

A note on the Mean Absolute Scaled Error

Philip Hans Franses

*Econometric Institute
Erasmus School of Economics*

Abstract

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

Keywords: Forecast accuracy, Forecast error measures, Statistical testing

This revised version: February 2015

Address for correspondence: Econometric Institute, Erasmus School of Economics, POB 1738, NL-3000 DR Rotterdam, the Netherlands, franses@ese.eur.nl

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

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 $t = 1, 2, \dots, T$. The forecasts bring along the forecast errors $\hat{\varepsilon}_{1,t}$ and $\hat{\varepsilon}_{2,t}$, respectively. To examine which of the two sets of forecasts provides most accuracy, the analyst can use criteria based on some average or median of loss functions of the forecast errors. Well-known examples are the Root Mean Squared Error (RMSE) or the Median Absolute Error (MAE), see Hyndman and Koehler (2006) for an exhaustive list of criteria and see also Table 1 below.

As there is always one set of forecasts that scores lower on some criterion, it seems wise to test if any observed difference in forecast performance is statistically significant. To statistically test whether the obtained values of these criteria the analyst can rely on the methodology initiated in Diebold and Mariano (1995) (DM), see for a recent review Diebold (2013). This methodology is based on the loss functions $l_{i,t} = f(y_t, \hat{y}_{i,t})$ for $i = 1, 2$. Denote \bar{d}_{12} as the sample mean loss differential, that is, $\bar{d}_{12} = \frac{1}{T} \sum_1^T (l_{1,t} - l_{2,t})$ and $\hat{\sigma}_{\bar{d}_{12}}$ as a consistent estimate of the standard deviation of \bar{d}_{12} , then 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, page 254) claim that this result holds for any arbitrary function f , it is quite clear that the function should allow for proper moment conditions to yield the asymptotic normality of the test. In fact, as will be argued in Section 2 below, many often applied functions in the forecast literature disqualify as useful functions for the DM methodology.

This note continues with a brief summary of typical functions in Section 2, and with a concise discussion which of these functions are useful in the DM framework. It is found that the Absolute Scaled Error (ASE) recommended in 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 from this note is that the use of the Mean ASE (MASE) criterion is recommended indeed.

Functions of realizations and forecasts

Hyndman and Koehler (2006) provide an exhaustive list of loss functions of realizations and forecasts, and for convenience a concise summary is presented in Table 1. Basically, there are six distinct loss functions $l_{i,t} = f(y_t, \hat{y}_{i,t})$ which are commonly used in a variety of criteria.

The squared error and absolute error have moment properties that match with 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, the asymptotic distribution may become a Cauchy distribution. The asymptotic normality of the DM test is guaranteed provided the first two moments exist and are finite because of the central limit theorem. These conditions do not hold for the loss functions symmetric absolute percentage and the absolute percentage error.

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

Simulations

To examine if the above arguments hold, consider the following simple simulation experiment. Assume the following data generating process (DGP), that is

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

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

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

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

The data are created for $t = 1, 2, \dots, 2000$. The first 1000 observations are used to estimate

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

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

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

The DM test value is computed using the six loss functions as in Table 1. This procedure is repeated 10000 times, and the corresponding empirical distributions of the DM test statistics are created. Figures 1 and 2 present the results. Evidently, for squared errors, absolute errors and absolute scaled errors, one can observe an empirical $N(0,1)$ distribution, whereas the bottom panel shows distributions that do not come near a $N(0,1)$ distribution. Also, these latter three distributions also vary across the two DGPs, suggesting that the associated DM test statistic does not have a unique distribution under the null hypothesis.

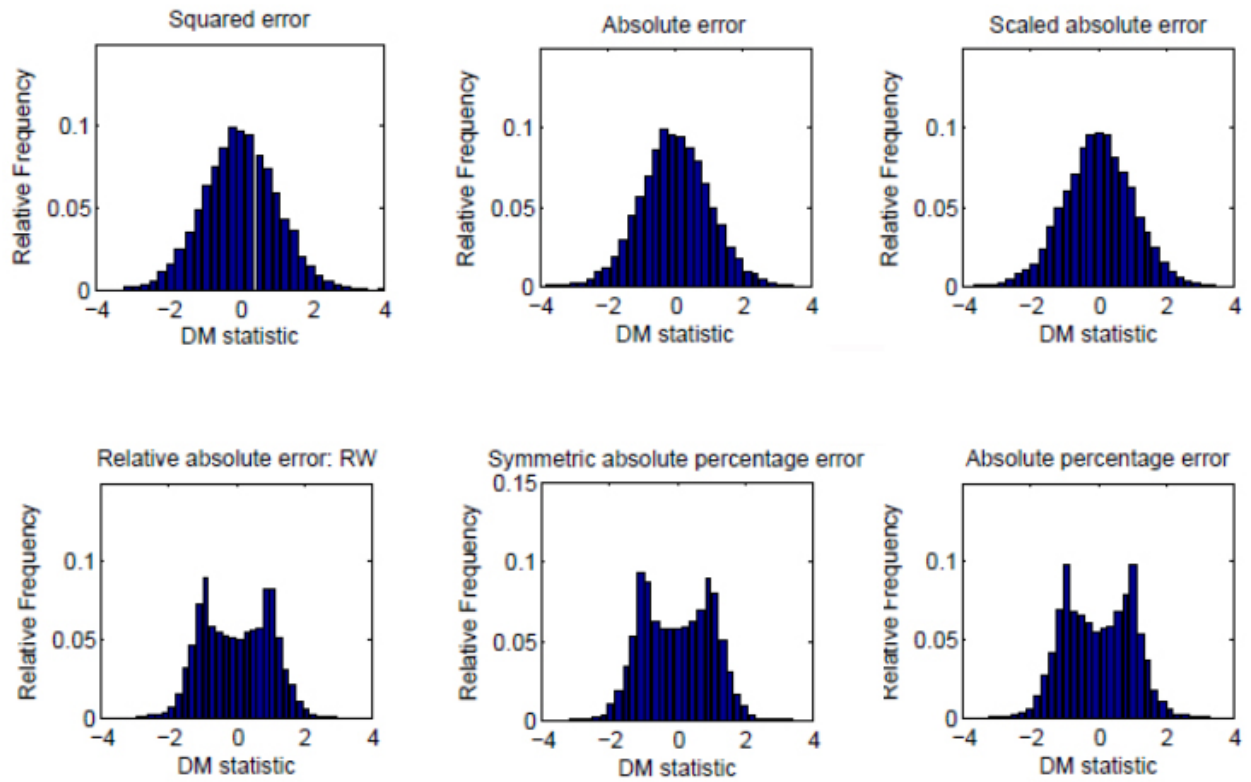


Figure 1: Empirical distribution of the DM test statistic, DGP1

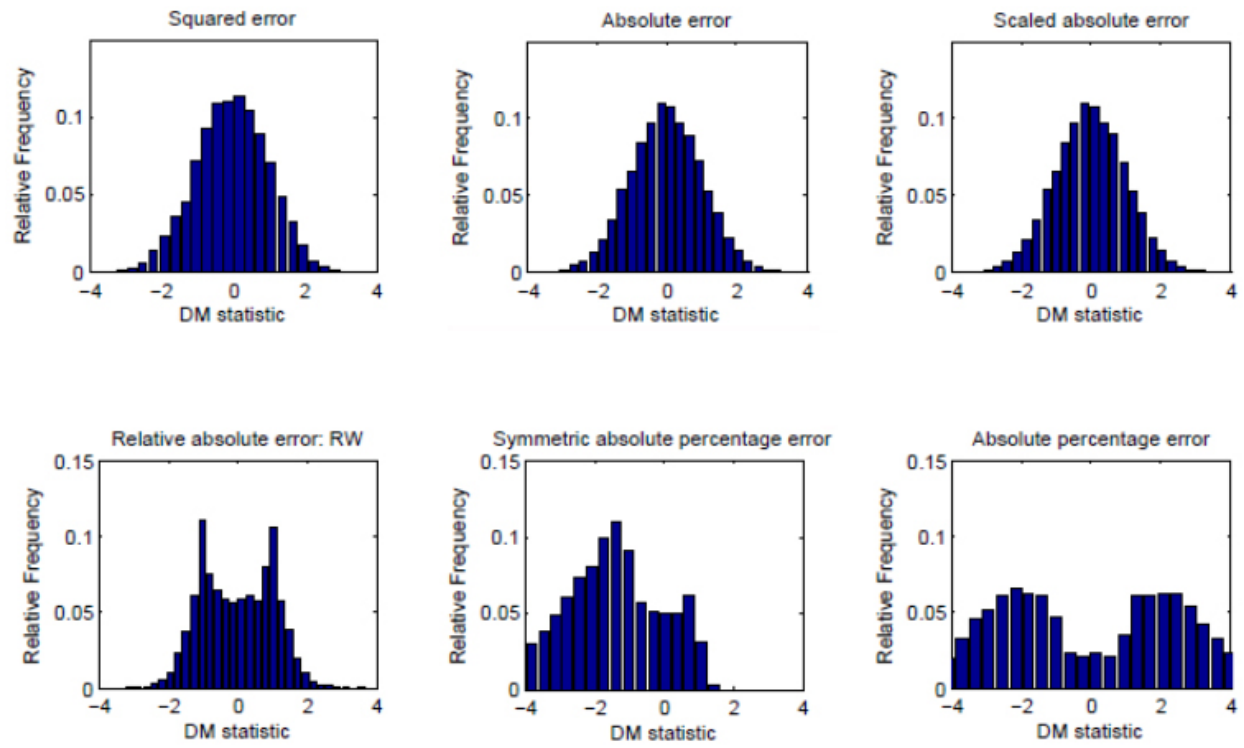


Figure 2: Empirical distribution of the DM test statistic, DGP2

Table 1: Various criteria to compare forecasts and realizations. References are given in Hyndman and Koehler (2006).

In words	Loss function	Summary statistics
<i>Squares</i>	$\hat{\varepsilon}_{i,t}^2$	Mean Squared Error (MSE) Root MSE (RMSE)
<i>Absolute</i>	$ \hat{\varepsilon}_{i,t} $	Mean Absolute Error (MAE) Median AE
<i>Absolute Scaled</i>	$\left \frac{\hat{\varepsilon}_{i,t}}{\frac{1}{T} \sum_{t=2}^T y_t - y_{t-1} } \right $	Mean Absolute Scaled Error (MASE)
<i>Relative to random walk</i>	$\left \frac{\hat{\varepsilon}_{i,t}}{y_t - y_{t-1}} \right $	Mean Relative Absolute Error (MRAE) Median RAE Geometric Mean RAE
<i>Symmetric Absolute Percentage</i>	$200 \frac{ y_t - \hat{y}_{i,t} }{ y_t + \hat{y}_{i,t} }$	Symmetric Mean Absolute Percentage Error
<i>Absolute Percentage</i>	$\left \frac{100 \hat{\varepsilon}_{i,t}}{y_t} \right $	Mean Absolute Percentage Error (MAPE) Median APE Root Mean Squared Percentage Error Root Median Squared Percentage Error

References

Diebold, Frank X. (2013), Comparing predictive accuracy, Twenty years later: A personal perspective on the use and abuse of the Diebold-Mariano tests, JBES Invited lecture ASSA meetings 2014, University of Pennsylvania, Working paper.

Diebold, Frank X. and Robert S. Mariano (1995), Comparing predictive accuracy, Journal of Business and Economic Statistics, 13, 253-263.

Hyndman, Rob J. and Anne B. Koehler (2006), Another look at measures of forecast accuracy, International Journal of Forecasting, 22, 679-688.