Out-of-Sample Forecasting Experiment

Out-of-sample forecasting experiments are used by forecasters to determine if a proposed leading indicator is potentially useful for forecasting a target variable. The steps for conducting an out-of-sample forecasting experiment are as follows:

1) Divide the available data on the target variable, y_t, (here we assume y_t is stationary) and the proposed leading indicator, x_t, (likewise we assume that x_t is stationary) into two parts: the **in-sample data set** (roughly 80% of the data) and the **out-of-sample data set** (the remaining 20% of the entire data set).

2) In consultation with the person who will be using your forecasts, choose an **appropriate forecast horizon and loss function** for the forecasting experiment. The forecast horizon is the number of steps ahead that one is most interested in forecasting the target variable. For example, if a person is in charge of managing the inventory of a firm, she might be only interested in obtaining accurate forecasts of sales one period ahead and the appropriate forecast horizon would be h = 1. On the other hand, if interest centers on the sales that will be present 8 periods from now, just in time for the completion of a new manufacturing facility then the appropriate forecast horizon for the out-of-sample forecasting experiment would be h = 8.

3) Once you have chosen the in-sample data set, you should use it to choose two competing forecasting models. The first model you should build is a **Box-Jenkins model** for the target variable, y_t, and then, separately, build a **Transfer Function model** for y_t that includes your proposed leading indicator, x_t. It is these two competing models that you are going to run an out-of-sample "horserace" with.

4) To run a **horserace** (i.e. forecasting competition) between these two models, you must **"roll"** each model through the out-of-sample data set **one observation at a time** while each time forecasting the target variable the chosen h periods ahead. (h is the forecast horizon of interest.) The term "rolling" means that you **re-estimate the parameters** (coefficients) of each model with one more observation added to your estimation data each time you forecast the target variable h periods ahead.

5) While you are rolling your competing models through the out-of-sample data set forecasting h periods ahead you need to record the errors of each model each time your forecast. Knowing the errors of each model, say e_t^{BJ} and e_t^{TF}, and the particular loss function that our boss has chosen for us, say, $L(e_t)$, we can calculate the respective loss for the Box-Jenkins model, $L(e_t^{BJ})$,

associated with a given forecast and the loss for the Transfer Function model, $L(e_t^{TF})$ for a given forecast. Let t_0 denote the last time period in the in-sample data set, h be the chosen forecast horizon, T be the total number of observations available (the sum of the number of observations in the in-sample and out-of-sample data set), and M be the number of observations reserved for the out-of-sample data set. It then follows that the in-sample data set contains T – M data points and we can forecast M – h + 1 times when rolling the competing forecasting models through the out-of-sample data set and with the chosen forecast horizon being h-steps ahead. Likewise, when we roll the two competing models through the out-of-sample data set we will correspondingly have (M – h + 1) losses associated with the Box-Jenkins model

$$L(e_t^{BJ}), t = t_0 + h, \cdots, T$$

and (M – h + 1) losses associated with the Transfer Function model

$$L(e_t^{TF}), t = t_0 + h, \cdots, T \quad .$$

6) Now to decide the winner of the horserace between the BJ and TF models we must calculate the **Average Loss** associated with the two models that occurs over the (M – h + 1) forecasts produced by each model. These Average Losses are calculated as the sample average of the (M – h + 1) losses associated with the (M – h + 1) forecasts produced by each forecasting model, namely,

$$\overline{L}(e_t^{BJ}) = \sum_{t=t_0+h}^{T} L(e_t^{BJ})/(M - h + 1)$$

and

$$\overline{L}(e_t^{TF}) = \sum_{t=t_0+h}^{T} L(e_t^{TF})/(M - h + 1) \quad .$$

Therefore the winner of the forecasting competition is the model that produces the smallest Average Loss in the out-of-sample forecasting experiment. If $\overline{L}(e_t^{BJ}) < \overline{L}(e_t^{TF})$, the BJ model is the winner and one would conclude that the leading indicator used in the TF model was not “potent” enough to offer a forecasting accuracy gain. We should then begin a search for a better leading indicator to use. On the other hand, if $\overline{L}(e_t^{TF}) < \overline{L}(e_t^{BJ})$ the TF model is the winner and we can conclude that we have found a leading indicator that is useful for forecasting the target variable y_t and we, as economists, have beaten the statistician in forecasting since he/she is not aware of the leading indicator and, in adopting the Box-Jenkins model, is working without it.

7) In case the "boss" does not have a specific loss function to describe the losses associated with forecast errors, one can always adopt the "standard" average loss functions, MAE and MSE. The **Mean Absolute Error** (MAE) average loss function is defined as

$$\text{MAE} = \sum_{t=t_0+h}^{T} |e_t| /(M - h + 1) \quad .$$

The **Mean Squared Error** (MSE) average loss function is defined as

$$\text{MSE} = \sum_{t=t_0+h}^{T} e_t^2 /(M - h + 1) \quad .$$

The forecasting method that has the smallest MAE **and** MSE average losses in the out-of-sample forecasting experiment is then the superior forecasting method. If one forecasting method has a better MAE measure while the other forecasting method has the better MSE measure then you have a **split decision**. Then the only way you can determine a winner between the two competing forecasting models is to make a decision and choose one of the average loss functions to base your choice on, either the MAE average loss function or the MSE average loss function.

Testing for Unbiasedness of Forecasts and the Statistical Significance in the Difference in Competing Forecasts

1. The Mincer-Zarnowitz test for the Unbiasedness of Forecasts