I) Title: Forecast Evaluation

II) Contents: Evaluating forecasts, properties of optimal forecasts, testing properties of optimal forecasts, statistical comparison of forecast accuracy

III) Documentation:
- Diebold, Francis X. (1998). *Elements of Forecasting*, Chapter 12, SW College Publishing.
- Hendry, David F., and Clements Michael P. (2001). Economic Forecasting: Some Lessons From Recent Research. *European Central Bank Working Paper Series*, No. 82.
- Ruoss, Eveline, and Savioz, Marcel. (2002). How accurate are GDP forecasts? An empirical study for Switzerland. *Quarterly Bulletin*, 3.
- Stekler, Herman O. (2002). The Rationality and Efficiency of Individuals' Forecasts. In *A Companion to economic forecasting*. Clements, M.C., and Hendry, D.F. Oxford: Blackwell.

Forecasting optimally

The conditional expectation is the optimal estimator for any random variable since it minimizes the average loss function of the prediction. If we use the squared loss function, i.e. $L(e)=e^2$ where e denotes the forecast error, we speak of the *mean squared error* (MSE). The conditional expectation is therefore the *minimum mean squared estimator* (MMSE). To see this, consider the example where we wish to forecast the value of the random variable y_{t+1} based on the vector of previous observations $\mathbf{X}_t=\begin{pmatrix} y_1 & \dots & y_t \end{pmatrix}'$. We are looking for a function such that

$$y_{t+1|t}=g(\mathbf{X}_t) \tag{0.1}$$

where $y_{t+1|t}$ denotes the forecast of y_{t+1} based on information available up to and including time t. Consider the MSE of the suggested function $g(\mathbf{X}_t)$

$$\begin{aligned}
E\left[y_{t+1}-y_{t+1|t}\right]^2 &= E\left[y_{t+1}-g(\mathbf{X}_t)\right]^2 \\
&= E\left[y_{t+1}-E\left[y_{t+1}|\mathbf{X}_t\right]+E\left[y_{t+1}|\mathbf{X}_t\right]-g(\mathbf{X}_t)\right]^2 \\
&= E\left[y_{t+1}-E\left[y_{t+1}|\mathbf{X}_t\right]\right]^2+E\left[E\left[y_{t+1}|\mathbf{X}_t\right]-g(\mathbf{X}_t)\right]^2 \\
&\quad +2E\left[\left(y_{t+1}-E\left[y_{t+1}|\mathbf{X}_t\right]\right)\left(E\left[y_{t+1}|\mathbf{X}_t\right]-g(\mathbf{X}_t)\right)\right] \\
&= E\left[y_{t+1}-E\left[y_{t+1}|\mathbf{X}_t\right]\right]^2+E\left[E\left[y_{t+1}|\mathbf{X}_t\right]-g(\mathbf{X}_t)\right]^2 \\
&\quad +2E\left[E\left\{\left(y_{t+1}-E\left[y_{t+1}|\mathbf{X}_t\right]\right)\left(E\left[y_{t+1}|\mathbf{X}_t\right]-g(\mathbf{X}_t)\right)\Big|\mathbf{X}_t\right\}\right] \\
&= E\left[y_{t+1}-E\left[y_{t+1}|\mathbf{X}_t\right]\right]^2+E\left[E\left[y_{t+1}|\mathbf{X}_t\right]-g(\mathbf{X}_t)\right]^2 \\
&\quad +2E\left[\left(E\left[y_{t+1}|\mathbf{X}_t\right]-g(\mathbf{X}_t)\right)E\left\{\left(y_{t+1}-E\left[y_{t+1}|\mathbf{X}_t\right]\right)\Big|\mathbf{X}_t\right\}\right] \\
&= E\left[y_{t+1}-E\left[y_{t+1}|\mathbf{X}_t\right]\right]^2+E\left[E\left[y_{t+1}|\mathbf{X}_t\right]-g(\mathbf{X}_t)\right]^2
\end{aligned} \tag{0.2}$$

where the last term drops out from the second last to the last line since we have $E\left\{\left(y_{t+1}-E\left[y_{t+1}\middle|\mathbf{X}_t\right]\right)\middle|\mathbf{X}_t\right\}=0$. Hence, if $g\left(\mathbf{X}_t\right)$ is to minimize the MSE, we need that

$$g\left(\mathbf{X}_t\right)=E\left[y_{t+1}\middle|\mathbf{X}_t\right] \tag{0.3}$$

which means that he conditional expectation is the MMSE.

Evaluating forecasts

The general concept of a measure for the goodness of forecast, i.e. the accuracy of the forecast, is the loss function.
Let $e_{t+h,t}=y_{t+h}-E\left[y_{t+h}\middle|y_t\right]=y_{t+h}-y_{t+h,t}$ be the forecast error, thus we can write the loss function $L\left(y_{t+h},E\left[y_{t+h}\middle|y_t\right]\right)=L\left(y_{t+h},y_{t+h,t}\right)=L\left(e_{t+h,t}\right)$
in function of the forecast error.

Rankings of forecasts may vary considerably, depending on the explicit form chosen for the loss function.

The most commonly used accuracy measures for forecasts are the following:

The **mean error** measures the bias of the forecast errors: $ME=\frac{1}{T}\sum_{t=1}^{T}e_{t+h,t}$

The **error variance** measures the dispersion of the forecast errors: $EV=\frac{1}{T}\sum_{t=1}^{T}\left(e_{t+h,t}-ME\right)^2$

The **mean squared error** provides an overall accuracy measure for a forecast:
$MSE=\frac{1}{T}\sum_{t=1}^{T}e_{t+h,t}^2$
Note that the measurement units of the MSE are the square (e.g. dollars^2) of the measurement units of the forecast error (e.g. dollars). It is therefore often convenient to take the **root mean squared error** in order to preserve the original measurement units:
$RMSE=\sqrt{\frac{1}{T}\sum_{t=1}^{T}e_{t+h,t}^2}$
An alternative but less popular overall measure of accuracy is the **mean absolute error**:
$MAE=\frac{1}{T}\sum_{t=1}^{T}\left|e_{t+h,t}\right|$
The task of comparing the accuracy of different forecast methods for varying time periods is more delicate since the mean and the volatility of y_t may be subject to great changes over time. **Theil's U** provides a reasonable solution to this kind of problem:

$$U=\frac{RMSE}{RMS}=\frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}{e_{t,t}}^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}{y_t}^2}} \text{ or } U_\Delta=\frac{RMSE}{RMS}=\frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\Delta y_t-\Delta\hat{y}_t\right)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\Delta{y_t}^2}}$$

Theil's U is zero if the forecast is perfect. If U is smaller (bigger) than one, it means that the forecasting model is performing better (worse) than the simple forecast "no change".

Note that those are only statistical measures which may poorly assess the economic loss of a bad forecast. When forecasts are used to guide decisions the loss is normally an asymmetric function of the error and a simple MSE will not capture this asymmetry.

Properties of optimal forecasts

1.) Optimal forecasts are unbiased, i.e. $E\left[e_{t+h,t}\right]=0$.

2.) The $e_{t+1,t}$ errors of an optimal forecast are white noise.

3.) The $e_{t+h,t}$ (h>1) errors of an optimal forecast are at most $MA(h-1)$.

4.) The forecast error variance of an optimal forecast is non-decreasing in h.

Testing properties of optimal forecasts

1.) Testing $e_{t+h,t}$ for bias

It is straightforward to use the standard t-stat in order to test the null that $E\left[e_{t+h,t}\right]=0$.

This can easily be done by running the OLS regression

(1) $e_{t+h,t}=\alpha+\varepsilon_t$

If the estimated coefficient $\hat{\alpha}$ is not significantly different from zero, we can conclude that the forecast is unbiased. The estimate not being different from zero is a sufficient condition for no bias.

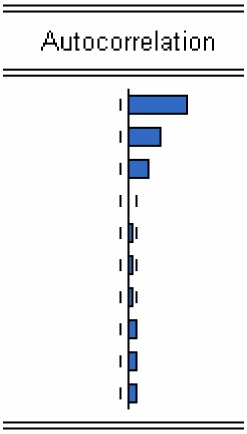
Note however that for h>1 we have $e_{t+h,t}=y_{t+h}-E\left[y_{t+h}\middle|y_t\right]=\varepsilon_{t+h}+\theta_1\varepsilon_{t+h-1}+\ldots+\theta_{h-1}\varepsilon_{t+1}$ and $e_{t+h,t}$ therefore follows (at most) an $MA(h-1)$ process.

This implies that the OLS error terms of (1) might show autocorrelation up to lag h-1. This phenomenon is due to the forecast period overlap associated with multi-step-ahead forecasts.

When regressing (1) we should therefore allow the error terms to follow an $MA(h-1)$ process and use AIC or SIC to find the correct q for the $MA(q)$ process.

2.) Testing $e_{t+1,t}$ for white noise

The standard tests for white noise may be used: Durbin-Watson test, Box-Pierce and Ljung-Box statistics or the sample autocorrelation and partial autocorrelation functions together with the Bartlett bands.



Autocorrelations of an $MA(3)$ process: The autocorrelation cuts off after lag 3.

3.) Testing $e_{t+h,t}$ (h>1) for $MA(h-1)$

We can test if the autocorrelations for lags longer than (h-1) are significantly different from zero by using the Bartlett standard errors.

One could also run a regression similar to (1):

$e_{t+h,t} = \alpha + \varepsilon_t$ where $\varepsilon_t \sim MA(q)$

If we choose $q > (h-1)$, we expect to find for an optimal forecast that the $MA(q)$ coefficients for lags longer than $(h-1)$ are not significantly different from zero.

4.) Testing for a non-decreasing forecast error variance σ_h^2

Recall that $\sigma_h^2 = \sigma^2 \left(1 + \sum_{i=1}^{h-1} \theta_i^2\right)$ and by using this formula it is straightforward to observe whether σ_h^2 is increasing in h or not.

Many of those properties of optimal forecasts can be summarized by the **unforecastability principle**: Optimal forecast errors should be unforecastable given all the information that was available up to the time when the forecast was made.

Contrary to the above mentioned tests, which only make use of univariate properties of the forecast errors, the unforecastability principle relates to their multivariate properties as well.

In order to fully test for the unforecastability principle one would run the regression

(2) $e_{t+h,t} = \alpha_0 + \sum_{i=1}^{k-1} \alpha_i x_{it} + u_t$

where the x_{it}'s contain all information available at time t. A necessary condition for optimality is that $\alpha_0 = \alpha_i = 0$.

One particular case of (2) is the **Mincer-Zarnowitz regression**:

$y_{t+h} = \beta_0 + \beta_1 y_{t+h,t} + u_t$

which can be obtained from (2) as follows:

$e_{t+h,1} = \alpha_0 + \alpha_1 y_{t+h,1} + u_t$

$y_{t+h} - y_{t+h,t} = \alpha_0 + \alpha_1 y_{t+h,t} + u_t$

$y_{t+h} = \alpha_0 + (\alpha_1 + 1) y_{t+h,t} + u_t$

$y_{t+h} = \beta_0 + \beta_1 y_{t+h,t} + u_t$

The necessary condition for an optimal forecast is now $\beta_0 = 0$ and $\beta_1 = 1$.

Note that those conditions are only necessary but not sufficient for optimality!

Rationality in forecasts

This discussion will follow the concepts outlined in Stekler (2002). The two main concepts used in forecasting literature are weak and strong rationality.

Weak rationality means that the forecasts are conditionally unbiased, i.e. there are no systematic errors. **Strong rationality,** which is also called **efficiency**, implies that the forecasts are neither biased nor correlated with any other information known at the time the forecast is made, i.e. the forecaster has efficiently used all available data.

Since it's impossible to test whether a forecaster has used all available information, one normally tests for efficiency by using only the forecaster's past forecasts and forecast errors: **Efficiency** requires that one-step-ahead forecast errors are neither serially correlated, nor correlated with past forecast values or errors. A forecast is then said to exhibit **weak form informational efficiency**. If in addition forecasts are unbiased, they may be called **weakly rational.**

One could think of a Mincer-Zarnovitz regression in order to test for unbiasedness but since this is only a necessary not a sufficient condition, one should rather use (1).

We can also use the regression approach to test for weak and strong efficiency. Consider the regression

(3) $y_{t+h} = \beta_0 + \beta_1 y_{t+h,t} + \beta_2 x_t + u_t$

where the vector x_t contains all information available to the forecaster at time t. Testing strong efficiency means testing the null that $\beta_2 = 0$. If only past values of the relevant variable are included in x_t, testing $\beta_2 = 0$ means testing for weak form efficiency. Note that in the case of $h = 1$ there should be no serial correlation in the error term of (3).

Statistical comparison of forecast accuracy

Diebold and Mariano propose a test for comparing the accuracy of forecasts. It is designed to determine whether two forecasts have the same accuracy. The null hypothesis "equal accuracy hypothesis" is $H_0 : E\left[L\left(e^a_{t+h,t}\right)\right] = E\left[L\left(e^b_{t+h,t}\right)\right]$ for two different forecasts a and b. It can be shown that

$$\sqrt{T}\left(\bar{d}-\mu\right) \sim N(0,f)$$

where $\bar{d} = \frac{1}{T}\sum_{t=1}^{T}\left[L\left(e^a_{t+h,t}\right) - L\left(e^b_{t+h,t}\right)\right]$ and f is the variance of the sample mean loss differential and μ the population mean loss differential.

Setting $\mu = 0$ and using the estimator $\hat{f} = \sum_{\tau=-M}^{M}\hat{\gamma}_d(\tau)$ with $\hat{\gamma}_d(\tau) = \frac{1}{T}\sum_{t=1}^{T}(d_t-\mu)(d_{t-\tau}-\mu)$ and $M = T^{1/3}$ leads to a statistic similar to a standard t-statistic:

$$B = \frac{\bar{d}}{\sqrt{\hat{f}/T}} \sim N(0,1)$$

The difference consists only in the use of f, which represents the fact that the loss differential series in not necessarily white noise.

A parametric way of testing the equal accuracy hypothesis is to regress $d_t = \alpha + \varepsilon_t$ with $\varepsilon_t \sim ARMA(p,q)$.