MFE 431 Quantitative Asset Management Problem Set 4

Vinamra Rai

May 26, 2024

1 Size and Value

I initially installed the required libraries and made a connection to the WRDS (Wharton Research Data Services) API to answer the questions in the problem set. I used the reference code provided by Prof. Herskovic to connect to the WRDS database and extract financial data from several CRSP (Centre for Research in Security Prices) data tables, Compustat and the link table connecting the two of them.

1. Obtaining Compustat Data

- The code connected to the Compustat database using the `conn.raw_sql()` function.
- It retrieved financial data from the `comp.funda` table, including variables such as total assets, liabilities, common equity, and more.
- The data was filtered based on specific criteria (e.g., `indfmt='INDL'`, `datafmt='STD'`, `popsrc='D'`, `consol='C'`).
- Additionally, pension data was retrieved from the `comp.aco_pnfnda` table.

2. Obtaining CRSP-Compustat Link Table

- The code retrieved the CRSP-Compustat link table from the `crspq.ccmxpf_linktable`.
- This table provided the mapping between CRSP `permno` and Compustat `gvkey`.
- The link table was filtered based on specific criteria (e.g., `linktype`, `linkprim`).

3. Obtaining CRSP Returns Data

- The code retrieved stock return data from the `crspq.msf` table.
- It also retrieved company information from the `crspq.msenames` table.
- The data was filtered based on specific criteria (e.g., `shrcd`, `exchcd`).
- The data was sorted by `permno` and `date`, and the `date` column was converted to the end of the month.

4. Obtaining CRSP Delisting Returns

- The code retrieved delisting returns from the `crspq.msedelist` table.
- It also retrieved company information from the `crspq.msenames` table for the delisting dates.
- The delisting returns data was merged with the CRSP returns data based on `permno` and `date`.
- Missing delisting returns were filled with zero.

5. Merging CRSP Data with Linktable

- The CRSP returns data was merged with the CRSP-Compustat link table based on `permno` and `permco`.
- Duplicates were checked and removed based on specific criteria.
- The merged data was filtered to keep only valid links within the specified date range.
- Additional filtering was applied to keep the most reliable links (e.g., `linktype='LC'`, `linkprim='P'`).

6. Obtaining Fama-French Factor Data

- The code retrieved Fama-French factor data using the `pandas_datareader.famafrench.FamaFrenchReader` function.
- It obtained the Fama-French 3 factors (market, size, and value) from the `F-F_Research_Data_Factors` dataset.
- Additionally, it retrieved book-to-market portfolios, size portfolios, and 25 book-to-market and size portfolios from their respective datasets.
- The factor data was merged with the previously obtained data based on the `date` column.

Conclusion

By following this workflow, the financial data from Compustat, CRSP, and Fama-French datasets was successfully obtained and merged. The resulting dataset contained the necessary information for further analysis and research purposes. The code provided a step-by-step guide to reproduce the data merging process, ensuring transparency and replicability.

1.1 Combining CRSP and Compustat to define size and book-to-market decile portfolios

1. The code computed returns for the CRSP data, including delisting returns. It filled missing returns with zero and calculated market equity (market capitalization) for each stock.
2. It then calculated lagged market equity by shifting the market equity values by one month within each security (permno). Missing lagged market equity values were dropped.
3. The Compustat data was merged with pension data on 'gvkey' and 'datadate'.
4. Several accounting variables were defined using the Compustat data:
   - Shareholders' equity (SHE) was defined using a hierarchy of available variables.
   - Deferred taxes and investment tax credit (DT) was defined using a hierarchy of variables.
   - Book value of preferred stock (PS) was defined using available variables or set to 0 if missing.
   - Book equity (BE) was calculated as SHE - PS + DT - prba, where prba (pension-related benefit assets) was filled with 0 if missing. Only positive book equity values were kept.
5. The CRSP and Compustat datasets were filtered to include data from 1972 onwards. The 'year' and 'month' columns were extracted from the 'date' column in CRSP data and 'datadate' column in Compustat data.
6. The CRSP and Compustat datasets were merged on 'gvkey' and 'year' using an outer join.
7. December market equity was obtained by filtering the combined data for December of each year. This was lagged by one year and merged back to the main dataset to get the lagged December market equity for the next year.
8. June market equity was obtained by filtering the combined data for June of each year and forward-filling the values.
9. Lagged book equity for December of the previous year (t-1) was calculated by shifting the book equity values by one row within each 'gvkey' group.
10. The book-to-market ratio was calculated as the lagged book equity divided by the lagged December market equity, multiplied by 1000.

11. Rows without the required data (lagged December market equity, June market equity, lagged book equity) were filtered out.

12. The 'year' and 'month' columns were reset based on the 'date' column.

13. Breakpoints for size deciles, book-to-market deciles, HML bins (high, medium, low), and SMB bins (small, big) were created based on NYSE stocks only. The breakpoints were then applied to the entire dataset to assign each stock to the respective bins.

The resulting combined dataset contains the necessary information for further analysis, including size and book-to-market decile assignments, HML bins, and SMB bins.

The following steps now outline the process of calculating the value weighted returns for all the deciles of Size and Value Portfolios:

1. The `calculate_value_weighted_returns` function is defined to calculate the value-weighted returns for each decile within a specified column (e.g., 'Size_Port' or 'BtM_Port'). It multiplies the returns by the lagged market equity to get the weighted returns and then divides the sum of weighted returns by the sum of lagged market equity for each decile and date.

2. The function is applied to calculate the monthly value-weighted returns for size deciles and book-to-market deciles separately, resulting in `size_vw_returns` and `bm_vw_returns` dataframes.

3. The risk-free rate ('RF') is merged into both `size_vw_returns` and `bm_vw_returns` dataframes based on the 'date' column.

4. Excess returns are calculated for both size and book-to-market deciles by subtracting the risk-free rate from the value-weighted returns.

5. The dataframes are filtered to include data from 1973 to 2023.

6. Long-short returns are calculated for size by subtracting the value-weighted returns of the largest size decile (decile 10) from the smallest size decile (decile 1). The resulting long-short returns are merged with the risk-free rate, and excess returns are calculated.

7. Similarly, long-short returns are calculated for book-to-market by subtracting the value-weighted returns of the lowest book-to-market decile (decile 1) from the highest book-to-market decile (decile 10). The resulting long-short returns are merged with the risk-free rate, and excess returns are calculated.

1.2 Annualized Stats and Correlation: Size Portfolios

Please note *that the process for computing the annualized statistics is outlined in this section once as the same process is followed for Size and Values portfolios and hence would only be mentioned in this subsection.*

1. The `calculate_annualized_statistics` function is defined to calculate the annualized statistics for a given set of monthly returns. It takes the monthly returns dataframe, an optional decile column, and the return column as input.

2. Within the function, if a decile column is provided, it iterates over each unique decile and calculates the statistics for each decile separately. Otherwise, it calculates the statistics for the entire dataframe (used for long-short portfolios).

3. For each decile or the entire dataframe, the following monthly statistics are calculated:
   - Mean return: Average of the monthly returns.
   - Volatility: Standard deviation of the monthly returns.
   - Sharpe Ratio: Mean return divided by volatility.
   - Skewness: Skewness of the monthly returns.

4. The monthly statistics are then annualized:
   - Annualized mean return: Monthly mean return multiplied by 12.
   - Annualized volatility: Monthly volatility multiplied by the square root of 12.
   - Annualized Sharpe Ratio: Annualized mean return divided by annualized volatility.

5. The calculated statistics for each decile or long-short portfolio are appended to a list of dictionaries.

6. The function returns a DataFrame containing the annualized statistics for each decile or long-short portfolio.

7. The `calculate_annualized_statistics` function is called separately for size deciles (`size_vw_returns`), book-to-market deciles (`bm_vw_returns`), long-short size portfolio (`long_short_size_returns`), and long-short book-to-market portfolio (`long_short_bm_returns`).

8. The resulting DataFrames are combined into a single DataFrame (`annualized_stats`) using `pd.concat()`. A 'portfolio' column is added to identify whether the statistics belong to the size or book-to-market portfolios.

9. The `french` DataFrame is filtered to include only dates after 1973 and stored in `french_filtered`.

10. The relevant columns for Book-to-Market and Size portfolios are selected from the `french_filtered` DataFrame and stored in `bm_columns` and `size_columns`, respectively.

11. The decile returns for size and book-to-market portfolios are extracted from the dataset (`size_vw_returns` and `bm_vw_returns`) using the `pivot` function. The resulting DataFrames (`size_decile_returns` and `bm_decile_returns`) have dates as the index and deciles as columns.

12. The Fama-French decile returns are merged with the decile returns based on the 'date' column using an inner join. This is done separately for size and book-to-market portfolios, resulting in `merged_size_returns` and `merged_bm_returns` DataFrames.

13. The correlations between the decile returns and the corresponding Fama-French decile returns are calculated using a dictionary comprehension. The correlations are stored in `size_correlations` and `bm_correlations` dictionaries.

14. The correlation dictionaries are converted to DataFrames (`size_correlations_df` and `bm_correlations_df`) for easier viewing.

15. To calculate the correlations for long-short returns, the Fama-French data is merged with the long-short returns data (`long_short_size_returns` and `long_short_bm_returns`) based on the 'date' column using an inner join. This results in `merged_size_long_short` and `merged_bm_long_short` DataFrames.

16. The long-short returns for the Fama-French portfolios are calculated by subtracting the returns of the smallest size decile (ME01) from the largest size decile (ME10) for size, and subtracting the returns of the lowest book-to-market decile (BM01) from the highest book-to-market decile (BM10) for book-to-market.

17. The correlations between the long-short returns and the Fama-French long-short returns are calculated using the `corr` function and stored in `size_long_short_correlation` and `bm_long_short_correlation`.

18. Finally, the long-short correlations for size and book-to-market are printed.

The final output (`annualized_stats`) contains the annualized mean return, annualized volatility, annualized Sharpe Ratio, and skewness for each size decile, book-to-market decile, and the long-short portfolios. These statistics can be used to compare the performance and characteristics of the portfolios constructed based on size and book-to-market factors. The subsequent steps calculate the correlations between the constructed decile portfolios and the corresponding Fama-French decile portfolios, as well as the correlations between the long-short portfolios and the Fama-French long-short portfolios. The correlations provide a measure of how closely the portfolios match the Fama-French portfolios in terms of returns.

Table 1: Replication, ME-sorted portfolios

	Decile										Long-Short
	1	2	3	4	5	6	7	8	9	10	1-10
Panel A: Ken French website											
ret	8.29	8.89	9.36	8.71	9.24	8.54	8.89	8.54	8.02	6.67	1.62
vol	21.78	22.61	21.39	20.73	20.23	18.84	18.77	18.02	16.68	15.40	16.25
SR	0.38	0.39	0.44	0.42	0.46	0.45	0.47	0.47	0.48	0.43	0.10
skewness	-0.16	-0.24	-0.47	-0.51	-0.44	-0.52	-0.48	-0.49	-0.45	-0.34	0.80
Panel B: Replication											
ret	6.28	6.95	7.36	7.49	8.18	7.87	7.60	7.92	7.34	6.44	-0.16
vol	22.55	22.78	21.46	20.90	20.21	19.12	18.77	18.15	16.74	15.33	16.52
SR	0.28	0.31	0.34	0.36	0.40	0.41	0.41	0.44	0.44	0.42	-0.27
skewness	-0.16	-0.30	-0.49	-0.45	-0.53	-0.47	-0.52	-0.44	-0.43	-0.33	0.82
corr w/ original	0.9796	0.9819	0.9926	0.9927	0.9929	0.9942	0.9942	0.9952	0.9964	0.9940	0.9512

1.3 Annualized Stats and Correlation: Book-to-Market Portfolios

Table 2: Replication, BM-sorted portfolios

	Decile										Long-Short
	1	2	3	4	5	6	7	8	9	10	10-1
Panel A: Ken French website											
ret	6.27	8.01	7.93	8.05	7.64	8.76	7.12	8.73	10.75	11.08	4.81
vol	18.28	16.68	16.38	16.56	15.77	16.34	16.62	17.41	18.39	22.21	17.54
SR	0.34	0.48	0.48	0.49	0.48	0.54	0.43	0.50	0.58	0.50	0.27
skewness	-0.21	-0.44	-0.53	-0.50	-0.50	-0.43	-0.44	-0.66	-0.40	-0.44	0.05
Panel B: Replication											
ret	5.35	5.90	6.89	7.22	7.83	8.70	8.63	10.26	12.31	17.76	-12.41
vol	18.29	17.04	16.39	16.57	16.92	16.92	16.38	17.39	18.60	21.57	18.29
SR	0.29	0.35	0.42	0.44	0.46	0.51	0.53	0.59	0.66	0.82	-0.68
skewness	-0.25	-0.38	-0.44	-0.44	-0.49	-0.34	-0.19	-0.11	-0.09	0.70	0.99
corr w/ original	0.9861	0.9678	0.9555	0.9508	0.9372	0.9338	0.9292	0.9232	0.9091	0.8882	0.7845

1.4 Performance of Value and Size Anomalies

To examine if the value and size anomaly has worked in the past few years, the following steps were performed using the data that has been constructed as part of this problem set. The inferences from the actual data (Fama-French) might differ:

1. **Plotting the excess returns series:**
   - The excess returns for the long-short size portfolio, long-short book-to-market portfolio, and the market (Mkt-RF) were plotted over time.
   - This provides a visual comparison of how the returns of these portfolios have evolved over the years.
2. **Calculating and plotting cumulative returns:**
   - Cumulative returns were calculated for the long-short size portfolio, long-short book-to-market portfolio, and the market (Mkt-RF).
   - The cumulative returns were plotted to observe the growth of each portfolio over time.
   - A separate plot was created with dual y-axes to compare the cumulative returns of the long-short size portfolio and the market (Mkt-RF) more clearly.
3. **Creating a table with annualized statistics:**
   - A function `calculate_annualized_stats` was defined to calculate the annualized mean return, volatility, and Sharpe ratio for a given series of returns.
   - The function was applied to the excess returns of the long-short size portfolio, long-short book-to-market portfolio, and the market (Mkt-RF).
   - A table (`stats_table`) was created to present the annualized mean return, annualized volatility, and annualized Sharpe ratio for each portfolio.

By analyzing the plots and the table, we can draw some empirical evidence regarding the performance of the value and size anomaly in recent years:

- The excess returns plot shows how the returns of the long-short size and book-to-market portfolios have compared to the market (Mkt-RF) over time. If the long-short portfolios consistently outperform the market, it suggests that the anomalies have persisted.

- The cumulative returns plot illustrates the growth of each portfolio over the years. If the cumulative returns of the long-short portfolios have increased more than the market, it indicates that the anomalies have been profitable.

- The table with annualized statistics provides a summary of the risk-adjusted performance of each portfolio. Higher annualized mean returns, lower volatility, and higher Sharpe ratios for the long-short portfolios compared to the market would suggest that the anomalies have been effective in generating excess returns.

By examining these empirical results, we can see that Value portfolios have mostly outperformed market returns. But on the other hand, the performance of size portfolios have been really dismal.

Table 3: Annualized Statistics

	Annualized Mean Return	Annualized Volatility	Annualized Sharpe Ratio
Size	-0.044640	0.165210	-0.270201
Value	0.081017	0.175804	0.460835
Market	0.082042	0.184971	0.443539

Figure 1: Monthly returns of the three portfolios

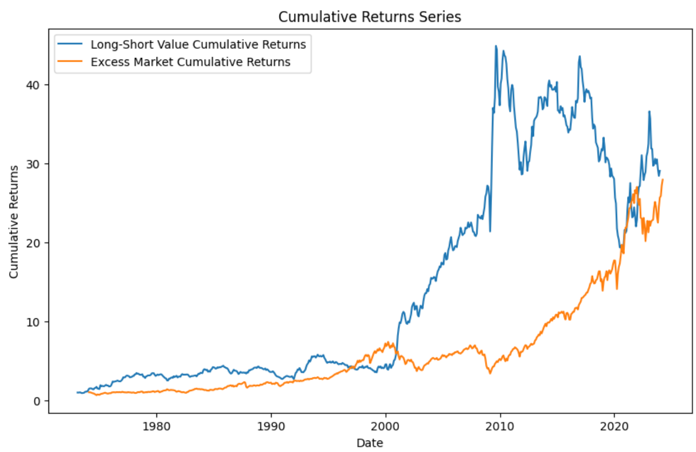

Figure 2: Value versus Market

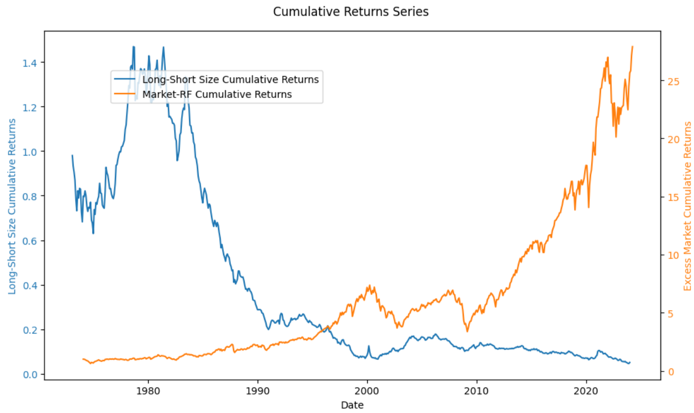

Figure 3: Size versus Market

1.5 Annualized Stats and Correlation: HML and SMB Portfolios

To report the annualized average excess returns, annualized volatility, Sharpe Ratio, and skewness for both the HML and SMB portfolios, as well as the correlations between the replicated factors and the factors from French's website, the following steps were taken:

1. **Calculating HML and SMB factors:**
   - The combined dataset was grouped by 'year', 'month', 'SMB_Port', and 'HML_Port', and the value-weighted returns were calculated for each group.
   - The resulting DataFrame was pivoted to compute the HML and SMB factors.
   - The HML factor was calculated as the average of the returns of the high book-to-market portfolios minus the average of the returns of the low book-to-market portfolios.
   - The SMB factor was calculated as the average of the returns of the small size portfolios minus the average of the returns of the big size portfolios.

2. **Merging risk-free rate and calculating excess returns:**
   - The risk-free rate ('RF') was merged from the 'french' dataset based on the 'date' column.
   - Excess returns for HML and SMB were calculated by subtracting the risk-free rate from the respective factor returns.

3. **Calculating annualized statistics:**
   - A function `calculate_hml_smb_annualized_statistics` was defined to calculate the annualized mean return, volatility, Sharpe ratio, and skewness for a given series of returns.
   - The function was applied to the excess returns of the HML and SMB factors.
   - The annualized statistics were stored in separate DataFrames for HML and SMB.

4. **Calculating correlations:**
   - The 'french' DataFrame was filtered for dates between 1973 and 2023.
   - The French HML and SMB returns were merged with the calculated HML and SMB returns based on the 'date' column.
   - Correlations between the French factors and the replicated factors were calculated using the `corr` function.

5. **Analyzing factor consistency across time:**

- Cumulative returns for the HML and SMB factors from the replicated data were calculated.
- A plot was created to visualize the cumulative returns of the HML and SMB strategies over time.
- The plot allows for a visual assessment of the consistency and performance of the factors across the time period.

The results of these steps provide the annualized statistics for the HML and SMB portfolios, both from the original French data and the replicated data. The correlations between the replicated factors and the original factors are also reported. The plot of cumulative returns offers empirical evidence regarding the consistency of the factors across time.

By examining these statistics and the plot, we can see that the HML returns are more stable and consistent over time.

Table 4: Annualized Statistics for HML and SMB

	HML	SMB
Panel A: Data from Ken French website		
ret	3.58	1.73
vol	10.82	10.59
SR	0.33	0.16
skewness	0.07	0.46
Panel B: Replication		
ret	6.98	-0.04
vol	11.17	10.20
SR	0.62	-0.38
skewness	0.61	0.27
corr w/ original	0.8434	0.9664

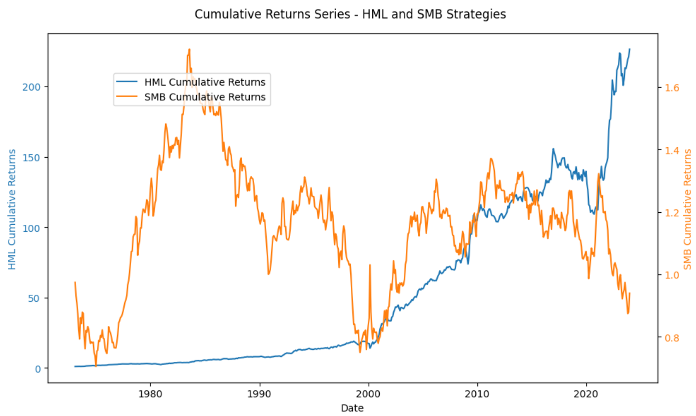

Figure 4: Size versus Market

1.6 Characteristic versus Factor Portfolios: A Qualitative Comparison

Construction

Characteristic Portfolios (Fama and French 1992)

- Formed based on firm-specific characteristics, such as size (market capitalization) and book-to-market ratio.
- Stocks are sorted into deciles or quantiles based on these characteristics.

Factor Portfolios (Fama and French 1993)

- Designed to capture systematic risk factors that influence stock returns.
- SMB (Small Minus Big) is constructed by taking the difference in returns between small-cap and large-cap stocks.
- HML (High Minus Low) is constructed by taking the difference in returns between high book-to-market and low book-to-market stocks.

Purpose

Characteristic Portfolios (Fama and French 1992)

- Examine the cross-sectional variation in average stock returns.
- Identify patterns in average returns based on firm-specific characteristics.

Factor Portfolios (Fama and French 1993)

- Explain the risk and return characteristics of individual stocks and portfolios.
- Capture systematic risk factors that can be used in asset pricing models to explain returns.

Implications

Characteristic Portfolios (Fama and French 1992)

- Help identify anomalies or patterns in stock returns that cannot be explained by the traditional Capital Asset Pricing Model (CAPM).
- Laid the groundwork for the development of multi-factor models.

Factor Portfolios (Fama and French 1993)

- Used in asset pricing models to provide a more comprehensive explanation of stock returns.
- Form the basis of the Fama-French three-factor model, which extends the CAPM by adding size and value factors to the market factor.
- Used in constructing passive investment strategies and evaluating the performance of active managers.

Scope

Characteristic Portfolios (Fama and French 1992)

- Focus on the cross-sectional variation in returns due to firm-specific attributes.

Factor Portfolios (Fama and French 1993)

- Focus on capturing broader systematic risk factors that affect returns across the market.

Application

Characteristic Portfolios (Fama and French 1992)

- Primarily used in empirical research to identify anomalies and patterns in stock returns.
- Useful for understanding the drivers of stock returns and testing asset pricing theories.

Factor Portfolios (Fama and French 1993)

- Used in risk management, portfolio construction, and performance evaluation.
- Applied in asset pricing models and investment strategies to account for systematic risk factors.

Conclusion

Both characteristic portfolios and factor portfolios are essential for understanding and modeling asset returns. Characteristic portfolios help identify the attributes that drive differences in returns, while factor portfolios translate these attributes into systematic risk factors. The insights gained from characteristic portfolios were crucial in developing factor portfolios and subsequent multi-factor models that provide a more complete explanation of asset returns. Together, these two types of portfolios have significantly contributed to the advancement of asset pricing theory and practice.