Notes for Hyndman and Athanasopoulos – Chapter 2

* Frequency of a time-series is the number of observations before the seasonal pattern repeats.
  1. Yearly data: Frequency = 1
  2. Quarterly data: Frequency = 4
  3. Monthly data: Frequency = 12
  4. Weekly data: Frequency = 52
  5. Daily data might have two types of seasonality:
     + Weekly seasonality (frequency = 7)
     + Annual seasonality (frequency = 365)
* Time series patterns
  1. *Trend:* A trend exists when there is a long-term increase or decrease in the data. It does not have to be linear.
  2. *Seasonal:* A seasonal pattern occurs when a time-series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality is always fixed and of known frequency.
  3. *Cyclic:* A cycle occurs when the data exhibit rises and falls that are not of a fixed frequency
* If the fluctuations are not of a FIXED FREQUENCY, then they are cyclic; if the frequency is unchanging and associated with some aspect of the calendar then the pattern is seasonal.
* An accurate forecasting method must take into account the time patterns in the data and be able to capture the patterns properly.
* A seasonal plot is similar to a time plot except that the data are plotted against individual seasons in which the data was observed.
* Scatterplots are useful for visualizing relationship between time-series. It is common in these cases to display the correlation coefficient to measure the strength of the relationship between these two variables.
* Autocorrelation measures the linear relationship between lagged values of a time-series.
* The autocorrelation coefficients are plotted to show the autocorrelation function (AFC). The plot generated is known as the *correlogram*.
* When data has a trend, the autocorrelation for small lags tend to be large and positive because observations nearby in time are also nearby in size (due to changing mean of the series as time goes by).
* When data is seasonal, the autocorrelations will be larger for the seasonal lags (at multiples of the seasonal frequency) than for other lags.
* Time series that show no autocorrelation are called **white noise**.

Notes for Hyndman and Athanasopoulos – Chapter 3

**Simple forecasting methods**

* Average method for forecasting:
  1. The forecasts of all future values are equal to the average of the historical data.
* Naïve method of forecasting:
  1. Naïve forecast consists in setting all forecasts to be the value of the last observation. This is an optimal technique when data follows a random walk.
* Seasonal naïve method:
  1. The forecast is equal to the last observed period from the same season of the past year.
* Drift method:
  1. A variation of the naïve method is to allow for forecasts to increase or decrease over time, where the amount of change over time (**drift**) is set to be the average change seen in historical data. Equivalent to drawing a line between the first and last observations and extrapolating into the future.
* These simple forecasting methods usually serve as benchmark to analyze more complicated models.

**Transformation and adjustment**

* Adjusting the historical data can often lead to a simpler forecasting task. The purpose of these adjustments and transformations is to simply the patterns in historical data by removing known sources of variation.
* Calendar adjustments:
  1. A common source of variation derives from the fact that some months have more days than others.
* Population adjustments:
  1. Any data that is affected by population changes can be adjusted to give per-capita data. This removes the effects of total population changes.
* Inflation adjustments:
  1. Financial time-series are usually adjusted so that all values are stated in dollar values from a particular year.

**Mathematical transformations**

* Logarithmic transformations.
* Power transformations.
* Box-cox transformation: a family of transformations that combines natural logarithms and power transformations. For different values of lambda, the series will change in shape. A good value of lambda makes the seasonal variation about the same across the whole series.
* One issue with mathematical transformations is that the back-transformed forecast will not be the mean of the forecast distribution. Therefore, the result might need to be adjusted to ensure that the mean of the distribution will be recovered. This process is called bias-adjusting.

**Residual diagnostics**

* The residuals in a time series model are what is left over after fitting a model.
* For many models, the residuals are equal to the difference between the observations and the corresponding fitted values.
* Residual analysis is useful for model diagnostics (whether it has adequately captured the information in the data). There are two characteristics of residuals generated by good forecasting methods:
  1. Residuals are uncorrelated. If there is correlation between residuals, it suggests that there is information that could be used in the forecasts.
  2. The residuals have zero mean. Otherwise the forecasts are biased.
* There are two other characteristics that are useful, although not necessary:
  1. Residuals have constant variance.
  2. They are normally distributed.
* Residual autocorrelation can be visually checked using the correlogram. However, it is also advisable to run tests that take into account the possibility that multiple comparisons might give rise to false positives. The box-pierce test is one of such tests (another alternative is the Ljung-Box test). Both tests assume that there is no autocorrelation between values in the series.
* Forecasts should be assessed (in terms of quality) against new information. Therefore, their accuracy can only be truly determined by considering how well a model performs on new data that was not used when fitting the model.
* When choosing models, the usual practice is to separate the available models into two parts:
  1. Training: used to learn model parameters.
  2. Test: used to evaluate the model accuracy.
* The size of the test set is typically about 20% of the total sample. However, it is good to practice for it to be as large as the maximum forecast horizon required.
* A forecast error is the difference between an observed value and its forecast.
* Forecast errors are on the same scale as the data. Therefore, accuracy measures that rely on forecast errors are scale-dependent and cannot be used to make comparison between series that involve different units.
* There are two scale-dependent measures that are frequently used:
  1. Mean absolute error (MAE): equal to the average absolute value of forecast errors.
  2. Root mean squared error (RMSE): equal to the square root of the average squared forecast error.
* A forecast method that minimizes the MAE will lead to forecasts of the median, while forecast methods that minimize the RMSE will lead to forecasts of the mean.
* Percentage error is how much, on a percent scale, the forecast is “wrong”. The mean absolute percentage error is the average of those. This is scale-independent.
* Scaled errors is an alternative to using percentage errors when comparing forecast accuracy across series with different units. They propose scaling the errors based on the *training* MAE from a simple forecast method.
* Cross-validation is another procedure to determine forecasting accuracy: In this procedure, there are a series of test sets, each consisting of a single observation. The corresponding training set consists only of observations that occurred prior to the observation that forms the test set.
* Since we need a minimum number of observations to generate a minimally reliable forecast, the earliest observations are not considered test sets.
* The forecast accuracy is computed by averaging over the test sets.
* A good way to choose the best forecasting model is to find the model with the smallest RMSE computed using time series cross-validation.
* A prediction interval provides an interval within which we expect our forecast to lie with a specified probability.
* The value of prediction intervals is that they express uncertainty in the forecasts.
* One-step prediction intervals:
  1. When forecasting one step ahead, the standard deviation of the forecast distribution is almost the same as the standard deviation of the residuals.
* For multi-step prediction intervals, it is necessary to estimate (the standard deviation of the forecast. For one-step ahead, the standard deviation of the residuals is a relatively good estimate).

Notes for Hyndman and Athanasopoulos – Chapter 5

* The basic concept behind time series regression models is that we forecast the time series of interest *y* assuming that it has a linear relationship with other time series x.
* When there are two or more predictor variables, the model is called a **multiple regression model**. The coefficients measure the effect of each predictor after taking into account the effects of all other predictors in the model. Thus, the coefficients measure the *marginal effects* of the predictor variables.
* Assumptions behind the linear regression model:
  1. The model is a reasonable approximation of reality; that is, the relationship between the forecast variable and the predictor variable satisfies the linear equation.
  2. There are three auxiliary assumptions about the errors:
     + They have mean zero; otherwise the forecasts will be systematically biased
     + They are not autocorrelated; otherwise the forecasts will be inefficient, as there is more information in the data that can be exploited.
     + They are unrelated to the predictor variables; otherwise there would be more information that should be included in the systematic part of the model.
     + Fixed predictors.
* We choose the coefficients that minimize the sum of squared residuals.
* The t-statistic and the associated p-values for the coefficients are useful if we are interested in studying the effect of each predictor and making causal claims about it. For forecasting these numbers are not particularly useful.
* A common way to summarize how well a linear regression model fits the data is via the coefficient of determination or R-squared. It reflects the proportion of the variation in the forecast variable that is accounted for by the regression model.
* Another measure of how well the model has fitted the data is the standard deviation of the residuals (the residual standard error). This measure is related to the size of the average error that the model produces.