MFE 431 Quantitative Asset Management Problem Set 1

Vinamra Rai

April 14, 2024

1 Replication of Market Porfolio

As a precursor, to solving the questions in the problem set, I installed all the necessary libraries and established a connect with the WRDS API. In this code, using the reference provided by Prof. Herskovic in the lecture, I first connected to the WRDS (Wharton Research Data Services) database to retrieve financial data from the CRSP monthly stock files. I ran an SQL query to select various stock-related data, including identifiers, dates, share codes, exchange codes, returns, shares outstanding, prices, and adjustment factors. This data was filtered for records between January 1, 1900, and December 31, 2023, and then sorted by stock identifier (permno) and date. Next, I ran another SQL query to fetch delisting returns data from CRSP, which includes stock identifiers, delisting returns, delisting dates, and delisting codes. This data was also sorted and processed to ensure consistency. To incorporate the Fama-French three-factor model data, I used the pandas' datareader library to load the necessary factors directly from Ken French's data library. These factors include market return over risk-free rate (MktRF), size premium (SMB), value premium (HML), and risk-free rate (RF). Afterward, I created a new column for the total market return (Mkt) by adding the risk-free rate. Throughout the process, I performed various data cleaning and preprocessing steps, such as converting data types, adjusting date formats, and resetting indexes. At the end, I merged the delisted returns with the monthly returns file downloaded from WRDS.

1.1 Construction of Market Portfolios

To construct the value-weighted market return and equal-weighted market return using CRSP data, replicating the market return time series available on Kenneth French's website, I followed these steps:

1. First, I restricted the sample to include only common shares traded on the NYSE, AMEX, and NASDAQ exchanges by filtering based on the exchange codes and share codes.

2. Next, I calculated the market capitalization column by taking the absolute value of the price multiplied by the shares outstanding. I also created a lagged market capitalization column by shifting the market capitalization column by one period for each stock identifier.

3. To handle missing data, I replaced missing returns with 0 and forward-filled missing prices and shares outstanding. I then dropped any remaining rows with missing values in the price, shares outstanding, or return columns.

4. I calculated the cumulative returns by combining the regular returns and delisting returns.

5. To calculate the value-weighted and equal-weighted returns, I grouped the data by date and applied lambda functions. For the value-weighted returns, I calculated the sum of cumulative returns multiplied by the lagged market capitalization divided by the sum of the lagged market capitalization for each date. For the equal-weighted returns, I simply took the mean of the cumulative returns for each date.

6. Finally, I combined the results into a single DataFrame, including the lagged total market capitalization, equal-weighted returns, and value-weighted returns. The lagged total market capitalization was calculated by summing the lagged market capitalization column and dividing it by 1,000,000 to convert it to millions.

The output of this section is a DataFrame containing the value-weighted market return, equal-weighted market return, and lagged total market capitalization from January 1926 to December 2023, at a monthly frequency, using the CRSP data.

1.2 Moments Comparison

To report the annualized return, annualized volatility, annualized Sharpe ratio, skewness, and excess kurtosis of the market excess returns for both the actual and estimated Fama-French time series, I followed these steps:

1. First, I merged the DataFrame containing the value-weighted market return with the Fama-French three-factor model data from Kenneth French's website, based on the date column.

2. Next, I calculated the excess returns for both the actual Fama-French market risk premium (which is the actual MktRF column) and the estimated Fama-French excess return derived from the value-weighted market return and risk-free rate.

3. To calculate the annualized moments, I created a separate function that takes the returns and the number of periods as input. Inside this function, I computed the annualized mean by multiplying the mean returns and the number of periods. The annualized standard deviation was calculated by multiplying the standard deviation of returns by the square root of the number of periods. The annualized skewness and excess kurtosis were computed using the inbuilt scipy functions.

4. I called this function separately for the actual Fama-French excess returns and the estimated Fama-French excess returns, passing in the respective return series and 12 as the number of periods as the data is monthly.

5. Next, I calculated the annualized Sharpe ratio for both series by dividing the annualized mean by the annualized standard deviation.

6. Finally, I created a DataFrame to store the results, with rows representing the annualized mean, annualized standard deviation, annualized Sharpe ratio, skewness, and excess kurtosis, and columns representing the actual Fama-French market excess returns and the estimated Fama-French market excess returns.

The output of this section is a DataFrame containing the annualized return, annualized volatility, annualized Sharpe ratio, skewness, and excess kurtosis for both the actual and estimated Fama-French market excess returns, covering the period from July 1926 to December 2023 at a monthly frequency. All the values matched the actual Fama-French data except for a minor difference in the kurtosis. Even after investigating the data, the source of this minor difference could not be traced.

Table 1: Comparative Annualized Statistics for Actual and Estimated FF Market Excess Return

	Fama French	**Replication**
Annualized Mean	8.14	8.14
Annualized Standard Deviation	18.51	18.51
Annualized Sharpe Ratio	0.44	0.44
Skewness	0.16	0.16
Excess Kurtosis	7.37	7.38

1.3 Correlation and Absolute Maximum Difference

To report the correlation between the estimated Fama-French market excess return time series and Kenneth French's actual time series, and the maximum absolute difference between the two, I followed these steps:

1. First, I merged the DataFrame containing the value-weighted market return with the Fama-French three-factor model data from Kenneth French's website, based on the date column.

2. Next, I calculated the correlation between the estimated Fama-French excess return (value-weighted market return minus risk-free rate) and the actual Fama-French market risk premium (MktRF) using the corr() function.

3. I also computed the maximum absolute difference between the two time series by subtracting the actual market risk premium from the estimated excess return, taking the absolute value, and finding the maximum value.

The output shows a correlation of 0.99999842 and a maximum absolute difference of 0.2061%.

Answering the questions:

1. **Is the difference zero?**
   No, the difference is not zero. The maximum absolute difference between the two time series is 0.2061%.

2. **If not, justify whether the difference is economically negligible or not.**
   The maximum absolute difference of 0.2061% is relatively small and likely economically negligible for most practical purposes. A difference of less than a quarter of a percentage point in monthly excess returns is generally considered acceptable, especially when compounded over long periods.

3. **What are the reasons for a nonzero difference?**
   There could be several reasons for a nonzero difference between the estimated and actual time series:

   - Treatment of corporate actions, adjustments, and other data-related issues, especially missing data.
   - Rounding errors or minor discrepancies in the way the portfolio is replicated.

   While a perfect replication is challenging due to these factors, the high correlation and relatively small maximum absolute difference indicate that the estimated time series closely tracks the actual Fama-French market excess return series over the given period.