The time has come for you (if you haven't done so already) to start working on your project for STAT 520. Remember that this project counts as 25 percent towards your final grade, so it is very important. By "project," I mean a written exposition that thoroughly describes the complete analysis of a data set. Of course, for us this will mean forecasting as well. After all, forecasting is an important part of time series analysis.

It is important to think about the **four major steps** of analyzing a time series data set.

1. Model specification (Chapter 6)
2. Parameter estimation; i.e., model fitting (Chapter 7)
3. Model diagnostics (Chapter 8)
4. Forecasting (Chapter 9)

In the model specification phase, your goal is to come up with a small set of candidate models for your data. In the estimation phase, you are to fit the models. In the diagnostics phase, you are to check the adequacy of the model fits (and possibly revisit the model specification phase at this point, based on what you learn from the diagnostics). After diagnosing the model fits, you should choose a final model and forecast future observations.

Here are some guidelines for choosing a data set:

- Use other sources (e.g., data sets on line, other research projects, etc.) to find an interesting data set. The more interesting the better! You could also collect your own data (but you may want to get my blessing first if you do this).
- The responses Y_t should be continuous in nature (or at least somewhat continuous; just not binary or strikingly discrete). Also, make sure you completely understand the sampling frequency; e.g., are the data collected every day? every week? every month? every year? every minute?
- Choose a data set in an area you are interested in! You should be able to demonstrate a working knowledge of the subject area (this may mean you have to do a considerable amount of research to get to this level).
- Don't choose a data set that is very small (e.g., let's stay away from series where $n < 50$). Ideally, we want $n > 75$ or so, but this is just a guideline. Remember that many statistical methods we have discussed exploit asymptotic distribution theory, so we want to apply these methods to suitably lengthy data sets.
- Because we won't discuss seasonal models until the end of the semester, you may want to strongly consider staying away from series that exhibit seasonality. **Exception:** If you want to work with a data set that has seasonality, you may be forced to use the trigonometric regression methods from Chapter 3 or you can read ahead and learn about seasonal ARIMA models (Chapter 10) on your own.

As a first step, I would recommend that you get your data and go through the methods in Chapter 6 to identify a small set of candidate models. At this point, if you want to bring your results to me and have me look at them, I can give you advice on your specifications. **Rationale:** If you do a bad job at this phase, it may make the rest of your work less meaningful. You could stop by my office (preferably during office hours) for a short meeting. Please be organized if you do this.

Another recommendation: It might be a good idea to "withhold" some of the data from your series towards the end of it so that you can compare your forecasts to the actual values of the process. For example, suppose that you have a series of length $n = 100$. Perform the specification, fit, and diagnostics on the first 95 observations and withhold the last 5. Then, when you forecast, you can compare your first 5 forecasts to the last 5 observations−this will give you an idea on how accurate/precise your forecasting is. Hopefully, your forecasts are "close" to the actual values in the series!

Outline of the written project (in this order):

- **Title page and abstract.** You must prepare a title page with an appropriate title and abstract. The abstract should go on the title page. An abstract is a very high-level written summary of the entire project. Main points and findings only. Aim for about 200 words.

- **Introduction**. This part introduces the reader to the data set and to the area to which it pertains. For example, if you are analyzing giraffe population growth data from southern Kenya, you should describe why this is an important problem to investigate and give the reader a review of pertinent background information about giraffes in Kenya. Basically, introduce the reader to the problem and why it is meritorious of investigation. This should be written at a very basic level (i.e., no mathematics or notation). Remember your reader (i.e., me!) may not know anything about the area in which you are writing. Aim for about 2 pages here.

- **Model specification.** This is the "meat" of the paper and will be the longest in length. In this section, you want to describe, in clear detail, the data analysis used to specify your candidate models. Pretend like you are taking me by the hand and leading me through your thought process which leads to your model selections.

- **Fitting and Diagnostics.** This part of the project should describe the model fitting and diagnostics techniques you used, with the goal of identifying a "final" model for forecasting. Identify also what possible deficiencies your final model has. Remember, no model is perfect.

- **Forecasting.** This section should describe the techniques you used to forecast future observations (see "Another recommendation" at top of this page). Why is forecasting important? What impacts could your forecasting have?

- **Discussion.** Here you want to offer a summary of what you did in the project and draw your main conclusions. Also, it is a good idea to discuss here other issues

related to the data analysis. For example, does your analysis have any shortcomings or lack of generalizability? What were the main problems you encountered? It is OK if your final model is not "picture-perfect." Few are. Real life data analysis is often more difficult than textbook problems.

- **Bibliography.** Cite all references carefully.
- **Appendices.** Use appendices to catalogue extra graphics/plots/output. You could also give the values of the series here as well (although this is not necessary). Basically, I use an appendix to house information that I want the reader to have access to, but feel that it would interrupt the flow of the main body of the paper.

Here is more general advice:

- Keep everything double spaced (my preference, so I can write comments). An abstract can be single spaced.
- Break your report into sections. Each section should have a title. Use subsections (with titles) if necessary. Avoid subsubsections.
- Integrate R graphics and output into the written text as you see fit. For example, if you want to show me the time series itself (you better!), embed it into the written work. Use the style of my course notes as a reference. It is important to strike a balance here!! You don't want too many graphics in the written text, because it may make your report seem "fragmented." Be judicious with your choices; don't be afraid to log ancillary graphs/plots/output in appendices.
- I am a very picky reader! I do not like errors in grammar, errors in the use of punctuation and capitalization, and spelling errors (I loathe these). These types of errors are clear signs of lack of interest and of the author not paying attention to details. My advice: Edit and reread your written project at least 10 times. Get other people (not in statistics) to read your project and offer comments/feedback.
- Adopt a writing style that you are comfortable with. Written projects do not have to be terse and cold. I think the best writing makes me feel like the author is right next to me reading it. There are no "correct" writing styles, but there are certainly bad ones.
- Have fun!! This is an opportunity for you to amalgamate all of your time series knowledge and apply it to a real problem. Show me what you have learned. Remember, the written project is how you disseminate your work. I think the most important part of a statistical analysis is clearly communicating it to others in writing.

Due Date: November 25, 2013. However, you can turn your final written project in anytime before December 6, 2013 at 1.00pm and receive full credit. Projects turned in after 12/6/13 at 1.00pm will receive at most 1/2 credit.