A survey of forecast error measures

Article *in* World Applied Sciences Journal · January 2013

DOI: 10.5829/idosi.wasj.2013.24.itmies.80032

Maxim Shcherbakov
Volgograd State Technical University

Adriaan Brebels
KU Leuven

Anton Tyukov
Volgograd State Technical University

World Applied Sciences Journal 24 (Information Technologies in Modern Industry, Education & Society): 171-176, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.24.itmies.80032

A Survey of Forecast Error Measures

Maxim Vladimirovich Shcherbakov, Adriaan Brebels,
Nataliya Lvovna Shcherbakova, Anton Pavlovich Tyukov,
Timur Alexandrovich Janovsky and Valeriy Anatol'evich Kamaev

Volgograd State Technical University, Volgograd, Russia

Submitted: Aug 7, 2013; **Accepted:** Sep 18, 2013; **Published:** Sep 25, 2013

Abstract: This article reviews the common used forecast error measurements. All error measurements have been joined in the seven groups: absolute forecasting errors, measures based on percentage errors, symmetric errors, measures based on relative errors, scaled errors, relative measures and other error measures. The formulas are presented and drawbacks are discussed for every accuracy measurements. To reduce the impact of outliers, an Integral Normalized Mean Square Error have been proposed. Due to the fact that each error measure has the disadvantages that can lead to inaccurate evaluation of the forecasting results, it is impossible to choose only one measure, the recommendations for selecting the appropriate error measurements are given.

Key words: Forecasting · Forecast accuracy · Forecast error measurements

INTRODUCTION

Different criteria such as forecast error measurements, the speed of calculation, interpretability and others have been used to assess the quality of forecasting [1-6]. Forecast error measures or forecast accuracy are the most important in solving practical problems [6]. Typically, the common used forecast error measurements are applied for estimating the quality of forecasting methods and for choosing the best forecasting mechanism in case of multiple objects. A set of "traditional" error measurements in every domain is applied despite on their drawbacks. These error measurements are used as presets in domains despite on drawbacks.

This paper provides an analysis of existing and quite common forecast error measures that are used in forecasting [4, 7-10]. Measures are divided into groups according to the calculating method an value of error for certain time t. The calculating formula, the description of the drawbacks, the names of assessments are considered for each error measure.

A Review

Absolute Forecasting Error: The first group is based on the absolute error calculation. It includes estimates based on the calculation of the value e_i

$$e_t = \left(y_t - f_t^{(m)}\right) \tag{1}$$

where - y_t is the measured value at time t, $f_t^{(m)}$ - predicted value at time *t*, obtained from the use of the forecast model *m*. Hereinafter referred to as the index of the model (*m*) will be omitted.

Mean Absolute Error, MAE is given by:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|e_i\right| = \underset{i=\overline{1,n}}{mean}\left|e_i\right|, \tag{2}$$

where *n* –forecast horizon, *mean*(•) – a mean operation.

Median Absolute Error, MdAE is obtained using the following formula

$$MdAE = \underset{i=\overline{1,n}}{median}\left|e_i\right|, \tag{3}$$

where *mean*(•) – operation for calculation of a median.

Mean Square Error, MSE is calculated by the formula

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(e_i^2\right) = \underset{i=\overline{1,n}}{mean}\left(e_i^2\right), \tag{4}$$

Corresponding Author: Shcherbakov, Volgograd State Technical University, Lenin avenue, 28, 400005, Volgograd, Russia.

hence, Root Mean Square Error, RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(e_i^2\right)} = \sqrt{\underset{i=\overline{1,n}}{mean}\left(e_i^2\right)} \quad (5)$$

These error measures are the most popular in various domains [8, 9]. However, absolute error measures have the following shortcomings.

- The main drawback is the scale dependency [9]. Therefore if the forecast task includes objects with different scales or magnitudes then absolute error measures could not be applied.
- The next drawback is the high influence of outliers in data on the forecast performance evaluation [11]. So, if data contain an outliers with maximal value (this is common case in real world tasks), then absolute error measures provide conservative values.
- *RMSE*, *MSE* have a low reliability: the results could be different depending on different fraction of data [4].

Measures Based on Percentage Errors: Percentage errors are calculated based on the value P_t

$$p_t = \frac{|e_t|}{y_t} \quad (6)$$

Also these errors are the most common in forecasting domain. The group of percentage based errors includes the following errors.
Mean Absolute Percentage Error, MAPE

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} 100\cdot|p_i| = \underset{i=\overline{1,n}}{mean}\left(100\cdot|p_i|\right) \quad (7)$$

Median Absolute Percentage Error, MdAPE is more resistant to outliers and calculated according to the formula

$$MdAPE = \underset{i=\overline{1,n}}{median}\left(100\cdot|p_i|\right) \quad (8)$$

Root Mean Square Percentage Error, RMSPE is calculated according to:

$$RMSPE = \sqrt{\underset{i=\overline{1,n}}{mean}\left(100\cdot|p_i|\right)^2}, \quad (9)$$

and the median percentage error of the quadratic

$$RMdSPE = \sqrt{\underset{i=\overline{1,n}}{median}\left(100\cdot|p_i|\right)^2}, \quad (10)$$

We note the following shortcomings.

- Appearance division by zero when the actual value is equal to zero.
- Non-symmetrical issue - the error values differ whether the predicted value is bigger or smaller than the actual [12-14].
- Outliers have significant impact on the result, particularly if outlier has a value much bigger then the maximal value of the "normal" cases [4].
- The error measures are biased. This can lead to an incorrect evaluation of the forecasting models performance [15].

Symmetric Errors: The criteria which have been included in this group are calculated based on the value:

$$s_t = \frac{|e_t|}{\left(y_t + f_t\right)}. \quad (11)$$

The group includes next measures. Symmetric Mean Absolute Percentage Error, sMAPE is calculated according to

$$sMAPE = \frac{1}{n}\sum_{i=1}^{n} 200\cdot|s_i| = \underset{i=\overline{1,n}}{mean}\left(200\cdot|s_i|\right), \quad (12)$$

and the median mean absolute percentage error

$$sMdAPE = \underset{i=\overline{1,n}}{median}\left(200\cdot|s_i|\right). \quad (13)$$

To avoid the problems associated with the division by zero, a modified sMAPE - Modified sMAPE, msMAPE has been proposed. Their denominators have an additional member:

$$msMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - f_i|}{\left(|y_i| + |f_i|\right)/2 + S_i}, \quad (14)$$

where $S_i = \frac{1}{i-1}\sum_{k=1}^{i-1}\left|y_k - \bar{y}_{i-1}\right|, \bar{y}_{i-1} = \frac{1}{i-1}\sum_{k=1}^{i-1} y_k$.

Developing the idea for the inclusion of an additional terms, more sophisticated measures was presented [16]: KL-N, KL-N1, KL-N2, KL-DE1, KL-DE2, IQR

The following disadvantages should be noted.

- Despite its name, this error is also non-symmetric [13].
- Furthermore, if the actual value is equal to forecasted value, but with opposite sign, or both of these values are zero, then a divide by zero error occurs.
- These criteria are affected by outliers in analogous with the percentage errors.
- If more complex estimations have been used, the problem of interpretability of results occurs and this fact slows their spread in practice [4].

Measures Based on Relative Errors: The basis for calculation of errors in this group is the value determined as follows:

$$r_t = \frac{|e_t|}{\left(y_t - f_t^*\right)}, \tag{15}$$

where f_t^* - the predictive value obtained using a reference model prediction (benchmark model). The main practice is to use a naive model as a reference model

$$f_t^* = y_{t-l}, \tag{16}$$

where l - the value of the lag and $l = 1$.

The group includes the next measures. Mean Relative Absolute Error, MRAE is given by the formula

$$MRAE = \underset{i=1,n}{mean}|r_i|, \tag{17}$$

Median Relative Absolute Error, MRAE is calculated according to

$$MdRAE = \underset{i=1,n}{median}|r_i|, \tag{18}$$

and Geometric Mean Relative Absolute Error, GMRAE), which is calculated similarly to (17), but instead of *mean*(•) the geometric mean is obtained *gmean*(•).

It should be noted the following shortcomings.

- Based the formulas (15-18), division by zero error occurs, if the predictive value obtained by reference model is equal to the actual value.
- If naive model has been chosen then division by zero error occurs in case of continuous sequence of identical values of the time series.

Scaled Error: As a basis for calculating the value of the scaled errors q_i is given by

$$q_t = \frac{|e_t|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|}. \tag{19}$$

This group contains Mean Absolute Scaled Error, MASE proposed in [9]. It is calculated according to:

$$MASE = \underset{i=1,n}{mean}|q_i|, \tag{20}$$

Another evaluation of this group is Root Mean Square Scaled Error, RMSSE is calculated by the formula [10]:

$$RMSSE = \sqrt{\underset{i=1,n}{mean}\left(q_i^2\right)}. \tag{22}$$

These measures is symmetrical and resistant to outliers. However, we can point to two drawbacks.

- If the forecast horizon real values are equal to each other, then division by zero occurs.
- Besides it is possible to observe a weak bias estimates if you do the experiments by analogy with [15].

Relative Measures: This group contains of measures calculated as a ratio of mentioned above error measures obtained by estimated forecasting models and reference models. Relative Mean Absolute Error, RelMAE is calculated by the formula.

$$RMAE = \frac{MAE}{MAE^*}, \tag{23}$$

where MAE and MAE^* the mean absolute error for the analyzed forecasting model and the reference model respectively, calculated using the formula (2).

Relative Root Mean Square Error, RelRMSE is calculated similarly to (23), except that the right side is calculated by (5)

$$RRMSE = \frac{RMSE}{RMSE^*}. \tag{24}$$

In some situations it is reasonable to calculate the logarithm of the ratio (23). In this case, the measure is called the Log Mean Squared Error Ratio, (LMR)

$$LMR = \log\left(\frac{RMSE}{RMSE^*}\right). \tag{25}$$

Syntetos *et al*. proposed a more complex assessment of the relative geometric standard deviation Relative Geometric Root Mean Square Error, RGRMSE [17].

The next group of measures counts the number of cases where the error of the model prediction error is greater than the reference model. For instance, PB (MAE) - Percentage Better (MAE), calculated by the formula:

$$PB(MAE) = 100\% \cdot mean\left(I\left\{MAE < MAE^*\right\}\right). \tag{26}$$

where $I\{\bullet\}$ - the operator that yields the value of zero or one, in accordance with the expression:

$$I(MAE) = \begin{cases} 0, if\ MAE < MAE^*; \\ \quad\quad 1. \end{cases} \tag{27}$$

By analogy with PB (MAE), Percentage Better (MSE) can be defined.

The disadvantages of these measures are the following.

- Division by zero error occurs if the reference forecast error is equal to zero.
- These criteria determine the number of cases when the analyzed forecasting model superior to the base but do not evaluate the value of difference.

Other Error Measures: This group includes measures proposed in various studies to avoid the shortcomings of existing and common measures.

To avoid the scale dependency, Normalized Root Mean Square Error (nRMSE) has been proposed, calculated by the formula:

$$nRMSE = \frac{1}{\bar{y}}\sqrt{\underset{i=\overline{1,n}}{mean}\left(e_i^2\right)}, \tag{28}$$

where $\bar{y}$ - the normalization factor, which is usually equal to either the maximum measured value on the forecast horizon, or the difference between the maximum and minimum values. Normalization factor can be calculated over the entire interval or time horizon or defined short interval of observation [18]. However, this estimate is affected by influence of outliers, if outlier has a value much bigger the maximal "normal" value. To reduce the impact of outliers, Integral Normalized Mean Square Error [19] have been proposed, calculated by the formula:

$$inRSE = \frac{1}{\sum_{i=1}^{n} y_i}\sqrt{\underset{i=\overline{1,n}}{mean}\left(e_i^2\right)}. \tag{29}$$

Some research contains the the ways of *NRMSE* calculation as [16]:

$$inRSE = \sqrt{\frac{\sum_{i=1}^{n}\left(e_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}, \tag{30}$$

where $\bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_k$.

Other measures are called normalized std_APE and std_MAPE [20, 21] and calculated by the formula

$$Std\_AE = \sqrt{\frac{\sum_{i=1}^{n}\left(e_i - MAE\right)^2}{n-1}} \tag{31}$$

and

$$Std\_APE = \sqrt{\frac{\sum_{i=1}^{n}\left(p_i - MAPE\right)^2}{n-1}} \tag{32}$$

respectively.

As a drawback, you can specify a division by zero error if normalization factor is equal to zero.

Recommendations How to Choose Error Measures: One of the most difficult issues is the question of choosing the most appropriate measures out of the groups. Due to the fact that each error measure has the disadvantages that can lead to inaccurate evaluation of the forecasting results, it is impossible to choose only one measure [5].

We provide the following guidelines for choosing the error measures.

- If forecast performance is evaluated for time series with the same scale and the data preprocessing procedures were performed (data cleaning, anomaly detection), it is reasonable to choose MAE, MdAE, RMSE. In case of different scales, these error measures are not applicable. The following recommendations are provided for mutli-scales cases.
- In spite of the fact that percentage errors are commonly used in real world forecast tasks, but due to the non-symmetry, they are not recommended. If the range of the values lies in the positive half-plane and there are no outliers in the data, it is advisable to use symmetric error measures.
- If the data are "dirty", i.e. contain outliers, it is advisable to apply the scaled measures such as MASE, inRSE. In this case (i) the horizon should be large enough, (ii) no identical values should be, (iii) the normalized factor should be not equal to zero.
- If predicted data have seasonal or cyclical patterns, it is advisable to use the normalized error measures, wherein the normalization factors could be calculated within the interval equal to the cycle or season.
- If there is no results of prior analysis and a-prior information about the quality of the data, it reasonable to use the defined set of error measures. After calculating, the results are analyzed with respect to division by zero errors and contradiction cases:
  - For the same time series the results for model m_1 is better than m_2, based on the one error measure, but opposite for another one;
  - For different time series the results for model m_1 is better in most cases, but worst for a few of cases.

CONCLUSION

The review contains the error measures for time series forecasting models. All these measures are grouped into seven groups: absolute forecasting error, percentage forecasting error, symmetrical forecasting error, measures based on relative errors, scaled errors, relative errors and other (modified). For each error measure the way of calculation is presented. Also shortcomings are defined for each of group.

ACKNOWLEDGMENTS

Authors would like to thank RFBR for support of the research (Grants #12-07-31017, 12-01-00684).

REFERENCES

1. Tyukov, A., A. Brebels, M. Shcherbakov and V. Kamaev, 2012. A concept of web-based energy data quality assurance and control system. In the Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services, pp: 267-271.
2. Kamaev, V.A., M.V. Shcherbakov, D.P. Panchenko, N.L. Shcherbakova and A. Brebels, 2012. Using Connectionist Systems for Electric Energy Consumption Forecasting in Shopping Centers. Automation and Remote Control, 73(6): 1075-1084.
3. Owoeye, D., M. Shcherbakov and V. Kamaev, 2013. A photovoltaic output backcast and forecast method based on cloud cover and historical data. In the Proceedings of the The Sixth IASTED Asian Conference on Power and Energy Systems (AsiaPES 2013), pp: 28-31.
4. Armstrong, J.S. and F. Collopy, 1992. Error measures for generalizing about forecasting methods: Empirical comparisons. International Journal of Forecasting, 8(1): 69-80.
5. Mahmoud, E., 1984. Accuracy in forecasting: A survey. Journal of Forecasting, 3(2): 139-159.
6. Yokuma, J.T. and J.S. Armstrong, 1995. Beyond accuracy: Comparison of criteria used to select forecasting methods. International Journal of Forecasting, 11(4): 591-597.
7. Armstrong, J.S., 2001. Evaluating forecasting methods. In Principles of forecasting: a handbook for researchers and practitioners. Norwell, MA: Kluwer Academic Publishers, pp: 443-512.
8. Gooijer, J.G.D. and R.J. Hyndman, 2006. 25 years of time series forecasting. International Journal of Forecasting, 22(3): 443-473.
9. Hyndman, R.J. and A.B. Koehler, 2006. Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4): 679-688.
10. Theodosiou, M., 2011. Forecasting monthly and quarterly time series using STL decomposition. International Journal of Forecasting, 27(4): 1178-1195.

11. Shcherbakov, M.V. and A. Brebels, 2011. Outliers and anomalies detection based on neural networks forecast procedure. In the Proceedings of the 31st Annual International Symposium on Forecasting, ISF-2011, pp: 21-22.
12. Goodwin, P. and R. Lawton, 1999. On the asymmetry of the symmetric MAPE. International Journal of Forecasting, 15(4): 405-408.
13. Koehler, A.B., 2001. The asymmetry of the sAPE measure and other comments on the M3-competition. International Journal of Forecasting, 17: 570-574.
14. Makridakis, S., 1993. Accuracy measures: Theoretical and practical concerns. International Journal of Forecasting, 9: 527-529.
15. Kolassa, S. and R. Martin, 2011. Percentage errors can ruin your day (and rolling the dice shows how). Foresight, (Fall): 21-27.
16. Assessing Forecast Accuracy Measures. Date View 01.08.2013 http:// www.stat.iastate.edu/ preprint/ articles/2004-10.pdf
17. Syntetos, A.A. and J.E. Boylan, 2005. The accuracy of intermittent demand estimates. International Journal of Forecasting, 21(2): 303-314.
18. Tyukov, A., M. Shcherbakov and A. Brebels, 2011. Automatic two way synchronization between server and multiple clients for HVAC system. In the Proceedings of The 13th International Conference on Information Integration and Web-based Applications & Services, pp: 467-470.
19. Brebels, A., M.V. Shcherbakov and V.A. Kamaev, 2010. Mathematical and statistical framework for comparison of neural network models with other algorithms for prediction of Energy consumption in shopping centres. In the Proceedings of the 37 Int. Conf. Information Technology in Science Education Telecommunication and Business, suppl. to Journal Open Education, pp: 96-97.
20. Casella, G. and R. Berger, 1990. Statistical inference. 2nd ed. Duxbury Press, pp: 686.
21. Kusiak, A., M. Li and Z. Zhang, 2010. A data-driven approach for steam load prediction in buildings. Applied Energy, 87(3): 925-933.