

Applied Forecasting

Instructions for Final Project

You've learned lots about doing time-series forecasting analyses. It's time to work without a net....

The plan

The only satisfactory way to learn how to forecast is to actually do it. It is also better to form forecasts and then to evaluate them in real time rather than using a contrived situation. A partial requirement for this course is to undertake a forecasting project of the type outlined below. You will present your results late in the semester to the class. Technically, the submission of the final project should include the following components:

- 1) The research-article-like report, which should be 1800-2200 word-long. (70% total grade).
- 2) The presentation poster: you can find the instruction for the poster generation in the later pages of this instruction (20% total grade).
- 3) The data file + code file RScript.

We have essentially a fourteen-week college semester. The outline we will follow for the requirement of the project will be as follows:

The Data

The most important decision to be made is what variables to use in the project. The series should be available **daily, monthly or quarterly**. In some rare cases daily or hourly data may be appropriate for the project (but daily, monthly or quarterly date is highly recommended). In any case, a good sequence of previous values must also be available. At least the previous 24 to 48 periods should be found. If the series is suspected to have strong seasonality, values for the corresponding periods for the last three years will be required.

If possible, it is best to forecast a series in which you have some real interest. Examples could be monthly sales in some industry, the number of telephone calls made from an office per quarter, or hours of mainframe computer time used monthly. The financial press is an excellent source of monthly data. Journals and newspapers such as *Barrons*, *The Economist*, *The Wall Street Journal*, and *The Financial Times* have many pages of suitable series, such as automobile sales, airline miles travelled, or deliveries of some produced items. For macroeconomics data, *WorldBank*, *IMF*, *ADB*,... would be excellent, while the trading daily data could be found at cophieu68.vn in Vietnam, and investing.com for international contexts. Furthermore, any time-series database that is reliable and transparent is good to use to as long as the data quality is relevant to support for your analysis.

It is strongly recommended that the series to be forecast is not a price taken from a highly speculative market such as stock or bond prices, the prices of major commodities, such as copper, silver or gold, an exchange rate or an interest rate. Although such series are certainly readily

available and are of general interest, it has been well established that **changes in any price or interest series are extremely difficult to forecast and so are inclined to make disappointing projects (i.e., do not use prices or interest rates as your data to be forecasted).**

You will need some data other than the values of the series to be forecasted in order to complete the project. This other (explanatory) data to be gathered is information that might be helpful in forecasting the series of interest. For example, if one is attempting to forecast electricity sales to households in some region, then a measure of local temperature could be a good explanatory variable. The search should be for a "causal" series or a "leading indicator" of the series to be forecast. All series should have data available over the same period (and of the same periodicity). One can use common sense or economic theory to choose likely explanatory series. Of course, several possible explanatory series could be utilized, and this would make the project rather more rewarding and, perhaps, produce a better forecast.

Initial Analysis

Each series should first be plotted through time. These plots give a good indication of the general properties of the series: are they smooth or jagged; is there a monthly or annual swing in the series; and does the series contain a trend in mean? It is also very important to ask if the series contain strange values, called outliers (Data visualization can help to detect these problems). If they exist, they can be seen easily from the plot. Outliers can arise either from a genuine strange period in the data--due to a strike, typhoon, or heavy snow storm, for example--or can be found because the data was misrecorded or incorrectly plotted.

If an outlier is found, it is suggested that for the purposes of the project, a simple method of removal should be used, such as replacing the outlier by the previous, acceptable value or by the average of the two terms on each side of the extraordinary value (be certain to advise the instructor that you are handling any outlier this way).

Moreover, the time series decomposition is also an ideal analysis to better understand the seasonality, trend-cycle, and remainder from the time line of the data.

Of course, as other quantitative data, time series can be analyzed simply via the descriptive analysis; therefore, some short analysis on the mean, max, min, standard deviation... of the data is also required.

The Various Forecasts

The text contains discussions of several forecasting methods. Some use just the past values of the series to be forecast as discussed in Chapters 3 (benchmarking methods, such as naïve, average, drift...), 5 (time series regression), 6 (decomposition, STL), 7 (exponential smoothing), 8 (ARIMA), and 9 (dynamic regression models) of the book and others also use other series (you can extend the advanced methods such as the Vector Autoregression (VAR), Neural Network (ANN) models (in chapter 11) or Long-short term memory methods (LSTM). Chapter 3 has a discussion about combining the results of the various forecasts and compare the accuracy (forecasting

effectiveness of the model) you have made in the project. In short, after implementing different methods, please provide the evidence for selecting the most effective model.

After having already the action plan for the project, the following part will instruct the requirement in the [Final Report](#) and the [Final Poster Presentation](#).

Final Project Report

In this project, I require an outcome of a research-paper-like report, which should follow the format of a journal article (you can find the sample in the Canvas Folder). A typical assignment report should content the following components, (but not limited to)

1. Executive Summary

Abstract
Intro 1 2 4 6

- 1 What is the topic of your project?
- 2 What are the main issues or problems the project to address?
- 3 What are your plans for obtaining background information (if needed) about your project?
- 4 Describe the data that you used or collected, including the variables measured. You don't have to give a detailed version of your data collection design; the detailed design plans should be represented in the later parts of the paper.
- 5 What questions and/or concerns do you have about your project?
- 6 What is your key findings from the empirical analysis?
- 7 What is your recommendations/ suggestions for managerial decision making related to the results of analysis?

2. Data collection and background writing

In this report, you collect your data by your own, so please feel free in choosing the way you want to conduct the data collection and present the data collection process.

The most important aspects of any statistical analysis are stating questions understanding data. Hence, to get the full experience of running your own study, the assignment requires you to write a data background to understand your data.

Moreover, although data cleaning is definitely a tedious task in data analytics, please spend enough time for this stages, checking your data validity, their format, cleaning the missing values (or replacing them); in short, try to do everything you can to prepare the data for the further steps.

3. Literature review III

As this is the final exercise of a course, a short paragraph of literature review is required. You should provide some background knowledge about the topic that the data will involve in, 3 to 4 journal research papers from Google Scholar will be a very good resource for these

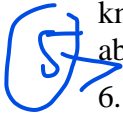
purposes. Remember, all good analysis come from excellent literature review foundation.

4. Data analysis result

5. Story telling/ Data interpretation and conclusion.

+ 7

As the basic implication of this course, I want you to tell me the story behind the data, what the graphs and tables imply, and there are the underlying relationships, linkages behind the variables. You could read some literature reviews (which should be available online, and some good academic sources might come from Google Scholar summary), based on your knowledge, and quick brain storming process, write out words to tell an interesting story about the data.



6. Reference.

For an academic writing paper, it will be highly comprehensive and relevant if the writers could make proper citation. Therefore, the paper citation with reference will be an extra if students can provide some.

Final Project Poster Presentation:

Besides the writing report, your project will be also presented in a poster session during the last week of lab sections. In a poster session, each group makes visual materials that explain the project. Then, people wander around looking at the posters and talking to the presenters, thereby learning about the various projects. Poster sessions are extremely common at professional conferences in many disciplines, including statistics. In our poster session, some members of each group are stationed at the poster to answer questions, while the others wander around to examine the projects. The poster-sitters and wanderers switch off after the wanderers have examined all the posters.

You can find the instruction of how to present your idea and the statistical analysis result on a research poster from these links:

How to make an academic poster in Powerpoint:

<https://www.youtube.com/watch?v=WnhoIbfc0M>

Some general guidelines for making a better research poster:

https://www.youtube.com/watch?v=AwMFhyH7_5g

An example of poster presentation:

<https://www.youtube.com/watch?v=vMSaFUrk-FA>

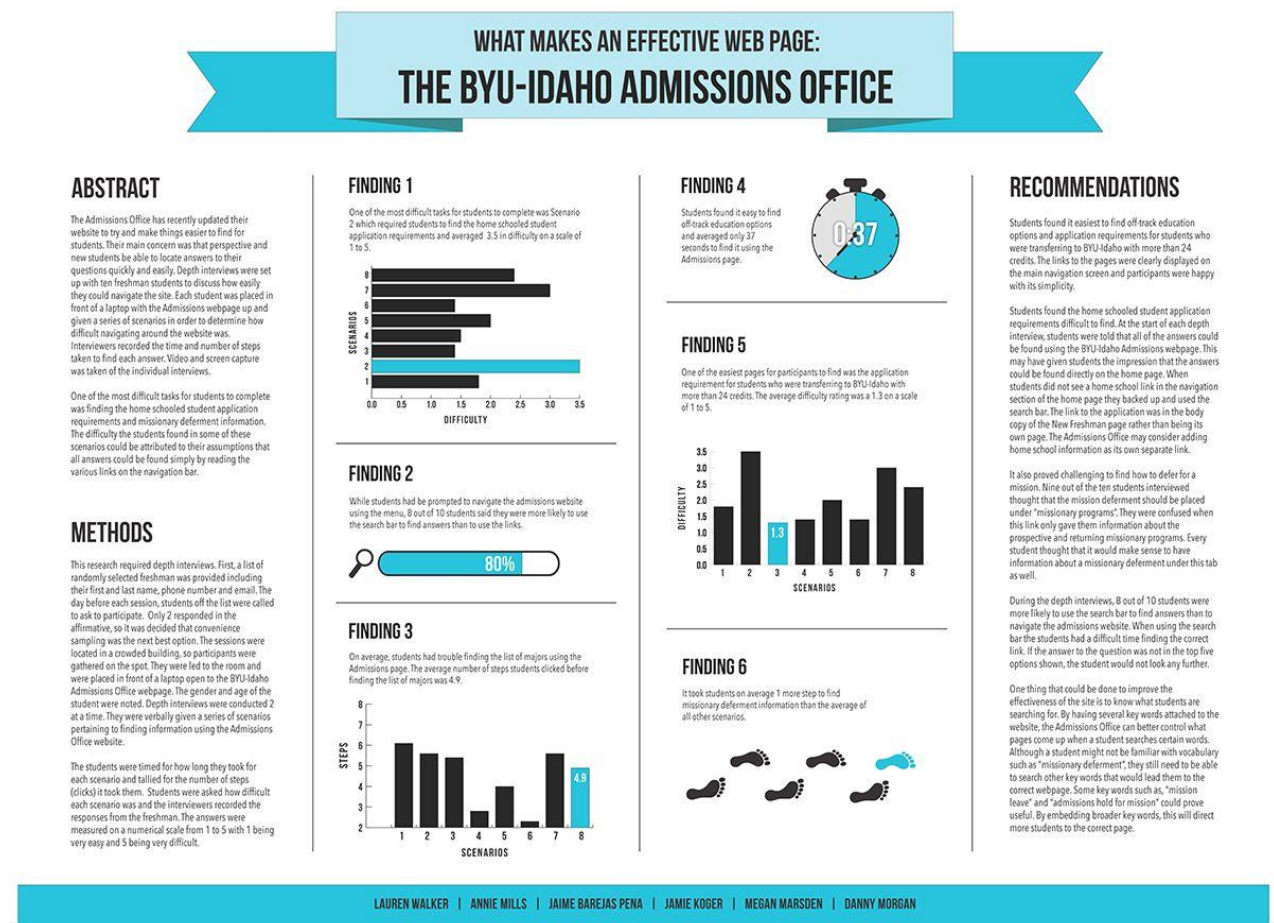
Moreover, here provides some resource for the effective poster templates that you can refer and follow:

https://www.makesigns.com/SciPosters_Templates.aspx

<https://www.posterpresentations.com/free-poster-templates.html>

<https://www.postermywall.com/index.php/posters/search?s=data%20science#>

This is an example of a research poster:



Assignment grading guidelines

You will be graded by your instructor, who will be looking for the following characteristics:

1. **Consistency:** Did you answer your question of interest?
2. **Clarity:** Is it easy for your reader to understand what you did and the arguments you made?
3. **Relevancy:** Did you use statistical techniques wisely to address your question?
4. **Interest:** Did you tackle a challenging, interesting question (good), or did you just collect very simple statistical analysis results (bad)?

Some suggestions for scoring high on these criteria, and suggestions you should keep in mind whenever you write anything, are the following:

1. State your question up front and use statistics to help answer it. The statistics should not drive the question; the question should drive the statistics.
2. Don't just collect data and publish it, rather have a specific question in mind. Otherwise, you wind up being hard-pressed to come up with something challenging and interesting.
3. Most importantly, talk to your instructor for advice. You can ask them, for example, about your planned methods of analysis and see what they think.

4. Be selective with computer output to help clarity.

If you are using techniques we learned in class, you do not have to re-explain the techniques. That hurts clarity. If you are using techniques that we did not cover in class, you should definitely explain the techniques. That is clarity!

Procedures for when group members are not contributing their fair share

Each group should spread the work among members so that everyone shares in the project. If some group members do not contribute their assigned workload, or are unwilling to take on work, your group may petition to have such group members dropped from the group. The process of this petition proceeds as follows:

- 1) Send an e-mail to the instructor explaining how the group members have not contributed adequately. **ALL MEMBERS OF THE GROUP MUST BE SENT THIS E-MAIL.** This is to ensure that everything is done openly.

