Original Research Article

Efficacy of Holt-Winters Forecast on Seasonal Time Series Crop Insurance Data with Structural Outliers

Abstract

The study analyses the components of time series data on crop insurance of the state of Tamil Nadu, India. The data was found to be seasonal and exponentially increasing. Multiplicative Holt-Winters Model was applied in order to do short range forecast. The model was run in Ms-Excel in order to understand the basics of times series forecasting. Minimum MSE value was the criteria used to find the better fitting smoothing values. The residual of the model was examined for the fit of the model. Residual mean value was close to zero. Residuals are tested for autocorrelation with Durbin-Watson test and Runs test. Histogram of residuals implies a normal distribution. Presence of outliers are detected using 3IQR method and the identified outliers are part of the structure of data and need not be removed. However, alternate models of Holt-Winters itself which are robust to work with outliers are reviewed and RHW model of Gelpers et al. (2010) was suggested.

Keywords: Holt-Winters, Forecast, Seasonal Data, Crop Insurance, Structural Outliers

Introduction

Crop Insurance in Tamil Nadu

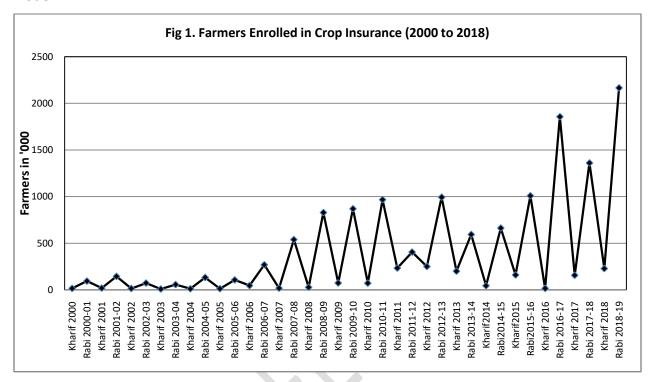
Agriculture is a risky business subject to price, climatic, geological and biological shocks Wenner (2005). Farmers face floods, drought, pests, disease, and a plethora of other natural disasters. As more than half of the cultivated area is rainfed (OECD, 2018), it is highly dependent on monsoon precipitation. Crop insurance provides risk sharing and risk balancing functions to farmer who is highly affected by seasonal calamities (Subash *et.al.*, 2017). Insurance could potentially help farmers cope with shocks without having to reduce investments in education or agriculture (Kramer and Cellabos, 2018). The Journey of crop insurance in India rolled out 48 years ago in 1971 when the First individual Approach Scheme

was introduced which existed till 1978. A Pilot Crop Insurance Scheme (PCIS) was orchestrated from 1979 to 1984. After which a Comprehensive Crop Insurance Scheme (CCIS) was introduced in 1985 till 1999. During Rabi 1999 has seen a new insurance scheme called National Agricultural Insurance Scheme (NAIS). In the state of Tamil Nadu, NAIS has covered 83.08 lakh farmers cumulatively till Rabi 2015-16. Based on the recommendations on NAIS a Modified National Agricultural Insurance Scheme (MNAIS) was came into force from Rabi 2010-11 and has covered 4.76 lakh farmers cumulatively till Rabi 2015-16 in the state (Ministry of Agriculture and Farmers Welfare, 2019). The year 2016-17 has seen the genesis of a new insurance scheme called Pradhan Mantri Fasal Bima Yojana (PMFBY) instead of NAIS and MNAIS. Pooling in the important learning from all the earlier schemes and taking into consideration of access to technology in the recent days, Pradhan Mantri Fasal Bima Yojana promises to take care of the loopholes of earlier schemes (OECD, 2018).

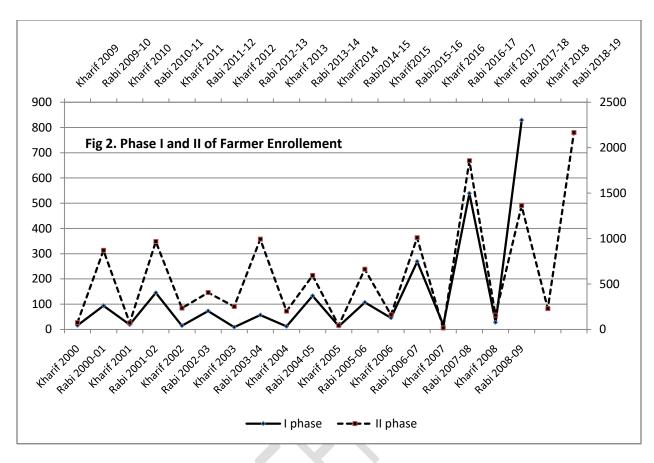
The data

The time series data on number of farmers enrolled in crop insurance was obtained from the Commissionerate of Agriculture, Chepauk, Chennai. The data used in the analysis pertains to farmers enrolled under NAIS, MNAIS and PMFBY over the time period from Kharif 2000 to Rabi 2018-19 in the state of Tamil Nadu. As given here the data contains two seasons for a year. Inherently the data is seasonal which is inevitable. The data from Kharif 2000 to Rabi 2010-11 consists of farmers enrolled under NAIS only. Whereas, from Kharif 2011-12 to Rabi 2015-16 consists of farmers enrolled under both NAIS and MNAIS. After Rabi 2015-16, both the schemes were up taken by PMFBY hence, the data is exclusive to PMFBY. Given the prelude, the time plot on farmers' enrollment over the years is presented in figure 1. When the data is plotted against time, it could be apparently noted that the data is not constant over time and changes (moreover increases) along with time, especially, the the variation in the seasonal pattern appears to be proportional to the level of the time series. Plotting the series and examining the structure is the most important and essential step in time-series analysis; preliminary even to adjusting and modeling the data (Chatfield and Mohammad, 1988). In this context, the paper has attempted to examine the crop insurance data and to forecast using the basic excel software. With modern machine learning for forecasting, the basic procedures of a

model has often regarded as automatic (Chatfield, 1978). Thus, this paper could help in understanding not only the behaviour of data but also the underlying basics of a forecast model.



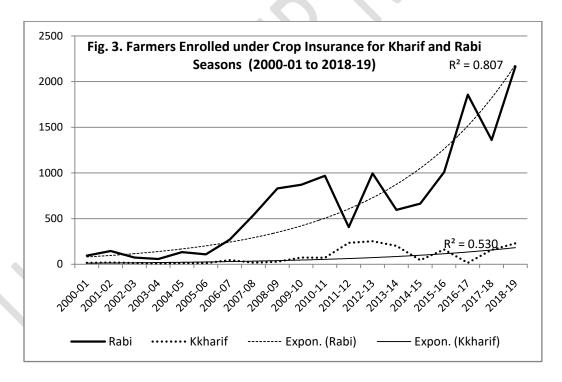
When closely observed, in the above time plot the initial pattern in the data has repeated itself. The first part of data with a pattern was classified as 'Phase I' (which doesn't mean it as the Phases of crop insurance in India but rather the phases within the given data). The second part of data, which closely resemble the pattern of Phase I, is classified as Phase II. The Phase I is from Kharif 2000-01 to Rabi 2008-09 and the Phase II is from Kharif 2009-10 to Rabi 2018-19 (Figure 2). Phase I is presented in Primary axes and Phase II is presented in Secondary axes. The exact reason behind this pattern arrangement within the data couldn't be established.



Kharif and Rabi Seasons

Another observation could be that Tamil Nadu farmers prefer to insure Rabi crops than Kharif crops. The 5 year ending area under crops 2017-18 has reported a conflicting fact. That is the area under food grains in Kharif season is more than Rabi season by more than 10 times and the area under oilseeds in Kharif season are twice than that of Rabi season (Government of India, 2019). One reason should be rainfall during Rabi season and Cauvery water being available during Kharif season. The state receives its highest showers during the Rabi season from North-East monsoon (October to December) just like most of South India. To emphasize, though Tamil Nadu is under bi-monsoon pattern, more share of rainfall is received from NE monsoon (48%) than SW monsoon (35%). And also, 2016-17 was a dry year in Tamil Nadu when the North-East monsoon deviated by around -62%. There is assurance of water availability for irrigation during Kharif season as Cauvery water is released during June and July. Farmers associate enrolment to crop insurance with water availability at least in delta region (based on author's survey). Seasonal variation in agricultural data is innate, though opportunity is given also to glance the data classified into Kharif and Rabi seasons (Figure 3). One other reason for

higher Rabi enrollment: the important paddy-growing season in Tamil Nadu such as, Samba, Thaladi, Pishanam are being counted under Rabi season. Even within southern India, the peculiarity of higher Rabi enrollment is very special to the state of Tamil Nadu alone for eg. the claims paid during Rabi season under PMFBY for two years (2016-17 and 2017-18) in Tamil Nadu contributed around 50% of claims paid throughout India during the same season and period. The disparity between Kharif and Rabi farmer numbers could not allow for a reliable trendline in Figure 1. While the Figure 3 could enable for an exponential trendline with R² values of 0.530 and 0.807 for Kharif and Rabi, respectively. Exponential trendlines are used when the data is either falling or rising at an increasing rate. Thus exponential type of trendlines are the most appropriate for the given data and also the R² value were the highest under the aforesaid type than any other. This inability to fit a trendline in Figure 1 is not sufficient to classify the data into two, as there are many well-developed models, which could work on a seasonal data.



Methodology Holt-Winter's Multiplicative Model

The basic Exponential smoothing methods were developed by Robert. G. Brown (1959 & 1964) and later expanded by Holts. C.C. (1957) and Winters (1960). In exponential smoothing,

one or more parameters assign exponentially decreasing weights for the older observations as the "future events usually depend more on recent data than on data from a long time ago" (Xie et.al., 1997). There is also a power of adjusting early forecast for older observations. Gardener (1985) claims that the popular reputation of exponential smoothing has attributed to the several practical considerations in short-range forecasting. When the stochastic processes of the data (Y_t) are not stationary, then the statistical properties of forecast cannot capture the estimates of original data ($E[Y_t]$). In such cases a sensible method could be to use weighted moving average with exponentially decreasing weights (Winters, 1960). Several remarkable econometricians such as Pegels (1969), Roberts (1982), Abraham and Ledolter (1983), Makridakis and Hibon (1979) were on the league of 'exponential smoothing' methods and also have made contributions to augment the methodology. One such augmentation to the equations, which is very widely used, is that of Winters (1960), which hereafter be called as Holt-Winters' model (HW).

Holt-Winter's model is the second extension of basic exponential smoothing model and it is optimal for a wide variety of time series. The basic exponential smoothing is also called as single exponential smoothing (SES) as it has a single equation for a 'level' component. The equation for SES can be written as follows,

$$S_t = \alpha Y_{t-1} + (1 - \alpha)S_{t-1}$$
 $0 \le \alpha \le 1$ (1)

 S_t is the smoothing or level equation at time t. The speed at which the older responses are dampened (smoothed) and adapt to changes in level is a function of the value of α . The Damping factor (1- α) assigns the least value to distant data points. When α is close to 1, dampening is quick and when α is close to 0, dampening is slow. Thus smoothing constant and damping factors are inversely proportional. Larger dampening factors smoothes out the peaks and valleys more than smaller damping factors. Smaller damping factors also mean that the smoothed values are closer to the actual data points than larger damping factors.

The first extension to SES is 'Holt's Linear trend' model which is also called as 'double exponential smoothing- DES' as it has two equations one for 'level' and other for 'trend' component. Thus when the data has a linear trend then DES is suitable. The equation for DES can be written as follows,

$$S_{t} = \alpha Y_{t} + (1-\alpha)(S_{t-1} + b_{t-1}) \qquad 0 \le \alpha \le 1 \qquad(2)$$

$$b_{t} = \gamma(S_{t} - S_{t-1}) + (1-\gamma)b_{t-1} \qquad 0 \le \gamma \le 1 \qquad(3)$$

Note that the current value (Y_t) of the series itself is used to calculate its smoothed value is the replacement in double exponential smoothing. In the 'trend' equation (3) a value of γ is chosen to allow the trend estimate to react to changes in the rate of growth of the series. There are two smoothing equations which will be combined later to form a forecast equation, $\widehat{Y_{t+1}} = S_t + b_t$.

As states earlier, HW model is the second extension of SES and also called as the triple exponential smoothing model. The HW model is a three parameter model which augments equations to include trend and seasonality in the data. Holt-Winter's watch out for trend and seasonality in the data, whose values increases over time and could reproduce the seasonal changes in the data. Forecast takes two forms given the movement of values in the data. Multiplicative model is chosen when the data is exponentially increasing with time and if not Additive model is chosen. HW algorithms also have two forms, multiplicative and additive. The HW multiplicative model could best forecast using the given data as the crop insurance data itself is inherently seasonal and exponentially increasing. The basic equation for TES is as follows,

Overall smoothing equation:
$$S_t = \alpha \frac{Y_t}{I_{t-l}} + (1-\alpha)(S_{t-1} + b_{t-1})$$
(4)

Trend Equation:
$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}$$
(5)

Seasonality equation:
$$I_t = \beta \frac{Y_t}{S_t} + (1-\beta) I_{t-L}$$
(6)

Forecast Equation:
$$F_{t} = \widehat{Y}_{t} = (S_{t-1} + b_{t-1})I_{t-L}$$
(7)

Where $0 < \alpha \le 1$, $0 \le \gamma \le 1$ and $0 \le \beta \le 1 - \alpha$. The small value of β for the multiplicative model means that the seasonal component hardly changes over time.

TES has one more additional equation (6) I_t for seasonality component. θ is the smoothing constant for seasonality whose value ranges between 0 and $1-\alpha$. When the level equation is substituted in equation (6), the usual parameter restrictions translates to this range (Hyndman and Athanasopoulos, 2018). L here denotes the frequency of the seasonality, *i.e.*, the number of seasons in a year. For example, for quarterly data L=4, and for monthly

data L=12. The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series. With the multiplicative method, the seasonal component is expressed in relative terms (percentages), and the series is seasonally adjusted by dividing through by the seasonal component. Within each year, the seasonal component will sum up to approximately L. Exponential smoothing is an interesting model and it predicts the future evolution of the time series via a simple extrapolation. Researchers have found Holt-Winters having better predicting power than complex Box-Jenkins methodology such as ARIMA (Roberts, 1982); SARIMA (Makatjane and Ntebogang, 2016) and also precise in short term forecasting (Makatjane and Ntebogang, 2016; Chatfield et.al., 1988). Tratar and Strmčnik (2016) has found Holt-Winters suitable for both short-term as well as long term forecasting. Makridakis and Hibon (1997) have found the problem in ARIMA models of Box and Jenkins is the way of making the series stationary in its mean. Whereas, Holt-Winters is the sensible method when the data is not stationary (Winters, 1960). The Holt-Winters model has been augmented and it was being proposed and done by various statisticians and econometricians over the years. Taylor (2003 & 2010) introduced the double and triple seasonal Holt-Winters models (HWT). These models are characterized by capturing the information contained in the seasonal component, split into several seasonalities of different lengths, as well as including an adjustment of the forecast including the first autocorrelation error. García-Díaz and Trull (2016) generalized these models to adapt to an indeterminate number of seasonalities, proposing the multiple seasonal Holt-Winters models (nHWT). Holt-Winters along with its variations have been found to be suitable for most socio-economic studies.

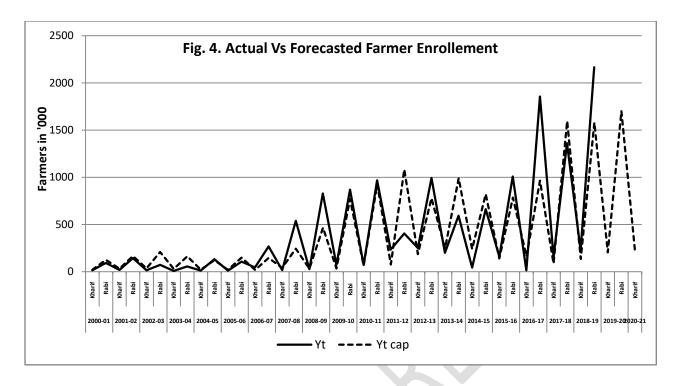
The initial values

The holt winter's model is a recursive model, it provides results which are responsive not only to the smoothing parameters chosen but also to the method of initializing the values. Given the equations above, the methodology of exponential smoothing operates to assign the weights to preceding values to obtain the forecast. Thus the initial values and the weights assigned to them are important especially when the dataset is less than 20 years as in this case

(and when alpha value is also small). There are several methods established to reach this initial values. For example, for S₀, the original method given by Brown is to take the mean of the data. To mention, Backcasting (Ledolter and Abraham, 1984), Bayesian methods combined with an average of the available data (Cohen, 1966; Jonhson and Montgomery, 1974 & Taylor, 1981), regression based procedure (Bowerman *et.al.*, 2005) are few other methods to estimate. Accurate estimates of initial conditions can result in better forecasting results (Vercher *et.al.*, 2012 & Hansun, 2017). The estimation procedure for initial conditions used in this paper is of the Hyndman and Athanasopoulos (2018). Hyndman's method consists of a moving average decomposition to attain the seasonally adjusted initial level 'S₀' and trend values 'b₀' as well as de-trended initial seasonal values 'l₋₁ and l₀'. The initial values are then placed before the time period t=1. The procedure is the one which is implemented in the 'HoltWinters' function in R software.

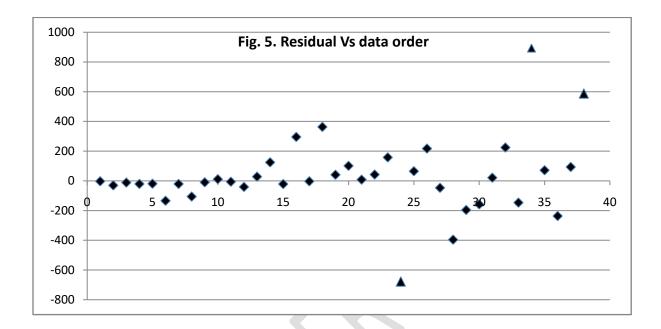
Results and discussion

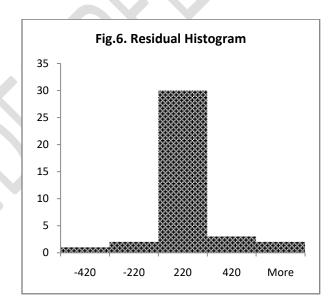
The result of the Holt-Winters model on crop insurance data is the predicted \widehat{Y}_t values. In this paper, the Holt-Winters procedure as in R software was followed and so the predicted numbers of farmers enrolled were calculated even from the first year. The predicted and actual numbers of farmer enrolled in the crop insurance were plotted against time (Figure. 4). The values for smoothing constants α , γ and β were chosen based on the minimum root square mean value as 1.414, 0.11 and 0.46, respectively. The Holt-Winters have forecasted the farmer enrollment for three future period Kharif and Rabi of 2019-20 and Kharif of 2020-21.



A time series forecasting model could be validated with the characteristics of their residuals. Predominantly, a residual should have mean zero and no autocorrelation. The mean of the residuals in this case is 27. The residual plot presented in Figure 5 is plotted against data order, which could aid in inferences towards locating outliers and autocorrelation also. The residual plot also shows that there is no autocorrelation. The Durbin-Watson statistics was very close to two at 1.92 and the p-value of Runs tests was higher than the significance level. Hence, along with scatter plot, Durbin-Watson test and Runs test also conclude that there is no autocorrelation. Other not so binding requirements are constant variance and normal distribution. A histogram of residuals infer that they are normally distributed (Figure 6). If all the other conditions are satisfied, a variation in mean could be adjusted by just adding the mean value to the predictions (Hyndman and Athanasopoulos, 2018). Another important finding from the residual plot is outliers. The outliers are identified using 3IQR (Inter quartile range) formula (NIST/SEMATECH, 2012). The upper lower fence and upper outer fence values are -416.18 and 465.51, respectively. These outliers are significant difference between actual and predicted values at certain points namely, at 2011-12 (strong growth performance in agriculture), 2016-17 (drought and PMFBY was introduced), 2007-08, 2018-19 etc. Outliers are marked with Δ symbol in the plot. There are methods to pre-clean the data for forecasting.

However, in the given crop insurance data, the outliers mentioned above and such are in the structure of crop insurance data and they are inevitable during forecast. The residual data points beyond the upper inner and outer fence are extreme outliers.





Any variation in the exponential smoothing method, which claims to be robust need to reduce the number of outliers beyond the IQR. It was Cipra (1992), who presented a variation of Holt–Winters method as a weighted regression problem with a robust regression fit (not the

traditional regression). Using quantile regression to obtain robust forecasts (Taylor, 2007) and method to classify each observation (Kirkendall, 2006) are few other variations. However, these methods do not use the basic recursive formulae of Holt-Winters, which are very important. The procedure of Cipra yields a reasonable method, but it can be improved upon. The method by Cipra consisted of a Huber- Ψ function (weights assigned to each data point decreases when the distance for mean increases) and iterated weighted least square algorithm. A augmentation to the Holt-Winters model was carried out by Gelper *et.al.*, (2010) and are called as Robust Holt-Winters (RHW). The RHW are suitable to forecast even in the presence of outliers. The methodology of Gelpers is the widely used one and could be run in *R* software with the code *Robets*. The method suggested by Gelpers consists of using pre-cleaned Y_t^* instead of actual Y_t . The data cleaning process of Gelper also uses the Huber- Ψ function to replace the unexpected values of more likely values. The pre-cleaned Y_t^* is used in forecasting equation of Exponential and Holt-Winters model. Due to these updates, not just the data but also the starting value and smoothing values need to be set. It is suggested to use repeated median estimator than OLS of Bowerman *et.al.* (2005) to find the starting values for RHW.

Conclusion

The study has attempted to explore the crop insurance farmers enrollment as a time series data. Crop insurance schemes itself had been changed twice within the data period. Seasonal fluctuations are explicitly found as farmers preferring insurance for Rabi crops. The reason for the same might be the assurance of water availability for irrigation during kharif season. The data has exhibited an exponentially increasing trend, which recommended a multiplicative model for forecast. Holt-Winters exponential smoothing was the chosen model for forecast as it best suits the seasonal data with exponential trend. The forecast is carried out in Excel 2007 for better understanding of the model. The model was run and the smoothing constants where chosen based lesser MSE value. The characteristics of residuals have complied with necessitates of fine forecasting model. However, when plotted against time the predicted values and actual values were significantly different in few instances. Outliers were identified using residual plot and interquartile range. Outliers are observed as those associated with structural changes in the crop insurance data. Hence, it is concluded that the conventional

model is not robust enough to handle the structural outliers which are required for forecasting. Few variations to exponential smoothing which could handle outliers were discussed and the Gelper's (2010) augmentation on Holt-Winters model (RHW) was suggested. A complex RHW model is automatic and could be run in *R* Software.

References

- 1. Abraham B. and Ledolter J. 1983. Statistical Methods for Forecasting. Wiley: New York.
- 2. Bowerman B., O'Connell R. and Koehler A. 2005. *Forecasting, Time Series, and Regression*. Duxbury Press. Pacific Grove. California: Thomson Books Cole.
- 3. Holt C.C. 1957. "Forecasting seasonals and trends by exponentially weighted averages." *Proceedings of the O.N.R. Memorandum*, Carnegie Institute of Technology, Pittsburgh. Volume 52.
- 4. Chatfield C. 1978. "The Holt-Winters Forecasting Procedures." *Applied Statistics*. 27: 264—279.
- 5. Chatfield C. and Mohammad Y. 1991. "Prediction Intervals for Multiplicative Holt Winters." *International Journal of Forecasting* 1 (31).
- 6. Cipra T. 1992. "Robust exponential smoothing." *Journal of Forecasting* 11: 57–69.
- 7. Cohen G.D. 1966. "Bayesian Adjustment of Sales Forecasts in Multi-item Inventory Control System." *Journal of Industrial Engineering*. 17: 474-479.
- 8. Gardener E.S.Jr. 1985. "Exponential Smoothing: The State of Art." *Journal of Forecasting* 4:1-28
- 9. García-Díaz J.C. and Trull O. 2016. "Competitive Models for the Spanish Short-Term Electricity Demand Forecasting." pp. 217–231. *Time Series Analysis and Forecasting: Selected Contributions* from the ITISE Conference. Rojas I. and Pomares H. Cham, Switzerland: Springer International Publishing. ISBN 978-3-319-28725-6.
- 10. Gelper S. E. C., Fried R. and Croux C. 2010. "Robust forecasting with exponential and holt-winters smoothing." *Journal of Forecasting*. 29(3): 285-300. https://doi.org/10.1002/for.1125
- 11. Hansun S. 2017. "New Estimation Rules for Unknown Parameters on Holt- Winters Multiplicative Method." J. Math. Fund. Sci. 49 (2): 127-135
- 12. Hyndman, R.J., and G. Athanasopoulos, 2018, *Forecasting: Principles and Practice*. 2nd Edition. Melbourne, Australia: OTexts. OTesxt.com/fpp2. Accessed on 3.4.2020.
- 13. Johnson L.A. and Montgomery D.C. 1974. *Operational Research in Production Planning, Scheduling and Inventory Control.* New York: Wiley.
- 14. Kirkendall N. 2006. "Monitoring for outliers and level shifts in kalman filter implementations of exponential smoothing". *Journal of Forecasting* 11: 543–560.
- 15. Kramer B. and Cellabos F. 2018. Enhancing adaptive capacity through climate-smart insurance: Theory and evidence from India. 30th International Conference of Agricultural Economists. Vancouver.

- 16. Ledolter J. and Abraham B. 1984. "Some Comments on the Initialisation of Exponential Smoothing." *Journal of Forecasting*. 3: 79-84.
- 17. Makridakis S. and Hibon M. 1979. "Accuracy of Forecasting: An Emprical Investigation (with discussion)." *Journal of Royal Statistical Society* 142: 97-145.
- 18. Ministry of Agriculture and Farmers Welfare. 2019. Agricultural Statistics at a Glance-2018. Department of Agriculture, Cooperation & Farmers Welfare. Directorate of Economics and Statistics. Government of India.
- 19. Makatjane K.D. and Ntebogang D.M. 2016. "Comparative Study of Holt-Winters Triple Exponential Smoothing and Seasonal ARIMA: Forecasting Short Term Seasonal Car Sales in South Africa." Risk governance & control: financial markets & institutions 6(1).
- 20. NIST/SEMATECH. 2012. E-Handbook of Statistical Methods, https://www.itl.nist.gov/div898/handbook/. 3.4.2020.
- 21. OECD. 2018. "Review of Agricultural Policies in India." Trade and Agriculture Directorate
 Committee for Agriculture.
 http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=TAD/CA(201
 8)4/FINAL&docLanguage=En
- 22. Pegels C. 1969. "Exponential Smoothing: Some New Variations." *Management Science*. 15: 311-315.
- 23. Roberts, S.A., 1982, A general class of Holt-Winters type forecasting models, Management Science, 28: 808-820.
- 24. Subash S.P., Pavithra S. and Suresh Pal. 2017. "Pradhan Mantri Fasal Bima Yojana: Challenges and Way Forward." https://www.researchgate.net/publication/322266079_Pradhan_Mantri_Fasal_Bima_Yojana_PMFBY_Challenges_and_way_forward
- 25. Taylor J.W. 1981. "Initialisation of Exponential Smoothing Forecast." *AllE Transactions* 13:199-205
- 26. Taylor, J.W. Short-term electricity demand forecasting using double seasonal exponential smoothing. J. Oper.
- 27. Res. Soc. **2003**, 54, 799–805.
- 28. Taylor J. 2007. "Forecasting daily supermarket sales using exponentially weighted quantile regression." *European Journal of Operational Research* 178: 154–167.
- 29. Taylor J.W. 2010. "Triple seasonal methods for short-term electricity demand forecasting." Eur. J. Oper. Res. 204: 139–152.
- 30. Tratar F. L. and Strmcnik E. 2016. "The Comparison of Holt-Winters method and Multiple regression method: A case study." *Energy* 109: 266-276.
- 31. Vercher E., Corberan-Vallet A., Segura J.V. and Bermudez J.D., 2012. "Initial Conditions Estimation for Improving Forecast Accuracy in Exponential Smoothing." *TOP* 20(2): 517-533.
- 32. Winters P. 1960. "Forecasting sales by exponentially weighted moving averages." *Management Science* 6: 324–342.

- 33. Wenner, Mark. 2005. "Agricultural Insurance Revisited: New Developments and Perspectives in Latin America and the Caribbean." Rural Development Unit Sustainable Development Department. Inter-American Development Bank. New York.
- 34. Xie M., Hong G.Y. and Wohlin C. 1997. "A study of exponential smoothing technique in software reliability growth prediction." *Quality and Reliability Engineering* 13(6):347-353.