Project

Due Date

The project is due at 5:00 PM Friday March 10 2017.

Do both of the following:

- Turn in a printed copy in lecture or in my mailbox in the stats office of the MSB.
- Send me an electronic submission by email.

Only one submission per group is necessary.

Guidelines and Format

You may work alone or in groups of no more than three students from the class. Use any time series techniques covered in this class to analyze the provided data set, and carry out the steps described below.

Each group should turn in one submission consisting of a report and an appendix with code. During grading, 20% of the total points will be assigned based on the groups' adherence to the following rules.

- Both the hard copy and electronic submissions are turned in on time.
- The submission must include a title page with the names and student ID numbers of all group members.
- The report must NOT contain any code. Only plots and R output should be included.
- All plots and figures must have appropriate titles, axis labels, etc.
- The submission must contain an appendix containing ALL code used.
- You may choose to structure your submission as one report, or as separate entries for the steps outlined below.
- You have to notify Dmitriy of any proposed group changes after March 1.

Data

The file *bostonArmedRobberies.txt* contains monthly records of the number of armed robberies in Boston from January 1966 until October 1975 (McCleary & Hay (1980)).

Steps

1. Use graphical techniques to inspect the data. Describe the behavior of the data. Mention any features you think may be important for analysis and forecasting.

2. In what way(s) is the data not stationary? Use transformations and/or differencing to make the series stationary. Include a time plot of the transformed and/or differenced data. If you chose a transformation, use this transformation for the remainder of the steps.

3. Using the transformed and/or differenced data from the previous part, obtain the sample ACF and PACF plots, as well as plots of the raw periodogram, and its smoothed version. Comment on the plots. Use the plots to make a preliminary guess for an appropriate ARIMA model. Keep in mind that differencing plays a part in determining whether the model should be ARMA or ARIMA.

4. Fit the model from the previous step. Include the model, and the parameter estimates. Plot the fitted values and the observed values on the same plot.

5. Examine the residuals. Provide necessary plots and/or hypothesis test results. Do the residuals resemble Gaussian white noise?

6. Use AICc to select an ARIMA model for the (possibly transformed) data. Keep in mind that differencing should be incorporated into the model. It is fine to use the funciton `auto.arima()` here. It is enough to consider $p = 0, \cdots, 8$, $q = 0, \cdots, 8$, and $d = 0, 1, 2$. Include the chosen model, and provide parameter estimates and their standard errors.

7. Inspect the residuals of this model. Provide necessary plots and/or hypothesis test results. Do the residuals resemble Gaussian white noise?

8. Plot the (theoretical) spectral density of the final model together with the smoothed periodogram. Comment on the plots. Describe the method you chose for smoothing the periodogram.

9. Now remove the data for 1975 (the last 10 observations). Using only data from 1966-1975 (the first 108 observations), fit an ARIMA model using AICc. Again, it is fine to use `auto.arima()`. Then do the following.
   - Write down the chosen model. Include parameter estimates.
   - Inspect the residuals of this model. Do they resemble Gaussian white noise?
   - Compute point forecasts of the values for January through October 1975. If you used a transformation, then be sure to compute forecasts of the *original* data, not the *transformed* data.
   - Make a time plot of the entire data set, the point forecasts, and 95% prediction intervals. Make another plot of just the observed values from 1975 along with the point forecasts and the prediction intervals. Comment on the forecast performance.