ISOM3530 HW3
Due: 5pm 1 May

- You can collaborate with your classmates for the assignments. If you work in groups, please list the names of your group members in the report. Submit by one member only.
- You need to submit both the report and the source code.

Case 1 (see Tractor-Sales.csv)

Thunder Horse is a farming tractor manufacturer. The company observes a consistent growth in its revenue from tractor sales recently. However, it is not easy to keep it's inventory and production cost down because of variability in sales and tractor demand. The management team wants to reduce the production/inventory cost. Forecasting sale/demand of tractors for next 3 years is a pro-active measure.

Preliminary study

1. Draw a timeseries plot of sales. By visual checking from the timeseries plot, comment on the constant variance across time.
2. If the constant variance is not met, try a log transformation on sales. Draw a timeseries plot of log(sales), any improvement?
3. Do you observe seasonal pattern of log(sales)? If yes, what is the length of a cycle?

Modeling

1. Apply ADF and KPSS test to check the need of differencing on log(sales). Keep applying the differencing until stationary.
2. Print the ACF plot of the stationary series from above, do you see any signal of seasonal feature?
3. Use AIC to determine the best SARIMA model. (Remark: directly use 'log(sales)' as the response in the modelling function).
4. Draw the forecasting for next 3 years.

Business Insights

1. From the forecasting, come with an inventory strategy (~50 words).

Case 2 (see bank_panel_data.csv)

A banker wants to know the influential factors on revenue of 20 banks from different sectors (investment bank and commercial bank) in recent years. He gets the following data from 2015 to 2023: Assets, Liabilities, Revenue, Loans, Interest_Rate, GDP_Growth and Ibank(is investment bank or not)

Preliminary study

1. Draw a side-by-side boxplot of Revenue (by Bank_ID). Repeat for Revenue by Year. Comment on the heterogeneity.
2. Write down the hierarchical level of this dataset.

Pre-processing

1. Standardize all the numerical variables. Use the standardized data all the way afterward.

Modeling

1. Fit a random slope model for Revenue with the following setting:
    a. Predictors: Assets, Loans, Interest_Rate, GDP_Growth
    b. Random slope on Loans and Interest_Rate by Bank_ID
    c. Random slope on Loans and Interest_Rate by Ibank
    d. Cross level interaction: Ibank*Loans
2. Check the significance of the interaction at 5% level. If it is not significant, refit the model without the interaction terms.

Business Insights

1. Use the final model from above, describe how the revenue is affected by other factors. (~50 words)