Project 3

Project 3 has a solid analysis and coherent narration in presenting the patterns of the Federal Funds Rate and unemployment data in the US. The report begins by a valid personal motivation on the issue of monetary policy and its relative unemployment level, which then affects and is affected by the global economy. The introduction part of this project also neatly discusses the organization of the paper, which helps the reader to navigate the report. Another area of strength is that this report seems to be very careful in determining what plots and which kind of information need to be presented. The limited space of the report is used efficiently and most points are clearly described and elaborated. Their source code shows their data analysis code is reproducible and readable. The choice of the variable name is intuitive, even though giving additional comments on their code can be better.

In the analysis part, the group has shown their careful assessment to determine whether the time-series data is stationary, before fitting the model. Adding KPSS and ADF tests help to emphasize their point about the stationary property of their data better. After indicating a short-term trend, supported with the evidence from the ADF test, the group appropriately decides to detrend the data by taking the first differences, which is a reasonably right step to do. The group has also made an appropriate discussion on model fitting and selection. They also correctly diagnose the violation of the model briefly and effectively. When faced with unusual behaviors in the distribution of residuals, the group also shows that they have done their best to examine if there is indeed a violation of their model assumption by computing profile likelihood. In the part on the regression with ARMA error between the Fed and Unemployment data, the group also shows their great understanding on fitting and diagnosing ARIMA models and seems to be careful with the reverse causality issue in their model.

The first weakness area in this report is the technical aspects. The report can be better if it adjusts the aesthetic and pays attention to details, such as labeling the axis, giving a plot with appropriate names, and giving labels names like "Figure 1" or so in each plot so it is easier for the readers to refer to which plot they are discussing. Further, the report can present a better aesthetic if it uses a table instead of a raw output from the R code in the analysis. For example, it is not hard to report the AIC values by creating an HTML table by using knittr package in their report. It also might be better to use a consistent theme for the plots. The group seems like using a basic plot from R but also uses the ggplot function from Tidyverse with a default theme, which yields an aesthetical inconsistency across the plots. The group can make the plot from the ggplot function with theme_minimal(), for example, so it is consistent with the R basic plot, even though implementing the ggplot function to all plots can also be an alternative.

In terms of substantial weakness, the group made a conceptual misunderstanding on the periodogram. Looking at the presented periodogram plot, it seems there is evidence of cyclical behavior of monthly interest rate, as the sliding confidence intervals show a significant difference between the highest and adjacent peaks. Therefore, it seems not correct to say "the frequency domains of the plots are excessively broad, suggesting a lack of inherent periodicity within the dataset." Ironically, the group also wrote this statement, despite the ACF plot presented before these periodograms showing a lag correlation in the data.

The group also needs to be careful with all the results of the hypothesis tests in the report. In total, there are four formal hypothesis tests introduced, this raises a question, to what extent is the Type I error inflated in their analysis? It might be better if the group does their hypothesis testing in the model selection because this is the main course of the report.

When it comes to analyzing the interaction of the Federal Rates and Unemployment, it might be helpful for the group to demonstrate their point by presenting the cross-spectrum and coherency plots between these. This plot can be helpful to demonstrate their point on the behavior of the two random processes.

Moreover, while the group correctly assessed the potential short-term trends in the annual rate, especially due to the stagflation in the 1980s and the global financial crisis in 2010, the group can actually introduce additional filters to smooth this short trend, so they can obtain the cyclical behavior in the observations. The lack of additional filtering might be the reason why they observed a long-tailed residual in the model.

Further, in the introduction part, it might be better to state briefly what is the finding that the group observes in their analysis. For example, before presenting the structure of the report, the group can write a sentence like, "in the following report we observe there is an interaction between the interest rates and unemployment data. To demonstrate this finding, the first section will talk about [...]", and so on and so on. This will help them to engage the readers with the main idea of the report.

Finally, in the exploratory data analysis part, instead of talking about where they download the data, the group might be better to briefly examine the generating process of their dataset. While it is clear the FRED is generated through the official report and announcement of the central bank, how was the sample of unemployment rate obtained? Does this data contain the total unemployment of the US or just in the urban or rural areas? Is this data discussing about voluntary or involuntary unemployment? By giving space to present the group's understanding on the data being used, the report can generate more credibility to the readers.

Project 16

This project is particularly strong in utilizing available time-series techniques from the course on their data. The group also seems to be careful in analyzing the data in the early part of the report. The group also demonstrates a careful analysis of determining the best model based on the selection of AIC values through explicit hypothesis testing. The group shows an understanding of model diagnosis technique by examining the residuals of the model. Forecasting is an additional value of this project because this technique is not emphasized in the course. Finally, the coding is reproducible, easy to follow, and applying an intuitive variable name, indicating the group has a sufficient understanding in writing a code for data analysis in R.

The weakness of this project is the presentation of the report, the lack of solid exploratory analysis of the timeseries, and the vagueness in the model diagnosis. In terms of report presentation, the project can be better if it can develop a careful narration by connecting their story on the exponential trends in the sales data to the research question of interest. More specifically, the group can make their research question more explicit and justify why this is important. Moreover, stating briefly their argument in the report can also be helpful for the readers. For example, the group can state, "in the following report we observe that [...]." This rhetorical repertoire can help the group to make the reader stay engaged with their narration.

Further, in the exploratory data analysis part, it seems trivial to present the plots of log-transformation and autocorrelation function of non-detrended data. In the first graph presented in the "Importing Data" section, it seems clear that there is an exponential trend with little spikes, so we will not expect different information in the plot on log-transformation of the data. Given the presence of the trend, it seems clear, the autocorrelation function plot will indicate a long-term lag correlation, so this plot also does not add new information to the readers.

Moreover, while the group has correctly decided to take the first difference of monthly data to filter out the exponential trend, the autocorrelation plot of the first difference data still shows short-term and systematic trends. This implies, the group might be better to apply additional filters, such as the Hodrick-Prescott filter, to extract the cyclical trends in the model. For the lack of stationarity in the first difference data, the ARIMA model fitted in the next section of the data exploration does not seem right.

Further, the part on the spectrum analysis might be better placed next to the data exploration, because the purpose of this part of the analysis is to examine the structure of the data in the frequency domain, which is "an exploration" of the data. The group also seems to incorrectly conclude the cyclical trends in the first-difference of the sales data because the confidence interval does not show there is a significant difference between the peak of the frequency and the adjacent local peaks. It also might be better to state explicitly what is the default span implemented to smooth periodogram and what the different information conveyed by the smoothed method by AR.

In the modeling diagnosis part, it seems unclear why the conclusion that "two roots of MA are lying on the unit circle," leads to the decision to fit the ARIMA(p, 0, q) model, while it is clear that the non-first difference data is non-stationary. Additionally, the group has a good intention to provide residual analysis plots in this part, but providing three plots to make one sentence for a point that the residual is a Gaussian white noise seems a waste of space. Instead, it might be better if the group states what is the implication if their model is causal but non-invertible. Can this model be useful or should we abandon it at all?

Further, the group can simplify the presentation of information better if they include their discussion on SARIMA in the part of their model analysis section and include the model diagnosis in a separate part.

In the discussion on the prediction, it might be better if the group makes it more explicit in what way they implement the forecasting. For example, they can state that the forecasting is computed using the forecast package in R, as they wrote in their source code, and discuss how this package operates briefly. Finally, interpretation-wise, even though the mean predicted values indicate an exponential trend, the confidence interval shows there is a declining trend as well, but the group remains silent about this part of their prediction. The group might be better to highlight why their model predicts declining trends in the confidence interval.