AAEC 4484/AAEC(STAT) 5484: Applied Economic Forecasting

Your Name Here

Homework #5 - Spring 2025

Instructions: Where necessary ensure that your graphs and visuals have proper titles and axis labels. Refer to the output, whenever appropriate, when discussing the results. **Creativity (coupled with relevance) will be rewarded.**

Question 1: Forecasting U.S. Leisure and Hospitality Employment

Hyojun (your classmate) is a Ph.D. student in the Department of Hospitality and Tourism Management at Virginia Tech. As a part of his blog, he is interested in forecasting employment in the Leisure and Hospitality sector in the US. Although he has a good understanding of the theory and practical application, he is strapped for time and asked you to help him with this task.

You did a quick Google search and realized that the U.S. Bureau of Labor Statistics (BLS) provides a wealth of data on employment in various economic sectors. The `FRED` also hosts this data. You decide to use the `quantmod` package to pull the `All Employees, Leisure and Hospitality (CEU7000000001)` data from `FRED`.

(a) In a single code, pull the series from FRED database using the `quantmod` package and convert to a `tsibble` object with the appropriate time index.

(b) Provide plots of the (i) data, (ii) ACF (with a *max lag of about 3 years*), (iii) `gg_subseries()`, and (iv) transformed data, if appropriate (see discussion that follows). Also comment on your observations.

In particular:

- Be sure to include a discussion of seasonality (is there strong evidence for seasonality?) and or trend.
- Specific to the variance, does the data need transforming before estimations? If so, what kind of transformation would you recommend? **A simple yes will not suffice. Be sure to discuss your reasoning as to what transformation is required in the first place. Note: I would like to see you use the patchwork package to combine the plots. If needed, please adjust the size of the figure in your code chunk to ensure that they are all legible to the reader.**

(c.i) Hyojun looked at the transformed series in (b) and immediately concludes that the series is non-stationary. Can you explain how he came to this conclusion? **Here, I do not need you to perform any tests, I just need you to use your intuition and knowledge of the theory of stationarity to support your answer.**

(c.ii) Proceed by conducting formal unit root testing. You will need to specify the null hypothesis (hypotheses) and corresponding conclusions of the unit root test(s) used.

(d) Perform the Box-Jenkins procedure to determine the candidate ARIMA models for this data. **You can skip any step that you have already explored above. Just make sure you reference that in your answer so that I can follow along.**

- **Be sure to:**
  - **Report your preferred model and explain how you arrived at this specification.**
    - **Note: Unlike our approach in class of writing down and individually coding all the models, I am OK with you testing over ranges (for the lags) and allowing `R` to minimize as appropriate. You will still need to explain your reasoning, though. Instruct `R` to use the `ic = bic` to determine the preferred model.**

- **Report the diagnostic checks to justify the adequacy of your model. Comment on the Ljung-Box (LB) test results. Be sure to account for the appropriate degrees of freedom and lags of the LB test.**

(e) You might notice and recall that COVID-19 had a significant impact on the Leisure and Hospitality sector. Let us account for this by including a `COVID-19` dummy as an exogenous variable in our model.

- Re-estimate your ARIMA model to include `COVID`, a dummy variable that takes the value of 1 from **March 2020 (when the first state of emergency was issued) to April 2023 (when the national emergency was lifted)** and 0 otherwise. **I would like you to use an `if_else` statement to create this dummy variable.**
- Report the results of your model. Is the COVID dummy a significant predictor of the employment series? Explain, in full detail, why or why not.

Note: You can allow `R` to auto-select the best model for you. However, be sure to force the seasonal (`D`) and non-seasonal (`d`) differences to be the same as you had above.

(f) Conduct diagnostic checks on the model in (e) and comment on its adequacy. Again, remember to account for the appropriate lags and degrees of freedom.

(g) Assuming that your model in (f) is adequate, provide a forecast of the employment series for the next three (3) years. If it was not, what would you do to improve the model? **Hint: Remember that you will have to supply the future values for the `COVID` dummy variable in the `new_data()` argument.**

Question 2: Forecasting C02

Carbon Emissions is an ever-evolving and interesting issue among environmental activists and academics alike.

(a) Use the `quantmod` package to pull Annual Transportation Carbon Dioxide Emissions, All Fuels for United States (`EMISSCO2TOTVTCTOUSA`) from `FRED` and declare as a `tsibble` object. **You will need to adjust the time index to match the appropriate years.** Next, drop all observations after 2019.

(b) Provide a time-series plot of the data and comment on its stationarity.

(c) Perform the Box-Jenkins procedure to determine the candidate models for this data. Unlike in Q1, you are not allowed to have `R` auto-select the best model for you. You will have to write down and individually code all the models.

- **Be sure to:**
  - **Perform formal unit root testing. You will need to specify the null hypothesis (hypotheses) and corresponding conclusions of the unit root test(s) used.**
  - **present the AIC and BIC statistics of your candidate model(s) in a table using the `kable()` function. Columns and Rows should be properly labeled.**
  - **Explain how you arrived at your preferred model.**
  - **Report the diagnostic checks to justify the adequacy of your model. Comment on the Ljung-Box test results.**

(d) Using your preferred model, present an `autoplot` of the forecast for the next five (5) years.