Assignment 4: Estimating Beta from CAPM and Volatility

Overview

Weight: 5%

Learning objective: Develop an understanding of capital asset pricing model (CAPM) and also calculate volatility

The goals of this assignment are for the students to

- Become familiar with the monthly stock file of the CRSP and with variables that can be used in in future analysis.
- Practice regression analysis as an important application of a Machine Learning (ML) tool, specifically focusing on linear regression within the context of finance.
- Please note that the most important part of the assignment is not the programming or data analysis. The primary learning should be to identify patterns in your analysis and draw meaningful conclusions from those patterns.

Required Files for Submission

- One Jupyter Notebook that includes all the data wrangling and computations
- A report with all the plots and analyses
  - A short note (1-2 pages, bullet points) on what insights you gained from analyzing the CAPM beta's over time and in the cross-section. More details are provided in the assignment below. We will cold call students during the class to ask for their findings
- Note: Please do **NOT** submit data

Data

In order to reduce the time demands for the assignment, we have uploaded the required dataset to OneDrive: Link

- The name of data file is *MSF_1996_2023.csv*
- Description of all the variable is provided in *CRSP_MSF_Variable_Definition.pdf*

Assignment

- Use monthly CRSP stock data for this assignment
- Use the value-weighted portfolio of all U.S. securities in the CRSP universe (VWRETD) from CRSP as the market portfolio (MKT)
- See the definitions of industry code provided at the end of the assignment

Beta estimation

From the Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965), and others, we have if $R_{i,t}$ is the return on the security i, $R_{f,t}$ is the return on the riskless security, and $R_{m,t}$ is the return on the market portfolio, then:

$$\mathbb{E}[R_{i,t}] = R_{f,t} + \beta_i \cdot (\mathbb{E}[R_{m,t}] - R_{f,t})$$

where

$$\beta_i = \frac{\text{Cov}(R_{i,t}, R_{m,t})}{\text{Var}(R_{m,t})}$$

A stock's market beta can be estimated using the following CAPM or one-factor market model regression:

$$r_{i,t} = \alpha_i + \beta_i \cdot MKT_t + \epsilon_{i,t}$$

where $r_{i,t}$ is the excess return of stock i during period t, MKT_t is the excess return of the market factor or portfolio during period t, and $\epsilon_{i,t}$ is the regression residual.

This estimation involves performing a linear regression, where the dependent variable is the excess return of stock i, and the independent variable is the excess return of the market. Regression analysis is a fundamental Machine Learning tool, and in this assignment, you will apply it to estimate the beta coefficients in the CAPM model.

0. To reduce work (computational effort), for each year, instead of analyzing the entire set of firms, select a sample of 10 firms per industry.
1. Using regression analysis, estimate the $\beta_{i,t}$ of each firm for the period 1996–2023 for every year, use the monthly stock returns to estimate the beta using previous 12, 24 and 36 months stock returns
2. 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 by industry (use SIC code provide in supplementary details) for the beta values calculated above.
3. Compute the mean and standard deviation of betas for each industry (using the SIC codes provide in the Supplementary Details) for each year and plot the beta's over the 1996–2023 time period.
4. In a few bullet points, briefly describe the findings from the beta computation and from the graphs.

Given the market model estimation, we are able to decompose the total volatility of an individual stock into the systematic volatility and the idiosyncratic volatility as follows

$$\sqrt{\text{var}(R_{i,t})} = \beta_{i,t}\sqrt{\text{var}(R_{m,t})} + \sqrt{\text{var}(\epsilon_{i,t})}. \tag{2}$$

That is, the total volatility of stock i, $TVOL_{i,t} = \sqrt{\text{var}(R_{i,t})}$, is a sum of the systematic volatility, $SVOL_{i,t} = \beta_{i,t}\sqrt{\text{var}(R_{m,t})}$, and the idiosyncratic volatility $IVOL_{i,t} = \sqrt{\text{var}(\epsilon_{i,t})}$.

The variance of the residuals from your regression analysis provides the idiosyncratic volatility for each stock.

Volatility estimation

5. For each stock, compute the total volatility, the systematic volatility, and the idiosyncratic volatility using the equation above and the betas estimated in part 1.
6. Do you see any patterns in the total volatility, the systematic volatility, and the idiosyncratic volatility over the sample period? Examine and comment

Beta, Volatility and Stock Returns

7. Sort all stocks based on their beta into quintile portfolios.

- Form equal weighted portfolios and compute the average beta and the equal weighted $\frac{1}{N}$ portfolio excess return for each decile and the difference between portfolio 5 (high beta) and portfolio 1 (low beta)

- Repeat the previous steps for the value-weighted portfolio (weighted by the market capitalization) returns

8. Sort all stocks based on their idiosyncratic volatility into quintile portfolios.

- Form equal weighted portfolios and compute the average beta and the equal weighted $\frac{1}{N}$ portfolio excess return for each decile and the difference between portfolio 5 (high beta) and portfolio 1 (low beta)
- Repeat the previous steps for the value-weighted portfolio (weighted by the market capitalization) returns

Supplementary Details

Some of the following commands may be useful

- Regression analysis is a fundamental tool in Machine Learning. For linear regression, you can use statsmodels.OLS() or sklearn.linear_model.LinearRegression().
- Calculate deciles using pandas.qcut()
- For decile calculation in Python, you can use pandas's qcut function. Example: pd.qcut(data, q=5)
- Alternative methods for calculating deciles: numpy.percentile()
- For data summary: pandas.DataFrame.describe()
- Please refer to the lab from Sep 5 for more on regression.

Industry code:

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