Project 1 – Due Tuesday, April 7

STAT-GB.2302, STAT-UB.0018: Forecasting Time Series Data

Use ARIMA methods to forecast a time series data set of your choice, subject to the requirements below. The data set should have at least 50 observations. It would be a good idea to show me a plot before you start the project to avoid potential data difficulties. Please provide a source for the data (web link, but not the course website). The last observation should be recent, and must be the most up-to-date observation available from the source.

Start by plotting the data. Briefly describe any patterns you see. Decide whether you need to take logs. If the data are clearly seasonal, e.g., monthly sales of beer, or if the plot seems to indicate a strong seasonal pattern, then it would be best to remove this component. (It can be put back in at the end.) The easiest way to remove a seasonal component is to subtract out the seasonal averages. For example, if your data is monthly, subtract from each data value the average for the corresponding month.

Use the ACF, PACF, and AIC_C to select an ARIMA model, and to decide whether a constant term should be included in the model.

Present the parameter estimate printout for the selected model. Write the complete form of the fitted model. (You *should not* give the printouts for the other models you tried. A table of the AIC_C values for all models will be sufficient. Furthermore, I do not need to see a printout of the data set. A plot is much more informative.)

For the select model compute and comment on the Ljung-Box statistics for the first 12, 24, 36, and 48 lags. Plot the residuals and the ACF and PACF of the residuals. Comment briefly on any problems revealed by this diagnostic checking.

Plot the data, together with the forecats (at lead times 1–50 or further) and the 95% forecast intervals. Comment briefly on whether the forecasts seem reasonable, and on whether the forecast intervals seem excessively wide.

Length of Project: Please be brief. Just tell me what you did and why. Long explanations are not needed. I think you should be able to do this project quite adequately with 5 pages of text, not including figures and tables. Please don't submit a huge number of figures and tables. I would hope that the whole project, including figures and tables will be at most 10 pages. This is not an absolute limit, but there is really no advantage to be gained by exceeding it.

Work Independently: You can discuss your project with other members of the class, but you should not share your report, your *Rmd* files or your dataset with other students. Copying and modifying another student's code will be considered to be plagiarism.

If You Need Help: I will be happy to talk to you about your project, after class or in my office.