Modelling and Forecasting International Tourism Demand – Evaluation of Forecasting Performance

Maja Mamula[1]

[1] Faculty of Tourism and Hospitality Management, Opatija, Croatia

Correspondence: Maja Mamula, Faculty of Tourism and Hospitality Management, Primorska 42, 51410 Opatija, Croatia. Tel: 385-51-294-684. E-mail: majam@fthm.hr

Received: April 13, 2015 Accepted: May 11, 2015 Online Published: May 17, 2015

doi:10.5430/ijba.v6n3p102 URL: http://dx.doi.org/10.5430/ijba.v6n3p102

Abstract

The paper examines the forecasting accuracy of different forecasting techniques in modelling and forecasting international tourism demand in Croatia. As tourist arrivals is the most commonly used measure of international tourism demand, the realized number of German tourists arrivals in the period from first quarter of 2003 to the last quarter of 2013 is taken as a measure of tourism demand in Croatia. In this paper following forecasting techniques are compared: the seasonal naïve model, the Holt-Winters triple exponential smoothing, the seasonal autoregressive integrated moving average model (SARIMA) and the multiple regression model. After approaching the forecasting procedure, all models are compared considering the *in sample* and the *out of sample* mean absolute percentage error (MAPE). All compared models show good forecasting performances. Although the diagnostics for the selected models reveals that the four models do not significantly differ, it can be concluded that multiple regression model perform a highly accurate forecasting of German tourists arrivals in Croatia.

Keywords: international tourism demand, tourist arrivals, forecasting, seasonal naïve model, Holt-Winters model, regression analysis, forecasting accuracy, MAPE

1. Introduction

In 2013, in the Republic of Croatia were realized 12 441 476 tourist arrivals and 64 827 814 overnight stays. Of the total number of tourist arrivals 88.1% were foreign guests, and 11.9% domestic, while of the total number of overnight stays 92.1% were foreign guests, and the remaining 7.9% domestic.

Croatian tourism statistics indicates that apart from a very pronounced seasonal character, tourism demand in the Republic of Croatia is affected by the dominant number of foreign tourists. In the overall structure of foreign tourist arrivals, German tourists are leading, with more than 18% of arrivals in 2013. German tourists realized the highest average length of stay, also, even 7.5 days.

It is obviously that the German tourism flows to Croatia represent a significant source of profit for the tourism sector in the Republic of Croatia. But, the fact is that there is a lack of systematic research of international tourism markets, especially German tourism market. Such, more comprehensive, detailed and systematic studies are useful in establishing of future macroeconomic and microeconomic development and tourism strategies in Croatia, which is a predominantly tourism oriented country.

Due to the importance of German tourism demand for the Croatian tourism sector, this research attempt to present and emphasize the necessity of implementation of quantitative methods, both time series and econometric analysis, in modelling and forecasting tourism demand and all its components.

Keeping in mind mentioned above and taking into account the complexity of tourism demand, as well as all its determinants, and the number of available forecasting techniques, the main hypothesis of this study can be highlighted; in assessing and forecasting the phenomenon being analysed, it is necessary to know and take into account all characteristics of the phenomenon. This approach ensures the implementation of adequate forecasting tools, which would result in accurate forecast values.

2. Theoretical Background and Literature Review

The most detailed review of forecasting research and studies gave Song and Li (2008). They reviewed 121 papers on domestic and international tourism demand modelling and forecasting which were published after 2000. Out of these

121 studies, 72 used time-series techniques to model demand for tourism, and more than 30 of them applied both the time series and econometric approaches in estimating the tourism demand models, and compared the forecasting performances of these models. The most of this studies paid particular attention to exploring the historical trends and patterns, such as seasonality, of the time series involved and to predict future values of this series based on the trends and patterns identified in the model.

Burger, et al. (2001) made a guideline for tourism forecasters for those who don´t have large datasets for creating structural models. Authors used the naïve methods, moving average, decomposition, single exponential smoothing, ARIMA, multiple regression, genetic regression and neural networks in tourism demand modelling and forecasting and then compared actual and predicted number of visitors.

Diviskera (2003) developed a model for international tourism demand based on the consumer theory of choice, reflecting the diversity of tourist preferences. ″The study has generated substantial new information on the effects and sensitivity of economic parameters on international tourism. Models which were estimated proved to be theoretically consistent and the derived elasticity estimates are statistically sound with empirically plausible magnitudes.″

Baldigara (2013) compared five time series forecasting methods in international tourism demand forecasting; the naïve 2, the double moving average with linear trend, the double exponential smoothing, the linear trend and the autoregressive method. That research showed that all used models have good forecasting performances, but the double moving average method performed the best forecasting performance due to the smallest mean absolute percentage error.

The problem of seasonality of Croatian tourism was analysed by Kožić (2013). He concluded that high seasonality of Croatian tourism is often emphasised as main undesirable characteristic of tourism demand, and that dealing with seasonality is the strategic aim of the Croatian tourism.

Considering that the Republic of Croatia is a tourism oriented country, where the most of the tourist traffic makes international tourist traffic, especially the arrivals and overnight stays of German tourists, and having in mind the lack of research dealing with modelling, forecasting and evaluating the seasonal phenomenon, given literature review highlights the lack of such research in Croatia. The research, conducted in Croatia and beyond, raise many questions and can be considered as the starting point in order to select appropriate models for modelling and forecasting Croatian tourism demand.

3. Data and Methodology

″Tourist arrivals is the most commonly used measure of international tourism demand, followed by tourist expenditure and tourist nights in registered accommodation (Song, et. al., 2012, 2).″ This study considers German tourists arrivals as a measure of international tourism demand in Croatia. Figure 1. shows the actual data used in the models estimation process.

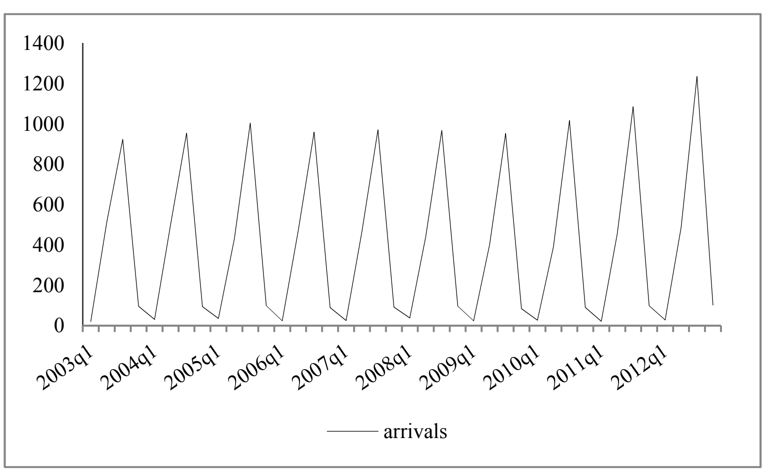

Figure 1. German tourists arrivals from 2003 – 2012 (in thousands)

The following figure shows the forecasting time horizon.

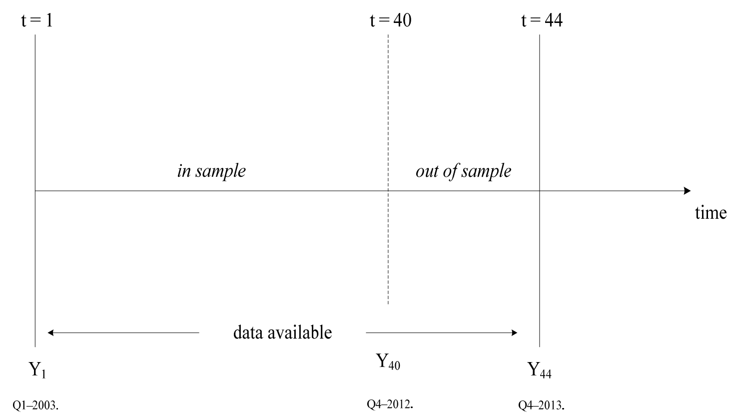

Figure 2. Time horizon of forecast

As shown in the figure above, the model estimation was based on the quarterly data of the realized number of German tourists arrivals in the Republic of Croatia in the period from the first quarter of 2003 to the last quarter of 2012 (*in sample data*). Forecasting for the period from the first to the last quarter of 2013 (*out of sample data*) was made on estimated model. Data were taken from the Croatian Bureau of Statistics database and publications Statistical Yearbook and First releases.

In modelling and forecasting of German tourists arrivals in the Republic of Croatia several statistical methods of time series analysis and econometric methods were used. Regardless Kožić, et. al (2013, 174) pointed out the smallest impact of seasonality in Croatia is obvious in domestic tourists arrivals, followed by tourists from Germany and Austria. Figure 1 shows that the realised number of German tourists arrivals in Croatia in the observed period indicate a pronounced seasonal character of the German tourist demand. For this reason, in this research models which take into consideration the seasonal nature of the phenomenon were used — the seasonal naïve model, the Holt-Winters model, the seasonal autoregressive integrated moving average model and the regression model

3.1 The Seasonal Naïve Model

″The seasonal naïve can be used with seasonal data and postulates that the next period´s value is equal to the same period in the previous year (Fretchling, 2001, 66).″

$$\hat{Y}_t = Y_{t-m} \tag{1}$$

where:

- Y – actual value
- $\hat{Y}$ – forecast value
- t – some time period
- m – number of period in a year

Seasonal naïve model is the basic kind of extrapolation of historical data, also called no-change model. No-changing models are used frequently in seasonal phenomenon forecasting. This model takes into account seasonality by using the value for the same time period in the following year, which is better than naïve 1 model which assumes that there is no change at all between time periods.

3.2 The Holt-Winters Model

In the Holt-Winters model a new estimate for dependent variable is the combination of the estimate for the present time period plus a portion of the random error generated in the present time period. When used for forecasting Holt-Winters model, also called triple exponential smoothing, uses weighted averages of past data. The Holt-Winters model employs triple exponential smoothing; one equation for the level, one for trend and one for the seasonality. Equations associated to each of these elements are as follows:

Level: $$L_t = \alpha \cdot \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{2}$$

Trend: $$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \tag{3}$$

Seasonal: $$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s} \tag{4}$$

Forecast: $$\hat{Y}_{t+h} = (L_t + hb_t)S_{t-s+h} \tag{5}$$

where:

Y – actual value
L – level of the series
S – seasonal component
α – level smoothing constant between 0 and 1
b – trend of the series
s – number of seasonal period in the year
t – some time period

Initial values can be calculated as follows:

Initial level: $$L_t = \alpha \cdot \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{6}$$

Initial trend: $$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \tag{7}$$

Seasonal indices: $$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s} \tag{8}$$

The smoothing constants determine the sensitivity of forecast to changes in tourism demand. According to Handanhal (2013, 348) ″large values of α make forecasts more responsive to more recent values, whereas smaller values have a dumping effect. Large values of β have a similar effect, emphasizing recent trend over older estimates of trend.″ There is no exact rule for constant determination; some literature provides general recommendations. Schroeder, et. al. (2013, 261) suggest value of α between 0.1 and 0.3, while Stevenson (2012, 87) suggests a wider range; from 0.05 to 0.5. Most of literature also recommends that smoothing constant can be chosen so that forecast is more accurate, with accuracy measured by some forecasting error.

3.3 The Seasonal Autoregressive Integrated Moving Average Model

″The Box-Jenkins methodology is based on the assumption that the underlying time series is stationary or can be made stationary by differencing it one or more times. This is known as the ARIMA (p,d,q) model, where *d* denotes the number of times a time series has to be differenced to make it stationary (Gujarati, 2015, 303). For seasonal time series literature suggests seasonal autoregressive integrated moving average models, also called SARIMA (p,d,q) (P,D,Q)s models. The general form of seasonal model can be given by:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Y_t = \Theta_0 + \theta(B)\Theta(B^s)\varepsilon_t \tag{9}$$

The general form from (9) can be rewritten as follows:

$$\begin{aligned} &\underbrace{\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)}_{AR(p)}\underbrace{\left(1-\Phi_1 B^s-\Phi_2 B^{2s}-\cdots-\Phi_P B^{Ps}\right)}_{AR_s(P)}\underbrace{(1-B)^d}_{I(d)}\underbrace{(1-B^s)^d}_{I_s(D)} Y_t \\ &= \Theta_0 + \underbrace{\left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right)}_{MA(q)}\underbrace{\left(1-\Theta_1 B^s-\Theta_2 B^{2s}-\cdots-\Theta_Q B^{Qs}\right)}_{MA_s(Q)}\varepsilon_t \end{aligned} \tag{10}$$

where:

$AR(p)$ – autoregressive component of order *p*
$MA(q)$ – moving average component of order *q*

$AR_s(P)$ – seasonal autoregressive compoeenent of order P

$MA_s(Q)$ – seasonal moving average component of order Q

$I(d)$ – integrated component of order d

$I_s(D)$ – seasonal integrated component of order D

The Box Jenkins methodology follows a four-step procedure:

– identification
– estimation
– diagnostic checking, and
– forecasting stage.

Once a particular seasonal autoregressive integrated moving average model is fitted, it can be used for forecasting.

3.4 The Regression Model

Regression analysis is concerned with how one or more variables affect the dependent variable. Trying to understand tourism demand as the dependent variable, on the basis of one independent variable would certainly create inaccurate an unreliable predictions, and the way to deal with this kind of problem is to include many more factors in regression analysis, known as the multiple regression model. The general form of a linear regression model is:

$$Y_t = \beta_0 + \beta_1 X_t + u_t \tag{11}$$

where:

Y – dependent variable

X – independent variable

β_0 – the intercept constant

β_1 – slope coefficient

u – residual

t – some time period

The objective in forecasting is to derive sound estimates of the coefficients of parameters so the forecast variable based on the values of explanatory variables can be estimated. Compared to the seasonal naïve model, the Holt-Winters model and the seasonal autoregressive integrated moving average model, regression models have several following advantages (Fretchling, 2001, 142-143):

– It explicitly addresses causal relationships that are evident in the real world.
– It aids assessment of alternative business plan.
– It provides several statistical measures of accuracy.
– Regression models accommodate a wide range of relationships.

As a general statistical procedure known as the general linear modelling, multiple regression models take many different forms. In practice, the estimated model should be correctly specified in terms of functional form, and should be able to compete all rival models. Generally, classical linear regression model makes the following assumptions (Gujarati, 2015, 8):

– The regression mode is linear in the parameters as in (11); it may or may not be linear in the variable Y and the Xs.
– The regressors are assumed to be fixed or nonstohastic in the sense that their values are fixed in repeated sampling.
– Given the values of the X variables, the expected value, or mean, of the error term is zero.
– The variance of each u_i, given the values of X, is constant, or homoscedastic.
– There is no correlation between two error terms. That is, there is no autocorrelation.
– There are no perfect linear relationships among the X variables. This is the assumption of no multicolinearity.
– The regression model is correctly specified.

On the basis of the above mentioned assumptions, the most popular method for parameters estimation — Ordinary Least Squares, provides estimators which have several desirable statistical properties, such as (Verbeek, 2012, 16-19):

– The estimators are linear, which means that they are linear functions of dependent variable Y.
– The estimators are unbiased, which means that in repeated applications of the method, on average, they are equal to their true values.
– The estimators are efficient, which means that they have minimum variance.

3.5 Measures of Forecasting Accuracy

The main purpose of whole modelling and forecasting process is to clearly discern the future values of tourism demand, and the most important criterion of all is how accurately a model does this. According to Fretchling (2001, 23) ″the most familiar concept of forecasting accuracy is called ′error magnitude accuracy′ and relates to forecast error with a particular forecasting model.″ This is defined as:

$$e_t = A_t - F_t \tag{12}$$

where:

t − some time period
e − forecast error
A − actual value
F − forecast value

Although there are a number of forecasting errors that can be used for accuracy evaluation, in this paper mean absolute percentage error (MAPE) is used. The mean absolute percentage error (MAPE) is expressed in generic percentage terms and it is computed by the following formula:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|(A_t - F_t)|}{A_t}\cdot 100 \tag{13}$$

The reason why MAPE is considered as good accuracy measure is that this measure doesn′t depend on the magnitudes of the demand variables being predict. According to Baggio and Klobas (2011, 151) ″a rough scale for the accuracy of a model can be based on MAPE″ following the suggestions given in the table below.

Table 1. Forecasting accuracy according to MAPE

MAPE	Forecasting accuracy
less than 10%	highly accurate
10-20%	good
20-50%	reasonable
greater than 50%	inaccurate

Source: Baggio R., and Klobas J. (2011, 151)

4. Results and Discussion

In this section the comparison of the selected models in forecasting Germans tourism demand in Croatia, expressed in Germans tourist arrivals, is made and the obtained results are discussed.

Firstly, seasonal naïve model was used for forecasting German tourist arrivals in the Republic of Croatia in observed period according to following equation:

$$\hat{Y}_t = Y_{t-4} \tag{14}$$

The Holt-Winters model, also known as triple exponential smoothing model, was used because of the upward trend and seasonal impact present in the time series. Smoothing constants were chosen according to following rule: different values of smoothing constants were tried out on past data and as the best one were chosen those constants which made minimum mean absolute percentage error. The parameter α smoothed the level equation and its value was set up at 0.3, the parameter β smoothed the trend equation and its value was set up at 0.4, and the parameter γ smoothed the seasonal equation and its value was set up at 0.4.

The Box-Jenkins method is developed throughout four interactive steps — the model identification, the parameter estimation, the diagnostic checking and the forecasting stage. As the Box-Jenkins methodology requires the time series to be stationary in its mean and variance; the empirical time series was tested for stationary. The analysis of the empirical autocorrelation functions in lags $k = 4,8,12,\ldots$ shows that the value of the autocorrelation function is significant for lag $k = 4$; therefore, it is evident that the time series of German tourists arrivals is seasonal nonstationary, and that a seasonal autoregressive model of order 4 should be appropriate. In order to eliminate the nonstationary, the time-series variance should be stabilised. The original time-series is therefore seasonally differenced and the stationary test is performed again. The seasonal differencing has removed quite all significant autocorrelations. The most appropriate seasonal autoregressive integrated moving average model (SARIMA) was identified in order to model and forecast the empirical data. Several models were computed and only models with statistically significant coefficient at 5% level were chosen, ensuring the normality and the non-autocorrelation of residuals at 5%. As the most appropriate model was identified seasonal $\mathrm{ARIMA}(0,0,0)(1,1,3)_4$. This model presented the smallest Akaike info criterion, Schwarz criterion and Hannan-Quinn criterion and the smallest mean absolute percentage error. The selected model can be written as follows:

$$(1-\Phi_1 B^4)(1-B^4)Y_t = (1-\Theta_1 B^4 - \Theta_2 B^8 - \Theta_3 B^{12})\varepsilon_t \tag{15}$$

where:

Y_t – the tourists' arrivals

B – the backshift operator

ε_t – the random noise

The estimation of Equation (15) gives: (Note 1)

$$(1-0.9098B^4)(1-B^4)Y_t = (1+0.3648B^4 - 0.4437B^8 + 0.8636B^{12})\varepsilon_t \tag{16}$$

$t =$ 4.5268 (0.0001)   −2.4473 (0.0209)   3.4485 (0.0018)   −14.8470 (0.0000)

$R^2 = 0.5304 \qquad \sigma = 27183.69 \qquad SSR = 2.07e^{10}$

$\chi^2_{AUTO}(4) = 2.7996$ (0.5919) $\qquad \chi^2_{WHITE}(4) = 3.1795$ (0.5282) $\qquad \chi^2_{NORM}(2) = 0.1405$ (0.9322)

where:

t – the t-statistics

R^2 – the adjusted coefficient of determination

σ – the standard error of model

SSR – the sum squared residuals

$\chi^2_{\mathrm{AUTO}}(4)$ – the Breusch-Godfreyev LM test for autocorrelation

$\chi^2_{\mathrm{WHITE}}(4)$ – the White test for heteroscedasicity

$\chi^2_{\mathrm{NORM}}(2)$ – the Jarque-Bera normality test

The adjusted R^2 of the model is 0.5304 and shows a quite good model fitting. In diagnostic checking stage model is tested for stationary and invertibility using the inverted AR and MA roots. As all the absolute values of the inverted AR and MA roots are smaller than one, it can be concluded that the estimated model is stationary and invertible. The estimated model is tested also for the residuals autocorrelation, the normality of residuals and the presence of heteroscedasticity. All the performed diagnostic statistics show that the model passes all the tests.

According to existing literature, as well as the availability of data, for the purposes of this research, in regression analysis of tourism demand of German tourists in Croatia following variables were selected: number of German tourists arrivals in the Republic of Croatia as dependent variable, and number of German tourists departures abroad, number of German tourists departures abroad in previous period, price variable (real costs of tourism services in the Republic of Croatia), and seasonal dummy variables (dummy variable 1, dummy variable 2 and dummy variable 3).

Before approaching the modelling and diagnostic testing, the dependent and the explanatory variables were retested for stationary. As variables weren't stationary, in order to stabilize the variance and obtain a stationary time series, all variables were logarithmic transformed. As logarithm values of dependent and independent variables were stationary, we were able to access the modelling process. The data generating process is therefore a model as follows:

$$\mathrm{lnARRIVALS} = \beta_0 + \beta_1 \mathrm{lnDEPARTURES} + \beta_2 \mathrm{lnDEPARTURES}_{t-1} + \beta_3 \mathrm{lnRCPI} + \beta_4 \mathrm{D}_2 + \beta_5 \mathrm{D}_3 \text{(Note 2)} \tag{17}$$

The OLS estimation of Equation (17) gives:

$$lnARRIVALS = -80.2062 + 2.5744\ lnDEPARTURES + 1.8879\ lnDEPARTURES_{t-1} \tag{18}$$

$t =$ −7.0507 4.2635 3.2105

(0.0000) (0.002) (0.0030)

$$-6.7012\ lnRCPI + 2.5191\ D_2 + 2.3603\ D_3$$

$t =$ −5.1723 9.4264 10.1898

(0.0000) (0.0000) (0.0000)

$R^2 = 0.9605$ $F = 185.7066$ $\sigma = 0.2777$ $SSR = 2.5449$

$\chi^2_{AUTO}(4) = 10.7746$ $\chi^2_{WHITE}(5) = 8.0409$ $\chi^2_{NORM}(2) = 1.8636$

(0.0292) (0.1540) (0.4309)

where:

- t – the t-statistics
- R^2 – the adjusted coefficient of determination
- F – the F-statistics
- σ – the standard error of model
- SSR – the sum squared residuals
- $\chi^2_{\mathrm{AUTO}}(4)$ – the Breusch-Godfreyev LM test for autocorrelation
- $\chi^2_{WHITE}(5)$ – the White test for heteroscedasicity
- $\chi^2_{NORM}(2)$ – the Jarque-Bera normality test

The estimation results show that model is well specified; the coefficient of all the variables have ″correct″ signs and are statistically significant at 5% level. After model estimation, assumptions of classical linear regression were tested. The performed tests indicated that there is presence of seasonal autocorrelation in estimated model. Knowing the consequences of the seasonal autocorrelation, the model is transformed using the Cochrane-Orcutt two-step procedure. After transformation, the following model was obtained:

$$lnARRIVALSCO = -112.7265 + 3.1673\ lnDEPARTURESCO + 1.8479\ lnDEPARTURES_{t-1}CO \tag{19}$$

$t =$ −3.3350 4.0026 3.0233

(0.0000) (0.004) (0.0052)

$$-8.0205\ lnRCPICO + 2.3833\ D_2CO + 2.0998\ D_3CO$$

$t =$ −5.3344 8.7167 8.0828

(0.0000) (0.0000) (0.0000)

$R^2 = 0.9090$ $F = 68.9839$ $\sigma = 0.2263$ $SSR = 1.4859$

$\chi^2_{AUTO}(4) = 6.2206$ $\chi^2_{WHITE}(5) = 4.7608$ $\chi^2_{NORM}(2) = 0.0431$

(0.1833) (0.4458) (0.9786)

where:

t	–	the t-statistics
R^2	–	the adjusted coefficient of determination
F	–	the F-statistics
σ	–	the standard error of model
SSR	–	the sum squared residuals
$\chi^2_{\text{AUTO}}(4)$	–	the Breusch-Godfreyev LM test for autocorrelation
$\chi^2_{WHITE}(5)$	–	the White test for heteroscedasicity
$\chi^2_{NORM}(2)$	–	the Jarque-Bera normality test

The transformed model fits the data well with high adjusted R^2 (R^2=90.90). Overall, the estimated demand model can be considered as well specified. All explanatory variables are consistent and significant at 5% level, and estimated parameters sign are correct and consistent with economic theory as expected. From the diagnostic statistics we can see that the model is well specified; the assumptions of classical linear regression model are not violated.

After modelling, and *in sample* as well as *out of sample* forecasting, forecasted values were compared to actual data, as shown on the figure below.

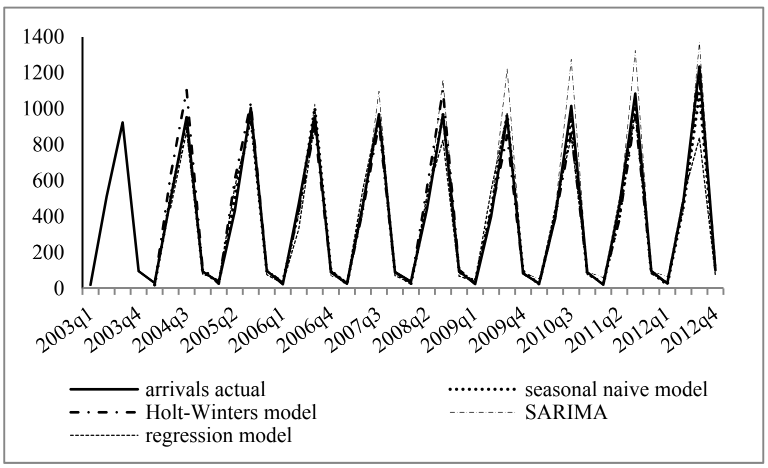

Figure 3. Actual and forecast value of German tourist arrivals

As shown on the figure above, the forecasts follow quite well the overall trend of the time series and show good correspondence with empirical data. After modelling and forecasting of German tourist arrivals in Republic of Croatia the *in sample* and the *out of sample* mean absolute percentage error of the selected forecasting techniques were calculated.

Table 2. Mean absolute percentage error of selected forecasting techniques

	seasonal naïve model	Holt-Winters model (α=0.3, β=0.4, γ=0.5)	SARIMA $(\mathbf{0,0,0})(\mathbf{1,1,3})_{\mathbf{4}}$	regression model
***in sample* MAPE**	20.217	19.280	26.135	1.689
***out of sample* MAPE**	23.912	7.603	3.801	1.672

The calculated error indicates that when examining the *out of sample* MAPE three of four forecasts are highly accurate; the Holt-Winters (MAPE=7.603%), the seasonal autoregressive integrated moving average (MAPE=3.801%) and the regression model (MAPE=1.672%).

The smallest mean absolute error is performed by the regression model (*in sample* MAPE=1.689%, and *out of sample* MAPE=1.672%). The reason for this is probably in the fact that estimated model explicitly addresses causal relationship between German tourists arrivals, as dependent variable, and number of German tourists departures abroad, number of German tourists departures abroad in previous period, price variable (real costs of tourism services in the Republic of Croatia), and seasonal dummy variables (dummy variable 2 and dummy variable 3), as independent variables.

As expected, the seasonal naïve model, as the simplest model has the greatest MAPE. Despite the slightly higher value of mean absolute percentage error, the seasonal naïve model can be considered as reasonable model (*in sample* MAPE=20.217%, *out of sample* MAPE=23.912%).

Generally, the most accurate model is the multiple regression model; model which involves a number of variables in the assessment, but also highlights the impact of the seasonal component in a time series through dummy variables. It can be concluded that the knowledge of the phenomenon being analysed is extremely important in order to take into account all the relevant factors in the modeling, and forecasting as well.

5. Conclusions

The aim of this study was to model the international tourist flows from Germany to Croatia and identify some of its key determinants. The presented research has attempted to examine different forecasting methods which take into account seasonal character of analysed phenomenon, and analyse how accurate are they. These forecasting techniques were seasonal naïve model, Holt-Winters triple exponential smoothing, seasonal autoregressive integrated moving average model and multiple regression model. In the forecasting process all models were statistically tested and passed all tests. After conducted research it can be concluded that the knowledge on analysed phenomenon is useful in the process of modelling and forecasting. After forecasting, forecast values were compared considering the *in sample* and the *out of sample* mean absolute percentage error. The comparison pointed out that the multiple regression model is the most accurate. This model was the most complex as well. The number of German arrivals in Croatia is sensitive to the number of German departures abroad (current and lagged values), price variable and seasonal dummy variables. Considering the fact that modelling and forecasting of tourism demand is a challenging topic, that the adequacy and accuracy of a forecasting model is valued according to its *in sample* and *out of sample* forecasts, and that is still difficult to indicate which modelling techniques are the most adequate for tourism demand modelling, the author´s intention was to highlight the necessity of systematic and comprehensive analysis of Croatia's international tourism demand in all its characteristics.

The assessment of adequacy of the estimated model, except forecast accuracy on one side, includes a number of other factors such as forecasting horizon, data frequency and other models included in the forecasting comparison, on the other side. According to this research, the future attention should be paid to further developments of forecasting techniques (especially regression models and its different functional forms), or to combination of different forecasts.

Acknowledgments

The author would like to express its thanks to the Faculty of Tourism and Hospitality Management in Opatija, University of Rijeka for providing facilities, as well as moral and financial support for this research.

References

Athanasopoulos, G., & Hyndman, R.J. (2008). Modelling and forecasting Australian domestic tourism. *Tourism Management, 29*(1), 19–31.

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

Baldigara, T. (2013). Forecasting Tourism Demand in Croatia: A Comparison of Different Extrapolative Methods. *Journal of Business Administration Research, 2*(1), 84–89. http://dx.doi.org/10.5430/jbar.v2n1p84

Burger, C.J.S.C. (2011). A practioners guide to time–series methods for tourism demand forecasting – a case study of Durban, South Africa. *Tourism Management, 22*(4), 403–409. http://dx.doi.org/10.1016/S0261-5177(00)00068-6

Croatian Bureau of Statistics. (2013). First Releases. Retreived February 15, 2015, from http://www.dzs.hr/Hrv_Eng/publication/2013/04–03–02_11_2013.htm

Croatian Bureau of Statistics. (2013). Statistical Yearbook. Retreived February 15, 2015, from http://www.dzs.hr/Hrv_Eng/publication/2013/04–03–02_11_2013.htm

Divino, A.J., & McAleer, M. (2010). Modelling and forecasting daily international mass tourism to Peru. *Tourism Management, 31*(6), 846–854. http://dx.doi.org/10.1016/j.tourman.2009.09.002

Diviskera, S. (2003). A model of demand for international tourism. *Annals of Tourism Research, 30*(1), 31–49. http://dx.doi.org/10.1016/S0160-7383(02)00029-4

Frechtling, D.C. (2001). *Forecasting Tourism Demand: Methods and Strategies*. Woburn: Butterworth Heinemann, Elseivier group.

Gujarati, D. (2015). *Econometrics by example*. UK: Palgrave Macmillan.

Gunter, U., & Önder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management, 46*, 123–135. http://dx.doi.org/10.1016/j.tourman.2014.06.017

Han, Z., Durbarry, R., & Sinclair, M.T. (2006). Modelling US tourism demand for European destinations. *Tourism Management, 27*(1), 1–10. http://dx.doi.org/10.1016/j.tourman.2004.06.015

Handahal, V.R. (2013). Determining the Optimal Values of Exponential Smoothing Constants. Does Solver Really Work?. *American Journal of Business Education, 6*(3), 347–360.

Kožić, I. (2013). Kolika je sezonalnost turizma u Hrvatskoj?. *Ekonomski vjesnik, 26*(2), 470–480.

Kožić, I., Krešić, D., & Boranić-Živoder, S. (2013). Analiza sezonalnosti turizma u Hrvatskoj primjenom Gini koeficijenta. *Ekonomski pregled, 64*(2), 159–182.

Lin, C.J., Chen, H.F., & Lee, T.S. (2011). Forecasting Tourism Demand Using Time Series, Artificial Neural Network and Multivariate Adaptive Regression Splines: Evidence from Taiwan. *International Journal of Business Administration, 2*(2), 14-24. http://dx.doi.org/10.5430/ijba.v2n2p14

Preez, J., & Witt, S.F. (2003). Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting, 19*(3), 435–451. http://dx.doi.org/10.1016/S0169-2070(02)00057-2

Schroeder, R., Rungtusanatham, M.J., & Goldsteing, S. (2013). *Operations Management in the Supply Chain: Decisions and Cases*. UK: McGraw-Hill Higher Eduction.

Shen, S., Li, G., & Song, H. (2009). Effect of seasonality treatment on the forecasting performance of tourism demand models. *Tourism Economics, 15*(4), 693–708.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting – A review of recent research. *Tourism Management, 29*(2), 203–220. http://dx.doi.org/10.1016/j.tourman.2007.07.016

Song, H., Witt, S., & Gang, L. (2012). *The Advanced Econometrics of Tourism Demand*. London: Routledge.

Stevenson, W. (2012). *Operations Management*. UK: McGraw-Hill Higher Eduction.

Verbeek, M. (2012). *A Guide to Modern Econometrics*. UK; A John Wiley & Sons.

Wong, K.F., & Song, H. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management, 28*(4), 1068–1078. http://dx.doi.org/10.1016/j.tourman.2006.08.003

Notes

Note 1. Values in parentheses are empirical p-values.

Note 2. Since the dummy variable 1 was not statistically significant at 5% level of significance, it was excluded from the model.