Module 4 Forecast Uncertainty and Evaluation. Appendix

Notation

- Actual series: y_t, forecast for period t: $\hat{y}_t$
- Average of y_t, $\bar{y} = \frac{1}{f}\sum_{t=1}^{f} y_t$, average of forecast series $\hat{y}_t$, $\bar{\hat{y}} = \frac{1}{f}\sum_{t=1}^{f} \hat{y}_t$,
- Standard error of y_t: $s_y = \sqrt{\frac{1}{f}\sum_{t=1}^{f}(y_t - \bar{y})^2}$, standard error of $\hat{y}_t$: $s_{\hat{y}} = \sqrt{\frac{1}{f}\sum_{t=1}^{f}(\hat{y}_t - \bar{\hat{y}})^2}$
- Forecast error: $FE_t = \hat{y}_t - y_t$
- Bias: $BIAS = \frac{1}{f}\sum_{t=1}^{f} FE_t = \frac{1}{f}\sum_{t=1}^{f}(\hat{y}_t - y_t) = \frac{1}{f}\sum_{t=1}^{f} \hat{y}_t - \frac{1}{f}\sum_{t=1}^{f} y_t$
- Standard forecast error $SE = \sqrt{\frac{1}{f}\sum_{t=1}^{f}(FE_t - BIAS)^2}$
- Mean squared error $MSE = \frac{1}{f}\sum_{t=1}^{f} FE_t^2$
- Root mean squared error $RMSE = \sqrt{MSE} = \sqrt{\frac{1}{f}\sum_{t=1}^{f} FE_t^2}$
- Mean absolute error $MAE = \frac{1}{f}\sum_{t=1}^{f} |FE_t|$
- Mean absolute percentage error $MAPE = \frac{1}{f}\sum_{t=1}^{f} |\frac{FE_t}{y_t}|$

Proof that $MSE = SE^2 + BIAS^2$

$$\begin{aligned}
SE^2 &= \frac{1}{f}\sum_{t=1}^{f}(FE_t - BIAS)^2 \\
&= \frac{1}{f}\sum_{t=1}^{f}(FE_t^2 - 2BIAS \cdot FE_t + BIAS^2) \\
&= \frac{1}{f}\sum_{t=1}^{f} FE_t^2 - \frac{1}{f}\sum_{t=1}^{f} 2BIAS \cdot FE_t + \frac{1}{f}\sum_{t=1}^{f} BIAS^2
\end{aligned}$$

$$
\begin{aligned}
&= MSE - 2BIAS \cdot \frac{1}{f}\sum_{t=1}^{f} FE_t + BIAS^2 \\
&= MSE - BIAS^2
\end{aligned}
$$

Proof that $MSE = BIAS^2 + (s_y - s_{\hat{y}})^2 + 2(1-r)s_y s_{\hat{y}}$

Since we already know that $MSE = SE^2 + BIAS^2$, we only need to show that $SE^2 = (s_y - s_{\hat{y}})^2 + 2(1-r)s_y s_{\hat{y}}$

$$
\begin{aligned}
(s_y - s_{\hat{y}})^2 + 2(1-r)s_y s_{\hat{y}} &= \\
&= s_y^2 - 2s_y s_{\hat{y}} + s_{\hat{y}}^2 + 2s_y s_{\hat{y}} - r s_y s_{\hat{y}} \\
&= s_y^2 + s_{\hat{y}}^2 - 2r s_y s_{\hat{y}}
\end{aligned}
$$

By definition of the correlation coefficient $r = \frac{\hat{Cov}(y_t, \hat{y}_t)}{s_y s_{\hat{y}}} = \frac{\frac{1}{f}\sum_{t=1}^{f}(y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})}{s_y s_{\hat{y}}}$
Then,

$$
\begin{aligned}
s_y^2 + s_{\hat{y}}^2 - r s_y s_{\hat{y}} &= \\
&= \frac{1}{f}\sum_{t=1}^{f}(y_t - \bar{y})^2 + \frac{1}{f}\sum_{t=1}^{f}(\hat{y}_t - \bar{\hat{y}})^2 - 2\frac{1}{f}\sum_{t=1}^{f}(y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}}) = \\
&= \frac{1}{f}\sum_{t=1}^{f}\Big\{(y_t - \bar{y})^2 + (\hat{y}_t - \bar{\hat{y}})^2 - 2(y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})\Big\} \\
&= \frac{1}{f}\sum_{t=1}^{f}\Big\{(y_t - \bar{y}) - (\hat{y}_t - \bar{\hat{y}})\Big\}^2 \\
&= \frac{1}{f}\sum_{t=1}^{f}\Big\{(\bar{\hat{y}} - \bar{y}) - (\hat{y}_t - y_t)\Big\}^2 \\
&= \frac{1}{f}\sum_{t=1}^{f}(BIAS - FE_t)^2 = SE^2
\end{aligned}
$$

Therefore bias proportion plus variance proportion plus covariance proportion sum uip to unity.

Example of incorrect forecast ranking based on the U1 statistic

This section provides an example of two forecasts, where the first forecast is clearly superior to the second, but the U1 statistic based on the first forecast is larger than for the second. In other words, if one makes his judgement solely based on U1 statistic, she will incorrectly prefer the inferior forecast.

Suppose we know that the actual series y_t are distributed with zero mean and variance σ^2. Consider the following forecasts for this test period:

1. Set $\hat{y}_t = 0$ for $t = 1..f$. Then

$$\begin{aligned}
U1 &= \frac{\sqrt{\frac{1}{f}\sum_{t=1}^{f}(y_t-\hat{y}_t)^2}}{\sqrt{\frac{1}{f}\sum_{t=1}^{f}\hat{y}_t^2}+\sqrt{\frac{1}{f}\sum_{t=1}^{f}y_t^2}} \\
&= \frac{\sqrt{\frac{1}{f}\sum_{t=1}^{f}(y_t-0)^2}}{\sqrt{\frac{1}{f}\sum_{t=1}^{f}0^2}+\sqrt{\frac{1}{f}\sum_{t=1}^{f}y_t^2}} = \\
&= \frac{\sqrt{\frac{1}{f}\sum_{t=1}^{f}{y_t}^2}}{\sqrt{\frac{1}{f}\sum_{t=1}^{f}y_t^2}} = 1
\end{aligned}$$

2. Independently draw $\hat{y}_t$from $N(0, \sigma^2)$ every $t = 1..f$. This forecast is a random variable. Then by law of large numbers and using that $\hat{y}_t$ is independent of y_t

$$\begin{aligned}
U1 &= \frac{\sqrt{\frac{1}{f}\sum_{t=1}^{f}(y_t-\hat{y}_t)^2}}{\sqrt{\frac{1}{f}\sum_{t=1}^{f}\hat{y}_t^2}+\sqrt{\frac{1}{f}\sum_{t=1}^{f}y_t^2}} \\
&= \frac{\sqrt{\frac{1}{f}\sum_{t=1}^{f}y_t^2-2\frac{1}{f}\sum_{t=1}^{f}\hat{y}_t y_t+\frac{1}{f}\sum_{t=1}^{f}\hat{y}_t^2}}{\sqrt{\frac{1}{f}\sum_{t=1}^{f}\hat{y}_t^2}+\sqrt{\frac{1}{f}\sum_{t=1}^{f}y_t^2}} \\
&\approx \frac{\sqrt{\sigma^2-2\cdot 0+\sigma^2}}{\sqrt{\sigma^2}+\sqrt{\sigma^2}} = \frac{\sqrt{2}}{2} < 1
\end{aligned}$$

Clearly, the second forecast is worse than the first. The root mean squared error of the first forecast is $RMSE_1 = \sqrt{\frac{1}{f}\sum_{t=1}^{f}(y_t-\hat{y}_t)^2} = \sqrt{\frac{1}{f}\sum_{t=1}^{f}{y_t}^2} \approx \sigma$. For the second forecast $RMSE_2 = \sqrt{\frac{1}{f}\sum_{t=1}^{f}(y_t-\hat{y}_t)^2} \approx \sigma\sqrt{2} > RMSE_1$ (see the numerator of the second U statistic). At the same time the first forecast recieves a worse ranking based on the U1 statistic.