

Problem Set 1

Niko Hiananto

Question 1. Replicating Value-Weighted Portfolio Returns

To replicate the value-weighted market returns using CRSP data, several data cleaning processes have been done. The sample period chosen for replication is the period of January 1926 - December 2019.

1. Universe of Stocks: The sample has been restricted to only include common shares (SHRCD = 10 and 11) and will also be restricted to securities traded on the New York Stock Exchange, American Stock Exchange, and NASDAQ (EXCHCD = 1, 2, and 3)
2. Missing Returns: The data contains missing returns specifically for RET with the characters (E, D, C, B, A), are treated as they are missing and will be considered to have 0 return. The characters represent no listing information, no valid previous prices, etc. The universe of stocks that have missing returns are quite small and will be assumed to have negligible impact when replicating the value-weighted returns.
3. Delisting return calculation: Similarly, the delisting returns could also be missing and DLRET with the characters (S, T, A, P) will also be considered as missing and will be assumed to have 0 return.
4. Total return calculation: To calculate the total monthly return for each stock I will consider both returns (with dividends) and total delisting returns (with dividends). More specifically, if both the returns and delisting-returns are not missing then I will add both returns in the calculation to calculate the total return, if the returns are not missing and the delisting returns are missing I will only use the returns, and finally if the returns are missing and the delisting returns are not, I will only use the delisting returns.
5. Market Capitalization Calculation: Several prices in the data are reported as negative which indicates that the prices are stale. I will still use the data on stale prices and take the absolute value of the price and calculate market capitalization as usual i.e. Market Cap = Price * Shares Outstanding ($|\text{PRC}| * \text{SHROUT}$). The data also contains missing prices, in which case, I will delete those stock data with missing prices. Finally, I will also create an additional variable called lagged market capitalization which measures the previous period's market capitalization for that PERMNO. I will also only consider the data with non-missing lagged market capitalization variable as the lagged variable is needed to replicate the value-weighted portfolio return.
6. Portfolio Weights: The equal-weighted portfolio return is calculated as the sum of all the firms' total return for that month divided by the number of existing firms (with non-missing prices and lagged market cap) for that month. Hence, the portfolio weights for the equal-weighted portfolio return changes depending on the amount of firms that exist for that month. The value-weighted portfolio return is calculated as the weighted mean of all the firms' returns using the lagged market capitalization variable (again only considering firms with non-missing prices and lagged market cap). The lagged variable is used instead of the current value of the market cap variable is used to avoid forward-looking bias.
7. Sample Period: The sample period considered for the replication strategy is for the period of January 1926 - December 2019.

Question 2. Moments Fama-French v. Replication

The resulting moments for the Fama-French portfolio and the replicated actual market excess return is summarised by the table below. I have reported the following five statistics: annualized mean, annualized standard deviation, annualized sharpe ratio, skewness and excess kurtosis. The first column shows the statistics for the Fama-French's value-weighted portfolio return of stocks and the second column shows the statistics for the replicated value-weighted market portfolio of stocks calculated in Question (1).

	Estimated FF Market Excess Return	Actual FF Market Excess Return
Annualized Mean	0.0800	0.0800
Annualized Standard Deviation	0.1844	0.1844
Annualized Sharpe Ratio	0.4339	0.4340
Skewness	0.1766	0.1767
Excess Kurtosis	7.8352	7.8360

From Question (1) we can obtain the replicated value-weighted market portfolio of stocks and using the Fama-French's portfolio return of stocks we will have 2 monthly time-series of value-weighted market returns and the summary statistics can then be calculated as follows:

1. Sample Period: Monthly from July 1926 to December 2019
2. Skewness: The excess skewness is calculated directly from the monthly time-series excess returns from both portfolios using the full sample
3. Excess Kurtosis: Similarly, the excess kurtosis is also calculated directly from both of the monthly time-series excess returns using the full sample
4. Annualized Mean: The annualized mean is calculated by taking the monthly average of the monthly time-series excess returns and is arithmetically annualized by multiplying it by 12.

$$\bar{\mu} = \frac{1}{T} \sum_{i=1}^T r_i$$
$$\mu_A = \bar{\mu} * 12$$

5. Annualized Standard Deviation: The annualized standard deviation is calculated by calculating the monthly standard deviation of the monthly time-series excess returns and is arithmetically annualized by multiplying it by $\sqrt{12}$.

$$\sigma_M = \sigma(\sum_{i=1}^T r_i)$$
$$\sigma_A = \sigma_M * \sqrt{12}$$

6. Sharpe Ratio: The sharpe ratio reported is annualized and is calculated as the annualized excess returns divided by the annualized standard deviation of both series

$$SR_A = \frac{\mu_A}{\sigma_A}$$

where μ_A and σ_A are defined as above.

Question 3. Correlation

The table below reports the correlation between the replicated and the Fama-French's value-weighted portfolio as well as the maximum absolute difference between the two monthly time-series for the period of July 1926 - December 2019.

	Return Series
Correlation	0.9999983
Maximum Absolute Difference	0.0017924

As seen from the table above, the correlation is very close to 1 but not exactly. The minor difference is most likely caused by the different assumptions that are used to construct the replicating portfolio returns, i.e. the way I have handled missing prices and/or missing lagged stock market values might be different than FF's assumptions on dealing with missing datas. Similarly, the assumptions I have used to calculate total return from returns and delisting returns might also be different than the one used to construct the FF's portfolio. The difference is very small, and can be considered as economically negligible as the correlation is really close to one, furthermore, all the moments computed in question (2) are also very similar to the actual FF's portfolio.