Problem Set 4
MFE 431: Quantitative Asset Management
Professor: Bernard Herskovic
Student: Xiahao Wang
Name of whom I discussed this problem set with: Mu Lin

Question 1:

Prepare data for analysis. Combine necessary CRSP and Compustat datasets needed to define size and book-to-market decile portfolios as defined in Fama and French (1992b)1, as well as the HML and SMB factors as defined in Fama and French (1993)2. Detail which datasets you use,how you merged them, how you calculated the portfolios, and any differences between the building of the decile portfolios and the factors. Output should be between January 1973 and December 2018.

Summary:

The resulting dataset:

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday,
##     week, yday, year

## The following object is masked from 'package:base':
##
##     date

##        Year Month port    Size_Ret     BtM_Ret     HML_Ret    SMB_Ret
##    1: 1973     1    1 -0.03132254 -0.01237214 0.027691849 -0.0373765
##    2: 1973     1    2 -0.04600440 -0.05189962 0.027691849 -0.0373765
##    3: 1973     1    3 -0.04881413 -0.05023729 0.027691849 -0.0373765
##    4: 1973     1    4 -0.07033170 -0.04292899 0.027691849 -0.0373765
##    5: 1973     1    5 -0.06512458 -0.05652230 0.027691849 -0.0373765
##   ---
## 5516: 2018    12    6 -0.11541152 -0.05911644 0.003611573  0.0107717
## 5517: 2018    12    7 -0.11694734 -0.08864151 0.003611573  0.0107717
## 5518: 2018    12    8 -0.10394757 -0.16457328 0.003611573  0.0107717
## 5519: 2018    12    9 -0.09452291 -0.11120628 0.003611573  0.0107717
## 5520: 2018    12   10 -0.08845702 -0.12557800 0.003611573  0.0107717
```

1. **Data Selection:**

Following the procedure from Fama French's paper, the datasets below are downloaded from CRSP and Compustat in WRDS and Fama French's data library.

- Monthly CRSP US equity Data (CRSP)
- Fundamental Annual Updates for North America (Compustat)
- Pension Annual (Compustat)
- Linking Table (CRSP Compustat merged)
- Historical Book Equity Data (Fama French)

As the problem set requests return data starting from 1973 Jan and 2018 Dec. Data from US equity and Fundamental updates are downloaded starting from 1970 til 2018 to ensure the return and decile sorting result starts from Jan 1973.

2. **Clean up US equity data in CRSP**

Following the Fama French paper, I restrict the sample to common shares(share code 10 and 11) and to securities traded in the New York Stock Exchange, American Stock Exchange, or the Nasdaq Stock Exchange (exchange codes 1, 2, and 3)

Scope: Dataset from CRSP Monthly from 1970 Jan to 2018 Dec

Then set standard format of "date" and extract "Year" and "Month" column.

2.a Handle missing data in Delisting Return (DLRET), and Return (RET):

The missing data in RET are denoted by 'A','C','P','S','T','B','C', The missing data in DLRET are denoted by -66, -77, -88, -99 Replace these missing data with NA.

2.b Delisting return calculation:

To calculate the cum-dividend return (Ret), let RET be cum-dividend return if DLRET is missing. If DLRET is not missing and RET is missing, let DLRET be just cum-dividend return. If both are not missing, use the formula from lecture notes: Ret = (1 + DLRET)(1 + RET) - 1 to get the cum-dividend return.

2.c Market Capitalisation calculation:

There are some price that are negative, I take the absolute value for both price and shares outstanding and multiply them together to get the market capitalisation. As the CRSP data needs to be merged with Compustat which uses millions as its base unit, divide the market cap by 1000 so as to have the same base unit in the merged data.

2.d Aggregation of Market cap by company

From WRDS: PERMCO is a unique permanent identifier assigned by CRSP to all companies with issues on a CRSP file. This number is permanent for all securities issued by this company regardless of name changes. The PERMNO identifies a firm's security through all its history, and companies may have several stocks at one time.

In short: A PERMCO can have multiple PERMNOs as a company can have multiple stocks.

Hence market cap of stocks from the same companies are aggregated together so as to reflect the right market value of the firm.

3. **Clean up Compustat datasets**

3.a Remove financial firms

As the paper has suggested, financial firms have to be excluded from the analysis. The variable *indfmt* identifies financial firm as *FS* and non-financial firm as *INDL*. Filter out the financial firms from the merged dataset.

3.b Merge compustat and pension

Annual fundamental accounting variables (compustat) and Pension information (PRBA) are found separately on compustat database, hence the two datasets needs to be merged.

- Extract the gvkey and the year from the pension dataset.
- Use the two variables extracted and merge it with the compustat data

3.c Calculation of book equity

Following the instructions on the problem set, there are the steps I used to calculate book equity.

- Calculate shareholder's equity (SHE): it is equal to stockholders equity - total (SEQ)
  - if not available, use common/ordinary equity - total (CEQ) + Preferred/Preference Stock (Capital) - Total (PSTK)
  - if not available, use Assets - Total (AT) - Liabilities - Total (LT) - Minority Interest (Balance Sheet) (MIB)
  - if not available, use AT - LT
- Calculate Deferred taxes (DT): it is equal to Deferred Taxes and Investment Tax Credit (TXDITC)
  - if not available, use Investment Tax Credit (Balance Sheet) (ITCB) + Deferred Taxes (Balance Sheet) (TXDB)
  - if not available, sum up what is not missing
- Calculate preferred stock (PS): it is equal to Preferred Stock Redemption Value (PSTKRV)
  - if not available, use Preferred Stock Liquidating Value (PSTKL)
  - if not available, use Preferred/Preference Stock (Capital) - Total (PSTK)
- Calculate book equity (BE): BE = SHE - PS + DT - PRBA, PRBA comes from the pension dataset which we have merged earlier on.
  - If SHE is not available, assign BE to NA, otherwise, include other variables in the calculation if not missing.
- Historical book equity data from Fama French website will be merged later on as it has only PERMNO as unique ID

4. Merge CRSP with Linking table and Compustat

Following the procedure that was demonstrated during the TA session for this part, first merge the clean up CRSP monthly stock data with the linking table from CRSP compustat merged database by PERMCO.

4.a Filter the data based on Linking table start and end date

Variables LINKDT and LINKENDDT indicate the start and end dates during which the PERMCO is available. Hence select only the PERMCO whose date is within the start and end date range. For PERMCO that still exists, the original data show 'E' instead of a date. Replace the character with NA for processing later on.

4.b Further processing on merged data

Then, follow the steps below to clean up the merged data:

- Step 1: Filter on variable LINKTYPE, if there are a few types on PERMCO and the type is not 'LC', only keep 'LC'.
- Step 2: Filter on variable LINKPRIM, same with step 1. Only keep 'P' if there are different values for LINKPRIM on the same PERMCO
- Step 3: Filter on variable LIID , same with step 2. Only keep '1' if there are different values for LIID on the same PERMCO
- Step 4: Filter on the variable LINKENDDT, the the link that's current
- Step 5: Find the difference between the start and end date of the link, use the link that's been around the longest
- Step 6: With the same logic on step 5, use the gvkey that has been around the longest
- Step 7: Use the smaller gvkey if there are multiple with same PERMCO on the same date

4.c Merge CRSP with Compustat

Upon completing part 4 a & b, merge the CRSP data with Compustat by variable gvkey and year. Do not remove the NAs in the dataset.

5 Calculation of size (ME) portfolio return

Followint the problem set, below are the steps to calculate portfolio return:

- Step 1: As there might be multiple securities for the same firm (PERMCO), first aggregate the market cap by PERMCO for the same year and month. Variables Year and Month are extracet from the date of the merged dataset.
- Step 2: size portfolio is calculated by matching Market cap at June of year t with return from July of year t to June of year t + 1. The rebalancing is done at the end of June annually. Note that only stocks with data on June will be used for size decile sorting.
- Step 3: The stocks are sorted into deciles from 1 to 10. The breakpoint is set based on Market cap of NYSE stock in June of that particular year.
- Step 4: By now, the size decile has now been obtained for each stock across years. Compute the value-weighted return using 1 month lagged market cap for all the companies within each decile to get monthly value-weighted return of the portfolio.
- Step 5: Restrict the data from 1973 Jan to 2018 Dec.
- Step 6: Create a long short portfolio by going long on the 1st decile and shorting the 10th decile

6 Calculation of Book to Market (BE/ME) portfolio return

Book value refers to the BE that we have calcuated earlier from compustat data. And Market value is the market cap for each firm. The firms are sorted into 10 deciles based on the ratio book value / market value.

Below are the steps to calculate BE/ME portfolio return:

- Step 1: As there might be multiple securities for the same firm (PERMCO), first aggregate the market cap by PERMCO for the same year and month. Variables Year and Month are extracet from the date of the merged dataset.
- Step 2: BE/ME portfolio is calculated by matching Market cap and book equity at December of year t-1 with return from July of year t to June of year t + 1. The rebalancing is done at end of June annually. Note that only stocks with data on December will be used for BE/ME decile sorting.
- Step 3: Load the historical book equity information from Fama French website, restrict data from 1970 onwards. Merge the data with the merged CRSP and compustat dataset. Assign the historical book equity as book equity if the current book equity is NA.
- Step 4: The stocks are sorted into deciles from 1 to 10. The breakpoint is set based on BE/ME of NYSE stock in December of that particular year. Note that book equity value less than 0, equal to 0 is excluded when finding the breakpoint. Book equity is set to NA as well if the total asset is NA for that period. Also the market cap needs to be available in order to have a BE/ME ratio.
- Step 5: By now, the BE/ME decile has now been obtained for each stock across years. Compute the value-weighted return using 1 month lagged market cap for all the companies within each decile to get monthly value-weighted return of the portfolio.
- Step 6: Restrict the data from 1973 Jan to 2018 Dec.
- Step 7: Create a long short portfolio by going long on the 1st decile and shorting the 10th decile

7 Calculation of SMB and HML portfolio return

Following the Fama French website, portfolio is double sorted into 6 portfolios based on ME and BE/ME.

- ME is sorted based on median of the market cap into 2 deciles
- BE/ME is sorted based on 30th and 70th percentile
- Formula : SMB = 1/3(Small Value + Small Neutral + Small Growth - Big Value - Big Neutral - Big Growth)
- Formula : HML = 1/2(Small Value + Big Value - Small Growth - Big Growth)

As the same rebalancing and lagging periods are applied for the characteritic portfolio as well as the size and BE/ME portfolio, the result from the previous portfolio ranking are be used to do the sorting for SMB and HML portfolios.

- Step 1: Sort ME portfolio. The current ME (size) ranking is from 1 to 10 for all stocks, sort into 2 deciles small (1 - 5), big (5 - 10)
- Step 2: Sort BE/ME portfolio. The current BE/ME ranking is from 1 to 10 for all stocks. sort into 3 deciles Low (1-3), Medium (4-7), High (8-10)
- Step 3: Merge the two portfolios so as to get 6 differnet combinations of portfolios, then calculate the value-weighted portfolio based on lagged market cap.
- Step 4: Restrict the data from 1973 Jan to 2018 Dec.

Question 2:

The summary for size decile and long short portfolio, shown in table 1.

1. Annualized Excess Return: Annualized excess return for decile 1 to 10 (Return above risk free) = Average of excess return multiply by 12. Annualize return for long short portfolio = average of return of (decile 1 - decile 10)

Table 1: Size decile and long short portfolios

	Size decile Portfolios										
Return_Statistics	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Long Short
Excess Mean Return	0.0828870	0.0840524	0.0922530	0.0833558	0.0887202	0.0781188	0.0829347	0.0783827	0.0727544	0.0579238	0.0249632
Volatility	0.2168964	0.2236369	0.2098146	0.2037008	0.1971657	0.1827887	0.1808507	0.1758024	0.1621834	0.1502269	0.1632802
SR	0.3821502	0.3758432	0.4396880	0.4092071	0.4499781	0.4273721	0.4585813	0.4458566	0.4485933	0.3855756	0.1528854
Skewness	-0.1927309	-0.2332042	-0.4678370	-0.5134567	-0.5323154	-0.5575531	-0.4828872	-0.5048574	-0.4578669	-0.3553916	0.8318901
Correlation	0.9979897	0.9977598	0.9981128	0.9980211	0.9978539	0.9976289	0.9977088	0.9985421	0.9988033	0.9993363	0.9956623

Table 2: BE/ME decile and long short portfolios

	BE/ME decile Portfolios										
Return_Statistics	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Long Short
Excess Mean Return	0.0537083	0.0677846	0.0737692	0.0801111	0.0636013	0.0732837	0.0778255	0.0683614	0.0892918	0.1039617	-0.0502534
Volatility	0.1801735	0.1640856	0.1605978	0.1683311	0.1586120	0.1571105	0.1527625	0.1659995	0.1683039	0.1980033	0.1601787
SR	0.2980923	0.4131050	0.4593408	0.4759138	0.4009866	0.4664468	0.5094538	0.4118170	0.5305390	0.5250506	-0.3137335
Skewness	-0.2012860	-0.5250051	-0.4607912	-0.4702241	-0.4781551	-0.3763718	-0.1320413	-0.5632884	-0.4194621	-0.2293717	-0.3415019
Correlation	0.9968267	0.9842816	0.9730382	0.9709959	0.9724531	0.9616792	0.9632471	0.9425558	0.9674015	0.9551014	0.9200403

2. Annualized Volatility: Compute portfolio standard deviation as the sample standard deviation across years for each decile and long short and annualize it by multiplying by square root of 12.
3. Sharp Ratio: Compute the sharp ratio using the following formula: Sharp ratio = Annualized Excess Return / Annualized Volatility
4. Skewness: Compute the skewness of the return using the following formula: skewness(monthly return), where skewness is the function from library moments to calculate the skewness.
5. Correlation: Find the correlation between the replicated decile and long short portfolio and the one on Fama French website named *Portfolios_Formed_on_ME.csv*

Question 3:

The summary for book-to-market decile and long short portfolio, shown in table 2

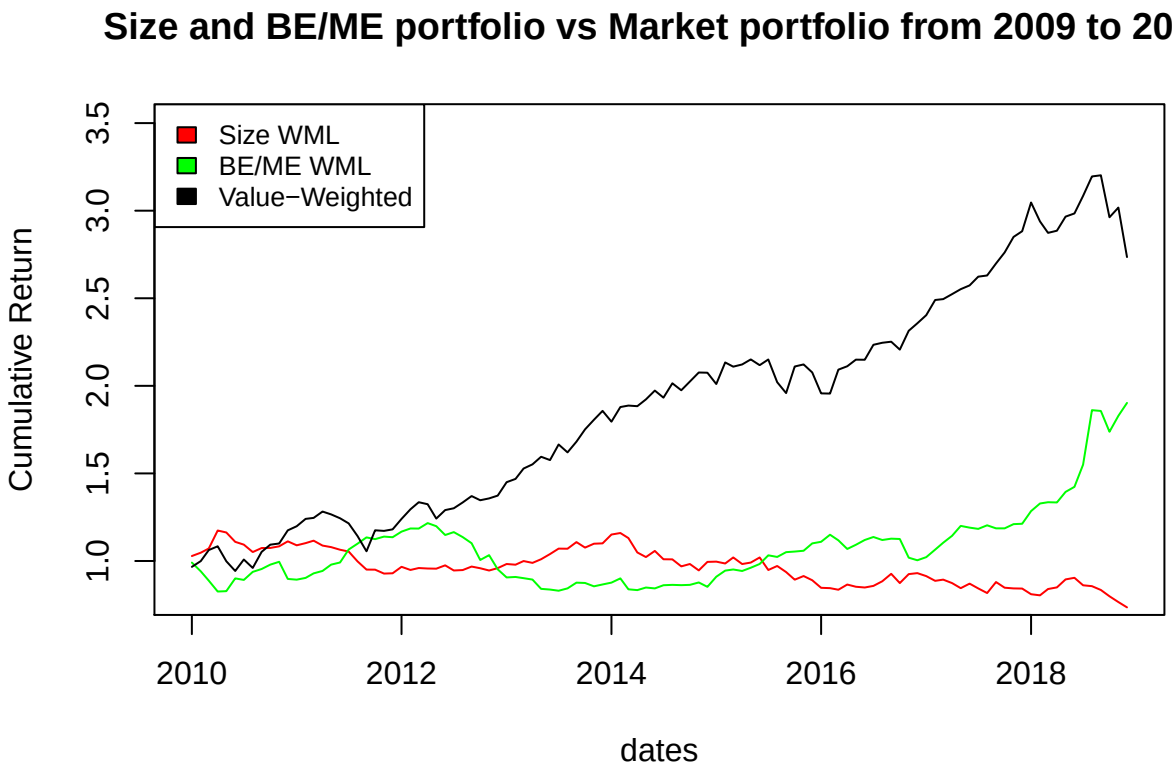
1. Annualized Excess Return: Annualized excess return for decile 1 to 10 (Return above risk free) = Average of monthly excess return multiply by 12. Annualize return for long short portfolio = average of return of (decile 1 - decile 10)
2. Annualized Volatility: Compute portfolio standard deviation as the sample standard deviation across years for each decile and long short and annualize it by multiplying by square root of 12.
3. Sharp Ratio: Compute the sharp ratio using the following formula: Sharp ratio = Annualized Excess Return / Annualized Volatility
4. Skewness: Compute the skewness of the return using the following formula: skewness(monthly return), where skewness is the function from library moments to calculate the skewness.
5. Correlation: Find the correlation between the replicated decile and long short portfolio and the one on Fama French website named *Portfolios_Formed_on_BE-ME.csv*

Question 4:

Compare the long short portfolio from size and BE/ME with market portfolio for the last 10 years. As observed, the return for the size, BE/ME portfolio are lower than the market portfolio.

Table 3: SMB portfolios

Return_Statistics	SMB
Excess Mean Return	-0.0227876
Volatility	0.1075147
SR	-0.2119490
Skewness	0.5724543
Correlation	0.9878124



Question 5:

Summarny for SMB portfolio show in table 3:

1. Annualized Excess Return: Annualized excess return = Average of monthly return minus Fama French risk free rate and multiply by 12.
2. Annualized Volatility: Compute portfolio standard deviation as the sample standard deviation across years for each decile annualize it by multiplying by square root of 12.
3. Sharp Ratio: Compute the sharp ratio using the following formula: Sharp ratio = Annualized Excess Return / Annualized Volatility
4. Skewness: Compute the skewness of the return using the following formula: skewness(monthly return), where skewness is the function from library moments to calculate the skewness.
5. Correlation: Find the correlation between the replicated SMB portfolio and the one in Fama French website named *F-F_Research_Data_Factors.csv*

Summarny for HML portfolio show in table 4:

Table 4: HML portfolios

Return_Statistics	HML
Excess Mean Return	-0.0114177
Volatility	0.1043528
SR	-0.1094142
Skewness	-0.0943849
Correlation	0.9729895

1. Annualized Excess Return: Annualized excess return = Average of monthly return minus Fama French risk free rate and multiply by 12.
2. Annualized Volatility: Compute portfolio standard deviation as the sample standard deviation across years for each decile annualize it by multiplying by square root of 12.
3. Sharp Ratio: Compute the sharp ratio using the following formula: Sharp ratio = Annualized Excess Return / Annualized Volatility
4. Skewness: Compute the skewness of the return using the following formula: skewness(monthly return), where skewness is the function from library moments to calculate the skewness.
5. Correlation: Find the correlation between the replicated HML portfolio and the one in Fama French website named *F-F_Research_Data_Factors.csv*

Question 6: Compare and Contrast the characteristics portfolio in 1992 paper and factors portfolio in 1993

Definition of characteristic vs factor portfolio: We often highlight that the value premium can be explained by either risk and/or mispricing. A core aspect of the risk argument is that a portfolio's factor "loading," or covariance, on a specific factor (e.g., Fama and French HML value factor) represent a proxy for some unobserved systematic risk. The characteristics argument claims that value firms earned a higher expected return simply because they have higher B/M ratios (which may be independent of systematic risk).

Characteristics (1992 paper)

Scope: Assets are common stocks only

Set of variables: size (ME) and book-to-market ratio (BE/ME)

The approach to testing asset-pricing model: Fama and French (1992a) use the cross-section regressions of Fama and MacBeth (1973): the cross-section of stock returns is regressed on variables hypothesized to explain average returns.

Summary of the result: Two easily measured variables, ME and BE/ME, provide a simple and powerful characterization of the cross-section of average stock returns for the 1963-1990 period.

The result suggests that:

- Beta doesn't seems to help explain the cross-section of average stock returns
- The combination of size and book-to-market equity seems to absorb the roles of leverage and E/P in average stock returns, at least during our 1963-1990 sample period.
- Stock risks are multidimensional. One dimension of risk is proxied by size, ME. Another dimension of risk is proxied by BE/ME,

Factors (1993 paper)

Scope: Assets are common stocks and U.S. government and corporate bonds

Set of variables: ME BE/ME and term structure variables for bonds

The approach to testing asset-pricing model:This paper uses the time-series regression approach of Black, Jensen, and Scholes (1972). Monthly returns on stocks and bonds are regressed on the returns to a market portfolio of stocks and mimicking portfolios for size, book-to-market equity (BE/ME), and term-structure risk factors in returns.

Summary of the result: For stocks, portfolios constructed to mimic risk factors related to size and BE/ME capture strong common variation in returns, no matter what else is in the time-series regressions. This is evidence that ME and BE/ME equity indeed proxy for sensitivity to common risk factors in stock returns. Moreover, for the stock portfolios we examine, the intercepts from three-factor regressions that include the excess market return and the mimicking returns for size and BE/ME factors are close to 0. Thus a market factor and our proxies for the risk factors related to ME and BE/ME equity seem to do a good job explaining the cross-section of average stock returns.

For bonds, the mimicking portfolios for the two term-structure factors (a term premium and a default premium) capture most of the variation in the returns on our government and corporate bond portfolios. The term-structure factors also 'explain' the average returns on bonds, but the average premiums for the term-structure factors, like the average excess bond returns, are close to 0. Thus, the hypothesis that all the corporate and government bond portfolios have the same long-term expected returns also cannot be rejected.