MFI Assignment 3: Understanding U.S. Stock Returns

Assignment Submission Details

1. Assignment deadline: 6:00pm on September 26, 2023
2. Submit through https://canvas.gatech.edu
3. Use Python for the programming and either Jupyter Notebooks or Quarto for the analysis and visualization
4. You have to submit ONLY
    - Jupyter notebook or Quarto file (well commented)
    - Output in PDF format
    - Submit the markdown file and the PDF file in one zip folder
    - Do not print any unnecessary output (pandas dataframes in the jupyter)
    - A short note (1-2 pages, bullet points) on what you learn from analyzing the stock returns over time and in the cross-section. More details in the assignment below. I will cold call students during the class to ask for their findings.
    - You don't need to submit any datasets

Assignment Details

The objective of this assignment is to get familiar with the stylized facts about the U.S. stock returns, both in the cross-section and overtime. The daily and monthly stock files from CRSP and the (original and derived) variables in these files, would be useful for your future assignments. So, please write well-documented and reusable code.

The Center for Research in Security Prices (CRSP) database contains end-of-day and month-end price on primary listings for the NYSE, NYSE MKT, NASDAQ, and ARCA exchanges along with basic market indices.

The SAS data provided in the server, **CRSP_DSF**, contains all end-of-day information on stocks in the daily security file (DSF) from the CRSP over the period from 2004 to 2022. The monthly data from 1926 - 2022 is in the file **CRSP_MSF**. Some other stock related information is available in **CRSP_MSENAMES**

Use the following for the definition of industry

SIC CODE	Industries
1 – 999	Agriculture, Forestry and Fishing
1000 – 1499	Mining
1500 – 1799	Construction
2000 – 3999	Manufacturing
4000 – 4999	Transportation and other Utilities
5000 – 5199	Wholesale Trade
5200 – 5999	Retail Trade
6000 – 6799	Finance, Insurance and Real Estate
7000 – 8999	Services
9000 – 9999	Public Administration

Descriptive Stats of U.S. Stock Returns

Section 1: MSF data analysis

Using the dataset, **CRSP_MSF**,

1.1 Data

- Restrict to SHRCD value of 10 or 11 (common stock sample)
- Compute the month end market capitalization ($MKTCAP$) as the product of the month end stock price and shares out standing. Make sure to take the absolute value of the price.
- Use the CPI from FRED to scale the market capitalization to December 2010 dollars. Give some thought to how to scale and the rationale for scaling. See https://fred.stlouisfed.org/series/CPIAUCSL You can use the unadjusted CPI as this series is available on FRED for the entire time period.

1.2 Plot by EXCHCD over time

- Plot a graph over the time period December 1925 to December 2019, at a monthly frequency, by exchange based on EXCHCD.

- Consider all exchanges apart from NYSE, NYSE MKT, NASDAQ and ARCA as "Other Exchange" category (Exchange Codes are respectively 1, 2, 3 and 4 for NYSE, NYSE MKT, NASDAQ, and Arca. An Exchange Code of zero indicates that a security is either trading on an unknown exchange, or is temporarily not trading at all. Adding 30 to the normal Exchange Codes (31, 32, 33 and 34) identifies when-issued trading, such as during a reorganization.) See http://www.crsp.org/products/documentation/data-definitions-e#exchange-code-header

1.3 Plot Number of Stocks and Market Capitalization over time

- Consider the following variables
  - the number of stocks in the sample
  - Market capitalization in December 2010 dollars
  - In a few bullet points, briefly describe the findings from the graphs
  - What patterns do you notice in the number of stocks and market capitalization over time? Can you relate the patterns to some economic, technological or regulatory events that occurred during this time period?

1.4 Plot by industry (HSICCD) over time

- Plot a graph over the time period December 1925 to December 2022, by industry
  - definition of industry is given in table above
  - the number of stocks in the sample
  - Market capitalization in December 2010 dollars
  - What patterns do you notice in the number of stocks and market capitalization across industries? Can you relate the patterns to some economic, technological or regulatory events that occurred during this time period?

1.5 Compute excess return and log of return

- Compute the excess return $(r_i, t$)for each stock for each month (note, we are ignoring the delisting returns. we will cover this in a future assignment) as return of the stock R_i, t (RET) - return on the risk-free security (from Ken French's data library, use R to download and process the file)

1.6 Compute CRSP_MSF descriptive stats and plot monthly

- Compute the descriptive stats - N, mean, standard deviation, skewness, kurtosis along with the minimum value, maximum value, 1%, 5%, 25%, 50%, 75%, 95%, 99% percentiles.
- Compute these descriptive statistics for the following time periods
  - for 1925–2022 time period
  - for 1963–2022 time period
  - Plot the mean and standard deviation at a monthly frequency.
  - In a few bullet points, briefly describe the findings from these tables. Are there major differences across the two time periods? Can you relate the patterns to any major economic, political or technological events?

1.7 Compute monthly VWRETD descriptive stats and plot monthly

- Use the value-weighted portfolio of all U.S. securities in the CRSP universe ($VWRETD$) from CRSP as the market portfolio (MKT)
  - Compute the excess returns of the MKT portfolio and log of the excess returns ($log(1 + \text{excess return})$)
  - Compute these descriptive statistics for the following time periods
    - for 1925–2022 time period (compute **one** descriptive table for entire period)
    - for 1963–2022 time period (compute **one** descriptive table for entire period)
    - Plot the mean and standard deviation at a monthly frequency. (needs daily $VWRETD$ data in **CRSP_DSF**)
    - In a few bullet points, briefly describe the findings from these tables. Are there major differences across the two time periods? Can you relate the patterns to any major economic, political or technological events?

1.8 Plot excess returns

- Plot the compounded excess returns and cumulative log excess returns (needs daily **CRSP_DSF** data and daily fama-french data)
  - Compute these descriptive statistics for the following time periods
    - for 1925–2022 time period
    - for 1963–2022 time period
    - Plot the mean and standard deviation at a monthly frequency.
    - In a few bullet points, briefly describe the findings from these tables. Are there major differences across the two time periods? Can you relate the patterns to any major economic, political or technological events?

Section 2: DSF data analysis

2.1 Compute descriptive stats and range

- Using the dataset, **CRSP_DSF**, please compute
  - descriptive statistics of daily returns, daily closing prices, daily bid-ask spreads, and trading volumes for the full sample period. (For daily closing prices, you may observe some negative prices, when there is no trading on a stock. In such cases, you should take absolute values.)
  - Compute the range of the daily stock prices (high price - low price) and also (closing price -opening price). Compute the descriptive stats of these measures along with the daily return and plot these for the entire time period. Describe your findings
- Returns of the stocks from the IPO to the next three years). Let us assume for the sake of simplicity that the first day the return data is available in CRSP is the IPO date when the firm starts trading. Please compute the following statistics using the dataset, **CRSP_DSF**

2.2 Plot number of IPO shares vs VWRETD per month

- During the 1963–2022 time period, for every month, check if there are any firms that have their first trading day i.e. likely to be the IPO. It is possible that some months may not have any IPOs
- Plot the "number of IPOs" per month over this time period along with the market returns (use VWRETD)- do you see any patterns?
- Are there **hot** and **cold** markets in IPOs that you can see? Discuss in a couple of bullet points

2.3 Plot average size of IPO vs VWRETD per month

- Plot the "average size of IPO firm" (using market capitalization as measure of size) per month over this time period along with the market returns (use VWRETD)- do you see any patterns?

2.4 Plot first day return of IPO vs VWRETD per month

- Plot the monthly mean first day returns of these IPOs along with the market returns for this time period? What do you observe?
- Do you see any differences in the first day returns - say over time, across industries or over size distribution?

2.5 IPO return analysis

- For every **IPO**, construct the first day, first month, 12 month, 24 month and 36 month stock returns and compute the descriptive stats (N, mean, standard deviation, skewness, kurtosis along with the minimum value, maximum value, 1%, 5%, 25%, 50%, 75%, 95%, 99% percentiles) over the entire 1963–2019 time period. In a few bullet points, briefly describe the findings from these descriptive stats Compute the excess returns over the market (say using VWRETD) for the post-IPO returns for various time horizons and compute the descriptive stats as in the previous item. In a few bullet points, briefly describe the findings from these descriptive stats
- Is adjusting using aggregate market returns the right bench mark to compute the excess returns of IPOs? If not, what would be a better benchmark?

Recommendations

- You can use garbage collector to manage your memory. Check gc package in python.
- Analysis Report:
  - Categorize your analysis in MSF and DSF
  - For MSF provide analysis based on:
    - Market cap vs Exchange (NYSE, NASDAQ etc.)
    - Market cap vs Industry (Construction, Services etc.)
    - Mean and SD for different time periods
  - For DSF provide analysis based on:
    - IPO data
    - Excess Return
  - You can add any other relevant points as well