Forecasting Women's Apparel Sales Using Mathematical Modeling

Celia Frank*[1], Balaji Vemulapalli[1], Les M. Sztandera[2], Amar Raheja[3]

[1]School of Textiles and Materials Technology
[2]Computer Information Systems Department,
Philadelphia University, Philadelphia, PA 19144
[3]Department of Computer Science,
California State Polytechnic, Pomona, CA 91768

Goal: The goal of the work is to demonstrate the effectiveness of soft computing methods like artificial neural networks and fuzzy logic models in apparel sales forecasting.

Project web site: http://vega.philau.edu/~ntc

1. Abstract

Traditionally, statistical time series methods like Moving Average (MA), Auto-regression (AR), or combinations of them are used for forecasting sales. Soft computing methods such as Fuzzy Logic, Artificial Neural Networks (ANNs), and Genetic Algorithms offer an alternative taking into account both endogenous and exogenous variables and allowing arbitrary non-linear approximation functions derived (learned) directly from the data. In this research three approaches have been investigated for forecasting women's apparel sales, statistical time series modeling, Univariate Time Series Fuzzy Logic modeling and modeling using Artificial Neural Networks (ANNs). Four years' sales data (1997 – 2000) was used as backcast data in the model and a forecast was made for two months of the year 2000. The performance of the models was tested by comparing one of the goodness-of-fit statistics, R^2, and also by comparing actual sales with the forecasted sales of different types of garments. On an average, an R^2 of 0.75 and 0.90 was found for Single Seasonal Exponential Smoothing and Winters' Three Parameter model, respectively. The model based on ANN gave a higher R^2 averaging 0.92 and the one based on time series fuzzy logic model showed an average R^2 of 0.75.Although, R^2 for ANN model was higher than that of statistical models, correlations between actual and forecasted were lower than those found with Winters' Three Parameter model.

2. Present Research

A foundation has been made for multivariate fuzzy logic based model by building an expandable database and a rule base. After a substantial amount of data is collected, this model can be used to make predictions for sales specific to a store, color, or size of garment.

2.1. Database

From January 1997 until February 2001, sales data was collected for each day for every class for every store. From March 2001 onward, more detailed information about each garment is available including its size and color. The data for the previous years (1997-2000) has much fewer independent variables as compared to the data for the current year 2001.

2.2. Pre-Analysis

Data was analyzed for trends and seasonality. This analysis helps in choosing an appropriate statistical model, although this kind of preparation is not necessary for soft computing based models. For pre-analysis, three classes were chosen, one each from the Spring, Fall and non-seasonal garment categories. Analysis was spread among all the categories to remove any bias due to type of class.

2.3. Weekly Trend

Sales data often showed a weekly trend with sales volume increasing during weekends as compared to weekdays. This observation was further supported by qualitative means by calculating Auto Correlations Functions (ACFs).

2.4. Annual Trend

Garment sales are generally seasonal with demand increasing for a particular type in one season and for a different type in another season. To investigate seasonality, the same methodology was used as was used to establish weekly trend. Both graphically as well as using ACF, it was shown that sales of all three classes under consideration show strong and distinct seasonal trend.

2.5. Methodology And Results

Evident from the format of the database for years 1997-2001 only sales information with respect to time for various garments is available. Hence, only univariate time series and soft computing models were investigated using this data.

From March 2001 onwards, much more elaborate sales data is available. Using this data set, a multivariate forecasting model can be implemented which could prove useful for inventory maintenance.

Six classes, two each from the Spring (AS and BS), the Fall (AF and BF), and the Non-seasonal (CN and DN) categories, were chosen for each model. They were built using four years sales data, and the next two months data was forecasted. The forecasted data was then compared with actual sales to estimate the forecasting quality of the model.

3. Univariate Time Series Model

3.1 Seasonal Single Exponential Smoothing

The equation for Seasonal SES is:

$$F_t = \alpha A_{t-s} + (1- \alpha) F_{t-s} \qquad \ldots(1)$$

where

F_t = Exponentially smoothed forecast for period t
s = Length of the seasonal cycle
A_{t-s} = Actual in the period t-s
F_{t-s} = Exponentially smoothed forecast of the period t-s
α = Smoothing constant, alpha

Weekly sales data was used for the forecast model given by the above equation. Hence, s was chosen to be 52 (number of weeks in a year). There are many ways of determining alpha.

Method chosen in the present work was based on Minimum Squared Error (MSE). Different alpha values were tried for modeling sales of each class and the alpha that achieved the lowest SE was chosen.

After choosing the best alpha value, a forecast model was built for each class using four years weekly data. Using the model, a weekly sales forecast was done for January and February of 2001. In order to forecast daily sales, the fraction contribution of each day was multiplied by total forecasted sales of each week.

Figure 1a. shows the actual versus fitted values of three year sales data and Figure 1b. shows the actual versus forecasted daily sales for class AS. The data for the remaining five classes are available on the project website. Table I gives the alpha value, R^2 of the model, and correlation coefficient between actual and forecasted daily sales for January 3 2001-February 27 2001 for all classes.

Table I. Values of Alpha, R^2, and Correlation Coefficients for Seasonal SES Model

Class	AS	BS	AF	BF	CN	DN
Alpha	1.4	1.4	0.9	1.3	1.3	1.0
R^2	0.738	0.832	0.766	0.872	0.762	0.831
Corr.[1]	0.906	0.893	0.862	0.910	0.892	0.722

1: Correlation coefficient between actual and forecasted sales of Jan 3 ‘01-Feb 27 ‘01

It can been seen that even with single parameter Seasonal SES, R^2 is on an average more than 0.75 implying that the model is able to explain 75% of the variation in the data. Correlation coefficients between actual and forecasted sales of January 3 ’01-February 27 ’01 are also quite high except for class DN.

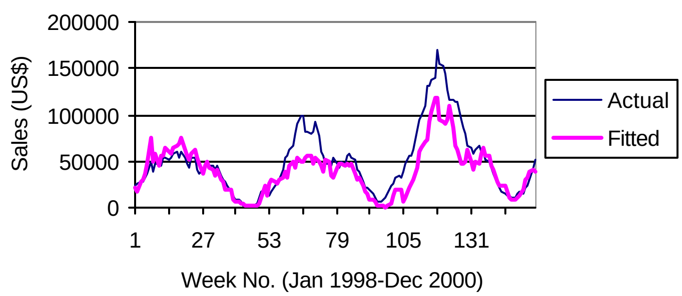

Figure 1a. Actual vs. Fitted sales value for Class AS using Seasonal SES Model

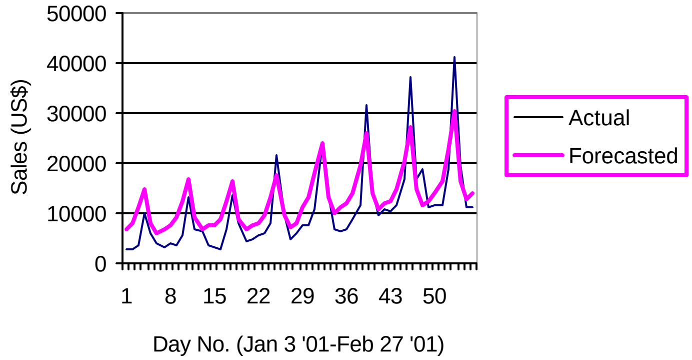

Figure 1b. Actual vs. Forecasted sales value for Class AS using Seasonal SES Model

3.2. Winters' Three Parameter Exponential Smoothing

Winters' powerful method models trend, seasonality, and randomness using an efficient exponential smoothing process. The underlying structure of additive trend and multiplicative seasonality of Winters' model assumes that:

$$Y_{t+m}=(S_t + b_t)\, I_{t-L+m} \qquad \ldots(2)$$

where

S_t = smoothed non seasonal level of the series at end of t

b_t = smoothed trend in period t

m = horizon length of the forecasts of Y_{t+m}

I_{t-L+m} = smoothed seasonal index for period t + m

That is, Y_{t+m} the actual value of a series, equals a smoothed level value S_t plus an estimate of trend b_t times a seasonal index I_{t-L+m}. These three components of demand are each exponentially smoothed values available at the end of period t. The equations used to estimate these smoothed values are:

$$S_t = á(Y_t/I_{t-L}) + (1 - á)\,(S_{t-1} + b_{t-1}) \qquad \ldots(3)$$

$$b_t = â(S_t - S_{t-1}) + (1 - â)b_{t-1} \qquad \ldots(4)$$

$$I_t = ã\,(Y_t/S_t) + (1 - ã)\, I_{t-L+m} \qquad \ldots(5)$$

$$Y_{t+m} = (S_t + b_t\, m)I_{t-l+m} \qquad \ldots(6)$$

where

Y_t = value of actual demand at end of period t

á = smoothing constant used for S_t

S_t = smoothed value at end of t after adjusting for seasonality

â = smoothing constant used to calculate the trend (bt)

b_t = smoothed value of trend through period t

I_{t-L} = smoothed seasonal index L periods ago

L = length of the seasonal cycle (e.g., 12 months or 52 weeks)

ã = smoothing constant, gamma for calculating the seasonal index in period t

I_t = smoothed seasonal index at end of period t

m = horizon length of the forecasts of Y_{t+m}

Equation 3 calculates the overall level of the series. S_t in equation 4 is the trend-adjusted, deseasonalized level at the end of period t. S_t is used in equation 6 to generate forecasts, Y_{t+m}. Equation 5 estimates the trend by smoothing the difference between the smoothed values S_t and S_{t-1}. This estimates the period-to-period change (trend) in the level of Y_t. Equation 5 illustrates the calculation of the smoothed seasonal index, I_t. This seasonal factor is calculated for the next cycle of forecasting and used to forecast values for one or more seasonal cycles ahead. For choosing á (alpha), â (beta), and ã (gamma) Minimum Squared Error (MSE) was used as criterion. Different combinations of alpha, beta, and gamma were tried for modeling sales of each class and the combination that achieved the lowest RSE was chosen. After choosing the best alpha, beta, and gamma values, the forecast model was built for each class using four years of weekly sales data. Using the model, a weekly sales forecast was done for Jan and Feb of 2001. In order to forecast daily sales afterwards, the fractional contribution of each day was multiplied by total forecasted sales of each week. Figure 2a. shows the actual versus fitted values of three year sales data and Figure 2b. shows the actual versus forecasted daily sales for class BS. This data for the five remaining classes is available on the project website. Table II gives the alpha, beta, gamma, R^2 of the model, and the correlation coefficient between actual and forecasted daily sales for Jan 3 2001-Feb 27 2001.

Table II. Values for Alpha, Beta, Gamma, R^2, and Correlation Coefficients for Winters' Model

Class	AS	BS	AF	BF	CN	DN
Alpha	0.60	0.50	0.50	0.80	0.60	0.50
Beta	0.01	0.01	0.01	0.01	0.01	0.01
gamma	1.00	0.47	0.91	0.72	1.00	0.82
R^2	0.923	0.969	0.951	0.685	0.941	0.933
Corr.[1]	0.903	0.920	0.869	0.667	0.927	0.777

1: Correlation coefficient between actual and forecasted sales of Jan 3 '01-Feb 27 '01

R^2 values for all the classes except BF are much higher than those obtained from Seasonal SES. Higher R^2 values and the ability of Winters' model to better define this is due to the additional parameter beta utilized for trend smoothing. Although curve fitting is very good, correlation coefficients between actual and forecasted sales are not as high. On observing the graphs of actual versus forecasted values for all the classes, it can be observed that trend (growth or decay) has always been over estimated and, hence, forecasted values are too high or too low. As in any multiplicative model, the division by very small numbers or multiplication by extremely large values is a problem with equation 3 which could have resulted in overestimation of the trend.

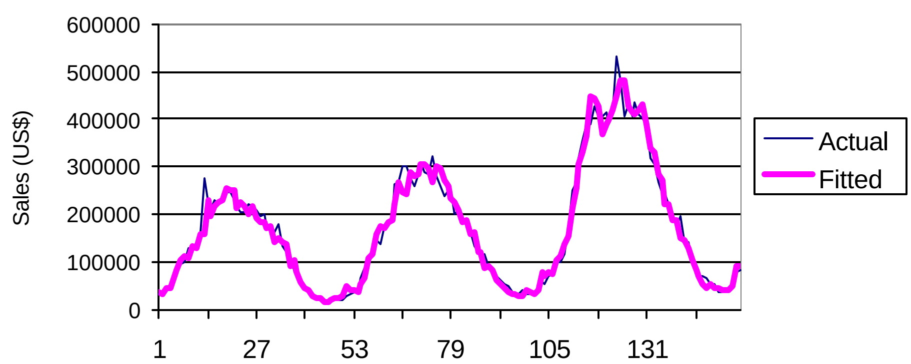

Figure 2a. Actual vs Fitted sales value for Class BS using Winters' Model (Jan 1998-Dec 2000).

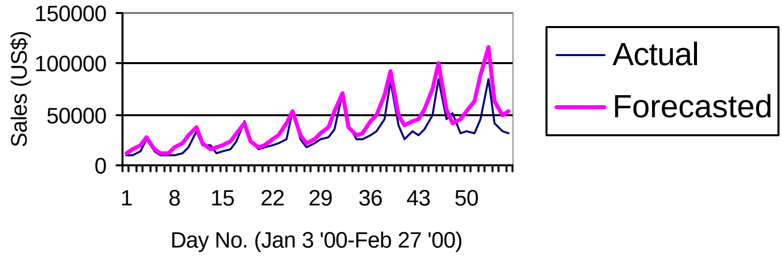

Figure 2b. Actual vs. Forecasted sales value for Class BS using Winters' Model

Soft computing methods are a rapidly growing area of computer science. These methods are being used to solve many problems such as optimization, pattern classification, forecasting, learning, etc. Artificial neural networks, genetic algorithms, fuzzy logic based reasoning, etc. are some of the popular soft computing methods being used to solve many real world problems. In our current research, we have used ANN to learn the patterns of sales of garments in the past to forecast sales in the future.

4.1. Artificial Neural Network (ANN) Model

Neural networks mimic some of the parallel processing capabilities of the human brain as models of simple and complex forecasting applications. These models are capable of identifying nonlinear and interactive relationships and hence can provide good forecasts.

In our research, one of the most versatile ANNs, the feed forward, back propagation architecture was implemented. The architecture of the feed forward neural network is shown in Figure 3. The hidden layers are the regions in which several input combinations from the input layer are fed and the resulting output is finally fed to the output layer. $\{x_1,,\ldots,x_M\}$ is the training vector and $\{z_1,,\ldots,z_M\}$ is output vector.

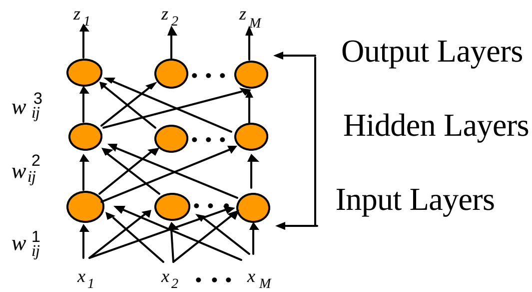

Figure 3. Multilayer Perceptron model used for forecasting sales

The error E of the network is computed as the difference between the actual and the desired output of the training vectors and is given in equation 7:

$$E = \frac{1}{P.N_L}\sum_{n=1}^{N_L}\sum_{p=1}^{P}(t_n^{(p)} - s_n^{(p)})^2 \quad \text{---(7)}$$

where $t_n^{(p)}$ is the desired output for the training data vector p and $s_n^{(p)}$ is the calculated output for the same vector. The update equation for the weights of individual nodes in the different layers is defined using the first derivative of the error E as given in equation 8:

$$w_i^{(n)}(l) = w_i^{(n)}(l) + \Delta w_i^{(n)}(l) \quad ; \qquad \text{where} \qquad \Delta w_i^{(n)}(l) = -h\frac{\partial E}{\partial w_i^{(n)}(l)} \qquad \ldots(\mathbf{8})$$

ANN consists of three layers the input layer, the hidden layer and the output layer with 10, 30 and 1 neuron respectively. 217 weeks sales data was divided into three parts. First part consisted of 198 weeks, which was used to train the network. The second part with 10 weeks data was used to test the network for its performance. The third part with 9 weeks data was used to compare forecasting ability of the network by comparing forecasted data with the actual sales data. In order to forecast daily sales afterwards, the fraction contribution of each day (given in Figure 3.2) was multiplied by total forecasted sales of each week. Figure 4a. shows the actual versus fitted values of three year sales data and Figure 4b. shows the actual versus forecasted daily sales for class CN. This data for the five remaining classes is available on the project website.

Table III gives the R^2 of the model, and correlation coefficient between actual and forecasted daily sales for Jan 3 2001-Feb 27 2001.

Table III Values of R^2, and Correlation Coefficients for ANN Model

Class	AS	BS	AF	BF	CN	DN
R^2	0.963	0.941	0.953	0.906	0.953	0.916
Corr.[1]	0.878	0.906	0.704	0.793	0.914	0.845

1: Correlation coefficient between actual and forecasted sales of Jan 3 ’01-Feb 27 ‘01

R^2 values for all the classes are much higher than those obtained from Seasonal SES, and Winters’ Model. High R^2 values and the strength of the ANN model are due to the ability of ANNs to learn nonlinear patterns.

Although curve fitting is very good, correlation coefficients between actual and forecasted sales are not that good. This might be due to over learning of the network. A potential problem when working with noisy data is the so-called over-fitting. Since ANN models can approximate essentially any function, they can also overfit all kinds of noise perfectly. Typically, sales data have a high noise level. The problem is intensified by a number of outliers (exceptionally high or low values). Unfortunately, all three conditions that increase the risk of over-fitting are fulfilled in our domain and have impacted correlations.

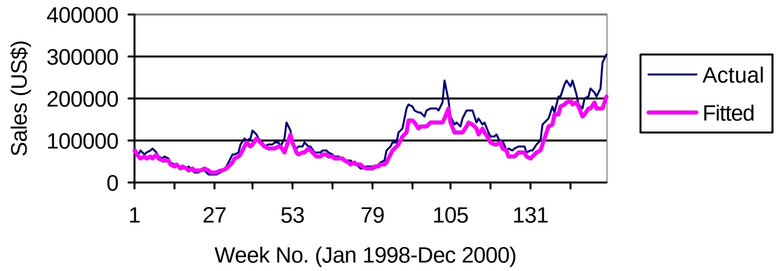

Figure 4a. Actual vs. Fitted sales value for Class CN using ANN Model

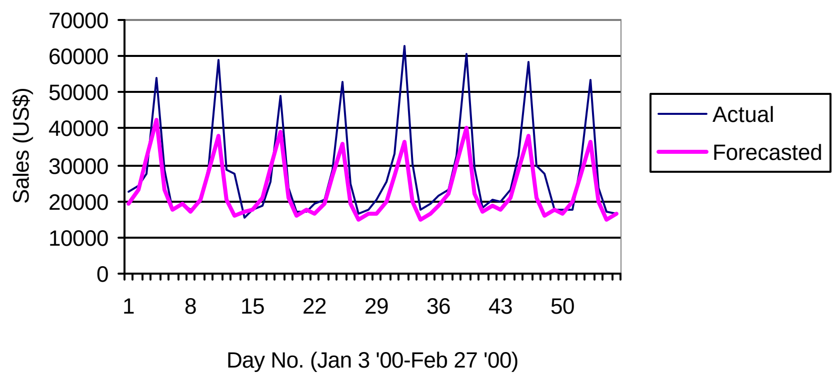

Figure 4b. Actual vs. Forecasted sales value for Class CN using ANN Model

4.2. Univariate Time Series Fuzzy Logic Model

Fuzzy logic allows the representation of human decision and evaluation process in algorithmic form. Fuzzy logic models are capable of deriving a solution for a given case from rules that have been defined for similar cases. In our research, univariate time series fuzzy logic model was implemented.

Fuzzy logic model was accomplished after 3 stages:

1**.** Fuzzification using linguistic variables: Linguistic variables were defined for all variables (day, month and sales) used in the forecasting.

2. Fuzzy inference: The rules that apply to the current situation were identified and the values of the output linguistic variable were computed. The computation of the fuzzy inference consists of two components:
Aggregation: Computation of the input variable using the minimum operator
Composition: Computation of the final output variable sales using the maximum operator.

3. De-fuzzification using linguistic variables: In this stage the linguistic sales value obtained in the previous stage was converted into a real value. This was accomplished by computing typical values and crisp result was found out by balancing out the conflicting results. The projected

value of sales was calculated in this step by approximating the typical values and then a final result was calculated.

Using this model, a weekly sales forecast was done for 1997,1998,1999 and 2000 years. In order to forecast daily sales afterwards, the fractional contribution of each day was multiplied by total forecasted sales of each week. The actual sales values were compared with the forecasted ones for the year 2000.Figure 5a shows the actual versus fitted values for year 2000 data and figure 5b shows the actual versus forecasted daily sales for class 41-Pant Sets.

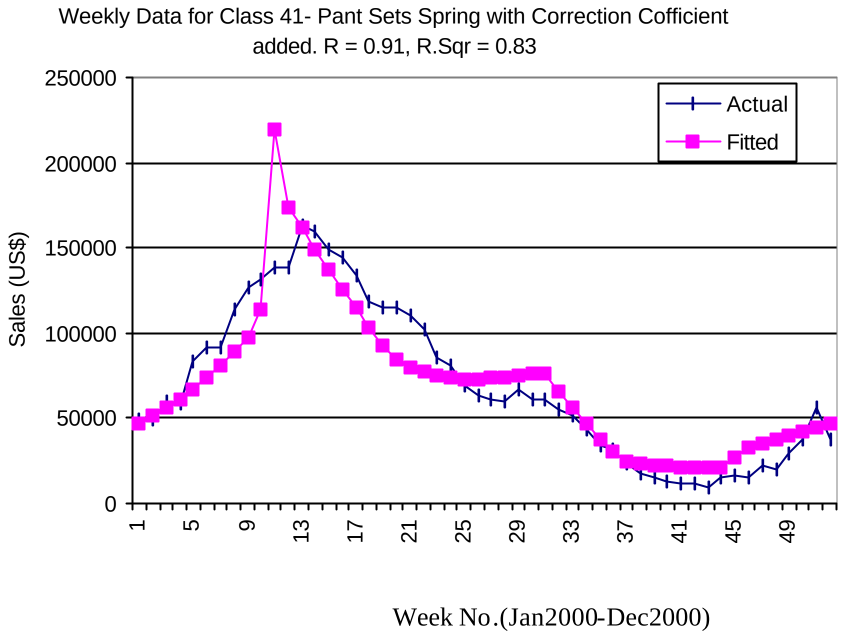

Figure 5a Actual versus fitted values for year 2000(weekly)

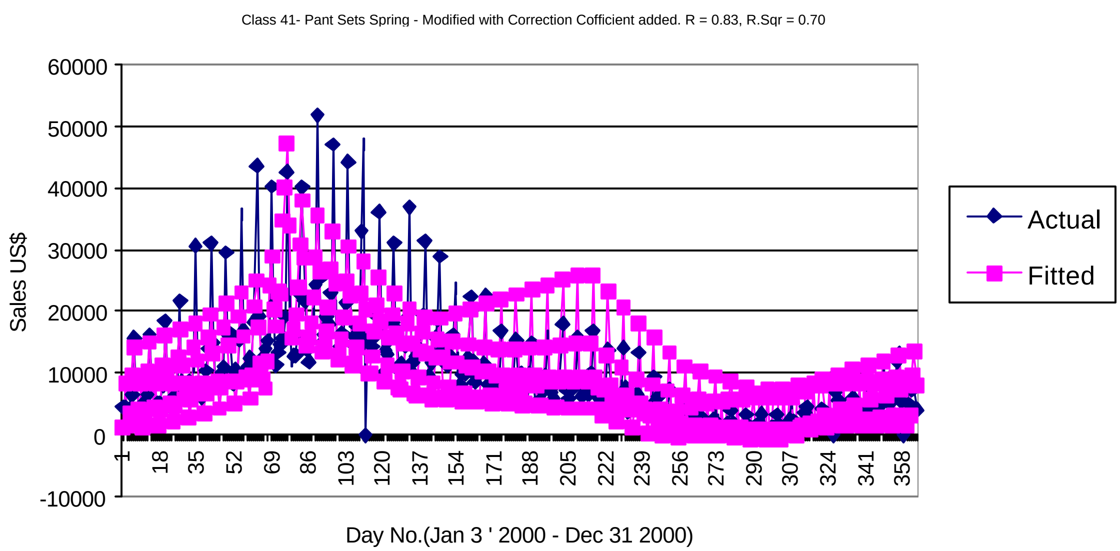

Figure 5b Actual versus fitted values for year 2000(weekly).

Low value of r square can be attributed to the univariate model not taking other factors like colour,size,class etc into consideration. A much more efficient way of applying the fuzzy logic model can be implemented by taking the above-mentioned factors into consideration. This can be implemented using the multivariate fuzzy logic model.

5. Conclusions

Time series analysis seemed to be quite effective in forecasting sales. In all of the four models, the R^2 and the correlation coefficient were significantly high. The three parameter Winters' model outperformed Seasonal SES in both explaining variance in the sales data (in terms of R^2) and forecasting sales (in terms of correlation coefficient). Univariate Time Series Fuzzy logic Model showed a very low R^2 value. This may be due to it taking only time as factor in predicting apparel sales.

ANN model performed best in terms of R^2 among three models. But correlations between actual and forecasted sales were not satisfactory. A potential problem when working with noisy data, a large number of inputs, and small training sets is the so-called over-fitting. Since big ANN models can approximate essentially any function they can also over fit all kinds of noise perfectly. Unfortunately, all three conditions that increase the risk of over-fitting are fulfilled in our domain. Typically, sales data have a high noise level. The problem is intensified by a number of outliers (exceptionally high or low values).

A multivariate fuzzy logic based model could model the sales very well, as it would take into account many more influence factors in addition to time. This naturally leads to the first extension of this work. Extensions of the concept of discovery learning are of current interest and are being investigated.

6. References

[1]. L. M. Sztandera, C. Frank, A. Garg, A. Raheja, A Computational Intelligence Forecasting Model for Apparel Sales, *Proceedings of the 29th International Aachen Textile Conference*, Aachen, November 2002.

[2]. A.Garg, Forecasting Women's Apparel Sales Using Mathematical Modeling, *Master's Thesis*, Philadelphia University, Philadelphia, April 2002.

[3]. C. Frank, A. Garg, A. Raheja, L. Sztandera, Forecasting Women's Apparel Sales Using Mathematical Modeling, *International Journal of Clothing Science and Technology*, 2002, under review.

[4]. J. S. Armstrong, *Principles of Forecasting*, Kluwer, Boston, 2001.

Acknowledgement We would like to acknowledge research support provided by Mothers Work, Inc.