Research paper

Positioning and clustering of the world's top tourist destinations by means of dimensionality reduction techniques for categorical data

Oscar Claveria [*,1], Alessio Poluzzi

[a] *AQR-IREA, University of Barcelona, 08034 Barcelona, Spain*
[b] *International Tourism Destinations Consultant, 08034 Barcelona, Spain*

ARTICLE INFO

ABSTRACT

This study aims to cluster the world's top tourist destinations based on the growth of the main tourism indicators over the period between 2000 and 2010. It ranks the destinations with respect to the average growth rate over the sample period. The results find that both China and Turkey are at the top of the rankings of all variables. By assigning a numerical value to each country corresponding to its position, a Spearman's coefficient is calculated and a negative correlation found between a destination's dependency on tourism and the profitability of the tourism activity. Finally, several multivariate techniques for dimensionality reduction are used to cluster all destinations according to their positioning. Three groups are obtained: China, Turkey, and the rest of the destinations. These results show that the persistent growth of the tourism industry poses different challenges in different markets regarding destination marketing and management.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Tourism is one of the most important economic activities worldwide. Travel and passenger transport represent 30% of the world's exports of services, and 6% of overall exports of goods and services (UNWTO, 2015). While other commodity prices show decreasing prices, international tourism expenditure increased by 3.7% in real terms in 2014 (UNWTO, 2015). As a result, tourist destinations have to make major efforts in order to develop and manage their brand within an increasingly competitive market (Datzira & Poluzzi, 2014; Mariani, Buhalis, Longhi & Vitouladiti, 2014; Wang & Pizam, 2011).

This study proposes a methodology to position and cluster tourist destinations according to the evolution of their main tourism indicators. The paper has a dual purpose. First, the study aims to contribute to destination research literature by analysing how the dynamic interactions between the main tourism indicators ultimately affect the positioning of destinations. Second, the paper aims to highlight the utility of multivariate techniques for destination marketing and management. With this aim, two different dimensionality reduction techniques for categorical variables: categorical principal component analysis (CATPCA) and

multidimensional scaling (MDS). To the authors' knowledge, this is the first study to compare the performance of both techniques in the clustering of tourist destinations.

First, a descriptive analysis of the annual percentage growth rates of the tourism indicators is conducted over the period 2000 to 2010. The analysis is complemented by graphing the evolution of the series so as to visually represent the co-movements between tourism variables and economic growth. The world's top tourist destinations are then ranked according to their average growth rate over this period. By assigning a numerical value to each destination corresponding to its position in the rankings, the relationship between all the variables by means of the Spearman correlation coefficient is analysed. Finally, the destinations are clustered according to their position in the rankings by means of two multivariate techniques of optimal scaling.

Data from the Compendium of Tourism Statistics provided by the World Tourism Organization (UNWTO – http://www2.unwto.org/content/data-0) are used. Data include the annual number of international overnight visitors, total expenditure, total number of rooms, and the percentage of the occupancy rates from 2000 to 2010. Table 1 presents the frequency distribution of international overnight visitors in the top ten world destinations during the sample period.

The information in Table 1 indicates that the international tourism sector is highly concentrated in few destinations, as the first five national markets (France, Spain, the United States, China and Italy) account for almost 50% of all international tourism. The

* Correspondence to: University of Barcelona, 08034 Barcelona, Spain.
E-mail addresses: oclaveria@ub.edu (O. Claveria), a.poluzzi@tin.it (A. Poluzzi).
[1] AQR-IREA, Department of Econometrics and Statistics, University of Barcelona, Diagonal 690, 08034 Barcelona, Spain. Tel.: +34 934021825; fax.: +34 934021821.

Table 1
Frequency distribution of international overnight visitors (2000-2010)

Country	Annual mean	Total 2000-2010	Relative frequency	Cumulative relative frequency
France	76,934	846,272	14.4%	14.4%
Spain	53,019	583,214	9.9%	24.4%
United States	50,719	557,911	9.5%	33.9%
China	44,269	486,960	8.3%	42.2%
Italy	40,731	448,042	7.6%	49.8%
United Kingdom	26,470	291,165	5.0%	54.8%
Germany	21,711	238,818	4.1%	58.9%
Mexico	21,167	232,842	4.0%	62.8%
Turkey	19,998	219,980	3.8%	66.6%
Austria	19,956	219,513	3.7%	70.3%

Note: International overnight visitors are measured in thousands.

next five destinations (United Kingdom, Germany, Mexico, Turkey and Austria) represent an additional 20% of total international overnight visitors.

Tourism demand is predominantly measured by the number of arrivals and the level of tourism expenditure. Some authors have made use of the length of stay (Claveria & Datzira, 2010). Given that ratios provide insight into the profitability and the sustainability of tourism activities, this study calculates the ratio of expenditure per tourist as a proxy of tourism demand.

As pointed out by Song, Dwyer, Li, and Cao (2012), one of the problems with the existing tourism literature is the omission of economic indicators and the lack of attention paid to economic return. With the aim of covering this deficit, this study incorporates economic information. We use the annual percentage growth rates of gross domestic product (GDP) and of total inbound expenditure over GDP. Additionally, we introduce the average growth of the human development index (HDI) so as to assess the potential effect of tourism on development beyond a strictly economic sense. Using annual percentage growth rates instead of levels avoids the issues derived from working with non-stationary time series, since most tourism variables are non-stationary due to the steady growth in tourism (Chu, Yeh & Chang, 2014).

On the one hand, this approach allows the use of official data. On the other hand, by ranking the destinations according to their average growth during the sample period, a dynamic perspective is indirectly added into the analysis. Given the changing tourism market, this methodology could be used in destination positioning studies, and consequently contribute to enhancing and sustaining destination competitiveness.

The remainder of this study is organized as follows. The next section reviews the existing literature. Section 3 presents the descriptive and graphical analysis of the data. In the next section, the rankings and the results of the correlation analysis are presented. In Section 5 the results of the multivariate analysis are discussed. The final section summarizes the findings and the limitations of the study, and offers suggestions for further research.

2. Literature review

As suggested by Pike (2008), the improved understanding of the market conditions allows the pursuit of competitiveness and sustainability at a destination level. There is abundant literature on the contribution of tourism to economic growth as well as to destination competitiveness (Balaguer & Cantavella-Jordá, 2002; Capó, Riera & Rosselló, 2007; Chou, 2007; Croes, 2011; Crouch & Richie, 2006; Durbarry, 2004; Oh, 2005; Pérez-Rodríguez, Ledesma-Rodríguez & Santana-Gallego, 2015; Schubert & Brida, 2009; Schubert, Brida & Risso, 2011). Skerritt and Huybers (2005)

examine the net effect of international tourism on GDP per capita in 37 developing economies, finding that tourism positively affects economic development. Tang and Tan (2005) test the tourism-led growth hypothesis in Malaysia and find that tourism is an effective long-term engine of growth. In a similar study, Hye and Khan (2005) confirm the long-run relationship between income from tourism and economic growth in Pakistan.

Recent literature highlights the role of capital formation, arguing that the mechanism underlying tourism's welfare-promoting effect heavily relies on capital goods imports (Cortes-Jimenez, Nowak & Sahli, 2011; Nowak, Sahli & Cortes-Jimenez, 2007). Foreign direct investment, trade volume, and exchange rates are also linked to tourism (Santana-Gallego, Ledesma-Rodríguez & Pérez-Rodríguez, 2010, 2011; Wong & Tang, 2010).

The most commonly considered determinants of tourism demand are the income of origin, the prices in the destination, and the substitute prices of alternative destinations (Song, Witt & Li, 2009). An additional variable that affects tourists' decisions is the marketing expenditure at the destination level (Kulendran & Dwyer, 2009; Zhang, Kulendran & Song, 2010). Globalisation has led to increasing market interdependence, as tourism demand in one destination tends to be affected by demand for alternative destinations. These interdependences have been addressed mostly by means of vector autoregression (VAR) models and co-integration techniques (Seo, Park & Boo, 2010; Torraleja, Vázquez & Franco, 2009), while a few studies have used multivariate techniques (Chandra & Menezes, 2001).

Multivariate analysis techniques can be classified into two major categories: dependency and interdependency techniques. While the former assume that a set of variables is explained by other variables, the latter involve the simultaneous analysis of all the variables in the dataset. Cluster analysis is an example of an interdependent procedure. The main purpose of these techniques is to reduce the dimensionality, and to detect underlying structures in the relationships between variables.

There are several multivariate techniques for dimensionality reduction, e.g. cluster analysis, multiple correspondence analysis (MCA), exploratory factor analysis (EFA), confirmatory factor analysis (CFA), principal components analysis (PCA). For a detailed description of these techniques see Hair, Black, Babin, and Anderson (2009), Jolliffe (2002), and Sharma (1996).

Dimensionality reduction techniques have been used in a wide range of tourism studies: from image and perception analyses to motivation studies. One of the main areas in which multivariate analysis is widely used is market segmentation studies (Dey & Sarma, 2010; Donaire, Camprubí & Galí, 2014; Keng & Cheng, 1999; Lee, Lee, Bernhard & Yoon, 2006; Park & Yoon, 2009; Rid, Ezeuduji & Pröbstl-Haider, 2014; Sinclari-Maragh, Gursoy & Vieregge, 2015; Upchurch, Ellis & Seo, 2004; Voges, 2007). Guo, Guillet, Kucukusta, and Law (2015) conduct conjoint and a cluster analyses to segment Chinese spa customers in Hong Kong. Arimond and Elfessi (2001) use MCA to spatially map attributes from a categorical survey data, and then cluster analysis to identify market segments.

In a recent study, Marcussen (2014) reviews 64 papers, published between 1975 and 2014, that apply MDS to tourism research. MDS is also known as Principal Coordinates Analysis or Torgerson scaling (Torgerson, 1952, 1958). The author uses MDS to evaluate the application of the technique in the tourism literature, finding that the most common themes are image and positioning of destinations.

For a review of the literature on destination image see Pike (2002), who reviewed 142 destination image papers published in the literature over the period 1973 to 2000. Gallarza, Gil, and Calderon (2002) propose a taxonomy of the methodological and statistical procedures for measuring the image of the destinations. Phau, Shanka, and Dhayan (2010) use factor analysis (FA) to