Homework 2

M7: Predictive Analytics

Due: Jan 10, 2020 (23:59)

This assignment is to be carried out individually. You can discuss your solutions with your friends, but please write your solutions individually. Submit your solution, including all the code scripts, as a single PDF or JPYNB file via LMS. (This exercises are from the textbook. All data are posted at LMS.)

Late Homework Policy: If you submit your homework late up to 7 days, there will be a 20% penalty. Any homework will not be accepted after 7 days.

1. Download retail time series data. These represent retail sales in various categories for different Australian states, and are stored in a MS-Excel file. Consider the series in Column Q.
    a. Why is multiplicative seasonality necessary for this series?
    b. Apply Holt-Winters' multiplicative method to the data. Experiment with making the trend damped.
    c. Compare the RMSE of the one-step forecasts from the two methods. Which do you prefer? Write down the mathematical equations for the preferred method.
    d. Check that the residuals from the best method look like white noise.

2. For your retail data, find the appropriate order of differencing (after transformation if necessary) to obtain stationary data.

3. Consider **wmurders**, the number of women murdered each year (per 100,000 standard population) in the United States.
    a. By studying appropriate graphs of the series in R, find an appropriate ARIMA(p,d,q) model for these data.
    b. Should you include a constant in the model? Explain.
    c. Write this model in terms of the backshift operator.
    d. Fit the model and examine the residuals. Is the model satisfactory?
    e. Forecast three times ahead. Check your forecasts by hand to make sure that you know how they have been calculated.
    f. Create a plot of the series with forecasts and prediction intervals for the next three periods shown.
    g. Does auto.arima() give the same model you have chosen? If not, which model do you think is better?

4. Consider **austourists**, the quarterly number of international tourists to Australia for the period 1999–2015.
    a. Describe the time plot.
    b. What can you learn from the ACF graph?
    c. What can you learn from the PACF graph?
    d. Produce plots of the seasonally differenced data $(1-B^4)Y_t$. What model do these graphs suggest?
    e. Does auto.arima() give the same model that you chose? If not, which model do you think is better?
    f. Write the model in terms of the backshift operator, then without using the backshift operator.