PSTAT 174/274 COURSE PROJECT

The course project is an opportunity for students to apply time series techniques to real-world problems. Data and software for the project can be obtained from various Internet sites, or developed by students.

You are encouraged to collaborate on this project with the rest of the class - feel free to discuss your ideas with other students. The "deliverables", however, should be unique to you. Your project will be graded on its merits; plagiarism will result in a score of 0.

Deliverables and deadlines. There will be two "deliverables" on this project.

Project report is due on **December 6th, 2017**. Submit in one pdf file by e-mail to feldman@pstat.ucsb.edu.

Project defense is required in addition to a written report. It should take place during finals week or before; **scheduling of the defense is your responsibility.** Please prepare a 5-minutes summary of your findings.

Project Report. The project report should contain the following:

1. Abstract or Executive summary should be one–two short paragraphs summarizing briefly the questions you addressed, your time series techniques, key results, and conclusions.

2. The main body of the report should cover the following:

Introduction. Restate your problem, including details. Describe the data set and explain why this data set is interesting or important. Provide a clear description of the problem you plan to address using this dataset (for example to forecast) and include techniques you use. Describe results (positive and negative) and briefly state your conclusions. Please acknowledge the source of your data and software used.

Sections.
- Plot and analyze the time series. Examine the main features of the graph, checking, in particular, whether there is (i) a trend; (ii) a seasonal component, (iii) any apparent sharp changes in behavior. Explain in detail.

- Use any necessary transformations to get a stationary series. Give a detailed explanation to justify your choice of a particular procedure. If you have used transformation, justify why. If you have used differencing, what lag did you use? Why? Is your series stationary now?

- Plot and analyze ACF and PACF to preliminary identify your model(s). Explain your choices of suitable p and q here.

- Fit your model(s): Estimate the coefficients and perform diagnostic checking. Compare at least two models to choose the final model and explain how you decided on your "best" model. Is the model obtained by using AIC(C) the same as one of the models suggested by ACF/PACF? Write the fitted model in algebraic form. Do you conclude from the analysis of residuals that your model is satisfactory?

- Do forecasting. Make sure to include confidence intervals. Make sure to return to original data. Plot the original series and the forecasts.

- **Only for PSTAT 274 students**: Perform spectral analysis of your model.

3. Conclusion Section. Reiterate your conclusions referring to the goals of your project. Were these goals achieved? Record the math formula for the model you chose. Acknowledge all individuals who helped you with this project.

4. References.

5. Appendix. Include your code with comments.

Report should not be long; please do not add words for the sake of being wordy. The report must be self-contained, i.e. if you use formulas, write them. Include all necessary plots either in the body of the report or in appendix. When including R outputs, analyze the outputs and explain what you plan to do next, why and how.

Time Series Data Libraries:

Prof. Rob Hyndmans, Time Series Data Library (TSDL)

`https://datamarket.com/data/list/?q=provider:tsdl` DataMarket

KDnuggets Datasets: `http://www.kdnuggets.com/datasets/index.html` KDnuggets

Please acknowledge the source of your data in the project.   **Project diagram – see page 2.**

GOOD LUCK !!!

Model building

DATA

$(X_1, \dots, X_n) = (x_1, \dots x_n)$

Postulate class of models

Class of models: log/power transforms, seasonal, non/stationary

Identify particular model

Identify p and q for ARMA using acf/pacf

Go back

Estimate model parameters

Estimate from data $\quad \hat{\varphi}, \hat{\theta}$
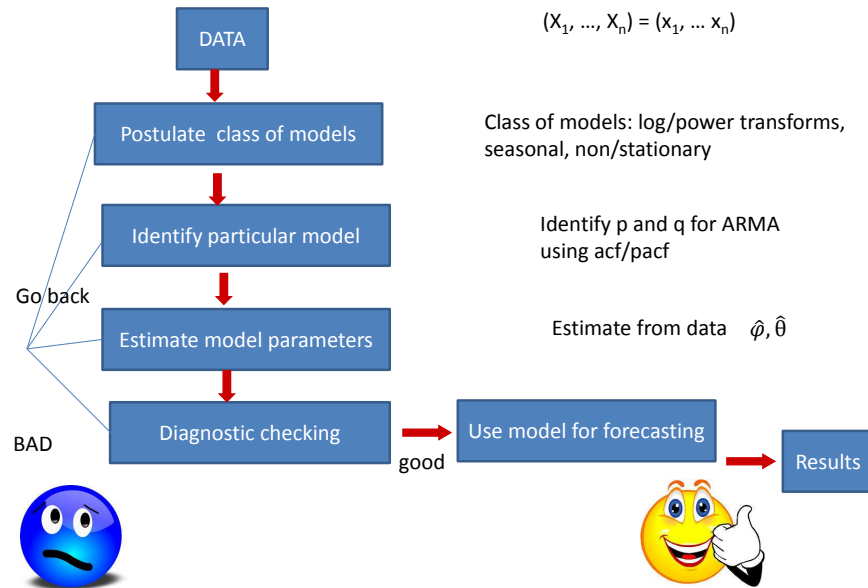
BAD

Diagnostic checking

good

Use model for forecasting

Results

Figure 1: Model Identification Diagram