Sales Forecasting Using Regression and Artificial Neural Networks

Conference Paper · November 2013

Jai Kedia
University of California, Irvine

Ryan Snyder
State Auto Insurance

R. Drew Pasteur
College of Wooster

Robert Wooster
Ancera, LLC

Sales Forecasting using Regression and Artificial Neural Networks

Giang H. Nguyen
College of Wooster
Wooster, OH
gnguyen14@wooster.edu

Jai Kedia
College of Wooster
Wooster, OH
jkedia15@wooster.edu

Ryan Snyder
College of Wooster
Wooster, OH
rsnyder14@wooster.edu

R. Drew Pasteur*
College of Wooster
Wooster, OH
rpasteur@wooster.edu

Robert Wooster, III*
College of Wooster [1]
Wooster, OH
rwooster@wooster.edu

ABSTRACT

The goal of this paper is to incorporate regression techniques and artificial neural network (ANN) models to predict industry sales, which exhibit a seasonal pattern, by using both historical sales and non-seasonal economic indicators. Both short-term and long-term predictive models were constructed, ranging from one-quarter predictions to twenty-quarter predictions. The step-by-step process was as follows: deseasonalize the data set, choose the relevant economic indicators using various statistical techniques, make predictions with ANNs, reseasonalize the predictions, and compute the errors of the predictions.

1. INTRODUCTION

Sales forecasting is an important aspect of many businesses today. Increasingly, companies are attempting to expand their forecasting abilities, in order to get the edge on their competitors. For example, a good forecasting model enables manufacturers to hold just the right amount of inventory to satisfy the demand for their product. In this paper, we construct a methodology that enables us to predict industry sales in a manufacturing sector. In order to avoid disclosing confidential information, we will refer to the input and output variables using generic terms and without units. Because we are concerned with relative error, the accuracy of the predictions does not depend on the magnitude of the data.

Many forecasting models use historical sales to predict future sales [13][14][15]. Our model is different in that we did not use historical sales as the sole input to the forecasting model. Instead, we utilized economic indicators as predictor variables, alongside historical sales. Since the economic indicators in our sample data do not exhibit a similar periodic pattern to the industry sales, we first deseasonalized the sales data using the LOESS smoothing method, as presented in section 2. Next, we chose relevant economic indicators by using statistical techniques. This is discussed in section 3. The chosen indicators were then input into ANNs to obtain predictions without seasonality. Our ANN methods are outlined in section 4. Finally, we reintroduced seasonality into our predictions and compared them to actual quarterly sales figures. This enabled us to calculate the percentage error of the model and accordingly evaluate the accuracy of our

---

[1] 1189 Beall Avenue, Wooster, OH 44691

*Faculty advisors

predictions. Sections 5, 6, and 7 illustrate these last steps. We present our results in section 8. Section 9 provides the conclusions and suggestions for future research.

2. SMOOTHING DATA

As mentioned above, the economic factors in our sample data do not display a seasonal trend, while industry sales do. Using non-seasonal economic factors to predict seasonal sales would not provide the best insight into the industry. Therefore, we decided to smooth (deseasonalize) the sales data using the LOESS method.

LOESS stands for "locally weighted scatterplot smoothing," and is a regression method used to smooth data [2]. With LOESS, we define a neighborhood for each point in the data set. The neighborhood is comprised of *k* points adjacent to the chosen point. LOESS then fits a quadratic regression curve to the points in the neighborhood. More weight is given points more closely adjacent in time than to more distant points. For example, we used an 11-point neighborhood to remove seasonality from the sales data. Thus, in order to find the value of the curve at a point A, the regression method will use five points prior in time to point A and five points subsequent to point A, as well as point A itself. Those points closer to point A will naturally receive higher weights when determining the smooth sales value at A. Figure 1 shows the raw sales data in blue and the smooth data in red.

In our data set, sales are recorded quarterly from the first quarter of 1996 to the first quarter of 2013 (1996 Q1-2013 Q1). Figure 1 shows that sales exhibit a strong seasonal pattern, with sales on average lowest in the first quarter and highest in the third quarter.

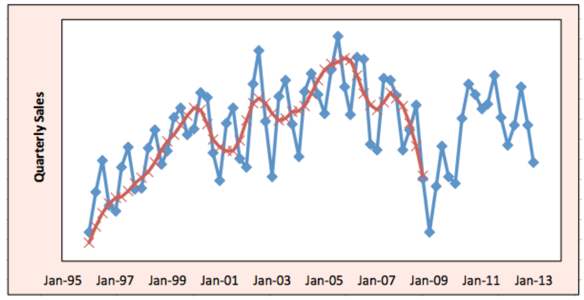

Figure 1. Raw and smoothed sales data

In Figure 1, there are two exceptions to the procedure of LOESS described above. First, the leftmost point in Figure 1 (1996 Q1) obviously does not have any points to its left. To counter this, the LOESS method simply gives more weight to the other six points. Similarly, the five data points from 2007 Q4 to 2008 Q4 were deseasonalized using fewer than eleven points. More weight is given to the other points in the neighborhood of the point chosen.

Figure 1 shows that the data is only smoothed up to the fourth quarter of 2008. This is because we fit our model on the first 52 observations (1996 Q1-2008 Q4), and then used that model to forecast the industry sales for the time period 2009 Q1-2013 Q1. Those predictions were then compared to the last 17 observations (2009 Q1-2013 Q1) in order to compute the error of our predictions. As such, had we deseasonalized the last 17 observations, we would be indirectly using future data in our forecasts. The first 52 observations served as a training set, whereas the last 17 observations were used to validate our predictions.

3. CHOOSING ECONOMIC INDICATORS

The next step in our forecasting process is to choose economic indicators that are relevant to the prediction of future industry sales. Based on the economic literature, we first identified potentially useful economic indicators, and then calculated the correlations between those economic indicators and the smooth historical industry sales. Strong correlation between an economic indicator and the historical sales suggests the potential relevance of that indicator. Economic indicators are then further narrowed down using statistical techniques, specifically, the Adjusted R-Squared, t-scores, correlation matrix, Variance Inflation Factor (VIF), and Akaike's Information Criterion (AIC) and Schwarz Criterion (BIC).

3.1 Regression Model

Regression analysis provides insight to determine the economic indicators that have the most predictive ability. Through the quantification of a single equation, regression can estimate the effect that a set of independent variables have on another (dependent) variable. In terms of our forecasting model, industry sales is the dependent variable, while economic indicators act as the independent variables, and regression estimates the strength and the direction of the economic relationships between them.

Building on a general multivariate linear regression model, we used the following regression equation,

$$Y_{t+1}=\beta_0+\beta_1 X_{1,t-s}+\ldots+\beta_k X_{k,t-s}+\varepsilon. \quad (1)$$

In (1), Y_{t+1} is the quantity of industry sales in the future period and X's are the past values of economic indicators. Since we are looking at the *predictive* power of economic indicators, we have to investigate the impact of the past values of the economic indicators on future sales, so we consider the time-lagged values of economic indicators. The subscript *t-s* means that the X's are lagged by *s* time periods.

In fact, we used the lagged values of economic indicators throughout all steps of our methodology: calculating the correlations, choosing economic indicators, and making predictions with ANNs.

We ran regression with different lag values, from 1 quarter to 20 quarters, in order to construct short-term as well as long-term predictive models. The next subsections present the statistical methods that were used to pick the economic indicators.

3.2 Adjusted R-Squared

The *adjusted coefficient of determination*, or the Adjusted R-Squared, is a commonly used measure of fit of a regression equation. It penalizes the addition of too many variables while rewarding a good fit of the regression equation [4] [8].[2] The Adjusted R-Squared is computed,

$$\bar{R}^2 = 1 - \frac{\sum e^2/(N-k-1)}{\sum(Y-\bar{Y})^2/(N-1)}. \quad (2)$$

Between two regression equations with the same dependent variable and different numbers of independent variables, the equation with higher Adjusted R-Squared has a better fit.

3.3 The t-scores

In a regression equation, if the t-score of an estimated coefficient is significant in the expected direction, the variable is more likely to be relevant to the equation. Further, if adding a variable to the equation significantly changes other variables' coefficients, the added variable probably should be included in the equation. This is because omitting a relevant variable will cause bias in other variables' coefficients [12].

In our variable selection process, we employed the t-score in the following way: suppose we are looking at the regression equation (1) with *k* independent variables. Suppose further that we need to determine if

[2] The R-Squared is a standard measure of fit in many texts. See, for example, [3] and [10]. Ohtani (2000) shows that adjusted R-squared is an unbiased estimator of the contribution of a set of explanatory variables X to the explanation of the dependent variable Y [8].

variable X_1 should belong to (1). We could run two regression equations: one with X_1 and the other without X_1. If the t-score for the estimated coefficient of X_1 is low and adding X_1 does not significantly affect the coefficients of other variables, X_1 probably should not be included in the regression.

3.4 Correlation Matrix

In building a forecasting model for sales, we frequently encountered issues of collinearity and multicollinearity among the independent variables. In the case of collinearity, one independent variable is a linear function of another variable. Multicollinearity means that one variable is a linear function of two or more variables. Due to the complexity of the economy, (multi)collinearity among economic factors is almost unavoidable.

The presence of multicollinearity in a regression equation might lead to false conclusions of insignificant estimates (type II error: accept a false null hypothesis). Multicollinearity also increases the likelihood of obtaining unexpected signs for the estimates [12]. Including irrelevant variables (that are already explained for the most part by other variables in the equation) will also add to the burden of heavy calculation. Therefore, we sought signs of multicollinearity in searching for the most predictive set of independent variables.

We detected collinearity in the equation by investigating the correlation matrix of the independent variables. The correlation matrix displays the correlation coefficients between all pairs of independent variables in the equation. If the magnitude of the correlation coefficient between two independent variables is high, these variables are strongly correlated; including both will result in collinearity in the equation. We also used the correlation matrix to determine which variables are highly correlated with a given variable. A variable that is highly correlated with many variables in the equation is likely to be dropped. Figure 2 presents an example of correlation matrix output from the statistical software Stata. High correlation values are highlighted.

	var1	l1gdp	l1dow	l1inve~t
var1	1.0000			
l1gdp	0.7927	1.0000		
l1dow	0.8262	0.8533	1.0000	
l1investment	0.8906	0.9198	0.8720	1.0000
l1industrial	0.8690	0.9549	0.9367	0.9560
l1m1	−0.8522	−0.9300	−0.8446	−0.8894
l1foodsales	0.8385	0.9868	0.8550	0.9460
l1exports	−0.7523	−0.9800	−0.7729	−0.9147
l1import	0.8039	0.9951	0.8644	0.9410

Figure 2. Correlation matrix.

3.5 High Variance Inflation Factors

The Variance Inflation Factor (VIF) allows us to detect multicollinearity among more than two variables [6] [12]. VIF investigates the extent to which an independent variable is explained by the rest of the independent variables in a regression equation. For each independent variable in the regression equation, a VIF is calculated.

Suppose we have built a regression equation with k independent variables:

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon \qquad (3)$$

As there are *k* independent variables in this equation, we are required to calculate *k* different VIFs, one for each X. The calculation of the VIF for a variable X_i involves two steps:

<u>Step 1</u>: Run a secondary (or auxiliary) ordinary least squares (OLS) regression, in which the chosen X_i is a function of the other independent variables in (3). For i=1, this regression equation is,

$$X_1 = \alpha_0 + \alpha_1 X_2 + \alpha_2 X_3 + \ldots + \alpha_{k-1} X_k + \tau \qquad (4)$$

in which α_i's are the regression estimates and τ is the error term.

<u>Step 2</u>: Calculate the VIF for the estimated coefficients of X_i, using the equation,

$$\mathrm{VIF}(\hat{\beta}_i) = \frac{1}{1-R_i^2}, \qquad (5)$$

in which R_i^2 is the unadjusted R-Squared of the auxiliary regression in (4).

A high VIF means that multicollinearity significantly impacts the equation. We paid particular attention to VIF values that are higher than 5. When VIF equals 5, the R-Squared of the auxiliary equation is 0.8. In this case, 80% of the movement of the chosen independent variable is explained by the other variables in the equation – a high degree of correlation.

3.6 Akaike's Information Criterion (AIC) and the Schwarz Criterion (BIC)

AIC and BIC are methods to compare different model specifications by adjusting the R-Squared for n observations and k independent variables [1] [11]. The equations are

$$AIC = \log(RSS/n) + 2(k+1)/n, \qquad (6)$$

$$BIC = \log(RSS/n) + \log(n)(k+1)/n. \qquad (7)$$

For each iteration of the model, involving a different set of economic indicators, we re-calculate the AIC and BIC. The regression equation with lower AIC is the better one. Similarly, the equation with lower BIC is the better model.

3.7 Using Statistical Criteria to Select Economic Variables

In this section we illustrate how we used all of the aforementioned statistical techniques to select indicators.

First, from a list of 90 economic indicators, we identified 56 whose correlation coefficients with industry sales are higher than 0.5.[3] From this list, we chose 18 indicators that represent different aspects of the industry and the economy. For example, if we had chosen Net Exports, we would leave Imports out of the model. Next, Ordinary Least Squares was used to estimate the regression coefficients. We recorded values for Adjusted R-Squared, mean VIF, AIC and BIC to decide how well the equation approximated the training data set. Variables that exhibit high VIF were then dropped from the equation, and the new equation was assessed with the same statistical criteria. Based on their correlation with sales, some variables were later introduced into the equation, and the same statistics were investigated. The process is repeated until a satisfactory level of fit is reached. Figure 3 summarizes the main regression equations of this selecting process. The equations that display improvement in the level of fit – represented by higher $\bar{R}^2$ and lower mean VIF, AIC and BIC – are in bold.[4]

Equation #	# of Variables	Adjusted R-Squared	Mean VIF	AIC	BIC	Variables Added/Dropped for Next Equation
1	18	0.9268	126.780	1334.14	1370.84	Drop only Industry Indicator III
2	**17**	**0.9290**	**70.000**	**1332.14**	**1366.91**	Drop only Industry Indicator IV
3	17	0.9289	126.720	1332.22	1366.99	Drop only Unemployment Rate
4	17	0.9283	113.190	1332.68	1367.45	Drop only Homeprice Index
5	17	0.9289	95.570	1332.24	1367.01	Drop Homeprice Index
6	**16**	**0.9310**	**42.000**	**1330.25**	**1363.09**	Drop Industrial Ratio
7	**15**	**0.9329**	**40.930**	**1328.27**	**1359.18**	Drop Industry Indicator IV
8	**14**	**0.9346**	**35.640**	**1326.38**	**1355.36**	Drop Dow Ratio
9	**13**	**0.9359**	**33.750**	**1324.73**	**1351.78**	Drop 10-year Treasury Rate
10	**12**	**0.9374**	**30.200**	**1322.89**	**1348.01**	Drop Commodity Price I
11	**11**	**0.9385**	**27.720**	**1321.32**	**1344.51**	Add Industrial Mining
12	**12**	**0.9455**	**28.080**	**1315.82**	**1340.93**	Drop Exports
13	**11**	**0.9468**	**7.390**	**1313.99**	**1337.17**	Add Employees in Trade
14	**12**	**0.9517**	**17.620**	**1309.71**	**1334.83**	Drop Unemployment Rate
15	**11**	**0.9527**	**8.980**	**1308.00**	**1331.18**	Drop Money Stock Acceleration
16	**10**	**0.9532**	**9.410**	**1306.70**	**1327.95**	

Figure 3. Summary of regression equations.

4. ARTIFICIAL NEURAL NETWORKS

Having chosen economic indicators to serve as inputs to the forecasting model, we used artificial neural networks (ANNs) to make predictions. ANNs are computational systems that are modeled on the human central nervous system itself, and are capable of pattern recognition, via machine learning, leading to predictive models [7].

There are three stages to making predictions with ANNs: training, validation and testing. In the training phase, the ANNs aim to identify patterns between the input data (the

[3] Data was obtained from the Federal Reserve Bank of St. Louis Economic Data [5].

[4] The t-scores were used to aid the decision of dropping or adding variables as discussed earlier. Due to space constraint, individual t-scores are not listed.

previously-chosen set of economic indicators) and the time-lagged smoothed sales. In the validation phase, ANN minimizes the error in its pattern identifications by avoiding overfitting, which could reduce predictive power of the model [9]. In the third stage, ANN makes predictions for sales in the remaining time periods (2009 Q1-2013 Q1). The actual data for sales during this period then allowed us to calculate the error of the model, assessing its merit. Figure 4 represents the structure of ANNs:

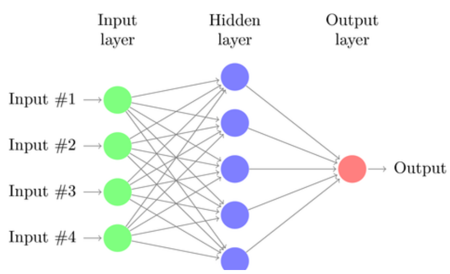

Figure 4. Structure of an ANN.

The inputs are the economic indicators coupled with smoothed sales data. Training and validation are carried out in the hidden layer. The hidden layer is comprised of *nodes*, represented by the blue circles in Figure 4. The number of nodes in the hidden layer can be modified from outside the ANN to make better predictions. The output is predicted sales data, which can be compared to actual sales to measure error.

To make better predictions, many ANNs may be used to make predictions, which are then combined, via measures of central tendency, such as the mean. Such a group of ANNs is called a *committee*. Increasing the size of the committee could also lead to reductions in error.

5. RESEASONALIZING PREDICTIONS

The biggest challenge of this project, mentioned earlier, was the difficulty in predicting seasonal sales from non-seasonal input variables, within our neural network model. We used the aforementioned deseasonalized sales data as our target values. In other words, the neural network fits the economic indicators to the smooth version of sales, rather than the actual sales. Thus the ANN output was deseasonalized sales. In order to make predictions for the actual sales, we reseasonalized our predictions by multiplying by appropriate coefficients. A reseasonalization coefficient was determined for each quarter. For example, third quarter sales were on average the highest, so the third quarter coefficient had the largest value.

To calculate the reseasonalization coefficients, we used sales data from our training set, which spanned the years 1996 to 2008, the same sales data used in training the ANNs. The process is as follows. First, for each quarter, we computed the quotient of raw and smooth sales. This gave us a coefficient for each quarter. For example, the quotient of the raw sales of 1996 Q1 to the smooth sales of 1996 Q1 equals 0.940. This means that the raw sales are roughly 94% of the smooth sales. This makes sense because in the first quarter, sales are generally lower, so the smooth sales would eliminate some of that seasonality by increasing sales.

After calculating the coefficient for each quarter as described above, we averaged all of the first-quarter coefficients, to determine the first quarter reseasonalization coefficient. We then repeated this process for the second, third, and fourth quarters. The coefficients are displayed in Table 1 below.

Table 1. Reseasonalization coefficients

Quarter	Coefficient
1	0.905
2	1.054
3	1.091
4	0.955

6. BASELINE PREDICTIONS

In this section, we discuss the determination of the best number of nodes and committee size to use when making the preliminary predictions for a particular lag (1 quarter to 20 quarters). These lags serve as 'prediction horizons', i.e. how far into the future we are looking to predict. In general, the further into the future we are trying to predict, the harder it is to make accurate predictions. The nodes and committee sizes to be tested were inputs for a Matlab code that used ANNs to make sales predictions. We determined the optimal numbers of nodes and committee members by observing the average predictive error, based on many trials of each combination. In general, larger committees led to more consistent results, but there is a diminishing-returns principle; very large committees add to computation time without greatly improving performance.

We initially tested ANNs with 5, 10, 15, 20, and 25 nodes, while committee sizes tested were 100, 250, and 500 networks. From these we determined which node and committee sizes gave the best results. We then ran the Matlab code again using five node values close to the one determined previously. For instance, if the value of 15 nodes is determined to be the best from the 5, 10, 15, 20 and 25 values tested initially, we would now test node values of 11, 13, 15, 17 and 19. This helps to hone in from the broader range of nodes to find a more specific node size. The best of these predictions served as a baseline error when testing other methods to reduce prediction errors. See the middle column in Table 2 for the baseline relative errors we obtained.

We now focus on the most significant of our prediction methodologies called the 'Repeated Training Model'

7. REPEATED TRAINING MODEL

In order to improve upon our baseline predictions, we used a method of *repeated training*. The idea is to re-train a committee of networks after each quarter. For example, we began by creating and training the committee the same way as the previous methods, using the data from 1996 Q1 through 2008 Q4. Then, we used that committee to predict industry market sales in the first quarter of 2009. Next, we retrained the committee, using data from the first quarter of 2009 in the training set. This newly trained committee was then used to predict sales in the second quarter of 2009. We continued this process until the committee was being trained on the data from 1996 Q1 through 2012 Q3 to predict the first quarter of 2013. This method can be adjusted for different lag periods.

It is important to note that the repeated training process involved updating the smooth data after predicting each quarter of the training data. Along with the data smoothing, reseasonalization coefficients were also updated using the most recent sales data. This technique was successful in reducing the relative error, and the results are given in the section 8.

8. RESULTS

Using the methods described in section 3, we identified a regression equation that best approximated future sales in the training set. The economic indicators used in that equation were:

- Industry Specific Indicator I
- Private Domestic Investment
- Real GDP
- Personal Saving Rate
- Concavity of the Yield Curve
- Moody's AAA Corporate Bond Yield
- Dow Jones Industrial Average
- Industry Specific Indicator II
- All Employees in Retail Trade
- Industrial Production Index: Mining

Baseline predictions were made using the methodology described in section 6. We then used the repeated training method to predict sales. Table 2 gives the relative errors using the baseline and the repeated training predictions associated with different lag times.

Table 2. Percentage error

Model	Error	Repeated Training Error
1Q	9.38%	4.89%
2Q	12.47%	6.50%
4Q	9.93%	8.15%
6Q	9.91%	5.49%
8Q	7.25%	4.70%
10Q	7.05%	5.58%
12Q	9.05%	7.00%
20Q	7.34%	5.51%

Figures 5 and 6 below highlight the improvement in performance of the repeated training model as compared to the baseline predictions in predicting sales. The first plot is for the 2Q lag model and the second is for the 10Q lag model.

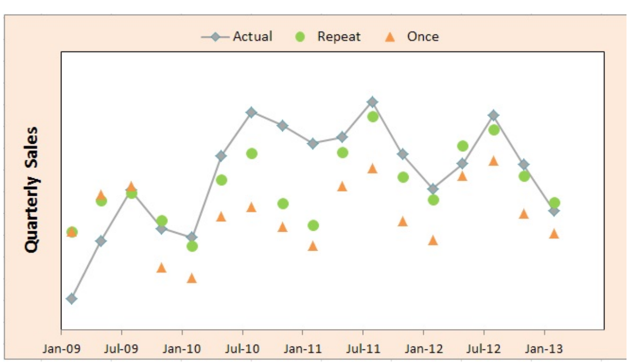

Figure 5. Plot for 2Q lag model.

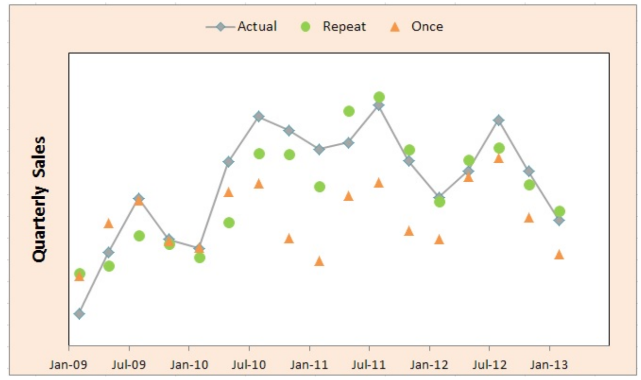

Figure 6. Plot for 10Q lag model.

Note that at the beginning of the validation set, the repeated training method made predictions similar to those made by the baseline method. However, as more data points were used in repeated training, the predictions improved.

9. CONCLUSION

Sales forecasting is a large and important field because it allows businesses to more accurately manage their inventory levels. The goal of our project was to apply the techniques of regression and ANN to predict future sales without using historical sales as the primary predictors. Instead of historical sales alone, we also used various economic indicators as the input variables of the artificial neural networks.

We overcame three obstacles in this project. First, we found a way to remove the seasonality from the sales data. Second, we used statistical techniques, including AIC, BIC, and VIF, to reduce the number of economic indicators from 90 to 10, retaining only those that were the most predictive for a given lag time. We input those indicators into an ANN and obtained unseasonalized sales predictions using repeated training. Finally, after reseasonalizing the predictions, we were able to predict future sales with relative error percentages ranging from 5 to

10 percent, depending on the lag time. These relative errors are a significant improvement over those obtained using the baseline predictions.

Our model takes into account what happens both in the industry and in the economy at large, as not only past trends but also changes in various economic factors affect future sales. Therefore, our methods could be applied in other business sectors. However, one would expect that models for different sectors use different sets of economic variables.

Future research in this topic could aim to automate the process of choosing relevant economic indicators used as inputs into the neural network. Since economic indicators are determined by comparing information about the level of fit of different models until a satisfactory level of fit is achieved, it might be worthwhile to create a computer program that executes all of the comparisons until the statistical measures reach a pre-established threshold. Such a technique would enable us to accomplish the determination of economic indicators in a more efficient manner.

10. REFERENCES

[1] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19, 6* (1974), 716-723.

[2] Cleveland, W.S. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74, 368* (Dec. 1979), 829-836.

[3] Diehr, G. and Hoflin, D.R. Approximating the distribution of the sample R2 in best subset regressions. *Technometrics*, *16, 2* (May 1974), 317–320.

[4] Ezekiel, M. M*ethods of Correlation Analysis*. J. Wiley & Sons, inc.; Chapman & Hall,limited, 1930.

[5] Federal Reserve Bank of St. Louis Economic Data. <http://research.stlouisfed.org/>, July 2013.

[6] Marquardt, D.W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, *12, 3* (Aug. 1970), 591–612.

[7] Mitchell, T. *Machine Learning.* WCB McGraw-Hill, 1997. 81-82.

[8] Ohtani, K. Bootstrapping R2 and adjusted R2 in regression analysis. *Economic Modelling*, *17, 4* (2000), 473–483.

[9] Priddy, K.L. and Keller, Paul E. *Artificial Neural Networks: An Introduction.* The International Society of Optical Engineering, 2005. 36 - 47.

[10] Rencher, A.C. and Pun, F.C. Inflation of R2 in best subset regression. *Technometrics, 22, 1* (Feb. 1980), 49–53.

[11] Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics, 6, 2* (Mar. 1978), 461-464.

[12] Studenmund, A.H. *Using Econometrics: a Practical Guide*. Addison-Wesley, Boston, MA, 2011.

[13] Taylor, J.W. Short-term electricity demand forecasting using double seasonal exponential smoothing. *The Journal of the Operational Research Society*, *54, 8* (Aug. 2003), 799–805.

[14] Winklhofer, H. et al. Forecasting practice: a review of the empirical literature and an agenda for future research. *International Journal of Forecasting*, *12, 2* (Jun. 1996), 193–221.

[15] Winters, P.R. Forecasting sales by exponentially weighted moving averages. *Management Science*, *6, 3* (Apr. 1960), 324–342.