Tourism Management Perspectives 16 (2015) 116–122

Contents lists available at ScienceDirect

Tourism Management Perspectives

journal homepage: www.elsevier.com/locate/tmp

Common trends in international tourism demand: Are they useful to improve tourism predictions?

Oscar Claveria [a,*], Enric Monte [b], Salvador Torra [c]

[a] AQR-IREA (Institute of Applied Economics Research), University of Barcelona, Spain
[b] Department of Signal Theory and Communications, Polytechnic University of Catalunya, Spain
[c] Department of Econometrics, Riskcenter-IREA, University of Barcelona, Spain

ARTICLE INFO

Article history:
Received 1 July 2015
Received in revised form 21 July 2015
Accepted 26 July 2015
Available online xxxx

Keywords:
Tourism demand
Forecasting
Multivariate
Multiple-output
Artificial Neural Networks

ABSTRACT

This study evaluates whether modelling the existing common trends in tourist arrivals from all visitor markets to a specific destination can improve tourism predictions. While most tourism forecasting research focuses on univariate methods, we compare the performance of three different Artificial Neural Networks in a multivariate setting that takes into account the correlations in the evolution of inbound international tourism demand to Catalonia (Spain). We find that the multivariate multiple-output approach does not outperform the forecasting accuracy of the networks when applied country by country, but it significantly improves the forecasting performance for total tourist arrivals.

1. Introduction

Tourism demand forecasting has become essential in one of today's fastest growing economic activities. Song and Li (2008) have acknowledged the importance of applying new approaches to tourism demand forecasting in order to improve the accuracy and the performance of the methods of analysis. Whilst most research efforts focus on conventional tourism forecasting methods (Gounopoulos et al., 2012) or a combination of them (Chan et al., 2010), in recent years the availability of more advanced forecasting techniques and the requirement for more accurate forecasts of tourism demand have led to a growing interest in Artificial Intelligence (AI) techniques (Wu, Law & Xu, 2012; Pai et al., 2014). The suitability of AI models to handle nonlinear behaviour is one of the reasons why Artificial Neural Networks (ANNs) are increasingly used for forecasting purposes.

In spite of the increasing interest in AI methods for time series forecasting, very few studies compare the accuracy of different ANN architectures for tourism demand forecasting. This study seeks to break new ground by comparing the performance of three different ANN models in a multivariate setting that takes into account the common trends in inbound international tourism demand shared by all visitor markets to a specific destination. We use three ANNs: the multi-layer perceptron (MLP) network, the radial basis function (RBF) network and the Elman network. ANNs are able to learn from experience. Each ANN architecture handles information in a different manner, so by comparing the different models we can evaluate the impact of alternative ways of processing data on forecast accuracy.

Given that univariate specifications are limited and unable to capture dynamic interrelationships between different countries of origin, we analyse whether a multivariate approach, in which information about tourist arrivals from all origin countries is simultaneously used, proves useful for forecasting purposes. With this aim, we design a multiple-output setting and compare the performance of three different ANN models. In order to evaluate the forecasting performance of the multivariate ANNs we also compare the forecasting accuracy of multiple-output predictions to those obtained country by country. To our knowledge, this is the first study to model tourism demand incorporating the common trends in international tourist arrivals from all visitor markets to a specific destination and to analyse whether this approach allows to improve the forecasting performance of ANN models.

The present study deals with tourist arrivals to Catalonia, which is a region of Spain. After France and the United States, Spain is the third most important destination of the world with almost 65 million tourist arrivals in 2014. Catalonia received 25% of all inbound tourist demand to Spain. Tourist spending in Catalonia grew by 14% in 2014. Barcelona is the capital of Catalonia, and the third most important destination in Europe in terms of tourist spending after London and Paris. It follows that tourism is one of the fastest growing industries in Catalonia, accounting for 12% of GDP and providing employment for 15% of the

* Corresponding author at: Department of Econometrics, University of Barcelona, Diagonal 690, 08034 Barcelona, Spain.
E-mail addresses: oclaveria@ub.edu (O. Claveria), enric.monte@upc.edu (E. Monte), storra@ub.edu (S. Torra).

http://dx.doi.org/10.1016/j.tmp.2015.07.013
2211-9736/

working population. These figures show the importance of accurate forecasts of tourism volume at the destination level for policy makers and professionals in the tourism industry. Capó et al. (2007) and Balaguer and Cantavella-Jordá (2002) have shown the important role of tourism in the Spanish long-run economic development.

The article proceeds as follows. The next section reviews the literature on tourism demand forecasting with AI-based techniques. Then we present the different NN architectures used in the analysis and describe the data. In the next section, results of the out-of-sample forecasting competition are discussed. Finally, the last section provides a summary, a discussion of the implications, and potential lines for future research.

2. Tourism demand forecasting with AI-based techniques

A growing body of literature has focused on tourism demand forecasting, but most research efforts apply conventional forecasting methods, either casual econometric models (Cortés-Jiménez & Blake, 2011; Page et al., 2012, Wu et al., 2012b; Lin et al., 2015), time series models (Assaf et al., 2011; Gounopoulos et al., 2012), or a combination of them (Shen et al., 2008; Coshall & Charlesworth 2010). See Song et al. (2012) and Peng et al. (2014) for a thorough review of tourism economics research and tourism demand forecasting studies. Nevertheless, the need for more accurate forecasts has led to an increasing use of AI techniques, such as fuzzy time series models and support vector machines (SVMs), or a mix of them (Pai et al., 2014; Cang & Yu, 2014), in order to obtain more refined predictions of tourist arrivals at the destination level.

Yu and Schwartz (2006) and Huarng et al. (2011) use fuzzy time series models in predicting annual U.S. tourist arrivals and monthly tourism demand in Taiwan respectively. Law et al. (2004) apply a rough sets algorithm to forecast Japanese tourism demand for Hong Kong. The use of genetic algorithms for parameter selection has led to the increased use of support vector machines (SVMs) (Pai & Hong, 2005) and their regression version (Chen & Wang, 2007; Chen, 2011; Hong et al., 2011). In a recent meta-analysis of published tourism forecasting studies, Kim and Schwartz (2013) find that forecast accuracy is closely associated with data characteristics. The fact that ANNs are data-driven procedures that learn from past experience explains the growing interest in ANNs for tourism demand forecasting (Lin et al., 2011; Teixeira & Fernandes, 2012; Claveria & Torra, 2014).

ANNs can be classified into two major types of architectures: feed-forward networks and recurrent networks. MLP networks are the most widely used feed-forward topology in tourism demand forecasting (Pattie & Snyder, 1996; Uysal & El Roubi, 1999; Law, 2000, 2001; Tsaur et al., 2002; Zhang & Qi, 2005). A class of multilayer feed-forward architecture with two layers of processing is the radial basis function (Broomhead & Lowe, 1988). RBF networks have the advantage of not suffering from local minima in the same way as MLP networks, which explains their increasing use in many fields. The first study to implement a RBF ANN for forecasting tourism demand is that of Kon and Turner (2005), who use a RBF network to forecast arrivals to Singapore. Cang (2014) generates predictions of UK inbound tourist arrivals and combines them in non-linear models. Çuhadar et al. (2014) compare the forecasting performance of RBF and MLP NNs to predict cruise tourist demand at the destination level (Izmir, Turkey).

Recurrent networks are models with bidirectional data flow which allow for a temporal feedback from the outer layers to the lower layers. This feature is specially suitable for time series modelling. A special case of recurrent network is the Elman network (Elman, 1990). Whilst MLP networks are increasingly used with forecasting purposes, the Elman neural networks have been scarcely used with forecasting purposes. The first study that uses the Elman ANNs for tourism demand forecasting is that of Cho (2003), who applies the Elman architecture to predict the number of arrivals from different countries to Hong Kong.

Multivariate approaches to tourist demand forecasting are also few and have yielded mixed results. Athanasopoulos and de Silva (2012) compare the forecasting accuracy of exponential smoothing methods in a multivariate setting against univariate alternatives. They use international tourist arrivals to Australia and New Zealand and find that multivariate models improve on forecast accuracy over the univariate alternatives. Contrary to what could be expected, Du Preez and Witt (2003) find that multivariate time series models did not generate more accurate forecasts than univariate time series models.

With regard to studies on tourism in Spain at the regional level, there have been several articles published in recent years (Aguiló & Rosselló, 2005; Rosselló et al., 2005; Garín-Muñoz & Montero-Martín, 2007; Bardolet & Sheldon, 2008; Santana-Jiménez & Hernández, 2011; Nawijn & Mitas, 2012; Andrades-Caldito et al., 2013).

Concerning tourism demand forecasting, Palmer et al. (2006) design a MLP neural network to forecast tourism expenditure in the Balearic Islands. Medeiros et al. (2008) develop a NN-GARCH model to estimate demand for international tourism also in the Balearic Islands. Bermúdez et al. (2009) calculate prediction intervals for hotel occupancy in three provinces in Spain by means of a multivariate exponential smoothing. Gil-Alana (2010) models international monthly arrivals in the Canary Islands using different time-series approaches to analyse the degree of persistence of the series. Claveria and Datzira (2010) use consumer expectations derived from tendency surveys to improve forecasts of tourism demand for Catalonia. Guizzardi and Stacchini (2015) also make use of business sentiment indicators from tendency surveys for real-time forecasting of hotel arrivals at a regional level, improving the forecasting accuracy of structural time series models.

3. Methodology

3.1. Neural network models

ANNs emulate the processing of the human neurological system to identify related spatial and temporal patterns from historical data. ANNs learn from experience and are able to capture functional relationships among the data when the underlying process is unknown. The data generating process of tourist arrivals is too complex to be specified by a single linear algorithm, which explains the great interest that ANNs have aroused for tourism demand forecasting.

ANNs are composed of interconnected processing units called neurons and can also be classified into feed-forward networks and recurrent networks depending on the connecting patterns of the different layers of neurons. In feed-forward networks the information runs only in one direction, whilst in recurrent networks there are feedback connections from outer layers of neurons to lower layers of neurons. ANNs can also be classified according to their learning paradigm: supervised learning and non-supervised learning. MLP networks are supervised learning models, while RBF networks, combine both learning methods (hybrid learning). The MLP network is the most widely used feed-forward topology in tourism demand forecasting.

In this study we use three ANN models: MLP, RBF and the Elman networks. Eqs. (1), (2) and (3) respectively describe the input/output relationship of the three architectures:

MLP

$$\begin{aligned}
&y_t = \beta_0 + \sum_{j=1}^{q} \beta_j g\left(\sum_{i=1}^{p} \varphi_{ij} x_{t-i} + \varphi_{0j}\right) \\
&\{x_{t-i} = (1, x_{t-1}, x_{t-2}, \cdots, x_{t-p})', i = 1, \ldots, p\} \\
&\left\{\varphi_{ij}, i = 1, \ldots, p, j = 1, \cdots, q\right\} \\
&\{\beta_j, j = 1, \ldots, q\}
\end{aligned} \tag{1}$$