Available online at www.sciencedirect.com
SCIENCE DIRECT®
Int. J. Production Economics 86 (2003) 217–231
international journal of production economics
www.elsevier.com/locate/dsw

A comparative study of linear and nonlinear models for aggregate retail sales forecasting

Ching-Wu Chu[a], Guoqiang Peter Zhang[b,*]

[a] Department of Shipping and Transportation Management, National Taiwan Ocean University, Keelung, Taiwan
[b] Department of Management, J. Mack Robinson College of Business, Georgia State University, 35 Broad Street, NW, Atlanta, GA 30303, USA

Received 30 January 2002; accepted 30 January 2003

Abstract

The purpose of this paper is to compare the accuracy of various linear and nonlinear models for forecasting aggregate retail sales. Because of the strong seasonal fluctuations observed in the retail sales, several traditional seasonal forecasting methods such as the time series approach and the regression approach with seasonal dummy variables and trigonometric functions are employed. The nonlinear versions of these methods are implemented via neural networks that are generalized nonlinear functional approximators. Issues of seasonal time series modeling such as deseasonalization are also investigated. Using multiple cross-validation samples, we find that the nonlinear models are able to outperform their linear counterparts in out-of-sample forecasting, and prior seasonal adjustment of the data can significantly improve forecasting performance of the neural network model. The overall best model is the neural network built on deseasonalized time series data. While seasonal dummy variables can be useful in developing effective regression models for predicting retail sales, the performance of dummy regression models may not be robust. Furthermore, trigonometric models are not useful in aggregate retail sales forecasting.

Keywords: Aggregate retail sales; Forecasting; Seasonality; ARIMA; Neural networks

1. Introduction

Forecasting of the future demand is central to the planning and operation of retail business at both macro and micro levels. At the organizational level, forecasts of sales are needed as the essential inputs to many decision activities in various functional areas such as marketing, sales, production/purchasing, as well as finance and accounting (Mentzer and Bienstock, 1998). Sales forecasts also provide basis for regional and national distribution and replenishment plans. The importance of accurate sales forecasts to efficient inventory management at both disaggregated and aggregate levels has long been recognized. Barksdale and Hilliard (1975) examined the relationship between retail stocks and sales at the aggregate level and found that successful inventory management depends to a large extent on the accurate forecasting of retail sales. Thall (1992)

*Corresponding author. Tel.: +1-404-651-4065; fax: +1-404-651-3498.
E-mail address: gpzhang@gsu.edu (G.P. Zhang).

0925-5273/03/$ - see front matter
doi:10.1016/S0925-5273(03)00068-9

and Agrawal and Schorling (1996) also pointed out that accurate demand forecasting plays a critical role in profitable retail operations and poor forecasts would result in too-much or too-little stocks that directly affect revenue and competitive position of the retail business.

Retail sales often exhibit strong seasonal variations. Historically, modeling and forecasting seasonal data is one of the major research efforts and many theoretical and heuristic methods have been developed in the last several decades. The available traditional quantitative approaches include heuristic methods such as time series decomposition and exponential smoothing as well as time series regression and autoregressive and integrated moving average (ARIMA) models that have formal statistical foundations. Among them, the seasonal ARIMA model is the most advanced forecasting model that has been successfully tested in many practical applications. In addition, it has been shown that the popular Winter's additive and multiplicative exponential smoothing models can be implemented by the equivalent ARIMA models (McKenzie, 1984; Bowerman and O'Connell, 1993).

One of the major limitations of the traditional methods is that they are essentially linear methods. In order to use them, users must specify the model form without the necessary genuine knowledge about the complex relationship in the data. Of course, if the linear models can approximate the underlying data generating process well, they should be considered as the preferred models over more complicated models as linear models have the important practical advantage of easy interpretation and implementation. However, if the linear models fail to perform well in both in-sample fitting and out-of-sample forecasting, more complex nonlinear models should be considered.

One nonlinear model that recently receives extensive attention in forecasting is the artificial neural network model (NN). Inspired by the architecture of the human brain as well as the way it processes information, NNs are able to learn from the data and experience, identify the pattern or trend, and make generalization to the future. The popularity of the neural network model can be attributed to their unique capability to simulate a wide variety of underlying nonlinear behaviors. Indeed, research has provided theoretical underpinning of neural network's universal approximation ability. That is, with appropriate architectures, NNs can approximate any type of function with any desired accuracy (Hornik et al., 1989). In addition, few assumptions about the model form are needed in applying the NN technique. Rather, the model is adaptively formed with the real data. This flexible data-driven modeling property has made NNs an attractive tool for many forecasting tasks as data are often abundant while the underlying data generating process is hardly known or changing in the real world environment.

Although numerous comparative studies between traditional models and neural networks have been conducted in the literature, findings are mixed with regard to whether the flexible nonlinear approach is better than the linear method in forecasting (Adya and Collopy, 1998). In addition, contradictory conclusions have been reported on when or under what conditions one method is better than the other (see Zhang et al., 1998). Several researchers have provided empirical evidence on the comparative advantage of one model over the other in various forecasting situations. For example, Elkateb et al. (1998) reported a comparative study between ARIMA models and neural networks in electric load forecasting. Their results showed NNs were better in forecasting performance than the linear ARIMA models. Prybutok et al. (2000) compared NNs with ARIMA and linear regression for maximum ozone concentrations and found that NNs were superior to the linear models. Although most of the published research indicates the superiority of the NN model in comparison to simpler linear models, several studies report different results. Church and Curram (1996) and Ntungo and Boyd (1998) showed that neural networks performed about the same as, but no better than, the econometric and ARIMA models. Callen et al. (1996) reported the negative findings about neural networks in forecasting quarterly accounting earnings. They showed that NNs were not as effective as the linear time series models in forecasting performance even if the data were nonlinear. Kirby et al. (1997) and Darbellay and Slama (2000) also reported similar findings in forecasting motorway traffic and short-term electricity demand.

While most of the above studies do not involve seasonal data, little research has been done focusing directly on seasonal time series modeling and forecasting. How to effectively model seasonal time series is a challenging task not only for the newly developed nonlinear models, but also for the traditional models. One popular traditional approach to dealing with seasonal data is to remove the seasonal component first before other components are estimated. Many practitioners in various forecasting applications have satisfactorily adopted this practice of seasonal adjustment. However, several recent studies have raised doubt about its appropriateness in handling seasonality. Seasonal adjustment has been found to lead into undesirable nonlinear properties, severely distorted data, and inferior forecast performance (Plosser, 1979; Ghysels et al., 1996; De Gooijer and Franses, 1997). De Gooijer and Franses (1997) pointed out that "although seasonally adjusted data may sometimes be useful, it is typically recommended to use seasonally unadjusted data." On the other hand, mixed findings have also been reported in the limited neural network literature on seasonal forecasting. For example, Sharda and Patil (1992) found that, after examining 88 seasonal time series from the M-competition (Makridakis et al., 1982), NNs were able to model seasonality directly and pre-deseasonalization is not necessary. Alon et al. (2001) also found that NNs are able to "capture the dynamic nonlinear trend and seasonal patterns, as well as the interactions between them." Based on a sample of 68 time series from the same database, Nelson et al. (1999), however, concluded just the opposite. Zhang and Qi (2002) confirmed conclusions in Nelson et al. (1999) with consumer retail sales.

The purpose of this paper is to compare the out-of-sample forecasting performance of aggregate retail sales between several widely used linear seasonal forecasting models and the nonlinear neural network models. Our focus will be on univariate time series forecasting. Motivated by the lack of general guidelines and clear evidence on whether the powerful nonlinear modeling capability of neural networks can improve forecasting performance for seasonal data, we would like to provide detailed analysis and empirical evidence on the effectiveness of different modeling strategies for seasonal time series forecasting. Although Alon et al. (2001) have studied aggregate retail sales forecasting issues with neural networks, their study does not specifically consider a number of modeling issues for seasonal time series. On the other hand, to our knowledge, no research has been conducted to investigate whether using auxiliary variables such as seasonal dummy and trigonometric variables is useful in improving retail sales forecasting.

The rest of the paper is organized as follows. The next section presents a review of the traditional methods as well as the nonlinear neural networks for direct seasonal time series modeling. The research methodology of the study is described in Section 3, which is followed by the discussion of empirical findings. The last section provides a summary of the results and concluding remarks.

2. Modeling seasonal variations

This section examines several linear and nonlinear models that have been commonly used in modeling and forecasting seasonal time series. Although the most popular traditional approach to handling seasonality is to remove the seasonal variations from the data, it is important to note that seasonal adjustment is an approximation method and forecasts based on seasonally adjusted data may have wider error margins than these based on the original data because of the more uncertainties involved in the seasonal adjustment process. Of course, if the seasonality is removed, seasonal models are not needed for the deseasonalized data. For this reason, our focus here will be on several direct seasonal modeling methods. Specifically, we consider three classes of general modeling approach to seasonal data: seasonal ARIMA, regression, and feedforward neural networks.

2.1. Box–Jenkins ARIMA modeling approach

ARIMA is the most versatile linear model for forecasting seasonal time series. It has enjoyed great success in both academic research and industrial applications during the last three decades. The class of ARIMA models is broad. It can represent many different types of stochastic seasonal and nonseasonal time series such as pure autoregressive (AR), pure moving average (MA), and mixed AR and MA processes. The theory of ARIMA models has been developed by many researchers and its wide application was due to the work by Box and Jenkins (1976) who developed a systematic and practical model building method. Through an iterative three-step model building process: model identification, parameter estimation and model diagnosis, the Box–Jenkins methodology has been proved to be an effective practical time series modeling approach.

The general seasonal ARIMA model has the following form:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \tag{1}$$

where

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p,$$

$$\Phi_P(B) = 1 - \Phi_s B^s - \Phi_{2s} B^{2s} - \cdots - \Phi_{Ps} B^{Ps},$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q,$$

$$\Theta_Q(B) = 1 - \Theta_s B^s - \Theta_{2s} B^{2s} - \cdots - \Theta_{Qs} B^{Qs}.$$

s is the season length, B is the backward shift operator, and ε_t is a sequence of white noises with zero mean and constant variance. The model order (p, d, q; P, D, Q) are determined during the model identification stage by use of various sample autocorrelation functions.

Although expression (1) is the most commonly used multiplicative form of the seasonal ARIMA model, other nonmultiplicative forms are also possible (Pankratz, 1983). Whatever the form used, all of the seasonal ARIMA models can express the future value as a linear combination of the past seasonal and nonseasonal lagged observations.

2.2. Regression approach to seasonal modeling

Multiple regression can be used to model seasonal variations. It evolves from the traditional decomposition method. The general additive decomposition model has the following expression:

$$Y_t = T_t + S_t + \varepsilon_t, \tag{2}$$

where T_t is the trend component and S_t is the seasonal component at time t. ε_t is the error term often assumed to be uncorrelated. This additive model is appropriate if the seasonal variation is relatively constant. If the seasonal variation increases over time, then the multiplicative model will be more appropriate. In this case, however, the logarithmic transformation can be used to equalize the seasonal variation and then the additive model (2) will be again suitable.

Traditionally the trend component can be modeled by polynomials of time t of some low orders. Here we consider the following linear trend model:

$$T_t = \beta_0 + \beta_1 t. \tag{3}$$

On the other hand, the seasonal component can be modeled by either seasonal dummy variables or trigonometric functions. With the seasonal dummy variable I_{ti} defined as $I_{ti} = 1$ if time period t corresponds to season i and $I_{ti} = 0$ otherwise, we have

$$S_t = \omega_1 I_{t1} + \omega_2 I_{t2} + \cdots + \omega_s I_{ts}. \tag{4}$$

When combining (3) and (4) into model (2), it is necessary to either omit the intercept β_0 or set one of the seasonal parameter ω to zero in order for the parameters to be estimated.

The seasonal component S_t can also be modeled as a linear combination of trigonometric functions:

$$S_t = \sum_{i=1}^{m} A_i \sin\left(\frac{2\pi i}{s} t + \phi_i\right), \tag{5}$$

where A_i and ϕ_i are the amplitude and the phase of the sine function. m is the number of sine functions used to represent the seasonal variation. In many cases, $m = 1$ or 2 is sufficient to represent complex seasonal patterns (Bowerman and O'Connell, 1993; Abraham and Ledolter, 1983).

An equivalent form to (5), which is often used in practice, is

$$S_t = \sum_{i=1}^{m}\left(\omega_{1i}\sin\left(\frac{2\pi i}{s}t\right) + \omega_{2i}\cos\left(\frac{2\pi i}{s}t\right)\right). \qquad (6)$$

It is important to note that unlike ARIMA models, regression models are deterministic in that model components or coefficients are constants over time. Thus the behavior of the regression method can be quite different from that of the stochastic models. If the model components are changing as in many economic and business time series, the deterministic models may not be appropriate.

2.3. Nonlinear modeling approach

A number of nonlinear time series models have been developed in the literature but few are specifically for seasonal modeling. Moreover, most of these models are parametric and the effectiveness of the modeling effort depends to a large extent on whether assumptions of the model are satisfied. To use them, users must have knowledge on both data property and model capability and the model form must be pre-specified. This is the major obstacle for general application of these models.

Neural networks are the most versatile nonlinear models that can represent both nonseasonal and seasonal time series. The most important capability of neural networks compared to other nonlinear models is their flexibility in modeling any type of nonlinear pattern without the prior assumption of the underlying data generating process. The most popular feedforward three-layer network for forecasting problems has the following specification:

$$y_t = \alpha_0 + \sum_{j=1}^{q}\alpha_j f\left(\sum_{i=1}^{p}\beta_{ij}x_{it} + \beta_{0j}\right) + \varepsilon_t, \qquad (7)$$

where p is the number of input nodes, q is the number of hidden nodes, f is a sigmoid transfer function such as the logistic: $f(x) = (1/1 + \exp(-x))$. $\{\alpha_j, j = 0, 1, \ldots, n\}$ is a vector of weights from the hidden to output nodes and $\{\beta_{ij}, i = 0, 1, \ldots, p,\ j = 1, 2, \ldots, q\}$ are weights from the input to hidden nodes. α_0 and β_{0j} are weights of arcs leading from the bias terms which have values always equal to 1. The input variables, $x_i, i = 1, 2, \ldots, p$, are the lagged past observations if the time series data are used in model building. In this case, model (7) acts as a nonlinear AR model. To model seasonality, seasonal lagged observations (observations separated by multiples of seasonal period s) should be used. However, selecting an appropriate NN architecture or more importantly lagged variables may require some experimental efforts and traditional modeling skills (Faraway and Chatfield, 1998). On the other hand, we can use seasonal dummy variables or trigonometric terms as predictor variables, in which case, the neural network (7) is equivalent to a nonlinear regression model. Williams (1997) found the encouraging results with trigonometric variables to model seasonal variations in an application of rainfall prediction.

In the neural network literature, there are different opinions with different empirical findings on how to best model seasonal variations. The fundamental difference is on whether neural networks are able to directly model seasonal patterns and whether seasonal adjustment is necessary. Gorr (1994) and many others believed that neural networks should be able to capture the seasonality in the data because of their universal approximation ability. Sharda and Patil (1992) found empirically that neural networks can model seasonality effectively and pre-deseasonalizing the data is not necessary. Franses and Draisma (1997) found that neural networks could also detect possible changing seasonal patterns. Encouraging results with direct seasonal modeling and forecasting are also reported by Tang and Fishwick (1993), Nam and Schaefer (1995), and Williams (1997). On the contrary, Nelson et al. (1999) found that data deseasonalization is critical to significantly improve the forecasting performance of neural networks.

3. Methodology

The data used in this study are monthly retail sales compiled by the US Bureau of the Census.

The total sampling period examined is from January 1985 to December 1999. Although longer data series are available, a pilot study shows that larger samples are not necessarily helpful in overall forecasting performance. In a comparative study for business sales forecasting, Luxhoj et al. (1996) used a sampling period of only 5 years. Fig. 1 plots the data, which clearly shows the increasing seasonal fluctuations over the sampling period.

Three major research questions are addressed in this study:

- Are auxiliary variables such as seasonal dummy variables or trigonometric variables helpful in forecasting seasonal time series with both linear and nonlinear models?
- Does the increased modeling power of the nonlinear neural network model improve the out-of-sample forecasting performance for retail sales?
- What is the best way to model seasonal time series with neural networks? Is seasonal adjustment useful to improve forecasting accuracy?

The first question is motivated by a recent study by Williams (1997) who found increasing accuracy by employing trigonometric variables in his neural network models in forecasting daily rainfalls. Since seasonal dummy variables are also commonly used in linear seasonal models, we are interested in knowing if using seasonal dummy variables can improve modeling and forecasting performance for retail sales. Furthermore, if auxiliary variables were able to improve forecasting accuracy, then we would like to know further if the improvement is bigger with linear method or nonlinear method.

These issues are investigated through a comparative study of out-of-sample forecasting between linear and nonlinear models discussed in the previous section. Three linear models—ARIMA with time series, regression with dummy variables, and regression with trigonometric variables—are built using the in-sample data. The forecasting performance of each model is then evaluated by results from the out-of-sample which is excluded in the in-sample fitting and model selection process. Since the retail sales series exhibits both trend and seasonality, the following regression model is established:

$$Y_t = \beta_0 + \beta_1 t + \omega_1 I_{t1} + \omega_2 I_{t2} + \cdots + \omega_{11} I_{t11} + \varepsilon_t \quad (8)$$

where the seasonal dummy variable I_{ti} defined as $I_{ti} = 1$ if time period t corresponds to month i and $I_{ti} = 0$ otherwise; $\beta_0, \beta_1, \omega_1, \omega_2, \ldots, \omega_{11}$ are model

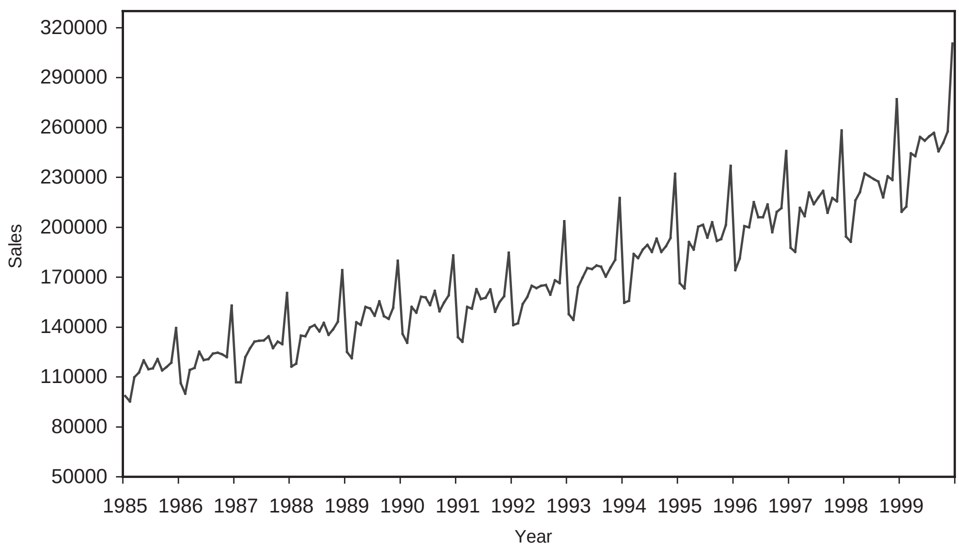

Fig. 1. Retail sales (1985–1999).

parameters. Note that the dummy variable for December is not explicitly defined.

We also consider the following regression models with two-term and four-term trigonometric functions, respectively:

$$Y_t = \beta_0 + \beta_1 t + \omega_1 \sin\left(\frac{2\pi t}{12}\right) + \omega_2 \cos\left(\frac{2\pi t}{12}\right) + \varepsilon_t, \quad (9)$$

$$Y_t = \beta_0 + \beta_1 t + \omega_1 \sin\left(\frac{2\pi t}{12}\right) + \omega_2 \cos\left(\frac{2\pi t}{12}\right) + \omega_3 \sin\left(\frac{4\pi t}{12}\right) + \omega_4 \cos\left(\frac{4\pi t}{12}\right) + \varepsilon_t. \quad (10)$$

Using models (8)–(10) requires that the time series have constant seasonal variations. Since the retail sales time series presents a clear increasing seasonality, the natural log-transformation is performed to stabilize the seasonal variations. The models are then fitted to the transformed data. Finally, the forecasts are scaled back to their original units.

On the nonlinear model side, we use the standard fully connected three-layer feedforward networks. The logistic function is used for all hidden nodes as the activation function. The linear activation function is employed for the output node. Bias terms are employed for both output and hidden nodes.

For a time series forecasting problem, NN model building is equivalent to determining both the number of input nodes and the number of hidden nodes. The input nodes are the past lagged observations through which the underlying autocorrelation structure of the data can be captured. However, there is no theoretical guideline that can help us pre-specify how many input nodes to use and what they are. Identifying the proper autocorrelation structure of a time series is not only a difficult task for nonlinear modeling, but also a challenge in the relatively simple world of linear models. On the other hand, it is not easy to preselect an appropriate number of hidden nodes for a given application. Although the NN universal approximation theory indicates that a good approximation may require a large number of hidden nodes, only a small number of hidden nodes are needed in many real applications (Zhang et al., 1998). While hidden nodes are important to capture the nonlinear structure in the data, they are not as important as input nodes for nonlinear time series forecasting (Zhang et al., 2001). Following the common practice, we use the cross-validation approach to select the best NN architecture. That is, the in-sample data are further split into a training set and a testing set. The training set is used to estimate the model parameter and the testing set is used to choose the final NN model. In this study, the last 2 years data in the in-sample are used as the testing set for model selection. We consider 10 different levels of input nodes: 1, 2, 3, 4, 12, 13, 14, 24, 25, and 36, and 7 hidden node levels from 2 to 14 with an increment size of 2. Thus, a total of 70 different networks are experimented in the model building process.

To see the effect of seasonal adjustment on the forecasting performance of neural networks, we use the most recent Census Bureau's X12-ARIMA seasonal adjustment program (Findley et al., 1996) to deseasonalize the original series. NNs are then fitted to the deseasonalized data and finally forecasts based on the nonseasonal data are transformed back to original scale using the forecast seasonal indices provided by the program.

We also investigate the issue of whether using dummy or trigonometric variables can enhance neural network's capability of modeling seasonal variations. Corresponding to dummy regression model, a 12-input NN with the same predictor variables used in (8) is constructed. The same idea is used to build NN models with input variables corresponding to those in trigonometric models (9) and (10). Of course, in these settings, since the input nodes are identified, the only thing left to be decided in NN modeling is the number of hidden nodes. The same experimental design for the hidden nodes as in the time series modeling described earlier is carried out to determine this parameter.

To ensure that the observed differences of performance between various models are not due to chance, we used a five-fold moving validation scheme with five out-of-sample periods from 1995 to 1999 in the study. Each out-of-sample (or validation sample) contains 12 observations, representing 12 monthly retail sales in a year. The length of the corresponding in-sample periods is fixed at 10 years. That is, for each validation sample, the previous 10 years of data are used as in-sample for model development. This ‘‘moving’’ validation approach with multiple overlapped in-sample data and different out-of-samples can provide useful information on the reliability of a forecasting model with respect to changing underlying structures or parameters over time. In this cross-validation analysis, all of the models are rebuilt or re-estimated each time a new validation sample is examined. Fig. 2 shows the five validation samples stacked over the 12-month forecasting horizon.

To evaluate and compare the forecasting performance of different models, we use three overall error measures in this study. They are the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE). Since there is no universally agreed-upon performance measure that can be applied to every forecasting situation, multiple criteria are therefore often needed to give a comprehensive assessment of forecasting models.

4. Results

Neural network training is conducted with a GRG2 based system (Hung and Denton, 1993; Subramanian and Hung, 1993). GRG2 (Lasdon and Waren, 1986) is a widely used optimization routine that solves general nonlinear optimization problems using the generalized reduced gradient method. As shown in a number of previous studies (Hung and Denton, 1993; Lenard et al., 1995), the GRG2 training algorithm has many advantages over the popular backpropagation based training systems. On the other hand, we use Forecast Pro to conduct the ARIMA model fitting and forecasting. In particular, we use the automatic model identification feature of Forecast Pro to choose the best model. The methodology is based on the augmented Dickey–Fuller test and the Bayesian Information Criterion (BIC). The capability of this package is documented in Goodrich (2000).

We first give a detailed analysis of various models on their performance for the 1999 retail sales forecasting. The in-sample period for model fitting and selection is from 1989 to 1998 while the out-of-sample consists 12 periods in 1999. As mentioned before, for NN time series modeling,

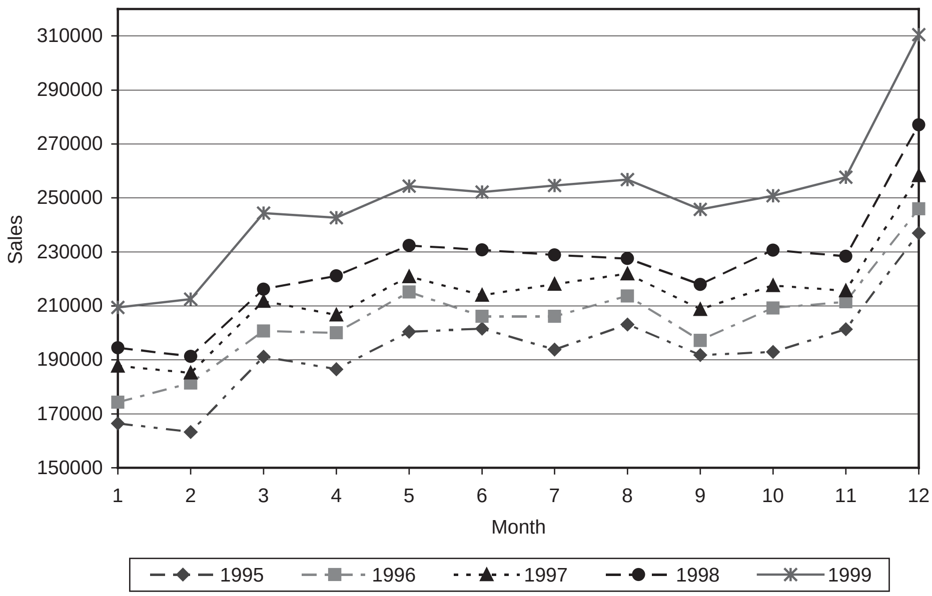

Fig. 2. Five out-of-samples.

the last 2 years in-sample data are used as the validation and testing sample and the rest of observations are used for model estimation. The model with the best performance in the testing sample will be selected as the final model for further validation in the out-of-sample. All model comparisons are based on the results for the out-of-sample.

The ARIMA model identified by the Forecast Pro is ARIMA$(0,1,1)(0,1,1)_{12}$ with the following mathematical relationship:

$$(1-B)(1-B^{12})y_t = (1-0.7613B)(1-0.5429B^{12})\varepsilon_t.$$

Fig. 3 shows the out-of-sample forecasts of this model. We find that overall, the ARIMA model follows the seasonal pattern exhibited in the sales data. However it does not provide good forecasts for the retail sales in 1999 because most of the forecasted values are below the actual, a clear under-forecasting situation. To see if neural networks are able to perform better, we plot the NN forecasts for the 1999 sales in Fig. 4. In addition to the direct modeling approach with the original time series observations, we also build neural networks with the transformed deseasonalized data. From Fig. 4, we find that the direct NN model performs even worse than the ARIMA model as almost all forecasts are relatively far below the actual values. On the other hand, it seems clear that the neural network model built with the deseasonalized data can improve the forecast accuracy dramatically although forecasts are still generally lower than the actual.

The under-forecast property of the above time series models leads us to examine the data more closely. Fig. 2 shows that the increases in almost all 1999 monthly sales from the previous year are significantly higher than those in all the previous 4 years. For example, retail sales in March for the years 1995–1998 are fairly close while the number for 1999 is much higher than that in 1998. This could be the major reason that causes the models built from the historical data before 1999 under-forecast the values in 1999.

Figs. 5 and 6 present the forecast comparisons between linear and nonlinear regression models with seasonal dummy variables and trigonometric functions, respectively. None of these models

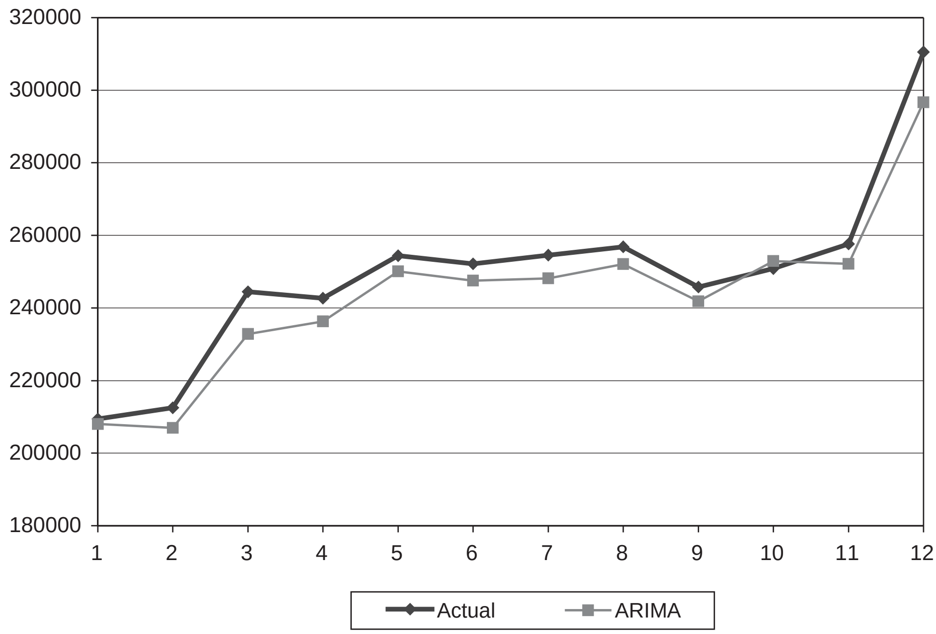

Fig. 3. ARIMA forecasts for 1999 sales.

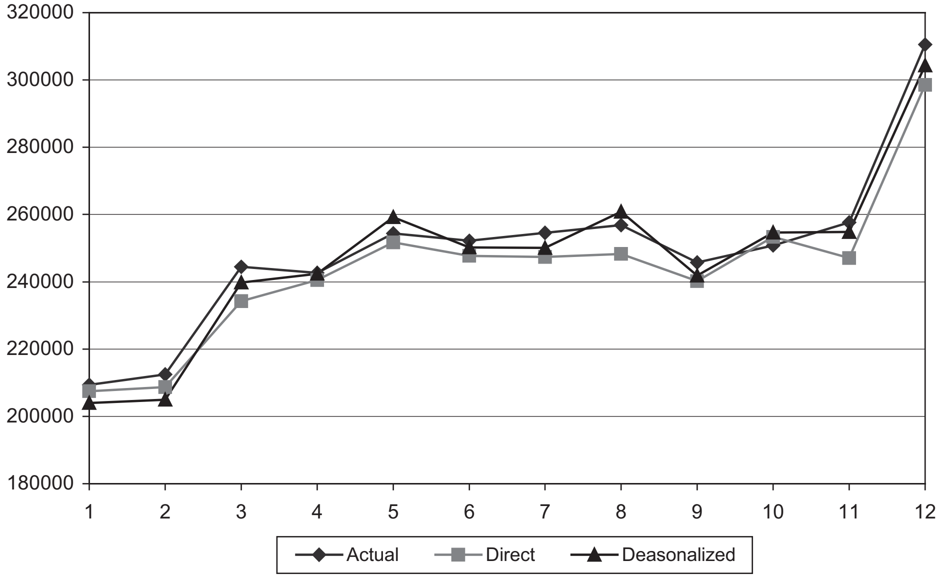

Fig. 4. NN time series forecasts for 1999 sales.

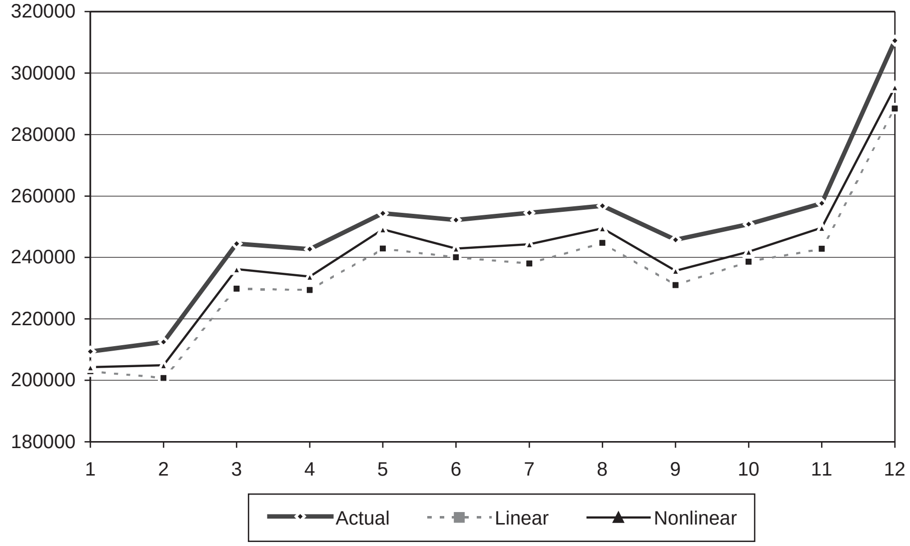

Fig. 5. Regression forecasts for 1999 sales.

provides satisfactory forecasts although the nonlinear regression models perform generally better than their linear counterparts. The dummy regression models consistently underforecast while forecasts from trigonometric models are not stable. In addition, while dummy regression models follow the general seasonal pattern fairly well, trigonometric models fail to capture the seasonal pattern in the out-of-sample.

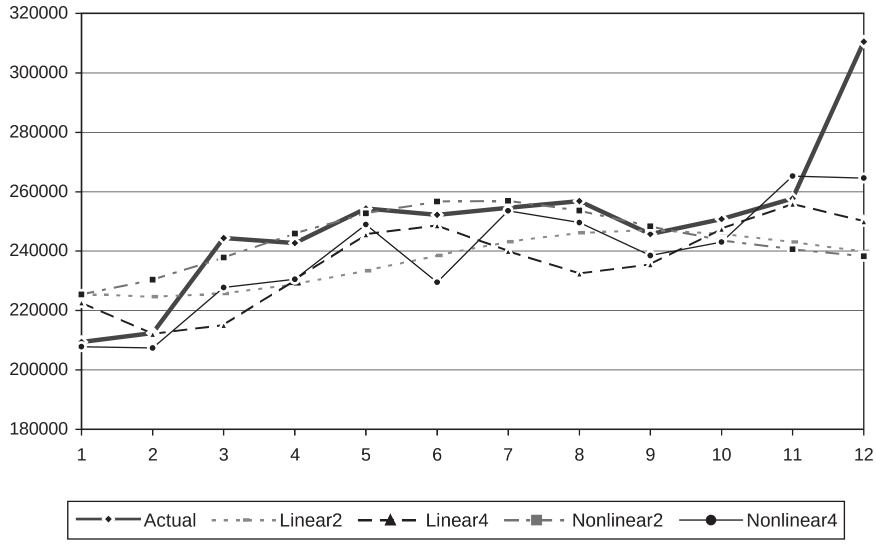

Fig. 6. Trigonometric forecasts for 1999 sales.

Table 1
Out-of-sample (1999) forecasting error measures of various models

Model	RMSE	MAE	MAPE
Linear			
ARIMA	6794.48	5866.51	2.30
Dummy regression	13 978.04	13 521.21	5.36
2-term trigonometric	24 233.05	17 423.35	6.67
4-term trigonometric	22 073.95	15 166.68	5.76
Nonlinear			
NN-direct	6825.46	5878.21	2.29
NN-deseason	4525.26	4147.26	1.69
NN-dummy variables	9013.17	8664.97	3.44
NN-2-term trigonometric	22 805.58	12 879.66	4.88
NN-4-term trigonometric	16 685.41	11 725.57	4.44

The summary statistics for the out-of-sample forecasting performance of all linear and nonlinear models are given in Table 1. All error measures confirm the impression we obtained with the time series plots presented earlier. Linear models as well as most of the nonlinear models do not perform well judged by all three criteria, though nonlinear models in general outperform linear models. The best model is the neural network built on deseasonalized data. This model has the most accurate forecast because of the lowest error measures. Moreover, compared to time series models, both linear and nonlinear regression models with dummy or trigonometric variables yield much worse forecasts for the 12 months in 1999.

Because of the possible change in the data structure or parameter of the model over time, we use a five-fold moving validation scheme described in the last section. From Fig. 2 and the discussion above, it seems quite possible that a one-size-fit-all model may not perform well over time. Therefore, it is important to demonstrate that the conclusion made from one particular sample can extend to others and is not due to chance alone. In this validation analysis, all of the models are rebuilt and parameters re-estimated each time a new validation sample is examined. It is important to note that the model does change each time a different in-sample is used to build ARIMA and NN models with regard to model structures and/or the parameters. For example, the ARIMA models used for validation samples of 1995 and 1996 are ARIMA$(0,1,1)(0,1,1)_{12}$, the same as that for 1999, but all with different parameters, and the models for both 1997 and 1998 are ARIMA$(0,1,1)(0,1,3)_{12}$, and again with different parameters.

Table 2
Forecasting comparison of time series models with five validation samples

Year	RMSE			MAE			MAPE		
	ARIMA	NN-direct	NN-deseason	ARIMA	NN-direct	NN-deseason	ARIMA	NN-direct	NN-deseason
1995	3708.23	3332.21	3242.05	3137.78	2786.50	2950.87	1.70	1.46	1.52
1996	4189.38	5281.78	4703.98	3239.66	4342.37	3631.94	1.61	2.16	1.80
1997	4499.30	4052.38	3919.85	3805.40	3246.29	3240.43	1.83	1.56	1.55
1998	5325.35	4458.28	3890.54	4111.72	3888.95	3194.99	1.77	1.75	1.45
1999	6794.48	6825.46	4525.26	5866.51	5878.21	4147.26	2.30	2.29	1.69
Mean	4903.35	4790.02	4056.33	4032.21	4028.47	3433.10	1.84	1.84	1.60
Std. dev.	1209.93	1338.30	580.63	1101.20	1193.09	467.95	0.27	0.37	0.14

Table 2 presents the overall results comparing ARIMA and NN time series models across the 5-year validation periods. Judged by all three overall accuracy measures and across five validation samples, neural networks with deseasonalized data (NN-deseason) perform the best overall, while ARIMA and neural networks modeled with original data (NN-direct) perform about the same. The overall superiority of the deseasonalized NN model is further confirmed by the summary statistics such as the mean and the standard deviation in Table 2 and the results from Wilcoxon rank-sum test presented in Table 3. The mean difference measures between NN-deseason and ARIMA and between NN-deseason and NN-direct are all significant at the 0.05 level while there is no significant difference between ARIMA and NN-direct. Note that the use of seasonal adjustment significantly reduces the variability of the neural network models in prediction as indicated by the standard deviations of the performance measures.

We notice that the forecasting performance of the NN-deseason model is not as good as that of the ARIMA model in 1996 across all three accuracy measures although it outperforms the ARIMA in other four validation samples. This observation suggests that though an overall best model can provide most accurate predictions over several forecasting horizons, judging by a specific portion of the forecasting horizon, it is possible that another model performs better. It further suggests that no forecasting model is always the best for all situations. Therefore, the importance of using multiple cross-validation samples to compare different forecasters becomes clearer.

Table 3
Mean difference between time series models

Comparison	Mean difference		
	RMSE	MAE	MAPE
ARIMA vs. NN-direct	113.33	3.75	0.00
ARIMA vs. NN-deseason	847.01[a]	599.11[a]	0.24[a]
NN-direct vs. NN-deseason	733.69[a]	595.37[a]	0.24[a]

[a] Significant at the 0.05 level with Wilcoxon rank-sum test.

Results of linear and nonlinear regression models are reported in Tables 4 and 5. Table 4 shows the results from regression models with seasonal dummy variables. Several observations can be made from this table. First, except for the 1999 case in which both linear and nonlinear regressions fail to predict well, the dummy variables are very helpful in improving forecasting performance. In fact, compared to the results in Table 2, dummy regression models forecast much better than all three linear and nonlinear time series models in the validation samples from 1995 to 1998. However, dummy regression models performed very poorly in forecasting 1999 sales with all three error measures more than doubling the size of those for time series models. Therefore, seasonal dummy variables may be useful for retail sales forecasting but the model may not be as robust as NNs built on deseasonalized data.

Second, with a flexible nonlinear model, the forecasting errors can be considerably reduced.

Table 4
Forecasting comparison of dummy regression models

Year	RMSE		MAE		MAPE	
	Linear	Nonlinear	Linear	Nonlinear	Linear	Nonlinear
1995	2452.43	2436.88	2010.12	1910.53	1.01	0.98
1996	4442.36	4442.70	3589.12	3589.36	1.76	1.76
1997	3367.96	3164.81	2965.38	2452.68	1.41	1.15
1998	3091.90	2834.43	2843.26	2453.65	1.26	1.10
1999	13978.04	9013.17	13521.21	8664.97	5.36	3.44
Mean	5466.54	4378.40	4985.82	3814.24	2.16	1.68
Std. dev.	4812.03	2697.75	4804.46	2779.81	1.81	1.03
Mean difference[a]	1088.14[b]		1171.58[b]		0.47[b]	

[a] Mean difference = linear measure–nonlinear measure.
[b] Significant at the 0.05 level with Wilcoxon rank-sum test.

Table 5
Forecasting comparison of trigonometric regression models with five validation samples

Year	RMSE				MAE				MAPE			
	Linear		Nonlinear		Linear		Nonlinear		Linear		Nonlinear	
	2-term	4-term	2-term	4-term	2-term	4-term	2-term	4-term	2-term	4-term	2-term	4-term
1995	15 216.88	12 784.23	14 394.33	9967.58	10 854.78	9772.07	10 590.19	8734.33	5.57	4.96	5.59	4.56
1996	14 780.34	12 527.97	15 255.59	10 663.07	10 192.38	9492.53	10 146.75	8574.49	4.91	4.53	4.67	4.13
1997	15 091.80	13 392.87	14 629.80	10 451.71	10 423.38	10 379.62	10 256.77	9248.37	4.79	4.74	4.55	4.33
1998	17 788.15	14 313.38	17 930.98	14 036.61	11 899.74	10 021.56	10 712.49	9551.54	5.21	4.35	4.68	4.02
1999	24233.05	22073.95	22805.58	16685.41	17423.35	15166.68	12879.66	11725.57	6.67	5.76	4.88	4.44
Mean	17 422.04	15 018.48	17 003.26	12 360.88	12 158.73	10 966.49	10 917.17	9566.86	5.43	4.87	4.87	4.29
Std. dev.	3993.59	4003.56	3536.30	2905.38	3015.03	2370.55	1121.37	1268.82	0.75	0.55	0.42	0.22

Not only do the nonlinear neural networks outperform their linear regression counterparts in almost all validation samples, but their results are also more stable as reflected by the standard deviation measure. The mean differences in three error measures between linear and nonlinear models are all positive and significant at the 0.05 level with Wilcoxon rank-sum test, suggesting the advantage of nonlinear model over its linear counterpart.

All performance measures reported in Table 5 are significantly worse than those obtained with the time series and dummy regression models. Thus using trigonometric functions to predict retail sales is not helpful at all. Although it is almost always the case that a nonlinear NN model performs better than its linear counterpart and a four-term trigonometric regression model is better than a two-term one, none of the trigonometric regression models examined is able to provide adequate forecasts for retail sales in all of the validation samples.

5. Conclusions

This paper presents a comparative study between linear models and nonlinear neural networks in aggregate retail sales forecasting. Accurate forecasts of future retail sales can help improve effective operations in retail business and retail supply chains. Since retail sales data present strong seasonal variations, we investigate the effects of different seasonal modeling strategies and techniques on their forecasting accuracy. Both time series approach and regression approach with seasonal dummy and trigonometric variables are examined in the study. Our results suggest that the nonlinear method is the preferred approach to modeling retail sales movement. The overall best model for retail sales forecasting is the neural network model with deseasonalized time series data.

Our study confirms the earlier work by Nelson et al. (1999) that prior seasonal adjustment of the data can significantly improve forecasting performance of the neural networks. While seasonal dummy variables can be very useful and promising in developing effective regression models for predicting retail sales, the performance of dummy regression models may not be robust and consistent. On the other hand, trigonometric models are not helpful in aggregate retail sales forecasting. This finding is contradictory to that in Williams (1997), which reports encouraging results by employing trigonometric variables in predicting daily rainfall. Examining the data plotted in Fig. 1 suggests that the retail sales series does not have a clear sinusoidal shape as in rainfall data. Therefore, this result may not be unexpected as trigonometric models are best for modeling sinusoidal behaviors.

In the forecasting literature, it is an established fact that no single forecasting model is the best for all situations under all circumstances (Makridakis et al., 1982). Therefore, the "best" model in most real world forecasting situations should be the one that is robust and accurate for a long time horizon and thus users can have confidence to use the model repeatedly. To test the robustness of a model, it is critical to employ multiple out-of-samples to ensure that the results obtained for one particular sample are not due to chance or sampling variations. The usefulness of this strategy in comparing and evaluating forecasting performance of various models is clearly demonstrated in our experiments.

References

Abraham, B., Ledolter, J., 1983. Statistical Methods for Forecasting. Wiley, New York.

Adya, M., Collopy, F., 1998. How effective are neural networks at forecasting and prediction? A review and evaluation. Journal of Forecasting 17, 481–495.

Agrawal, D., Schorling, C., 1996. Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model. Journal of Retailing 72 (4), 383–407.

Alon, I., Qi, M., Sadowski, R.J., 2001. Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. Journal of Retailing and Consumer Services 8 (3), 147–156.

Barksdale, H.C., Hilliard, J.E., 1975. A cross-spectral analysis of retail inventories and sales. Journal of Business 48 (3), 365–382.

Bowerman, B.L., O'Connell, R.T., 1993. Forecasting and Time Series: An Applied Approach, 3rd Edition. Duxbury Press, Belmont, CA.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Callen, J.L., Kwan, C.Y., Yip, C.Y., Yuan, Y., 1996. Neural network forecasting of quarterly accounting earnings. International Journal of Forecasting 12, 475–482.

Church, K.B., Curram, S.P., 1996. Forecasting consumers' expenditure: A comparison between econometric and neural network models. International Journal of Forecasting 12, 255–267.

Darbellay, G.A., Slama, M., 2000. Forecasting the short-term demand for electricity: Do neural networks stand a better chance? International Journal of Forecasting 16, 71–83.

De Gooijer, J.G., Franses, P.H., 1997. Forecasting and seasonality. International Journal of Forecasting 13, 303–305.

Elkateb, M.M., Solaiman, K., Al-Turki, Y., 1998. A comparative study of medium-weather-dependent load forecasting using enhanced artificial/fuzzy neural network and statistical techniques. Neurocomputing 23, 3–13.

Faraway, J., Chatfield, C., 1998. Time series forecasting with neural networks: A comparative study using the airline data. Applied Statistics 47, 231–250.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., Chen, B.C., 1996. New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. Journal of Business and Economic Statistics 16, 127–152.

Franses, P.H., Draisma, G., 1997. Recognizing changing seasonal patterns using artificial neural networks. Journal of Econometrics 81, 273–280.

Ghysels, E., Granger, C.W.J., Siklos, P.L., 1996. Is seasonal adjustment a linear or nonlinear data filtering process? Journal of Business and Economics Statistics 14, 374–386.

Goodrich, R.L., 2000. The Forecast Pro methodology. International Journal of Forecasting 16, 533–535.

Gorr, W.L., 1994. Research prospective on neural network forecasting. International Journal of Forecasting 10, 1–4.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Hung, M.S., Denton, J.W., 1993. Training neural networks with the GRG2 nonlinear optimizer. European Journal of Operational Research 69, 83–91.

Kirby, H.R., Watson, S.M., Dougherty, M.S., 1997. Should we use neural networks or statistical models for short-term motorway traffic forecasting? International Journal of Forecasting 13, 43–50.

Lasdon, L.S., Waren, A.D., 1986. GRG2 User's Guide. School of Business Administration, University of Texas at Austin.

Lenard, M.J., Alam, P., Madey, G.R., 1995. The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. Decision Science 26, 209–226.

Luxhoj, J.T., Riis, J.O., Stensballe, B., 1996. A hybrid econometric–neural network modeling approach for sales forecasting. International Journal of Production Economics 43, 175–192.

Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibdon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R., 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. Journal of Forecasting 1, 111–153.

McKenzie, E., 1984. General exponential smoothing and the equivalent ARIMA process. Journal of Forecasting 3, 333–444.

Mentzer, J.T., Bienstock, C.C., 1998. Sales Forecasting Management. Sage, Thousand Oaks, CA.

Nam, K., Schaefer, T., 1995. Forecasting international airline passenger traffic using neural networks. Logistics and Transportation Review 31 (3), 239–251.

Nelson, M., Hill, T., Remus, T., O'Connor, M., 1999. Time series forecasting using neural networks: Should the data be deseasonalized first? Journal of Forecasting 18, 359–367.

Ntungo, C., Boyd, M., 1998. Commodity futures trading performance using neural network models versus ARIMA models. The Journal of Futures Markets 18 (8), 965–983.

Pankratz, A., 1983. Forecasting with Univariate Box–Jenkins Models: Concepts and Cases. Wiley, New York.

Plosser, C.I., 1979. Short-term forecasting and seasonal adjustment. Journal of the American Statistical Association 74, 15–24.

Prybutok, V.R., Yi, J., Mitchell, D., 2000. Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. European Journal of Operational Research 122, 31–40.

Sharda, R., Patil, R.B., 1992. Connectionist approach to time series prediction: An empirical test. Journal of Intelligent Manufacturing 3, 317–323.

Subramanian, V., Hung, M.S., 1993. A GRG2-based system for training neural networks: Design and computational experience. ORSA Journal on Computing 5, 386–394.

Tang, Z., Fishwick, P.A., 1993. Feedforward neural nets as models for time series forecasting. ORSA Journal of Computing 5 (4), 374–385.

Thall, N., 1992. Neural forecasts: A retail sales booster. Discount Merchandiser 23 (10), 41–42.

Williams, P.M., 1997. Modelling seasonality and trends in daily rainfall data. In: Jordan, M., Kearns, M., Solla, S. (Eds.), Advances in Neural Information Processing Systems, Vol. 10. MIT Press, Cambridge, MA, pp. 985–991.

Zhang, G., Patuwo, E.P., Hu, M.Y., 1998. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting 14, 35–62.

Zhang, G.P., Patuwo, E.P., Hu, M.Y., 2001. A simulation study of artificial neural networks for nonlinear time series forecasting. Computers & Operations Research 28, 381–396.

Zhang, G.P., Qi, M., 2002. Predicting consumer retail sales using neural networks. In: Smith, K., Gupta, J. (Eds.), Neural Networks in Business: Techniques and Applications. Idea Group Publishing, Hershey, PA, pp. 26–40.