International Journal of Grid and Distributed Computing
Vol. 10, No. 8 (2017), pp.37-48
http://dx.doi.org/10.14257/ijgdc.2017.10.8.04

Comparative Study on Performance Analysis of Time Series Predictive Models

Syed Muzamil Basha[1], Yang Zhenning[1], Dharmendra Singh Rajput[2],
Ronnie D. Caytiles* and N. Ch. S.N Iyengar**

[1]SCOPE/ [2]IT, VIT University, Vellore
**Multimedia Eng., Dept, Hannam University, Daejeon, Korea*
***Dept. of I.T, Sreenidhi Institute of Science and Technology., Hyderabad, India*
{[1]muza.basha,[1]yangzhenning}@gmail.com, [1]dharmendrasingh@vit.ac.in,
rdcaytiles@gmail.com, srimannarayanach@sreenidhi.edu.in

Abstract

Time series models the analyses of data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. While regression analysis is often employed in such a way as to test theories that the current values of one or more independent time. Here five-time series datasets with different range of observation are considered to study its performance. In this paper, moving averages (MA) of series with different periods to average over are calculated; plotted series for forecasted data against original data; compared the performance of HOLT-WINTERS with the Auto Regressive Integrated Moving Average (ARIMA) model with non-zero mean; and computed the statistic test to examining the null hypothesis for the considered time series datasets.

Keywords: *standard regression techniques, Moving averages, HOLT-WINTERS, ARIMA*

1. Introduction

Any ordered sequence of variable values at equally time intervals is termed as Time series. The application of time series models is to obtain an understand the structure that produced the observed data. The methods used to model and forecast time series are, Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) uni-variate models, Multivariate Models, Holt-Winters Exponential Smoothing and Simple Average moving, in which arithmetic mean of the series over the past n observations considering equal weight for both recent and current observations. In our research we would like to compare the performance of the models Holt-winter and ARIMS with respect to their p-values and X-Squared. The additive model of Holt-Winters prediction function for time period p is described as

$$X[t] = (Y[t] - Z[t-p]) + (1-a)(X[t-1] + Y[t-1]) \quad (1)$$
$$Y[t] = b(X[t] - X[t-1]) + (1-b)Y[t-1] \quad (2)$$
$$Z[t] = c(Y[t] - X[t]) + (1-c)Z[t-p] \quad (3)$$

substitute Eq. (1-3) to obtain Y as in Eq. (4).

$$\hat{Y}[t+h] = X[t] + h \times Y[t] + s[t - p + 1 + (h-1) \bmod p] \quad (4)$$

where a, b and c are used as filtering parameters. Similarly, multiplicative Holt-Winters prediction function with period length p is derived as

ISSN: 2005-4262 IJGDC

$$X[t] = a\left(\frac{Y[t]}{Z[t-p]}\right) + (1-a)(X[t-1] + Y[t-1]) \tag{5}$$

$$Y[t] = \beta(X[t] - X[t-1]) + (1-b)Y[t-1] \tag{6}$$

$$Z[t] = \gamma\left(\frac{Y[t]}{X[t]}\right) + (1-C)Z[t-p] \tag{7}$$

substitute Eq. (5-7) to obtain Y as in Eq. (8).

$$\hat{Y}[t+h] = (X[t] + h \times Y[t]) \times Z[t-p+1+(h-1) \bmod p] \tag{8}$$

The level, trend and seasonal components are to be estimated at time (t) of the series.

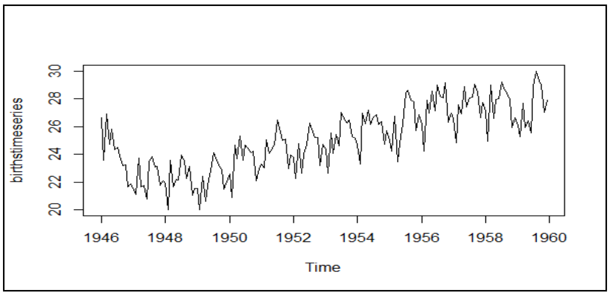

Figure 1. Time Series Plot of the Dataset

In Figure 1, the x-axis is used to plot time which is independent of nature and the y-axis is used to plot the observed data which is dependent in nature.

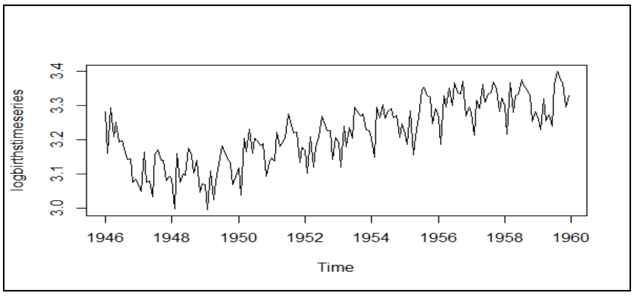

Figure 2. Log Time Series Plot of the Dataset

In Figure 2, the x-axis is used to plot the time which is independent in nature and the y-axis is used to plot the Log values of data which is dependent in nature.

Our research in this paper novel in the has considered five time series datasets with different range of observation namely Kings-time-series-forecasts(1), Births-time-series-forecasts(2), Souvenir-time-series-forecasts(3), Rain-time-series-forecasts(4), Volcanodust-time-series-forecasts(5), Calculate moving averages (SMA) of a series with different number of periods to average over, plotted the time series of forecasted data against original data, compared the performance of HOLT-WINTERS with the ARIMA models with non-zero mean and finally Computed the Ljung–Box test statistic [10] where made to examined the null hypothesis of independence for the considered time series datasets.

Our paper is organized as, Section 2 gives idea to research carried out in the field of Predictive analytics, In Section 3 describe the way the research is designed and implemented, In Section 4 Results details are given for better understanding the research and describes the performance evaluation of Time series prediction models with respect to Sum of Squared, P-value and Section 5 contains the conclusion and future scope left out in the area of predictive analytics on time series dataset.

2. Related Work

Nogales *et. al.*, [1] had proposed two techniques Dynamic regression and Transfer function, addressing serial correlation problem, using dynamic regression model that produce accurate results on hourly based electricity price data and achieved average forecasting errors on the Spanish market around 5% and around 3% in the Californian market considered in the experiment.

Hamilton *et. al.*, [2] presented an algorithm for plotted the mean growth rate of a non-stationary series occasional undergoes discrete shifts and estimated the population parameter by the method of maximum likelihood to provide the foundation for forecasting values of time series.

Beck *et. al.*, [3] considered cross related Binary time series data (BTSCS) and stated that this data is mostly widely used in International conflicts. The proposed method uses Logit analytics to address temporal dependences in event history analytics of BTSCS data.

Pivarski *et. al.*, [4] have introduced a language Portable Format for Analytics (PFA) for describing statistical and data mining models. Developed and deployed two implementations of PFA-compliant scoring engines, described two deployments into production sites of the Scala scoring engine.

Gupta *et. al.*, [5] presented an analytical framework for proactive Electrical submersible pump health monitoring system based on predictive model and made analysis that can identify real-time patterns and assess Electrical submersible pump health in real time.

Williamson *et. al.*, [6] made a survey and mapped the background of digital policy instrumentation in education, by providing two case studies of digital data systems.

Duan *et. al.*, [7] presented a complete system that combines different sources of data in to an intermediate representation in the Resource Description Framework. Tested the proposed system and obtained 66.7% and 83.3% of accuracy by using the SVM and Multiple Hidden Markov Models. Park *et. al.*, [8] introduced a method called parameter inference to estimate the performance of the likelihood function general function using Gaussian inputs, which was predictable in the process of correlating the numerical and the text data in the datasets, used to diverse the use cases for joining two datasets belongs from the continuous and discrete time series domains. Ahmed *et. al.*, [9] have performed investigation on the area time-series predictability related to commodity futures, adopted the statistical evaluation metrics to find weakness of the factor models.

3. Design and Implementation

In this section, we will discuss the approach followed to apply the forecasting techniques on time series dataset.

ARIMA model can be represented in the equation by using the parameters as an X,Y,Z, where X is count of auto-regressive terms, Y is count of non seasonal differences, Z is count of lagged forecast error in the prediction equation.

$$\begin{aligned} Z_t &= \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} \\ &+ a_t - \theta_1 a_{t-1} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \end{aligned} \tag{9}$$

where δ is mean of z_t , $t=\{1,2\ldots\}$, z_t is observed value at time t, z_1 is the expected value of z_t , $a_t= z_t- z_1$is considered as noise at the first time period.

The Box Jenkins approach is described as follows:

1. The series has to be stationary.
2. For non-stationary process, find the differences are computed to determine that the operation performed results in stationary series.
3. Step2 is repeated until a stationary series is found.
4. Determine the value of d.
5. Decide the P value less than 0.05.
6. Auto correlation function is to defined the correlation between two values of same variable at different time, which represents degree of persistence.
7. Fit the better model in to time series using values obtained by ACF.
8. Plot the residual of the model parameters obtained to verify that no lag occurs for each observation.
9. If lag occurs in residual, repeat the step 7 until we get good residual plot.
10. Calculate the forecast.

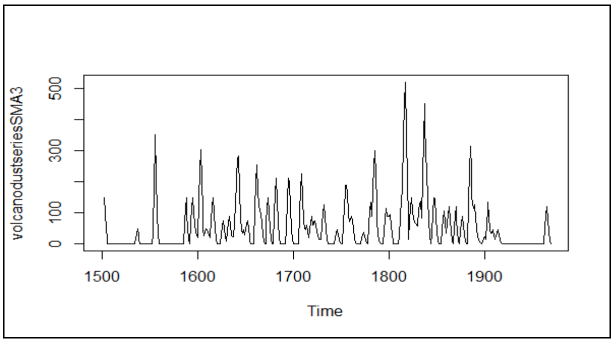

Figure 3. Time Series Plot of SMA with p=3 on 5

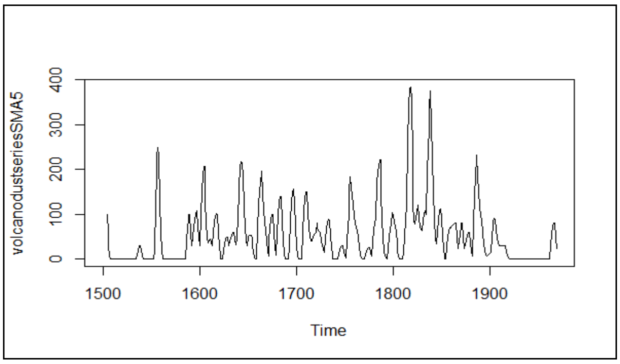

Figure 4. Time Series Plot of SMA with p=5 on 5

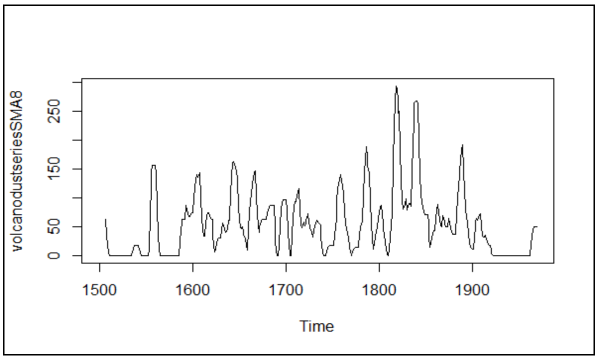

Figure 5. Time Series Plot of SMA with p=8 on 5

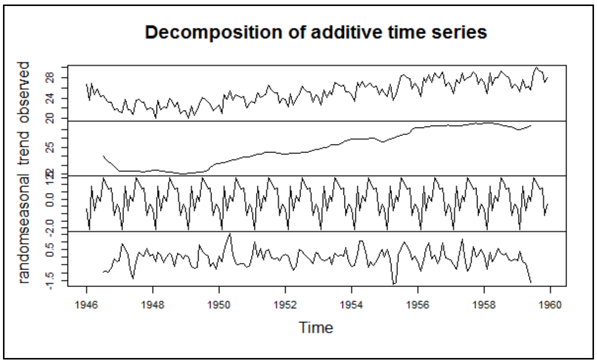

Figure 6. Decomposition of Additive Time Series on 2

Figure (3-5) represent the plot of Simple Moving Average (SMA), which is a very basic forecasting method with time period *p*. In figure 6, the components observed, trend, seasonal and random that effects the performance of time series analysis are plotted individually.

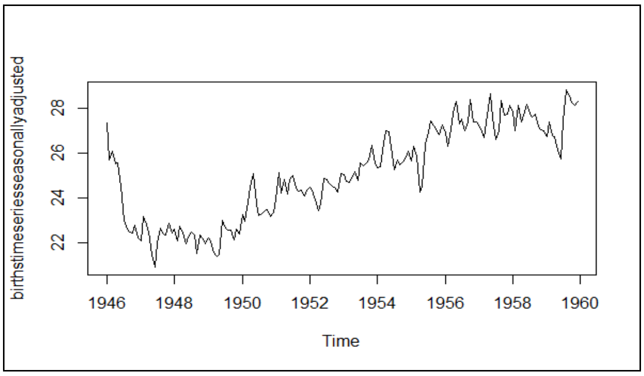

Figure 7. Adjustment of Additive Time Series on 2

$$PA = TS - TSP \tag{9}$$

where PA is parameter adjustment, obtained from the difference of total time series data with time series data with parameters. The value obtained from Eq. (9) is plotted in figure 7. The final sum of squared errors is used to predict the deviations from actual empirical values of data. A small RSS is used as in indicator to check whether the mode is tightly fitted to the model or not. It is a measure of the inconsistency between the data and an inference model. In Table 1 the Computed Holt-Winters exponential smoothing values without trend and without seasonal component are listed assuming the beta and gamma parameters as zero.

Table 1. Holt-Winters Exponential Smoothing

	(1)	(2)	(3)	(4)	(5)
Alpha	0.26893	0.4522	0.42649	0.02412	0.91711
Coefficients	67.792	27.981	65495.2	24.678	0.02483
SSE	9511.6	281.87	1445.88	1828.8	2767.44

4. Result

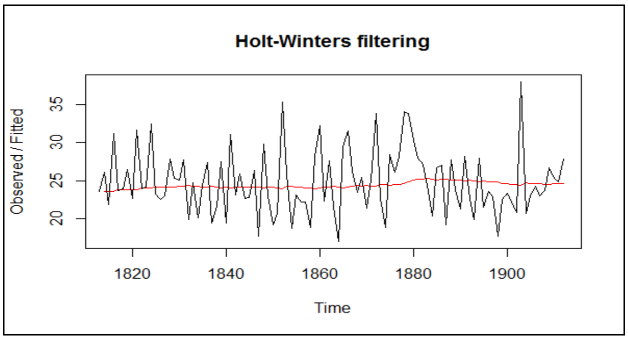

Figure 8. Plot of Original with Forecasted Data on 4

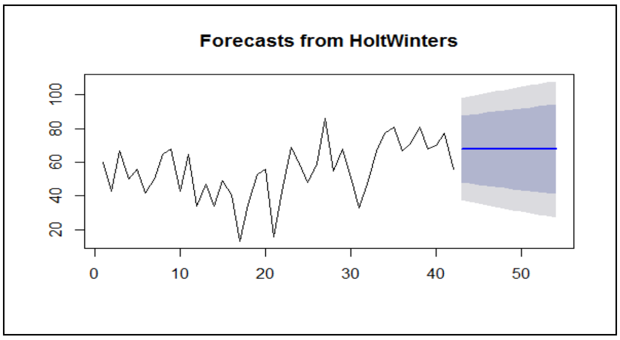

Figure 9. Plot of Original with Forecasted Data on 1 using HoltWinters

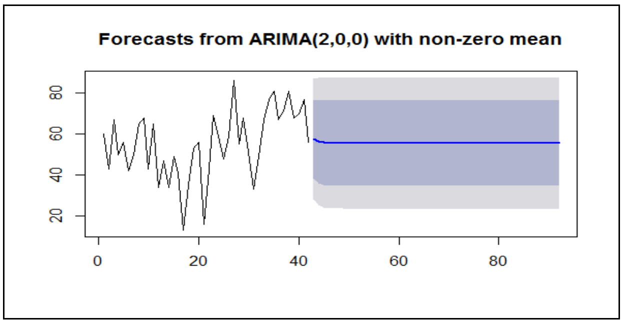

Figure 10. Plot of Original with Forecasted Data on 1 using ARIMA

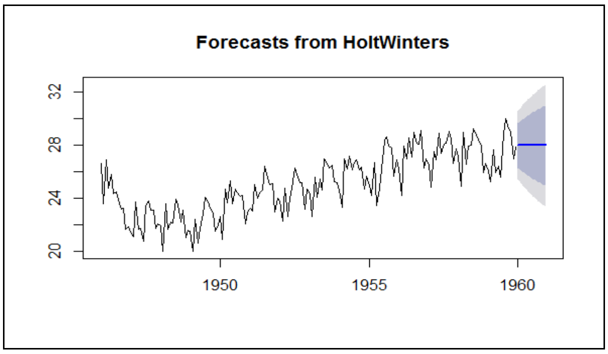

Figure 11. Plot of Original with Forecasted Data on 2 using HoltWinters

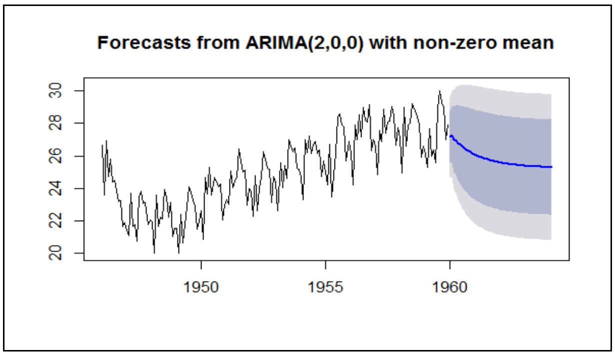

Figure 12. Plot of Original with Forecasted Data on 2 using ARIMA

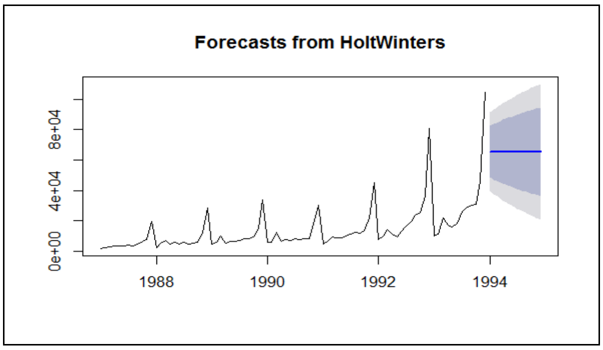

Figure 13. Plot of Original with Forecasted Data on 3 using HoltWinters

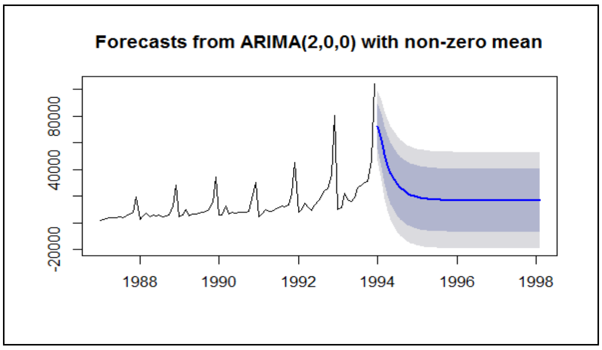

Figure 14. Plot of Original with Forecasted Data on 3 using ARIMA

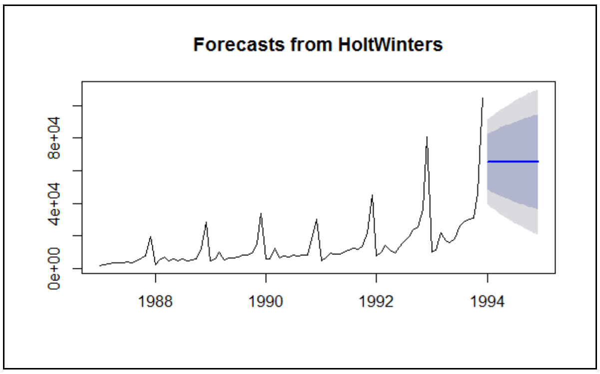

Figure 15. Plot of Original with Forecasted Data on 3 using HoltWinters

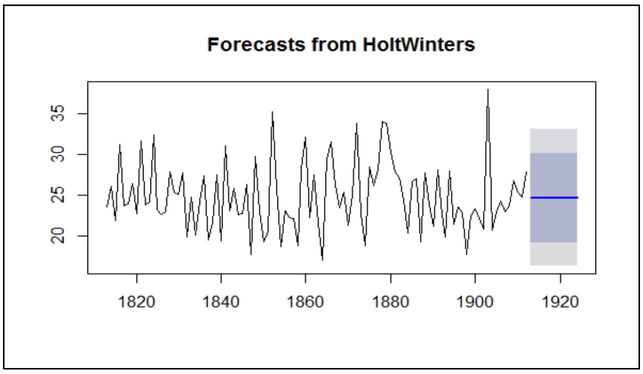

Figure 16. Plot of Original with Forecasted Data on 3 using ARIMA

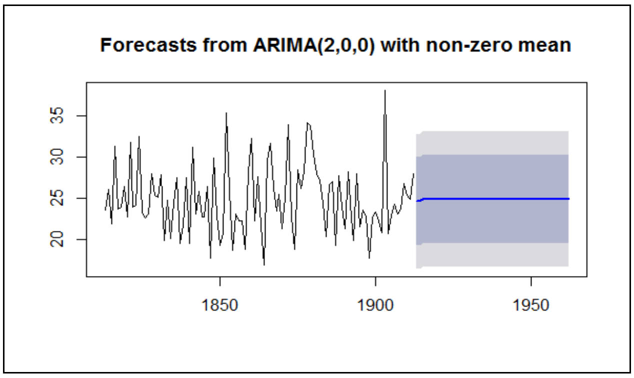

Figure 17. Plot of Original with Forecasted Data on 4 using HoltWinters

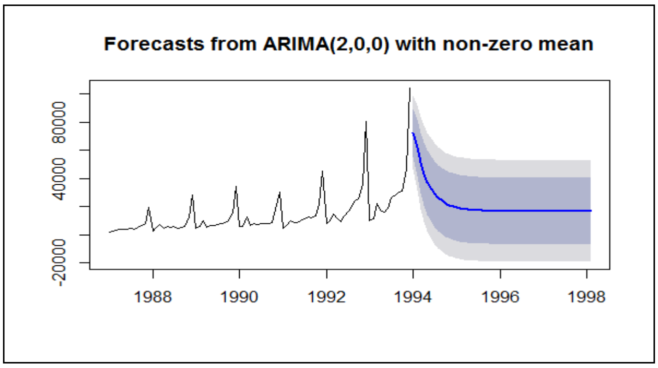

Figure 18. Plot of Original with Forecasted Data on 4 using ARIMA

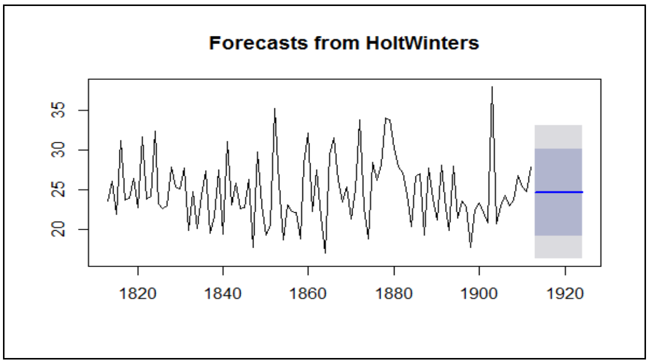

Figure 19. Plot of Original with Forecasted Data 5 using HoltWinters

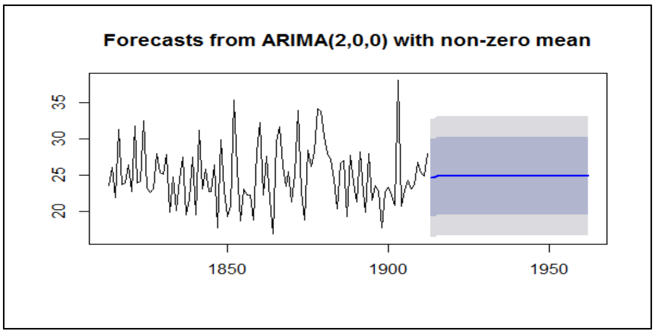

Figure 20. Plot of Original with Forecasted Data on 5 using ARIMA

In Figure 9, the forecasted value is represented as Blue Line, using Holt-Winters model (HWM) on one of the dataset considered for the experiment. Similarly, in Figure 10, the forecasted values are represented using ARIMA model with mean 2, on one of the dataset considered for the experiment and forested the data from 1920 to 2000.

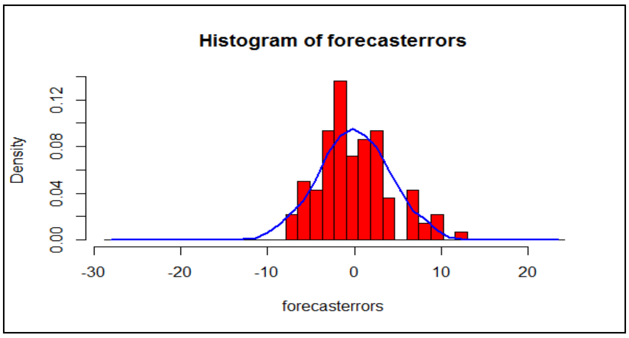

Figure 21. Forecast-Error using ARIMA with Non-Zero Mean on 4

In Figure 11, we obtained the mean error of forecasted model ARIMS on dataset 4 as -0.223, mean error of forecasted model ARIMS on dataset 2 as -0.193 form Figure 12 and mean error of forecasted model ARIMS on dataset 5 as -0.220 form Figure 13. The function forecast is used to plot of the forecasts and prediction intervals.

Table 2. Test Statistic with Degree of Freedom 10

DoF=10			**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
X-squared	HWM	F	97.638	1115.8	222.36	609.2	328.16
		R	4.5023	50.966	9.1874	10.683	34.545
	ARIMA	F	34.727	949.94	83.511	25.388	203.59
		R	6.5271	23.224	7.0753	8.0505	6.176
p-value	HWM	F	<0.05	<0.05	<0.05	<0.05	<0.05
		R	0.9219	1.772e-07	0.5144	0.3827	0.0001493
	ARIMA	F	0.000139	<0.05	1.026e-13	0.004657	<0.05
		R	0.7692	0.009951	0.7183	0.6239	0.8003

In Table 2, the statistical values obtained from Ljung-Box test on Holt- Winters and ARIMA model with respect to X-squared and p-value on attributed Fitted(F) and Residuals(R), with Degree of Freedom equal to ten were listed.

In Table 3, with Degree of Freedom equal to twenty. The P-Value of residuals of both ARIMA and Holt-Winters are compared, from Table 3 and Table 4 one can easily interpret that ARIMA is resulting in the small value. and X-squared of ARIMA is resulting in large value.

Table 3. Test Statistic with Degree of Freedom 20

DoF=20			(1)	(2)	(3)	(4)	(5)
X-squared	HWM	F	180.31	1947.8	291.1	752.3	360.64
		R	13.833	183.12	57.757	17.401	58.88
	ARIMA	F	55.763	1683.6	142.7	35.56	230.67
		R	17.715	122.46	58	15.133	24.364
p-value	HWM	F	<0.05	<0.05	<0.05	<0.05	<0.05
		R	0.8389	<0.05	1.575e-05	0.6268	1.06e-05
	ARIMA	F	3.155e-05	<0.05	<0.05	0.01732	<0.05
		R	0.6062	<0.05	1.446e-05	0.7688	0.2268

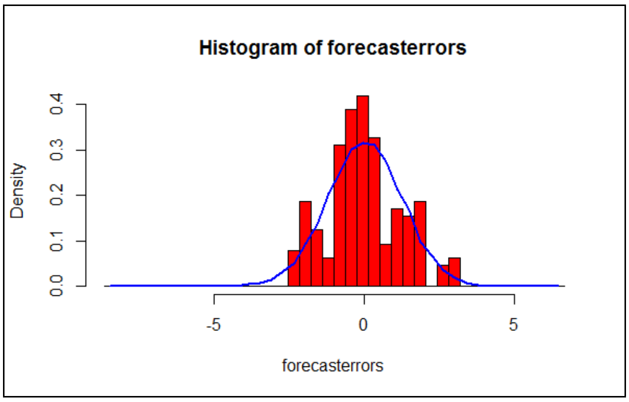

Figure 22. Forecast-Error using ARIMA with Non-Zero Mean on 2

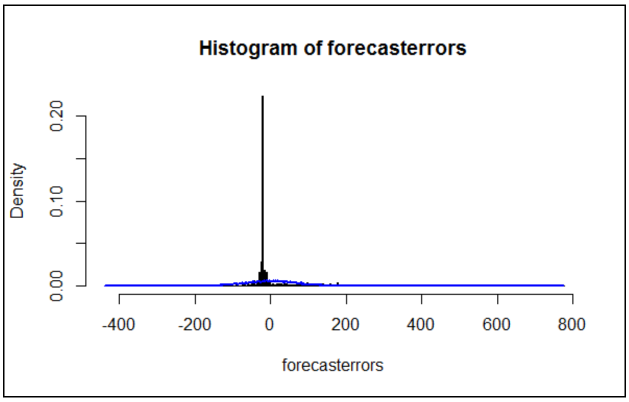

Figure 23. Forecast-Error using ARIMA with Non-Zero Mean on 5

6. Conclusion

In this paper, the forecasting techniques like: HOLT-WINTERS and ARIMA performance are compared with respect to X-squared, Degree of freedom and p-value. In the comparison, it is found that the ARIMA performance is better than that of HOLT-WINTERS with the degree of freedom 10 and 20. In the future, some other techniques used to perform forecasting on time series data will be performed and achieve a minimum sum of squared error in foresting the values.

Acknowledgments

The author Dr. N.Ch.S.N. Iyengar is thankful to Dr. P.Narasimha Reddy, Executive Director, SNIST and Dr. G. Viswanathan, Chancellor**,** VITU for their support and encouragement.

References

[1] F. J. Nogales, J. Contreras, A. J. Conejo and R. Espínola, "Forecasting next-day electricity prices by time series models", IEEE Transactions on power systems, vol. 17, no. 2, **(2002)**, pp. 342-8.
[2] J. D. Hamilton, "A new approach to the economic analysis of non stationary time series and the business cycle", Econometrica: Journal of the Econometric Society, vol. 1, no. 1, **(1989)**, pp. 342-8.
[3] N. Beck, JN. Katz and R. Tucker, "Taking time seriously: Time-series-cross-section analysis with a binary dependent variable", American Journal of Political Science, vol. 1, no. 1, **(1998)**, pp. 1260-88.
[4] J. Pivarski, C. Bennett and RL. Grossman, "Deploying Analytics with the Portable Format for Analytics (PFA)", Proceedings of the International Conference of Knowledge Discovery and Data Mining, **(2016)**.
[5] S. Gupta, M. Nikolaou, L. Saputelli and C. Bravo, "ESP Health Monitoring KPI: A Real-Time Predictive Analytics Application", Proceedings of the International Conference of SPE Intelligent Energy, **(2016)**.
[6] B. Williamson, "Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments", Journal of Education Policy, vol. 31, no. 2, **(2016)**, pp. 123-41.
[7] W. Duan and Y. Y. Chiang, "Building knowledge graph from public data for predictive analysis: a case study on predicting technology future in space and time", Proceedings of the International Conference of Analytics for Big Geospatial Data, **(2016)**.
[8] S. Park, W. Lee and IC. Moon, "Associative topic models with numerical time series", Information Processing & Management, vol. 51, no. 5, **(2015)**, pp. 737-55.
[9] S. Ahmed and D. Tsvetanov, "The predictive performance of commodity futures risk factors", Journal of Banking & Finance, vol. 71, no. 2, **(2016)**, pp. 0-36.
[10] G. M. Ljung and G. E. P. Box, "On a Measure of a Lack of Fit in Time Series Models", Biometrika. vol. 65, no. 2, **(1978)**, pp. 297-303.

Authors

Syed Muzamil Basha, he had his Bachelor of Science in Information Technology at SITAMS, MTech in Information Technology (Networking) at VIT University and currently doing his research at VIT University. His research area are Wireless Sensor Networks, Text Mining and Big Data Predictive Analytics.

Yang Zhenning, he is pursuing M.ScComputer Science at School of Computing Science and Engineering, VIT University, Vellore. His area of interests are Algorithm design an Pattern Recognition, operating Systems and cloud computing.

Dharmendra Singh Rajput working as Associate Professor in the Department of Software and Systems Engineering, School of Information Technology and Engineering, VIT University. His research area are Data Mining and Big Data Predictive Analytics.

Ronnie D. Caytiles, he had his Bachelor of Science in Computer Engineering- Western Institute of Technology, Iloilo City, Philippines, and Master of Science in Computer Science– Central Philippine University, Iloilo City, Philippines. He finished his Ph.D. in Multimedia Engineering, Hannam University, Daejeon, Korea. Currently, he serves as an Assistant Professor at Multimedia Engineering department, Hannam University, Daejeon, Korea. His research interests include Mobile Computing, Multimedia Communication, Information Technology Security, Ubiquitous Computing, Control and Automation.

N. Ch. S. N. Iyengar (b 1961), currently Professor, Department of Information Technology, Sreenidhi Institute of Science and Technology Hyderabad, India. His research interests include Agent-Based Distributed Computing, Intelligent Computing, Network Security, Secured Cloud Computing and Fluid Mechanics. He had 32+ years of experience in teaching and research, guided many scholars, has authored several textbooks and had nearly 200+ research publications in reputed. peer reviewed international journals. He served as PCM/reviewer/keynotespeaker/Invited speaker for many conferences. He serves as editorial board member for many international journals, reviews papers for many conferences with an interest of serving to the education community

