Hyndman and Athanasopoulos – chapter 7 answers

1. PIGS series analysis.
   1. The data shows no signs of a trend or of a seasonality. Alpha is roughly equal to 0.3, while the initial state is calculated as 77,260. The standard deviation of the residuals is equal to 10,308.58 pigs slaughtered.
   2. The 95% prediction interval for Sep-1995 was [78,199 - 119,433]. It is wider than the interval produced by R.
      * From Gelman: the residual standard deviation, sigma, can be thought as a measure of the average distance each observation falls from its prediction from the model.
2. Done exclusively in R. There is some difference, but I suspect it is a product of rounding the initial conditions and the alpha. For all intents and purposes, they are the same.
3. The results are the same as obtained with SES. There is a small difference, but again I think this is a consequence of rounding.
4. Done in R
5. Book series analysis
   1. Both series express a very erratic behavior, although they seem to be trending up in both cases. Since the frequency of the data is 1, it is hard to discuss any type of seasonality. A cycle also seems not be clear from the data.
   2. For the paperback series, we expect the next four days to sell 207 books. For the hardcover series, we expect to sell 240 books in the next four days.
   3. Paperback RMSE: 33.6 / Hardcover RMSE: 31.9
6. Using holt’s linear method.
   1. Done in R.
   2. RMSE for paperback using Holt’s method: 31.1/ RMSE for hardcover using Holt’s method: 27.2. The RMSE is smaller using Holt’s method, though this is expected since it is a more complex model. The Holt’s method seems to be more adequate, because both series appear to be trending up. However, the estimates of beta for both series are really small, meaning that there is little evidence that the trend is changing over time.
   3. The Holt’s forecasts seem to be better for Hardcover series. For the paperback series, the reduction in RMSE was not as pronounced, so there isn’t much evidence that the more complicated model provides a better fit.
   4. For both SES and Holt’s linear method, the 95% CIs produced by using the RMSE are narrower than the ones produced by the functions.
7. Eggs dataset.
   1. As expected, the forecasts without dumping have a linear downward trend. The estimate for beta is small, meaning that the variation in the trend is almost negligible as time goes by. The estimate for alpha is very large, suggesting that the method relies considerably on new information introduced. Using a damping parameter of 0.85 makes the forecasts barely change. Introducing a box-cox transformation does not change the overall trends.
8. Retail dataset.
   1. Because the seasonal variability is not constant.
   2. Using a damp parameter prevents the forecasts from growing indefinitely into the future.
   3. Using time-series cross validation, the RMSE of the method without dumping was smaller.
   4. They do not look like White-Noise. The Ljung-Box text rejects the null hypothesis of no residual autocorrelation, while the correlogram shows strong correlations especially at lags multiples of 6.
   5. The model with a damp term performs better than the seasonal naïve method.
9. Retail dataset
10. UK cars
    1. The series does not appear to have a clear trend, because it decays significantly in the beginning and then starts to grow. The decrease could be related to the economic crises of the UK in the late 1970s. There is a clear seasonal pattern though.
    2. As the STL decomposition show, the seasonal variation has been mostly constant throughout the period.
    3. Done in R
    4. Done in R
    5. Done in R
    6. The model without the damped parameter on the seasonally adjusted data
    7. The ets() one because it accounts for seasonality in the forecasts.
    8. The residuals display some autocorrelation, especially at lags 4 and 8.