Comparison study on univariate forecasting techniques for apparel sales

Min Li, W.K. Wong*, S.Y.S Leung

Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong, P. R. China

*Corresponding author:

Tel.: 852 27666471, Fax: 852 27731432,
E-mail address: tcwongca@inet.polyu.edu.hk

Abstract

This paper compared the forecasting performance of several commonly used univariate forecasting techniques for apparel sales forecasting. Extensive comparison experiments were conducted based on a large number of real-world apparel sales time series, including trend, seasonal, irregular and random data patterns. Comparison results showed that (1) For different data patterns, the forecasting performance generated by different univariate forecasting models are mixed. MA always generated worse forecasting results whichever data pattern was consider whereas NN(3) model generates the worst performance for random data pattern; (2) Different numbers of input variables could have large effects on forecasting results; (3) Different accuracy measures affects forecasting results largely.

Keywords: univariate sales forecasting, apparel retailing, neural networks, exponential smoothing, moving average

1. Introduction

Sales forecasting is the foundation for planning various phase of the firm's business operations [1, 2], which is a crucial task in dynamic supply chain and greatly affects retailers and other channel members in various ways [3]. Due to the ever-increasing global competition, sales forecasting plays a more and more prominent role in supply chain management. Recent research has shown that effective sales forecasting enabled large improvements in supply chain performance [4, 5].

Research in sales forecasting can be traced back to 1950s [1]. Since then sales forecasting has attracted extensive attention from academia. A large number of sales forecasting papers have been published, in which various forecasting techniques have been proposed, including Naïve, exponential smoothing, regression, moving average, autoregressive, and Autoregressive integrated moving average (ARIMA), neural networks.

It is well accepted that there does not exist a forecasting technique appropriate to all sales time series [6]. For example, for non-linear forecasting tasks, linear forecasting models cannot perform well and generate ideal results. However, it is unknown how each forecasting technique fits different apparel sales data.

In the apparel retail industry, sales forecasting activities mainly rely on qualitative methods, including panel consensus and historical analogy. These methods are mostly based on subjective assessment and experience of sales/marketing personnel with simple statistical analysis of limited historical sales data. Forecasting models tailor-made for specific apparel sales data patterns are desirable since forecasting apparel product sales is a challenging task.

Apparel retailers must improve their sales forecasting performance in order to deliver appropriate apparel finished product in appropriate time. Research in apparel sales forecasting have attracted some researchers' attention.

In the literature of apparel sales forecasting, the most majority of researchers employed NN technique to develop forecasting models [7-10]. It is well-known that NN model can perform better if more training samples are available. Unfortunately, available historical sales data of apparel products are usually insufficient due to frequent product changes and short selling season in apparel retailing, which will probably detract from the credibility of forecasts generated by NN-based models.

Previous studies in apparel sales forecasting usually utilized several sets of data to compare the forecasting performances of proposed intelligent models and several classical techniques. It is questionable if these comparisons are sufficient. In addition, it is still open and desirable to investigate whether the classical techniques used for performance comparison in these studies are fair and reasonable.

To a certain extent, the nature of data can determine what forecasting method can be used. For example, it is impossible to use ARIMA forecasting techniques if sufficient sample data are unavailable; it is also unnecessary to use a complicated nonlinear technique to forecast a simple linear time series. Witt et al. [11] have reported that different forecasting techniques might perform differently in handling stable vs. unstable data. It is well accepted that no forecasting technique is appropriate to all data patterns. However, no research has investigated and compared the effects of different techniques on different sales data patterns from apparel retailing, which leaves much room for further research exploration.

It is thus desirable to compare the adaptability of different forecasting techniques on different apparel sales data patterns based on a large number of experimental data. This research will conduct a comparison study on different univariate techniques for apparel retailing. Undoubtedly, this research will enrich greatly the study on forecasting techniques for apparel sales and it is helpful to identify and select benchmark forecasting techniques for different data patterns.

2. Methodology for forecasting performance comparison

This research will investigate the performances of different types of forecasting techniques when handling different types of sales data patterns. It includes these five steps:

(1) Collect a large number of apparel sales data from point-of-sales (POS) databases of a couple of apparel retailers headquartered in Hong Kong.

(2) Based on extensive analyses on real apparel sales data, several sales data patterns will then be used to represent change trends of most apparel sales data.

(3) Forecasting performances of several commonly used forecasting techniques will then be compared in terms of each data pattern.

(4) This research will also consider the effects of different numbers of input variables and accuracy measures on forecasting performance.

(5) Finally, the appropriateness and adaptability of different forecasting techniques on different apparel sales data patterns will be identified.

The details of the methodology are explained as follows.

2.1. Data patterns in apparel retail sales data

In apparel retailing, there exist different sales forecasting tasks, including sales forecasting of one or more products, one or more product categories, one or more shops, one or more cities. In these forecasting tasks, time series of sales data involve various data patterns. It is known that no forecasting technique is effective to all data patterns. To identify and compare the performances of different forecasting techniques for different types of sales data, apparel sales data is classified into four sales data patterns, including trend pattern, seasonal pattern, irregular pattern and random pattern, based on comprehensive analysis on a great number of apparel sales data..

2.2. Forecasting techniques used

Several representative univariate forecasting approaches, including Naïve, moving average(MA), autoregressive(AR), ARMA, exponential smoothing(ES) and neural networks (NNs), will be used and their forecasting performances will be compared so as to evaluate each technique's performance on each data pattern. For each model, different number of input data and different model settings will be used.

A time-series is a collection of observations taken sequentially at specified times, usually at 'equal intervals' (e.g. sales of an apparel product in successive months, seasons or years). Suppose we have an observed time series $(x_1, x_2, x_3, \ldots, x_T)$ and wish to forecast the future values such as $x_{T+1}, \ldots, x_N$. The $x_1, x_2, x_3, \ldots, x_T$ is called the in-sample data or training sample for model creation. The $x_{T+1}, \ldots, x_N$ is called out-of-sample data or testing sample for model testing.

$$time \quad series: \quad \{x_1, x_2, ..., x_t, \ldots, x_T, x_{T+1}, \ldots, x_N\}$$

Suppose the observed data are divided into a training sample of length T and a test sample of length N. Typically T is much larger than N. Let that $\hat{x}_{t+1}$ denotes the forecast for the period $t+1$.

3. Experiments

To compare forecasting performances of different univariate forecasting techniques, extensive experiments have been conducted based on real-world apparel sales data. This section presents how experimental data are collected and selected and how experiments are conducted

3.1 Experimental design

3.1.1 Apparel sales data collection

Appropriate experimental data are the basis of reliable experimental results. A large variety of real-world apparel sales data were collected from different apparel retail companies located in Hong Kong and Mainland China. The raw data are point-of-sales (POS) data from retail shops of different cities from 01/2000 through 05/2009, which are actual sales records of each apparel item in each retail shop. In apparel retail, it is extremely difficult and even impossible to predict the short-term sales of each apparel item by using time series forecasting techniques due to the highly uncertainties and randomness of their short-term sales. This research thus uses time series of medium-term aggregate sales, i.e., aggregate sales amount of an apparel product (or product category) in one or more retail shops (or cities) on monthly, quarterly or yearly basis. Incomplete data is an inevitable problem in handling most real-world data sources. Raw sales data are often incomplete in retailing practice. This research thus tries to collect complete sales data as experimental data for performance comparison. For each data pattern, a specified number of time series have been selected out for comparison experiments.

In this research, 105 time series are used in total for performance comparisons of univariate forecasting techniques, in which some time series consist of 103 observations. Due to page limit, these time series will not be presented in this paper. For each time series, the last 15% observations are used as out-of-sample to compare and evaluate the accuracy of forecasting models. For each out-of-sample observation, its previous sales data are used as training samples to set the forecasting model for making one-step-ahead forecast.

3.1.2 Four types of data patterns

Trend pattern: In this research, for a time series, we use a linear function to fit it's all observations. If all absolute percentage errors between observation points and the corresponding outputs of the linear function are less than 5%, the time series is identified as trend pattern. There are 15 time series of yearly sales data of one or more product categories (or cities). Although a yearly time series with more observations is more appropriate for performance comparison, it is hard to find trend time series with more observation due to the incompleteness and unavailability of raw sales data.

Seasonal pattern: In this research, for a time series with periodic changes, we use a linear function to fit the values on the same quarters (or months) of different years, if the absolute percentage error value between observation points and the corresponding outputs of the function are less than 5%, this time series can be identified as seasonal pattern. There are 30 time series of quarter or monthly sales data of one or more product categories (or cities).

Irregular pattern: If the time series data set partly includes the features of trend or seasonal series, the time series is identified as irregular data pattern. There are 30 time series of quarter or monthly sales data of one or more product categories (or cities).

Random pattern: If the time series data set does not includes any of the features of above three data patterns, the time series is identified as irregular data pattern. There are 30 time series of quarter or monthly sales data of one or more product categories (or cities).

3.1.3 Univariate forecasting models adopted

This research adopted a wider variety of models in greater depth and then makes some general comments on models and the model-building process. The univariate forecasting approaches introduced in section 2.2 will be adopted for performance comparison. In detail, the following models will be adopted.

(1) Naïve model: $\hat{x}_{t+1} = x_t$

(2) AR(2) model: It is an AR model using the latest two observations as input variables to forecast next data point. That is $\hat{x}_{t+1} = \alpha_1 x_t + \alpha_2 x_{t-1}$

(3) AR(3) model: It is an AR model using the latest three observations as input variables to forecast next data point. That is $\hat{x}_{t+1} = \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 x_{t-2}$

(4) MA(2) model: It is an MA model using the latest two observations as input variables to forecast next data point. That is $\hat{x}_{t+1} = \frac{x_t + x_{t-1}}{2}$

(5) MA(3) model: It is an MA model using the latest three observations as input variables to forecast next data point. That is $\hat{x}_{t+1} = \frac{x_t + x_{t-1} + x_{t-2}}{3}$

(6) ARMA(1,1) model: It is an AR model with one autoregressive term and one moving average term. That is $\hat{x}_{t+1} = c + \varepsilon_2 + \varphi_1 x_1 + \theta_1 \varepsilon_1$

(7) ARMA(1,2) model: It is an AR model with one autoregressive terms and two moving average term. That is $\hat{x}_{t+1} = c + \varepsilon_t + \varphi_1 x_{t-1} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$

(8) DES model: It is an ES model using the past observations as input variables to forecast next data point. That is $\hat{x}_{t+1} = \frac{2-\alpha}{1-\alpha} s'_t - \frac{1}{1-\alpha} s''_t$.

(9) TES model: It is an ES model using the past observations as input variables to forecast next data point. That is $\hat{x}_{t+1} = a_t + b_t + c_t$.

(10) NN(2) model: It is an NN model using the latest two observations as input variables to forecast next data point. That is $\hat{x}_{t+1} = f(x_t, x_{t-1})$

(11) NN(3) model: It is an NN model using the latest three observations as input variables to forecast next data point. That is $\hat{x}_{t+1} = f(x_t, x_{t-1}, x_{t-2})$

In the two NN models, the conjugate gradient backpropagation algorithm with Fletcher-Reeves updates is used as learning algorithm. The maximum of training epochs are 2000. The number of hidden neurons is 3 if the length of training data of time series is equal to or less than 15; otherwise it is equal to 2×the number of input variables+1. For each time series in the experiments, 30 different trials, each with randomly generated initial weights, are run so as to avoid randomicity of forecasts. The final forecast of each time point is the mean of forecasts generated by the 30 repetitive trials.

3.1.4 Accuracy measures

The investigated training accuracy measures include mean absolute deviation (MAD), Mean absolute error (MAE), mean absolute percentage error (MAPE) and mean absolute scaled error (MASE), Root mean square error (RMSE), proposed by Hyndman and Koehler [12, 13].

3.2. Experimental results

The objective of this comparison experiment is to evaluate the forecasting performances of different univariate forecasting techniques for four typical types of apparel sales data patterns. Numerical experiments have thus been conducted based on each type of data patterns respectively. The experiments are carried out on a desktop computer with an Intel□ Core™2 Duo 3 GHz processor and 2 GB of RAM, running MATLAB version 7.0 (R14).

The comparison results for each type of data pattern are described as follows. Due to page limit, the forecasts generated by each forecasting model will not be detailed in this paper. Instead, this research will present the values of each accuracy measure generated by each forecasting technique for each time series and their comparison results of these techniques.

3.2.1. Trend pattern

1) Forecasting performances generated by different models

To evaluate the effectiveness of each forecasting model, the forecasting performances generated by 10 forecasting models for trend data pattern are presented in Table A1 in Appendix. The value in this table represents the value of an accuracy measure generated by a forecasting model for a time series. Columns 3-17 show the forecasting performances of time series 1-15 respectively. For example, the value in the 3rd column and the 2nd row, 58744520.0, is the MAE value generated by Naïve model for time series 1. For each model, 4 performance values are shown according to 4 different forecasting accuracy measures. The 15 time series of trend pattern are all yearly sales data. The number of observations in each time series is less than 10. It is insufficient for such a small number of observations to establish an ARMA(1,2) model. Thus, Table A1 does not include the results generated by this model.

Take MAPE performances as an example to describe the forecasting performances generated by different models. When MAPE is used, the performances for 15 time series can be summarized in Table 1. In this table, the second and the third rows shows the minimal and the maximal MAPE values generated by each forecasting model for 15 time series respectively; the 4-6 rows show the number of time series for which the MAPE values generated by corresponding models are greater than 10%, 15% and 20% respectively. For example, for 15 time series, the minimal and maximal MAPEs, generated by the AR(2) model, are 01.% and 16.6% respectively. In addition, MAPEs of two series are greater 15% but MAPEs of all series are less than 20%. It can be easily found from this table that two AR models and two ES models generate very good forecasts while two MA models are unacceptable for these time series of trend pattern.

Table 1. Summary of forecasting performances in terms of MAPE

	Naïve	AR(2)	AR(3)	MA(2)	MA(3)	ARMA(1,1)	DES	TES	NN(2)	NN(3)
Min.	6.9%	0.1%	0.4%	13.5%	13.7%	1.4%	0.1%	0.1%	3.3%	2.1%
Max.	62.3%	16.6%	20.0%	100.6%	135.8%	29.0%	15.9%	25.5%	46.7%	73.5%
>10%	11	3	3	15	15	5	2	3	7	8
>15%	6	2	3	12	14	2	1	2	6	6
>20%	4	0	1	8	13	1	0	1	4	5

2) Forecasting performance comparisons in terms of different accuracy measures

For the 15 trend time series investigated, there is only one out-of-sample point for performance evaluation because time series with more yearly sales cannot be obtained. As a result, the comparison results generated by different forecasting accuracy measures are the same, which are shown in Table 2. The value in this table represents the number of time series for which the corresponding forecasting model generates the best forecasting performance. For example, the value '3' in the 3rd row and 2nd column represents that AR(2) model generates the best forecasting results for 3 time series. From Table 2, we can found that: (i) No model can generate obviously better performance than others; (ii) Naïve model and two MA models are the worst three. (iii) Two NN models perform poorly, which do not show any superiority than others.

Table 2. The number of time series for which the forecasting model generates the best forecasting performance

Performance Ranking	1	2	3	4	5	6	7	8	9
Naïve	0	0	1	0	0	3	2	9	0
AR(2)	3	7	1	3	0	0	1	0	0
AR(3)	3	1	2	5	2	1	0	1	0
MA(2)	0	0	0	0	0	0	0	3	12
MA(3)	0	0	0	0	0	0	0	0	1
ARMA(1,1)	1	2	1	0	6	2	2	1	0
DES	4	2	4	3	1	1	0	0	0
TES	2	2	4	2	4	0	1	0	0
NN(2)	1	1	1	1	2	6	1	1	0
NN(3)	1	0	1	1	0	2	8	0	2

3.2.2. Seasonal pattern

1) Forecasting performances generated by different models

To evaluate the effectiveness of each forecasting model, the forecasting performances generated by 11 forecasting models for seasonal data pattern are presented in Table A2 in Appendix. In this paper, the structures of Tables A2-A4 are the same with that of Table A1 except for 30 time series included. The 30 time series of seasonal pattern include monthly and quarterly sales data.

Take MASE performances as an example to describe the forecasting performances generated by different models. When MASE is used, the performances for 30 time series can be summarized in Table 3. In this table, the second and the third rows shows the minimal and the maximal MASE values generated by each forecasting model for 30 time series respectively; the 4-6 rows show the number of time series for which the MASE values generated by corresponding models are greater than 0.5, 1 and 2 respectively. Take the results generated by AR(2) model as an example. For 30 time series, the minimal and maximal MASEs are 0.33 and 1.64 respectively. In addition, MASEs of 8 series are greater than 1 but MASEs of all series are less than 2. For the results generated by NN(2) model, the minimal and maximal MASEs are 0.00 and 2.80 respectively whereas MASEs of 6 series are greater than 1 and MASEs of 2 series are greater than 2. It can be easily found from this table that two MA models generate the worst forecasts in all these models for these time series of seasonal pattern. Some results generated by the two NN models are very good (almost zero) while some results are not ideal because NN models are prone to over-fitting.

Table 3. Summary of forecasting performances in terms of MASE

	Naïve	AR(2)	AR(3)	MA(2)	MA(3)	ARMA(1,1)	ARMA(1,2)	DES	TES	NN(2)	NN(3)
Min.	0.47	0.33	0.15	0.65	0.69	0.29	0.42	0.52	0.54	0.00	0.00
Max.	1.87	1.64	1.69	2.44	2.99	1.76	1.74	1.63	1.87	2.80	3.12
>0.5	29	27	18	30	30	27	29	30	30	20	20
>1	17	8	7	25	20	10	10	11	11	6	9
>2	0	0	0	1	6	0	0	0	0	2	1

2) Forecasting performance comparisons in terms of different accuracy measures

If more than one out-of-samples are forecasted, different comparison results can be used when different accuracy measures are used to evaluate forecasting accuracy. Tables 4-7 show the performance comparison results of the 11 forecasting models when using MAE, MAPE, RMSE and MASE, respectively, as forecasting accuracy measures.

Table 4 shows the comparison results generated by different forecasting models in terms of MAE: (i) NN(2) model provides better forecasts although it also generates the worst forecasts for one time series; (ii) Naïve model, two MA models and two ARMA models perform poorly which cannot generate best forecast for even one time series. (iii) AR(3) model generates obviously better forecasts than AR(2) does.

Table 4. The number of time series for which the forecasting model generates the best forecasting performance (MAE)

Performance Ranking	1	2	3	4	5	6	7	8	9
Naïve	0	1	2	0	2	1	5	5	7
AR(2)	2	5	5	5	4	4	1	2	1
AR(3)	8	3	6	5	6	1	1	0	0
MA(2)	0	0	0	0	0	0	0	3	9
MA(3)	0	0	0	0	0	1	1	3	2
ARMA(1,1)	0	1	3	9	3	4	6	3	1
ARMA(1,2)	0	1	4	4	2	6	6	5	2
DES	2	5	1	5	4	7	3	3	0
TES	1	1	5	1	4	4	5	3	4
NN(2)	10	6	0	1	4	1	2	1	1
NN(3)	7	7	4	0	1	1	0	2	3

Table 5 shows the comparison results generated by different forecasting models in terms of MAPE: (i) NN(2) model is still the best and two MA models are still the worst. (ii) Unlike those in Table 4, Naïve model and two ARMA models perform better than two ES models. (iii) AR(3) model generates slightly better forecasts than AR(2) does.

Table 5. The number of time series for which the forecasting model generates the best forecasting performance (MAPE)

Performance Ranking	1	2	3	4	5	6	7	8	9
Naïve	1	3	1	3	4	2	1	3	5
AR(2)	1	3	9	5	9	2	1	0	0
AR(3)	4	4	7	4	4	5	2	0	0
MA(2)	0	0	0	1	0	0	3	7	6
MA(3)	0	0	0	0	0	0	2	7	3
ARMA(1,1)	3	6	4	9	3	1	0	2	0
ARMA(1,2)	4	4	0	2	1	3	5	5	4
DES	1	0	3	1	5	6	5	4	5
TES	0	0	0	3	1	3	8	1	3
NN(2)	10	5	1	0	3	3	1	1	3
NN(3)	6	5	5	2	0	5	2	0	1

It can be clearly found from Table 6 that the comparison results generated by MASE is completely the same with those shown in Table 4. That is, MASE and MAE generate the same comparison results for the 30 time series of seasonal pattern investigated. For other data patterns, the comparison results generated by MASE and MAE are also the same. We thus do not present the results generated by MASE in the rest of this paper.

Table 6. The number of time series for which the forecasting model generates the best forecasting performance (MASE)

Performance Ranking	1	2	3	4	5	6	7	8	9
Naïve	0	1	2	0	2	1	5	5	7
AR(2)	2	5	5	5	4	4	1	2	1
AR(3)	8	3	6	5	6	1	1	0	0
MA(2)	0	0	0	0	0	0	0	3	9
MA(3)	0	0	0	0	0	1	1	3	2
ARMA(1,1)	0	1	3	9	3	4	6	3	1
ARMA(1,2)	0	1	4	4	2	6	6	5	2
DES	2	5	1	5	4	7	3	3	0
TES	1	1	5	1	4	4	5	3	4
NN(2)	10	6	0	1	4	1	2	1	1
NN(3)	7	7	4	0	1	1	0	2	3

When RMSE is used as accuracy measure, the comparison results are shown in Table 7, which are closer to those generated by MAE than those generated by MAPE: (i) NN(2) model cannot show superiority when it is compared with AR(3) model. (ii) Naïve model, two MA models and two ARMA models perform poorly which cannot generate best forecast for even one time series. (iii) AR(3) model generates obviously better forecasts than AR(2) does.

Table 7. The number of time series for which the forecasting model generates the best forecasting performance (RMSE)

Performance Ranking	1	2	3	4	5	6	7	8	9
Naïve	0	0	1	1	1	2	4	7	7
AR(2)	1	6	5	7	5	1	2	2	1
AR(3)	10	4	9	5	2	0	0	0	0
MA(2)	0	0	0	0	0	0	0	3	6
MA(3)	0	0	0	0	0	2	2	2	7
ARMA(1,1)	0	2	2	5	5	9	4	2	1
ARMA(1,2)	0	1	4	4	8	5	4	3	0
DES	2	3	3	4	6	6	1	2	3
TES	1	2	2	2	2	2	10	4	2
NN(2)	9	6	1	1	0	2	1	5	0
NN(3)	7	6	3	1	1	1	2	0	3

Due to page limit, the results for irregular pattern and random pattern will not be detailed in this paper. From the experimental results described above, the performance comparison of different univariate forecasting techniques can be summarized in Table 8. Each value (number) in this table represent the performance ranking of the corresponding forecasting model among all forecasting models used for specific data pattern and accuracy measure. For example, the number '8' in the second row and the third column represents that, for all trend time series, the forecasting performance generated by Naïve model ranks eighth when MAE is used as testing accuracy measure. In addition, in each row, the cells with green background indicate several corresponding techniques generate best results while the cells with olive green background indicate corresponding techniques generate worst results.

Table 8. Summary of performance comparison

		Naïve	AR(2)	AR(3)	MA(2)	MA(3)	ARMA (1,1)	ARMA (1,2)	DES	TES	NN(2)	NN(3)
Trend	MAE	8	2	3	9	10	5	/	1	4	6	7
	MAPE	8	2	3	9	10	5	/	1	4	6	7
	MASE	8	2	3	9	10	5	/	1	4	6	7
	RMSE	8	2	3	9	10	5	/	1	4	6	7
Seasonal	MAE	9	4	2	11	10	8	7	5	6	1	3
	MAPE	7	6	3	10	11	5	4	8	9	1	2
	MASE	9	4	2	11	10	8	7	5	6	1	3
	RMSE	9	5	1	11	10	7	8	4	6	2	3
Irregular	MAE	2	5	7	10	11	9	3	6	8	1	4
	MAPE	5	1	4	10	11	7	2	8	6	3	9
	MASE	2	5	7	10	11	9	3	6	8	1	4
	RMSE	6	1	5	10	11	7	2	9	8	3	4
Random	MAE	7	1	4	6	9	2		8	5	3	10
	MAPE	7	2	3	4	8	1		9	5	6	10
	MASE	7	1	4	6	9	2		8	5	3	10
	RMSE	9	7	3	5	6	1		8	4	2	10

Best results   Worst results

In summary, the following conclusions can be drawn:

(1) For different data patterns, the forecasting performance generated by different forecasting models are mixed.

(i) For trend data pattern, the forecasting results generated by the AR, ARMA, ES and NN models are acceptable in retailing practice. Among these models, Naïve and MA models generate the worst forecasts while AR and ES models generate the best;

(ii) For seasonal data pattern, the forecasting results generated by the Naïve and MA models are unacceptable in retailing practice. The results generated by the ARMA and ES models are acceptable but they perform not very well while the AR and NN models generate the obviously better results;

(iii) For irregular data pattern, among these models, the Naïve model generates better results than MA. In addition, among the models used, no one can generate obviously better results than others;

(iv) For random data pattern, ARMA(1,1) and AR(2) models generate slightly better results than others. NN(3) model generates the worst performance.

It is clear that MA always generates worse forecasting results whichever data pattern is consider. In addition, NN models cannot exhibit obviously better performances than other traditional models.

(2) Even for the same model, different numbers of input variables could have large effects on forecasting results. For instance, for seasonal data pattern, AR(3) generates obviously better results than AR(2); However, for irregular data pattern, AR(2) generates better results than AR(3).

(3) Different accuracy measures affects forecasting results largely. Take irregular data pattern as an example, Naïve and AR(2) generate similar forecasts when MAE is used as accuracy measure; however, AR(2) generates much better forecasts when MAPE is used.

4. Conclusions

This research aimed at addressing the performance comparison of several commonly used univariate forecasting techniques for apparel sales forecasting. A large number of apparel sales time series are used, which are categorized as trend, seasonal, irregular and random data patterns. This research also investigated the effects of different numbers of input variables and different accuracy measures on sales forecasting performances. The comparison presented in this paper can provide a theoretical basis for forecasting researchers and practitioners, and help them select the appropriate forecasting models or benchmark models for different apparel sales forecasting tasks.

Further research will compare the forecasting performances of different univariate and multivariate forecasting techniques for apparel sales data in terms of different sales data patterns and accuracy measures.

References

[1] J. B. Boulden, "Fitting the sales forecast to your firm," *Business Horizons*, vol. 1, pp. 65-72, 1958.

[2] G. Lancaster and P. Reynolds, *Marketing: Made simple*. Oxford, U.K.: Elsevier, 2002.

[3] T. Xiao and D. Yang, "Price and service competition of supply chains with risk-averse retailers under demand uncertainty," *International Journal of Production Economics*, vol. 114, pp. 187-200, 2008.

[4] E. Bayraktar, S. Koh, A. Gunasekaran, K. Sari, and E. Tatoglu, "The role of forecasting on bullwhip effect for E-SCM applications," *International Journal of Production Economics*, vol. 113, pp. 193-204, 2008.

[5] X. Zhao, J. Xie, and J. Leung, "The impact of forecasting model selection on the value of information sharing in a supply chain," *European Journal of Operational Research*, vol. 142, pp. 321-344, 2002.

[6] A. GarciaFerrer, J. DelHoyo, and A. MartinArroyo, "Univariate forecasting comparisons: The case of the Spanish automobile industry," *Journal of Forecasting*, vol. 16, pp. 1-17, 1997.

[7] K. Au, T. Choi, and Y. Yu, "Fashion retail forecasting by evolutionary neural networks," *International Journal of Production Economics*, vol. 114, pp. 615-630, 2008.

[8] Z. Sun, T. Choi, K. Au, and Y. Yu, "Sales forecasting using extreme learning machine with applications in fashion retailing," *Decision Support Systems*, vol. 46, pp. 411-419, 2008.

[9] S. Thomassey and A. Fiordaliso, "A hybrid sales forecasting system based on clustering and decision trees," *Decision Support Systems*, vol. 42, pp. 408-421, 2006.

[10] S. Thomassey and M. Happiette, "A neural clustering and classification system for sales forecasting of new apparel items," *Applied Soft Computing*, vol. 7, pp. 1177-1187, 2007.

[11] C. A. Witt and S. F. Witt, "Forecasting International Tourist Flows," *Annals of Tourism Research*, vol. 21, pp. 612-628, 1994.

[12] J. De Gooijer and R. Hyndman, "25 years of time series forecasting," *International Journal of Forecasting*, vol. 22, pp. 443-473, 2006.

[13] R. Hyndman and A. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, pp. 679-688, 2006.