Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

Document de Treball 2014/17, 21 pàg.
Working Paper 2014/17, 21 pag.

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

Document de Treball 2014/10 21 pàg.
Working Paper 2014/10, 21 pag.

"A multivariate neural network approach to tourism demand forecasting"

Oscar Claveria, Enric Monte and Salvador Torra

Research Institute of Applied Economics

Regional Quantitative Analysis Research Group

I|R|E|A|
Institut de Recerca en Economia Aplicada Regional i Públic
Research Institute of Applied Economics
WEBSITE: www.ub-irea.com • CONTACT: irea@ub.edu

AQR
Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group
WEBSITE: www.ub.edu/aqr/ • CONTACT: aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

Abstract

This study compares the performance of different Artificial Neural Networks models for tourist demand forecasting in a multiple-output framework. We test the forecasting accuracy of three different types of architectures: a multi-layer perceptron network, a radial basis function network and an Elman neural network. We use official statistical data of inbound international tourism demand to Catalonia (Spain) from 2001 to 2012. By means of cointegration analysis we find that growth rates of tourist arrivals from all different countries share a common stochastic trend, which leads us to apply a multivariate out-of-sample forecasting comparison. When comparing the forecasting accuracy of the different techniques for each visitor market and for different forecasting horizons, we find that radial basis function models outperform multi-layer perceptron and Elman networks. We repeat the experiment assuming different topologies regarding the number of lags used for concatenation so as to evaluate the effect of the memory on the forecasting results, and we find no significant differences when additional lags are incorporated. These results reveal the suitability of hybrid models such as radial basis functions that combine supervised and unsupervised learning for economic forecasting with seasonal data.

JEL classification: L83; C53; C45; R11

Keywords: forecasting; tourism demand; cointegration; multiple-output; artificial neural networks

Oscar Claveria. AQR Research Group-IREA, Universitat de Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain. E-mail: oclaveria@ub.edu

Enric Monte. Department of Signal Theory and Communications, Polytechnic University of Catalunya. E-mail: enric.monte@upc.edu

Salvador Torra. Riskcenter-IREA, Universitat de Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain. E-mail: storra@ub.edu

I. Introduction

The availability of more advanced forecasting techniques and the requirement for more accurate forecasts of tourism demand at the destination level has led to a growing interest in tourism demand forecasting over the past decades. Despite there is no consensus on the most appropriate approach to forecast tourism demand (Kim and Schwartz, 2013; Song and Li, 2008), it is generally believed that the nonlinear methods outperform the linear methods in modelling economic behaviour (Cang, 2013). As stated by Granger and Terasvirta (1993), real world systems are often nonlinear, so that their responses are not proportional to changes in the inputs.

During the 80s, several nonlinear models time series models were developed. See De Gooijer and Kumar (1992) for a review of this field. These nonlinear models are still limited in that an explicit relationship for the data series has to be assumed with little knowledge of the underlying data generating process. Since there are too many possible nonlinear patterns, the specification of a nonlinear model to a particular data set becomes a difficult task. The suitability of artificial intelligence techniques to handle nonlinear behaviour explain why Artificial Neural Networks (ANNs) have become an essential tool for economic forecasting. ANNs can be regarded as one of the multivariate nonlinear nonparametric statistical methods.

As data characteristics are associated with forecast accuracy (Kim and Schwartz, 2013), nonlinear data-driven approaches such as ANNs represent a flexible tool for forecasting, allowing for nonlinear modelling without a priori knowledge about the relationships between input and output variables. The introduction of the backpropagation algorithm fostered the use of ANNs for forecasting (Santín *et al.*, 2004; Binner *et al.*, 2005; Vlastakis et al., 2008; Madden and Tan, 2008; Lin *et al.*, 2011; Choudhary and Haider, 2012; Teixeira and Fernandes, 2012). Zhang *et al.*(1998) review the literature comparing ANNs with statistical models in time series forecasting.

Many different ANN models have been developed since the 1980s. ANNs can be classified into two major types of architectures depending on the connecting patterns of the different layers: feed-forward networks, where the information runs only in one direction, and recurrent networks, in which there are feedback connections from outer layers of neurons to lower layers of neurons. Feed-forward networks were the first ANNs devised. The most widely used feed-forward topology in time series forecasting is the multi-layer perceptron (MLP) network. MLP networks have been widely used for tourism demand forecasting (Pattie and Snyder, 1996; Uysal and El Roubi, 1999;

Law, 1998, 2000, 2001; Law and Au, 1999, Burger *et al.*, 2001; Tsaur *et al.*, 2002; Kon and Turner, 2005; Palmer *et al.*, 2006; Claveria and Torra, 2014).

A class of multi-layer feed-forward architecture with two layers of processing is the radial basis function (RBF) network (Broomhead and 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. Cang (2013) has recently compared the forecast accuracy of RBF networks to that of MLP and Support Vector Machine (SVM) networks.

Recurrent networks are models with bidirectional data flow: they propagate data linearly from input to output but also allow for a temporal feedback from the outer layers to the lower layers. This feature is specially suitable for time series modelling. There are many recurrent architectures. A special case of recurrent network is the Elman network (Elman, 1990). Whilst MLP networks are increasingly used with forecasting purposes, Elman neural networks have been scarcely used in tourism demand forecasting. Cho (2003) used the Elman architecture to predict the number of arrivals from different countries to Hong Kong.

Regarding their learning strategy, ANNs can also be classified into two major types of architectures: supervised and unsupervised learning networks. In supervised learning networks, weights are adjusted to approximate the output to a target value for each pattern of entry. SVMs and MLP networks are examples of supervised learning models. In non-supervised learning networks, the subjacent structure of data patterns is explored so as to organize such patterns according to their correlations. Kohonen self-organizing maps (SOM) are the most used non-supervised models. Some ANNs combine both learning methods, so part of the weights are determined by a supervised process while the rest are determined by unsupervised learning. This is known as hybrid learning. An example of hybrid model is the RBF network.

In spite of the increasing interest in machine learning methods for time series forecasting, very few studies compare the accuracy of different ANN architectures for tourism demand forecasting. This study focuses on the implementation of three different ANNs (MLP, RBF and Elman) so as to evaluate how different ways of handling information affect forecast accuracy. We use a multiple-output approach to predict international tourism demand in order to compare the forecasting performance of the three different architectures. The motivation for applying a multiple-output framework is twofold. On the one hand, there are no studies analyzing the forecasting performance of multiple-output ANNs. On the other hand, a multivariate approach is especially suited when the evolution of tourist arrivals from all the different countries of origin share a common stochastic trend.

The fact that tourism data is characterised by strong seasonal patterns and volatility, make it a particularly interesting field in which to apply different types of NN architectures. International tourism is one of the fastest growing industries and accounts for almost 10% of total international trade (Eilat and Einav, 2004). Balaguer and Cantavella-Jordá (2002) showed the important role of tourism in the Spanish long-run economic development. Catalonia is a region of Spain and one of the world's major tourist destinations. Tourism represents 12% of Catalonian GDP and provides employment for 15% of the working population. These figures show to what extent accurate forecasts of tourism volume play a major role in tourism planning at the destination level.

We use official statistical data of tourist arrivals from all countries of origin to Catalonia over the period 2001 to 2012. By means of the Johansen test we find correlated accelerations between the different markets, which lead us to apply a multiple-output approach to obtain forecasts of tourism demand for different forecast horizons (1, 3 and 6 months). To assess the effect of expanding the memory on forecast accuracy, we repeat the experiment assuming different topologies with respect to the number of lags used for concatenation. Finally, we compute several measures of forecast accuracy and the Diebold-Mariano test for significant differences between each two competing series.

The structure of the paper is as follows. Section II briefly describes the different neural networks architectures used in the analysis. Section III analyses the data set. In Section IV results of the forecasting competition are discussed. Finally, concluding remarks are given in Section V.

II. Methodology

Neural networks are flexible structures capable of learning sequentially from observed data. This feature makes ANNs specially suitable for time series forecasting. As opposed to traditional approaches to time series prediction, the specification of ANN models does not depend on a previous set on assumptions. Nevertheless, obtaining a reliable neural model involves selecting a large number of parameters experimentally: determining the number of input nodes, hidden layers, hidden nodes and output nodes, the activation function, the training algorithm, the training and the test samples, as well as the performance measures for cross-validation (Zhang *et al.*,1998). This range of different choices allows to chose the optimal topology of the ANN, while the weights of the model are estimated by gradient search. A complete summary on ANNs modelling issues can be found in Bishop (1995) and Haykin (1999).

Therefore, each network is suited to a combination of a learning paradigm (supervised and non-supervised learning), a learning rule related to the gradient cost function (Boltzmann, Hebbian, etc.) and a learning algorithm (forward-propagation, back-propagation, self-organization, etc.). The different learning paradigms represent alternative approaches to the treatment of information. In this study we focus on three ANN architectures (MLP, RBF and Elman), each of which deals with data in a different manner.

Multi-layer perceptron neural network

MLP networks consist of multiple layers of computational units interconnected in a feed-forward way. MLP networks are supervised neural networks that use as a building block a simple perceptron model. The topology consists of layers of parallel perceptrons, with connections between layers that include optimal connections. The number of neurons in the hidden layer determines the MLP network's capacity to approximate a given function. In order to solve the problem of overfitting, the number of neurons was estimated by cross-validation. In this work we used the MLP specification suggested by Bishop (1995) with a single hidden layer and an optimum number of neurons derived from a range between 5 and 25:

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

Where y_t is the output vector of the MLP at time t; g is the nonlinear function of the neurons in the hidden layer; x_{t-i} is the input value at time $t-i$ where i stands for the memory (the number of lags that are used to introduce the context of the actual observation.); q is the number of neurons in the hidden layer; φ_{ij} are the weights of neuron j connecting the input with the hidden layer; and β_j are the weights connecting the output of the neuron j at the hidden layer with the output neuron.

Note that the output y_t in our study is the estimate of the value of the time series at time $t+1$, while the input vector to the neural network will have a dimensionality of $p+1$.

We considered a $\text{MLP}(p;q)$ architecture that represented the possible nonlinear relationship between the input vector x_{t-i} and the output vector y_t. The parameters of the network (φ_{ij} and β_j) were estimated by means of the Levenberg-Marquardt algorithm, which is a quasi Newton algorithm. The training was done by iteratively estimating the value of the parameters by local improvements of the cost function. To avoid the possibility that the search for the optimum value of the parameters finishes in a local minimum, we used a multi-starting technique that initializes the neural network several times for different initial random values returning the best result.

Radial basis function neural network

RBF networks consist of a linear combination of radial basis functions such as kernels centred at a set of centroids with a given spread that controls the volume of the input space represented by a neuron (Bishop, 1995). RBF networks typically include three layers: an input layer; a hidden layer and an output layer. The hidden layer consists of a set of neurons, each of them computing a symmetric radial function. The output layer also consists of a set of neurons, one for each given output, linearly combining the outputs of the hidden layer. The output of the network is a scalar function of the output vector of the hidden layer. The equations that describe the input/output relationship of the RBF are:

$$
\begin{aligned}
& y_t = \beta_0 + \sum_{j=1}^{q} \beta_j g_j\left(x_{t-i}\right) \\
& g_j\left(x_{t-i}\right) = \exp\left(- \frac{\sum_{j=1}^{p} \left(x_{t-i} - \mu_j\right)^2}{2\sigma_j^2} \right) \\
& \left\{ x_{t-i} = \left(1, x_{t-1}, x_{t-2}, \ldots, x_{t-p}\right)', i = 1, \ldots, p \right\} \\
& \left\{ \beta_j, j = 1, \ldots, q \right\}
\end{aligned}
\qquad (2)
$$

Where y_t is the output vector of the RBF at time t; β_j are the weights connecting the output of the neuron j at the hidden layer with the output neuron; q is the number of neurons in the hidden layer; g_j is the activation function, which usually has a Gaussian shape; x_{t-i} is the input value at time $t-i$ where i stands for the memory (the number of lags that are used to introduce the context of the actual observation); μ_j is the centroid vector for neuron j; and the spread σ_j is a scalar that measures the width over the input space of the Gaussian function and it can be defined as the area of influence of neuron j in the space of the inputs. Note that the output y_t in our study is the estimate of the value of the time series at time $t+1$, while the input vector to the neural network will have a dimensionality of $p+1$.

In order to assure a correct performance, before the training phase the number of centroids and the spread of each centroid have to be selected. The spread σ_j is a hyper parameter selected before determining the topology of the network, and it was determined by cross-validation on the training database. The training was done by adding the centroids iteratively with the spread fixed. Then a regularized linear regression was estimated to compute the connections between the hidden and the output layer. Finally, the performance of the network was computed on the validation data set. This process was repeated until the performance on the validation database ceased to decrease.

Elman neural network

An Elman network is a special architecture of the class of recurrent neural networks, and it was first proposed by Elman (1990). The architecture is also based on a three-layer network but with the addition of a set of context units that allow feedback on the internal activation of the network. There are connections from the hidden layer to these context units fixed with a weight of one. At each time step, the input is propagated in a standard feed-forward fashion, and then a back-propagation type of learning rule is applied. The output of the network is a scalar function of the output vector of the hidden layer:

$$
\begin{aligned}
&y_t = \beta_0 + \sum_{j=1}^{q} \beta_j z_{j,t} \\
&z_{j,t} = g\left(\sum_{i=1}^{p} \varphi_{ij} x_{t-i} + \varphi_{0j} + \delta_{ij} z_{j,t-1} \right) \\
&\left\{ x_{t-i} = \left(1, x_{t-1}, x_{t-2}, \cdots, x_{t-p}\right)', i = 1, \ldots, p \right\} \\
&\left\{ \varphi_{ij}, i = 1, \ldots, p, j = 1, \ldots, q \right\} \\
&\left\{ \beta_j, j = 1, \ldots, q \right\} \\
&\left\{ \delta_{ij}, i = 1, \ldots, p, j = 1, \ldots, q \right\}
\end{aligned}
\tag{3}
$$

Where y_t is the output vector of the Elman network at time t; $z_{j,t}$ is the output of the hidden layer neuron j at the moment t; g is the nonlinear function of the neurons in the hidden layer; x_{t-i} is the input value at time $t-i$ where i stands for the memory (the number of lags that are used to introduce the context of the actual observation); φ_{ij} are the weights of neuron j connecting the input with the hidden layer; q is the number of neurons in the hidden layer; β_j are the weights of neuron j that link the hidden layer with the output; and δ_{ij} are the weights that correspond to the output layer and connect the activation at moment t. Note that the output y_t in our study is the estimate of the value of the time series at time $t+1$, while the input vector to the neural network will have a dimensionality of $p+1$.

The training of the network was done by back-propagation through time, which is a generalization of back-propagation for feed-forward networks. The parameters of the Elman neural network were estimated by minimizing an error cost function, which takes into account the whole time series. In order to minimize total error, gradient descent was used to change each weight in proportion to its derivative with respect to the error. A major problem with gradient descent for standard recurrent architectures is that error gradients vanish exponentially quickly with the size of the time lag. Recurrent neural networks cannot be easily trained for large numbers of neuron units and may behave chaotically.

III. Data

In this study we made use of tourism data. We used the number of tourist arrivals (first destination) provided by the Institute of Tourism Studies (IET) disaggregated by each visitor market over the period 2001:01 to 2012:07. The first four visitor markets (France, the United Kingdom, Belgium and the Netherlands and Germany) account for more than half of the total number of tourist arrivals to Catalonia, although Russia and the Northern countries are the ones experiencing the highest growth in tourist arrivals.

First, we tested the unit root hypothesis. In Table 1 we present the results of the augmented Dickey-Fuller (ADF), the Phillips–Perron (PP) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. While the ADF and the PP statistics test the null hypothesis of a unit root in x_t, the KPSS statistic tests the null hypothesis of stationarity. As it can be seen in Table 1, in most countries we cannot reject the null hypothesis of a unit root at the 5% level. Similar results are obtained for the KPSS test, where the null hypothesis of stationarity is rejected in most cases. When the tests were applied to the first difference of individual time series, the null of non-stationarity is strongly rejected in most cases. In the case of the KPSS test, we cannot reject the null hypothesis of stationarity at the 5% level in any country. These results imply that differencing is required in most cases and prove the importance of deseasonalizing and detrending tourism demand data (Zhang and Qi, 2005). In order to eliminate both linear trends as well as seasonality we used the first differences of the natural log of tourist arrivals.

Table 1. Unit root tests on the trend-cycle series of tourist arrivals

Country	Test for I(0)			Test for I(1)			Test for I(2)		
	ADF	PP	KPSS	ADF	PP	KPSS	ADF	PP	KPSS
France	-2.19	-3.41	0.32	-3.32	-2.48	0.15	-5.17	-3.53	0.04
United Kingdom	-1.71	-2.28	0.35	-2.72	-2.88	0.15	-18.98	-2.37	0.06
Belgium and the NL	-3.53	-2.55	0.21	-2.56	-3.42	0.10	-8.36	-4.46	0.02
Germany	-2.28	-3.61	0.23	-3.36	-3.70	0.15	-9.07	-4.35	0.05
Italy	-0.78	-0.99	0.33	-3.96	-2.46	0.08	-5.45	-3.24	0.24
US and Japan	-1.29	-2.40	0.33	-7.16	-4.06	0.03	-6.90	-2.32	0.02
Northern countries	-3.26	-2.16	0.17	-3.86	-3.61	0.07	-11.36	-2.50	0.03
Switzerland	-1.80	-2.99	0.16	-7.11	-4.10	0.07	-6.65	-4.41	0.06
Russia	0.25	0.82	0.30	-5.01	-3.70	0.09	-8.31	-4.07	0.02
Other countries	-2.04	-1.96	0.20	-4.56	-4.23	0.06	-9.84	-2.50	0.02
Total	-2.14	-1.76	0.30	-2.99	-2.91	0.14	-12.47	-2.34	0.05

Notes: Estimation period 2002:01-2012:07.
Tests for unit roots. Intercept included in test equation. Critical values for I(0) and I(1): ADF – Augmented Dickey and Fuller (1979) test, the 5% critical value is -2.88; KPSS – Kwiatkowski *et al* (1992) test, the 5% critical value is 0.46. Critical values for I(2): ADF – Augmented Dickey and Fuller (1979) test, the 5% critical value is -3.44; PP – Phillips and Perron (1988) test, the 5% critical value is -3.44; KPSS – Kwiatkowski *et al* (1992) test, the 5% critical value is 0.15.

Given the common patterns displayed by most countries, we tested for cointegration using Johansen's (1988, 1991) trace tests (Lee, 2011; Dritsakis, 2004). Trace tests test the null hypothesis of r cointegrating vectors against the alternative hypothesis of n cointegrating vectors. In Table 2 we present the results of five different unrestricted cointegration rank trace tests. It can be seen that we can only reject the null hypothesis of nine cointegrating vectors with two of the tests. The fact that the evolution of tourist arrivals is multicointegrated has led us to apply a multiple-output neural network approach to obtain forecasts of tourism demand.

Table 2. Unrestricted Cointegration Rank Tests

Hypothesized number of CE(s)	Type of test				
	Assume no deterministic trend in data		Allow for linear deterministic trend in data		Allow for quadratic deterministic trend in data
	No intercept in CE	Intercept in CE	Intercept in CE	Intercept in CE	Intercept and trend in CE
	No test VAR	No intercept in VAR	Test VAR	No trend in VAR	Linear trend in VAR
$H_0 : r = 0$	856.6229	969.8334	946.8238	1085.223	1012.763
$H_0 : r \le 1$	642.9016	741.4399	719.5322	857.7293	785.4048
$H_0 : r \le 2$	489.0577	586.3624	566.5598	676.3885	604.4294
$H_0 : r \le 3$	358.9547	452.6527	432.9569	541.7908	471.6636
$H_0 : r \le 4$	267.2172	344.7378	327.2923	412.1319	342.0272
$H_0 : r \le 5$	186.4016	256.9106	240.6405	314.3369	245.5905
$H_0 : r \le 6$	118.7815	176.0951	162.8499	227.6863	160.0873
$H_0 : r \le 7$	59.45009	110.2719	97.56685	149.9044	92.67206
$H_0 : r \le 8$	20.81093	56.72323	47.79385	85.37519	38.75788
$H_0 : r \le 9$	0.041106* (0.8681)	18.08417	10.98843	35.64879	0.944397* (0.3311)

Notes: Estimation period 2002:01-2012:07.
* Denotes rejection of the hypothesis at the 0.05 level.
** MacKinnon-Haug-Michelis (1999) *p*-values.
p-values in parentheses when different from zero.

IV. Results

In this section we implemented a multiple-output approach to predict arrivals to Catalonia from the different visitor countries. Since growth rates of tourist arrivals from all the different countries of origin share a common stochastic trend, we applied a multivariate forecasting framework. While a single-output approach requires to implement the experiment for each visitor market, the multiple-output approach allows to simultaneously obtain forecasts for all countries. We compared the forecasting performance of three different multiple-output ANN architectures: multi-layer perceptron, radial basis function and Elman recursive neural networks.

Following Bishop (1995) and Ripley (1996), we divided the collected data into three sets: training, validation and test sets. This division is done in order to asses the performance of the network on unseen data. The assessment is undertaken during the training process by means of the validation set, which is used in order to determine the epocs, the topology of the network and, in the case of the RBF the spread. The initial size of the training set was determined to cover a five-year span in order to accurately train the networks and to capture the different behaviour of the time series in relation to the economic cycle. After each forecast, a retraining was done by increasing the size of the set by one period and sliding the validation set by another period. This iterative process is repeated until the test set consisted of the last sample of the time series.

Based on these considerations, the first sixty monthly observations (from January 2001 to January 2006) were selected as the initial training set, the next thirty-six (from January 2007 to January 2009) as the validation set and the last 20% as the test set. Note that the sets consist of consecutive subsamples, and the resulting validation and test sets at the beginning of the experiment correspond to different phases of the economic cycle. All neural networks were implemented using Matlab™ and its Neural Networks toolbox.

To make the system robust to local minima, we applied the multistartings technique, which consists on repeating each training phase several times. In our case, the multistartings factor was three. The selection criterion for the topology and the parameters was the performance on the validation set. The Elman networks' parameters and topology had to be optimized taking into account that it could yield an unstable solution such as divergent training due to the fact that during the training the weights of the feedback loop could give rise to an unstable network.

Using as a criterion the performance on the validation set, the results that are presented correspond to the selection of the best topology, the best spread in the case of the RBF neural networks, and the best training strategy in the case of the Elman neural networks. Forecasts for 1,3 and 6 months ahead were computed in a recursive way. To summarise this information, two measures of forecast accuracy were computed: the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) (Tables 3 and 4). We repeated the experiment assuming different topologies regarding the memory values, which refer to the number of past months included in the context of the input, ranging from one to three months. Therefore, when the memory is zero, the forecast is done using only the current value of the time series, without any additional temporal context.

Table 3. RMSE (2010:04-2012:02)

	Memory (0) – no additional lags			Memory (3) – 3 additional lags		
France	MLP	RBF	Elman	MLP	RBF	Elman
1 month	0.21	0.09	0.59	0.28	0.09	0.41
3 months	0.23	0.09	0.47	0.23	0.09	0.46
6 months	0.16	0.08	0.29	0.33	0.09	0.47
United Kingdom						
1 month	0.16	0.16	0.46	0.35	0.16	0.50
3 months	0.31	0.16	0.46	0.35	0.16	0.41
6 months	0.22	0.15	0.40	0.46	0.15	0.54
Belgium and the NL						
1 month	0.21	0.12	0.39	0.28	0.12	0.34
3 months	0.13	0.11	0.34	0.24	0.12	0.34
6 months	0.23	0.12	0.28	0.38	0.12	0.48
Germany						
1 month	0.19	0.18	0.43	0.22	0.18	0.59
3 months	0.27	0.18	0.43	0.28	0.18	0.46
6 months	0.23	0.18	0.37	0.32	0.18	0.56
Italy						
1 month	0.32	0.08	0.63	0.43	0.09	0.44
3 months	0.37	0.09	0.43	0.33	0.09	0.49
6 months	0.29	0.09	0.43	0.52	0.09	0.57
US and Japan						
1 month	0.18	0.13	0.55	0.30	0.13	0.35
3 months	0.32	0.13	0.28	0.30	0.13	0.39
6 months	0.21	0.13	0.28	0.49	0.13	0.59
Northern countries						
1 month	0.29	0.19	0.55	0.37	0.17	0.41
3 months	0.33	0.17	0.34	0.41	0.18	0.34
6 months	0.16	0.18	0.23	0.31	0.18	0.39
Switzerland						
1 month	0.28	0.19	0.67	0.34	0.19	0.49
3 months	0.40	0.19	0.70	0.42	0.19	0.55
6 months	0.65	0.20	0.45	0.45	0.18	0.49
Russia						
1 month	0.38	0.34	1.00	0.72	0.36	1.01
3 months	0.86	0.38	1.04	0.76	0.37	0.86
6 months	0.90	0.39	0.92	0.90	0.36	1.07
Other countries						
1 month	0.18	0.08	0.36	0.26	0.08	0.36
3 months	0.28	0.08	0.27	0.25	0.08	0.31
6 months	0.23	0.08	0.21	0.21	0.08	0.35
Total						
1 month	0.11	0.05	0.33	0.14	0.05	0.30
3 months	0.20	0.05	0.17	0.15	0.04*	0.31
6 months	0.15	0.04*	0.23	0.19	0.05	0.28

Notes: * Best model.

Table 4. MAE (2010:04-2012:02)

	Memory (0) – no additional lags			Memory (3) – 3 additional lags		
France	MLP	RBF	Elman	MLP	RBF	Elman
1 month	0.15	0.08	0.48	0.23	0.08	0.34
3 months	0.15	0.07	0.35	0.18	0.07	0.36
6 months	0.12	0.07	0.24	0.25	0.07	0.41
United Kingdom						
1 month	0.12	0.14	0.37	0.27	0.13	0.41
3 months	0.20	0.13	0.32	0.26	0.14	0.33
6 months	0.19	0.14	0.33	0.38	0.13	0.43
Belgium and the NL						
1 month	0.17	0.11	0.30	0.21	0.10	0.24
3 months	0.11	0.10	0.26	0.19	0.10	0.29
6 months	0.18	0.10	0.23	0.28	0.11	0.37
Germany						
1 month	0.15	0.14	0.32	0.17	0.14	0.45
3 months	0.19	0.14	0.36	0.23	0.14	0.36
6 months	0.18	0.14	0.26	0.28	0.14	0.43
Italy						
1 month	0.23	0.06	0.48	0.31	0.07	0.36
3 months	0.25	0.07	0.30	0.26	0.06	0.40
6 months	0.22	0.07	0.31	0.42	0.06	0.50
US and Japan						
1 month	0.14	0.11	0.44	0.25	0.11	0.28
3 months	0.23	0.11	0.22	0.24	0.11	0.29
6 months	0.17	0.11	0.21	0.34	0.11	0.45
Northern countries						
1 month	0.20	0.16	0.44	0.29	0.14	0.32
3 months	0.26	0.15	0.29	0.33	0.15	0.27
6 months	0.13	0.15	0.19	0.25	0.15	0.32
Switzerland						
1 month	0.23	0.16	0.53	0.29	0.16	0.37
3 months	0.32	0.16	0.51	0.32	0.16	0.42
6 months	0.38	0.16	0.36	0.34	0.15	0.39
Russia						
1 month	0.31	0.30	0.80	0.57	0.33	0.86
3 months	0.62	0.35	0.81	0.66	0.34	0.69
6 months	0.64	0.35	0.70	0.81	0.32	0.82
Other countries						
1 month	0.14	0.07	0.26	0.19	0.07	0.28
3 months	0.16	0.06	0.20	0.19	0.06	0.24
6 months	0.17	0.06	0.18	0.17	0.06	0.26
Total						
1 month	0.09	0.04	0.24	0.11	0.04	0.22
3 months	0.12	0.04	0.12	0.11	0.03*	0.25
6 months	0.12	0.03*	0.18	0.14	0.04	0.22

Notes: * Best model.

Table 5. Diebold-Mariano loss-differential test statistic for predictive accuracy

	Memory (0) *versus* Memory (3)		
	MLP	RBF	Elman
France			
1 month	-2.68*	0.47	2.42*
3 months	-1.11	0.62	-0.22
6 months	-2.74*	-0.65	-2.92*
United Kingdom			
1 month	-3.98*	0.32	-0.70
3 months	-1.19	-0.37	-0.17
6 months	-3.75*	1.07	-1.35
Belgium and the Netherlands			
1 month	-0.85	0.53	0.75
3 months	-2.63*	-2.72*	-0.73
6 months	-1.47	-2.19*	-2.24*
Germany			
1 month	-0.95	0.12	-1.36
3 months	-0.92	-0.47	0.04
6 months	-1.75	-0.40	-2.24*
Italy			
1 month	-1.29	-0.55	1.44
3 months	-0.07	1.84	-1.38
6 months	-3.09*	1.31	-2.08*
US and Japan			
1 month	-1.93	-0.22	2.08
3 months	-0.09	-1.43	-0.97
6 months	-1.39	-0.64	-3.62*
Northern countries			
1 month	-1.30	1.98	1.27
3 months	-1.17	-1.95	0.38
6 months	-2.25*	-0.92	-2.54*
Switzerland			
1 month	-1.48	0.08	1.25
3 months	0.06	-0.52	0.95
6 months	0.36	3.01*	-0.38
Russia			
1 month	-2.66*	-0.66	-0.38
3 months	-0.29	1.64	0.82
6 months	-1.16	3.41*	-0.75
Other countries			
1 month	-1.75	-0.07	-0.51
3 months	-0.41	-0.97	-0.66
6 months	-0.10	-0.24	-1.24
Total			
1 month	-0.78	0.46	0.25
3 months	0.20	0.62	-3.55*
6 months	-0.75	0.53	-0.78

Notes: Diebold-Mariano test statistic with NW estimator. Null hypothesis: the difference between the two competing series is non-significant. A negative sign of the statistic implies that the second model has bigger forecasting errors.
* Significant at the 5% level.

We also use the Diebold-Mariano test for significant differences between each two competing series for each forecast horizons in order to assess the effect of different memory values on the forecasts (Table 5). When analysing the forecast accuracy, MLP and RBF networks show lower RMSE and MAE values than Elman networks. RBF networks display the lowest RMSE and MAE values in most countries both when the memory is zero and when is set to three. When the forecasts are obtained incorporating additional lags of the time series, the forecasting performance of RBF networks significantly improves in Switzerland and Russia for 6 months ahead. The lowest RMSE and MAE values are obtained with the RBF network for total tourist arrivals, for 3 months ahead when the memory is zero, and for 6 months ahead when using a memory of three lags.

When testing for significant differences between each two competing series (Table 5), we find that in most cases, as the number of previous months used for concatenation increases, the forecasting performance of the different networks shows no significant improvement. This result can in part be explained by the pre-processing (detrending) of the original time series and the cross-correlations accounted for in the multiple-output approach.

V. Conclusion

The main objective of the study is to evaluate the forecasting performance of three artificial neural networks models: the multi-layer perceptron neural network, the radial basis function neural network and the Elman recursive neural network. The seasonal patterns and the volatility that characterizes tourism data constitute an enabling field in which to compare the forecast accuracy of different neural network architectures that treat information in a different way. We use official statistical data of inbound international tourism demand to Catalonia. By means of the Johansen test we find that the evolution of arrivals from all countries of origin are multicointegrated. Since all markets share a stochastic trend, we apply a multivariate approach to obtain forecasts of tourism demand for all different countries and different forecast horizons.

When comparing the forecasting accuracy of the different techniques, we find that radial basis function neural networks outperform both multi-layer perceptron and Elman neural networks. This result shows that hybrid models, which combine supervised and non-supervised learning, are more indicated for economic forecasting with seasonal data than models using supervised learning alone.

Our results also suggest that when using dynamic or recurrent neural networks with forecasting purposes scaling issues may arise, which can give rise to divergence in the learning algorithm.

In order to evaluate the effect of the memory on the forecasting results, we repeated the experiment assuming different topologies regarding the number of lags used for concatenation. No significant differences are found when additional lags are incorporated in the feature vector, especially in the case of multi-layer perceptron neural networks. The explanation for this result is that the increase in the weight matrix is not compensated by the more complex specification and leads to overparametrization. The fact that increasing the dimensionality of the input does not have a significant effect on forecast accuracy is indicative that the pre-processing of the raw data conditions the forecasting results.

Summarizing, the forecasting out-of-sample comparison shows the suitability of applying hybrid models such as radial basis function neural networks to economic forecasting with seasonal time series. The study also reveals that the implementation of multiple-output architectures, taking into account the connections between the different time series, improves the forecasting performance of practical neural network forecasting. A question to be considered in further research is whether these results apply to different data pre-processing methods.

References

Balaguer, J. and Cantavella-Jordá, M. (2002) Tourism as a long-run economic growth factor: the Spanish case, *Applied Economics*, **34**, 877-884.

Bishop, C. M. (1995) *Neural networks for pattern recognition*, Oxford University Press, Oxford.

Binner, J. M., Bissoondeeal, R. K., Elger, T., Gazely, A. M. and Mullineux, A. W. (2005) A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro Divisia, *Applied Economics*, **37**, 655-680.

Broomhead, D. S. and Lowe, D. (1988) Multi-variable functional interpolation and adaptive networks, *Complex Syst.*, **2**, 321-355.

Burger, C., Dohnal, M., Kathrada, M. and Law, R. (2001) A practitioners guide to time-series methods for tourism demand forecasting: A case study for Durban, South Africa, *Tourism Management*, **22**, 403-409.

Cang, S. (2013) A comparative analysis of three types of tourism demand forecasting models: individual, linear combination and non-linear combination, *International Journal of Tourism Research*, **15.**

Cho, V. (2003) A comparison of three different approaches to tourist arrival forecasting, *Tourism Management*, **24**, 323-330.

Choudhary, M. A. and Haider, A. (2012) Neural network models for inflation forecasting: an appraisal, *Applied Economics*, **44**, 2631-2635.

Claveria, O. and Torra, S. (2014) Forecasting tourism demand to Catalonia: Neural networks vs. time series models, *Economic Modelling*, **36**, 220-228.

De Gooijer, J. G. and Kumar, K. (1992) Some recent developments in non-linear time series modelling, testing and forecasting, *International Journal of Forecasting*, **8**, 135-156.

Dickey, D. A. and Fuller, W. A. (1979) Distribution of the estimators for autoregressive time series with a unit root, *Journal of American Statistical Association*, **74**, 427-431.

Dritsakis, N. (2004) Cointegration analysis of German and British tourism demand for Greece, *Tourism Management*, **25**, 111-119.

Diebold, F. X. and Mariano, R. (1995) Comparing predictive accuracy, *Journal of Business and Economic Statistics*, **13**, 253-263.

Eilat, Y. and Einav, L. (2004) Determinants of international tourism: a three dimensional panel data analysis, *Applied Economics*, **36**, 1315-1327.

Elman, J. L. (1990) Finding structure in time, *Cognitive Science*, **14**, 179-211.

Granger, C. W. J. and Terasvirta, T. (1993) *Modelling nonlinear economic relationships*, Oxford University Press, Oxford.

Haykin, S. (1999) *Neural networks. A comprehensive foundation*, Prentice Hall, New Jersey.

Johansen, S. (1988) Statistical analysis of co-integration vectors, *Journal of Economic Dynamics and Control*, **12**, 231-254.

Johansen, S. (1991) Estimation and hypothesis testing of co-integration vectors in Gaussian vector autoregressive models, *Econometrica*, **59**, 1551-1580.

Kim, D. and Shwartz, Y. (2013) The accuracy of tourism forecasting and data characteristics: a meta-analytical approach, *Journal of Hospitality Marketing amd Management*, **22**, 349-374.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*, **54**, 159-178.

Kon, S. C. and Turner, W. L. (2005) Neural network forecasting of tourism demand, *Tourism Economics*, **11**, 301-328.

Law, R. (1998) Room occupancy rate forecasting: A neural network approach, *International Journal of Contemporary Hospitality Management*, **10**, 234-239.

Law, R. (2000) Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting, *Tourism Management*, **21**, 331-340.

Law, R. (2001) The impact of the Asian financial crisis on Japanese demand for travel to Hong Kong: a study of various forecasting techniques, *Journal of Travel & Tourism Marketing*, **10**, 47-66.

Law, R. and Au, N. (1999) A neural network model to forecast Japanese demand for travel to Hong Kong, *Tourism Management*, **20**, 89-97.

Lee, K. N. (2011) Forecasting long-haul tourism demand for Hong Kong using error correction models, *Applied Economics*, **43**, 527-549.

Lin, C. J., Chen, H. F. and Lee, T. S. (2011) Forecasting tourism demand using time series, artificial neural networks and multivariate adaptive regression splines: Evidence from Taiwan, *International Journal of Business Administration*, **2**, 14-24.

MacKinnon, J. G., Haug, A. and Michelis, L. (1999) Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration, *Journal of Applied Econometrics*, **14**, 563-577.

Madden, G. and Tan, J. (2008) Forecasting international bandwidth capacity using linear and ANN methods, *Applied Economics*, **40**, 1775-1787.

Palmer, A., Montaño, J. J. and Sesé, A. (2006) Designing an artificial neural network for forecasting tourism time-series, *Tourism Management*, **27**, 781-790.

Pattie, D. C. and Snyder, J. (1996) Using a neural network to forecast visitor behavior, *Annals of Tourism Research*, **23**, 151-164.

Phillips, P. C. B. and Perron, P. (1988) Testing for a unit root in time series regression, *Biometrika*, **75**, 335-346.

Ripley, B. D. (1996) *Pattern recognition and neural networks*, Cambridge University Press, Cambridge.

Santín, D., Delgado, F. J. and Valiño, A. (2004) The measurement of technical efficiency: a neural network approach, *Applied Economics*, **36**, 627-635.

Song, H. and Li, G. (2008) Tourism demand modelling and forecasting – a review of recent research, *Tourism Management*, **29**, 203-220.

Teixeira, J. P. and Fernandes, P. O. (2012) Tourism Time Series Forecast – Different ANN Architectures with Time Index Input, *Procedia Technology*, **5**, 445-454.

Tsaur, S. H., Chiu, Y. C. and Huang, C. H. (2002) Determinants of guest loyalty to international tourist hotels: a neural network approach, *Tourism Management*, **23**, 397-405.

Uysal, M. and El Roubi, M. S. (1999) Artificial neural Networks versus multiple regression in tourism demand analysis, *Journal of Travel Research*, **38**, 111-118.

Vlastakis, N., Dotsis, G. and Markellos, R. (2008) Nonlinear modelling of European football scores using support vector machines, *Applied Economics*, **40**, 111-118.

Zhang, G. and Qi, M. (2005) Neural network forecasting for seasonal and trend time series, *European Journal of Operational Research*, **160**, 501-514.

Zhang, G., Putuwo, B. E. and Hu, M. Y. (1998) Forecasting with artificial neural networks: the state of the art, *International Journal of Forecasting*, **14**, 35-62.

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona