PERGAMON

Journal of Retailing and Consumer Services 8 (2001) 147–156

JOURNAL OF RETAILING AND CONSUMER SERVICES

www.elsevier.com/locate/jretconser

Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods

Ilan Alon[a,*], Min Qi[b], Robert J. Sadowski[c]

[a]*Department of Business and Economics, State University of New York Brockport, 350 New Campus Drive, Brockport, NY 14420, USA*
[b]*College of Business Administration, Kent State University, Kent, OH 44242-0001, USA*
[c]*Market Analysis, Greater Cleveland Growth Association, 200 Tower City Center, 50 Public Square, Cleveland, OH 44113, USA*

Abstract

Like many other economic time series, US aggregate retail sales have strong trend and seasonal patterns. How to best model and forecast these patterns has been a long-standing issue in time-series analysis. This article compares artificial neural networks and traditional methods including Winters exponential smoothing, Box–Jenkins ARIMA model, and multivariate regression. The results indicate that on average ANNs fare favorably in relation to the more traditional statistical methods, followed by the Box–Jenkins model. Despite its simplicity, the Winters model was shown to be a viable method for multiple-step forecasting under relatively stable economic conditions. The derivative analysis shows that the neural network model is able to capture the dynamic nonlinear trend and seasonal patterns, as well as the interactions between them.

Keywords: Aggregate retail sales; Forecasting; Box–Jenkins modeling ; Winters exponential smoothing; Artificial neural networks; Multiple regression; Time-series analysis

1. Introduction

This study is concerned with forecasting US aggregate retail sales, a time-series with trend and seasonal patterns. Artificial neural networks (ANNs) are compared to the more traditional time-series forecasting methods, including Winters' exponential smoothing, ARIMA model, and multivariate regression. These methods are chosen because of their ability to model trend and seasonal fluctuations present in aggregate retail sales data. The objectives of this article are two-fold: (1) to show how to forecast aggregate retail sales using ANN and (2) to display how various time-series forecasting methods compare in their forecasting accuracy of aggregate retail sales.

The reasons for this article are both theoretical and practical. Theoretically speaking, how to improve the quality of forecasts is still an outstanding question (Granger, 1996). For data containing trend and seasonal patterns, failure to account for these patterns may result in poor forecasts. Over the last few decades several methods such as Winters exponential smoothing, Box–Jenkins ARIMA model and multiple regression have been proposed and widely used to account for these patterns. ANN is a new contender in forecasting trend and seasonal data. Franses and Draisma (1997) suggested that ANNs be used to investigate how and when seasonal patterns change over time.

Industry forecasts are especially useful to big retailers who may have a greater market share. For the retailing industry, Peterson (1993) showed that larger retailers are more likely to use time-series methods and prepare industry forecasts, while smaller retailers emphasize judgmental methods and company forecasts. Better forecasts of aggregate retail sales can improve the forecasts of individual retailers because changes in their sales levels are often systematic. For example, around Christmas time, sales of most retailers increase. Moreover, models of forecasting individual store sales will often include assumptions about industry-wide sales and market-share. Indeed, these predictive models may embody an aggregate-retail-sales variable as a predictor variable. Accurate forecasts of aggregate retail sales have the potential to improve individual stores' sales forecasts, especially of larger retailers who may have a significant market share.

* Corresponding author. Tel.: + 1-607-436-3509.
E-mail addresses: i_aloni@yahoo.com (I. Alon), mqi@bsa3.kent.edu (M. Qi), sadowski@ix.netcom.com (R.J. Sadowski).

0969-6989/01/$ - see front matter
PII: S0969-6989(00)00011-4

"Accurate demand forecasting is crucial for profitable retail operations because without a good forecast, either too-much or too-little stocks would result, directly affecting revenue and competitive position" (Agrawal and Schorling 1996, p. 383).

This article shows the relative forecasting accuracy of ANNs in comparison to the more traditional time-series methods, including Winters', ARIMA, and regression models. The remainder of the paper is organized as follows. Section 2 provides a brief literature review which shows how ANNs compared to the more traditional methods of forecasting. Section 3 discusses the experimental plan. Section 4 gives an introduction to the multilayer feedforward ANN that is used in this research. The benchmark traditional statistical methods are also reviewed in this section. Section 5 reports the results of the comparisons. Finally, the conclusions are given in Section 6 and the managerial implications in Section 7.

2. Literature review

In the past decade, ANNs have emerged as a technology with a great promise for identifying and modeling data patterns that are not easily discernible by traditional statistical methods in fields as diverse as cognitive science, computer science, electrical engineering and finance. Qi and Maddala (1999) and Qi (1999), for example, showed that many studies in the finance literature evidencing predictability of stock returns by means of linear regression can be improved by a neural network (See Qi, 1996, for a comprehensive survey of ANN applications in finance).

ANNs have also been increasingly used in management, marketing and retailing. The types of applications include market response forecasting (Curry and Moutinho, 1993; Hruschka, 1993; Dasgupta et al., 1994; van Wezel and Baets, 1995; Agrawal and Schorling, 1996; Ainscough and Aronson, 1999), consumer choice forecasting (West et al., 1997; Davies et al., 1999), tourism marketing (Mazanec, 1992, 1994, and 1999; Davies et al., 1999), buyer and seller relationship analysis (Wray et al., 1994), and market segmentation analysis (Fish et al., 1995; Mazanec, 1999; Natter, 1999). The reader is referred to Krycha and Wagner (1999) for a comprehensive survey of ANN applications in management and marketing. Zhang et al. (1998) provided a comprehensive review of ANNs' usage in forecasting.

Despite the great potential of ANNs in time-series forecasting, the empirical findings, thus far, are somewhat mixed. In comparing ANNs and ARIMA models on the 50 M-competition series that are designated as most appropriate for the Box–Jenkins technique, Kang (1991) found that although the ARIMA model has a superior or equivalent mean absolute percentage error (MAPE) to that of the ANNs, the forecast error for the ANNs is lower when trend and seasonal patterns are in the data. Hill et al. (1996) showed that ANNs significantly outperform traditional methods of forecasting when forecasting quarterly and monthly data. Although theoretically speaking ANN may improve on the traditional time-series methods in forecasting a series with trend and seasonal patterns, Nelson et al. (1994) found that ANNs do not model the seasonal fluctuations in the data very well.

Foster et al. (1992) found that exponential smoothing is superior to ANNs in forecasting yearly data, and comparable in forecasting quarterly data. Monthly data were not used in their study. Winters' exponential smoothing model, in particular, has been found to provide superior forecasts in a variety of contexts. Chen and Winters (1966) used it in the context of predicting peak demand for an electric utility company. Dugan et al. (1994) showed that the Winters' model can outperform both the Census X-11 and the random walk models in predicting a variety of income statement items (i.e., sales, earnings before interest and taxes, interest expenses, earnings before taxes, tax expenses, and earnings before extraordinary items). This result was derived from a 15-year sample (1971–1985) of 127 manufacturing and retailing firms. Using aggregate level data, Alon (1997) found that the Winters' model forecasts aggregate retail sales more accurately then simple exponential and Holt's models. The Winters' model was shown to be a robust model that can accurately forecast individual product sales, company sales, income statement items, and aggregate retail sales.

In the 1980s, the overall impression was that for immediate and short-term forecasts ARIMA models provide more accurate forecasts than other econometric models (O'Donovan, 1983). This perception was reaffirmed when more recently Dugan et al. (1994) showed that the ARIMA model forecasted income statement items more accurately than Census X-11 and random walk models. Using 75 out of 111-series, Sharda and Patil (1990) found that Box–Jenkins performed as well as ANN models. For time series with a long history, Box–Jenkins and ANNs provided comparable results (Sharda and Patil, 1992; Tang et al., 1990). Kang (1991) also showed that the ARIMA model is the same or superior to ANNs in terms of MAPE in a variety of applications. Despite its old tradition, the Box–Jenkins approach is a formidable competitor in the forecasting arena.

Because both the ARIMA and the Winters' exponential smoothing models were shown to be powerful time series forecasting methods, they are used as benchmarks of comparison to the relatively new ANN. Multiple regression is also used because of its popularity in both industry and academia. The following section provides the experimental plan of this paper.

3. Experimental plan

Forecasts of monthly aggregate retail sales are made using three traditional statistical methods including Winters' exponential smoothing, Box–Jenkins ARIMA model, and multivariate regression. These forecasts are used as benchmarks to be compared to the forecasts made by ANNs. ANNs are the most recent and most sophisticated statistical technique of the group and, therefore, are expected to outperform the traditional methods of forecasting.

The case of aggregate retail sales is chosen for several reasons. First, aggregate retail sales time series contain both trend and seasonal patterns, providing a good testing ground for comparing forecasting methods (see Fig. 1). Second, practitioners can benefit from more accurate forecasts of aggregate retail sales. Better forecasts of aggregate retail sales have the potential to increase the profitability of retailers and their suppliers because the sales of their companies may partly follow aggregate patterns. Third, forecasting aggregate retail sales is important to governmental officials designing and implementing optimal public policy for the retailing industry. A forecasted decline in aggregate retail sales, for example, can be offset by expansionary policy. Finally, more accurate forecasts of aggregate retail sales may improve portfolio investors' ability to predict movements in the stock prices of retailing chains. Particularly, unexpected forecasted changes in aggregate consumer-retail spending can create opportunities for capital gains.

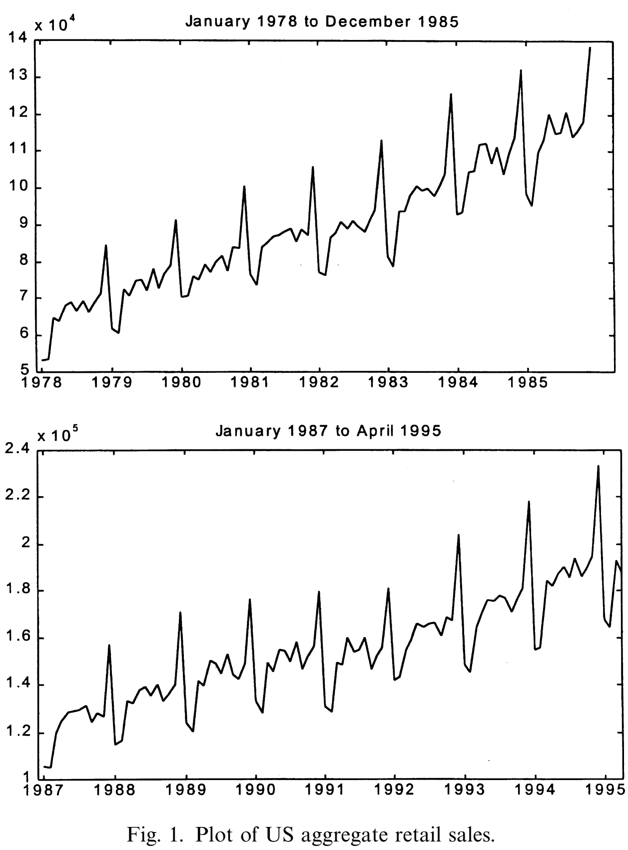

Fig. 1. Plot of US aggregate retail sales.

Both one-step and multiple-step forecasts are employed. Multiple-step forecasts provide up to 12-month projection into the future, while the one-step forecasts project the data only one month into the future. Multiple-step forecasts are important in facilitating longer-term planning and decision making. They simulate the real-world forecasting environment in which data need to be projected for longer than one period. For Winters model, multiple-step forecasts are accomplished by inserting a forecasted value in place of an actual observation in the smoothing exponential.

For the one-step forecast, each model is estimated recursively using all the data that are observable before the month to be forecasted. To minimize the impact of possible structural changes, the estimation and forecasting are carried out recursively for the ARIMA, regression, and ANN models.

The robustness of the alternative forecasting methods are tested in two time periods with different economic conditions: from January 1978 to December 1985 (period 1), and from January 1986 to April 1995 (period 2). Data are obtained from the *Current Business Report, Monthly Retail Trade: Sales and Inventories*, compiled by U.S. Department of Commerce. The first time period was marked with supply push inflation, two recessions, high interest rates, and high unemployment (Branson, 1989). The second time period is characterized by less fluctuations in the macroeconomy. Because aggregate retail sales are affected by macroeconomic instability, we expect the forecasts in the second time period to be more accurate.

Forecasts are compared via mean absolute percentage error (MAPE). This measure is used because of its popularity in the forecasting literature and because it is not prone to changes in the magnitude of the time series to be forecasted. Since aggregate retail sales tend to increase over time, measures of prediction error such as the root mean square error may be misleading when two periods are compared. The last 12 observations in each period are reserved to test the out-of-sample forecasting accuracy of both the one-step and the multiple-step forecasts.

To test whether the forecasts from two competing models are equally accurate, we use the Diebold and Mariano (1995, DM hereafter) test for the significance of the difference between the squared percentage errors of the two models. The advantage of the DM test over the commonly used t-test is that the DM test is applicable even when the errors are non-Gaussian, nonzero mean, serially correlated, and contemporaneously correlated. We denote the percentage forecast error of model A by $\hat{\varepsilon}_{A,t}$, $t = 1,2,\ldots,N$, the percentage forecast error of model B by $\hat{\varepsilon}_{B,t}$, the difference of square percentage errors by

$d_t = (\hat{\varepsilon}^2_{B,t} - \hat{\varepsilon}^2_{A,t})$, and the spectral density of d_t at frequency zero by $f_d(0)$. The DM test is based on the statistic

$$DM = \bar{d}/\sqrt{2\pi \hat{f}_d(0) N^{-1}},$$

where $\bar{d} = N^{-1}\sum_{t=1}^{N} d_t$, and $\hat{f}_d(0)$ is a consistent estimate of $f_d(0)$. Under the null hypothesis of equal forecast accuracy, the mean difference of square percentage errors is zero, and the asymptotic distribution of DM is standard normal. We use Newey and West's (1987) method to obtain a consistent estimate of the spectral density at frequency zero. Andrews' (1991) approximating rule is used to set the truncation lag. The idea is to calculate the truncation lag using the Bartlett kernel

$$lag = Int\left\{1.1447\left[4\left[\frac{\hat{\rho}}{(1-\hat{\rho})(1+\hat{\rho})}\right]^2 N\right]^{1/3}\right\},$$

where Int takes the integer part of a number, and $\hat{\rho}$ is the first-order serial correlation coefficient of d_t.

An alternative method to compare forecast accuracy is to employ Wilcoxon's signed-ranks test (SR), which is distribution free. The test statistic is

$$SR = \sum_{t=1}^{N} I_+(d_t) rank(|d_t|),$$

where

$$I_+(d_t) = \begin{cases} 1 & \text{if} \quad d_t > 0, \\ 0 & \text{otherwise} \end{cases}$$

where $rank(|d_t|)$ denotes the rank of the absolute value of d_t. The SR test gives an observation with a larger absolute-squared percentage-error difference a higher weight than that with a smaller difference. Upon scaling, this statistic is asymptotically standard normal

$$\frac{SR - N(N+1)/4}{\sqrt{N(N+1)(2N+1)/24}} \tilde{=} aN(0,1).$$

4. Models description

4.1. *Artificial neural networks* (*ANNs*)

ANNs are a class of generalized nonlinear nonparametric models inspired by studies of the brain and nerve system. The comparative advantage of ANNs over more conventional econometric models is that they can model complex, possibly nonlinear relationships without any prior assumptions about the underlying data-generating process (See Hornik et al., 1989, 1990; White, 1990). The data-driven nature of ANNs makes them appealing in time series modeling and forecasting. ANN models overcome the limitations of traditional forecasting methods, including misspecification, biased outliers, assumption of linearity, and re-estimation (Hill et al., 1996). They have been shown to be universal approximators, a property which makes them attractive in most forecasting applications. In addition, ANNs are more parsimonious than linear subspace methods such as polynomial, spline and trigonometric series in approximating unknown functions.

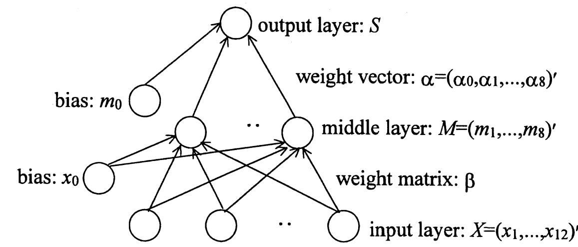

Fig. 2. A three-layer feedforward neural network.

Despite the many desirable features of ANNs, constructing a good network for a particular application is a non-trivial task. It involves choosing an appropriate architecture (the number of layers, the number of units in each layer, and the connections among units), selecting the transfer functions of the middle and output units, designing a training algorithm, choosing initial weights, and specifying the stopping rule.

It is widely accepted that a three-layer feedforward network with an identity transfer function in the output unit and logistic functions in the middle-layer units can approximate any continuous function arbitrarily well given sufficient amount of middle-layer units (White, 1990). Thus, the network used in this research is a three-layer feedforward one (Fig. 2). The inputs (similar to the regressors used in the multiple regression model) are connected to the output (similar to the regressand) via a middle layer. The network model can be specified as

$$\begin{aligned} S_t &= f(X_t, \alpha, \beta) + \varepsilon_t \\ &= \alpha_0 + \sum_{j=1}^{n} \alpha_j F\left(\sum_{i=1}^{12} \beta_{ij} x_{it} + \beta_{oj}\right) + \varepsilon_t, \end{aligned}$$

where S_t is the aggregate retail sales at time t; X is a vector of regressors that are exactly the same as the ones used in the multiple regression, which include a trend variable and eleven monthly dummy variables; n is the number of units in the middle layer; F is a logistic transfer function $F(a) = 1/(1 + \exp(-a))$, α_t represents a vector of coefficients (or weights) from the middle to output layer units; and β_t represents a matrix of coefficients from the input to middle-layer units at time t. The details of the specification and estimation of our ANN model is listed below:

(1) *Initial parameter values*: The initial values of α_t and β_t are generated using a technique suggested by Nguyen and Widrow (1990) so that the active regions of the layer's units will be distributed fairly evenly

over the range of inputs. The Nguyen–Widrow initial values have two advantages over purely random initial values. First, few units are wasted since the active regions of all the units are in the input space. Second, training works faster since each area of the input space has active regions.

(2) *Training algorithm*: The ANN network is trained using the Levenberg–Marquardt's algorithm which has been found to be the fastest method for training moderate-sized feedforward neural networks of up to several hundred weights (Demuth and Beale, 1997).

(3) *Number of middle-layer units*: Bayesian regularization has been implemented in the training algorithm because it provides a measure of how many network parameters are being effectively used by the network regardless of the total number of parameters in the network (MacKay, 1992). This procedure prevents overfitting and produces networks which generalize well, thus eliminating the guesswork required in determining the optimal number of middle-layer units (Interested readers should refer to Foresee and Hagan (1997) for a detailed discussion of the use of Bayesian regularization, in combination with Levenberg–Marquardt training algorithm.) Nonetheless, we experimented with several different numbers of middle-layer units and found that the results were not very sensitive to the number of middle-layer units because of the Bayesian regularization. The final number of middle-layer units was set to eight.

4.2. Benchmark statistical methods

4.2.1. Winters exponential smoothing

Historically, exponential smoothing models were used by approximately 13% of industry (Lilien and Kotler, 1983). These models use a weighted average of past values, in which the weights decline geometrically over time to suppress short-term fluctuations in the data. The accuracy of the forecasts from these techniques reflected the conformity of reality with the assumptions that the data follow some historical pattern and that the more current the observation the more relevance it has in predicting the future (Granger, 1980; Wilson and Keating, 1990).

Winters' model is a three parameter exponential smoothing model which incorporates a simple smoothing series, a trend effect, and a seasonal effect. The trend and the seasonal effects are smoothed in the same fashion as the series. Winters (1960) showed that his model is superior to the simple exponential model and the naive model in predicting sales of three products: kitchen utensils, paint, and cellars.

4.2.2. Box–Jenkins ARIMA model

Since its introduction in the 1970s, the Box–Jenkins approach has become one of the most popular method for time series forecasting. Box and Jenkins (1970) combined autoregressive (AR) and moving average (MA) techniques to model a time series. Box–Jenkins methodology requires several steps: model identification, estimation, evaluation, diagnostic checking, and prediction.

4.2.3. Multiple regression

Multiple regression can also be used to model time series with trend and seasonal patterns. The advantage of multiple regression is that it can incorporate more information about the time series than its past observations. The idea is to use a time trend variable and seasonal dummy variables as exogenous regressors to explain the series of interest.

5. Out-of-sample forecast results

The out-of-sample forecast performance of various models for the aggregate retail sales for the first and second periods are reported in Tables 1 and 2, respectively. In general, all four models performed well, with mean absolute percentage errors ranging from 1.16 (Winters

Table 1
MAPE (%) of various models for the first out-of-sample forecasting period (Period 1: January 1985 to December 1985)

Model	One-step forecast (Rank)	Multi-step forecast (Rank)	Average (Rank)
ANN	1.79 (1)	1.67 (1)	1.73 (1)
Box-Jenkins	2.20 (2)	2.02 (2)	2.11 (2)
regression	2.44 (3)	2.66 (4)	2.55 (3)
Winters	3.11 (4)	2.27 (3)	2.69 (4)
Average	2.39	2.15	2.27
	DM test	SR test	
Regression vs. ANN	2.4432 (0.0073)	2.0396 (0.0207)	
Winters vs. ANN	1.5711 (0.0581)	0.8629 (0.1941)	
Box–Jenkins vs. ANN	1.1558 (0.1239)	1.0983 (0.1360)	

Table 2
MAPE (%) of various models for the second out-of-sample forecasting period (Period 2: May 1994 to April 1995)

Model	One-step forecast (Rank)	Multi-step forecast (Rank)	Average (Rank)
Box–Jenkins	1.18 (1)	1.26 (2)	1.22 (1)
ANN	1.26 (2)	1.29 (3)	1.27 (2)
Winters	2.22 (3)	1.16 (1)	1.69 (3)
Regression	2.93 (4)	2.97 (4)	2.69 (4)
Average	1.90	1.67	1.79
	DM test	SR test	
Regression vs. ANN	1.8450 (0.0326)	1.8043 (0.0356)	
Winters vs. ANN	− 1.1461 (0.8741)	− 0.3922 (0.6526)	
Box–Jenkins vs. ANN	− 0.5972 (0.7248)	− 0.8629 (0.8059)	

multi-step forecast in period 2) to 3.11% (Winters one-step forecast in period 1). The forecasting accuracy of the various methods are compared in three dimensions: one-step vs. multiple-step, period 1 vs. period 2, ANNs vs. the traditional statistical methods. The test statistics as well as the *p*-values (in parenthesis) of the DM and SR tests are also reported in Tables 1 and 2 to compare the ANN to the three traditional statistical models for the multi-step forecasts.

5.1. One-step vs. multiple-step forecasts

Overall, the results of our analysis show that multiple-step forecasts are better than one-step forecasts for both time periods. This is a surprising result because one-step forecasts incorporate information that is more updated. This generalization may mask the reality of forecasts examined at the model level. For example, the Winters' model consistently displayed better results using multiple-step forecasts, while the regression model performed better using single-step forecasts.

In the first time period, with the exception of multiple regression, the multi-step forecasts were more accurate than one-step forecasts. The results are opposite in the second time period. With the exception of the Winters' model, the one-step forecasts outperformed the multiple-step forecasts. This suggests that when the macroeconomic conditions are less stable, multiple-step forecasts may provide better results than one-step forecasts. One possible explanation is that new data during turbulent times add more noise to the model rather than improving its efficiency in forecasting.

5.2. Period 1 vs. period 2

The MAPE was smaller in the second forecasting period than in the first period for all methods except multiple regression. This was true for both one-step and multi-step forecasts. The results confirm our expectation that the forecasts in the second period will be more accurate due to the more stable macroeconomic conditions. More stable economic conditions allow for more accurate forecasting of aggregate retail sales.

5.3. Neural networks vs. traditional methods

To get a sense of the overall performance of each method, the average MAPE across the two different forecasting horizons and the two different forecasting periods are reported in Table 3 for all the methods. ANN had the smallest average MAPE (1.50%), followed by Box–Jenkins ARIMA (1.67%), Winters exponential smoothing (2.19%), and multiple regression (2.75%).

In the first period (Table 1), ANN generated the smallest MAPE for both the one-step and multi-step forecasts, followed by Box–Jenkins model. The Winters exponential smoothing ranked third in the multi-step forecast and fourth in the one-step forecast, while multiple regression ranked third in one-step forecast and fourth in the multi-step forecast. Statistically speaking, for the multi-step forecast ANN outperformed the multiple regression at 5% significance level based on both the DM and SR tests (the *p*-values are 0.73 and 2.07%, respectively). It also performed better than Winters at 10% significance level based on the DM test (the *p*-value is 5.81%). The improvement of ANN over Box–Jenkins is not statistically significant at 10% level based on both

Table 3
Average MAPE of each model across two out-of-sample forecasting periods and two forecasting horizons

Model	Average MAPE (%)
ANN	1.50
Box–Jenkins	1.67
Winters	2.19
Regression	2.75

DM and SR tests. The data suggest that during turbulent economic times, ANNs generally provide superior forecasts over the traditional methods, but the Box–Jenkins model remains a formidable competitor.

In the second period (Table 2), Box–Jenkins generated the smallest MAPE for the one-step forecast while the Winters' model displayed the best results for the multi-step forecasts. ANN ranked second in the one-step forecast and third in the multi-step forecast. However, the forecasting improvements provided by the traditional methods were not statistically significant. Multiple regression had the largest MAPE for both forecasting horizons. ANN outperformed the multiple regression at 5% significance level for the multi-step forecast based on both DM and SR tests (the p-values are 3.25 and 3.56%, respectively).

The changes in the relative ranks of various methods between the two periods suggest that the traditional statistical methods, such as Box–Jenkins and Winters exponential smoothing models, may perform well during relatively stable economic conditions. The results also indicate that ANNs provide significantly better forecasts when the economy experiences large fluctuations.

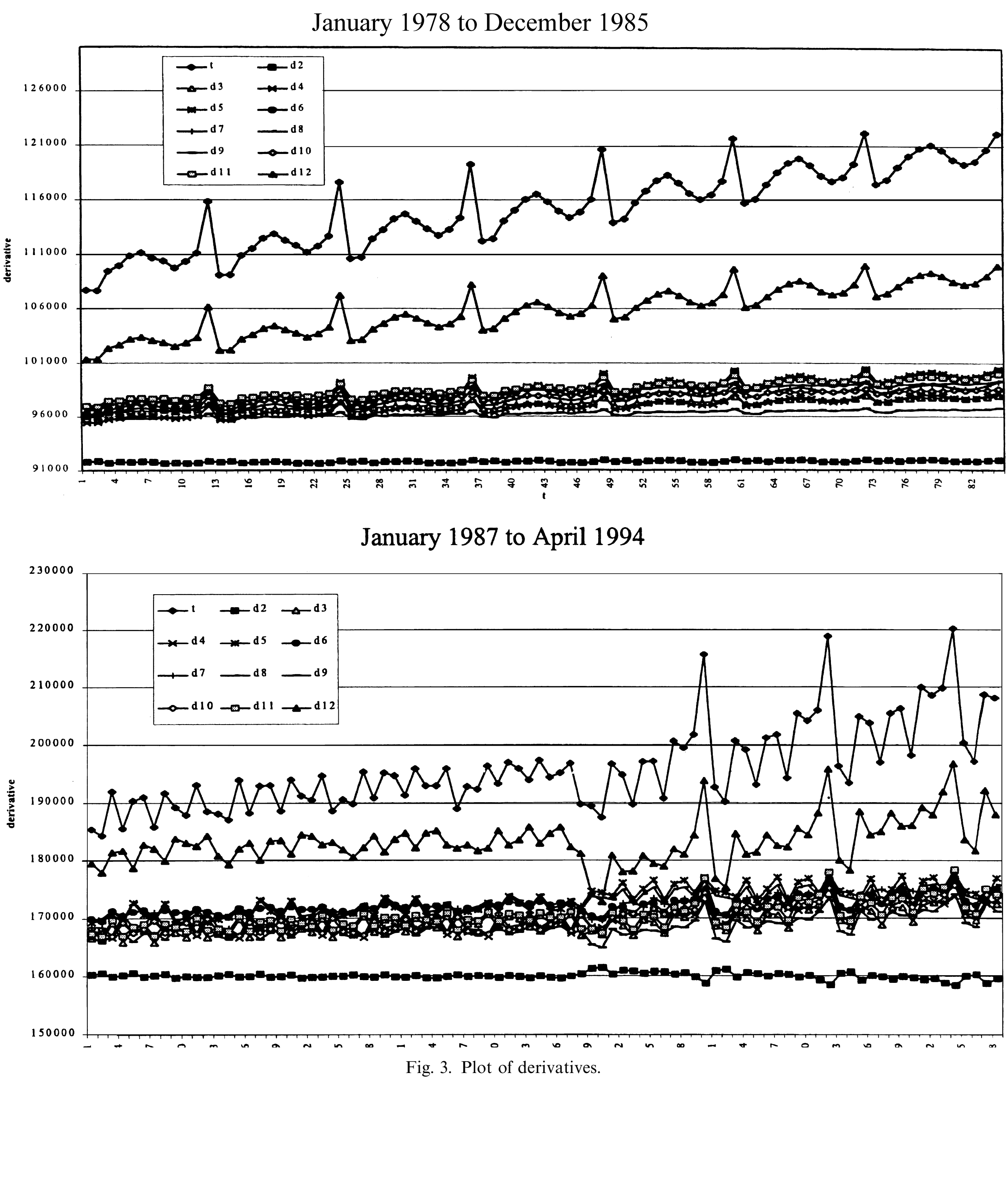

Fig. 3. Plot of derivatives.

5.4. Trend and seasonal patterns captured by ANN

Although most studies involving ANN acknowledge and utilize its ability to model complex nonlinear patterns, few have illustrated the captured patterns. This has led some to perceive ANN as a "black box". The complex interactions among the units in different layers make interpretation of the estimated weights difficult, if not impossible. Given this difficulty, in the present study, we examine the first-order derivative of the ANN output with respect to the input variables (the trend, t, and seasonal dummy variables, $d2$, $d3$, … $d12$, indicating February, March, December, respectively). These derivatives are constants in linear regression models, but are likely to change from observation to observation in nonlinear models such as ANN. To illustrate the dynamics in the trend and seasonal patterns captured by our ANN model, plots of these derivatives over time are provided in Fig. 3 for the multi-step forecasts. The upper plot shows how the derivatives change in the first period, while the lower plot shows the second period.

From Fig. 3, we observe the following patterns. First, the derivative with respect to the trend variable is the largest in magnitude compared to those with respect to the seasonal dummy variables for all observations in both periods. Among the seasonal dummy variables, the December dummy variable has the largest derivative, while the February dummy variable has the smallest derivative. The derivatives for the March through November dummies are similar in magnitude. This is consistent with the seasonal patterns observed in Fig. 1, which show peaks in December and troughs in February. Second, in both plots, for both the trend and the seasonal dummy variables, the derivatives appear to be increasing over time. More interestingly, for both the trend and the seasonal dummy variables, the derivatives themselves show seasonal patterns, again reaching a peak in December and a trough in February. This seasonal pattern of the derivatives is more obvious in the first period. The trend and seasonal influences and their interactions shown in Fig. 3 further demonstrate the advantage of ANN over the traditional statistical methods, i.e., ANN is capable of capturing dynamic nonlinear trend and seasonal patterns.

6. Conclusions

In this paper, we explored the goodness of fit of various time-series models outside their behavioral components. Aggregate retail sales are forecasted out of sample on a monthly basis using ANN and three traditional statistical techniques. Based on the empirical results, we make four conclusions. First, across different forecasting periods and different forecasting horizons the ANN performed the best, followed by Box–Jenkins and Winters exponential smoothing. Multiple regression with trend and seasonal dummy variables fared the poorest. Second, the ANN outperformed the traditional statistical methods in the first period in which the economic conditions were relatively volatile. When the macroeconomic conditions are relatively stable, Box–Jenkins and Winters exponential smoothing models delivered viable performance. Third, multiple-step forecasts may be preferred under volatile macroeconomic conditions because new data may not add much useful information to the forecasting model. Finally, the derivative plots show that ANN is able to capture the dynamic nonlinear trend and seasonal patterns and their interactions. One direction of future research is to include structural components into both the multiple regression and ANN models to see if the forecasting accuracy will improve. This will indicate whether the forecaster is better off using univariate analysis or structural analysis to forecast aggregate retail sales.

7. Managerial implications

The findings of this study have important managerial and practical implications. Companies should evaluate the costs and benefits of each model before choosing an appropriate forecasting tool. The four methods in this study differ in the necessary amount of historical data, forecasting horizon, personnel background, and software requirements (see Table 4). While the ANN model performs well overall, particularly under changing economic conditions, it is computationally difficult and requires special software, much computing time and great amount of in-house expertise. Furthermore, it is normally difficult

Table 4
Comparison of various methods of forecasting time series with trend and seasonal patterns

Forecasting method	Minimum data needed	Forecast horizon	Personnel background[a]	Standard software
Winters'	4–5 per season	Short to medium	2	No
ARIMA	50–100	Short to long	3	Yes
Regression	10–20	Short to medium	2	Yes
ANN	12–50	Short to long	5	No

[a]Personnel background = 1 (lowest) to 5 (highest).

to interpret an estimated ANN model. Companies need to evaluate the trade-off between forecasting accuracy and the above-mentioned costs and limitations of the method. Although this study showed that ANNs provided more accurate forecasts on average, ANNs lack parsimony. Therefore, it is prudent of companies to carefully evaluate the opportunity costs associated with switching their traditional forecasting models to ANNs.

Acknowledgements

The authors wish to thank Michael Y. Hu, Harry Timmermans, and the two anonymous referees for their helpful comments. Financial support from the College of Business Administration and the University Research Council at Kent State University is acknowledged. The usual disclaimer applies.

References

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.

Ainscough, T.L., Aronson, J.E., 1999. An empirical investigation and comparison of neural networks and regression for scanner data analysis. Journal of Retailing and Consumer Services 6, 205–217.

Alon, I., 1997. Forecasting aggregate retail sales: the Winters' model revisited. In: Goodale, J.C. (Ed.), The 1997 Annual Proceedings. Midwest Decision Science Institute, pp. 234–236.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.

Box, G.E.P., Jenkins, G.M., 1970. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.

Branson, W.H., 1989. Macroeconomic Theory and Policy, 3rd Edition. Harper and Row Publishers, New York.

Chen, G.K.C., Winters, P.R., 1966. Forecasting peak demand for an electric utility with a hybrid exponential model. Management Science 12, 531–537.

Curry, B., Moutinho, L., 1993. Neural networks in marketing: modeling consumer responses to advertising. European Journal of Marketing 27, 5–20.

Dasgupta, C.G., Dispensa, G.S., Ghose, S., 1994. Comparing the predictive performance of a neural network model with some traditional market response models. International Journal of Forecasting 10 (2), 235–244.

Davies, F., Goode, M., Mazanec, J., Moutinho, L., 1999. LISREL and neural network modeling: two comparison studies. Journal of Retailing and Consumer Services 6, 249–261.

Demuth, H., Beale, M., 1997. Neural Network Toolbox User's Guide, Version 3.0. The MathWorks, Inc., pp. 5–35.

Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13 (3), 253–263.

Dugan, M., Shriver, K.A., Peter, A., 1994. How to forecast income statement items for auditing purposes. Journal of Business Forecasting 13, 22–26.

Fish, K.E., Barnes, J.H., Aiken, M.W., 1995. Artificial neural networks: a new methodology for industrial market segmentation. Industrial Marketing Management 24, 431–438.

Foresee, F.D., Hagan, M.T., 1997. Gauss–Newton approximation to Bayesian regularization. Proceedings of the 1997 IJCNN.

Foster, B., Collopy, F., Ungar, L., 1992. Neural network forecasting of short, noisy time series. Computers and Chemical Engineering 16 (12), 293–297.

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

Granger, C.W.J., 1980. Forecasting in Business and Economics. Academic Press, Inc., New York.

Granger, C.W.J., 1996. Can we improve the perceived quality of economic forecasts?. Journal of Applied Econometrics 11, 455–473.

Hill, T., O'Connor, M., Remus, W., 1996. Neural network models for time series forecasts. Management Science 42 (7), 1082–1092.

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

Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Network 3, 551–560.

Hruschka, H., 1993. Determining market response functions by neural network modeling: a comparison to econometric techniques. European Journal of Operational Research 66, 27–35.

Kang, S., 1991. An investigation of the use of feedforward neural networks for forecasting. Ph.D. Dissertation, Kent State University, Kent, Ohio, 1991.

Krycha, K.A., Wagner, U., 1999. Applications of artificial neural networks in management science: a survey. Journal of Retailing and Consumer Services 6, 185–203.

Lilien, G.L., Kotler, P., 1983. Marketing Decision Making: A Model Building Approach. Harper and Row Publishers, New York.

Mazanec, J., 1992. Classifying tourists into market segments: a neural network approach. Journal of Travel and Tourism Marketing 1, 39–59.

Mazanec, J., 1994. Image measurement with self-organizing maps: a tentative application to Austrian tour operators. The Tourist Review 49 (3), 9–18.

Mazanec, J.A., 1999. Simultaneous positioning and segmentation analysis with topologically ordered feature maps: a tour operator example. Journal of Retailing and Consumer Services 6, 219–235.

MacKay, D.J.C., 1992. Bayesian interpolation. Neural Computation 4, 415–447.

Natter, M., 1999. Conditional market segmentation by neural networks: a Monte-Carlo study. Journal of Retailing and Consumer Services 6, 237–248.

Nelson, M., Hill, T., Remus, B., O'Connor, M., 1994. Can neural networks be applied to time series forecasting learn seasonal patterns: an empirical investigation. Proceedings of the 27 Annual Hawaii International Conference on System Sciences, pp. 649–655.

Newey, W., West, K., 1987. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Nguyen, D., Widrow, B., 1990. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. Proceedings of the International Joint Conference of Neural Networks 3, 21–26.

O'Donovan, T.M., 1983. Short Term Forecasting: An Introduction to the Box–Jenkins Approach. Wiley, New York.

Peterson, R.T., 1993. Forecasting practices in the retail industry. Journal of Business Forecasting 12, 11–14.

Qi, M., 1996. Financial applications of artificial neural networks. In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, Statistical Methods in Finance, 4. North-Holland, Elsevier Science Publishers, Amsterdam, pp. 529–552.

Qi, M., 1999. Nonlinear predictability of stock returns using financial and economic variables. Journal of Business and Economic Statistics 17 (4), 419–429.

Qi, M., Maddala, G.S., 1999. Economic factors and the stock market: a new perspective. Journal of Forecasting 18 (3), 151–166.

Sharda, R., Patil, R., 1990. Neural networks as forecasting experts: an empirical test. Proceeding IJCNN Meeting Vol. 2, pp. 491-494.

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

Tang, Z., de Almeida, C., Fishwick, P., 1990. Time series forecasting using neural networks vs. Box–Jenkins methodology. Simulation 57 (5), 303–310.

van Wezel, M.C., Baets, W.R.J., 1995. Predicting market responses with a neural network: the case of fast moving consumer goods. Marketing Intelligence and Planning 13 (7), 23–30.

West, P.M., Brockett, P.L., Golden, L., 1997. A comparative analysis of neural networks and statistical methods for predicting consumer choice. Marketing Science 16 (4), 370–391.

White, H., 1990. Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings. Neural Networks 3, 535–549.

Wilson, J.H., Keating, B., 1990. Business Forecasting. Richard Irwin, Inc, Homewood, IL.

Winters, P.R., 1960. forecasting sales by exponentially weighted moving averages. Management Science 6, 324–342.

Wray, B., Palmer, A., Bejou, D., 1994. Using neural network analysis to evaluate buyer-seller relationships. European Journal of Marketing 28 (10), 32–48.

Zhang, G.P., Patuwo, B.E., Hu, M.Y., 1998. Forecasting with artificial neural networks: the state of the art. International Journal of Forecasting 14 (1), 35–62.