Model estimation – MLE

- Maximum likelihood estimation (MLE) is a common procedure for estimating the parameters of a model
- Given a set of observations and an underlying probability model, maximum likelihood identifies the values of the model parameters that are most likely to have generated the observations, y_t
- To illustrate the concept of parameter estimation using MLE, we consider the following simple model

$$y_t = \alpha_0 + \alpha_1 x_t + \varepsilon_t$$

for $\varepsilon_t \sim N(0, \sigma)$, we have $y_t \sim N(\alpha_0 + \alpha_1 x_t, \sigma)$. Hence, we can write the conditional probability density function of y_t as

$$f(y_t|x_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_t - \alpha_0 - \alpha_1 x_t)^2}{2\sigma^2}\right]$$

- If the ε_t are independent, the joint PDF for a set of y_i for $i = 1,2,\ldots,n$ can be expressed as

$$f(y_1, y_2,\ldots, y_t|x_1, x_2,\ldots,x_t) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \alpha_0 - \alpha_1 x_i)^2}{2\sigma^2}\right] = (2\pi\sigma^2)^{-n/2} \exp\left[-\sum_{i=1}^{n} \frac{(y_i - \alpha_0 - \alpha_1 x_i)^2}{2\sigma^2}\right] \ldots$$

- The focus of MLE approach is to find the models parameters (α_0, α_1) that maximize the likelihood of drawing the observed sample of observation.
- It is usually more convenient to maximize the log of likelihood expression

$$\max log\left((2\pi\sigma^2)^{-n/2} \exp\left[-\sum_{i=1}^{n} \frac{(y_i - \alpha_0 - \alpha_1 x_i)^2}{2\sigma^2}\right]\right)$$
$$= \max\left(-\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(y_i - \alpha_0 - \alpha_1 x_i)^2}{2\sigma^2}\right)$$

Note that finding the value of α_0 and α_1 that maximize the likelihood expression is equivalent to finding the values that minimize the following expression:

$$\min \sum_{i=1}^{n} (y_i - \alpha_0 - \alpha_1 x_i)^2$$

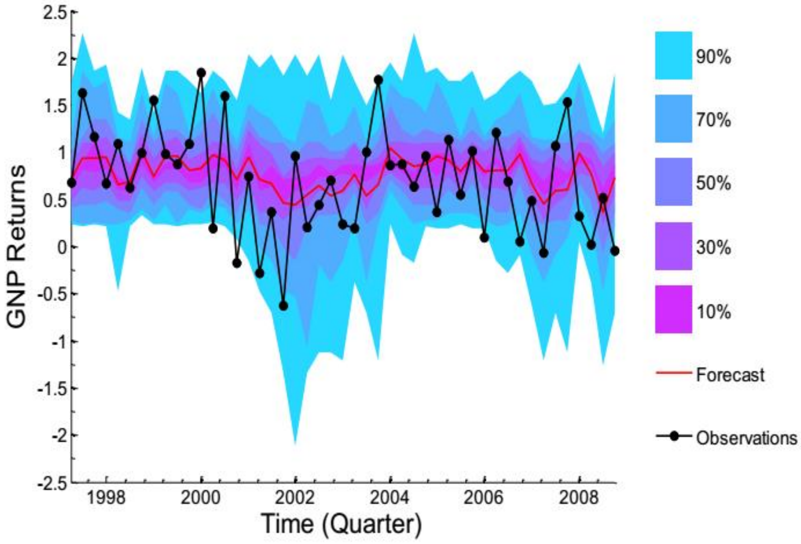
Interestingly, the above expression is equivalent to **Ordinary Least Squares** (OLS)! Maximizing the log likelihood corresponds to minimizing least squares when the errors are independent and normally distributed

Model evaluation

- We use the in-sample data (*training set*) to estimate the model parameters (using MLE, OLS etc.)
- Using the optimized parameter vector, we generate forecasts for the out-of-sample data (*testing data*). Estimate parameters based on in-sample probabilistic forecasts (minimize CRPS).
- A model needs to be evaluated in terms of its *point*, *quantile*, and *probability density* forecast accuracy using a range of different *performance scores.*
- A performance score basically quantifies the difference between the true observation (y_t) and the corresponding forecast estimate ($\hat{y}_t$).
- In practise, the performance scores are computed for each forecast horizon under consideration. Plotting out-of-sample Forecast Error versus Forecast Horizon can be very useful to gauge how the model's performance deteriorates with longer horizons.
- Look for the horizon where the model's accuracy is no better than a simple benchmark.

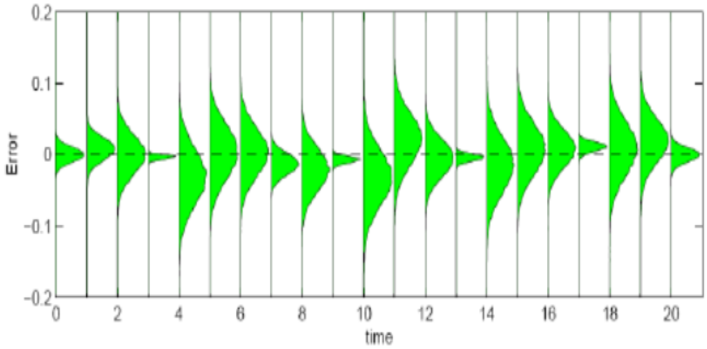
Model evaluation

- Given a set of out-of-sample/test observations, and corresponding *k*-step ahead forecasts, how do we assess the model's performance?



Model evaluation

- In essence, model evaluation is all about quantifying prediction errors
- Useful to employ a range of different performance scores for evaluation
- Overall, the forecast accuracy would deteriorate with longer horizons
- We discuss some of the most widely used performance scores used for assessing point, quantile and density forecast accuracy



Point forecast evaluation – RMSE

- The Root Mean Square Error (RMSE) is given by

$$RMSE_h = \sqrt{\frac{1}{N-T-h+1}\sum_{i=T+h}^{N}(y_i - \hat{y}_i)^2}$$

where $RMSE_h$ is the RMSE at horizon h, y_i is the actual observation, $\hat{y}_i$ is the corresponding forecast, T is the forecast origin, and N is the length of the time series

- RMSE represents a measure of forecast performance that is analogous to the least square parameter estimation technique

- However, RMSE is sensitive to outliers, may give misleading results if the forecast errors are not normally distributed, and is scale dependent

Point forecast evaluation – MAE/MAD

- The Mean Absolute Error (MAE) is given by

$$MAE_h = \frac{1}{N-T-h+1}\sum_{i=T+h}^{N} |y_i - \hat{y}_i|$$

where MAE_h is the MAE at horizon h, y_i is the actual observation, $\hat{y}_i$ is the corresponding forecast, T is the forecast origin, and N is the length of the time series

- This forecast measure focuses on the absolute magnitude of the errors.
- MAE is more robust than RMSE as the large errors are not squared
- Perhaps also worth computing Median Absolute Deviation (MAD), which is a robust measure of the variability

Point forecast evaluation – MAPE

- The Mean Absolute Percentage Error (MAPE) is given by

$$MAPE_h = \frac{100}{N-T-h+1}\sum_{i=T+h}^{N} \left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

where $MAPE_h$ is the MAPE at horizon h, y_i is the actual observation, $\hat{y}_i$ is the corresponding forecast, T is the forecast origin, and N is the length of the time series

- Focusing on the percentage error is useful as a means of standardising the results
- It should only be used if the dependent variable is positive definite
- Commonly used in energy forecasting