Forecast Evaluation Statistics

Let $\{y_t\}$ denote the series to be forecast and let $y_{t+h|t}$ denote the out-of-sample forecasts of $y_{t+h}$ based on $I_t$.

Out-of-sample forecasts are typically computed using one of two methods.

- **Recursive (expanding window) forecasts**: An initial sample using data from $t = 1, \ldots, T$ is used to estimate the models, and $h$–step ahead out-of-sample forecasts are produced starting at time $T$. The sample is increased by one, the models are re-estimated, and $h$–step ahead forecasts are produced starting at $T + 1$.

  $$ [1, \ldots, t, \ldots, t + h] $$
  $$ [1, \ldots, t + 1, \ldots, t + h + 1] $$
  $$ \vdots $$
  $$ [1, \ldots, t + N, \ldots \ldots \ldots, t + h + N] $$
Rolling (moving window) forecasts. An initial sample using data from \( t = 1, \ldots, T \) is used to determine a window width \( T \), to estimate the models, and to form \( h \)-step ahead out-of-sample forecasts starting at time \( T \). Then the window is moved ahead one time period, the models are re-estimated using data from \( t = 2, \ldots, T + 1 \), and \( h \)-step ahead out-of-sample forecasts are produced starting at time \( T + 1 \).

\[
[1, \ldots, t, \ldots, t + h] \\
[2, \ldots, t + 1, \ldots, t + h + 1] \\
\vdots \\
[N, \ldots, t + N, \ldots, t + h + N]
\]
Traditional Forecast Evaluation Statistics

• Let \( y_{t+h|t} \) denote the \( h \)-step ahead forecast of \( y_{t+h} \) based on recursive or rolling methods.

• Define the corresponding forecast error as \( e_{t+h|t} = y_{t+h} - y_{t+h|t} \).

• Common forecast evaluation statistics based on \( N \) \( h \)-step ahead forecasts

\[
\begin{align*}
\text{MSE} &= \frac{1}{N} \sum_{j=t+1}^{t+N} e_{j+h|j}^2, \\
\text{MAE} &= \frac{1}{N} \sum_{j=t+1}^{t+N} |e_{j+h|j}|, \\
\text{MAPE} &= \frac{1}{N} \sum_{j=t+1}^{t+N} \frac{|e_{j+h|j}|}{y_{j+h}}.
\end{align*}
\]
• For $h > 1$ the forecast errors $\{e_{j+h|j}^2\}_{t+N}^{t+1}$ are serially correlated and follow an MA($h - 1$) process.

• A model which produces small values of the forecast evaluation statistics is judged to be a good model.

• Of course, the forecast evaluation statistics are random variables and a formal statistical procedure should be used to determine if they are “small”.
Diebold-Mariano Test for Equal Predictive Accuracy

Let \( \{y_t\} \) denote the series to be forecast and let \( y_{t+h|t}^1 \) and \( y_{t+h|t}^2 \) denote two competing forecasts of \( y_{t+h} \) based on \( I_t \). For example, \( y_{t+h|t}^1 \) could be computed from an AR\((p)\) model and \( y_{t+h|t}^2 \) could be computed from an ARMA\((p, q)\) model. The forecast errors from the two models are

\[
\varepsilon_{t+h|t}^1 = y_{t+h} - y_{t+h|t}^1
\]
\[
\varepsilon_{t+h|t}^2 = y_{t+h} - y_{t+h|t}^2
\]

The \( h \)-step forecasts are assumed to be computed for \( j = t, \ldots, t + N \) for a total of \( N \) forecasts giving

\[
\{\varepsilon_{j+h|j}^1\}_{t}^{t+N}, \{\varepsilon_{j+h|j}^2\}_{t}^{t+N}
\]
Note: because the $h$-step forecasts use overlapping data the forecast errors in

$$\{\varepsilon_{j+h|j}^{1}\}_{t}^{t+N} \text{ and } \{\varepsilon_{j+h|j}^{2}\}_{t}^{t+N}$$

will be serially correlated.
The accuracy of each forecast is measured by a particular loss function

\[ L(y_{t+h}, y^i_{t+h|t}) = L(\varepsilon^i_{t+h|t}), \ i = 1, 2 \]

Some popular loss functions are:

\[ L(\varepsilon^i_{t+h|t}) = (\varepsilon^i_{t+h|t})^2 : \text{ squared error loss} \]
\[ L(\varepsilon^i_{t+h|t}) = |\varepsilon^i_{t+h|t}| : \text{ absolute value loss} \]

To determine if one model predicts better than another we may test null hypotheses

\[ H_0 : E[L(\varepsilon^1_{t+h|t})] = E[L(\varepsilon^2_{t+h|t})] \]

against the alternative

\[ H_1 : E[L(\varepsilon^1_{t+h|t})] \neq E[L(\varepsilon^2_{t+h|t})] \]
The Diebold-Mariano test is based on the loss differential
\[ d_{t+h|t} = L(\varepsilon_{t+h|t}^1) - L(\varepsilon_{t+h|t}^2) \]
The null of equal predictive accuracy is then
\[ H_0 : E[d_{t+h|t}] = 0 \]
The Diebold-Mariano test statistic is
\[ S = \frac{\bar{d}}{\left( \text{avar}(\bar{d}) \right)^{1/2}} = \frac{\bar{d}}{(LRV_{\bar{d}}/T)^{1/2}} \]
where
\[ \bar{d} = \frac{1}{N} \sum_{j=t}^{t+N} d_{j+h|j} \]
\[ LRV_{\bar{d}} = \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k, \quad \gamma_k = \text{cov}(d_{t+h|t}, d_{t+h-j|t-j}) \]
Note: The long-run variance is used in the statistic because the sample of loss differentials $\{d_{j+h|j}^{t+N}\}$ are serially correlated for $h > 1$. 
Diebold and Mariano (1995) show that under the null of equal predictive accuracy

\[ S \overset{A}{\sim} N(0, 1) \]

So we reject the null of equal predictive accuracy at the 5% level if

\[ |S| > 1.96 \]

One sided tests may also be computed.