Midterm

Parameters

This should be submitted via blackboard. I am happy to answer any questions you may have.

Please treat this like a professional report for someone who is not familiar with time series analysis.

All answers should be concise but as informative as possible. All graphs and tables should be legible and labeled.

Please Limit your response to 15 pages.

Please provide three files when you submit your midterm:

1. A PDF/Word document with your answers.
2. A .Rda/.xlsx/.csv file with your data.
3. A script with clean, commented code that replicates your analysis.

Grading will be based on the following criteria:

- Report structure and clarity - 20 points.
- Literature review - 20 points.
- Data description - 30 points.
- Stationarity - 15 points.
- Modeling (be detailed) - 50 points.
- Serial correlation - 10 points.
- Forecasting and forecast combination - 15 points.
- Code accuracy and clarity - 20 points.

The Midterm

Think of a question you would like to explore that involves four or more time series. For example, you may want to explore the relationship between GDP, unemployment, inflation, and monetary supply. Alternatively you could think of something less traditional, such as the relationship between search history for a particular term, sales of a particular product or products, and advertising expenditures. Be creative!

After you have thought of your question you will need to do the following:

1. Write a few paragraphs explaining your question and why you find it interesting. What do you expect to learn from it?
2. Search for academic articles that may have already explored your question or a related question using the data.
    - Write a brief literature review (1 paragraph per article is more than enough) of the articles you found.
    - Remember to cite your articles in an appropriate manner (APA, Chicago, etc.).
3. Identify data sources and obtain each of the time series in your question.
    - Remember, you need to have four time series.
    - Your series should be at least 100 observations in length and of the same frequency.
    - Avoid stocks and indices, as they are typically random walks.
4. Write a few paragraphs about the data you have collected.
    - Who collects it and how is it collected? Survey, administrative data, etc.
    - What are the limitations of the data? Are there imputed data, missing data, measurement issues, etc.? You may need to search the technical notes for the data to find this type of information. It can be important!
    - Is it seasonally adjusted? Is it transformed in any way (e.g., year-over-year growth, log transformation, etc.)?
    - What are the units of measurement? Are they consistent across series?
5. Define strict and weak stationarity. Determine if each time series is stationary. Use all available evidence to support your conclusion.
6. Choose one of your time series. Split the data into a train (1:(T-6)) and test ((T-5):T) set it. Model the training set using the Box-Jenkins methodology (ARIMA).
    - Define the model selection criteria you will use to determine the best model and describe its relative strengths/weaknesses.

- Evaluate the ACF and PACF of the stationary data to determine what model to start your analysis with, be specific about why you choose your starting point.
- Estimate the starting model and move from general to specific using the ARIMA framework. Provide a table of at least three different models including your "best" model as indicated by your information criteria. Be detailed in explaining your decision process.

7. Define serial correlation and how to test for it. Include the appropriate maths in your explanation. Test your models for serial correlation and discuss the results and how they impact the choice of "best" model.
8. Using each of the three models you provided in the table, forecast the next six observations of the time series and evaluate the forecast.
   - You will need to choose a loss function for evaluating the forecast and define it.
   - Which model is the better forecast based on your loss function?
9. Define the concept of forecast combinations and provide at least two specific examples with a detailed explanation.
10. Choosing one of the forecast combination methods create a combined forecast for the next six observations of the time series using the two models that are not the best model. Compare the combined forecast to the best model forecast and discuss the results.