Hyndman and Athanasopoulos – Answers to chapter 8

1. – ACF for random numbers.
   1. All three figures suggest that the data is white noise. Their pattern suggests that there is no autocorrelation between values of the series.
   2. Because the critical values are a function of the length of the series. The critical values are equal to 1.96/√T. That is why the critical values have different distances from zero. The autocorrelations are different because we are dealing with random numbers, so the existence of autocorrelations is a product of change and not of any characteristic in the data.
2. IBM closing price analysis.
   1. The ACF pattern suggests that that the data is non-stationary due to the fact that it decreases slowly, showing that more recent values depend heavily on past values. The plot also makes it clear that there is a possible negative trend. The PACF also contains a unique significant spike at lag 1, which suggests that the data is basically a random walk (which is non-stationary). After differencing the ACF does not show the decreasing pattern anymore. The high correlation at lag-1 on the PACF also disappears.
3. Appropriate Box-Cox and differencing.
   1. Box-cox
      1. usnetelec = 0.517
      2. usgdp = 0.366
      3. mcopper = 0.192
      4. enplanements = -0.227
      5. visitors = 0.278
   2. Differencing
      1. Usnetelec = 1
      2. Usgdp = 2
      3. mcopper = 1
      4. enplanements = 1
      5. Visitors = 1
4. Done in notebook
5. Seasonal differencing is not necessary, while we need only one regular differencing.
6. ARIMA simulation.
   1. Done in R.
   2. The variance increases and the resulting series becomes smoother.
   3. Same as above.
   4. The variance increases and the resulting series becomes smoother.
   5. Done in R
   6. Done in R
   7. The AR (2,0) model is much less smooth than the ARMA (1,1) model.
7. Women Murdered in the US series
   1. After twice- differencing the series to make it stationary, an analysis of the ACF suggests an MA (2) process and an AR (2) process.
   2. A non-zero constant and a twice differenced series will cause long-term forecasts to follow a quadratic trend, which seems to be unwarranted given the data at hand.
   3. Done in notebook
   4. The residuals pass the test. The Ljung-box test fails to reject the null hypothesis. The residuals also seem to not deviate from normality.
   5. They check
   6. Done in R. Notice how the forecasts follow a straight line
   7. Auto-arima chose a different model: ARIMA (0,2,3). Notice that double-differencing remains.
8. USGDP quarterly analysis
   1. Box-Cox with lambda equals to 0.36
   2. Exhaustive search found an ARIMA (0,1,2) model with a drift parameter.
   3. Done in R
   4. Using time-series cross-validation, the auto-arima model is better than the fitted model.
   5. The residuals seem to follow a white-noise process.
   6. They seem reasonable.
   7. The ETS fits a model with no seasonality, but with an additive trend and additive errors. The ARIMA model forecasts slightly lower future levels of GDP. The ETS (A,A,N) show an upward trend into the future as we would expect from a model without a damping parameter.
9. Quarterly visitor nights in Australia data
   1. This time series has quarterly frequency. The patterns in the data suggest strong seasonality and an upward trend. The seasonal variation is much larger than the rise in the trend. Also, it seems that the data requires some transformation to stabilize the variance.
   2. The correlation between present values and past values is statistically significant for the seasonal lags (1-3) and for adjacent lags to seasonal components.
   3. The PACF suggests significant lags 1 and 2 and significant seasonal lags 1 and 2 as well.
   4. A Box-Cox transformation and one seasonal differencing seem to create a stationary time-series. We fail to reject the null hypothesis of the KPSS unit root test. A good model seems to be SARIMA (1,0,1), (1,1,1)
   5. No, it chose an SARIMA (1,0,0) (0,1,1). An ARIMA model with one non-seasonal autoregressive component, but no moving average or differencing; and a seasonal part comprised of no autoregressive terms, but that differences the data once and also contains one moving average component. Using cross-validation, the model chosen by auto-arima has a smaller MSE.
   6. Done in notebook
10. Usmelec series
    1. The data clearly has an upward trend.
    2. The plot of the data shows that the seasonal variation increases with time. A Box-Cox with lambda -0.57 will be used.
    3. Using the ndiff command after taking first-seasonal differencing seems to suffice to create a stationary series.
    4. Using auto-arima, we found an ARIMA (1,1,1) (2,1,1) model, which has a lower AICc than the model fitted manually ARIMA(1,1,3)(0,1,3)
    5. The residuals resemble white-noise.
    6. Checking against actual values, the model has the following error metrics
       1. MAE: 8.17
       2. MAPE: 2.41
       3. RMSE: 10.33
    7. Around 2020.