Lesson19: Comparing Predictive Accuracy of two Forecasts: The Diebold-Mariano Test

Umberto Triacca

Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica
Università dell'Aquila,
umberto.triacca@univaq.it

The Diebold-Mariano Test

In empirical applications it is often the case that two or more time series models are available for forecasting a particular variable of interest.

- actual values $\{y_t; t = 1, ..., T\}$
- two forecasts: $\{\hat{y}_{1t}; t = 1, ..., T\}$ $\{\hat{y}_{2t}; t = 1, ..., T\}$

Question: Are the forecasts equally good?

The Diebold-Mariano Test

Define the forecast errors as

$$e_{it} = \hat{y}_{it} - y_t, \quad i = 1, 2$$

The loss associated with forecast i is assumed to be a function of the forecast error, e_{it}, and is denoted by $g(e_{it})$.
The function $g(.)$ is a loss function, that is a function such that

1. takes the value zero when no error is made;
2. is never negative;
3. is increasing in size as the errors become larger in magnitude.

Typically, $g(e_{it})$ is the square (squared-error loss) or the absolute value (absolute error loss) of e_{it}.

The Diebold-Mariano Test

$$g(e_{it}) = e_{it}^2$$

$$g(e_{it}) = \left| e_{it}^2 \right|$$

A problem with these loss function is that they are symmetric functions. In fact, in some case, the symmetry between positive and negative forecast errors could be inappropriate. When it is more costly to underpredict y_t than to overpredict it, the following loss function can be used

$$g(e_{it}) = \exp(\lambda e_{it}) - 1 - \lambda e_{it}$$

where λ is some positive constant.

The Diebold-Mariano Test

We define the loss differential between the two forecasts by

$$d_t = g(e_{1t}) - g(e_{2t})$$

and say that the two forecasts have equal accuracy if and only if the loss differential has zero expectation for all t.

The Diebold-Mariano Test

So, we would like to test the null hypothesis

$$H_0 : E(d_t) = 0 \ \ \forall t$$

versus the alternative hypothesis

$$H_1 : E(d_t) \neq 0$$

The null hypothesis is that the two forecasts have the same accuracy. The alternative hypothesis is that the two forecasts have different levels of accuracy

The Diebold-Mariano Test

Consider the quantity

$$\sqrt{T}\left(\bar{d}-\mu\right)$$

where

$$\bar{d}=\sum_{t=1}^{T} d_t$$

is the sample mean of the loss differential,

$$\mu = E(d_t).$$

is the population mean of the loss differential,

$$f_d(0)=\frac{1}{2\pi}\left(\sum_{k=-\infty}^{\infty}\gamma_d(k)\right)$$

is the spectral density of the loss differential at frequency 0, $\gamma_d(k)$ is the autocovariance of the loss differential at lag k.

The Diebold-Mariano Test

It is possible to show that if the loss differential series $\{d_t; t = 1, ..., T\}$ is covariance stationary and short memory, then

$$\sqrt{T}\left(\bar{d} - \mu\right) \rightarrow N(0, 2\pi f_d(0))$$

In the sequel we will assume that the loss differential series $\{d_t; t = 1, ..., T\}$ is covariance stationary and short memory.

The Diebold-Mariano Test

$$\sqrt{T}\left(\bar{d}-\mu\right) \rightarrow N(0, 2\pi f_d(0))$$

$$\Downarrow$$

$$\frac{\bar{d}-\mu}{\sqrt{\frac{2\pi f_d(0)}{T}}} \rightarrow N(0,1)$$

The Diebold-Mariano Test

Under H_0

$$\frac{\bar{d}}{\sqrt{\frac{2\pi f_d(0)}{T}}} \to N(0,1)$$

The Diebold-Mariano Test

Suppose that the forecasts are $h(>1)$-step-ahead. In order to test the null hypotesis that the two forecasts have the same accuracy, Diebold-Mariano utilize the following statistic

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}}$$

where $\hat{f}_d(0)$ is a consistent estimate of $f_d(0)$ defined by

$$\hat{f}_d(0) = \frac{1}{2\pi} \sum_{k=-(T-1)}^{T-1} I\left(\frac{k}{h-1}\right) \hat{\gamma}_d(k)$$

The Diebold-Mariano Test

where

$$\hat{\gamma}_d(k) = \frac{1}{T}\sum_{t=|k|+1}^{T}\left(d_t - \bar{d}\right)\left(d_{t-|k|} - \bar{d}\right)$$

and

$$I\left(\frac{k}{h-1}\right) = \begin{cases} 1 \ \text{ for } \ \left|\frac{k}{h-1}\right| \leq 1 \\ 0 \ \text{ otherwise} \end{cases}$$

The Diebold-Mariano Test

We note that

$$\hat{\gamma}_d(-k) = \frac{1}{T}\sum_{t=|-k|+1}^{T}\left(d_t - \bar{d}\right)\left(d_{t-|-k|} - \bar{d}\right) = \hat{\gamma}_d(k)$$

and that

$$I\left(\frac{k}{h-1}\right) = 0 \ \text{ for } \ |k| > h-1$$

Thus

$$\hat{f}_d(0) = \frac{1}{2\pi}\left(\hat{\gamma}_d(0) + 2\sum_{k=1}^{h-1}\hat{\gamma}_d(k)\right)$$

The Diebold-Mariano Test

If $h = 1$

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}}$$

where $\hat{f}_d(0)$ is a consistent estimate of $f_d(0)$ defined by

$$\hat{f}_d(0) = \frac{1}{2\pi}\hat{\gamma}_d(0)$$

The Diebold-Mariano Test

Hence, for $h \geq 1$, we have

$$DM = \frac{\bar{d}}{\sqrt{\frac{\hat{\gamma}_d(0)+2\sum_{k=1}^{h-1}\hat{\gamma}_d(k)}{T}}}$$

In practice, using

$$\sum_{k=-M}^{M} \hat{\gamma}_d(k),$$

where $M = T^{1/3}$, provides an adegauate estimator of $2\pi f_d(0)$ in many cases. Thus

$$DM = \frac{\bar{d}}{\sqrt{\frac{\sum_{k=-M}^{M}\hat{\gamma}_d(k)}{T}}}$$

The Diebold-Mariano Test

Under the null hypothesis, the test statistics DM is asymptotically $N(0,1)$ distributed. The null hypothesis of no difference will be rejected if the computed DM statistic falls outside the range of $-z_{\alpha/2}$ to $z_{\alpha/2}$, that is if

$$|DM| > z_{\alpha/2},$$

where $z_{\alpha/2}$ is the upper (or positive) z-value from the standard normal table corresponding to half of the desired α level of the test.

The Diebold-Mariano Test

Suppose that the significance level of the test is $\alpha = 0.05$.
Because this is a two-tailed test 0.05 must be split such that 0.025 is in the upper tail and another 0.025 in the lower. The z-value that corresponds to -0.025 is -1.96, which is the lower critical z-value. The upper value corresponds to 1-0.025, or 0.975, which gives a z-value of 1.96.
The null hypothesis of no difference will be rejected if the computed *DM* statistic falls outside the range of -1.96 to 1.96.

The Diebold-Mariano Test

As the simulation experiments in Diebold and Mariano (1995) show, the normal distribution can be a very poor approximation of the DM test's finite-sample null distribution. Their results show that the DM test can have the wrong size, rejecting the null too often, depending on the degree of serial correlation among the forecast errors and the sample size, T.

The Diebold-Mariano Test

Harvey, Leybourne, and Newbold (1997) (HLN) suggest that improved small-sample properties can be obtained by:
(i) making a bias correction to the DM test statistic, and

(ii) comparing the corrected statistic with a Student-t distribution with (T-1) degrees of freedom, rather than the standard normal.

The corrected statistic is obtained as

$$HLN - DM = \sqrt{\frac{T + 1 - 2h + h(h-1)}{T}} DM$$

The Diebold-Mariano Test

A problem: The Diebold-Mariano test should not be applied to situations where the competing forecasts are obtained using two nested models
What are the reasons for this?
The root of the problem is that, at the population level, if the null hypothesis of equal predictive accuracy is true, the forecast errors from the competing models are exactly the same and perfectly correlated, which means that the numerator and denominator of a Diebold-Mariano test are each limiting to zero as the estimation sample and prediction sample grow.

The Diebold-Mariano Test

However, when the size of the estimation sample remains finite as the size of the prediction sample grows, parameter estimates are prevented from reaching their probability limits and the Diebold-Mariano test remains asymptotically valid even for nested models, under some regularity assumptions (see Giacomini and White 2003).

Essentially, this means that model parameters are estimated using a rolling window of data, rather than an expanding one.

The Diebold-Mariano Test

The conventional *DM* test is conservative when applied to short-horizon forecasts.

The Diebold-Mariano Test

References

Diebold, F.X. and R.S. Mariano. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13: 253-63.

Giacomini, R. and H. White (2003). Tests of Conditional Predictive Ability. Working Paper No. 572, Boston College.

Harvey, D., S. Leybourne, and P. Newbold. (1997). Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*, 13: 281-91.