

PSTAT W 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. Please read carefully policies on use of automated writing and coding technologies. Instructor reserves the right to request an interview for clarification of the answers.

Deliverables and deadlines.

Project report is due on the **last day of classes**. Submit as attachment in one pdf file by e-mail to raya@ucsb.edu. The subject line of your e-mail must start with **your first and last name**, e.g., the subject line for a student named Tim Ser would be *Tim Ser - PSTATW 274 final project submission*. Please use the following format for the name of the file, e.g., *Tim Ser-174 final project.pdf*.

Project Report. The project report should contain the following:

1. The project title and the name of the author. The title should reflect what your project is about. Just like books have individual titles, your project should have a unique name.
2. Abstract or Executive summary should be one–two short paragraphs summarizing briefly the questions you addressed, your time series techniques, key results, and conclusions.
3. 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. Modeling and forecasting should include:

- Divide data to training and test sets. Use training set for modeling and tests set for validation.
- Plot and analyze the time series (training set). 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 AICc 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 W 274 students:** Perform spectral analysis of your model.
- **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 including fellow students, teaching assistants, etc.*

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, that is, 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.

Please see a separate document with suggested websites for time series data. Do not use datasets identical to those used in the lecture slides, homework or labs, for example, tsdl file 484 on Boston robberies or monthly milk production tsdl file. You may use similar datasets, e.g., different city, different years, etc.

Please acknowledge the source of your data in the project.

Use of auto.arima and other automated technologies.

The use of AI coding and writing technologies falls within the purview of the Student Conduct Code and the Student Guide to Academic Integrity. It states that “Materials (written or otherwise) submitted to fulfill academic requirements must represent a student’s own efforts unless otherwise permitted by an instructor.” In this course, student use of automated technology for coding and writing is not allowed.

Students are expected to follow all steps of time series analysis on their own while using auto.arima, checkresiduals or other automated codes is not acceptable. To get credit, students are required to provide detailed explanations on how they read the outputs and how they choose the next steps.

GOOD LUCK !!!

Project flowchart

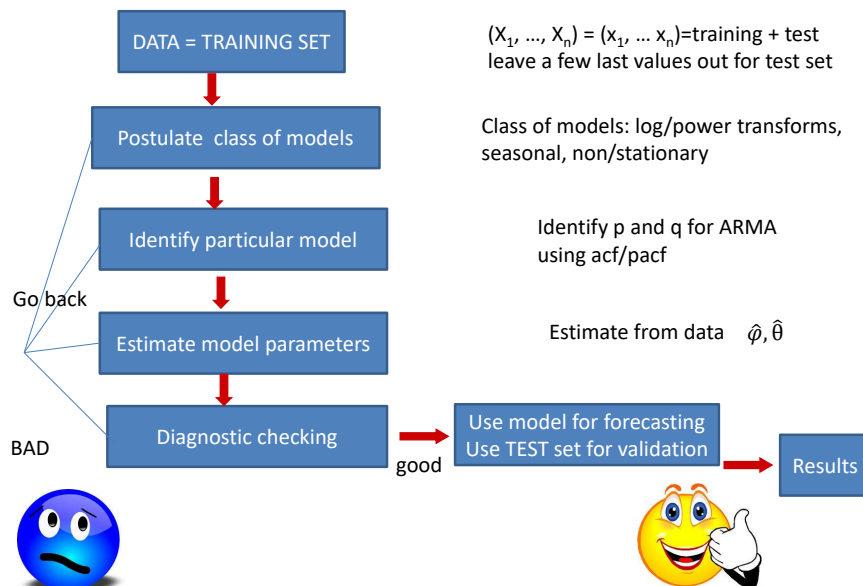


Figure 1: Diagram: Project Steps