

Tea Baldigara, Ph.D., Associate Professor

Marija Koić, graduate student

University of Rijeka, Faculty of tourism and hospitality management in Opatija
UDC 338.482:303.447(497.5)“2003/2013”

MODELLING THE INTERNATIONAL TOURISM DEMAND IN CROATIA USING A POLYNOMIAL REGRESSION ANALYSIS

Abstract: German tourist flows to Croatia represent a significant source of profit for the tourism sector. It is therefore crucial to analyze the determinants and the core patterns of the German tourism demand in Croatia. The paper aims to model international tourism demand in Croatia using the polynomial regression analysis. The second degree polynomial model is esti-

ated in the attempt to model German tourist arrivals and German tourists' overnight stays in Croatia in the period from 2003 to 2013.

Key words: modelling tourism demand, tourist arrivals, tourists' overnight stays, polynomial regression analysis, second-degree polynomial regression model.

Introduction

Tourism is a significant source of profit for a wide range of activities in Croatia. According to the data obtained from the Croatian Bureau of Statistics for the first three months of 2013, the Croatian tourism sector registered an income of 306 million euros, which is 3.1% of the total Croatian GDF. It is well known that Croatia is predominantly an international tourist destination. In fact, according to the Croatian Ministry of Tourism in 2012 there were 62.7 million of tourists' overnight stays registered and among these 57.5 million were realized by foreign tourists. Since 2009 the number of foreign tourist nights has revealed a slightly upward trend with a growth rate of 4.5%. Foreign tourists account for a total of 85% of the total Croatian tourist activity. The German tourist flows to Croatia obviously represent a significant source of profit for the tourism sector. In fact, Germany has traditionally been one of the most important generators of tourist flows in Croatia.

In 2013, most of the foreign tourists came from Germany and they realized 14.4 million of overnight stays, registering a growth rate of 3.5%. German tourists stayed in Croatia on average of 7.5 days and realized 24.2% of the total foreign tourists' overnight stays and 17.9% of the total foreign tourist arrivals. In 2013, there were a total 1.9 million of German tourist arrivals in Croatia. Therefore, it is crucial to analyze the determinants and the core patterns of the German tourist demand for Croatia. Such more comprehensive and detailed studies could be used in the formulation of future macroeconomic development strategies, pricing strategies and tourism sector routing strategies in Croatia as a predominantly tourism-oriented country.

Data and methodology

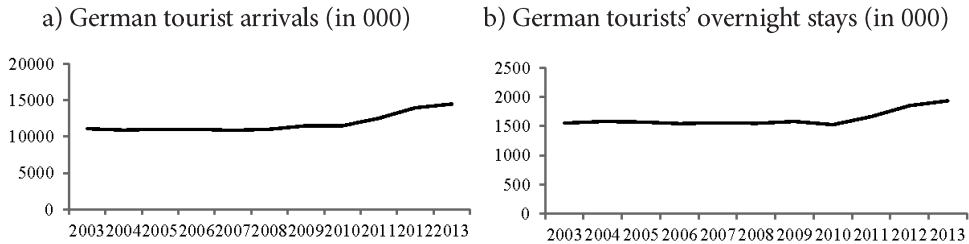
According to Song, Witt and Li (2012), tourist arrivals is the most commonly used



measure of international tourism demand, followed by tourist expenditure and tourists' overnight stays. This study considers German tourist arrivals and German tourists' overnight stays as a measure of international tourism demand in Croatia from 2003 to

2013. The data are provided on the annual base and were collected from the Croatian Central Bureau of Statistics. Figure 1 shows the actual data used in the forecasting model building process.

Figure 1. German tourist arrivals and German tourists' overnight stays



The plots of both German tourist arrivals and German tourists' overnight stays show a slight upward nonlinear trend. Therefore, in the attempt to model the empirical time-series the polynomial regression analysis will be applied. Chu (2004) used a cubic polynomial approach in predicting international tourism arrivals to Singapore and to forecast the volume of tourist arrivals from January 1989 to July 1990.

A polynomial regression is a special case of multiple regression, in which the relationship between the independent variable and the dependent variables is modelled as an n th order polynomial. It fits a nonlinear relationship between the observed variables, but as a statistical estimation problem it is linear, because the regression function is linear in the unknown regression parameters. The goal of regression analysis is to model the expected value of a dependent variable y in terms of the value of an independent variable (or vector of independent variables) x . In general, the expected value of y as an n th order polynomial, yielding the general polynomial regression model can be expressed as it follows:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_n x^n + \varepsilon \quad (1)$$

Where n is some integer.

Polynomial regression is a particular form of a multiple linear regression. A polynomial expression in (1) is an equation that is linear in parameters having an independent variable that is raised to a power included in the model. When using polynomial regression models, the classical assumptions for regression analysis should be applied. In order to obtain valid estimates of the parameters, the following assumptions have to be made (Song, Witt, Li, 2012):

- The expected value of the dependent variable depends on the values of the explanatory variables and the unknown $\hat{\alpha}$ parameters. This is equivalent to $E(e_j) = 0$.
- The sample variance of the dependent variable or the variance of the error term remains constant over time. If this assumption does not hold, the model suffers from the problem of heteroscedasticity.
- Any two observations on the dependent variable or the residuals are not correlated. Violation of this assumption results

in a model with autocorrelated residuals.

- The values of the residuals are normally distributed about their mean: $e_{it} = N(0, \sigma^2)$.
- The values of the explanatory variables are known and there is no linear relationship among the explanatory variables. If this condition is not true, the model suffers from the problem of multicollinearity.

Modelling tourism demand in Croatia using a second-order polynomial regression

The evolution of the annual number of German tourist arrivals as well as German tourists' overnight stays in Croatia in the period from 2003 to 2013, in dependence of time t (the independent variable x) can be approximated with a second-order polynomial (quadratic) function as follows:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \tag{2}$$

Where:

- Y = the dependent variable
- β_0 = intercept constant
- β_1 = linear effect parameter
- β_2 = quadratic effect parameter
- X = time, explanatory variable
- e = residual

The model in (2) is said to be hierarchical because it contains all terms of order two and one. Only hierarchical models are invariant under linear transformation.

Polynomial models can be useful in situations where curvilinear effects are present in the response function and in approximating functions to unknown and possibly very complex nonlinear relationships. All lower order terms must be in the equation in order for the polynomial function to be represented, but the statistical tests should focus on the second order term, as the highest term in the model. In essence, the tests are performed to establish whether the highest term in the model is significantly different from zero and thus whether the polynomial fits the data. There are also arguments from physics which suggest that first-degree and second-degree polynomials in t , which in fact are linear and quadratic time trends, are common in the natural world and therefore such trends might also arise in the social world.

Results and discussion

The second-order polynomial regression model in (2) is fitted using the method of ordinary least squares and the estimated quadratic polynomial trend models for the number of German tourist arrivals (equation 3) as well as for the number of German tourists' overnight stays (equation 4) are obtained:

$$GTA = 1678,873 - 75,98811x + 8,793706x^2$$

(61,99596)(23,74494) (1,927237)

$t = 1$ in 2003
unit for $t = 1$ year
unit for $Y =$ thousands of tourist arrivals

(3)

$R^2_{adj} = 0,83$ $\sigma = 56,45$ $RSS = 25,49$ $\chi^2_{\text{vuto}}(4) = 6,292$ $\chi^2_{\text{norm}}(2) = 1,951$
 $\chi^2_{\text{White}}(4) = 3,203$ $VIF =$



$$\begin{aligned}
 GTN &= 11756,48 - 555,6769x + 72,95338x^2 \\
 &(323,5495)(123,922) (10,058) \\
 &t = 1 \text{ in } 2003 \\
 &\text{unit for } t = 1 \text{ year} \\
 &\text{unit for } Y = \text{thousands of tourist arrivals}
 \end{aligned}
 \tag{4}$$

$$\begin{aligned}
 R^2_{\text{adj}} &= 0,947 & \sigma &= 294,62 & \text{RSS} &= 694388,2 & \chi^2_{\text{vuto}}(4) &= 8,25 & \chi^2_{\text{Norm}}(2) &= 0,038 \\
 \chi^2_{\text{White}}(4) &= 3,68 & \text{VIF} &= 19,46
 \end{aligned}$$

where

- GTA is the annual number of German tourist arrivals
- GTN is the annual number of German tourists' overnight stays
- x is the predictor, time
- σ is the standard error of the regression
- RSS is the sum squared residuals
- $\chi^2_{\text{Norm}}(2)$ is the Jarque-Bera normality test
- $\chi^2_{\text{Autocorr}}(4)$ is the Breusch-Godfrey LM test f or autocorrelation
- $\chi^2_{\text{White}}(4)$ is the White test for heteroscedasticity
- VIF are the Variance Inflation Factors

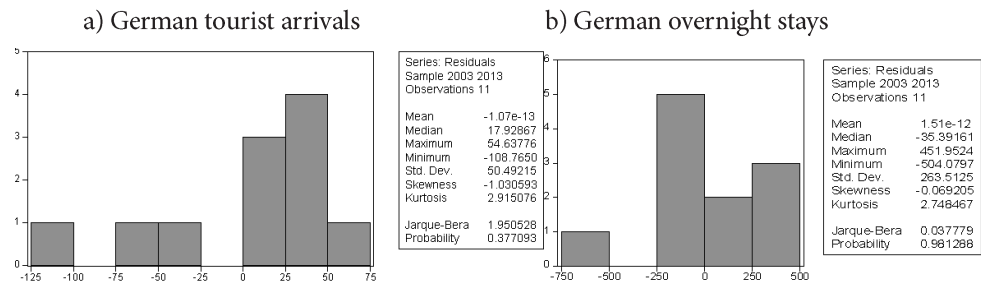
The diagnostic statistics shows that all the coefficient of all dependent variables are

significant at 5% level and have the correct sign.

In the attempt to investigate the performance of the specified model some basic diagnostic statistics are performed.

The residuals of both models are tested for normality. The residuals are assumed to follow the assumptions of a stationary univariate process. They should be white-noise with constant mean and variance ($\epsilon_t \sim \text{WN}(0, \sigma^2)$). The normality test is performed using the Jarque-Bera test statistics, as a goodness-of-fit test of whether sample data have the skewness and kurtosis matching a normal distribution.

Figure 2. Residuals Testing – The Jarque-Berra Test



The computed JB-statistics are 1,950528 for the quadratic polynomial model of German tourist arrivals and 0,037779 for the quadratic polynomial model of German tourists' overnight stays respectively, and are both smaller than the critical value of $\chi^2(2) = 5,99$ and therefore the hypothesis of nor-

mally distributed residuals for both models can be accepted.

The presence of serial correlation is tested using the Breusch-Godfrey test, also known as the LM test for autocorrelation. The Lagrange Multiplier (LM) test is not limited to testing the presence of the first-

order of autocorrelation and it is valid when there is present a lagged dependent variable. The usually used Durbin-Watson test is biased towards to 2 when a lagged dependent variable is included in the model as an explanatory variable. The null hypothesis to be tested is that:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0$$

That is, there is no serial correlation of any order. The test results of the applied Breusch-Godfrey test are shown in the table below.

Table 1. The Breusch-Godfrey Test

a) German tourist arrivals

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.336422	Prob. F(4,4)	0.3928
Obs*R-squared	6.291947	Prob. Chi-Square(4)	0.1784

b) German tourists' overnight stays

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	3.001321	Prob. F(4,4)	0.1562
Obs*R-squared	8.250908	Prob. Chi-Square(4)	0.0828

The performed LM test on both models showed that the null hypothesis of no serial correlation up to order 4 can be accepted, in fact both calculated White statistics, 6,2 for the model of German tourist arrivals and 8,25 for the model of German tourists' overnight stays respectively are smaller than the critical value of $\chi^2(4) = 9,48$.

White test for heteroscedasticity is the LM statistic for testing that all of the δ_i coefficients, except for the intercept, are zero. The White's statistics has a χ^2 distribution, with degrees of freedom equal to the number of regressors excluding the intercept. The results of the performed White test are shown in the table below.

To test the model for the presence of heteroscedasticity the White test is used. The

Table 2. The White Test

a) German tourist arrivals

Heteroskedasticity Test: White

F-statistic	0.616302	Prob. F(4,6)	0.6672
Obs*R-squared	3.203381	Prob. Chi-Square(4)	0.5244
Scaled explained SS	1.622405	Prob. Chi-Square(4)	0.8048

b) German tourists' overnight stays

Heteroskedasticity Test: White

F-statistic	0.754552	Prob. F(4,6)	0.5902
Obs*R-squared	3.681474	Prob. Chi-Square(4)	0.4508
Scaled explained SS	1.702330	Prob. Chi-Square(4)	0.7903

The null hypothesis of no heteroscedasticity can be accepted for both models as the calculated White statistics, 3,203381 and 33,681474 respectively, are smaller than the critical value of $\chi^2(4) = 9,48$.

The Variance Inflation Factors are used to test the estimated model for the presence of collinearity. The variance inflation factor

estimates how much the variance of a coefficient is "inflated" because of linear dependence with other predictors. A general rule states that, if the VIF of a variable exceeds 10, that variable is considered to be highly collinear. The minimum possible value of the VIF is 1. The testing results are presented in the tables below.

Table 3. The VIF

Variance Inflation Factors

Date: 10/25/14 Time: 13:46

Sample: 2003 2013

Included observations: 11

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	3843.499	13.26667	NA
X	563.8222	89.52308	19.46154
X2	3.714244	46.58974	19.46154

As shown above, the calculated VIF values indicate that there is a collinearity problem between explanatory variables in the estimated model. As it is known, collinearity is a common problem in polynomial regressions, where the predictors are usually high-

ly correlated. To reduce the problem of high collinearity between the predictors in both estimated models, independent variables are centred around the mean. After centering, the correlation between the predictors is calculated.

Table 4. Correlation between predictors before and after centering

	X	X2		Z	Z2
X	1.000000	0.973970	Z	1.000000	0.000000
X2	0.973970	1.000000	Z2	0.000000	1.000000

The correlation between the predictors x and x^2 is 0,97, while the correlation between z and z^2 is 0,00 as shown in the table above. Moreover, the regression of the used dependent variables on the predictors x and x^2 are virtually identical to the regression of

the same dependent variables on the centered predictors z and z^2 . After centering the independent variables, i.e. x and x^2 , by subtracting the variable's mean the models are re-estimated and the results are the following second-degree polynomials:

$$\begin{aligned}
 \text{GTA} &= 1539,517 + 29,53636z + 8,793706z^2 \\
 &\quad (61,99596)(23,74494) (1,927237) \\
 &\quad t = 1 \text{ in } 2003 \\
 &\quad \text{unit for } t = 1 \text{ year} \\
 &\quad \text{unit for } Y = \text{thousands of tourist arrivals}
 \end{aligned} \tag{5}$$

$$\begin{aligned}
 R^2_{\text{adj}} &= 0,83 & \sigma &= 56,45 & \text{RSS} &= 25,49 & \chi^2_{\text{Auto}}(4) &= 6,292 & \chi^2_{\text{Norm}}(2) &= 1,951 \\
 \chi^2_{\text{White}}(4) &= 3,203 & \text{VIF} &= 1 & & & & & &
 \end{aligned}$$

$$\begin{aligned}
 \text{GTN} &= 11048,74 + 3197636z + 72,95338z^2 \\
 &\quad (323,5495)(123,922 (10,058) \\
 &\quad t = 1 \text{ in } 2003 \\
 &\quad \text{unit for } t = 1 \text{ year} \\
 &\quad \text{unit for } Y = \text{thousands of tourist arrivals}
 \end{aligned} \tag{6}$$

$$\begin{aligned}
 R^2_{\text{adj}} &= 0,947 & \sigma &= 294,62 & \text{RSS} &= 694388,2 & \chi^2_{\text{Auto}}(4) &= 8,25 & \chi^2_{\text{Norm}}(2) &= 0,038 \\
 \chi^2_{\text{White}}(4) &= 3,68 & \text{VIF} &= 1 & & & & & &
 \end{aligned}$$

where

- GTA is the annual number of German tourist arrivals
- GTN is the annual number of German tourist nights
- z is the centred predictor
- $\hat{\sigma}$ is the standard error of the regression
- RSS is the sum squared residuals
- $\chi^2_{\text{Norm}}(2)$ is the Jarque-Bera normality test

- $\chi^2_{\text{Auto}}(4)$ is the Breusch-Godfrey LM test for autocorrelation
- $\chi^2_{\text{White}}(4)$ is the White test for heteroscedasticity
- VIF are the Variance Inflation Factors

After centering, the Variance Inflation Factors are calculated again. The results are listed below.



Table 5. The VIF after centering the predictors

Variance Inflation Factors

Date: 10/25/14 Time:

14:16

Sample: 2003 2013

Included observations: 11

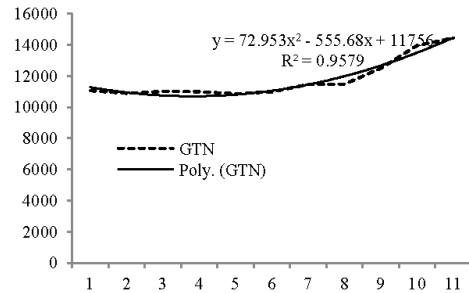
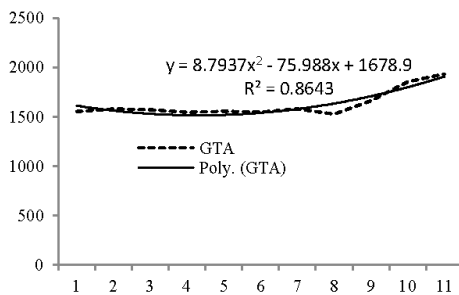
Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	661.1354	2.282051	NA
Z	28.97110	1.000000	1.000000
Z2	3.714244	2.282051	1.000000

As shown above, the values of the VIF are 1 and there is no evidence of a significant problem of collinearity. Moreover, the re-estimated model has the same predictive capability; all the re-performed perform-

ance tests showed that the models fit the data well.

As shown in Figure 3 the models fit the data reasonably well throughout the sample period.

Figure 3. Annual German tourist arrivals and tourists' overnight stays in Croatia: fitted versus actual; Time bounds: 2003-2013



The models fit the general movement of the analysed series during the entire sample

period. The forecasted results are presented in the table below.

Table 6. Forecasted results of the estimated models

	German tourist arrivals	German tourists' overnight stays
R^2_{adj}	0,83	0,94
MAPE	2,4222	1,614
RMSE	48,142	251,249
Theil'U	0,014	0,011

The forecast results are reasonably good; the predicted values, in fact, are quite close to the actual values.

Conclusion

Although there are other economic activities, Croatia's economy is largely based on tourism. International tourist flows in particular, represent a significant source of profit. Germany is among the most important generators of foreign tourist flows to Croatia. It is therefore crucial to analyze the determinants and the core patterns of German tourism demand in Croatia. The aim of this study was to establish an econometric model in order to explain the core determinants of German tourism demand to Croatia. In literature, tourist arrivals and tourists overnights stays are the most commonly used measure of international tourism demand. This study assumed that the data generating process of international tourism demand in Croatia can be approximated by a second degree polynomial regression model. The paper investigated the applicability of using polynomial analysis in modelling tourism demand and a second-degree polynomial regression model was established in order to explain the core determinants of German tourism demand for Croatia. A polynomial regression is a special case of multiple regression, in which the relationship between the independent variable and the dependent variables is modelled as an n th order polynomial. The paper assumed that the evolution of the annual number of German tourist arrivals as well as German tourists' overnight stays in Croatia in the period from 2003 to 2013, in dependence of time t (the independent variable x) can be approximated with a second-order polynomial. Therefore, two second degree polynomial models were estimated in the attempt to model German tourist arrivals and German tourists' overnight stays in Croatia in the pe-

riod from 2003 to 2013. The estimated coefficients and all performed diagnostic statistics showed that the specified models fit the classical assumptions and that the models pass all the usually used tests. The diagnostic statistics showed also, that the models fit the data reasonably well throughout the sample period and that the model performs well in modelling German tourists arrivals and German tourists' overnight stays in Croatia in the period from 2003 to 2013. Thus, more comprehensive and detailed studies, could be used in formulation of future macroeconomic development strategies, pricing strategies and tourism sector routing strategies in Croatia, as a predominantly tourism oriented country.

References

- Baggio, R. and Klobas, J. (2011), *Quantitative Methods in Tourism*, Channel View Publications. Bristol.
- Bozarth, C. (2013), Measuring Forecast Accuracy: Approaches to Forecasting : A Tutorial. Retrieved February 2, 2013, from <http://scm.ncsu.edu/scm-articles/article/measuring-forecast-accuracy-approaches-to-forecasting-a-tutorial>.
- Burger, C.J.S.C. (2011), A practitioners guide to time – series methods for tourism demand forecasting – a case study of Durban, South Africa, *Tourism Management*, No. 22, pp.403-409.
- Chang, C., Sriboonichtta, S., and Wiboobongse, A. (2009), "Modelling and forecasting tourism from East Asia to Thailand under temporal and spatial aggregation", *Mathematics and Computers in Simulation*, Vol. 79. No. 5, pp. 1730-1744.
- Cho, V. (2003), "A comparison of three different approaches to tourist arrivals forecasting" *Tourism Management*, Vol. 24 No.3, pp. 323-330.
- Croes, R. R. and Vanegas, M. (2005), „An econometric study of tourist arrivals in



- Aruba and its implications”, *Tourism Management*, No. 26, pp. 879-890.
- Gerard S., and Dharmaratne, G. S. (1995), „Forecasting tourists’ arrivals in Barbados” *Annals of Tourism Research*, Vol. 22, No.4, pp. 804-818.
- Teresa Garín Muñoz, (2007), „German demand for tourism in Spain,” *Tourism Management*, Vol. 28. No.1, pp. 12-22.4
- Gujarati, D. N. and Porter, D. C. (2009), *Basic Econometrics*, McGraw-Hill - International Edition, Boston.
- Ismail, J. H., Iverson, T. J., and Cai, L. A. (2000), „Forecasting Japanese arrivals to Guam: An empirical model”, *Journal of Hospitality and Leisure Marketing*, No. 7, pp. 381-396.
- Jaffe, R.A., and McGee, M. (2000), *An Introduction to Time Series Analysis and Forecasting: With Application of Sas and Spss*, Academic Press, New York.
- Lee, C.K., Song, H.J., and Mjelde, J.W. (2008), „The forecasting of International Expo tourism using quantitative and qualitative techniques”, *Tourism Management*, Vol. 29 No. 6, pp. 1084-1098.
- Song Li, G., and Witt, H. S. (2005), „Recent Developments in Econometric Modeling and Forecasting”, *Journal of Travel Research*, No. 44, pp. 82-99.
- Lim, C., and McAleer, M. (2001), *Time series forecasts of international tourism demand to Australia - Discussion paper 533*, The Institute of Social and Economic Research, Osaka University, Osaka.
- Lim, C. (1999), „A meta-analytic review of international tourism demand”, *Journal of Travel Research*, Vol. 37 No. 3, pp. 273-284.
- Lim, C. (2004), „The major determinant of Korean outbound travel to Australia”, *Mathematic and Computers Simulation*, No. 64, pp. 447-485.
- Lin, C., Chen, H. and Lee, T. (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. 14-24.
- Loganathan, N., and Yahaya, I. (2010), „Forecasting International Tourism Demand in Malaysia Using Box-Jenkins Sarima Application”, *South Asian Journal of Tourism and Heritage*, Vol. 3 No. 2, pp. 55-70.
- Nau, R.F. (2013), „What’s the bottom line? How to compare models”, Retrieved March 1, 2013, from <http://people.duke.edu/~rnau/compare.htm>
- O’Niel, M. (2003), „*Tourism Maturity and Demand: Jamaica*”, Research Services Department, Research and Economic Programming Division, Work in Progress, Bank of Jamaica.
- Shen, S., Li, G., and Song, H. (2011), „Combination Forecasts of International Tourism Demand” *Annals of Tourism Research*, Vol. 38 No.1, pp. 72-89.
- Song, H., and Li, G. (2008), „Tourism demand modelling and forecasting – A review of recent research”, *Tourism Management*, No. 29, pp. 203-220.
- Song, H., Witt, S., and Gang, L. (2012), *The Advanced Econometrics of Tourism Demand*, Routledge, London.