International Journal of Forecasting 32 (2016) 20–22

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

A note on the Mean Absolute Scaled Error

Philip Hans Franses *

Econometric Institute, Erasmus School of Economics, The Netherlands

ARTICLE INFO

Keywords:
Forecast accuracy
Forecast error measures
Statistical testing

ABSTRACT

Hyndman and Koehler (2006) recommend that the Mean Absolute Scaled Error (MASE) should become the standard when comparing forecast accuracies. This note supports their claim by showing that the MASE fits nicely within the standard statistical procedures initiated by Diebold and Mariano (1995) for testing equal forecast accuracies. Various other criteria do not fit, as they do not imply the relevant moment properties, and this is illustrated in some simulation experiments.

1. Introduction

Consider the case where an analyst has two competing one-step-ahead forecasts for a time series variable y_t, namely $\hat{y}_{1,t}$ and $\hat{y}_{2,t}$, for a sample $t = 1, 2, \ldots, T$. The forecasts have the associated forecast errors $\hat{\varepsilon}_{1,t}$ and $\hat{\varepsilon}_{2,t}$, respectively. To examine which of the two sets of forecasts provides the best accuracy, the analyst can use criteria based on some average or median of loss functions of the forecast errors. Well-known examples include the root mean squared error (RMSE) and the median absolute error (MAE); see Hyndman and Koehler (2006) for an exhaustive list of criteria, and also Table 1.

As there is always one set of forecasts that scores lower on some criterion, it seems wise to test whether any observed differences in forecast performances are statistically significant. To test statistically whether the obtained values of these criteria are equal, the analyst can rely on the methodology proposed by Diebold and Mariano (1995) (DM); see also Diebold (2015) for a recent review. This methodology is based on the loss functions $l_{i,t} = f(y_t, \hat{y}_{i,t})$ for $i = 1, 2$. Denoting the sample mean loss differential by $\bar{d}_{12}$, that is, $\bar{d}_{12} = \frac{1}{T}\sum_1^T (l_{1,t} - l_{2,t})$, and a consistent estimate of the standard deviation of $\bar{d}_{12}$ by $\hat{\sigma}_{\bar{d}_{12}}$, the DM test for one-step-ahead forecasts is

$$DM = \frac{\bar{d}_{12}}{\hat{\sigma}_{\bar{d}_{12}}} \sim N(0, 1),$$

under the null hypothesis of equal forecast accuracy. Even though Diebold and Mariano (1995, p. 254) claim that this result holds for any arbitrary function f, it is quite clear that the function should allow for proper moment conditions in order to yield the asymptotic normality of the test. In fact, as will be argued in Section 2 below, many of the functions that are commonly applied in the forecast literature fail to qualify as useful functions for the DM methodology.

This note continues with a brief summary of typical functions in Section 2, along with a concise discussion of which of these functions are useful in the DM framework. It is found that the absolute scaled error (ASE) recommended by Hyndman and Koehler (2006) does have the favorable properties, while various other criteria do not. Section 3 reports on limited simulation experiments which support these insights. The main conclusion of this note is to confirm that the use of the Mean ASE (MASE) criterion is recommended.

2. Loss functions of realizations and forecasts

Hyndman and Koehler (2006) provide an exhaustive list of the loss functions of realizations and forecasts, and a

* Correspondence to: Econometric Institute, Erasmus School of Economics, POB 1738, NL-3000 DR Rotterdam, The Netherlands.
E-mail address: franses@ese.eur.nl.

http://dx.doi.org/10.1016/j.ijforecast.2015.03.008
0169-2070/

Table 1
Various criteria for comparing forecasts and realizations. The references are provided by Hyndman and Koehler (2006).

In words	Loss function	Summary statistics
Squares	$\hat{\varepsilon}_{i,t}^2$	Mean squared error (MSE) Root MSE (RMSE)
Absolute	$\lvert\hat{\varepsilon}_{i,t}\rvert$	Mean absolute error (MAE) Median AE
Absolute scaled	$\left\lvert\frac{\hat{\varepsilon}_{i,t}}{\frac{1}{T}\sum_{t=2}^{T}\lvert y_t-y_{t-1}\rvert}\right\rvert$	Mean absolute scaled error (MASE)
Relative to random walk	$\left\lvert\frac{\hat{\varepsilon}_{i,t}}{y_t-y_{t-1}}\right\rvert$	Mean relative absolute error (MRAE) Median RAE Geometric mean RAE
Symmetric absolute percentage	$200\frac{\lvert y_t-\hat{y}_{i,t}\rvert}{\lvert y_t+\hat{y}_{i,t}\rvert}$	Symmetric mean absolute percentage error
Absolute percentage	$\left\lvert\frac{100\hat{\varepsilon}_{i,t}}{y_t}\right\rvert$	Mean absolute percentage error (MAPE) Median APE Root mean squared percentage error Root median squared percentage error

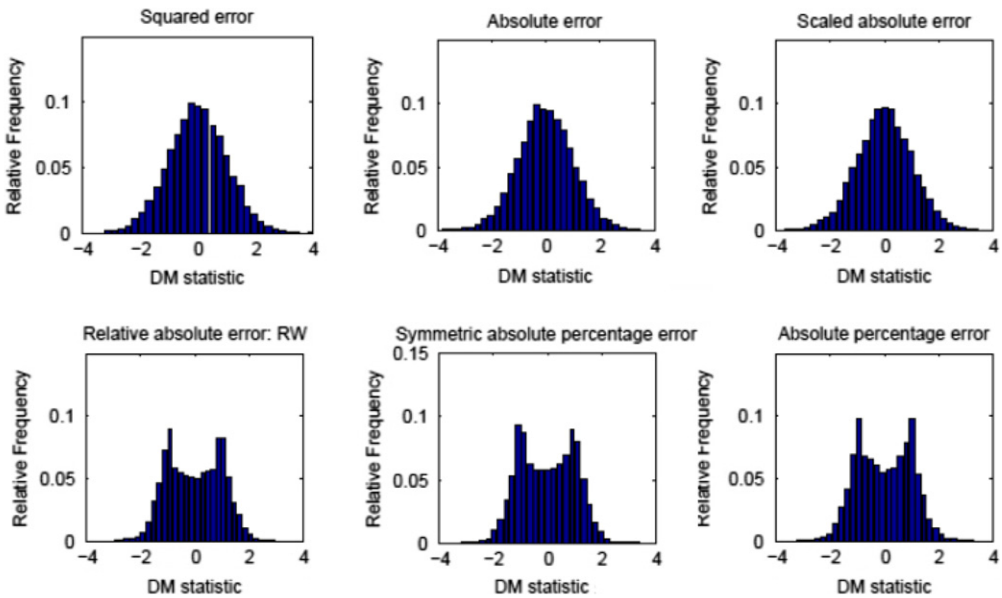

Fig. 1. Empirical distribution of the DM test statistic, DGP1.

concise summary of them is presented in Table 1 for convenience. Basically, there are six distinct loss functions $l_{i,t} = f(y_t, \hat{y}_{i,t})$ that are commonly used in a variety of criteria.

The squared error and the absolute error have moment properties that match the assumptions underlying the asymptotic theory of the DM test. The same obviously holds for the absolute scaled error, as the forecast errors are all divided by the same number, and hence these absolute scaled errors have the same moment properties as the absolute error.

In contrast, the second set of three functions of realizations and forecasts in Table 1 do not have these nice properties. The random walk forecasts can become very close to zero, and hence, the errors scaled by random walk forecasts have infinite moments. Only in very well-behaved cases may the asymptotic distribution become a Cauchy distribution. The asymptotic normality of the DM test is guaranteed provided that the first two moments exist and are finite, because of the central limit theorem, and these conditions do not hold for the symmetric absolute percentage and absolute percentage error loss functions.

In summary, one can only expect the familiar DM test to have an asymptotic $N(0, 1)$ distribution for squared errors, absolute errors and absolute scaled errors.

3. Simulations

To determine whether the above arguments hold, consider the following simple simulation experiment. Assume the data generating process (DGP):

$$y_t = 5x_{1,t} + 5x_{2,t} + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, 0.25)$ and

DGP1: $x_{1,t} \sim N(0, 1)$ and $x_{2,t} \sim N(0, 1)$

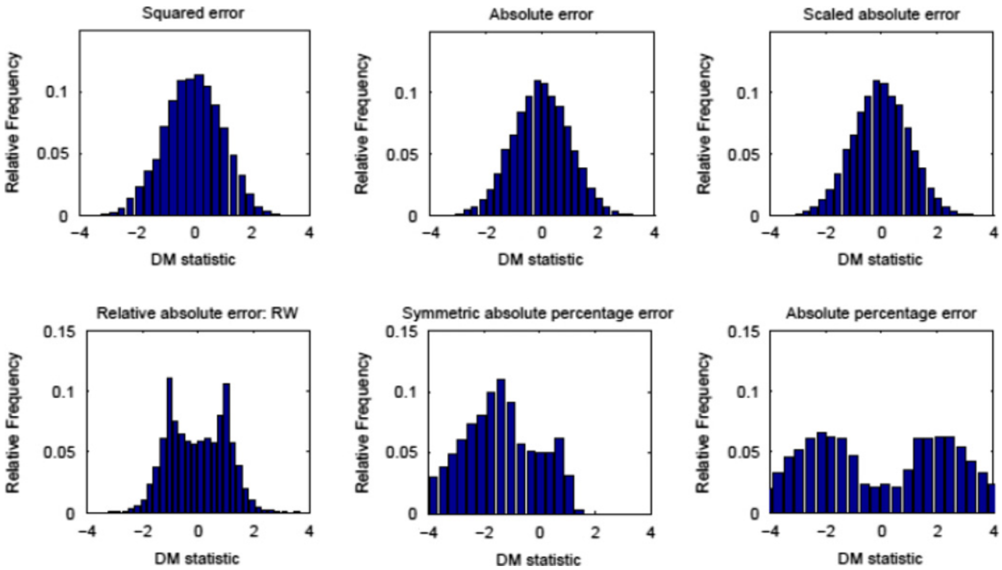

Fig. 2. Empirical distribution of the DM test statistic, DGP2.

DGP2: $x_{1,t} \sim N(10, 1)$ and $x_{2,t} \sim N(-10, 1)$.

The data are created for $t = 1, 2, \ldots, 2000$, where the first 1000 observations are used to estimate

Model 1: $y_t = \beta_1 x_{1,t} + \varphi_{1,t}$

Model 2: $y_t = \beta_2 x_{2,t} + \varphi_{2,t}$.

Then, one-step-ahead forecasts are created using a recursive procedure. This procedure entails the estimation of the parameters using the first T observations, after which a forecast for $T + 1$ is created using the estimated parameters for T observations, and the actual true value of $x_{1,t}$ or $x_{2,t}$ at $T + 1$. Then, the sample is shifted to 1001 observations and the procedure is repeated. In the end, there are 1000 one-step-ahead forecasts $\hat{y}_{1,t}$ and $\hat{y}_{2,t}$.

The DM test value is computed using the six loss functions in Table 1. This procedure is repeated 10,000 times, and the corresponding empirical distributions of the DM test statistics are created. Figs. 1 and 2 present the results. Evidently, one can observe an empirical $N(0, 1)$ distribution for squared errors, absolute errors and absolute scaled errors, but the distributions in the bottom panel do not come near a $N(0, 1)$ distribution. Also, these latter three distributions vary across the two DGPs as well, suggesting that the associated DM test statistic does not have a unique distribution under the null hypothesis.

Acknowledgments

Thanks to the Editor, an anonymous Associate Editor and an anonymous reviewer for helpful comments, and to Victor Hoornweg for excellent research assistance.

References

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: a personal perspective on the use and abuse of the Diebold–Mariano tests. *Journal of Business & Economic Statistics*, *33*(1), 1–9.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 679–688.

Philip Hans Franses is a Professor of Applied Econometrics and Professor of Marketing Research, both affiliated with the Erasmus School of Economics. His research interests are in time series, forecasting, marketing and empirical finance.