MODELLING PORTUGUESE TOURISM DEMAND:
A PANEL DATA APPROACH

Nuno Carlos Leitão
nuno.leitao@esg.ipsantarem.pt
MODELLING PORTUGUESE TOURISM DEMAND:  
A PANEL DATA APPROACH

Nuno Carlos Leitão *

Abstract: This study examines the features and determinants of tourism demand between Portugal and European Union (EU) in the period 1995-2006, using a static and dynamic panel data analysis. The findings indicate that Portuguese tourism increased significantly during the period in accordance with the values expected for a developed country. The regression results show that bilateral trade, income, border, and geographical distance between Portugal and countries of origin are the main determinants of tourism to Portugal. The dynamic panel data approach indicates that trade, population, and income are more important determinants than relative price. The central theme of this manuscript is to show that it may be preferable to use the GMM approach in empirical studies of tourism demand rather than static panel (polluted OLS, fixed effects or random effects estimators).

Keywords: Tourism demand; panel data; tourism forecasting; Portugal.

Introduction


* ESGTS, Polytechnic Institute of Santarém and Centre for Social Studies (CES), Migration Studies, School of Economics, University of Coimbra.

Nuno Carlos Leitão, Ph.D. in Economics. He teaches economics at Polytechnic Institute of Santarém, Portugal since 1998. His research interest include international economics, trade and immigration and, more recently, tourism studies. He has published papers in professional journals and collective volumes.
According to Michael Porter (1994) Portuguese economy should be developing a tourism cluster. Porter’s study of Portugal showed that tourism cluster was localized in four regions: Algarve, Alentejo, Madeira and Lisbon.

Tourism is the largest service-based industry and, as such, has been partly responsible for the service sector growth. Tourism has played a highly significant role in the economic development of Portugal.

The literature on tourism demand emerged in the past two decades. The empirical models have been developed on the application of econometric models for estimating tourism demand functions, elasticity values and forecasting (Lim 1997; Song and Witt, 2006). However, research on the determinants of inbound tourism has focused mainly on explaining tourism demand developed countries.

This study uses a static and dynamic panel data set between Portugal and the European Union (EU) for the period 1995-2006. In static panel data models, Pooled OLS, fixed effects (FE) and random effects have some problems, like serial correlation, heteroskedasticity and endogeneity of some explanatory variables. In dynamic panel data models, the GMM-SYS (Arellano, M. Bover, 1995; Blundell and Bond, 1998, 2000) estimator eliminates the unobserved specific effects through the equations in differences. The GMM-SYS estimator also controls for the endogeneity of explanatory variables.

In dynamic panel data models, the GMM-SYS estimator eliminates the unobserved country-specific effects through the equations in first differences. The GMM-SYS estimator also controls the endogeneity of the explanatory variables. A standard assumption on the initial conditions allows the use of the endogenous lagged variables for two or more periods as valid instruments if there is no serial correlation (Blundell and Bond 1998, 2000). If we assume that the first differences of the variables are orthogonal to the country-specific effects, this additionally allows the use of lagged first differences of variables for one or two periods as instruments for equations in levels (Arellano and Bover 1995; Blundell and Bond 1998, 2000).

The validity of instruments is testing using a Sargan test of the over-identifying restrictions and serial correlation. First-order and second-order serial correlation in the first – differenced residuals is tested using m1 and m2 statistics (Arellano and Bond 1991). The GMM system is consistent if there is no second – order serial correlation in the residuals (m2 statistics). The dynamic panel is valid if the estimator is consistent and the instruments are valid.

As far as we know, dynamic panel data analysis as been used in empirical studies of tourism demand (Mervan and Payne 2007; Phakdisoth and Kim 2007; Vogt 2008; Brida and Risso, 2009).

1 The GMM system estimator that we report was computed using DPD for OX (Doornik et al. 2002).
To estimate the dynamic model, we apply the methodology of Blundell and Bond (1998, 2000). The results presented in this article are generally consistent with the predictions of the theory of tourism demand.

In this manuscript we apply the Newton's gravitational law. Gravity models are based on the gravity law of spatial interaction. These models have been popular in tourism studies.

This paper argues and provides evidence that trade has a positive effect on tourism demand. Because trade can influence tourism demand, bilateral trade brings a preference for home-country products and can reduce transaction costs between home and host country.

The remainder of the paper is organized as follows: section 2 presents the theoretical background; in section 3 we reflect about tourism demand estimation; section 4 we formalized the econometrical model; section 5 shows the estimation results; the final section concludes.

Literature Review

Tourism as a form of temporary international migration can, like other types of movement, shifts in the distribution of population.

Tourist visits can take place for various reasons: holidays, business trips, visit to friends and relatives, and others.

The review of the empirical literature on tourism demand by Crouch (1994), Witt and Witt (1995), Mervan and Payne (2007), Carey (1991), and Lim (1997) suggest that the tourist demand is measured as in tourist arrivals – departures and tourists receipts as dependent variable. The number of nights spent in the country has also been included, but much less frequently.

For explanatory variables, the empirical models of tourism demand use consumer’s income, the price of services, exchange rate, and distance.

Some studies as in Mervan and Payne (2007) incorporate the dynamics of tourism demand for lagged effects on supply constrains. These authors study the demand on tourism for the Croatian Destination. We can refer other studies as in Carey (1991) who studied the Caribbean tourism demand.

More recently Phakdisoth and Kim (2007) specifies static and dynamic panel data (GMM-DIF) models for tourism Laos and estimates tourism demand equations. The authors concluded that communication for transportation infrastructures; destination risk, bilateral trade, and geographical distance are the main determinants of tourism Laos.

The study of Vogt (2008) estimates real income and relative price of demand US exports and imports of tourism. The author concluded that the US trading partners appear to be more sensitive to the determinants of international tourism than in US.
BRIDA and RISSO (2009) applied a dynamic panel data study of the German demand for tourism in South Tyrol. In this paper the authors investigate the determinants of the German demand for tourism in Italian region. The dynamic panel data approach analyses short and long-term effects. BRIDA and RISSO (2009) concluded that the cost of travel and the relative price have a negative and significant impact on tourism demand. The authors also showed that the lagged dependent variable (demand for tourism) has a positive and relevant impact on actual demand, reflecting according to the authors the loyalty of German tourists.

Modelling Tourism Demand

The pioneering empirical models of the demand tourism equations were estimated by ordinary least squares (OLS).


Our study uses the GMM- system estimator (GMM-SYS) was proposed by ARELLANO and BOVER (1995) and Blundell and Bond (1998, 2000). The GMM-SYS estimator permits efficient estimates to be obtained.

In general, the literature of our modelling tourism demand focuses on the determinants, as in transport, price, income in tourism generating countries, and population.

We consider that tourism demand is equal to:

\[ \text{TOUR}_t = f(Y, P^*, R, T, \text{POP}) \]  \hspace{1cm} (1)

Where:

\[ \frac{\partial f}{\partial Y} > 0, \frac{\partial f}{\partial P^*} R < 0, \frac{\partial f}{\partial T} < 0, \frac{\partial f}{\partial \text{POP}} > 0 \]

and:

- TOUR is the number of tourist arrivals;
- Y is the income in tourist generation countries;
- P is the price of tourist services;
- R is the exchange rate
- T is the transportation cost
- POP is the population of the country of origin
Econometric Model

The dependent variable used is the number of arrivals at destination \((i)\) from origin \((j)\) in year staying in hotels \((TOUR)\). Tourist nights spent in the destination country are an alternative measure. The data explanatory variables are sourced from World Bank, World Development Indicators (2008), and Bank of Portugal. The source used for dependent variable was INE- National Institute of Statistics.

Explanatory Variables

**Hypothesis One: The tourism demand will be influence by income in the origin tourists**

LogGDP is the logarithm of absolute in GDP per capita in tourist countries (PPP, in current international dollars). According to the literature we expected a positive sign. Income in the origin country is the most frequently used explanatory variable. Most researchers have relied on nominal or real (per capita) personal or national income, and GDP or GNP as measures (or proxies) for income in the origin. The studies consider that income is highly significant as a determinant of demand.

**Hypothesis Two: Trade partners are an important vehicle to expand the tourism**

Following Eliat and Einav (2004) and Phakdisoth and Kim (2007) we consider:

\[
LogTR_{it} = \left( \frac{X_{ij} + M_{ij}}{GDP_{Portugal} + GDP_{k}} \right) \]

Where:
- \(X_{ij}\) represents the annual exports of Portugal to the country of origination of each tourist at time \(t\);
- \(M_{ij}\) represents the annual import of Portugal from each tourist's country of origination at time \(t\);
- \(GDP_{k}\) is the absolute in GDP per capita in tourist countries (PPP, in current international dollars).

EILAT and EINAV (2004), and PHAKDISOTH and KIM (2007) found a positive sign.
Hypothesis Three: The population changes are a result of high immigration from a particularly country

According to the literature (Witt and Witt, 1995, and Oigenblick, and Kirschbaum, 2002) we expected a positive sign. As in the authors shows the level of tourism depend not only the population of origin country, but also the immigrants from the destination country.

Population of the country of origin is also included as an explanatory variable, LogPOP. Although it is theoretically incorrect to exclude population. This proxy is important to analyse the impact of population changes. Most of Studies do not consider this variable, because population tends to be highly correlated with income (multicollinearity problems).jud and Joseph (1974), or Fuji and Mark (1981) found a positive sign.

Hypothesis Four: Macroeconomics variables influence the price of tourist service

\[
\text{LogPR}_i = \frac{\text{CPI}_\text{Portugal}}{\text{CPI}_k} \times \frac{E_\text{Portugal}}{E_k}
\]

Where:
• LogPR\(_i\) is the logarithm of relative price, CPI, index of consumer price (Portugal, and partner) and E\(_i\) represent the exchange rate (Portugal, and partner). According to the literature, negative sign is expected. Bria and Rissi (2009) found a negative sign.

Hypothesis Five: The tourism increases if the transportation cost decreases

LogDIST\(_i\) is the logarithm of geographical distance between the Portugal and the partner country. The cost of transports is important as a policy variable. According to the literature we expected a negative sign.

BORDER\(_i\) is a dummy variable that equals 1 if the partner-country shares a border with Portugal (i.e, Spain) and 0 otherwise. The expected sign is positive for all models.

Model Specification

\[
\text{TOUR}_i = \beta_0 + \beta_1 X_i + \delta_i + \eta_i + \epsilon_i
\]

Where \(\text{TOUR}_i\) is the number of tourist arrivals, X is a set of explanatory variables. All variables are in the logarithm form; \(\eta_i\) is the unobserved time-
-invariant specific effects; \( \delta t \) captures a common deterministic trend; \( \epsilon_{it} \) is a random disturbance assumed to be normal, and identically distributed (IID) with \( \text{E}(\epsilon_{it})=0; \text{Var}(\epsilon_{it})=\sigma^2 \neq 0 \).

The model can be rewritten in the following dynamic representation:

\[
TOUR_t = \rho TOUR_{t-1} + \beta_t X_t - \rho \beta_t X_{t-1} + \delta_t + \eta_t + \epsilon_{it} 
\]

(5)

**Estimation Results**

Table 1 present three equations using pooled OLS. In 3rd equation we observe that the variable income per capita (GDP) is statically significant with an expected positive sign. Phakdisoth and Kim (2007) also found this result.

As the variables are in log form, the coefficient of LogGDP gives the elasticity. So, when the tourist per capita increased 1% the Portuguese demand tourism increased 0.46%.

As expected, the variable LogTRADE has a significant and positive effect on tourism demand (Eliat and Einav, 2004, and Phakdisoth and Kim, 2007). The coefficient of relative price (LogPR) is significant and has a negative sign. This result is according to the finds of Vanegas Sr. (2009), and Naude Saayman (2005).

The coefficient of LogDIST (Distance) is negative as expected. This result confirms the gravitational model and the importance of the neighbourhood. The proxy BORDER also confirms the classic hypothesis of gravity model.

The variable population (LogPOP) finds a positive sign (at 1% level of significant). Jud and Joseph (1974), or Fuji and Mark (1981) also found a positive sign.

**Table 1 – Modelling Portuguese Demand: Pooled OLS**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Coefficient</th>
<th>Coefficient</th>
<th>Expected Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogGDP</td>
<td>-0.097(-0.344)</td>
<td>0.036 (0.130)</td>
<td>0.4638 (2.407)**</td>
<td>(+)</td>
</tr>
<tr>
<td>LogTRADE</td>
<td>0.055 (2.850)***</td>
<td>0.082 (5.438)***</td>
<td>0.073 (5.388)***</td>
<td>(+)</td>
</tr>
<tr>
<td>LogPR</td>
<td>-4.725 (-5.180)***</td>
<td>-4.735 (-4.932)***</td>
<td>-1.863 (-2.525)***</td>
<td>(-)</td>
</tr>
<tr>
<td>LogDIST</td>
<td>-0.796 (-5.220)***</td>
<td>0.413 (4.666)***</td>
<td>0.494 (2.869)***</td>
<td>(+)</td>
</tr>
<tr>
<td>BORDER</td>
<td>0.413 (4.666)***</td>
<td>0.494 (2.869)***</td>
<td>0.6512 (9.809)***</td>
<td>(+)</td>
</tr>
<tr>
<td>POP</td>
<td>8.516 (7.263)***</td>
<td>4.949 (4.280)***</td>
<td>2.671 (1.305)</td>
<td>(+)</td>
</tr>
<tr>
<td>Adj. R²</td>
<td>0.655</td>
<td>0.223</td>
<td>0.580</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>196</td>
<td>196</td>
<td>196</td>
<td></td>
</tr>
</tbody>
</table>

T-statistics (heteroskedasticity corrected) are in round brackets.

***/***/. statistically significant, respectively at the 1%, and 5% levels.

In table 2 we see the results with fixed effects estimator. The explanatory power is very high (Adjusted $R^2=0.801$). All explanatory variables are significant (LogGDP at 5%, LogTRADE, LogDIST, and LogPOP at 1% level), with the exception of LogPR.

The variable GDP is statically significant, with an expected positive sign. One of the main determinants of tourism demand is the positive impact in the Portuguese economy.

The variable LogTRADE is statistically significant with a correct sign. This results demonstrates that trade partners are an important vehicle to expand the tourism.

As in PHAKDISOTH and KIM (2007), the logarithm of consumer price (Log PR) is a negative effect on tourism demand. According to PHAKDISOTH and KIM (2007) this results shows how many baskets of goods a tourist must refrain from buying in his home in order to buy a basket of good in host country.

<table>
<thead>
<tr>
<th>Table 2 – Modelling Portuguese Demand: Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>LogGDP</td>
</tr>
<tr>
<td>LogTRADE</td>
</tr>
<tr>
<td>LogPR</td>
</tr>
<tr>
<td>LogDIST</td>
</tr>
<tr>
<td>LogPOP</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

T-statistics (heteroskedasticity corrected) are in round brackets. ***/** statistically significant, respectively at the 1%, and 5% levels

The dynamic panel data model, presented in table 3 is valid if the estimator is consistent and the instruments are valid. The Sargan test of over-identifying restrictions test validity of instruments used. The first- and second order serial correlation in residuals is tested by $M_1$ and $M_2$ statistics. The GMM system estimator is consistent if there is no second-order serial correlation. The Sargan test and $M_2$ statistic show that the instruments used are valid.

However, the statistically significance and positive sign of the lagged TOURit variable confirms the hypothesis that the tourism demand has a dynamic nature. According to WITT and WITT (1995) the inclusions of a lagged dependent variable in tourism demand functions express the supply side.

The variable GDP is statically significant with an expected positive sign. One of the main determinants of tourism demand is the positive impact in Portuguese economy.
The coefficient of Bilateral trade (TRADE) is positive as expected, which is confirmed by the dynamic panel data.

The geographical distance (DIST), the dominant paradigm predicts a negative sign. The result confirms this.

Table 3 – Modelling Portuguese Demand: GMM-System Estimator

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Coefficient</th>
<th>Expected signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogTOURit-1</td>
<td>0.459 (2.71)***</td>
<td>0.480 (2.89)***</td>
<td>(+)</td>
</tr>
<tr>
<td>LogGDP</td>
<td>0.329 (2.11)**</td>
<td>0.374 (2.30)**</td>
<td>(+)</td>
</tr>
<tr>
<td>LogTRADE</td>
<td>0.042 (2.30)**</td>
<td>0.047 (2.61)***</td>
<td>(+)</td>
</tr>
<tr>
<td>LogPR</td>
<td>-0.440 (-0.512)</td>
<td>-0.348 (-0.389)</td>
<td>(-)</td>
</tr>
<tr>
<td>LogDIST</td>
<td>-0.339 (-2.06)**</td>
<td>0.208 (1.84)*</td>
<td>(+)</td>
</tr>
<tr>
<td>BORDER</td>
<td>0.344 (3.40)***</td>
<td>0.314 (3.30)***</td>
<td>(+)</td>
</tr>
<tr>
<td>LPOP</td>
<td>0.004 (0.004)</td>
<td>-1.314 (-1.79)*</td>
<td>(-)</td>
</tr>
<tr>
<td>C</td>
<td>-1.408 [0.159]</td>
<td>-1.510 [0.131]</td>
<td>(-)</td>
</tr>
<tr>
<td>M1</td>
<td>1.127 [0.260]</td>
<td>1.147 [0.251]</td>
<td>(+)</td>
</tr>
<tr>
<td>M2</td>
<td>647.3 [0.000]</td>
<td>716.6 [0.000]</td>
<td>(+)</td>
</tr>
</tbody>
</table>

The null hypothesis that each coefficient is equal to zero is tested using a one-step robust standard error. T-statistics (heteroskedasticity corrected) are in round brackets.

***/*** - statistically significant, respectively at the 1%, 5%, and 10% level. P-values are in square brackets. Year dummies are included in all specifications (this is equivalent to transforming the variables into deviations from time means, i.e. the mean across the fourteen countries for each period). M1 and M2 are tests for first-order and second-order serial correlation in the first-differenced residuals, asymptotically distributed as N(0,1) under the null hypothesis of no serial correlation (based on the efficient two-step GMM estimator). W15 is the Wald statistic of joint significance of independent variables (for first-steps, excluding time dummies and the constant term). Sargan is a test of the over-identifying restrictions, asymptotically distributed as $\chi^2$ under the null of the instruments’ validity (with the two-step estimator).

The instruments in levels used are: LogTOURit (3,7), LogGDP (3,7), and LogTRADE (3,7) for first differences. For levels equations, the instruments used are first differences of all variables lagged t-2. The equations present five significant variables: (TOURit-1, TRADE, PR, DIST (or BORDER), POP).
The coefficient of LogPOP (Population) confirms a positive sign. This result is according to the empirical studies as in Witt and Witt (1995). Our result shows confirmed that proxy tend to highly correlated with income per capita (GDP). According to the literature this results express population changes and to indicate a high immigration flows.

As expected, the variable BORDER has a significant and positive effect on tourism demand.

Conclusions

In this manuscript we consider static and dynamic panel data models of tourism demand and apply them to Portuguese tourism data. Our study find income and relative price are significant, demonstrating the expected signs with their coefficients.

The paper argues and provides that bilateral trade has a positive effect on the share of tourism demand. This proxy permits concluded about the economic relationships between Portugal and countries of origination. This result demonstrates the trade partners are an important vehicle to expand the tourism. Einaiv (2004) and Naidu and Saayman (2005) also found this result.

The destination-specific variables, geographical distance, and border are also significant determinants of total tourism flow to Portugal.

On the other hand, the tourism presents a dynamic nature. The dynamic results confirm this hypothesis.

The lagged tourism demand presents an expected positive sign. Other explanatory variables such as geographical distance and consumer price are statistically significant. These results prove the dynamic nature of tourism demand and suggest that a dynamic approach is needed in order to better understand for demand tourism determinants. In the static model we find empirical evidence for the effect of economic variables on Portuguese tourism demand.

The high growth rate in the number of international arrivals in recent years indicates the opportunities that Portugal has to develop as a tourist destination.

This study has some limitations. In the future, we need to include for other control variables such as in language, culture, and immigration. The increase in immigration flows and in immigrant population is the main change on economics, political and sociological issues for European Union countries.

Finally, although the use of more recent econometrical techniques should not be an in itself, and it would be dangerous to generalized from this empirical study, it may be preferable to use the GMM approach in empirical tourism demand studies, rather than pooled OLS, fixed-effects or random-effects estimators. Their results should at least be compared.
Acknowledgments: This paper has benefited from helpful comments made by the participants at the MSKE 2009 Conference, July 15-17, Universidade Lusíada de Famalicão, Portugal.

References


