"A study of accuracy assessment in time series sales forecasting models for seasonal food products"

Saurabh Gupta[*], Nishant Kumar[#]

* Research Scholar, Department of Business Administration, University of Lucknow, Lucknow, UP, India.

Assistant Professor, Department of Business Administration, University of Lucknow, Lucknow, UP, India

Email: saurabhgupta2dams@gmail.com

ABSTRACT

Sales forecasting is one of the most common phenomena observed in industry, as it assists other subsidiary department of the industry such as finance, human resources, marketing, supply chain etc. Although forecasted values are obtained through several qualitative and quantitative methods, each method has its own pros and cons. The selection of these models depends upon the knowledge, availability of data and context of forecasting. The purpose of this research paper is to check the accuracy of forecasting methods based on the use of time-series analysis for seasonal food products. Effort has been made to concentrate on **time-series forecasting methods** which are applicable quite generally. The scope of the subject is wide and the techniques chosen reflect particular interests and concerns.

Methodology: Historical sales data have been used to check the forecast accuracy. In order to identify an acceptable forecast method based on a check on its accuracy, ex post indicators are used such as **forecast error, mean absolute percentage error and mean percentage error**.

Findings: Among these models Autoregressive Integrated Moving Average (ARIMA) model performs better than other competing models in forecasting sales of seasonal food products. It provides lowest MAPE, MPE value.

Keywords: Sales Forecasting, Time Series Analysis, Forecast accuracy.

I. INTRODUCTION

Due to the strong competition that exists today, most manufacturing organizations give more consideration on continuous effort for increasing their profits and reducing their costs. Accurate sales forecasting is certainly helps the organization to meet the aforementioned goals, since this leads to improved customer service, reduced lost sales and product returns and more efficient production planning.

An efficient forecasting system can improve machine utilization, reduce inventories, achieve greater flexibility to changes and increase profits. Sales forecasting is very important, as its outcome is used by many functions in the organization. Finance and accounting departments are able to project cost, profit levels and capital needs based on a sales forecast. The sales department requires a good knowledge of sales volume of each product, in order to organize the job of sales force. Production needs a long-term forecast for planning the development of the plant and equipment and a more detailed short-term forecast for arranging the production plan. Marketing needs a view of the future market in order to plan its actions and assess the impact of changes in the marketing strategy on sales volumes. Finally, logistics also needs accurate sales forecasts of different horizon lengths: a long-term forecast in order to develop and organize logistics infrastructure and a short-term forecast to define specific logistics needs.

A perfect forecast is usually impossible because too many factors in the business environment cannot be predicted with certainty. Therefore, rather than search for the perfect forecast, it is far more important to establish the practice of continual review of forecasts and to learn to live with inaccurate forecasts. Because forecasts based on past data, it is not possible to predict the future cent percent accurate. Forecast accuracy decreases as time horizon increases. The accuracy of the forecast and its costs are interrelated. In general, the higher the need for accuracy translates to higher costs of developing forecasting models.

The products which show seasonal fluctuation due to the essence of its use on particular point of time are known as seasonal products. Some examples of seasonal food products are: soft drinks, ice-creams, energy drinks, mosquito repellent, coffee, etc. The big names in the food segments include **Nestle, Amul, Unilever, Mother dairy, Coca-Cola, Pepsi,** etc. As we know that these seasonal products have high demand in the peak period and low demand/ no demand in the off period, that’s why it is very necessary to forecast accurately the future expected sale of that kind of products in order to fulfil the demand of customers as well as reducing cost of inventory handling and customers' lost.

Food companies who offered seasonal products are more concerned with sales forecasting due to their special characteristics, such as the short shelf-life of their products, they need to maintain high product quality and the uncertainty and fluctuations in consumer demands. As products can only be sold for a limited period of time, both shortage and surplus of goods can lead to loss of income for the company. The variations in consumer demand are caused by many factors other than climate changes like price, promotions, competitor's activity, consumer's income, changing consumer preferences etc.

II. Review of Literature

The methodologies that have been used in sales forecasting are typically time series algorithms that can be classified as linear or nonlinear, depending on the nature of the model they are based on. Linear models, like simple and double moving average, single exponential smoothing, holt exponential smoothing, winter's exponential smoothing, decomposition method, autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) (**Box, Jenkins, & Reinsel, 1994**) are the popular methodologies, but their forecasting ability is limited by their assumption of a linear behaviour and thus, it is not always satisfactory (**Zhang, 2003**). In order to address possible nonlinearities in time series modelling, researchers introduced a number of nonlinear methodologies, including nonlinear ARMA time series models. Their main drawback is that the type of nonlinearity is not known in advance and the modeller needs to select the structure of the model by trial and error.

Obviously, the key question concerns the accuracy of each modelling method. To this end, a number of studies have been conducted to compare the aforementioned methods; the results are not clearly in favour of one particular method. **Zhang (2003)** pointed out that no single method is best in every situation and that combining different models is an effective and efficient way to improve forecasting accuracy.

For forecasting productivity, demand and sales of different products there are ample evidence of applications of various time series techniques. **While Ghosh (2008)** applied univariate time-series techniques such as **Multiplicative Seasonal Autoregressive Integrated Moving Average (MSARIMA)** and **Holt-Winters Multiplicative Exponential Smoothing (ES)** for seasonally unadjusted monthly data spanning from April 2000 to February 2007 to forecast the monthly peak demand of electricity in the northern region of India. **Mandal (2005)** have applied **ARIMA** model for forecasting sugarcane productions in India with annual data from 1950-51 to 2002-03.**Wankhade et. al. (2010)** have considered **ARIMA** model for forecasting pigeon pea production in India with annual data from 1950-51 to 2007-08. Using univariate time series analysis, **Rothman (1999), Johnes (1999), Proietti (2001), Gil-Alana (2001),** have forecasted the employment rate in US and UK.

III. PROBLEM OF THE STUDY

Seasonal products are such a short life span type of product, the company wishes to have a good forecasting that could be used to support decision making in the operational level of supply chain. The tolerance of excess stocks for this type of product is smaller than for a non-perishable product. Besides the obvious risk of financial lost that involved due to an overstock condition, a problem related to an environmental issue may be raised. Meanwhile, out of stock condition has never been an option in the company's policy for any type of its product at all. Especially in a new development stage, out of stock would harm the product penetration both to the customers and consumers.

However, to have a good forecasting is not an easy task. The uncertainty that is always involved in forecasting, for instance people's behaviour in refrigerated prepared food consumption, competitors' activity, price elasticity, increases the difficulties in forecast these seasonal products.

IV. OBJECTIVE OF THE STUDY

Therefore, the aim of this research is to develop a forecasting concept that is able to provide a realistic prediction of the future and valuable for users since it supports them in decision making, in particular within the supply chain area. Though many forecasting theories have been developed, however, it is realized that every forecasting method has trade-offs. The first trade-off occurs in accuracy of the method and effort to perform it (**Makridakis and Wheelwright, 1989**). The more sophisticated the method used, it might provide better accuracy, but it might involve more cost. Another important trade-off occurs in model generality. A more detailed model, with additional forecasting parameters, may provide better forecasts in sample. However, these models are usually less able to generalize out of sample, and to new and unexpected circumstances. Therefore, this study should also take into account these trade-offs, so that the result will be realistic to be applied in the company.

V. TIME SERIES FORECASTING METHODS

A time series is a set of observations measured sequentially through time. These measurements may be made continuously through time or be taken at a discrete set of time points. By convention, these two types of series are called continuous and discrete time series, respectively, even though the measured variable may be discrete or continuous in either case.

COMPONENTS OF TIME SERIES

The pattern or behaviour of data in a time series has several components:

The **trend** component accounts for the gradual shifting of the time series to relatively higher or lower values over a long period of time. The **seasonal** component accounts for regular patterns of variability within certain time periods, such as a year. Any regular pattern of sequences of values above and below the trend line lasting more than one year can be attributed to the **cyclical** component. The **irregular** component is caused by short-term, unanticipated and non-recurring factors that affect the values of the time series.

In general, the **seasonal factor** for any period of a year (a quarter, a month, etc.) measures how that period compares to the overall average for an entire year. Specifically, using historical data, the seasonal factor is calculated to be:

Seasonal factor = average for the period / overall average

Seasonally adjusted value = actual value / seasonal factor

The available traditional time series forecasting approaches are divided into two groups i.e. the **univariate** time series model and **multivariate** time series model. One of the major limitations of traditional statistical methods is that they are essentially linear methods.

1. Simple and Double Moving Average

The method of moving average eliminates randomness by taking a set of observed values, finding their average and then using that average as a forecast for the coming period. The term moving average is used because as each new observation becomes available, a new average can be computed and used as a forecast. To determine the appropriate periods of moving average, it is useful to perform forecast by using different average periods and then compute the forecasts errors of each forecast.

Double Moving Average method calculates a second moving average from the original moving average which has been presented previously as an attempt to eliminate systematic error. Simple and Double Moving average methods are appropriate to handle a horizontal data series, but these techniques may be ineffective in handling data series which involves trends and seasonality patterns.

2. Single Exponential Smoothing

A strong argument can be made that since the most recent observations contain the most current information about what will happen in the future, thus they should be given relatively more weight than the older observations. Exponential smoothing satisfies this requirement and eliminates the need for storing the historical values of the variable likewise in the moving average method. This method is appropriate in handling data series that contains a horizontal pattern. However, this technique may not be effective in handling trends and seasonal patterns.

3. Linear (Holt's) Exponential Smoothing

If a single exponential smoothing is used with a data series that contains a consistent trend, the forecast will trail behind (lag) that trend. In this case, the linear (Holt's) Exponential smoothing performs well in handling a consistent trend in data series.

4. Winter's Linear and Seasonal Exponential Smoothing

The Winter's Exponential Smoothing involves three parameters. Besides parameters for smoothing the series and trend factor that have been mentioned in the Holt's exponential smoothing, another parameter incorporated in this model is a parameter to smooth the seasonality index.

5. Decomposition Method

Decomposition methods identify three separates components of the basic underlying pattern that characterize series. These are the trend, cycle and seasonal factors. There are several approaches to decomposing a time series, all of which aim to isolate each component of the series as accurately as possible. The basic concept in such separation is empirical and consists of removing firstly the seasonality, then trend, and finally cycle.

6. ARIMA (Auto Regressive, Integrated and Moving Average)

ARIMA model, which also called Box-Jenkins method, has three components that are auto regressive, integrated and moving average. The purpose of ARIMA is to find a model that accurately represents the past and future patterns of a time series where the pattern can be random, seasonal, trend, cyclical, promotional or a combination of patterns until the errors are distributed as white noise. ARIMA models resemble other univariate forecasting method because they include trend, seasonal, and random components.

VI. RESEARCH METHODOLOGY

The purpose of this research study is to find a suitable model for seasonal food products to forecast the expected future sale of goods. Various univariate time series forecasting models for forecasting seasonal food product sales have been applied in this paper. It compares the out-of-sample forecast accuracy of different models using forecast error, mean absolute percentage error, and mean percentage error. Quarterly sales data of Coca-Cola in hectare litre since 2003 to 2012 are used to check forecast accuracy. The statistical tool **SPSS 20** and Microsoft Excel 2007 Add-Ins **Mega Stat** is used for computation. The forecasted error estimate results are shown in **Table 1** & **Table 2** for Coca-Cola data. The descriptive statistics of the sales data is displayed in **Table 3** for Coca-Cola in **Appendix**. The normal distribution curve is displayed in **Figure 1** for Coca-Cola in **Appendix**.

VII. FINDINGS AND ANALYSIS

In this research paper the Time series Forecasting Methods have been studied which require the past years sales data for computation. There are some other factors also which affect the sale of the product but as per the requirement of the study the focus is given on the previous year sales data. In order to identify the most accurate forecast method, quite often **ex post indicators** are used. The main **ex post indicators** defining the forecast accuracy are as follows **(Boguslauskas, 2007)**:

1. **Forecast error** e_t, which is calculated as a difference between actual economic indicator value y_t and forecast economic indicator value :

$$e_t = y_t - \hat{y}_t$$

2. **Mean absolute percentage error** (MAPE) that reflects the relative forecast accuracy, is calculated as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \frac{|e_t|}{y_t} \times 100\%$$

3. **Mean percentage error** (MPE) demonstrates the forecast deviation and is calculated in the following way:

$$\text{MPE} = \frac{1}{n} \sum_{t=1}^{n} \frac{e_t}{y_t} \times 100\%$$

Forecast error value demonstrates the smallest difference between sales forecast and fact; mean absolute percentage error discloses high forecast accuracy; and mean percentage error describes a small positive forecast deviation.

Forecast percentage error estimate results

Table 1

Time Period	Type of error	Simple Moving Average Method where k=3	Simple Moving Average Method where k=5	Double moving average where k=3	Double moving average where k=5	ARIMA Method
Q1(2010)	Forecast error (e_t)	-214.29	-170.97	-271.64	-53.80	-84.56
	MAPE	7.014293	5.59646	8.42027	1.66782	2.42859
	MPE	-7.014293	-5.59646	-8.42027	-1.66782	-2.42859
Q2(2010)	Forecast error (e_t)	490.38	438.63	522.38	664.67	-51.48
	MAPE	12.60617	11.27578	13.42879	17.08654	1.323393
	MPE	12.60617	11.27578	13.42879	17.08654	-1.323393

Q3(2010)	**Forecast error (e_t)**	98.00	158.83	227.03	323.74	-239.61
	MAPE	2.702703	4.380254	6.2612	8.928232	6.608108
	MPE	2.702703	4.380254	6.2612	8.928232	-6.608108
Q4(2010)	**Forecast error (e_t)**	-261.67	-68.17	-148.43	16.95	-147.26
	MAPE	7.759984	2.02171	4.40181	0.502609	4.367141
	MPE	-7.759984	-2.02171	-4.40181	0.502609	-4.367141
Q1(2011)	**Forecast error (e_t)**	-100.33	-110.80	-186.67	-58.30	-78.29
	MAPE	2.994131	3.30648	5.57048	1.73970	2.336318
	MPE	-2.994131	-3.30648	-5.57048	-1.73970	-2.336318
Q2(2011)	**Forecast error (e_t)**	652.00	622.00	749.67	833.10	-67.24
	MAPE	15.01958	14.3285	17.26945	19.19136	1.548952
	MPE	15.01958	14.3285	17.26945	19.19136	-1.548952
Q3(2011)	**Forecast error (e_t)**	391.33	611.60	712.33	859.09	218.37
	MAPE	8.807863	13.76547	16.03271	19.33584	4.9149223
	MPE	8.807863	13.76547	16.03271	19.33584	4.9149223
Q4(2011)	**Forecast error (e_t)**	-266.33	98.00	2.33	334.73	-39.13

	MAPE	6.646702	2.44572	0.058231	8.353521	0.976541
	MPE	-6.646702	2.44572	0.058231	8.353521	-0.976541
Q1(2012)	**Forecast error (e_t)**	-256.67	-151.20	-291.22	68.72	24.43
	MAPE	6.659747	3.923197	7.55636	1.783083	0.6338869
	MPE	-6.659747	-3.9232	-7.55636	1.783083	0.6338869
Q2(2012)	**Forecast error (e_t)**	677.00	624.00	728.33	988.68	127.66
	MAPE	13.68783	12.61626	14.7257	19.98949	2.5810756
	MPE	13.68783	12.61626	14.7257	19.98949	2.5810756
Q3(2012)	**Forecast error (e_t)**	331.33	460.40	579.22	792.96	6.09
	MAPE	6.771578	9.40936	11.83777	16.20601	0.1244635
	MPE	6.771578	9.40936	11.83777	16.20601	0.1244635
Q4(2012)	**Forecast error (e_t)**	-409.33	-95.00	-205.33	96.84	-3.74
	MAPE	9.492888	2.20315	4.7619	2.245826	0.086735
	MPE	-9.492888	-2.20315	-4.7619	2.245826	-0.086735

Table 2

Time Period	Type of error	Exponential Smoothing Method where α=0.3	Exponential Smoothing Method where α=0.5	Holt Exponential smoothing where α=0.3,β=0.5	Winter's Exponential smoothing	Decomposition method
Q1(2010)	Forecast error(e_t)	-141.57	-244.01	-473.73	-426.86	-340.53
	MAPE	4.55384176	7.849	14.3597625	12.93888	9.780253311
	MPE	-4.55384176	-7.849	-14.3597625	-12.93888	-9.780253311
Q2(2010)	Forecast error(e_t)	735.60	712.49	483.92	330.81	410.87
	MAPE	18.91002571	18.31594	12.44010283	8.5040716	10.56209794
	MPE	18.91002571	18.31594	12.44010283	8.5040716	10.56209794
Q3(2010)	Forecast error(e_t)	248.22	87.75	-21.06	-26.03	261.16
	MAPE	6.845559846	2.420022	0.5808053	0.718	7.202428045
	MPE	6.845559846	2.420022	-0.5808053	-0.718	7.202428045
Q4(2010)	Forecast error(e_t)	-81.15	-211.63	-360.04	-248.11	-103.20
	MAPE	2.40658363	6.2761	10.6773428	7.3580288	3.060499941
	MPE	-2.40658363	-6.2761	-10.6773428	-7.358029	-3.060499941
Q1(2011)	Forecast error(e_t)	-78.10	-127.31	-309.87	-261.24	-407.87
	MAPE	2.33064757	3.79916	9.24709042	7.795873	12.17159198
	MPE	-2.33064757	-3.79916	-9.24709042	-7.795873	-12.17159198

Q2(2011)	**Forecast error(e_t)**	935.03	925.84	782.58	622.62	470.29
	MAPE	21.53950703	21.3278	18.0276434	14.342786	10.83358738
	MPE	21.53950703	21.3278	18.0276434	14.342786	10.83358738
Q3(2011)	**Forecast error(e_t)**	756.22	564.42	541.76	536.44	383.36
	MAPE	17.02048166	12.70358	12.19356291	12.073793	8.628318859
	MPE	17.02048166	12.70358	12.19356291	12.073793	8.628318859
Q4(2011)	**Forecast error(e_t)**	87.95	-162.79	-253.88	-126.09	-15.46
	MAPE	2.194908909	4.06264	6.33591215	3.146802	0.385813594
	MPE	2.194908909	-4.06264	-6.33591215	-3.146802	-0.385813594
Q1(2012)	**Forecast error(e_t)**	-94.43	-239.39	-488.85	-431.16	-363.75
	MAPE	2.45018163	6.21147	12.6842242	11.18735	9.438303299
	MPE	-2.45018163	-6.21147	-12.6842242	-11.18735	-9.438303299
Q2(2012)	**Forecast error(e_t)**	1,025.90	972.30	668.00	475.69	598.79
	MAPE	20.74201375	19.65831	13.50586332	9.6177407	12.10664898
	MPE	20.74201375	19.65831	13.50586332	9.6177407	12.10664898
Q3(2012)	**Forecast error(e_t)**	668.13	438.15	237.10	230.75	444.11
	MAPE	13.654813	8.954629	4.845697936	4.7159055	9.076508281

	MPE	13.654813	8.954629	4.845697936	4.7159055	9.076508281
Q4(2012)	**Forecast error(e_t)**	-113.91	-362.92	-632.00	-483.72	-71.28
	MAPE	2.641697588	8.416512	14.6567718	11.21811	1.652951183
	MPE	-2.64169759	-8.41651	-14.6567718	-11.21811	-1.652951183

From the analysis of the given data set it is found that the **ARIMA** Method is best suitable model for forecasting the sales of seasonal food products. **The result have shown that the values of forecast error, MAPE, MPE for the period 2010-2012 for Coca-Cola data are lowest in ARIMA Method.**

In this study it is found that using Single and Double moving average method there is a problem of determining the optimal number of periods. Single exponential smoothing method and Linear (Holt's) exponential method do not model the seasonality of a series. Winter's Linear and Exponential method requires a large number of data and also this method might be too complex for data that do not have identifiable trends and seasonality. In Decomposition method it is very difficult to simultaneously decompose trend and seasonality in a series when only a few seasonal cycles exist. ARIMA model is capable of describing all patterns involved in a series which may be revealed through further statistical analysis.

VIII. CONCLUSION

Seasonal products are such a short life span type of product. The tolerance of excess stocks for this type of product is smaller than for a non-perishable product. These food products show frequent changes in sales in different seasons. Therefore it is very necessary to forecast the sale of these products accurately. Forecasting accurate seasonal food products sales is extremely complicated and requires lots of coordination from both IT systems and human interaction. But with the help of the use of ARIMA model it is possible to forecast the sales for seasonal food products accurately because ARIMA model included the trend, seasonal and random components.

REFRENCES

1. Armstrong, J. Scott (2001), Principles of Forecasting, a handbook for researchers and practitioners, Kluwer Academic Publisher, USA.
2. Box, G.E.P., Jenkins, G.M & Reinsel, G.C. (1994), Time series analysis: Forecasting and control (3rd Ed.) Englewood cliffs, NJ: Prentice Hall.
3. Fildes, R., Hibon, M., Makridakis, S., Meade, N., 1998. Generalising about univariate forecasting methods: Further empirical evidence, International Journal of Forecasting 14, 339-358.
4. Gardner, E.S., Jr., McKenzie, E., 1985. Forecasting trends in time series, Management Science 31, 1237-1246.
5. Ghosh, S, 2008. "Univariate Time-Series Forecasting of Monthly Peak Demand of Electricity in Northern India". International Journal of Indian Culture and Business Management, 1(4); 466 - 474.
6. Harvey, A.C., 1990. Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, New York.
7. Johnes, G.1999,"Forecasting Unemployment". Applied Financial Economics, 6, 605-607.
8. Lundholm, Russell J. and McVay, Sarah E, “Forecasting Sales: A model and some evidence from the retail industry.” Jan 2004.
9. Makridakis, S. Forecasting: ‘Methods and Applications’, 3rd ed. / S. Makridakis, S.C. Wheelwright, R. J. Hyndman, New York: John Wiley & Sons, 1998.
10. Mandal, B. N., 2005. "Forecasting Sugarcane Productions in India with ARIMA Model".

Appendix

Table 3

Sales of Coca-Cola

Descriptive Statistics										
	N	Minimum	Maximum	Mean		Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error	Statistic	Std. Error
Sales	40	1547.82	4936.00	2907.8955	144.05013	911.05302	.649	.374	-.531	.733
Valid N (listwise)	40									

Figure 1

Coca-Cola