XVII CONGRESO INTERNACIONAL DE INVESTIGACIÓN EN CIENCIAS ADMINISTRATIVAS A.C. (ACACIA)

# Modelling Tourism Revenue for the North Region of Portugal

**MESA DE TRABAJO: 10. Finanzas y Economía**

Vítor Fernando Fontes Costa; victorcosta82@hotmail.com; Instituto Politécnico de Bragança.

Paula Odete Fernandes (responsable); pof@ipb.pt; Instituto Politécnico de Bragança; NECE-Research Unit in Business Sciences (Universidade da Beira Interior).

Ana Paula Monte; apmonte@ipb.pt; Instituto Politécnico de Bragança; NECE-Research Unit in Business Sciences (Universidade da Beira Interior).

(Telf.: +351.273.303000; Fax: +351.273.313051)

Instituto Politécnico de Bragança (IPB)

Campus de Sta. Apolónia, Apartado 1134

5301-857 Bragança, Portugal.

Universidad del Valle de Atemajac (UNIVA), Campus Guadalajara, Mexico.

23 al 26 de abril de 2013.

# Modelling Tourism Revenue for the North Region of Portugal


## Abstract

In the last 50 years, the sector of the Tourism in Portugal has had an extraordinary economic importance to the level of financial revenue (contribution for the balance of payments and equilibrium) and by the creation of employment and sustained development of tourist regions. Due to the importance reached by the sector of the Tourism in Portugal, the goal of this work is the analysis of the tourism revenues for the North Region of Portugal, in the period of 2006 to 2011. The tourism revenues allow measuring the performance of the sector and the economic impact of several policies for the Tourism.

It follows that the first differences of logarithm model is the econometric model that explains better the tourism revenue in the North Region. The estimated model allows saying that the total tourism revenue are influenced by variations occurred in the number of national and foreigners tourists, in the available income of the national and foreigners tourists and still in the typology of the hotel establishment. In spite of the high increase of the foreign tourists in the North Region and hotel investment, the main contribution for the tourism revenue was the national tourists (internal market).


**Keywords**: Tourism Revenue, North Region of Portugal, Econometric Models.

## Introduction

The Tourism is by definition the set of “activities realised by the visitors during them trips and stays in places distinguished of them usual residence, by a continuous period less than 12 months, with aim of leisure, businesses or other reasons” (Tourism of Portugal, 2008, p. 18). This is an economic sector that grew significantly and sustained in the worldwide. Leitão (2011) refers that in the last 50 years the worldwide economic growth large extension due the tourism sector. According with the same author, it is expected the number of people that travel internationally scope 1,6 bilions in 2020. This shows the increasing importance of the tourism and its weight for the economic development of the countries, especially countries of small dimension and with high resources and touristic offer.

Santos and Fernandes (2010) state that tourism is considered one of the most important sectors of the Portuguese economy, because plays an important role for the economy, both at the level of production or employment.

Makhlouf (2012) adds that the investment made in this sector generates two large multiplier effects: i) increased spending of tourists and increasing employment in local and national sectors and ii) results in the construction of tourist infrastructure such as roads, airports, museums, amusement parks, establishments health and hotels. These effects have a positive impact in the life quality of the resident people allowing relate tourism with the economic and regional development (Santos, 2011, p.32).

In accordance with of the POPH (2012), POFC (2012), PRODER (2012) and CCDRN (2012) data, in Portugal, since 2007 there were approved 496 million euros of public and private investment, of which 190 millions been intended for the construction or requalification of four and five stars hotels. The tourism capacity increased by 4,652 beds between 2006 and 2011, reaching 40,156 beds in 2011, highlighting the hotels that have increased 7.6% annually and their number has almost doubled in value (205 units in 2011). The guests' number increased in 498 thousand (23,2 %), so it is an annual growth rate of 4,3 % and the overnight reached 4,5 million, more 702,6 thousand overnight than 2006. In 2011, the North Region reached 223,8 million euros of tourism revenue, of which 179,4 million euros are from hotels. The Overnight Average Stays of the national and foreign tourists was 1,5 and 1,9 overnights, respectively.

Given the evolution of these tourism indicators, the objective was to build a model that would allow estimate tourist revenue for the North Region and evaluate if the typology of hotel establishments, the GDP per capita of Portugal, Spain, France and the UK have effects in capturing tourism revenue as well as the number of domestic and foreign overnights. In the construction and modelling tourism revenue it used a sample of monthly data for the period between January 2006 to December 2011, collected from the National Statistical Institute, Eurostat and Bank of Portugal (BP).

This paper was structured in four sections. In the first section, it is established the econometric model used and its assumptions. In the second section, shows the variables included in the model and its behavior over the period of analysis. The next section explains the main results founded for the used models and in the last section it is pointed out the main conclusions and future research lines of investigation.

## 1. Theoretical framing

The tourist to be able to satisfy their necessities needs to have available goods and tourism services, hereinafter referred to as tourism products. The tourism supply is, by definition, “the quantity of tourism products that the enterprises are able to supply at given price, in a specific period of time” (Lage & Milone, 2001, p. 72).

In this way, the tourism supply can be defined as the set of natural and artificial attractions of a region, as well as all the tourism products available to consumers to satisfy their needs. Each of these natural and artificial elements induces in providing a distinct tourism supply at regional, national and global level. This tourism supply can be defined through a supply curve that is defined by a set of ordered pairs price-quantity for which the suppliers are satisfied (Frank, 2000). Another important variable of the tourism market is the “Tourism Demand”, which can defined as “the quantity of tourist goods and services that the people want and are able to consume at given price, in specific period of time” (Lage & Milone, 2001, p. 56). The tourist acquires a set of tourism products available in a specific time and geographical space, with the goal of satisfying their needs. That is, the demand of tourism products for the (consumer) tourists is represented by the curve of the tourism demand. This curve indicates the quantities that the buyers are willing to buy to several prices (Frank, 2000).

As said by Lage and Milone (2001, p. 92), the “tourism market can be considered a set of informations that allows to the economical agents (consumers and producers) to take tem rational decisions resolving the economical problems of the sector”. This value results from the multiplication of the quantity of tourism products sold at the respective price. It is hard or almost impossible to quantifying the true value of market, an approximation will be to consider the amount of total revenues in tourism establishments. This value will result from the balance between the quantity demanded by the national and foreign tourists (number of overnight, food industry, others services) and the supply quantity (capacity of accommodation, available services, infrastructures, etc.), for a given price established by the tourism company.

### 1.1. Model of Linear Multiple Regression

The modelling and forecasts of the tourism revenue for the North Region of Portugal is important for future strategic plannings and evaluation of the objectives established for the development of the Tourism sector.

This study seeks to estimate the tourism revenue for the North Region using an econometric model. For that, the methodology proposed envolved the collection data and estimation of the Multiple Linear Regression Model, which is an econometric model. These models postulate, generally, the joint and simultaneous existence of several relations, each one of them containing more than two variables. According to Oliveira et al. (1997), Johnston and DiNardo (2000), Maroco (2003), Pestana and Gageiro (2008) and Zhihua and Qihua (2009), the last goal of the econometrics is the analysis of systems of simultaneous equations how there is the case of the Multiple Linear Regression Model (GLM). In multiple linear regression there are *p* explanatory variables, and the relationship between the dependent variable and the explanatory variables is represented by the equation (Johnston & DiNardo, 2000):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \mu_i \qquad i = 1, \ldots, n \qquad [1]$$

Where: $\beta_0$ is the constant term and $\beta_1$ to $\beta_p$ are the coefficients relating the *p* explanatory variables to the variables of interest. So, multiple linear regression can be thought of an extension of simple linear regression, where there are *p* explanatory variables, or simple linear regression can be thought of as a special case of multiple linear regression, where *p*=1. The term ‘linear’ is used because in multiple linear

regression it was assume that *Y* is directly related to a linear combination of the explanatory variables.

The Multiple Linear Regression Model allows to obtain a set of estimated coefficients wich will be used in the forecasts of a set of data *Y* for an interval of confidence of (1-α) %, normally 95 %, (Pestana & Gageiro, 2008) with which it will be possible to build future hypotheses or forecasting.

### 1.2. A Brief Overview of the Assumptions of Multiple Linear Regression Models

The estimated coefficients through the method Ordinary Least Squares (OLS) allows to minimizing the Residual Sum of Squares. However, to build the desired model is necessary to check the following assumptions (Gujarati, 1995, pp. 192-193):

*i.* The average of the errors terms has to be zero

$$E(\mu_i) = 0 \quad [2]$$

*ii.* Does not exist correlation between the errors terms

$$cov(\mu_i, \mu_j) = 0, with\ i \neq j \quad [3]$$

*iii.* Homoscedasticity, the variance of the errors terms will be

$$var(\mu_i) = \sigma^2, \quad i = 1, \dots, n. \quad [4]$$

*iv.* The independent variables $X_j$ are fixed, in other words, the covariance between $\mu_i$ and any independent variable $X_j$ will be null (0).

$$cov(\mu_i, X_{2i}) = cov(\mu_i, X_{3i}) = \cdots = cov(\mu_i, X_{ni}) = 0 \quad [5]$$

*v.* The econometric model must be very well specified (characteristic of the matrix *X* is *k*<*n*)

*vi.* Does not exist multicollinearity between Independent variables $X_j$.

*vii.* If the errors terms, $\mu_i \sim N(0, \sigma^2 I)$, *i = 1,…, n* [6]

errors terms will be normal multivaried distribution, so matrix will be zero.

If any of these assumptions is violated, then the forecasts, confidence intervals, and economic insights yielded by a regression model may be (at best) inefficient or (at worst) seriously biased or misleading.

### 1.3. Violation of the Presuppositions of Multiple Linear Regression Model

To guarantee that the model is adapted to the study of the phenomenon in question, it is necessary to apply tests which they allow to check if the basic assumptions were or not infringed, in other words, if some assumptions are not valid, that means that there is an error of specification (Jonhston & DiNardo, 2000). For getting the regressor estimated through the method of the OLS it is important not violated that assumptions. Consistent with Gunst and Mason, (1980), Oliveira et al. (1997), Gujarati (1995) and Johnston and DiNardo (2000), the violation of assumption (*vi*) indicates that there will be linear dependence between columns of the matrix *X*, in other words, exists $cov\ (\beta_i, \beta_j) \neq 0$ with $i \neq j$. To overcome this, it is required to exclude the variable or variables that are causing the problem, since these variables are being explained by other independent variables. The exclusion of these variables makes the estimated model no longer the original, and in some cases it may lose the explanatory power of the economical phenomenon. One way to detect the presence of multicollinearity is through variance inflation factor (VIF). As a rule of thumb, variables whose VIF values are greater than 10, demonstrates that these explanatory variables are strongly correlated, and adds that the greater this VIF greater the presence of multicollinearity. One of the suggestions to eliminate multicollinearity is based on applying the Model First Differences (Johnston & DiNardo, 2000).

Regarding the violation of the error term ($\mu$), as said by Johnston and DiNardo (2000), there are 3 types of violation of the error terms: i) the violation of the assumption of normality of errors; ii) the existence of heroscedasticity and iii) the presence of autocorrelation of the error term.

After the estimated econometric model being designed and tested all assumptions, it is needed to check the quality or accuracy of the estimation. The use of the Adjusted R-Squared ($\bar{R}^2$), allows to know the explanatory power of the model (Johnston & DiNardo, 2000). This coefficient varies between 0 and 1, where 0 is no power explanatory and 1 explains the overall variability of the model.

## 2. Moddeling Tourism Revenue for the North Region of Portugal

### 2.1. Presentation and Characterization of Variables

This study aims to build an econometric model that defines the main variables that influence or explain the Tourism Revenue for the North Region of Portugal. Analyzing tourism revenue through tourism supply (number of establishments, number of rooms and accommodation capacity,…) and tourism demand (number of beds per establishment, number of guests, average length of stay of tourists, etc..), it was identified the following variables that was included in the original model (see Table 1):

**Table 1.**
*Variables in Econometric Model Study.*

| | |
|---|---|
| Explained variable | Tourism revenue in tourism establishments |
| Explanatory variables | Domestic overnights tourism; Foreigner overnights tourism; Portuguese GDP per capita; Spanish GDP per capita; France GDP per capita; United Kingdom GDP per capita; Domestic occupancy rate; Foreign occupancy rate; Number of accomodation establishments-hotels; Number of accomodation establishments-Others; Domestic overnight average length of stay; Foreign overnight average length of stay; Tourist function rate. |

The data variables: overnights, occupancy rate-bed places, number of accomodation establishments, average length of stay and tourism revenue were collected through the Tourism Statistics from the Portuguese National Institute of Statistics (INE) for the years 2006 to 2011. Data collected for the variables GDP per capita of Spain, Portugal, France and the United Kingdom (UK) were by EUROSTAT, while the variable population, to calculate the Tourist function rate, was used the Statistical Yearbook for the North Region of INE.

The selection of Spain, France and the UK were selected because they are the markets with the largest market share for years under review. The data correspond to 72 monthly observations between January 2006 and December 2011. Noticed that not all values of the variables were available collected in monthly values. It was necessary to make secondary transformations and calculations that enabled display monthly data without loss of its significant value and its eventually variability.

It was selected the total tourism revenue as dependent variable because it is one of the tourism indicators that directly quantifies the spendinf incurred by the domestic and foreign tourists in the region. There is another significant part of expenditure that occurs associated with tourism but that is not easily measurable (spending on restaurants, cafes, culture, transport, travel agency or communications).

### 2.1.1. Total Tourism Revenue in the North of Portugal

Analyzing the behavior of total tourism revenue between 2006 and 2011, it may be observed that registered positive percentage changes almost every year, especially in hotels. Tourism revenue registered a growth rate of 4.1%. The Hotels are the most significant weight to the total tourism revenue (Figure 1) and can be pointed three reasons: i) the number of hotels and hotel capacity installed (number of beds) it is clearly more than other types of establisments, ii) a Revenue per Available Room (RevPar) high due to the high number of guests and overnights and iii) the acquisition of a greater number of complementary services associated with the overnight stay. Thus, the hotels increased their weight from 72.2% in 2006 to 80.2% in 2011.

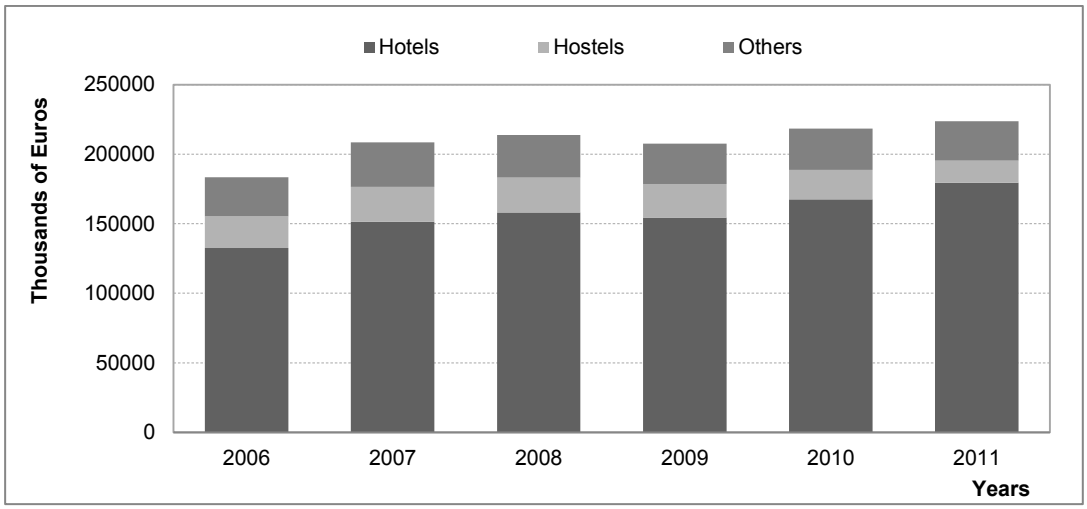


***Figure 1.*** Total Tourism Revenue in Accommodation Establishments.

The Figure 2 shows that total tourism revenue, in Accommodation Establishments in the North of Portugal, will increase during the study period and found an increased size of the intervals of seasonality from 2009, i.e., a higher concentration of total tourism revenue during the summer, reaching the maximum value in August 2011 with 28.7 million.

This increase may be the result of the policy of public and private investment undertaken in four and five star hotels and advertising campaigns to national and international tourists, which may be based on appropriate and adjusted marketing strategies to the new preferences of tourism consumers, based on unique and differentiators tourism resources, such as the cultural-historical heritage, landscape, nature and gastronomy.

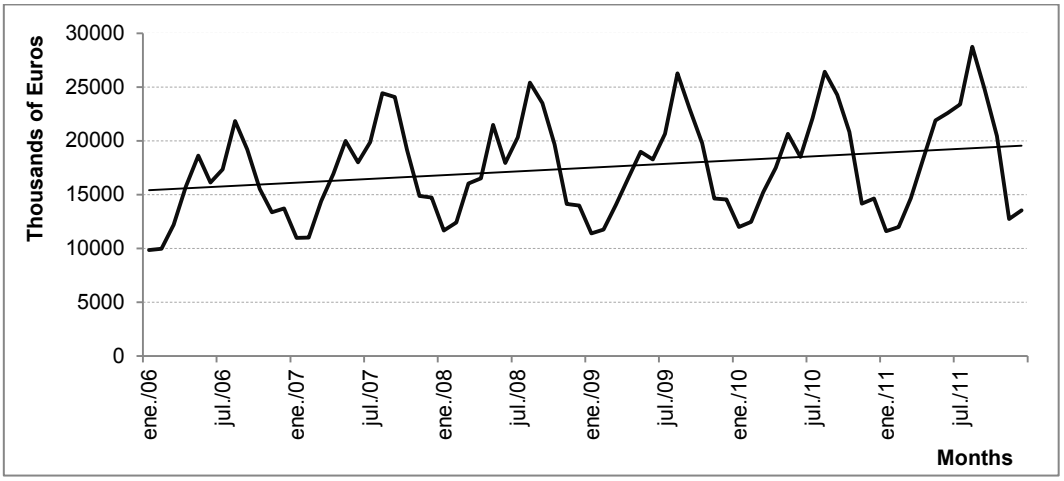


***Figure 2.*** Total monthly Tourism Revenue in Accommodation Establishments.

### 2.1.2 Domestic and Foreigner Overnights Tourism

The Figure 3 showed an increase in domestic and foreign overnights over time. The effect of seasonality it was obvious in the summer months, with 338,600 domestic (national) overnight stays registered in August 2009, while foreign overnight stays was 184,000 in August 2011. The foreign overnight stays increased every year, except 2009 which recorded a strong decrease compared to 2008 of 5.34%.

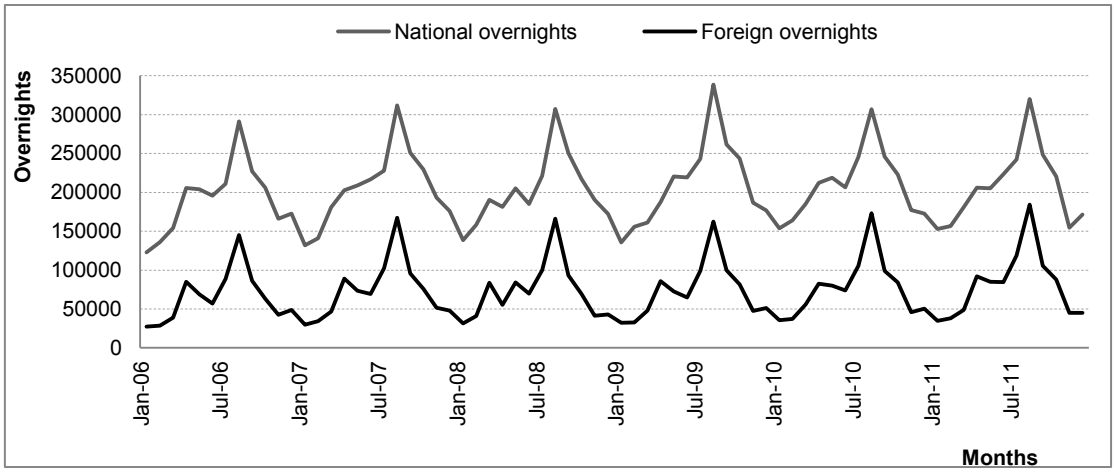


***Figure 3.*** Monthly Overnight of Domestic and Foreign Tourists.

### 2.1.3 Gross Domestic Product per Capita

The Gross Domestic Product per capita (GDP per capita) is the average wealth created per capita in a given period of time, i.e., the average income that every citizen was able to produce over a given time period (monthly, quarterly or yearly).

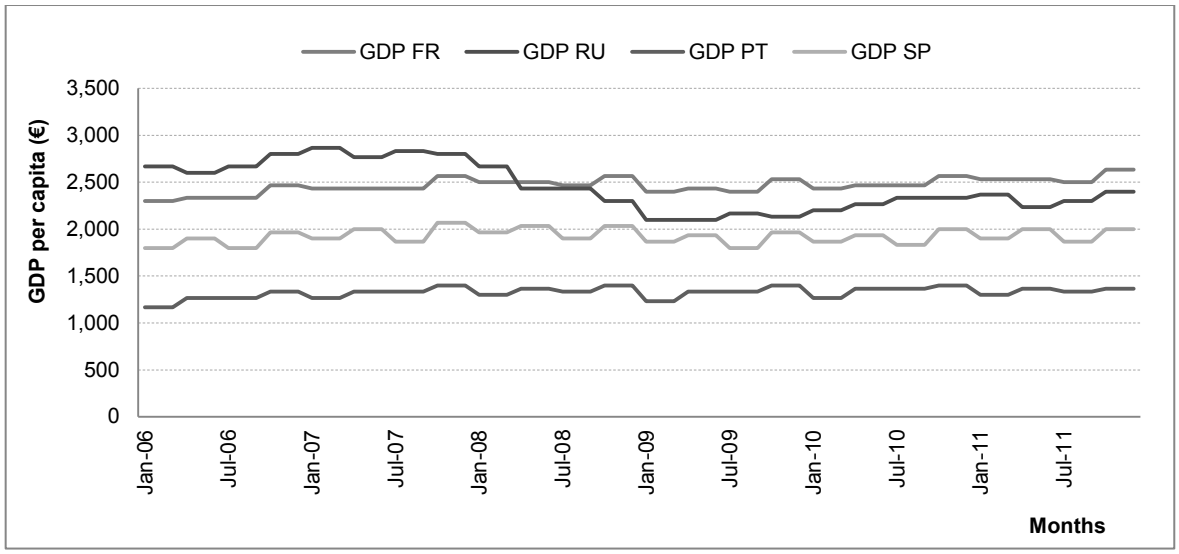


***Figure 4.*** GDP per capita of Spain, Portugal, France and the UK.

The Figure 4 demonstrates that the Portuguese GDP per capita increased from € 1,167 to € 1,367 between January 2006 to December 2011, but had 7 quarters with negative effect associated "Q1" (see Figure 5). This shortfall is due to lower tourism consumption of goods and services in the months of January and February that are associated with the period balances in sectors such as clothing and footwear or less disposable income due to high spending Christmas season.

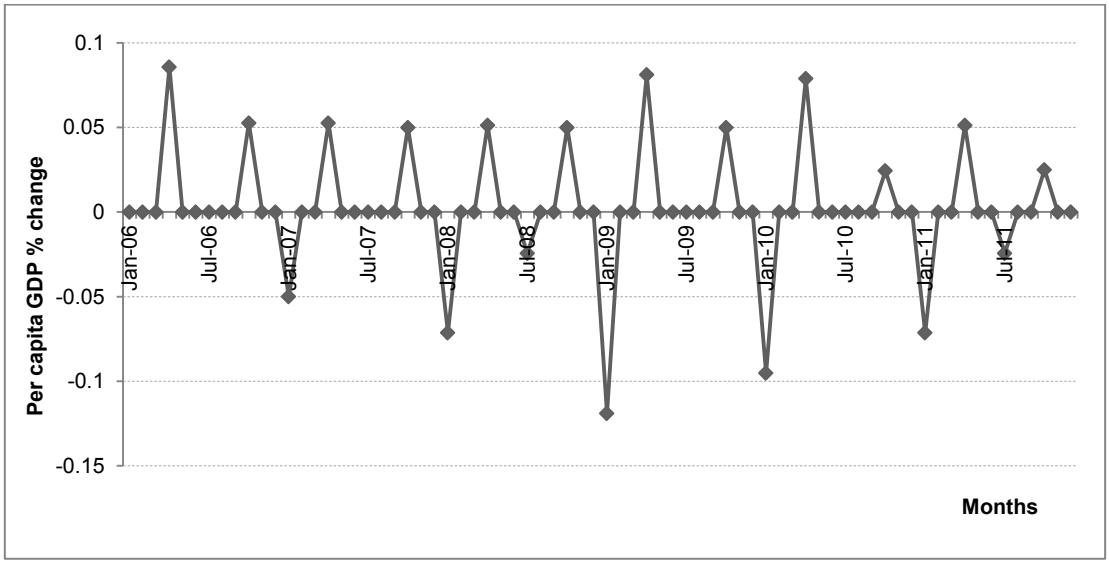


***Figure 5.*** Percentage of Change in GDP per capita of Portugal.

For Spain, Figure 4 shows a slight increase for the period under analysis but with several periods of alternation. The positive growth monthly of Spanish GDP per capita was close to 1% between September 2006 and September 2010. Thereafter notices a slowdown, which also occurred between January 2008 and July 2009 (financial crisis) and thought to increase over the coming years.

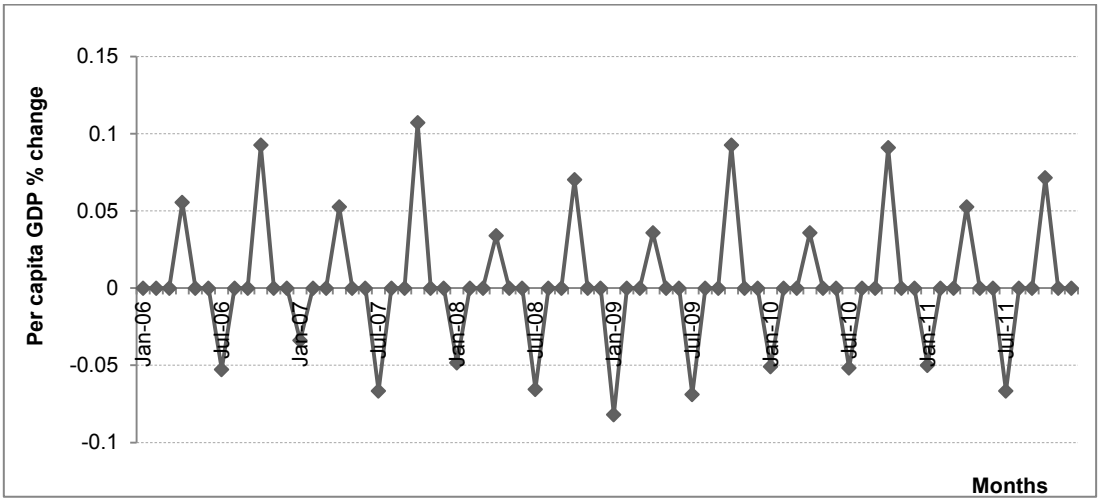


***Figure 6.*** Percentage of Change in GDP per capita of Spain.

The Figure 4 shows that France had a positive trend of growth that because its GDP per capita rose from € 2,300 in January 2006 to € 2,633 in December 2011, i.e., an annual average growth rate of 2.9%. The period of larger oscillations and falls in GDP of France also coincided with that occurred to Spain (January 2008 to July 2009).

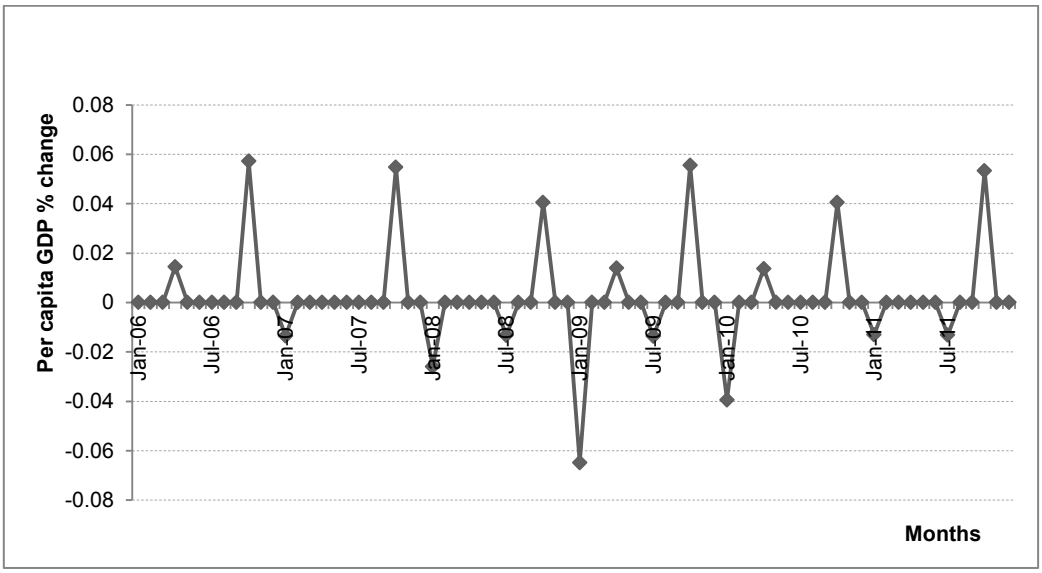


***Figure 7.*** Percentage of Change in GDP per capita of France.

Finally, UK GDP per capita was one that grew less during the period under review. Figures 4 and 8 show a clear downward trend for the period October 2007 to January 2009. This period was critical for all countries, especially Spain and the UK, due to the financial collapse that happened in the United States with the Subprime which quickly spread to other countries who were too exposed to financial markets. Thus, it may be concluded that one of the main causes that led to a significant drop in foreign tourism revenue in accommations establishments and the number of overnight stays in 2009, was the fall in GDP per capita of the countries under review.

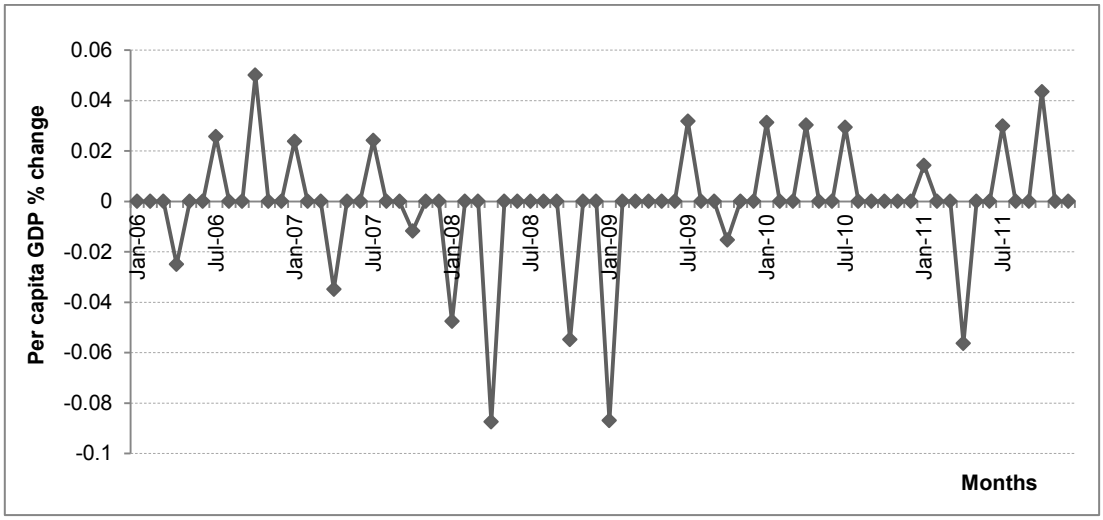


***Figure 8.*** Percentage of Change in GDP per capita of UK.

### 2.1.4 Accommation establishments

Figure 9 shows that the number of hotels increase 118 in January 2006 to 205 in 2011 representing an increase of 74% in 6 years. Instead, the other collective accommodation establishments decreased substantially from 332 to 248, representing a decrease of 25%. This tourism indicator is important to calculate the net bed occupancy rate.

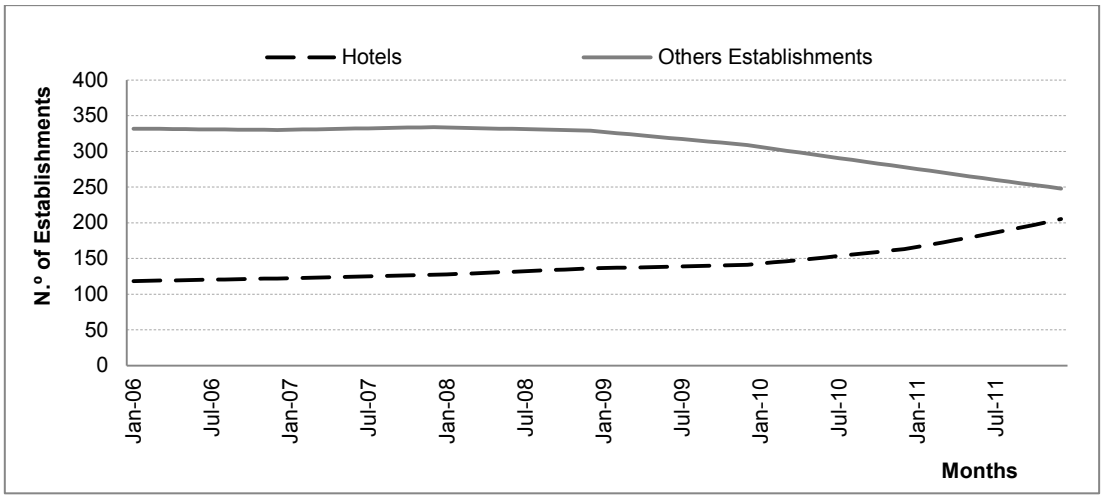


***Figure 9.*** Number of accommodation establishments in the Northern Region.

### 2.1.5 Average length of stay

Analysing graphically the average length of stay of tourists in hotels, it is observed that the domestic tourists spent 1.5 nights and that value has remained stable over time. This research also concluded that the average stay increase during periods of summer (June-September) due to the seasonality effect 'holiday season'. In the months of August 2006 and 2008 reached the highest value with 1.8 overnights. In turn, according to Figure 10, the foreign tourists are on average more nights in hotels than domestic tourists (1.9 nights). It can be observed also that from July 2009 until the end of 2011 there is a trend towards increasing the number of days of stay of foreigners in the North and greater distribution of overnight stays throughout the year. The introduction of the variable "average length of stay" in the model is justified with the importance that it has for the local, regional or national economy development (Santos, 2011). Therefore stay longer in a region, normally induces higher spending from tourists, i.e., it may represent more tourism revenue for commercial and hotels establishments, which can generate increased wealth or GDP to region where the tourists are carrying their trip.

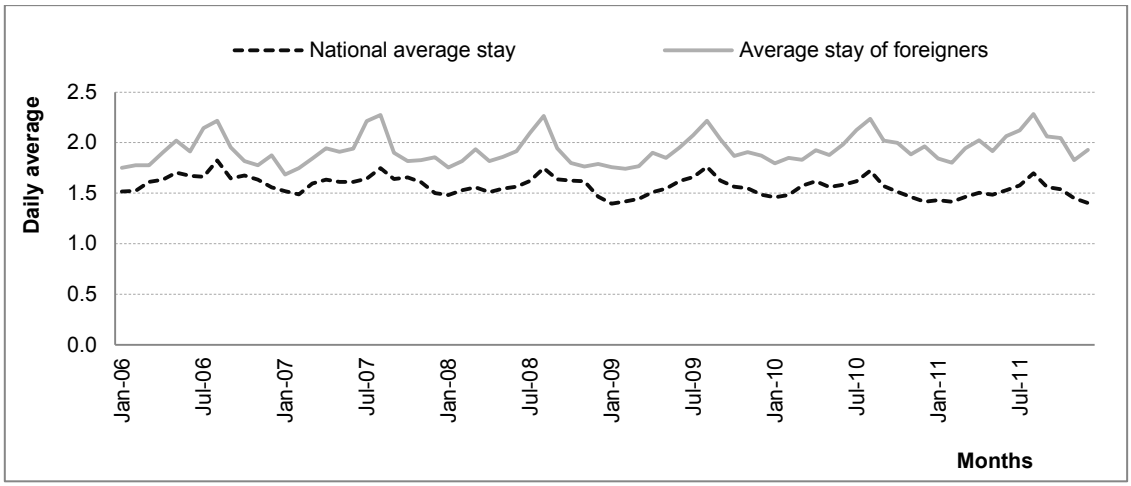


***Figure 10.*** Average length of stay of tourists' accommodation.

### 2.1.6 Tourist Function Rate

The tourist function rate proposed by Defert (1966) to determine areas receiving tourists and tourism dimension of the regions is based on the following expression (Equation 9), which relates the capacity of tourist accommodation with the number people living in region.

$$Tourist\ Function\ Rate\ = \frac{No.\ of\ beds\ in\ hotel\ establishments}{No.\ of\ local\ residents} \quad [9]$$

This rate indicates the number of beds available or the maximum number of tourists per 100 residents. So let's evaluate the maximum increase of tourists and their potential impact on the ability of the infrastructure and space available for tourism. In the North Region, the Tourist Function Rate had a positive growth over time, increasing from 0.93% in 2006 to 1.09% in 2011 (see Figure 11). Comparing this rate with the recorded in Portugal (2.9%), it may be concluded that there is still a huge potential for increased accommation capacity and hence the need to invest further in tourism establishments in the North Region of Portugal.

In future research it shall be able to perform this analysis by county and tourist destination in order to verify the potential capacity of each municipality in the Northern Region. Simultaneously with the definition of tourism products to these regions shall be able to encourage or discourage further investment in accommodation.

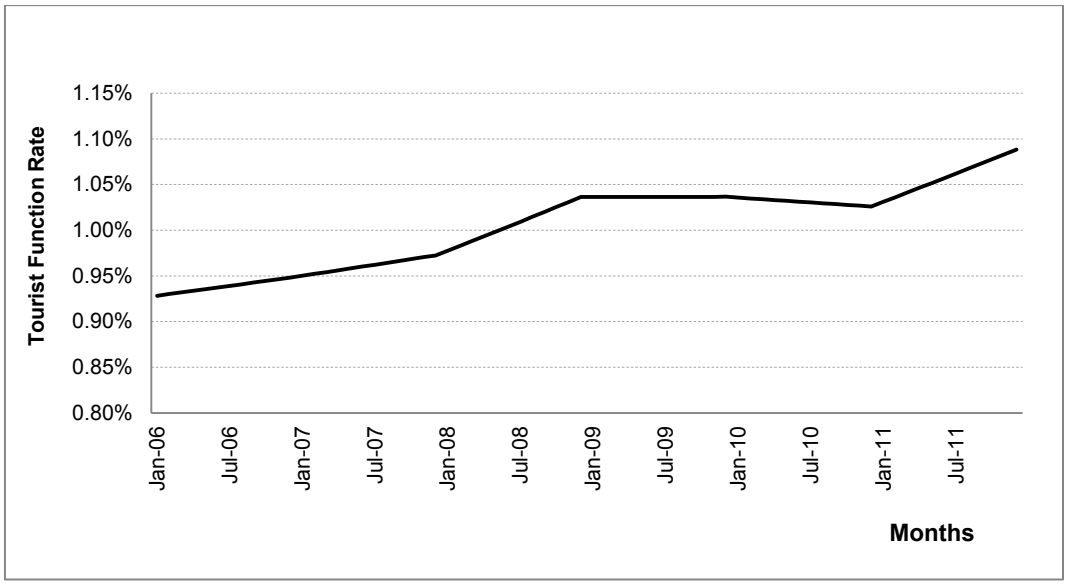


***Figure 11.*** Monthly Tourist Function Rate in the North of Portugal.

## 2.2. The Tourism Revenue Model

To estimate the Tourism Revenue from North Region of Portugal it was used the Multiple Linear Regression Model and its equation is as follows (equation 10):

$$TUR_t = a + b_1 DP_t + b_2 DE_t + b_3 PIBPT_t + b_4 PIBSP_t + b_5 PIBFR_t + \\ + b_6 PIBUK_t + b_7 TON_t + b_8 TOE_t + b_9 NE_t + b_{10} NER_t + \\ + b_{11} PMN_t + b_{12} PME_t + b_{13} TFT_t + \mu_t \quad [10]$$

where,

$Y_t$ - Tourism Revenue at time *t* [$TUR_t$];

$X_{1t}$ - Domestic overnights tourism at time *t* [$DO_t$];

$X_{2t}$ - Foreigner overnights tourism at time *t* [$DE_t$];

$X_{3t}$ - Portugal GDP per capita at time *t* [$PIBPT_t$];

$X_{4t}$ - Spain GDP per capita at the time *t* [$PIBSP_t$];

$X_{5t}$ - France GDP per capita at the time *t* [$PIBFR_t$];

$X_{6t}$ - UK GDP per capita at time *t* [$PIBUK_t$];

$X_{7t}$ - Domestic occupancy rate-bed places at the moment *t* [$TON_t$];

$X_{8t}$ - Foreigner occupancy rate-bed places at the moment *t* [$TOE_t$];

$X_{9t}$ - Number of accomodation establishments-hotels at the moment *t* [$NE_t$];

$X_{10t}$ - Number of accomodation establishments-others at the moment *t* [$NER_t$];

$X_{11t}$ - Domestic overnight average length of stay at time *t* [$PMN_t$];

$X_{12t}$ - Foreign overnight average length of stay at time *t* [$PME_t$];

$X_{13t}$ - Tourist function rate at time *t* [$TFT_t$];

$\mu_t$ - Term of the error at time *t* [$\mu_t$].

## 2.3. Model Estimated by Ordinary Least Squares Method

After the construction of the database was applied method of Ordinary Least Squares to estimate the Multiple Linear Regression Model for Tourism Revenue.

Analysing the data in Table 2, it conclude that the model has high explanatory power in relation to Total Tourism Revenue because the Adjusted R-squared was 0.92. This means that the explanatory variables of the model explain 92% of variations occurred in the variable total tourism revenue.

**Table 2.**

*Performance Measures Multiple Linear Regression Model.*

| | ***Coefficient*** | ***Error Standard*** | ***ratio-t*** | ***p-value*** | |
|---|---|---|---|---|---|
| const | 2,47368e+07 | 3,63955e+07 | 0,6797 | 0,49942 | |
| DP | 212,698 | 303,957 | 0,6998 | 0,48687 | |
| DE | 107,916 | 401,776 | 0,2686 | 0,78919 | |
| PIBPT | 28539 | 8724,5 | 3,2711 | 0,00181 | *** |
| PIBSP | 5940,04 | 5797,33 | 1,0246 | 0,30980 | |
| PIBFR | -31861 | 8399 | -3,7934 | 0,00036 | *** |
| PIBUK | 2377,49 | 2549,62 | 0,9325 | 0,35495 | |
| TON | -1,46776e+08 | 3,43823e+08 | -0,4269 | 0,67104 | |
| TOE | -1,0009e+08 | 4,55751e+08 | -0,2196 | 0,82694 | |
| NE | 170318 | 78877,7 | 2,1593 | 0,03498 | ** |
| NER | 84062,1 | 47246,1 | 1,7792 | 0,08044 | * |
| PMN | -8,49098e+06 | 4,66201e+06 | -1,8213 | 0,07372 | * |
| PME | -2,6433e+06 | 2,96182e+06 | -0,8925 | 0,37584 | |
| TFT | -3,37509e+09 | 4,58343e+09 | -0,7364 | 0,46448 | |

| | | | |
|---|---|---|---|
| Average var. dependent | 17476806 | S.D. var. dependent | 4582188 |
| Sum squared residuals | 9,94e+13 | E.P. regression | 1308843 |
| R-squared | 0,933350 | Adjusted R-squared | 0,918411 |
| F(11, 59) | 62,47844 | P value (F) | 3,68e-29 |
| Log of likelihood | -1108,475 | Akaike criterion | 2244,949 |
| Schwarz criterion | 2276,823 | Hannan-Quinn criterion | 2257,638 |
| rho | -0,028448 | Durbin-Watson | 1,984545 |

Note: *p<0,10; **p<0,05; ***p<0,01.

After estimated the Multiple Linear Regression Model is necessary to test whether or not the violation of model assumptions. Thus then analyzes each of the main shift assumptions of the model and presents the results:

- It appears that there is violation of the assumption of Multicollinearity, since the VIF values are above 10 values for almost all the explanatory variables, except PIBSP, PMN and PME (see Table 3). Concluded that there are explanatory variables that are correlated with each other.

**Table 3.**
*VIF of Multiple Linear Regression Model.*

| **Variables** | ***VIF*** |
|---|---|
| DP | 8333,049 |
| DE | 9015,051 |
| PIBPT | 10,658 |
| PIBSP | 8,404 |
| PIBFR | 18,8 |
| PIBRUK | 16,84 |
| TON | 8139,914 |
| TOE | 8911,611 |
| NE | 132,971 |
| NER | 67,425 |
| PMN | 7,844 |
| PME | 7,77 |
| TFT | 173,117 |

- Testing normality of the error term by the test statistic $\chi^2$=0.394 (see Figure 12), it may conclude that this model follows a normal distribution with a significance level of 1% (p-value=0.8210). The Figure 12 also shows graphically the mean equals μ=-1.0089 e-010. This value is approximately zero, then it can be concluded that the assumption of zero vector is not violated;

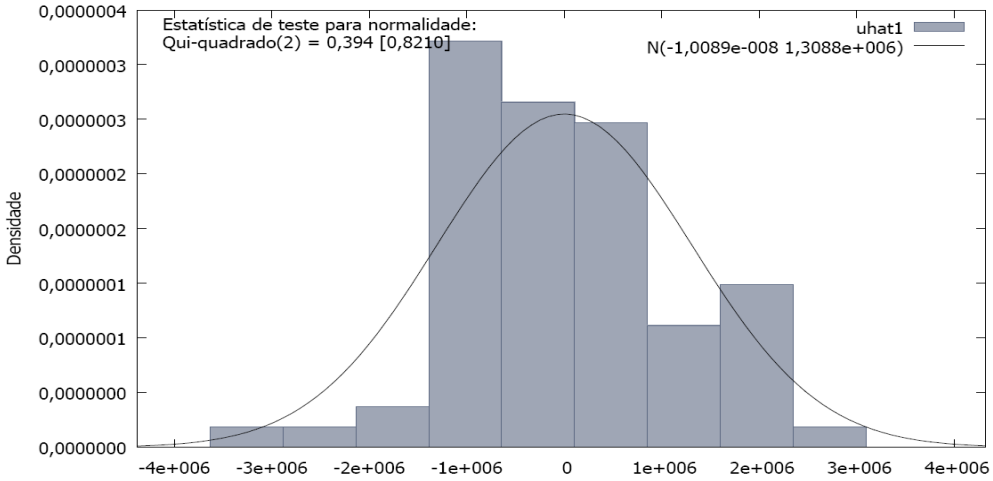


***Figure 12.*** Normal distribution of the residues of Multiple Linear Regression Model.

- Test White was used to test the homoscedasticity of error term. As the test statistic obtained was $TR^2$ =37.243196 was higher than the p-value=0.071065, it can be concluded that there is no violation of the assumption of homoscedasticity. The characteristics of the estimators OLS remain BLUE;
- It was obtained the following Durbin-Watson Statistic=1.984545 which was calculatead for twelve independent variables (excluding the constant). So its possible say this model doesn't have autocorrelation of error terms.

It was found the presence of Multicollinearity among explanatory variables on the original model, so it required estimating a new model that would guarantee the assumptions through the OLS method, which is guaranteeing estimators *Best Linear Unbiased Estimators* (BLUE). Therefore, it was chosen to estimate the First Differences Model (FDM) that, according to Fernandes (2005) and Johnston and DiNardo (2000), can be obtained by the method of differentiation regular series many times as necessary. The model is given by the following expression (equation 11):

$$\Delta Y_t = b_{0t} + b_{1t}\Delta X_{1t} + b_{2t}\Delta X_{2t} + \ldots + b_{pt}\Delta X_{pt} + \Delta \mu_t \quad [11]$$

where:

$$\Delta Y_t = Y_t - Y_{t-1} \quad [12]$$

$$\Delta X_{pt} = X_{pt} - X_{pt-1} \quad [13]$$

$$\Delta \mu_t = \mu_t - \mu_{t-1} \quad [14]$$

Thus, the First Differences Model includes lagged values and current values or one or more explanatory variables lagged in the regressors (Johnston & DiNardo, 2000). In this particular case is given by the following expression:

$$\Delta TUR_t = a + b_1 \Delta DP_t + b_2 \Delta DE_t + b_3 \Delta PIBPT_t + b_4 \Delta PIBSP_t + b_5 \Delta PIBFR_t + b_6 \Delta PIBUK_t + b_7 \Delta TON_t + b_8 \Delta TOE_t + b_9 \Delta NE_t + b_{10} \Delta NER_t + b_{11} \Delta PMN_t + b_{12} \Delta PME_t + b_{13} \Delta TFT_t + \Delta \mu_t \quad [15]$$

Using *Gretl*[1] econometric program, it was estimated the regression of FDM with the implementation of OLS and obtained an Adjusted R-Squared of 76.2%. It means that 76.2% of the variations that occurred in total tourism revenue were explained by variations that occurred in the independent variables in month t and previous month (*t-1*).

After having tested the assumptions of the OLS estimator, it is concluded that the model of the First Differences can not be used because there is violation of the assumption of multicollinearity of the explanatory variables and violation of the independence of errors.

The VIF values for multicollinearity indicate that the correlated variables of the model are correlated with variables TON and TOE. One reason for this is related with estimation model the logarithm of variables by OLS which indentify and removed the TOE by perfect multicollinearity.

The final model chosen does not include variables TON and TOE and called First Differences Logarithm Model, which is nothing more that a lag of one period in logarithmic variables. This model can overcome the multicollinearity of the explanatory variables and violation of the independence of errors.

The Table 4 presents the results obtained for the First Differences Logarithm Model.

The Adjusted R-Squared was 0.79 and using the significance test for all variables F (11.59)=24.92707 for a significance level of 1%, we can conclude that there is adequate statistical evidence for assert that the regressors have different values of zero and together satisfactorily explain the variations in Total Tourism Revenue in Hotels.

[1] Open-source statistical package.

**Table 4.**
*Performance Measures* First Differences Logarithm Model.

| | ***Coefficient*** | ***Error Standard*** | ***ratio-t*** | ***p-value*** | |
|---|---|---|---|---|---|
| const | 0,00460865 | 0,018405 | 0,2504 | 0,80315 | |
| d_l_DP | 0,834053 | 0,153919 | 5,4188 | <0,00001 | *** |
| d_l_DE | 0,170947 | 0,0655265 | 2,6088 | 0,01149 | ** |
| d_l_PIBPT | 1,02014 | 0,594118 | 1,7171 | 0,09121 | * |
| d_l_PIBSP | -0,47949 | 0,72129 | -0,6648 | 0,50879 | |
| d_l_PIBFR | -1,74777 | 1,48898 | -1,1738 | 0,24519 | |
| d_l_PIBUK | -0,154584 | 0,613225 | -0,2521 | 0,80185 | |
| d_l_NE | 1,72779 | 4,84461 | 0,3566 | 0,72263 | |
| d_l_NER | 2,38194 | 6,85866 | 0,3473 | 0,72961 | |
| d_l_PMN | -0,693299 | 0,384735 | -1,8020 | 0,07665 | * |
| d_l_PME | -0,623187 | 0,265514 | -2,3471 | 0,02230 | ** |
| d_l_TFT | -3,21334 | 7,44831 | -0,4314 | 0,66774 | |

| | | | |
|---|---|---|---|
| Average var. dependent | 0,004493 | S.D. var. dependent | 0,181643 |
| Sum squared residuals | 0,408963 | E.P. regression | 0,083256 |
| R-squared | 0,822928 | Adjusted R-squared | 0,789915 |
| *F*(11, 59) | 24,92707 | P value (F) | 3,32e-18 |
| Log of likelihood | 82,32218 | Akaike criterion | -140,6444 |
| Schwarz criterion | -113,4922 | Hannan-Quinn criterion | -129,8468 |
| rho | -0,389369 | Durbin-Watson | 2,766212 |

Note: *p<0,10; **p<0,05; ***p<0,01.

After checking that the regressors were statistically valid, it was necessary to test whether or not the violation assumptions in the model was presented below:

- There was no violation of the assumption of Multicollinearity because all VIF were below the value 10;
- Testing normality of the residue by test statistic $\chi^2$=8059, it is concluded this model does not follow normal distribution at a significance level of 5% (p-value=0.01778);

  The mean residue equals was approximately zero μ=-4.64222 *e*-019, so concluded the assumption of zero vector is not violated;
- The Test White it was used to test homoscedasticity of the error term. As the statistic test obtained $TR^2$=22.170227 was higher than p-value=0.389758, it concluded there is no violation of the assumption of homoscedasticity;
- Testing the non-presence the autocorrelation of errors term, it was obtained the following statistic Durbin-Watson = 2.766212. In the tables Durbin-Watson statistics for 10 independent variables (excluding the constant), the $d_L$ is equal to 1.12, the $d_U$ equal to 1802, the *4*-$d_U$ equal to 2.198 and finally the *4*-$d_L$ is

equal to 2, 88. Thus, as $4\text{-}d_U$ = 2.198 < DW = 2.77 < $4\text{-}d_L$ = 2.88 means that the statistic test Durbin-Watson was inconclusive for the independence errors terms (Table 4). The application of the Breusch-Godfrey Test shows that effectively occurs violation the assumption of independence errors terms, because the statistic test obtained $TR^2$ =35.970399 was higher than the p-value=8.52 e-005 to a level of 5% significance.

To overcome the presence of autocorrelation errors terms it was used the *Cochrane-Orcutt* Estimation, which is a repeatedly estimation iterative process model, using and incorporating residue after each new estimation even eliminate the autocorrelation. It concludes with the use of the test that the Adjusted R-Squared was 86.7% (see Table 5), higher than that for First Differences Model and below the Original Model.

**Table 5.**
*Performance Measures* First Differences Logarithm Model*, Cochrane-Orcutt* estimation.

| | ***Coefficient*** | ***Error Standard*** | ***ratio-t*** | ***p-value*** | |
|---|---|---|---|---|---|
| const | 0,00395635 | 0,00952739 | 0,4153 | 0,67948 | |
| d_l_DP | 0,905274 | 0,141854 | 6,3817 | <0,00001 | *** |
| d_l_DE | 0,151031 | 0,0572846 | 2,6365 | 0,01073 | ** |
| d_l_PIBPT | 1,73471 | 0,5648 | 3,0714 | 0,00324 | *** |
| d_l_PIBSP | 0,528653 | 0,57293 | 0,9227 | 0,35998 | |
| d_l_PIBFR | -4,29439 | 1,19533 | -3,5926 | 0,00068 | *** |
| d_l_PIBUK | -0,0680283 | 0,467448 | -0,1455 | 0,88480 | |
| d_l_NE | 1,84758 | 2,55902 | 0,7220 | 0,47320 | |
| d_l_NER | 2,13007 | 3,56888 | 0,5968 | 0,55293 | |
| d_l_PMN | -0,525635 | 0,357054 | -1,4721 | 0,14639 | |
| d_l_PME | -0,705504 | 0,213291 | -3,3077 | 0,00162 | *** |
| d_l_TFT | -2,57176 | 4,08471 | -0,6296 | 0,53143 | |

Statistics based on the differentiated data-rho:

| | | | |
|---|---|---|---|
| Average var. dependent | 0,004389 | S.D. var. dependent | 0,182952 |
| Sum squared residuals | 0,271684 | E.P. regression | 0,068441 |
| R-squared | 0,887746 | Adjusted R-squared | 0,866457 |
| F(11, 59) | 73,31534 | P value (F) | 7,59e-30 |
| rho | -0,089347 | Durbin-Watson | 2,108600 |

Based on the results obtained in Table 5, the regressors' interpretation of the First Differences Logarithms Model can be:

- If domestic and foreign overnight stays vary by 1%, in total tourism revenue increase in North Region hotels, respectively, 0.905 and 0.151 units;

- If PIBPT and PIBSP per capita increase by 1%, in total tourism revenue will increase by 1.73 and 0.528 units, respectively and if PIBFR and PIBUK per capita increase by 1%, in total tourism revenue will decrease by 4.29 and 0.068 units, respectively;
- If the number of accommodation establishments (hotels and other) increased by 1%, in total tourism revenue will increase by 1.85 and 2.13 units, respectively;
- If domestic overnight average length of stay (PMN) and foreign (PME) growth 1%, in total tourism revenue in hotels will decrease, respectively, 0.526 and 0.706 units;
- If the Tourist Function Rate increase 1%, in total tourism revenue in hotels will decrease in 2.572 units;
- Only variables overnights domestic and foreign, PMN and PIBFR were statistically significant for a significance level of 5% and the PIBPT variable for a significance level of 1%. The remaining variables have no explanatory power for a significance level of 5%.

The model guarantees the non-presence of multicollinearity explanatory variables of the First Differences of Logarithms Model because all VIF were below 10.

Finally, through test statistic $\chi^2$=2578 with p-value=0.2756 concluded there was normality of errors terms, so this model follows a normal distribution at a level of significance of 5%, then this assumption is not violated.

In conclusion, with the interpretation of the results presented above, the best model for estimation the Total Tourism Revenue in North Region of Portugal will be the First Differences Logarithm Model with the exclusion of variables TON and TOE and with the *Cochrane-Orcutt* estimation.

**Conclusions and Future Research**

As mentioned the aim of this study it was to build a model that allows estimating the Total Tourism Revenue for the North Region of Portugal. To achieve this model, it was useded the theoretical models of Tourism Supply and Demand to select the possible variables to include in the model. Thus, were selected 13 independent

variables and collected 72 monthly observations of North Region for the period between January 2006 and December 2011.

Although the inclusion a lag in the model has solved significantly the problem of multicollinearity, it originated autocorrelation on the errors terms in the new model. It was decided, therefore, to exclude the variables that originated the multicollinearity of the explanatory variables which were TON and TOE.

The estimation of First Differences Logarithms Model without variables TON and TOE allowed the elimination of multicollinearity but not a violation of the independence of errors terms. The application of the Cochrane-Orcutt helped to eliminate the autocorrelation of the error terms and obtained a Adjusted R-Squared of 79%, demonstrating the high quality of the fit of the model build to explain the dependent variable Total Tourism Revenue for the North Region of Portugal.

The model estimated to tourism revenue suffered some limitations however throughout their study. One limitation of this study concerned with the lack of monthly data for Investment apply in the North Region during the study period. The lack of data for the Rural Tourism, not allowed the extension of this study to all accommodation establishments in the North Region of Portugal.

Looking for these limitations, it is suggested that in future investigations should try to overcome these limitations, taking into account the results obtained in this study. One of the suggestions to improve the Tourism Revenue Model will go through the inclusion of other variables such as Human Capital in hotels, Number of accommodation establishments with online booking, marketing investment by property or a variable dummy that measure the impact of tourism strategic plan for the region. Another line of investigation is the possible application of this analysis by city or county or tourist destination in order to verify the potential tourism capacity of each city or county in the North Region of Portugal. At the same time, it is possible development tourism products to these regions and to encourage or discourage further investment in accommodation.

## References

Defert, P. (1966). *La localisation touristique: problemes theoriques et practiques*. Gurten. Berna.

Fernandes, P. (2005). *Modelación, Predicción y Análisis del Comportamiento de la Demanda Turística en la Región Norte de Portugal*. Dissertação de Doutoramento, Universidad de Valladolid, Espanha.

Fernandes, P., Monte, A., & Castro, J. (2004). A Região Norte de Portugal e a preferência da procura turística: Litoral versus Interior. *Revista Portuguesa de Estudos Regionais INE/APDR,* 4*,* 57-73.

Frank, R. (2000), *Microeconomics and Behavior. (*4.ª ed.*)*. New York. McGraw-Hill.

Gujarati, D. (1995). *Basic Econometrics*. (3.ª ed.). McGraw-Hill.

Gunst, R., & Mason, L. (1980). *Regression Analysis and Its Application: A Data-Oriented Approach*. Marcel Dekker, New York.

INE, (2006-2011). *Estatísticas do Turismo*. Instituto Nacional de Estatística. Lisboa, Portugal.

INE, (2007-2012). *Anuário Estatístico da Região Norte (2006 a 2011)*. Instituto Nacional de Estatística. Lisboa, Portugal.

INE, (2012). *Estatísticas do Turismo 2011*, Instituto Nacional de Estatística. Lisboa, Portugal.

Johnston, J., & Dinardo, J. (2000). *Métodos Econométricos*. (4.ª ed.), McGraw-Hill.

Lage, B., & Milone, P. (2001). *Economia do Turismo.* (7.ª ed.). São Paulo: Atlas S.A.

Leitão, N. (2011). Tourism and Economic Growth: A Panel Data Approach, Instituto Politécnico de Santarém. *Actual Problems of Economics*, 9, 343-349.

Makhlouf, H. (2012). The Multi-Dimensional Impact Of International Tourism. *International Business & Economics Research Journal*, University of the District of Columbia, USA. *11*(2), 233-241.

Maroco, J. (2003). *Análise Estatística com utilização do SPSS*. Lisboa: Edições Sílabo, Lda.

Oliveira, M., Aguiar, A., Carvalho, A., Martins, F., Mendes, V., & Portugal, P. (1997). *Econometria Exercícios*. Lisboa: McGraw-Hill.

Pestana, M., & Gageiro, J. (2008). *Análise de Dados para Ciências Sociais. A complementaridade do SPSS*. (5ª Edição); Lisboa: Edições Sílabo, Lda.

POFC (2012). *Projectos aprovados*. Programa Operacional Factores de Competitividade, Consultado en 25 Agosto 2012, Ministério da Economia e do Emprego: http://www.pofc.qren.pt/.

POPH (2012). *Apoios Concedidos*. Programa Operacional Potencial Humano, Consultado en 10 Junio 2012, Ministério da Economia e do Emprego: http://www.pofc.qren.pt/.

PRODER, (2012). *Projectos Contratados*. Programa de Desenvolvimento Rural, Consultado en 10 Junio 2012, Ministério da Agricultura, do Mar, do Ambiente e do Ordenamento do Território: http://www.proder.pt/conteudo.aspx?menuid=2068.

Santos, N., & Fernandes, P. (2010). Análise e Previsão da Procura Turística na Região Norte de Portugal. *Proceedings of APDR*. Madeira, Portugal.

Santos, N. (2011). *Modelação da Procura Turística para a Região Norte de Portugal - Dissertação de Mestrado*, Instituto Politécnico de Bragança. Bragança.

Turismo de Portugal, (2007). *Plano Estratégico Nacional do Turismo - para o desenvolvimento do turismo em Portugal*. Turismo de Portugal, I.P. Lisboa.

Turismo de Portugal, (2008). Conceitos Estatísticos para o Turismo, obtido em 25 Julho 2012, Turismo de Portugal,IP: www.turismodeportugal.pt/.

Zhihua, S., & Qihua, W. (2009). Checking the adequacy of a general linear model with responses missing at random. *Journal of Statistical Planning and Inference*. 139, 3588-3604.