Transactions on Engineering and Sciences
Vol. 2, Issue 5, May 2014
ISSN: 2347-1964 Online 2347-1875 Print

Comparing SARIMA and Holt-Winters' forecasting accuracy with respect to Indian motorcycle industry

Deepa Puthran[1] H.C.Shivaprasad[2] K.S.Keerthesh Kumar[3] M.Manjunath[4]

[1, 3&4]Research Scholar, Dept. of Humanities and Management, Manipal Institute of Technology, Manipal.
[2]Professor, Dept. of Humanities and Management, Manipal Institute of Technology, Manipal.

Abstract—**Indian automotive industry is one of the largest in the world and has been growing at a very rapid pace. The industry is dominated by motorcycles with a market share of more than 70%. The Indian motorcycle industry has direct as well as indirect influence on the growth of the Indian economy, hence understanding and forecasting the performance of this industry is very critical. The key purpose of this journal paper is to compare the accuracy of Holt-Winters and Autoregressive integrated moving average (ARIMA) model which is popularly known as Box-Jenkins auto regressive model, in relation to Indian motorcycle industry and possibly suggest the best model. Currently, there are no studies exploring the forecasting accuracy, of two models with reference to Indian motorcycle sales. In this journal paper we equate the forecasted values of both the models and we choose the best model based on the least mean square error (MSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).**

Index Terms—**Forecast, Holt-Winters, Autoregressive integrated moving average (ARIMA), Indian motorcycle, Mean absolute percentage error (MAPE), Mean absolute error (MAE), Mean square error (MSE), Society of Indian Automobile Manufacturers (SIAM).**

I. INTRODUCTION

Planning is a crucial element in any industrial activity. However, for a business to be successful it requires goals that are built on sales projections, which intern depends on accurate demand forecast. Moreover, various other business assessments are too reliant on the prediction of future sales. Therefore, sales prediction becomes very critical for all planning and development activities. Accurate forecast helps companies to uphold ideal inventory level and sustain daily operations which in turn affect the profit margins. Therefore, companies and strategists benefit a lot from accurate demand forecast.

Forecasters and statisticians use wide variety of forecasting techniques which differ in their complexity, amount of data required and ease of use [5]. Out of these forecasting techniques, judgmental technique was found to be more dominant [9], [10]. However, several studies have also shown that judgmental technique is less precise, prejudice and more likely to generate poor estimates than other methods [11], [12]. Choosing suitable forecasting methods is very important to generate reliable and accurate forecasts [16], [17]. However, each technique has its own limitation and fits only limited set of situations, thus forecaster have to choose techniques as per situations in order to get maximum accuracy. This journal paper thus efforts to weigh the two forecasting models and relate the outcomes and propose the best model for the Indian motorcycle industry.

II. RESEARCH METHODOLOGY AND DATA ANALYSIS

The Indian motorcycle industry's historic sales data was gathered from Society of Indian Automotive Manufacturers (SIAM). There has been lot of discussion on the least number of sales data needed to create an ARIMA model, most researchers have suggested that at least 50 observations are required in order to build an accurate model [13], [14]. For this study monthly sales of motorcycle from April-2006 to February-2014 are used which amounts to a total of 94 observations. The Figure 1 plots the monthly sales of motorcycles in India from April 2006 to February 2010. The analysis of the series was conducted using R 3.0.2 software and Microsoft Excel 2010.

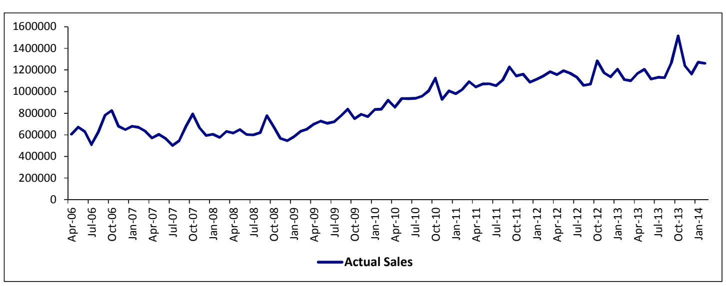

Figure 1: Indian motorcycle sales in units

The Holt-Winters forecasting technique involves an overall smoothing (Y_t), trend Smoothing (x_t) and seasonal smoothing (b_t). The relationship of these three components can be represented by the following equations:

$$Y_t= \alpha\ (y_t/b_{t-L}) + (1-\alpha)\ (Y_{t-1} + x_{t-1}) \quad (1)$$
$$B_t= \gamma(Y_t- Y_{t-1})+(1- \gamma)b_{t-1} \quad (2)$$
$$B_t=\beta\ (y_t/S_t) + (1-\beta)b_{t-L} \quad (3)$$
$$F_{t+m}= (Y_t+mx_t)_{Lt-L+m} \quad (4)$$

Where y is the number of observations, Y is the smoothed observation, x is the trend factor, b is the seasonal index, F is the forecast for the m periods ahead and t is the index denoting a time period. The appropriate Holt-Winters smoothing factors (alpha=α, beta=β and gamma=γ) for best fit model was chosen based on least MSE, MAE and MAPE values. The presence of seasonality in the series is evident from gamma value.

Table I: Holt-Winters smoothing factors

Smoothing factors			
Smoothing parameter	***Alpha (α)***	***Beta (β)***	***Gamma***
Values	0.6007	0.0242	0.4175

An Autoregressive integrated moving average (ARIMA) model is a univariate model associated three factors (p, d, & q). If F_t is the forecast of ARIMA, then the progressions for ARIMA model will be as follows:

$$F_t = a_1F_{t-1}+ a_2F_{t-2} +\ldots+ a_pF_{t-p}+\theta_0+ \theta_1b_{t-1}+ \theta_2b_{t-2}+\ldots+ \theta_qb_{t-q} \quad (5)$$

Where p is the order of the AR, q the order of the MA, d the order of differencing, b is the error term, θ_0 is a constant [7]. The time series used in this study exhibits seasonality; as a result the variant of ARIMA model used is called Seasonal ARIMA (SARIMA). SARIMA is primarily used when the time series indicates the presence seasonal variations. The SARIMA model will comprise of a seasonal autoregressive factor (P), a seasonal moving average factor (Q), seasonal differencing (D), a length of seasonal period component (s) and is represented as (p,d,q)(P,D,Q)[s] [6]. Augmented Dickey-Fuller (ADF) unit root test was used to test the time series for stationarity and then differenced accordingly [7]. Equation for Augmented Dickey-Fuller unit root test is given below:

$$\Delta y_t = a_0+ pY_{t-1}+\varepsilon_t \quad (6)$$

The table II and III gives the result of ADF test. The null hypothesis of the test is that unit root exists in the series and alternative hypothesis is no unit root exists in the data which in other words means that the series is stationary.

Table II: Summary of unit root test for original series

Unit root test values						
Unit Root test	**Test Statistics**	**Probability**	**Test critical value (1%)**	**Test critical value (5%)**	**Test critical value (10%)**	**Results**
ADF	-2.0541	0.5541	-4.04	-3.45	-3.15	Fail to reject null hypothesis

Table III: Summary of unit root test for first difference values

Unit root test values						
Unit Root test	**Test Statistics**	**Probability**	**Test critical value (1%)**	**Test critical value (5%)**	**Test critical value (10%)**	**Results**
ADF	-6.3735	0.01	-4.04	-3.45	-3.15	Reject null hypothesis

The ADF test for original series is insignificant at 1, 5 and 10 percent of significance level which results in failure to reject null hypothesis. Differencing is a technique used by many ARIMA forecasters to make the series stationary. The ADF test after first differencing shows that the data is significant at all the levels of significance and hence leads to rejection of null hypothesis.

The second step is to determine the p and q component of the ARIMA model. Using R 3.0.2 software, best fit model was selected on minimum value of Akaike Information Criterion (AIC) and the Schwartz

Bayesian Information Criterion (BIC). AIC is a measure of the relative quality of an arithmetical model, for a particular set of facts and figures available and proposes a relative estimate of the data lost when a particular model is used to signify the process. The AIC can be calculated using the formula:

$$AIC = N \ln(R^2) + 2n \tag{7}$$

BIC is a standard for model assortment among a determinate set of models. It is built on the probability function and it is closely linked to the Akaike information criterion (AIC). The BIC can be calculated using the formula:

$$BIC = N \ln(R^2) + n \ln(N) \tag{8}$$

In both the equations (7) & (8), N represents sample size, R^2 represents residual sum of squares and n represents the amount of regressors [7]. The model SARIMA $(2,1,1)(0,0,1)_{[12]}$ has the least AIC (2347.36) and BIC (2362.62) values as a result this model was considered for the analysis. Figure 2 compares the both best fit Holt-Winters and SARIMA models with actual values.

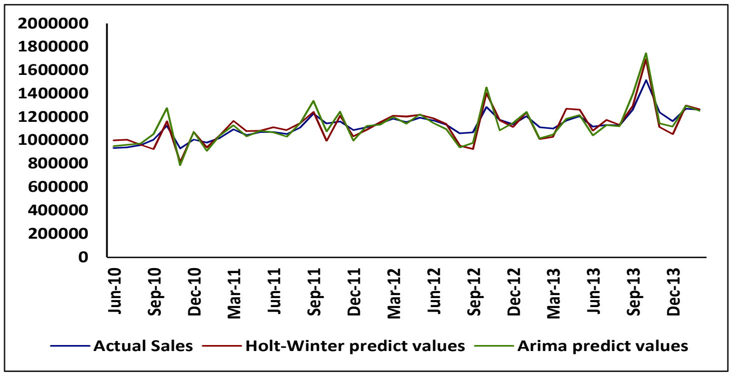

Figure 2: Compares Holt-Winters and SARIMA model vs Actual Values

A. *Error Measure*

As specified earlier, an ARIMA model needs a minimum of 50 samples, therefore for the purpose of consistency residual values from 51st sample (Jun 2010) onwards were considered for MAPE, MSE and MAE. The model's forecast accuracies were calculated, tested and compared by means of MSE, MAE and MAPE.

Table IV: Forecast Accuracy Measure

Error measure			
	MSE	*MAE*	*MAPE (%)*
Holt-	0.50	54702.4	10.8
SARIMA	0.58x	55053.1	10.87

III. CONCLUSION

This paper has compared the forecasting ability of Holt-Winters and SARIMA models with respect to their Indian motorcycle industry. The study results demonstrate that both models are pretty effective; however Holt-Winters model seems to be a more precise and accurate model. From table IV we understand that Holt-Winters model has the minimum MSE, MAE, and MAPE values when compared with SARIMA model. The Holt-Winters model's relative ease of use makes the model useful in forecasting comprehensive market trends. The study can be further enhanced by comparing other forecasting techniques with respect to Indian motorcycle sales in order to obtain better accuracy. The results will help Indian motorcycle and parts manufacturers to build effective strategy.

REFERENCES

[1] S. Makridakis, S.C. Wheelwright and V.E. McGee, Forecasting: Methods and Applications, 2nd ed., John Wiley & Sons, New York, NY, 1983.
[2] E. Mahmoud,."Accuracy in forecasting: a survey", Journal of Forecasting, Vol. 3, pp. 139-59, June 1984.
[3] J. P LeSage, "A comparison of the forecasting ability of ECM and VAR models", The Review of Economics and Statistics, Vol. 72 No. 4, pp. 664-71, November 1990.
[4] James M.W. Wong, Albert P.C. Chan, and Y.H. Chiang, "Construction manpower demand forecasting", Engineering, Construction and Architectural Management, Vol. 18 No. 1, pp. 7-29, June 2010.

[5] Syed Shahabuddin, "Forecasting automobile sales", Management Research News, Vol. 32 No. 7, pp. 670-682, 2009.

[6] Adhistya Erna Permanasari, Dayang Rohaya Awang Rambli and P. Dhanapal Durai Dominic, "Prediction of Zoonosis Incidence in Human using Seasonal Auto Regressive Integrated Moving Average (SARIMA)", International Journal of Computer Science and Information Security, Vol. 5, No. 1, 2009.

[7] Simon Stevenson, "Comparison of the forecasting ability of ARIMA models", Journal of Property Investment & Finance, Vol. 25 No. 3, pp. 223-240, 2007.

[8] Erie Febrian and Aldrin Herwany, "Volatility Forecasting Models and Market Co-Integration: A Study on South-East Asian Markets", The Indonesia Capital Market Review (ICMR), Vol.1, No. 1, 2009.

[9] N.R. Sanders, ''The status of forecasting in manufacturing firms'', Production and Inventory Management Journal, Vol. 5 No. 1, pp. 32-36, 1997.

[10] E. Mahmoud, ''Accuracy of forecasting: a survey'', Journal of Forecasting, Vol. 3 No. 2, pp. 139-51, 1984.

[11] Armstrong J.S., ''Research on forecasting; a quarter-century review, 1960-1984'', Interface, Vol. 5 No. 1, pp. 89-103, 1986.

[12] D. Bunn and G. Wright, ''Interaction of Judgemental and Statistical forecasting methods: issues and analysis'', Management Science, Vol. 37 No. 5, pp. 501-20, May 1991.

[13] T. McGough and S. Tsolacos, "Forecasting commercial rental values using ARIMA models", Journal of Property Valuation & Investment, Vol. 13, pp. 6-22, 1995.

[14] R. Y. C. Tse, "An application of the ARIMA model to real estate prices in Hong Kong", Journal of Property Finance, Vol. 8, pp. 152-63, 1997.

[15] B. L. Bowerman and R. T. O'Connell, Forecasting and Time Series: An Applied Approach, 3rd ed., Duxbury Thomson Learning, Pacific Grove, CA, 1993.

[16] B. H. Goh, "Forecasting residential construction demand in Singapore: a comparative study of the accuracy of time series, regression and artificial neural network techniques", Engineering, Construction and Architectural Management, Vol. 5 No. 3, pp. 261-75, 1998.

[17] D.Venkata Ratnam, G.Sivavaraprasad and Harshali Mane, "Forecasting of Ionospheric TEC Values Using Holt-Winters Method Over Low Latitude Station", International Journal of Electronics & Communication Technology, Vol. 5, pp. 27-30, March 2014.