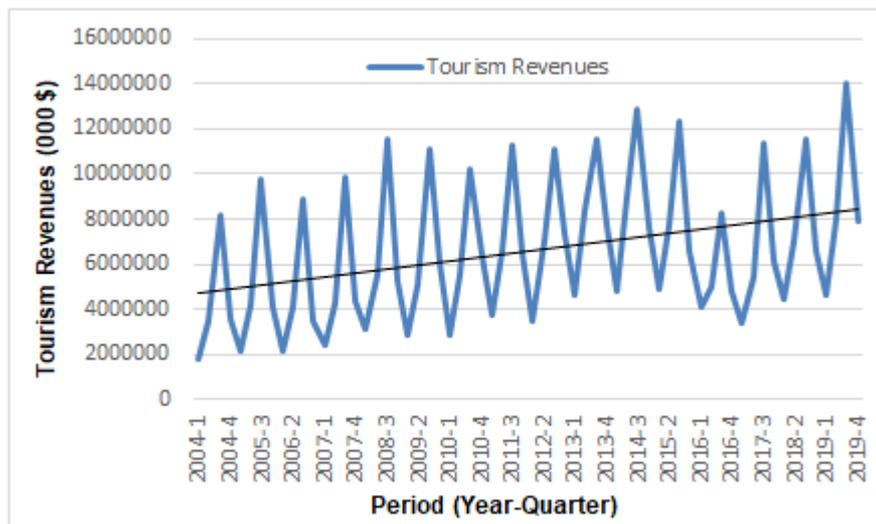


Figure 3. Time Series Graph of Tourism Revenues Data (2003-Q1 – 2019-Q4)

When the graph is analysed, it is observed that the data is influenced by a positive trend and seasonal component with irregular fluctuations in some years. Seasonal fluctuations start to increase in April and reach the highest value in July, August and September (3rd Quarter). It reaches the lowest values in January, February and March (1st Quarter). In order to determine the forecasting models suitable for the structure of the data used in the study, linear and non-linear trend analyses were performed. As a

result of the applied trend analysis, the data showed a positive trend. F test was performed to examine the statistical significance of trend analysis. It was observed that it was statistically significant at the level of 0.05. Model summary of the trend analysis is given in Table 1.

Table 1. *Model Summary of Trend Analysis*

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	0.143	11.153	1	67	0.001	4305361.802	58326.519	
Logarithmic	0.174	14.146	1	67	0.000	1682720.893	1422786.086	
Quadratic	0.178	7.161	2	66	0.002	2950057.901	172859.243	-1636.182
Compound	0.203	17.103	1	67	0.000	3689949.662	1.012	
Exponential	0.203	17.103	1	67	0.000	3689949.662	0.012	

Similarly, in order to reveal the seasonal factors in the series, the data were subjected to seasonal decomposition using the X-12 ARIMA technique. During the seasonal decomposition process, moving average weights were calculated with the interval of "Period + 1 (Endpoints Weighted by 0.5)". The seasonal factor values obtained reveal that the series is under the effect of recurrent seasonal fluctuation every four quarters. The seasonal pattern is formed by reaching the lowest value in the first quarter of each year and the highest value in the third quarter. Season factor values obtained are given in Table 2.

Table 2. *Seasonal Factor Values of the Data*

Period	Months	Seasonal Factors (%)	Period	Months	Seasonal Factors (%)
I. Quarter	January	53.5	III. Quarter	July	168.6
	February			August	
	March			September	
II. Quarter	April	90.7	IV. Quarter	November	87.2
	May			October	
	June			December	

### Application of the Exponential Smoothing Method

Since the data used in the study are under the influence of trend and seasonal fluctuation and seasonal patterns have different magnitudes in periods, "Multiplicative-Seasonal" models were applied in practice, and additive models were not included in the analyses. In the model determination process, statistical significance of t-tests of parameter

estimates of the models, error squares and MAPE values of the "Simple Seasonal, Multiplicative-Seasonal Exponential Trend, Multiplicative-Seasonal, Damped Trend" and Multiplicative-Seasonal Linear Trend (Holt-Winter's) models were examined. As a result of the evaluations, it was seen that the optimum exponential smoothing model is the "Multiplicative-Seasonal Holt-Winter's" model for the series. In the model, season factor values obtained by seasonal decomposition are used as seasonal factors. Model summary is given in Table 3.

Table 3. *Exponential Smoothing Model Parameters*

Parameter	Estimate	SE	t	Sig.
Alpha (Level)	0.499	0.085	5.863	0.000
Gamma (Trend)	0.0014	0.029	2.182	0.000
Delta (Season)	1.000	0.227	4.406	0.000

Initial values of the model are calculated as follows:

3302267.57439 (Level)

80724.47668 (Trend)

### Application of the Box-Jenkins Method

In the application of the Box-Jenkins method, the stationarity analysis of the data was carried out with the Augmented Dickey-Fuller (ADF) test. As a result of the ADF analysis, it was determined that there is a high association between neighbouring seasonal observations in the series and the series is out of stationary. As a result of the analysis, the trend stationarity has been provided by taking the first-order seasonal (D=1) difference. The orders of model processes were determined by analysing the autocorrelation (ACF) and partial autocorrelation (PACF) functions calculated from the data. As a result of the attempts to establish various models, it was determined that the Box-Jenkins model suitable for the series is "Multiplicative-Seasonal (1,0,0)(1,1,1)<sub>4</sub>" model. The final parameter estimates and the general summary of the determined model are given in Table 4.

When Table 4 is examined, it is seen that the t-statistics of the parameter estimates in the determined SARIMA model are significant at the 0.05 significance level.

Table 4. Model Summary of SARIMA (1,0,0)(1,1,1)<sub>4</sub> Model

Variable	Estimate	Std. Error	t-Statistic	Prob.
Constant	336967.061	85054.148	3.962	0.000
AR(1)	0.605	0.093	6.477	0.000
SAR(1)	0.670	0.121	5.522	0.000
SMA(1)	0.999	0.184	5.436	0.000
Innovational Q2-2016	-2346399,710	637883.106	-3.678	0.001
Additive Q3-2016	-2150002,488	431506.184	-4.983	0.000
R-squared	0.961	Schwarz Bayes Criterion		27.087
Seasonal Differencing: 1	Non-Seasonal Differencing: 0	Transformation: No Transformation		

After determination and parameter estimates, residues of the model were analysed. The Ljung-Box ( $Q^*$ ) statistics were computed for checking residuals. The Ljung-Box statistics, for the seasonal series calculating as follows:

$$Q^* = n(n+2) \sum \frac{r_k^2}{n-k} \sim \chi^2(k-p-q-P-Q)$$

It is a diagnostic measure of white noise for a time series, assessing whether there are patterns in a group of autocorrelations under the hypotheses;

H<sub>0</sub>: ACFs are not significantly different than white noise ACFs (i.e., ACFs = 0).

H<sub>1</sub>: ACFs are statistically different than white noise ACFs (i.e., ACFs ≠ 0).

Table 5. Ljung-Box Statistics and X<sup>2</sup> Values of the Model Residues Series

Lag	Autocorrelation	Std. Error <sup>a</sup>	Ljung-Box Statistic		
			Value	df	Sig. <sup>b</sup>
4	-0.016	0.119	4.033	4	0.402
8	0.041	0.115	5.176	8	0.739
12	0.132	0.111	9.070	12	0.697
16	-0.025	0.107	11.643	16	0.768
20	-0.052	0.102	12.626	20	0.893
24	0.093	0.097	14.426	24	0.936

a. The underlying process assumed is independence (white noise).

b. Based on the asymptotic chi-square approximation.

The Ljung-Box ( $Q^*$ ) statistics and X<sup>2</sup> values of the model for the varied lags of the residue series are given in Table 5. In the calculations

made for the varied delays of the model's series of residuals, it was found to be  $Q^* < X^2$ , therefore, null hypothesis of 0.05 significance level was accepted. The calculated  $Q^*$  statistics show that there is no significant autocorrelation between the model's residuals; confirms that the series has a random process and therefore the suitability of the determined SARIMA model.

### **Application of the ANN Method**

In modelling the data with ANN, the forecasting performances of alternative ANN architectures created using different data sets were analysed. 16 observations of the current 69 quarter-year data are grouped as training, and 53 observations are grouped as test data. In the input layer of each network established for trial purposes, 3 lagged observation values ( $y_{t-3}, y_{t-6}, y_{t-9}, y_{t-12}$ ) recommended for the prediction of quarterly series, and the original observation values ( $y_t$ ) were used in the output layer. The data was normalized in the range [0:1] before keying into the computer using the equation below,

$$x_n = \frac{x_0 - x_{\min}}{x_{\max} - x_{\min}}$$

where:

$x_0$  = original values,

$x_n$  = normalized values,

$x_{\min}$  = the lowest number included in the data, and

$x_{\max}$  = the biggest number included in the data.

Within the scope of the study, model alternatives created with 3, 4 and 5 layer architectures were tried. The learning process of MLP network was realized by presenting the training data to the network. The implementation of the method was carried out with the Neural Network Toolbox of the "MATLAB – Simulink (version: r2019a)" computer program. The stop limit of the training process of the established networks was determined as 30.000 iterations for each trial. In the ANN application performed, the predictive performances of candidate ANN models with 3 observations lagged datasets in the input layer and various number of neurons (1-7) in the hidden layer were examined. Since there is no generally accepted rule in the literature to determine the number of neurons in hidden

layers, the network has been trained with different neuron number alternatives. In the process of modelling the data with alternative multi-layered ANN structures, all the different ANN architectures were tested using the data determined subsequently for this purpose. The forecasting performances of ANN models for different architectures were examined by comparing the forecasted values found as a result of the test process with real values.

As a consequence of several attempts, it has been observed that 3 observations lagged MLP model which has [4-5-1] architecture performed best in forecasting competition. During the training of the ANN model, the connection weights were updated with the "back propagation" method. The "Logarithmic Sigmoid" algorithm was used as the activation function and the "Levenberg-Marquardt" algorithm was used as the training function. Figure 4 represents the ANN model created.

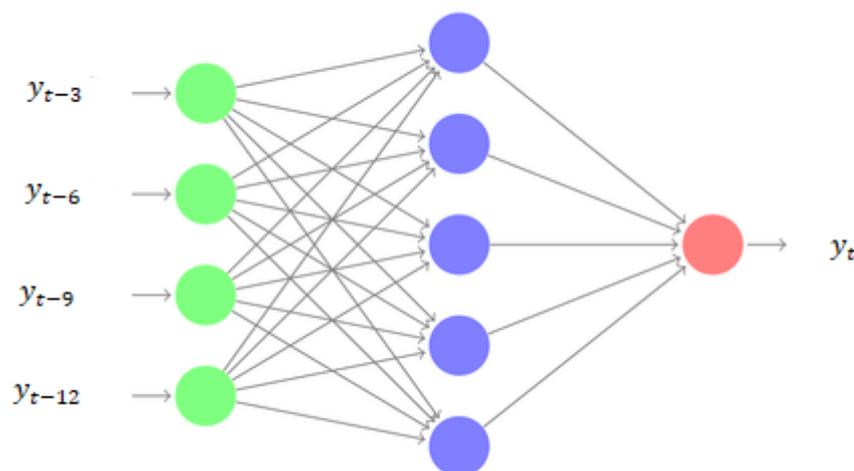


Figure 4. *Developed ANN Model as a Result of Experiments*

## EMPIRICAL RESULTS

The empirical results of the examined forecasting performances of ANN, exponential smoothing and Box-Jenkins (SARIMA) models are given in Table 6. According to the criteria proposed by Lewis (1982), it can be said that all applied models successfully produce high accuracy estimates, since the MAPE values of each model are less than 10%. Low MAPE indicates that the deviations between the predicted values obtained from the model and the actual values are very small. Among the alternative forecasting models obtained, the model that provides the lowest MAPE value is the

ANN model. The ex-post forecasting performances of three different models indicated that constructed MLP-ANN model outperforms the other models with the smallest MAPE of 8.06.

Table 6. *Accuracy Comparisons of Forecasting Models*

<b>Forecast Model</b>	<b>MAPE (%)</b>
MLP-ANN [4-5-1] Model	8.06
Box-Jenkins (1,0,0)(1,1,1) <sub>4</sub> Model	8.1
Holt-Winter's Multiplicative-Seasonal	9.34

When the results of the forecasting accuracy of each model tested are examined, it is possible to say that all the applied models produce highly successful forecasting results. As stated earlier, in the classification in the prediction modelling literature, models with 50% MAPE statistic value are misidentified models, models with 20% margin of error are models with acceptable accuracy, and models with 10% or less error margin are classified as highly accurate models.

The Box-Jenkins method has been successfully applied in time series forecasting studies due to its advantages, such as, following a gradual path in achieving the most appropriate solution, controlling the candidate model studied at all levels, providing the opportunity to model according to the characteristics of the data and especially the short and medium term prediction successes. Similarly, Holt-Winter's seasonal exponential smoothing method has also been successfully applied to modelling data under the influence of the trend and seasonal component. However, the model with the lowest margin of error among all tested models is the (MLP) ANN model. ANNs can learn and generalize the nonlinear patterns in the data and thus create solutions with a reasonable margin of error for the problems that they have not encountered before. These findings reveal the ability of ANNs, a machine learning technique, to learn complex and nonlinear models effectively, which is a clear advantage over linear models. For this reason, ANN is also known as a successful method in forecasting time series. Considering both the empirical findings obtained from this study and previous studies on tourism forecasting, it can be seen that ANN models that do not have any negativities (such as over-training, faulty architecture etc.) produce successful forecasting results when compared with models generated by conventional statistical methods.

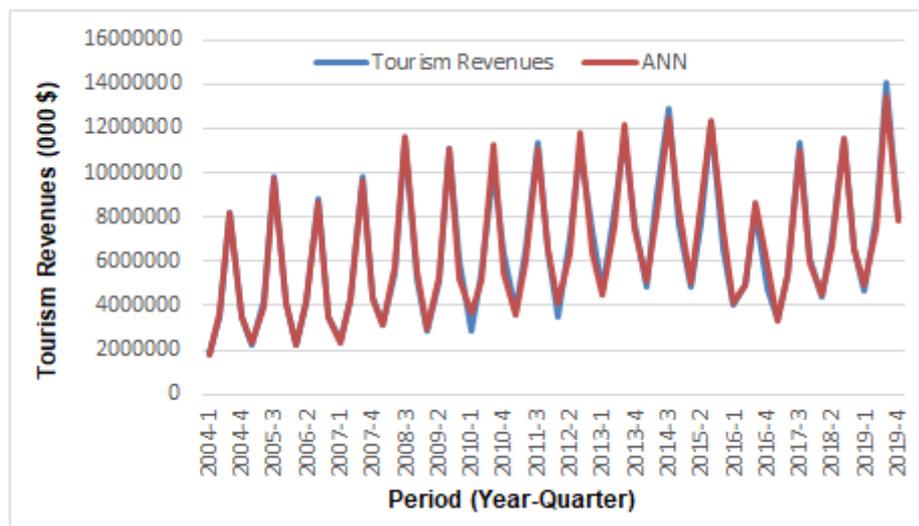


Figure 5. *Time Graph of Tourism Revenues Series and ANN Forecasts*

The time graph of the original tourism revenues series with the forecast values generated by MLP-ANN model is given in Figure 5. When the graph is examined, it is seen that the generated forecasting values series is in harmony with the real values series and the deviations remain at a very small level. These results demonstrate that constructed [4:5:1] MLP-ANN model can be successfully applied in forecasting future (ex-ante) tourism revenues in Turkey.

## DISCUSSION AND CONCLUSIONS

International tourist arrivals to a country increase its foreign currency supply. As a result, the balance of payments deficits diminish, the value of the national currency increases, leading to a revival of other sectors, as tourism revenues change hands within the economy. Tourism sector is one of the most important sectors in countries with high revenue potential. Despite the various crises and adversities experienced, tourism is among the sectors that have been growing rapidly all over the world in recent years. Today, many developed or developing countries are in a fierce competition in order to attract international visitors, gain foreign currency, increase international tourism revenues and open new business opportunities in an increasing global competitive environment. As in many developing countries, the tourism industry in Turkey is one of the leading sectors of the economy. Tourism as a source of foreign exchange in Turkey is important in terms of being both direct and indirect employment provider. Therefore, the need for accurate tourism revenue forecasting is

particularly important due to the industry's contribution to the economy. Estimates of the tourism market and revenues with scientific methods play a guiding role in the decision-making processes of managers at all levels related to tourism. In fact, forecasting is a significant part of the whole planning process in the tourism sector. Accurate forecasting of tourism revenues is also important for the tourism industry because it can help reduce risk and uncertainty as well as effectively provide basic information for better tourism planning. In this study it is aimed to develop optimal forecasting model that yields the highest accuracy when compared the forecast performances of various models. Due to the perishable nature of tourism services, accurate and reliable forecasts are of great importance for stakeholders and decision makers in the tourism industry. However, forecasting is often carried out under uncertain conditions, so modelling and forecasting developments in tourism is a difficult task for both practitioners and academics.

The epidemic of “New Type Coronary Virus (Covid-19)”, which emerged in China and spread throughout the world, started to have profound effects on the tourism sector. 209 countries in the world imposed travel restrictions as a precaution against the epidemic between January 2020 and 6 April 2020, this number corresponds to 96% of all destinations in the world (UNWTO, 2020b). Travel restrictions have brought tourism and related transportation activities to a halt. In the press release of the UNWTO on 7 May 2020, it was stated that the number of tourists in the world decreased by 22 percent in the first three months of 2020 compared to the same period of the previous year, and this decline may vary between 60 and 80% by the end of 2020. Tourism sector, which started hopefully in 2020, is one of the sectors that are most affected by the epidemic of COVID-19, which emerged in China and spread all over the world in a short time. The tourism sector has a critical importance in terms of both the employment it provides and the increase in foreign exchange reserves needed. Managers operating in tourism sector are in a difficult period in terms of determining the operational strategies for the next period, as well as cancellations, personnel management and financing configurations. The impact of COVID-19 on both revenues and supply chains is enormous and immeasurable. Considering that the share of tourism revenues in the national income in 2019 is 4% on average, it is likely that negative developments in tourism will have a direct impact on growth. Therefore, future tourism forecasts can be revised and used by experts in the light of possible scenarios envisaged by UNWTO.

## LIMITATIONS AND IMPLICATIONS

This study has some limitations that allow further research. First, in the tourism demand forecasting models constructed, only monthly tourism income data and its lagged values were used without considering other important factors affecting tourism demand. Therefore, subsequent studies can examine the factors that affect tourism revenues and whether integrating internet-based Big Data into the forecasting models will improve forecast accuracy. Second, as with most tourism demand forecasting and modeling studies, the research is based on a case study for a single country. Therefore, the findings of this study should not be generalized. The aim of the study is to show the potential benefit of using Artificial Neural Networks in tourism revenue estimates comparatively. The study is explorative by its nature. More research needs to be done by including more than one case in order to obtain a more generalizable result.

It is possible to say that the results of the study have practical contribution, but some areas are open for future research. Prospectively, comparative studies in which various methods of Artificial Intelligence such as Adaptive Neural Fuzzy Inference System (ANFIS), Support Vector Machines (SVMs), Rough Sets, Fuzzy Logic and Genetic Algorithms together with the methods used in this study can be proposed to researchers. Additionally, comparing forecasting accuracy with other forecasting techniques and other accuracy measurement dimensions can yield beneficial results. Considering the limited number of studies on modelling and forecasting tourism revenues, it can be said that the proposed studies will form the basis for the planning activities of the practitioners and public administrators operating in tourism sector and will contribute to filling the gap in this field.

## REFERENCES

- Akay, G. H., Cifter, A., & Teke, O. (2017). Turkish tourism, exchange rates and income. *Tourism Economics*, 23(1), 66–77.
- Aktaş, A., Ozkan, B., Kaplan, F., & Brumfield, R. (2014). Exchange rate volatility: Effect on Turkish tourism incomes. *Management Studies*, 2(8), 493-499.
- Anvari, S., Tuna, S., Canci, M., & Turkay, M. (2016). Automated Box–Jenkins forecasting tool with an application for passenger demand in urban rail systems. *Journal of Advanced Transportation*, 50, 25-49.
- Bayramoğlu, M. F., & Başarir, Ç. (2018). International diversified portfolio optimization with artificial neural networks: An application with foreign companies listed on NYSE. In D. Kumar G. (Ed.), *Machine Learning Techniques for Improved Business Analytics*, (pp. 201-223). IGI-Global.

- Çalışkan, U., Saltık, I. A., Ceylan, R., & Bahar, O. (2019). Panel cointegration analysis of relationship between international trade and tourism: Case of Turkey and silk road countries. *Tourism Management Perspectives*, 31, 361-369.
- Chen, B. F., Wang, H. D., & Chu, C. C. (2007). Wavelet and artificial neural network analyses of tide forecasting and supplement of tides around Taiwan and South China Sea. *Ocean Engineering*, 34(16), 2161-2175.
- Egrioglu, E., Aladag, Ç. H., Yolcu, U., Baş, E., & Dalar, A. Z. (2017). A new neural network model with deterministic trend and seasonality components for time series forecasting. In Ç. H. Aladağ (Ed.), *Advances in Time Series Forecasting*, (Vol. 2) (pp. 76-92). Sharjah, UAE: Bentham Science Publishers.
- Ertugrul, H. M., & Mangir, F. (2015). The tourism-led growth hypothesis: empirical evidence from Turkey. *Current Issues in Tourism*, 18(7), 633-646.
- Golam, K., & Hasin, M. A. H. (2013). Comparative analysis of artificial neural networks and neuro-fuzzy models for multicriteria demand forecasting. *International Journal of Fuzzy System Applications*, 3(1), 1-24.
- Gökovalı, U. (2010). Contribution of tourism to economic growth in Turkey. *Anatolia: An International Journal of Tourism and Hospitality Research*, 21(1), 139-153.
- Gunduz, L., & Hatemi-J, A. (2005). Is the tourism-led growth hypothesis valid for Turkey?. *Applied Economics Letters*, 12(8), 499-504.
- Höpken, W., Eberle, T., Fuchs, M., & Lexhagen, M. (2020). Improving tourist arrival prediction: A big data and artificial neural network approach. *Journal of Travel Research*, 1-20, DOI: 10.1177/0047287520921244
- Hüseyini, İ., Doru, Ö., & Tunç, A. (2017). The effects of tourism revenues on economic growth in the context of neo-classical growth model: In the case of Turkey. *Ecoforum*, 6(1), 1-6.
- Kaplan, F., & Aktas, A.R. (2016). The Turkey tourism demand: A gravity model. *The Empirical Economics Letters*, 15(3), 265-272.
- Kara, A., Tarim, M., & Tatoglu, E. (2003). The economic, social and environmental determinants of tourism revenue in Turkey: Some policy implications. *Journal of Economic and Social Research*, 5(2), 61-72.
- Kon, S. C., & Turner, W. L. (2005). Neural network forecasting of tourism demand. *Tourism Economics*, 11(3), 301-328.
- Law, R., & Pine, R. (2004). Tourism demand forecasting for the tourism industry: A neural network approach. In G.P. Zhang (ed.), *Neural Networks in Business Forecasting*, (pp. 121-141). Idea Group Publishing, Hershey: PA.
- Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. London. UK: Butterworth Scientific.
- Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and Applications* (3rd Edition). New York, USA: John Wiley and Sons.
- Ministry of Culture and Tourism (2020). *Tourism Statistics 2019: General Overview*. Retrived 16 June, 2020, from <https://testsite.ktb.gov.tr/kultursurasi/Eklenti/69320,turizmistatistikleri2019-4pdf.pdf?0>
- Moeeni, H., & Bonakdari, H. (2016). Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model. *Stochastic Environmental Research Risk Assessment*, 31(8), 1997-2010.
- Moreno, J. J. M., Poll, A. P., & Gracia, P. M. (2011). Artificial neural networks applied to forecasting time series. *Psicothema*, 23(2), 322-329.

- Önder, I. (2017). Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities. *International Journal of Tourism Research*, 19, 648–60.
- Ongan, S., Işık, C., & Özdemir, D. (2017). The effects of real exchange rates and income on international tourism demand for the USA from some European Union Countries. *Economies*, 51(5), 1-11.
- Öztük, İ., & Acaravcı, A. (2009). On the causality between tourism growth and economic growth: Empirical evidence from Turkey. *Transylvanian Review of Administrative Sciences*, 25E, 73-81.
- Payne, J. E., & Mervar, A. (2002). A note on modelling tourism revenues in Croatia. *Tourism Economics*, 8(1), 103–109.
- Peng, B., H. Song, H., & Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, 45, 181–93.
- Qin, Y., Luo, Y., Zhao, Y., & Zhang, J. (2018). Research on relationship between tourism income and economic growth based on meta-analysis. *Applied Mathematics and Nonlinear Sciences*, 3, 105–114.
- Smith, K. A. (2002). Neural networks for business: An introduction. In K. A. Smith & J. T. D. Gupta (Eds.), *Neural Networks in Business: Techniques and Applications* (pp. 1-25). Idea Group Publishing.
- Song, H., & Li, G. (2008). Tourism demand modelling and forecasting: A review of recent research. *Tourism Management*, 29, 203–20.
- Teixeira, J. P., & Fernandes, P. O. (2012). Tourism time series forecast - different ANN architectures with time index input. *Procedia Technology*, 5, 445–454.
- UNWTO-World Tourism Organization (2020a). *World Tourism Barometer*, January, 18(1).
- UNWTO-World Tourism Organization (2020b). *COVID-19 Related Travel Restrictions. A Global Review for Tourism*. First Report as of 16 April 2020.
- United Nations (2010). *International Recommendations for Tourism Statistics 2008*, Department of Economic and Social Affairs, New York.
- Wong, B. K., Jiang, L., & Lam, J. (2000). A bibliography of neural network business application research: 1994-1998. *Computers and Operations Research*, 27(11), 1045-1076.
- WTTC-World Travel and Tourism Council (2020). *Travel, & Tourism Economic Impact Report*.
- Wu, T. P., & Wu, H. C. (2018). The influence of international tourism receipts on economic development: Evidence from China's 31 major regions. *Journal of Travel Research*, 57(7), 871–882.
- Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2), 502-514.
- Zhang, Y., Li, G., Muskat, B., & Law, R. (2020). Tourism demand forecasting: A decomposed deep learning approach. *Journal of Travel Research*, 1-17. DOI: 10.1177/0047287520919522