**Time Series Project: Fall 2024**

**The project:**

1. Pick a data set that has two or more variables recorded over time (similar to the Schumway LA air quality data from Unit 12. (dataset: *lap* from package: *astsa*) (can’t use this one ☺ )

2. Select a response from the data set.

3. Be creative and come up with a scenario as to why a client would want to analyze this data and why this response is important! Or better yet, use a real problem that you are interested in!

4. Fit models from the following four categories (provide all plots and tables needed to ID these models: acfs, spectral density, factor tables, etc.):

a. At least 2: ARMA / ARIMA / ARUMA / Signal Plus Noise (univariate analysis)

b. At least 1: VAR with at least one explanatory variable.

c. At least 1: Neural Network (mlp)

d. At least 1: Ensemble model using at least two of the above. (this model does not have to “beat” your other models.

5. Pick a short and long term forecast horizon based on your “problem” from part 3 above and compare all models with the ASE and the rolling window RMSE for both the short and long term forecasts … this does not mean you have to choose the model with the lowest ASE or rwRMSE.

6. Provide the forecasts and prediction limits (when possible) for both the short and long term forecasts.

7. Create a ppt and a 7-minute video (with Zoom or YouTube) describing your analysis (more info below).

8. Post that video to you-Tube and the (private) link (or the Zoom link) to the Google-Doc and submit your ppt and Rmd File (or Jupyter notebook) to 2DS. Please leave the link on the Google Doc for a week so others can learn from your presentation. Please check out at least 3 of your peer’s presentations and please watch your own presentation as well. It is often very useful (although always a bit awkward for me at least ;) to watch yourself present! (Note: if you use the Zoom link, make sure you make it public so that I and your peers don’t need a password.)

**Individual:**

This is a group of two project.

**Intended Audience:**

You can assume your audience have taken an introductory course in statistics and another in time series and are familiar with ACFs, spectral density plots, ARMA/ARIMA models, AIC, ASE, and the tests that are associated with these models (ie Dickey-Fuller Test). However, they may be not be as familiar with VAR modeling and deep learning models as they were not as accessible when they went to school.

**Sources of Data**

See chapter 1 of the textbook.

**Deliverables:**

**Sunday, December 8th at 11:59pm CST**

Deliverables:

1. 5 to 8-minute YouTube (or Zoom) video:
2. Strong Introduction
3. Describe Data Set / Time Series (Who, What, When, Where, Why and How)
4. Stationary / Non-Stationary
5. ACFs and Spectral Densities
6. Univariate: At least 2 candidate ARMA / ARIMA models
   1. The models in factored form or at least separate the stationary and non-

stationary factors with standard deviation or variance of the white noise.

* 1. AIC
  2. ASE (short and long term forecasts)
  3. Rolling Window RMSE (short and long term forecasts)
  4. At least 10 superimposed spectral densities from 10 generated

realizations like we did in Unit 11. Use these to help choose between the at least two candidate univariate models.

* 1. Visualization of Forecasts for both the short- and long-term Horizons.
  2. Be sure and include confidence intervals when possible (I don’t have

code for confidence intervals from MLP models at the moment… but that would be a good thing to work on! ☺ )

1. Multivariate: At least one multivariate model (VAR or MLR with Correlated Errors)
   * 1. Include an ASE (rolling window is not yet available in multivariate models)
        1. Short Horizon (you pick the length.. could be one step ahead)
        2. Long Horizon (you pick ... just must be longer than the short horizon.)
     2. Describe the explanatory variable(s) used in the model and why you felt they were significant / important.
     3. Visualization of Forecasts for both the short- and long-Horizons.
     4. Be sure and include confidence intervals if using VAR …
2. At least 1 MLP Model
   * 1. ASE (short and long term forecasts)
     2. Rolling Window RMSE (short and long term forecasts (only if univariate)
     3. Visualization of Forecasts for both the short- and long-term Horizons.
     4. Confidence / Prediction intervals are not required (I don’t have

code for confidence / prediction intervals (bootstrap intervals) for MLP models at the moment… but that would be a good thing to work on! ☺ )

1. Model Comparison and Final Forecasts
   * 1. Provide a table comparing all models on at least ASE and rwRMSE (if available).
     2. Include at least one ensemble model in addition to the models above.
     3. Make a case as to why you feel one of your candidate models is the most useful.
     4. Provide your final short and long term forecasts with that model.
2. Strong, polished exit
3. Submit your slides to Canvas
4. Submit your RMD and Knit RMD File to Canvas
5. Make sure your video URL is on the Google Doc.

**Google Doc Link:**

<https://docs.google.com/document/d/1yWSMFXJr9MEsM8mfnjCz2Uh0yEQKnIJH_0aByfuFSmA/edit?usp=sharing>

**Rubric for both EDA and Final Presentation;**

Knit RMD: 20%

Well organized and documented RMD knit to pdf or html or docx.

Results / Analysis: 40%

Correct Interpretation

Creating Useful Models

Performing a Complete Analysis: Model ID, Model Building, Forecasting, Cross Validation

Presentation: 40%

* Communication and presentation of your findings are critical to being a successful data scientist. You will be graded on:
  + Voice inflection
  + Slide Organization / Content
  + Visualization
  + Animation (example: Faded bullet points.)
  + Readable and Labeled Plots
  + Composure: This will include **not reading** off of the slides and smoothness of delivery.
  + Pace: Not going a second over time. (8 min for Final Presentation.) Your client is very strict on this point.
  + Polished intro and exit.

**Examples:**

Link to project from Fall 2019:

<https://docs.google.com/document/d/1DbgQO551orifyjHJBXe-lORp-CFa8PQ0JuHNCwqRZe4/edit?usp=sharing>

A particularly good one is Kristen Rollins’ presentation although there are several that are great examples.

Groups:

A screenshot of a chat

Description automatically generated