ANALYSING SELECTED DEMANDS FOR TOURISM INDUSTRY USING SIMULTANEOUS EQUATION MODELS

Václav Žák – Lukáš Malec

Abstract

A majority from EU countries are significant source covering the tourism industry revenues. However, tourism is a dynamic sector strongly connected to the service industry for many effects which should be deeply studied. This paper is concentrated on the 2003–2018 Czech Republic demand-side accommodation data related to various exogenous factors, e.g. destination living cost, wages and salaries, labour productivity, trade openness and consumer confidence. One branch of structural equation modelling, i.e. simultaneous equations, is applied to differenced inputs moreover including the economy-driven tourism processes. Although structural equations have an undeniable place in multivariate statistics, they are rarely applied in tourism, probably due to requirements for confirmatory analyses. Data inputs based on first and seasonal differencing are accommodated for their diverse effects. The majority from predetermined causal relations remain confirmed excluding some specific cases. From many fundamental and derived factors used in this study, the especially significant variables are trade openness and labour productivity.

Key words: structural models, differencing, tourism

JEL Code: C31, Z30

1 Introduction

Tourism is a part of the economy being heavily dependent on human capital. It is a dynamic social activity with many direct, indirect, and induced effects covering specific kinds of a market segment. Tourism generally promotes investment (Krueger, 1980; Balaguer & Cantavella-Jorda, 2002), introduces the economy of scale and reduces unemployment. There exist many directional assumptions as economic-driven tourism growth assuming tourism development through certain economic steps. By VAR-based spillover index is proven the bidirectional causality between economy and tourism strongly dependent on various economic events (Antonakakis, Dragouni & Filis, 2015). However, mass tourism usually results in price competition for large number of visitors generating congestion problem at an elevated environmental, cultural, and social cost. For understanding the tourism processes and consequent growth sustainability, high-quality tourism development strategy based on adequate modelling techniques are essential. Even though structural equation methodology is suitable for studying tourism demand due to directly reflects consumer behaviour and potential causal fits, its applications are relatively sparse. It is especially true for Central European case from which reason we fill the gap in the economic literature. This study and conclusions are based predominantly on the results of Žák, V. (2020) diploma thesis, University of Economics, Prague.

The fundamental aim of structural equations modelling is to describe the relational pattern of endogenous (dependent) variables simultaneously between latent constructs and exogenous (independent) items. Thus, the variables are divided into manifest (observed) and latent. This is based on regression equations describing how exogenous and latent constructs relate endogenous ones (Byrne, 2001). However, there exist many variants for such approach, e.g. confirmatory factor analysis and full structural model. According to Bollen (1989), simultaneous equations dealing with exclusively manifest variables construct are also classified under econometric modelling. In any case, the distinction for any form of structural equations is straightforward as general linear models or regression methods (Bílková, 2019) do not correct the measurement error directly assuming the errors in independent variables vanish. Contrary to this, the structural approach usually estimates the error variance in the model. A variant approach, the partial least squares path modelling should be mentioned. This type of methods is not based on the covariance matrix and belongs to a family of soft-modelling techniques. In the case of a complex model structure with many latent and manifest variables and small sample problem, this type of modelling should be preferred (Henseler, Ringle & Sinkovics, 2009).

In the following, the simultaneous equation system is applied to Czech tourism information gathered from selected databases. This forms a suitable alternative to other methods applied to this branch predominantly based on panel analyses. The two differencing approaches are used, i.e. first and seasonal differencing for which the models are separately discussed.

2 Experimental

2.1 Data used

The Czech Republic input data originates predominantly from the European Statistical Office (Eurostat database, 2019) gathered on knowledge from demand modelling practice (Turner & Witt, 2001; Lim, 2006). They cover the period from 2003 to 2018. Exceptional sources are

Czech National Bank for domestic loans and the number of arrivals and nights spent covering national statistical office. The data are moreover improved by disaggregation and in some cases forecasted. Due to time series processes covering various lags and forecasting, the quarterly observations are generated from monthly data. The data are logarithmically transformed for their heteroscedasticity and normality violation. To achieve close to stationary inputs and to remove seasonal fluctuations, the first or seasonal differencing is applied. We expect the short-term shifts are measured by first differenced data while seasonal ones record long-term changes in the examined time series. The input variables with different scale and fluctuating variances result in consequent problems for maximum likelihood numerical estimation which is the target for this study. Although the results can be distorted in some degree differently to the ones for using correlation matrix (Ramlall, 2016), the last step of data pre-processing is min-max normalization. It is used to analyse relative effects overcoming the numerical obstacles.

In the case of relative variables, they are weighted by the three most significant source countries in 2018 based on overnights. Some indicators are expected acting only in a long horizon, e.g. balance of payments, excluded from first differencing. Destination living cost (DLC) is the first exogenous variable from eighth and covers adjusted consumer prices of weighted source countries to the target one. Adjusted consumer price corresponds to the harmonized index of consumer price divided by Euro/ECU exchange rates in some cases. Relative wages and salaries (RWS) is the parameter derived from GDP weighted for source countries relatively to the target destination. The parameter is introduced in current prices and million Euros. The disaggregation is used here for indicator wholesale and retail trade. Balance of payments (BAP) is a weighted form for three most significant source countries to the target one. The parameter is based on goods and services in million Euros covering total economy for the rest of world partner. Labour productivity (LAP) is a part of quarterly national accounts and consists of real labour productivity per person. Here, the disaggregation was used by industrial production. Trade openness (TRO) is based on total import and export sum for the trade outside EU28 divided by GDP of the target destination. The GDP is measured at market prices, current prices in million Euros on which disaggregation is also used. Harmonized unemployment (HAU) covers total unemployment measured in thousand persons. Domestic loans (DOL) data are recomputed to Euro and consist of liabilities of households and non-profit institutions serving households in loans. Consumer confidence (COC) is the indicator in balance.

The two endogenous variables are gathered from the Czech Statistical Office and cover the widest selection of accommodation establishments. Non-resident’s and resident’s ratio (NRR) is compiled as arrivals from the three most important source countries to resident’s arrivals for accommodation establishments of the target destination. Nights spent (NTS) is a sum of overnights for three source countries and target one. The expected causal relations between exogenous and endogenous variables are summarized in Table 1. These are considered with inclusion *a priori* known strong negative correlation between NRR and NTS for first differenced data. Inspecting the Czech Statistical Office database, this is predominantly due to the short length of stay for non-residents.

Tab. 1: Directions of expected causal relations

	NRR	NTS
DLC	pos./neg.	pos./neg.
RWS	pos./neg.	pos./neg.
BAP	pos./neg.	pos./neg.
LAP	pos.	neg.
TRO	pos./neg.	pos.
HAU	pos.	neg.
DOL	pos.	neg.
COC	pos.	pos.

Source: author

2.2 Methods

A structural equations modelling is a popular multivariate method used for complex models. Contrary to explanatory approaches, however, the relational patterns have to be established on fundamental theory before analysis due to its placement under causal modelling. There are plenty of estimation methods used in the structural approach, e.g. maximum likelihood, generalized least squares, weighted and unweighted least squares, asymptotically distribution-free criterion, and ordinary least squares (Bollen, 1989; Byrne, 2001). Maximum likelihood approach is the dominant approach based on an assumption for multivariate normality, while robust with a slight violation of such an assumption. A numerical realization of the technique given by corresponding fit function consists from standard approaches as Newton-Raphson algorithm, S or R optimization. The linearity of relations between all variables is assumed, generally due to using sample covariance matrix estimates. From the most common fit indices should be mentioned chi-square statistic based on a likelihood ratio value. For p-value >0.5, a far stronger assumption the model is consistent with the observed data, is accepted. Besides the Akaike information criterion, we also use Bayesian statistics. However, the huge amount of possibilities for a testing model is one from the most important issues in practice. Covering R2 as the proportion of explained variance, there are also the cases of acceptable model fit, accounting 1% of the variance in endogenous variables, see, e.g., (Tomarken & Waller, 2003). To reveal the significance of the individual coefficients, the t-tests are applied for individual relations.

In case of justification for a modified model exists, it should be rather specified before analysis. Some authors argued that respecification of the model as a confirmatory process for explanatory purposes (MacCallum et al. 1993) is not sufficient until the new model is validated by new sample data. We utilize the modification indices based on Lagrange-multiplier one-df chi-square score test for the fixed and constrained parameters. The R software (R Core Team, 2019) is used in this study with several packages as *sem*, *tempdisagg* and *forecast*.

3 Results and discussion

Initially, the first difference is applied to reveal short-term relations for tourism demand. BAP and HAU are expected to influence tourism in longer-term scale, so incorporated specifically in seasonal time series processing. Moreover, DOL is excluded from analysis due to collinear with many exogenous variables. DLC for first differenced data is considered excluding NTS, and for seasonal differencing, LAP covers only NRR variable. For both models, the value of Bayesian criterion is negative which according to Raftery (1995) supports the correct choice of these models. Covariance between variables is modelled in case of the absolute value of correlation coefficient exceeds 0.5 or in the case of recommendation when modifying the model. Although endogenous variables are strongly related for first differenced data, we analyse both for their importance simultaneously. In this case, the chi-square statistic equals 9.569 with corresponding p-value 0.297, and Akaike criterion has a value 49.57. According to the R2, 38% of variability for NRR is explained by the model, while 29% for NTS.

The resulting model is shown in Table 2 and corresponding Figure 1 for graphical expression in the sense of significant paths. In the table are introduced relations, standardized coefficients (StdCoeff), estimation errors (EstError) and p-values of corresponding t-tests. Moreover, the covariances RWS ↔ DLC, LAP ↔ RWS and NRR ↔ NTS are modelled with individual p-values $< 10e-4$, and then for TRO ↔ LAP with p-value $< 10e-2$. We generally interpret variables that are statistically significant at the 10% level. The effect of the TRO variable is important on both endogenous variables at a level of less than 1%. While it has a positive effect on the NRR variable, it negatively affects the NTS variable. The NRR variable is also positively influenced by LAP. Considering the NTS endogenous variable, RWS variable has a negative effect while positively affected by the COC. The negative relation of the TRO and RWS on the NTS variable can be partly explained by the negative correlation between the endogenous NRR and NTS variables in the data set.

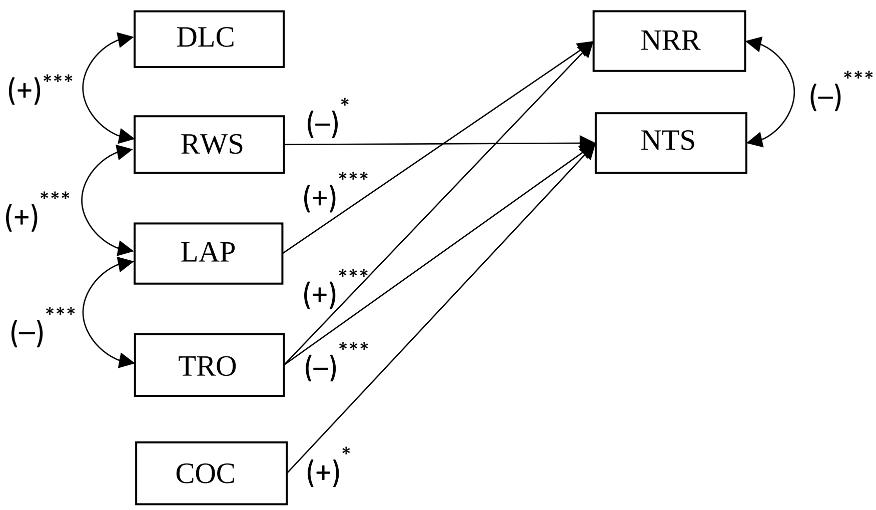
Table 3 consists of chi-square statistic processes, its *p*-values and indexes of determination for both endogenous variables. Time lag (months) is introduced in the sense of negative lag for delayed endogenous tourism variables. It is interesting the ratio of explained variability increases remarkably for lags 2 and 3 months. No interesting trend is evident in the chi-square statistic process. This is despite fact that relational pattern is relatively similar in time process to some degree of shift.

Tab. 2: Results for first differenced data

Relations	StdCoeff	EstError	*p*-value
DLC → NRR	-0.043	0.092	0.642
RWS → NRR	0.270	0.190	0.155
LAP → NRR	0.568	0.135	0.000
TRO → NRR	0.632	0.187	0.001
COC → NRR	-0.310	0.258	0.230
RWS → NTS	-0.354	0.192	0.066
LAP → NTS	-0.173	0.142	0.223
TRO → NTS	-0.787	0.199	0.000
COC → NTS	0.521	0.275	0.058

Source: author

Fig. 1: Structural diagram



Significant at ***(1%), **(5%) and *(10%). Source: author

Tab. 3: Chi-square time processes

Time lag	Chi-square	*p*-value	R2-nrr	R2-nts
-4	19.3	0.013	0.65	0.33
-3	21.2	0.007	0.28	0.18
-2	22.8	0.004	0.34	0.26
-1	16.0	0.042	0.39	0.38
0	9.6	0.297	0.38	0.29
1	9.0	0.339	0.37	0.67
2	7.5	0.481	0.92	0.85
3	30.8	0.001	0.80	0.78

Source: author

Table 4 and Figure 2 demonstrate the data processing input for seasonal differencing. In this case, the chi-square statistic equals 18.22 with corresponding *p*-value 0.051 and Akaike criterion has a value 54.22. From R2, 37% of variability for variable NRR is explained by the model, while 16% variability for NTS. This ratio for NTS is relatively lower than in the case of first differenced data. The modelled covariances are for LAP ↔ BAP and LAP ↔ HAU with *p*-values 0.002 and 0.006, respectively. All exogenous variables to NRR variable are statistically significant. The NRR variable is positively affected by the variables LAP, COC, HAU and BAP, and negatively by TRO. The LAP and TRO variables are especially significant. NTS is negatively affected by the variables HAU and COC.

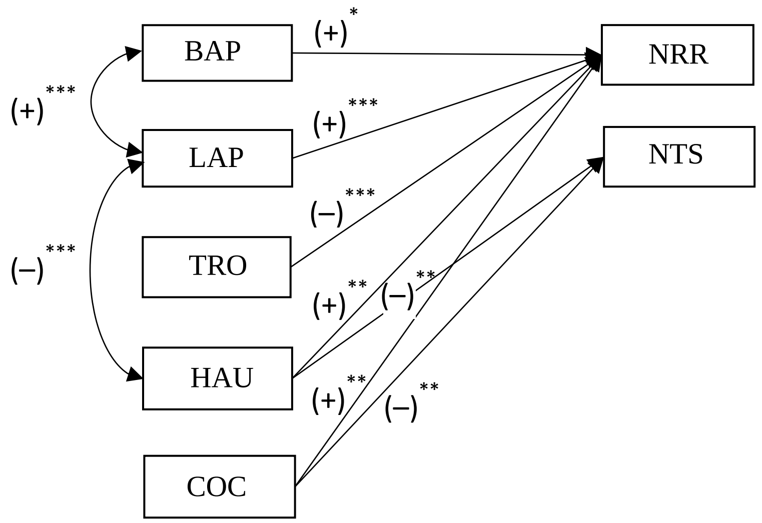
In Table 5 are introduced processes of the relations for exogenous variables lag. Chi-square statistic value slightly increases while R2 demonstrates no visible trend.

Tab. 4: Results for seasonally differenced data

Relations	StdCoeff	EstError	*p*-value
BAP → NRR	0.249	0.137	0.070
LAP → NRR	0.381	0.113	0.001
TRO → NRR	-0.309	0.118	0.009
HAU → NRR	0.247	0.120	0.040
COC → NRR	0.327	0.146	0.025
BAP → NTS	0.113	0.147	0.441
TRO → NTS	-0.054	0.141	0.699
HAU → NTS	-0.332	0.133	0.013
COC → NTS	-0.369	0.173	0.033

Source: author

Fig. 2: Structural diagram



Significant at ***(1%), **(5%) and *(10%). Source: author

Tab. 5: Chi-square time processes

Time lag	Chi-square	*p*-value	R2-nrr	R2-nts
-4	13.77	0.184	0.26	0.15
-3	14.45	0.153	0.35	0.21
-2	15.85	0.104	0.34	0.22
-1	18.05	0.054	0.38	0.19
0	18.22	0.051	0.37	0.16
1	20.98	0.021	0.38	0.12
2	20.75	0.023	0.31	0.15
3	23.12	0.010	0.30	0.17

Source: author

4 Conclusion

The simultaneous equation modelling proves to be successful in tourism data processing. It is interesting to see the weak influence of the DLC, RWS and BAP to NRR for both, first and seasonally differenced data. Due to NRR is a relative endogenous variable, we would rather expect a stronger relation to the ratio exogenous ones. On the other hand, the NRR relational patterns are stronger than for NTS and covers exogenous variables joined dominantly to the target country. Regarding the construction of TRO variable, TRO and LAP are the most important. The fundamental results are summarized below:

- In the case of first differenced data, the relationships are influenced in general, by a strong negative correlation between NRR and NTS. TRO is the most significant variable for the relations. In the short-term shifts in time series, it has a positive influence on non-residents, whereas the overnights are reduced due to different behaviour of non-residents and residents. RWS parameter is also influenced by negative relation for both endogenous variables as increasing revenues in source countries increase non-residents at simultaneous negative effect to NTS. The positive influence of LAP variable to NRR is evident due to workload. NTS are moreover positively influenced by COC which ranks safety among the important effects. Omitting the nonsignificant ones, here differ the expected relations of TRO negative to NTS.

- The situation for seasonal differencing is fundamentally different and covers a long horizon of relational patterns. NRR is strongly explained by many exogenous variables. Especially LAP has a negative effect on residents, while TRO has a positive influence. So, in the long-lasting changes without seasonal fluctuations, TRO rather influences residents, contrary to the first differenced case. While the negative effect of HAU on overnights is evident, the COC influence is remarkable. Covering discussion of only significant expected relations, all fit the predetermined directions with COC exceptional case.

References

Antonakakis, N., Dragouni, M., & Filis, G. (2015). How strong is the linkage between tourism and economic growth in Europe? *Economic Modelling*, 44, 142–155.

Balaguer, J., & Cantavella-Jorda, M. (2002). Tourism as a long-run economic growth factor: The Spanish case. *Applied Economics*, 34(7), 877–884.

Bílková, D. (2019). Assessment of living standard indicators in the OECD member countries. *ACC Journal*, 25(2), 21–45.

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley: New York.

Byrne, B. (2001). *Structural equation modeling with AMOS*. NJ: Lawrence Erlbaum Associates.

Eurostat (2019). European Statistical Office, European Commission, Luxembourg. Retrieved November 21, 2019, https://ec.europa.eu/eurostat/data/database

Henseler, J., Ringle, M., & Sinkovics, R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, 20, 277–319.

Krueger, A. O. (1980). Trade policy as an input to development. *American Economic Review*, 70(2), 288–292.

Lim, C. (2006). A survey of tourism demand modelling practice: Issues and implications. In L. Dwyer, & P. Forsyth (Eds.), *International handbook on the economics of tourism* (pp. 45–72). Edward Elgar.

MacCallum, R. C., D. T. Wegener, B. N. Uchino, & L. R. Fabrigar (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, 114, 185–99.

R Core Team (2019). R: A language and environment for statistical computing. Vienna: Foundation for Statistical Computing.

Raftery, A. E. (1995). Bayesian model selection in social research (with discussion). In V. Marsden (Ed.), *Sociological Methodology* (pp. 111–196). Blackwell.

Ramlall, I. (2016). *Applied structural equation modelling for researchers and practitioners: Using R and Stata for behavioural research*. Bingley: Emerald Publishing.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, 112(4), 578–98.

Turner, L. W., & Witt, S. F. (2001). Factors influencing demand for international tourism: Tourism demand analysis using structural equation modelling, revisited. *Tourism Economics*, 7(1), 21–38.

Contact

Václav Žák

Department of Statistics and Probability, University of Economics

W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic

E-mail: zakv00@vse.cz

Lukáš Malec

Department of Statistics and Probability, University of Economics

W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic

E-mail: lukas.malec@vse.cz