ISE-5970: Energy Analytics

Homework 3

Due: Thursday October 3rd, 11:59 p.m.

Before start, please read the following.

1. The questions in this homework allow you to practice your R skills for time series regression forecasting models.
2. For all questions, you must submit 1) the source file that contains the R commands, and 2) the snapshot of what R outputs after you run your R program.
3. I strongly prefer if you **electronically submit** your homework through Canvas by putting all files in a zip folder.
4. Please assign numbers for each solutions, so it would be easy for me to read the answers.

Good Luck! ☺

Question 1 (45 = 5 + 5 + 5 + 5 + 15 + 5 + 5 credits):

The data set fancy concerns the monthly sales figures of a shop which opened in January 1987 and sells gifts, souvenirs, and novelties. The shop is situated on the wharf at a beach resort town in Queensland, Australia. The sales volume varies with the seasonal population of tourists. There is a large influx of visitors to the town at Christmas and for the local surfing festival, held every March since 1988. Over time, the shop has expanded its premises, range of products, and staff. Produce a time plot of the data and describe the patterns in the graph. Identify any unusual or unexpected fluctuations in the time series.

a. Explain why it is necessary to take logarithms of these data before fitting a model.
b. Use R to fit a regression model to the logarithms of these sales data with a linear trend, seasonal dummies and a "surfing festival" dummy variable.
c. Plot the residuals against time and against the fitted values. Do these plots reveal any problems with the model?
d. What does the Breusch-Godfrey test tell you about your model?
e. Regardless of your answers to the above questions, use your regression model to predict the monthly sales for 1994, 1995, and 1996. Produce prediction intervals for each of your forecasts.
f. Transform your predictions and intervals to obtain predictions and intervals for the raw data.
g. How could you improve these predictions by modifying the model?

Question 2 (25 = 5 + 5 + 10 + 5 credits):

The gasoline series consists of weekly data for supplies of US finished motor gasoline product, from 2 February 1991 to 20 January 2017. The units are in "thousand barrels per day". Consider only the data to the end of 2004.

a. Fit a harmonic regression with trend to the data. Experiment with changing the number Fourier terms. Plot the observed gasoline and fitted values and comment on what you see.
b. Select the appropriate number of Fourier terms to include by minimizing the AICc or CV value.
c. Check the residuals of the final model using the checkresiduals() function. Even though the residuals fail the correlation tests, the results are probably not severe enough to make much difference to the forecasts and forecast intervals. (Note that the correlations are relatively small, even though they are significant.)
d. To forecast using harmonic regression, you will need to generate the future values of the Fourier terms. This can be done as follows.

   ```
   fc <- forecast(fit, newdata=fourier(x, K, h))
   ```

   where fit is the fitted model using tslm , K is the number of Fourier terms used in creating fit , and h is the forecast horizon required.

   Forecast the next year of data.

Question 3 (30 = 15 + 15 credits):

Using matrix notation it was shown that if $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, where $\mathbf{e}$ has mean $\mathbf{0}$ and variance matrix $\sigma^2\mathbf{I}$, the estimated coefficients are given by $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and a forecast is given by $\hat{y} = \mathbf{x}^*\hat{\beta} = \mathbf{x}^*(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ where $\mathbf{x}^*$ is a row vector containing the values of the regressors for the forecast (in the same format as $\mathbf{X}$), and the forecast variance is given by

$$var(\hat{y}) = \sigma^2\left[1 + \mathbf{x}^*(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{x}^*)'\right].$$

Consider the simple time trend model where $y_t = \beta_0 + \beta_1 t$. Using the following results,

$$\sum_{t=1}^{T} t = \frac{1}{2}T(T+1), \quad \sum_{t=1}^{T} t^2 = \frac{1}{6}T(T+1)(2T+1)$$

derive the following expressions:

a. $\mathbf{X}'\mathbf{X} = \frac{1}{6}\begin{bmatrix} 6T & 3T(T+1) \\ 3T(T+1) & T(T+1)(2T+1) \end{bmatrix}$

b. $(\mathbf{X}'\mathbf{X})^{-1} = \dfrac{2}{T(T^2-1)}\begin{bmatrix} (T+1)(2T+1) & -3(T+1) \\ -3(T+1) & 6 \end{bmatrix}$