**MMA 823 2025 Winter**

**Assignment 2**

**Create 1 Jupyter Notebook that includes your code and output of all 3 parts in a clear, orderly, professional quality format. Illegible, unorganized output will not be graded. Make use of proper commenting and markdown cells.**

**Part 1: Historical Data and Stock Betas Analysis (12 points)**

In this part of the assignment, you'll gain practical skills in financial data analysis by working with real-world market data. Focusing on the concept of **beta**, a measure of stock risk relative to the market, you'll learn how to gather, manipulate, and analyze data, and interpret your findings.

The primary objective is to understand how beta changes over time and in different market conditions. You will collect and work with data from different types of stocks, allowing you to see how these stocks respond to market movements. You will also learn about moving window regression, a useful tool in time series analysis, and observe how the beta of various stocks has evolved over time.

* Data Collection (2 points)

Your first task is to collect data for your analysis.

Collect 10 years (start date: 2015-01-02) of weekly historical price/return data for the following stocks identified by these ticker symbols: "RACE", "TSLA", "JNJ", "XOM". Include data for the S&P 500 Index, “^GSPC", as well.

Also, collect the Risk-Free Interest Rate data. A good measure to consider for this data could be the 3-month US Treasury Rate. (Attention: Interest rates are quoted as annual rates. Convert/scale the number accordingly)

* Beta Estimation (4 points)

Next, using the data collected (10-year sample), estimate the beta (in the Capital Asset Pricing Model or CAPM) by running an Ordinary Least Squares (OLS) regression with the weekly returns.

*Rt -rf,t = a + b (Rm,t – rf,t) + et*

Show the estimated coefficients and regression statistics for each stock separately.

Compare the beta of each stock with the others. Which stock is the most sensitive to market changes and which one is the least? Explain your findings.

* Moving Window Regression (6 points)

As a next step, run regressions over moving date windows. Begin with a 3-year window (approximately 157 weeks) starting from 2015-01-02. Then, repeat this regression by moving the 3-year window forward by 1 week each time. For this process, consider using a 'for-loop' in your preferred programming language.

From the moving window regression, collect the estimated betas for each stock. Plot these beta values on a graph to visualize how they change over time for each stock.

**Part 2: Logit Regression for Credit Default (10 pts)**

Use the attached data file to create a predictive model for credit card defaults. Use Jupyter Notebook.

**2.1** Apply the Logit regression and follow the specific instructions.

* Train Sample: First 20000 rows, Test Sample: The rest of the sample (next 10000)
* Make sure you encode the dummy variables (categorical variables)
* Make sure you normalize/scale the numerical values
* Run Exploratory Data Analysis (EDA) to show the properties of the data sample
* Show correlations and other statistical properties along with outliers if any
* Clean/treat the outliers and missing variables if any

Present results on the accuracy of the model and show the Feature Importance with a chart.

(Use the online sources and examples as needed)

**2.2** Repeat (2.1) with a different Train-Test sample. Make the last 20000 rows the Train Sample and the first 10000 the Test set. Highlight the differences in results if any.

**Part 3. Yield curve data (8 points)**

* Collect Canadian zero-coupon yield curve data for the 2015-Jan-30 to 2024-Dec-31 period. Data link: <https://www.bankofcanada.ca/rates/interest-rates/bond-yield-curves/>)

(The example old data file is attached for your reference. Download the latest version that includes the recent data up to the end of 2024)

* Show a simple chart of the most recent yield curve (latest date point in the data file) and the yield curve by the end of 2023-Dec. How did the interest rates change since then? Compare the values with yields you see now.
* What factors were important in that change in your view? Briefly explain.
* Run a PCA analysis.
* How many principal components explain 99% of variance.
* Plot a time-series chart of the 2 main PCAs.