Cuadernos de Turismo, nº 48, (2021); pp. 525-529 Universidad de Murcia
eISSN: 1989-4635

EXTENDED ABSTRACT

FORECASTING MODELS FOR THE ANDALUSIAN TOURISM SECTOR USING ADVANCED STATISTICAL METHODS

Francisca J. Sánchez-Sánchez
Universidad Pablo de Olavide
fsansan@upo.es
http://orcid.org/0000-0001-5325-3667
Ana M. Sánchez Sánchez
Universidad Pablo de Olavide
amsansan@upo.es

1. INTRODUCTION

The growing importance of the sector has caused an increasing number of works related to the modeling and forecast of tourist demand, both nationally and internationally (Garín-Muñoz, 2011; Petrevska, 2012; Cuhadar, Cogurcu and Kukrer, 2014; Yang *et al*., 2015). The growing interest in the tourism sector is directly related to the great expansion that has taken place in the industry of the sector, a global effect both in countries with developed economies and in other developing ones (Song, Witt and Jensen, 2003). In order to forecast the expected demand, the tourism demand models have focused on the search for variables that explain the evolution of the tourism sector. The main objective of these models is to determine the reason for the variations that occur in tourist demand, assess the repercussions of public policies and forecast future demand. A detailed review of the literature on tourism demand modeling and forecast can be found in the works of Crouch, 1995; Witt and Witt, 1995; Lim, 2006 and Song and Li, 2008 and Wanhill, 2011.

It would be very useful for companies and organizations to have forecast on tourism demand as accurate as possible, simplifying tactical and strategic decision-making. Despite the consensus on the need to build accurate forecasting models given the benefits they provide, there is no methodology that provides the best model in terms of forecast accuracy (Law and Au, 1999). One of the most widely used methodologies for the forecast of time series is the Box-Jenkins method (Box and Jenkins, 1976), which is based on a linear statistical model known as ARIMA (Autoregressive Integrated Moving Average). An alternative to this methodology are the classic decomposition methods, including the Holt-Winters procedure. Due to the diversity of forecasting methodologies (Song and Li, 2008), some authors advocate the combination of forecasts in order to obtain more precise models and better long-term estimates (Shen, Li and Song, 2011; Wong *et al*., 2007; Song *et al*., 2008).

This work develops a study of the tourist demand in Andalusia in the period 1999-2017. Therefore, in the first place it is necessary to clarify the concept of tourist demand that is used. In the literature, tourism demand has been defined in different ways (see for example Cooper *et al*., 1993; Song and Witt, 2000). As in Cooper *et al*. (1993), tourist demand is considered as the actual or current demand, formed by the total number of visitors who travel to a certain place.

Given the diversity of methodologies available to make forecast, this work aims to determine different forecasting models for the time series "Tourists visiting Andalusia" and to perform an analysis of the forecast capacity of these models to obtain quarterly forecasts.

2. METHODOLOGY

The quarterly time series "Tourists visiting Andalusia" is analyzed, considering the aggregate serie of both national and international visitors. The period studied runs from 1999 to 2017. The data set comes from official statistics published by the Andalusian Institute of Statistics and Cartography (IECA). The methodologies of Box-Jenkins, Holt-Winters and a combined method of the previous ones will be used for the determination of the forecasting models.

2.1. Box-Jenkins

The Box-Jenkins methodology (Box and Jenkins, 1976) is applied to the well-known ARIMA models. The basic argument of this methodology is that the forecasting time series comes from a stochastic process. The objective is to find a mathematical model that collects the behavior of the time series, in order to make forecasts.

In the univariate ARIMA model, which will be the one applied in our study, the behavior of the time series is explained from its own values in the past.

2.2. Holt-Winters

The Holt-Winters method belongs to the classical time series decomposition methodology. This method originally presented by Holt (1959) and by Winters (1960), consists of the decomposition of the time series into 4 components: trend, cyclical variations, seasonal factor and irregular component. The method indicates that the time series can be additive, in which case the fluctuations are not influenced by the trend, or of a multiplicative type, in which case they would be affected.

When the time series presents a multiplicative scheme with seasonality, to eliminate the seasonal component, the ratio method is usually applied to the moving average because it is the most consistent and most widespread. When the series is seasonally adjusted it can be used to make forecast of future values.

2.3. Forecast combination

The combination of different procedures can generate better and more accurate forecasts than those obtained individually from each method (Shen, Li and Song, 2011; Wong *et al*., 2007).

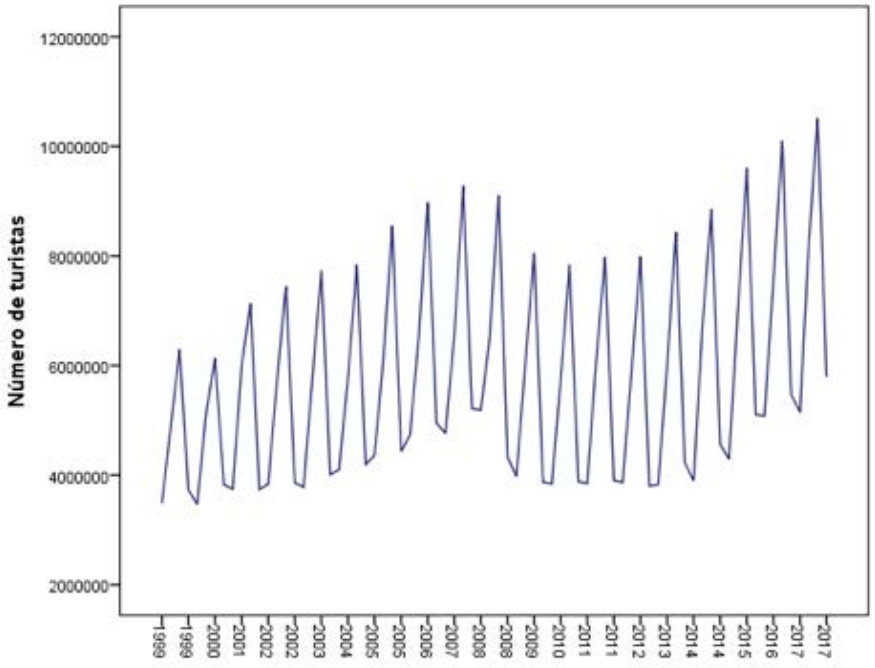
When forecasts are combined, it is usual to apply two types of rules: 1) The arithmetic mean of the forecasts reached by different methods and, 2) the weighted average of the forecasts obtained by the different methodologies, where the weights applied will depend on the relative precision. individual methodologies. In the study, three combined procedures will be applied: the arithmetic mean, the linear and the regression.

2.4. Forecastcability

For the comparison of the forecast capacity of the three proposed methodologies, three statistics of forecast reliability are applied that allow evaluating the forecast errors and selecting the best model: root mean square error (RECM), mean absolute error (EAM) and absolute mean percentage or relative error (EAMP).

3. RESULTS

Figure 1
TOURISTS VISITING ANDALUSIA



Source: Authors' own.

The seasonal factor is very marked in the series of analyzes; this factor forces us to treat the series so that the forecasts are as reliable as possible. According to the applied methodology, this treatment will be different.

Table 1 shows the forecasts for each quarter of 2018, made by applying each of the methodologies considered. The real quarterly values of the series for the year 2018 are also provided. It is observed that the results of the forecasts by any method are very similar to the real ones.

Table 1
FORECASTS OF TOURISTS VISITING ANDALUSIA IN 2018

Quarter	Real tourists year 2018	Forecasts year 2018				
		Box-Jenkins	Holt-Winters	Arithmetic mean combined	Linear weighted combined	Regressionweighted combined
I	5435522	5354243	5342117	5348423	5348180	53743234
II	8248510	8453575	8314197	8386674	8383886	8410119
III	10902313	10710123	10798704	10752642	10754414	10820827
IV	6000632	5871378	5859224	5865544	5865301	5894113

Source: Authors' own.

Table 2 shows the statistics used to measure the goodness of the forecasts made. In general, the differences in the forecast capacity between the three forecasting models are scarce, offering very similar results, although in the application carried out, it would be the Holt-Winters methodology that offers the best results, selecting it to make forecast because it presents better results.

Table 2
RELIABILITY OF THE FORECASTS OF THE NUMBER OF TOURISTS VISITING ANDALUSIA IN 2018

Errors Forecast Statistics	Box-Jenkins	Holt-Winters	Arithmetic mean combined	Linear weighted combined	Regression weighted combined
RECM	301351,424	289342,419	286570,454	286770,642	284189,248
EAM	238275,689	219877,905	225117,643	225360,258	223490,792
EAMP	4,241%	4,037%	4,039%	4,040%	4,041%

Source: Authors' own.

4. CONCLUSIONS

The coexistence in the literature of different procedures for obtaining forecasts of variables related to the tourism sector shows the complexity of selecting a methodology

that adequately adjusts to the needs of the sector. For the forecast of the variable "Tourists visiting Andalusia", the classical Holt-Winters procedure offers better results than the Box-Jenkins method or the combined one, although the three methodologies used show a similar forecast capacity. This result is a consequence of the strong seasonal component of the variable studied, since usually in the tourism sector the values of visits suffer a very high rebound in holiday periods. In order for the series not to be affected by this factor, it would have to be seasonally adjusted beforehand, obtaining much more reliable and exact forecasts. The three methods could be considered valid for decision-making in companies and organizations, from a strategic and tactical point of view.

In our study, the evaluation period of the forecasts made (year 2018) for the series of the number of tourists, confirms that the adjustment of the forecasts is very good, since the results of these forecasts are very similar to the real data.

It would be advisable to continue the search for procedures that allow obtaining better forecasts. For this, other forecasting procedures different from those used here can be applied, such as neural networks or genetic algorithms.