TIME SERIES MODELING AND FORECASTING WITH FEEDFORWARD NEURAL NETWORKS: A COMPARATIVE STUDY USING THE RESEX DATA

Silvia Joekes, Fac. Ciências Econômicas, Univ. Nac. Córdoba.
e-mail : joekess@eco.unc.edu.ar
and
Emanuel Pimentel Barbosa, IMECC, Unicamp.
e-mail: emanuel@ime.unicamp.br

ABSTRACT

It is considered in this paper the modeling and forecasting, through feedforward neural networks, of a time series previously studied in the literature (Brubacher, 1974; Martin, 1980; Stahlbut, 1985; Allende, 1989) called RESEX series, which presents seasonality and outliers. Some elements of the network architecture such as the input variables definition are suggested by a previous time series analysis and other elements such as the number of intermediate layers and corresponding knots are defined through numerical methods. Not only traditional backpropagation based algorithms are used but also a robust learning algorithm is considered in order to deal more properly with the outliers. All models and methods (traditional and NN based) are then compared considering different predictive performance measures and the results are discussed.

Contents

1 Introduction

It is considered in this paper the modeling and forecasting, through feedforward neural networks, of a time series previously studied in the literature, called RESEX series, which presents seasonality and outliers. This series measures the monthly internal conections of residential telephone extensions in a certain area of Canada. It has been considered as a reference for evaluating robust estimation methods in time series analysis because of its very prominent additive outliers.

Some authors have analysed these data considering different robust methods for dealing with the outliers, among others, pre-filtering the extreme data before using more standard methods (Brubacher, 1974; Martin, 1980; Stahlbut, 1985; Allende, 1989). Such previous analysis of the RESEX data include the identification of a stationary second order auto-regressive AR(2) model for the series after its seasonal differencing (Brubacher,1974). The results of the RESEX data analysis using the seasonal AR(2) model and different outlier correction or robust methods, are presented in Allende (1989).

We propose in this paper, an alternative approach for modeling and forecasting the RESEX data, dealing with the outliers in a different way, through feedforward neural networks, which present some interesting caracteristics or advantages.

First, the considered NN arquitecture (with 3 lagged variables as inputs and one intermediate layer) is such that it can be interpreted as a non-linear extension of the previous seasonal AR(2) model. That is, the feedforward NN considered can be seen as a way to represent and implement a seasonal NAR (Non-linear Auto-Regressive) model to the series, generalizing the previously proposed models. Since the neural network can be seen as an 'universal function aproximator', according to the Kolmogorov-Nielsen theorem (Bishop, 1995), it express and approximate the unknow non-linear function associated to the NAR model.

Second, the neural networks, and particularly the ones considered in this paper, are more resistent to outliers in the data than traditional or standard time series methods, even if the network is estimated or trained through standard procedures such as usual backpropagation algorithms based on gradient methods.

This is shown in this paper and is related with the fact that the network is a two-stage model with many parameters and great flexibility, restricting the outliers influence to a local efect. In fact, the network parameters or weights are estimated here considering two different algorithms , not only the standard backpropagation (Haykin, 1999) but also an alternative learning algorithm, even more robust, based on iterative robust filtering (Connor, 1993).

The paper main objectives are to propose alternative methods based on neural networks for the analysis of time series with outliers such as the RESEX data and to compare the proposed models and corresponding estimation methods with some previous analysis based on traditional time series methods presented in the literature for the mentioned data.

The organization of the paper is as follows. In section 2 we present the RESEX series and its preliminary analysis based on previous literature and our re-analysis. The feedforward neural networks are introduced in section 3, with the two learning algorithms considered and the practical implementation with the time series data. The comparative results from both methods (traditional and NN based) are presented and discussed in section 4, followed by the bibliographic references in section 5.

2 The RESEX Series and Its Preliminary Analysis

2.1 Data description and initial preparation

The RESEX series measures the monthly internal conections of residential telephone extensions in a certain area of Canada, from January 1966 to May 1973, with a total of 89 observations. The more prominent caracteristics of this time series are its yearly seasonality and the presence of a couple of outliers (extremely large values) near the end of the series. In fact, as we can see from the time series plot presented in Fig. 1 below, the outliers correspond

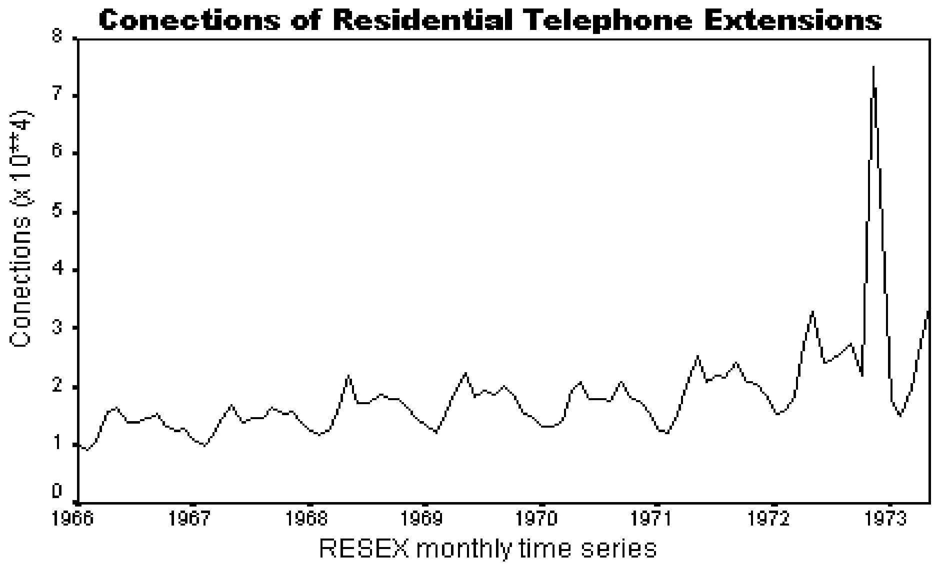

Figure 1: RESEX series

to the months of November and December 1972. The reason for that was a special promotion (low prices) offered by the telephone company at that months.

The seasonality, with period of one year or 12 months is very clear from the figure above, and could be easily verified using other exploratory tools (such as a periodogram) although it is not necessary.

Another caracteristic possibly suggested by the graphic, although not so clear as the other 2 elements just discussed, would be an eventual small trend component. However, we have tested the hypothesis of stationarity versus the presence of unit root or trend, using the Dickey & Fuller (1979, 1981) test after a seasonal differencing, and it was not detected evidence (at a 5% level of significance) of the presence of a trend component. Therefore, after the seasonal differencing, the series can be considered as stationary (for details, see Joekes, S.,2002), what is consistent with some previous work, as for instance, Brubacher, 1974, and others.

In order to identify a stationary model for the data, a preliminary filtering and interpolation of the extreme data points is recommendable in order to avoid a possible model mispecification.

In this way, the series extreme values were corrected using the Kalman filter (Joekes, S.,2002), resulting in an outliers free version of RESEX, presented below in Fig. 2.

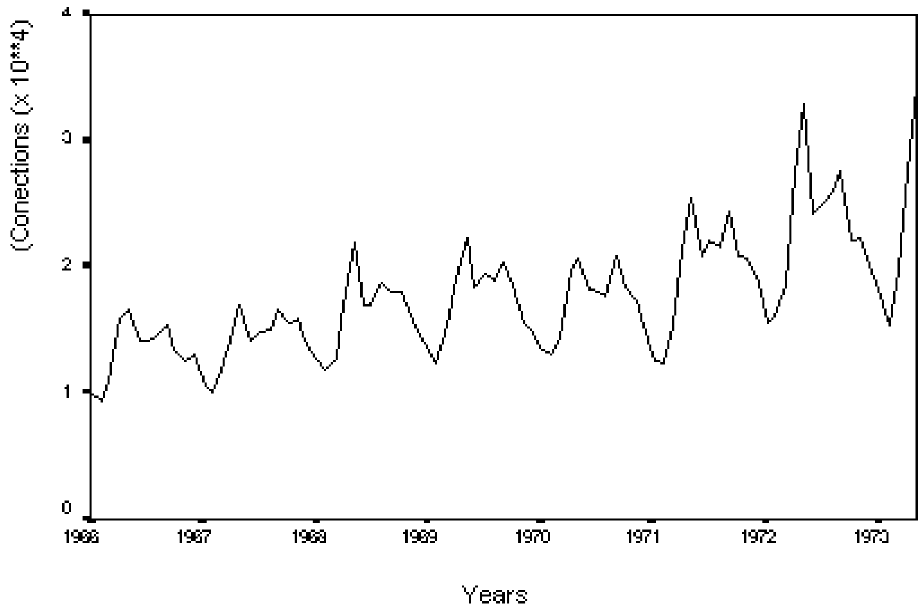

Figure 2: RESEX series using Kalman filter

As we can see from the picture above, the seasonal pattern shown is not stable or stationary, suggesting the application of a seasonal differencing. After this data initial preparation (outlier correction and seasonal differencing), we can now look for a stationary time series model for the corrected RESEX.

2.2 Basic Model Identification

A standard auto-correlation analysis of the (free of outliers) seasonaly differenced RESEX series through simple and partial correlogram graphics suggest an AR(2) model (auto-regressive of order 2) for the data. Or, using the Box-Jenkins notation for seasonal ARIMA models, we have identified an ARIMA(2,0,0)(0,1,0)12 model for the series, which is in accordance with the literature (Brubacher,1974 and others).

The parameter estimates (ML) and corresponding standard errors, are in accordance with previous results from the mentioned literature, and also, the obtained residuals can be considered satisfactory (residual auto-correlations non significative, etc.), confirming the identified model.

The standard seasonal AR(2) model is considered here as a reference, in order to assess the predictive performance of the proposed NN methods. In this way, the reference model is fitted to the first 7 years of data (84 observations), leaving the last 5 months for predictive purposes . The predictive results (one step ahead and multi step ahead) will be measured through the RSE (Residual Standard Error) and MAPE (Mean Absolute Percentage Error) and presented at section 4, jointly with the NN results.

3 Feedforward Neural Networks for Time Series Data

3.1 Basic elements

In its simplest form, a feedforward NN relates a response or output variable to n predictors or input variables through a non-linear relationship represented as a two level composition of generalized linear relations-GLR's. (A GLR is a non-linear transformation applied to a linear combination of its inputs). That is, the output is a GLR of intermmediate variables, where each one is a GLR of the input variables (Haykin, S., 1999).

In a time series context, it is common to relate a given series to its past (lags) through a linear auto-regressive process, which can be extended to a non-linear form (NAR model), with an implementation via FNN, as represented in Fig. 3 below,

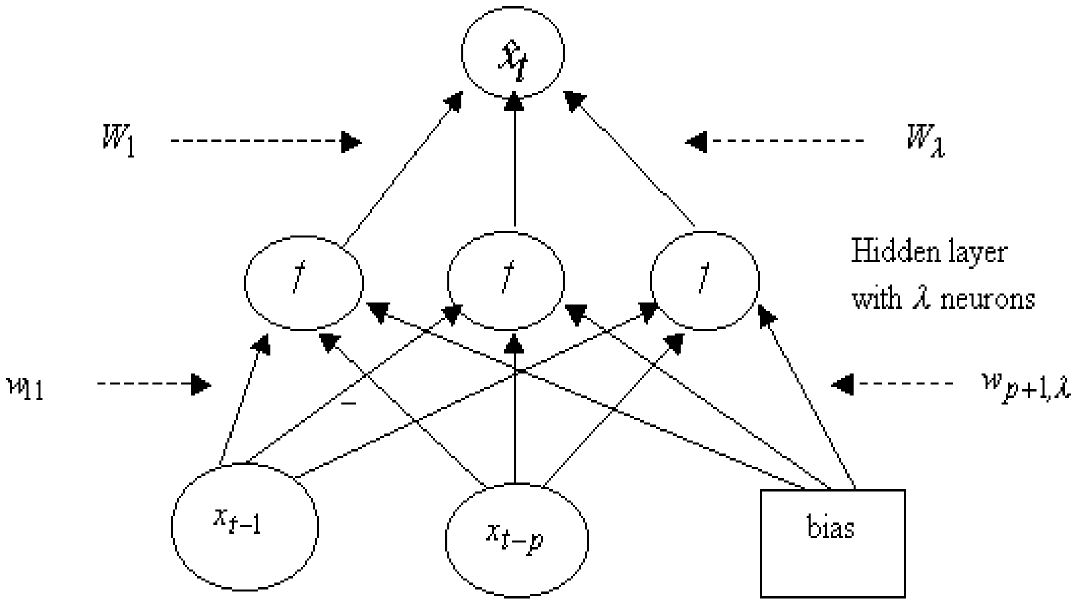

Figure 3: Feedforward network for NAR(p) modeling

The network arquitecture specification involves the definition of the input variables (the number p of lags: the NAR order), the number of GLR's or nodes in the intermmediate layer, and the non-linear transformations in the intermmediate layer (called link or activation function, usually taken as a sigmoidal function), since the activation for the output is usually identity for time series data. Once fully defined the network arquitecture, the objective is to estimate the weights based on data.

3.2 Learning Algorithms

The more common iterative learning methods for estimating the NN parameters or weights, are known as backpropagation algorithms (Bishop, C., 1995; Haykin, S., 1999) and are based on gradient methods, mainly stepest descent or conjugate gradient.

Among these, we consider the backpropagation algorithm based on conjugate gradient as our basic method for NN implementation. This is because it has known advantages in relation to its simpler version (Haykin, S., 1999), at the only extra computational cost of calculating Hessian matrices.

One important practical aspect of implementation of these algorithms is the adequate learning stopping in order to avoid overtraining, and the called 'early stopping' criteria is considered, that is, to stop when the validation error, monitored on line, reaches its minimum.

3.3 Network Arquitecture and Implementation

One of the first elements to be specified in the network are the input variables, and using the information that a reference for the data is a seasonal AR(2) model, these variables are defined as lagged versions (lags 1, 2 and 12) **of the output. ???? of the series.** A second element is the specification of the number of intermmediate layers (one or two) and corresponding nodes (between 1 and 4, since the expected optimal value would be 2). A total of 8 or 9 possible cases or combinations of these values are considered in order to choose the one with better performance for fitting and predicting. All the cases were implemented, using the software SPSS (Statistical Package for the Social Science) and its module for NN. Also, the sample data was divided in 3 parts: the trainning sample (first 72 observations), the validation sample (next 12 observations) and the test or prediction sample (last 5 observations or months). The fitting and prediction (test set) results for each plausible arquitecture are presented in terms of error measures such as the RSE and MAPE, as shown in the Tab.1 below,

Table 1: Comparison of various feedforward neural models according to number of hidden layers and number of hidden neurons by layer. Measures of fit and predicting results (no correction for outliers)
(The RSE values are expressed in 10^3 units and MAPE in percents)

Number of hidden layer	Number of hidden neurons	Number of parameters	Measures of fit				Testing set	
			Training set		Validation set			
			RSE	MAPE	RSE	MAPE	RSE	MAPE
1	1	6	1.66	7.28	1.55	7.58	1.44	4.75
1	2	11	1.60	7.16	1.56	7.56	1.20	4.14
1	3	16	1.67	7.31	1.54	7.49	1.73	5.96
1	4	21	1.49	6.46	1.53	6.57	9.94	36.67
2	1-1	8	1.82	8.22	1.59	8.17	3.12	10.70
2	2-2	17	1.86	8.14	1.31	6.00	4.30	15.13
2	2-3	21	1.71	7.82	1.57	7.92	2.22	8.76
2	3-2	23	1.49	6.65	1.58	7.23	7.34	20.86
2	3-3	28	1.69	7.51	1.55	7.58	2.56	7.82

From the table above it can be verified that the network with just one intermmediate (hidden) layer and 2 nodes or neurones in this layer is the one with better predictive measures and one of the best fitting measures; therefore, this is the chosen arquitecture for the RESEX data. The results are very consistent in many ways, as for instance, the number of nodes in the hidden layer is expected to be between 3 (inputs) and 1 (output). This arquitecture is shown in Fig.4 below,

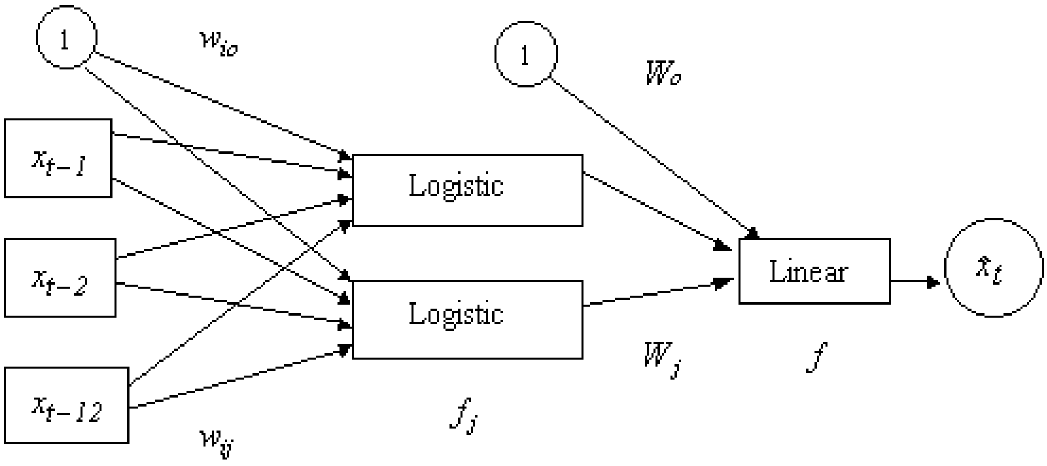

Figure 4: Architecture of the neural network for the RESEX series

4 Comparative Results and Discussion

We present here a comparative table with the predictive results obtained from the application of the 3 considered methods for the modeling of the RESEX data: the (linear) seasonal AR(2) model with outlier correction, the FNN with standard learning (without outlier correction) and the FNN with robust learning. The first method is not considered without outlier correction because its fitting and predictive performance are too poor to be compared.

The models were fitted considering the first 7 years or 84 observations, and they were used to predict the last 5 months (multi step ahead). For one-step-ahead prediction, at each point after the first prediction, a new data point is added to the learning sample. The predictive performance measures considered were the RSE and the MAPE, as shown in Tab. 2 below,

Table 2: Comparative Results
(The RSE values are expressed in 10^4 units and MAPE in percents)

Models	Measures Of fit		One step ahead prediction		Multi step ahead prediction	
	RSE	MAPE	RSE	MAPE	RSE	MAPE
AR(2) - with outlier correction	1.34	5.78	1.48	5.40	1.12	4.20
FNN - without outlier correction	1.60	7.16	1.20	4.14	1.17	4.00
FNN - with robust learning	1.37	6.48	1.20	4.33	1.06	3.33

5 Bibliographic References

Allende,H.(1989). Robust Recursive Estimation of Autoregressive Models. Rev. Soc. Chilena Estadist., 6, 3-19

Bishop,C.(1995). Neural Networks for Pattern Recognition. Claredon Press, Oxford, U.K.

Brubacher,S.R.(1974). Time series outlier detection and modeling with interpolation. Bell-Labs. Tech.memo,USA

Connor,J.(1993). Time Series and Neural Network Modeling. Ph.D. Thesis, Univ. Washington, USA.

Dickey,D.W. and Fuller,W.(1979). Distribution of the estimates of AR time series with unit root. JASA, vol.74, 427-431.

Dickey,D.W. and Fuller,W.(1981). Likelihood Ratio Statistics for AR Time Series with a Unit Root. Econometrica, vol 49, pp 1057-1072.

Haykin,S.(1999). Neural Networks: a comprehensive foundation. Macmillan College Publ. Comp., NY, USA.

Joekes,S.(2002). Aplicación de Redes Neuronales Robustas en Series de Tiempo: un estudio comparativo. Tesis de Maestria en Estadística Aplicada. Universidad Nacional de Córdoba.

Martin,R.D.(1980). Robust Estimation of Autoregressive Models (with discussion) In Directions in Time Series. Dr. Brilliger and G.C. Tiao eds. Inst. of Math. Statist., Hayward, California, 228-262.