Table of Contents

From Point Forecasts to Distributional Forecasts

- A forecast $\hat{y}_{T+h|T}$ is (usually) the mean of the conditional distribution: $Y_{T+h}|Y_1, \ldots Y_T$
- Most models produces Gaussian distributed forecast
- Because models assumes Gaussian residuals: remember that the residuals are the stochastic components that are determining the distribution of both the estimators and the forecasts
- The forecast distribution describes the probability to forecast any future value

Distributional Forecasts

Assuming the **residuals are uncorrelated** with variance σ^2, and with an estimate $\hat{\sigma^2}$

Model Typology

- **Mean:** $Y_{T+h|T} = \mathcal{N}(\overline{Y}, (1 + \frac{1}{T})\hat{\sigma^2})$
- **Naive:** $Y_{T+h|T} = \mathcal{N}(Y_T, h\hat{\sigma^2})$
- **Seasonal Naive:** $Y_{T+h|T} = \mathcal{N}(Y_{T+h-m(k+1)}, (k+1)\hat{\sigma^2})$
- **Drift:** $Y_{T+h|T} = \mathcal{N}(Y_T + \frac{h}{T-1}(Y_T - Y_1), h\frac{T+h}{T}\hat{\sigma^2})$

- where k is the integer part of $\frac{h-1}{m}$
- Note that when $h = 1$ and T is large, these all give the same approximate variance $\hat{\sigma^2}$

Prediction Intervals

Definition

A prediction interval gives **a region** within which we expect Y_{T+h} to lie with **a specified probability**

- Assuming that the forecasting errors are normally distributed, then a 95% prediction interval is:

$$Y_{T+h|T} \pm 1.96\hat{\sigma}_h$$

- where $\hat{\sigma}_h$ is the standard deviation of the h-step distribution
- when $h = 1$, $\hat{\sigma}_h$ can be estimated from the residuals

Prediction Interval Fit

```
brick_fc %>% hilo(level = 95)
```

```
## # A tsibble: 80 x 5 [1Q]
## # Key:       .model [4]
##    .model         Quarter       Bricks .mean        `95%`
##    <chr>            <qtr>       <dist> <dbl>       <hilo>
##  1 Seasonal_naive 2005 Q3 N(428, 2336)   428 [333, 523]95
##  2 Seasonal_naive 2005 Q4 N(397, 2336)   397 [302, 492]95
##  3 Seasonal_naive 2006 Q1 N(355, 2336)   355 [260, 450]95
##  4 Seasonal_naive 2006 Q2 N(435, 2336)   435 [340, 530]95
##  5 Seasonal_naive 2006 Q3 N(428, 4672)   428 [294, 562]95
##  6 Seasonal_naive 2006 Q4 N(397, 4672)   397 [263, 531]95
##  7 Seasonal_naive 2007 Q1 N(355, 4672)   355 [221, 489]95
##  8 Seasonal_naive 2007 Q2 N(435, 4672)   435 [301, 569]95
##  9 Seasonal_naive 2007 Q3 N(428, 7008)   428 [264, 592]95
```

Why Prediction Intervals Matter

- **Point forecasts are often useless without a measure of uncertainty** (such as prediction intervals)
- Prediction intervals require a **stochastic model** (with random errors, etc.)
- For most models, prediction intervals get wider as the forecast horizon increases
- The degree of confidence (the probability) impacts the width of the prediction interval
- Usually too narrow due to unaccounted uncertainty: pay attention

Difference between Confidence Interval and Prediction Interval

- A confidence interval informs about where the true parameter of a model can be
  - **Confidence interval quantifies the uncertainty about the model**, or the distance between the model and reality
  - Confidence interval are associated with a wide range of parameters, values, etc.
  - They inform about how the model represents well the reality
  - Wide confidence intervals are associated with a less accurate model, and/or a very volatile model

- The prediction interval predicts in what range a future individual observation will fall
  - **Prediction interval quantifies the uncertainty about the future**, or the distance between today and the future
  - Prediction interval are not about the parameters of the model, but about the dependent variable (y_t)
  - The problem is that prediction intervals tend to neglect the uncertainty about the parameters used to generate the forecasts...

Difference between Confidence Interval and Prediction Interval

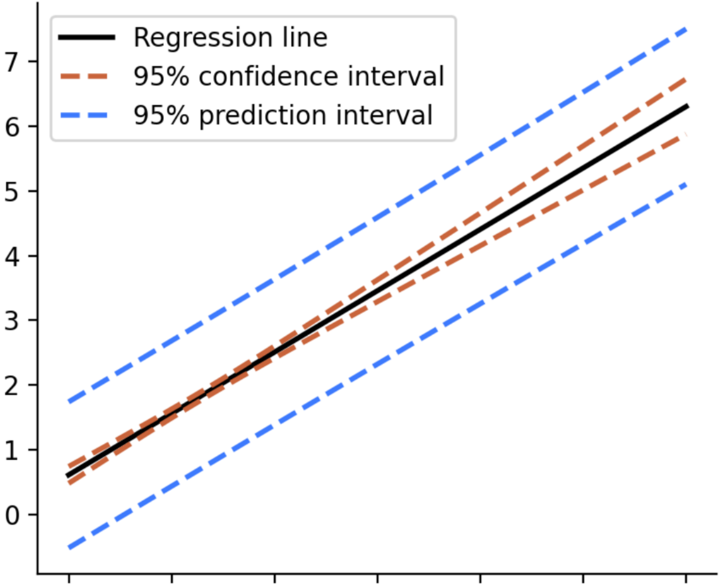


Table of Contents

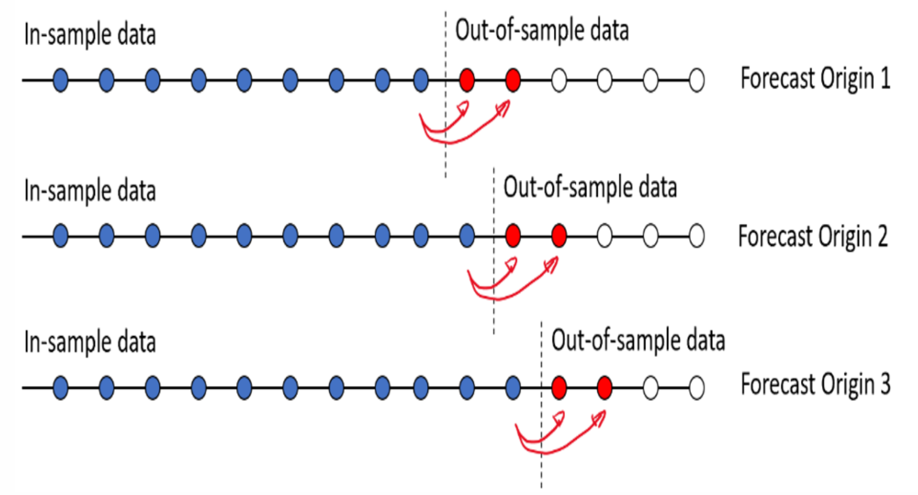
Fitting and Forecasting

Be careful

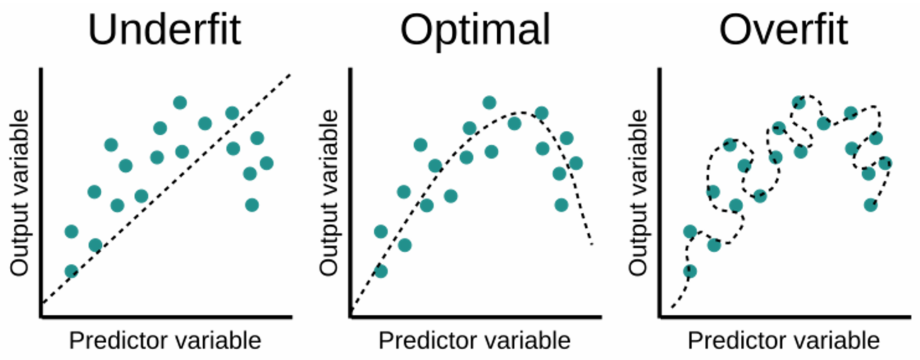
A model that fits the data well (in sample) might not necessarily forecast well

- A perfect in-sample fit can always be obtained by using a model with with enough parameters
- Over-fitting a model to data is just as bad as failing to identify a systematic pattern in the data
- Need to split the model between
- The test set must no be used to *any* aspect of model development or calculation of forecasts
- Forecast accuracy is only based on the test set

Train and Test Set



Underfit, Optimal, Overfit



Forecast Errors

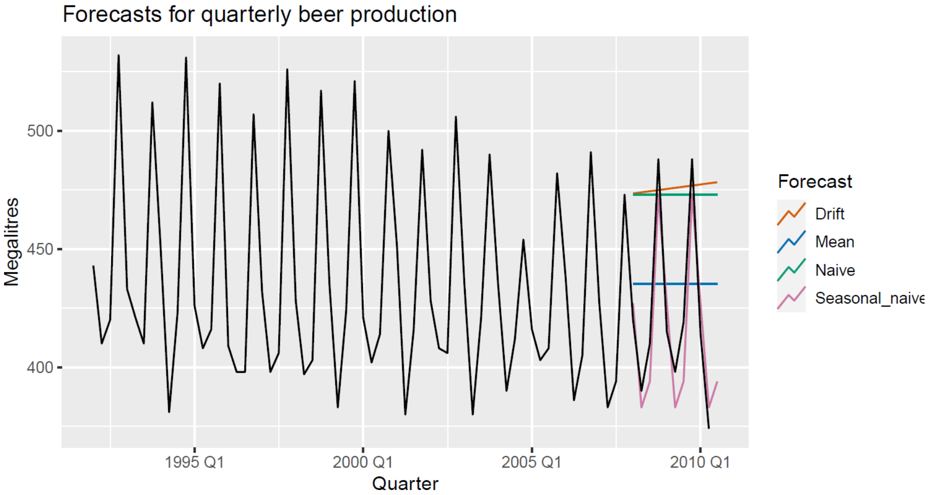
Definition: Forecast Errors

A forecast error is the difference between an observed value and its forecast

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h}|Y_T, \ldots, Y_1$$

- The conditional set $Y_T, \ldots, Y_1$ should only be taken from the training dataset
- The true value y_{T+h} is taken from the test set
- Unlike residuals, forecast errors on the test involve multi-step forecasts
- These are the **true** forecast error, as the test data is not used to compute $\hat{y}_{T+h}$

Example: Forecasting Beer Production



Measures of Forecast Accuracy

Main Metrics

- **MAE**: mean absolute errors $\frac{1}{S}\sum_{s \in S} |e_{s,T+h}|$
- **MSE**: mean squared errors $\frac{1}{S}\sum_{s \in S} (e_{s,T+h})^2$
- **MAPE**: mean absolute percentage errors $\frac{1}{S} 100 * \sum_{s \in S} \frac{|e_{s,T+h}|}{|y_{s,t+h}|}$
- **RMSE**: root mean squared errors: $\sqrt{\frac{1}{S}\sum_{s \in S} (e_{s,T+h})^2}$

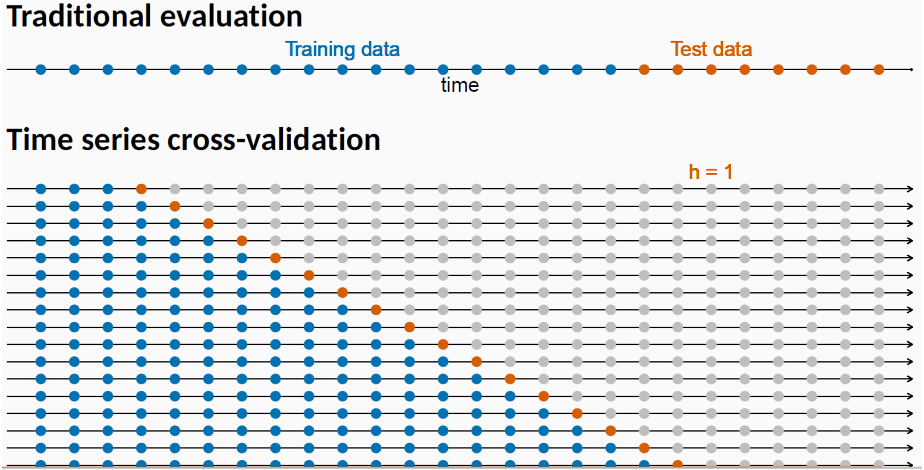
With:

- y_{T+h}: T+h observation, h being the horizon (h = 1, 2, ..., H)
- $\hat{y}_{T+h|T}$: the forecast based on data up to time T
- $e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$: The forecast errors
- S is the testing sample

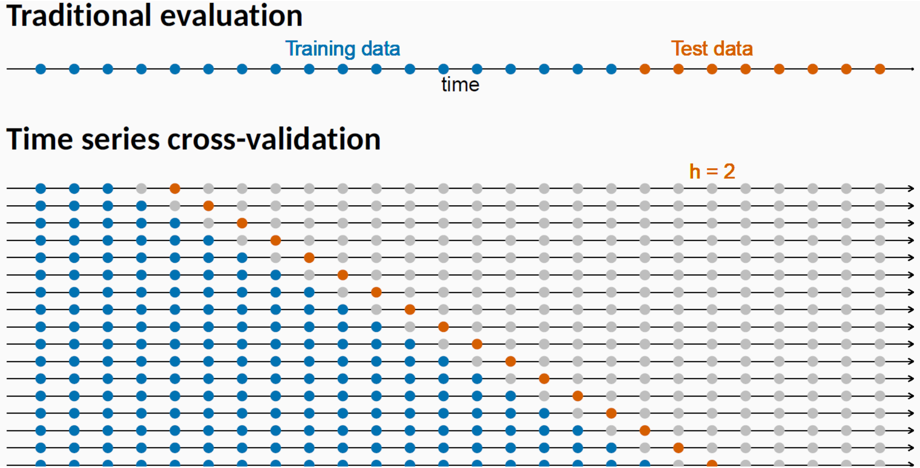
Note:

- MAE, MSE and RMSE are all **scale dependent**

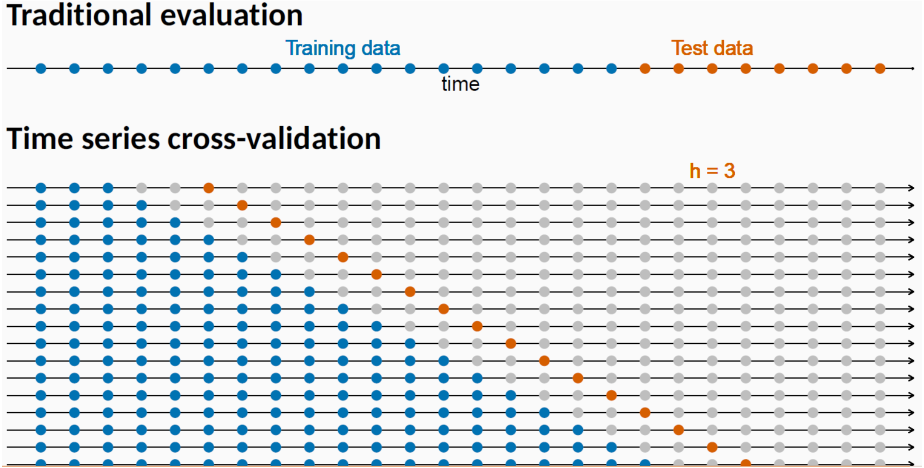
Time Series Cross-Validation



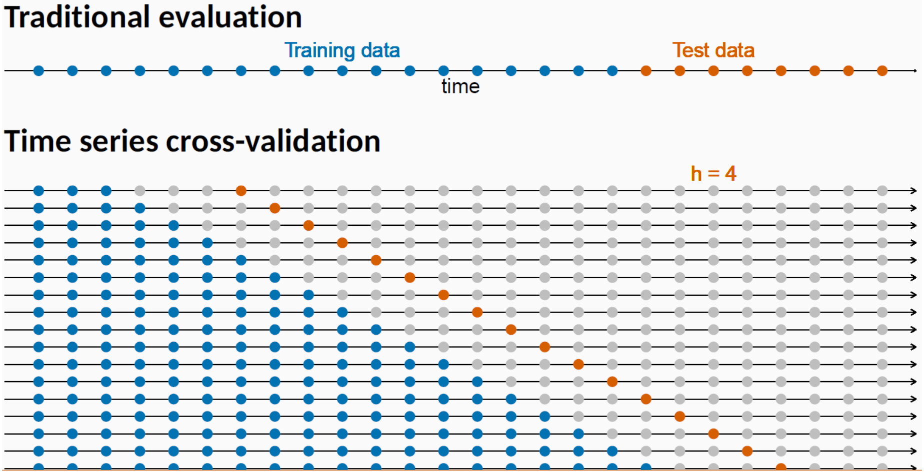
Time Series Cross-Validation



Time Series Cross-Validation



Time Series Cross-Validation



Time Series Cross-Validation