Problem Set 4

Name: Demi He 306079038
Names of whom I discussed this problem set with: Sumeng Wu

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 2022.

Before calculating the portfolio time series, I conduct a series of data cleaning as part of my PS_Q1 function. Next, I describe my data-cleaning process and their respective assumptions separated into 3 general parts: Retrieving and cleaning data, creating signals, and calculating portfolio returns.

0.0.1 Retrieving and cleaning data

This step includes downloading raw data, cleaning and examining missing returns, and some basic calculations to get data necessary for creating signals in the next step

1. **Compustats:** COMPUSTAT includes annual industrial files of income statement and balance-sheet data
   (a) **Data Period:** January 1973 - December 2022
   (b) **Raw Data:**
      i. **Compustat North America annual fundamental statistics :** Variables include gvkey, fiscal year, SHQ, CEQ, PSTK, AT, LT, MIB, TXDITC, ITCB, TXDB, PSTKRV, PSTKL, PSTK.
      ii. **Postretirement Benefit Asset values:** gvkey, datadate, prba from Compustat Pension Annual data.
   (c) **Calculate BE:** To calculate Book Equity, we merge PRBA dataset with Compustat dataset by GVKEY and followed steps as following:
      i. The shareholder's equity is "Stockholders' Equity - Total" (SEQ). If not available, we use "Common/ Ordinary Equity - Total" (CEQ) plus "Preferred/Preference Stock (Capital) - Total" (PSTK). If not available, we use

"Assets - Total" (AT) minus "Liabilities - Total" (LT) minus "Minority Interest (Balance Sheet)"(MIB). If not available, then use AT minus LT. If data is missing, we will drop this company from our portfolio.

ii. Deferred taxes and investment tax credit (DT) is reported in Compustat as "Deferred Taxes and Investment Tax Credit" TXDITC. If not available, use "Investment Tax Credit (Balance Sheet)" (ITCB) plus "Deferred Taxes (Balance Sheet)" (TXDB). If not available, sum what is not missing. If DT data is missing, we set it to be 0.

iii. Book value of preferred stock (PS) is redemption value, which is variable "Preferred Stock Redemption Value" (PSTKRV). If not available, use liquidation value, which is "Preferred Stock Liquidating Value" (PSTKL). If not available, use par value, which is "Preferred/Preference Stock (Capital) - Total" (PSTK). If PS data is missing, we set it to be 0.

iv. BE = SHE - PS + DT - PRBA. We need SHE to calculate BE, other variables will be set to 0 if missing.

(d) **Check duplicate GVKEY:** I noticed there are some GVKEY with 2 multiple lines of data from the same year, this might be caused by changing of reporting day. For future merging data purposes, I kept only the latter BE in the same year if there are duplicates.

(e) **Cleaned dataset:** Cleaned and checked data set for merging in the following steps is **company_cleaned** with variables gvkey datadate BE SHE PS and year.

2. **CRSP monthly stock:**

(a) **Data Period:** January 1973 - December 2022

(b) **Raw Data:** CRSP monthly stock files from 1926/01-2019/12 with universe of stocks. Variables include PERMCO, PERMNO,date, share code, exchange code, delisting return, holding period return, price and shares outstanding.

(c) **Universe of stocks:** Following Ken French procedure, I restrict the sample to common shares (share codes 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).

(d) **Missing returns:** Since assigning the average of return to any missing value is the same as dropping when calculating EW and VW and considering a large amount of data, I simply dropped any NA return.

(e) **Delisting return calculation:** First I got the delisting return from crspq.msedelist and merged the delisting return to crsp.msf if the delisting happened during that month for a specific permno. Then I used cum-dividend total return to get

$$r_{i,t}: \begin{cases} r^h_{i,t} & \text{if } r^h_{i,t} = \text{missing} \\ r^d_{i,t} & \text{if } rd_{i,t} = \text{missing} \\ (1 + r^h_{i,t})(1 + rd_{i,t})^{-1} & \text{if both not missing} \end{cases}$$

(f) **Lagged market cap and June market cap:** Since we will need lagged Decembe market cap and June market cap for future signal calculations, I created lag_me_dec and me_june accordingly.

(g) **Cleaned dataset:** Cleaned and checked data set for merging in the following steps is **CRPS_3**. I also created dataframe **CRPS_cleaned** with only June and December data for signal calculation.

3. **Linktable:** Linking table data from CRSP/COMPUSTAT merged with variables gvkey, CIK, SIC, LINKPRIM, LIID, LINKTYPE, LPERMNO, LPERMCO, LINKDT, LINKENDDT

(a) **Data Period:** January 1973 - December 2022

(b) **Linktype:** To best replicate Fama French paper, we will constrain the linktype is 'L', 'C' and 'P', which indicates that there are primary securities in Compustats and exist link between compustat and CRSP.

(c) **Missing data - linkenddt:** Fill NA linkenddt with 12/31/2023 indicating that specific gvkey is linked to that permno until today.

4. **Merging data:**

(a) **Merge linktable with compustat:** So that we have permno for each gvkey

(b) **Multiple gvkey:** For the situation that gvkey in compustats changed during time for the same permno, we checked if the date in compustat lands within linkdt and linkenddt in linktable.

(c) **Merge compustat with CRSP:** After cleaning, we merged the **comp_clean** with **CRSP_clean** by matching gvkey and peremno.

(d) **Check for redundancy:** I grouped the dataframe by 'gvkey' and check for redundant 'year' within each group.

(e) **Subsetting for signal calculation:** According to Fama French's paper: to be included in the return tests for July of year t, a firm must have a CRSP stock

price for December of year t - 1 and June of year t. All time series with applicable data is subsetted into dataframe **comp_merge_CRSP**.

0.0.2 Creating signals

1. **Compute BE ratio:** B/E ratio = $\frac{bookedequity}{lag\_me\_dec}$ In order to standardize the unit, we times 1000 to make all units into thousands.

2. **Creating bins:**

   (a) **NYSE data:** To best replicate Fama French's paper, I kept only data from NYSE for signal creation (subsetting exchcd == 1)

   (b) **size bins:** According to Fama French's paper, size bins are created using market cap data in June of each year, rebalanced annually.

   (c) **Value bins:** According to Fama French's paper, value bins are created using BM ratio we just calculated, rebalanced annually.

   (d) **Size Factor strategy:** According to Fama French's paper, any corp with lagged market cap larger than median is considered large corp and other companies are considered small cap. Both rebalanced annually.

   (e) **Value Factor strategy:**According to Fama French's paper, the cutoffs for high/ medium/ low value company are 30% and 70%. All rebalanced annually.

3. **Merge back to full data set:** Since we only used data from June and December for afore signal creation. We merged assigned bins and factors to full data set by matching year and permno. The resulting dataframe is **full_df**.

0.0.3 Calculating portfolio returns

1. **Size and Value bins:**

   (a) **Portfolio weights:** Calculated portfolio weights of each bin of each permno using the equation below:
   $w_{i,t} = \frac{me_{i,t-1}}{\sum_i me_{i,t-1}}$

   (b) **Definition of portfolio weights:** Value weighted return is calculated by getting the weight return of each bin of each permno according to the weight of its individual market equity divided by the total market equity. As such, within each bin, all stocks are assigned weights according to it's individual market cap.

(c) **VWAP:** I calculated 1 VWAP return for each bin of each month. Resulting a total of 6000 lines of data, representing 50 years, 12 month each year and 10 bins each month.

2. **Factor strategy returns:**

   (a) **Cetegorize company factors:** All companies are allocated in one of the three HML identifications and one of two SB identifications. Combining these 2 factors will give up six categories: a company will fit in either small value(SL), smalle neutral(SM), small growth(SH), big value(BL), big neutral(BM), and big growth(BH).

   (b) **Factor return calculation:**I followed instructions on FF website to get factor strategy returns:

      i. **SMB:** SMB = $\frac{1}{3}$ (SL+SM+SH) - $\frac{1}{3}$ (BL+BM+BH)

      ii. **HML:** SMB = $\frac{1}{2}$ (SL+BL) - $\frac{1}{2}$ (SH+BH)

All returns are matched and merged according to year and month to dataframe **port_ret**

Question 2

For each size decile and the long-short portfolio, report the annualized average excess returns, annualized volatility, Sharpe Ratio, and skewness. Also report the correlation between the portfolios that you have constructed (the 10 portfolios and the long-short portfolio) and those from French's website. I got risk free return and FF decile return from French's website and performed comparison from my replicated return recorded in **port_ret**

0.0.4 Steps description:

1. **Data Period**: January 1973 - December 2022

2. **Excess log return:**

   (a) Excess return: Replicated return - risk free rate

   (b) Excess log return: log(1+Excess return)

3. **WML return:** For each month, WML return is the Excess log return of bin 10 minus Excess log return of bin 1.

4. **Annualized Mean:** I calculate annulized mean from the monthly time series for each bin with formula below:
   $\mu_{annulized} = 12 * \mu_M$

5. **Annualized Standard Deviation:** I calculate annulized mean from the monthly time series for each with formula below:
   $\sigma_{annulized} = \sqrt{12} * \sigma_M$

6. **Sharpe Ratio:**

$$SharpeRatio_{annulized} = \frac{\mu_{annulized}}{\sigma_{annulized}}$$

7. **Skewness:** I calculate skewness of r_t from the monthly time series directly (no annualization) for each bin.

0.0.5 Result presentation:

Bins	1	2	3	4	5	6	7	8	9	10	WML
Excess return	-0.032525	0.028214	0.053225	0.069437	0.08268	0.089982	0.087289	0.094267	0.100486	0.101062	0.108138
sigma	0.22624	0.231967	0.219303	0.210931	0.205646	0.19265	0.187627	0.182289	0.167352	0.151875	0.153691
SR	-0.143765	0.12163	0.242701	0.329193	0.402049	0.467077	0.465226	0.517133	0.600446	0.665425	0.703609
sk(m)	-0.835369	-0.903897	-0.914516	-0.861405	-0.905467	-0.856662	-0.935392	-0.809744	-0.743003	-0.539688	-0.659696
Corr	0.939383	0.964044	0.980565	0.984524	0.98134	0.987464	0.983305	0.984628	0.990132	0.992422	0.863374

Table 1

Question 3

For each book-to-market decile and the long-short portfolio, report the annualized average excess returns, annualized volatility, Sharpe Ratio, and skewness. Also report the correlation between the portfolios that you have constructed (the 10 portfolios and the long-short portfolio) and those from French's website.

0.0.6 Steps description:

Similiar to Question 3, just replace size vwap return with value vwap return

Bins	1	2	3	4	5	6	7	8	9	10	WML
Excess return	0.090023	0.089346	0.099993	0.096353	0.103547	0.101361	0.1077	0.105731	0.115036	0.141264	0.042239
sigma	0.184821	0.167813	0.164155	0.173629	0.163738	0.168839	0.162786	0.170195	0.194226	0.216458	0.177378
SR	0.487082	0.532411	0.609136	0.554937	0.632395	0.600341	0.661604	0.621232	0.592279	0.652614	0.238127
sk(m)	-0.442324	-0.58009	-0.698429	-0.886442	-0.698825	-0.755327	-0.304868	-0.816305	-1.056586	-0.32264	0.576112
Corr	0.987082	0.974276	0.962966	0.955838	0.948913	0.93025	0.944249	0.930074	0.921933	0.918759	0.824912

Table 2

0.0.7 Result presentation:

Question 4

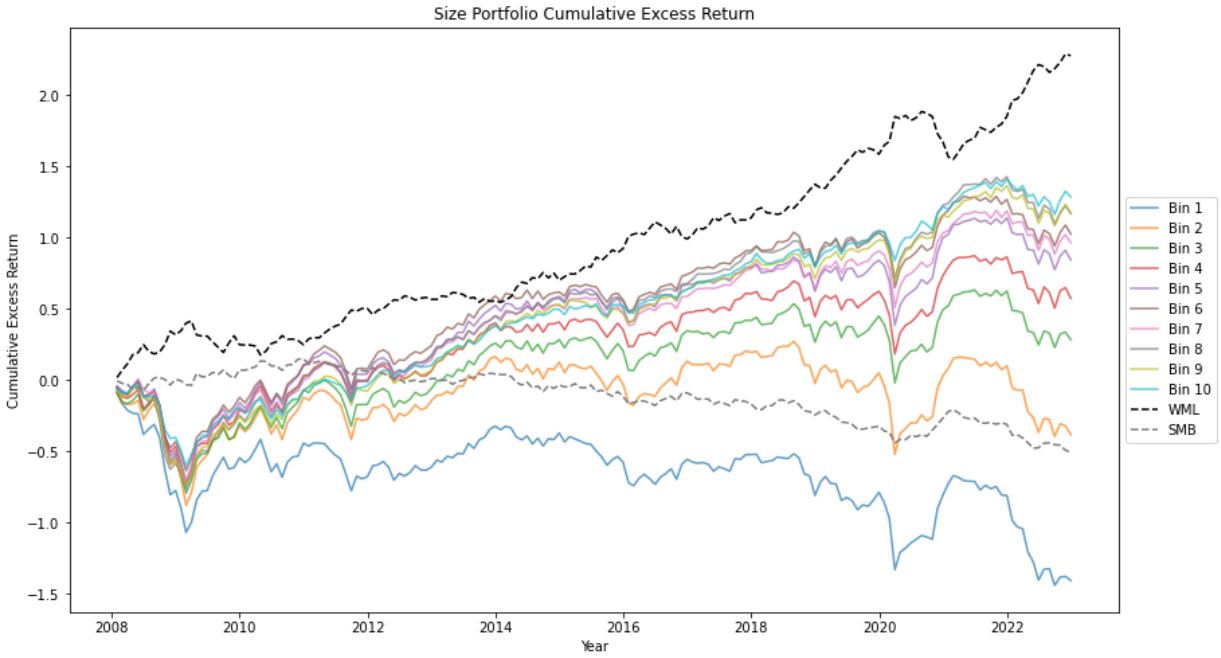
Has the value and size anomaly worked in the past few years? Show some empirical evidence.

0.0.8 Size anomaly:

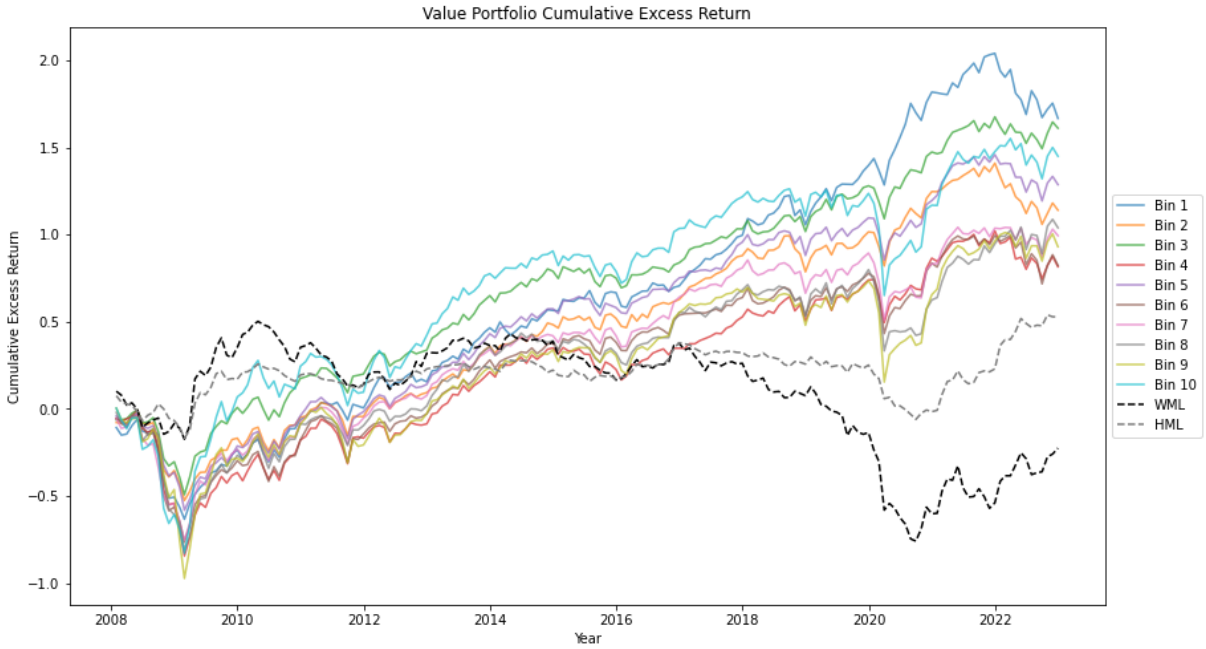
For size anomaly graph starting from 2008 we can observe that: 1. Small cap corp has always outperformed large cap corp despite market conditions; 2. Small cap corps have remained strong during COVID; 3. WML portfolio have consistent positive returns, which further proves our serious 2 observations.

0.0.9 Value anomaly:

For value anomaly graph starting from 2008 we can observe that: 1. Growth and value stock returns are not as monotonic as size stocks, meaning that growth stocks have been outperforming bins with stocks of lower book to market ratios; 2. Growth stocks seems to be performing extra well during COVIS crisis; 3. Value stock has been strong since 2018; 4. WML portfolio has consistent negative returns, which further proves our serious 3 observations.



Graph 1



Graph 2

Figure 1: Size and Value returns starting from 2008

Question 5

For both HML and SMB portfolios, report the annualized average excess returns, annualized volatility, Sharpe Ratio, and skewness. Report correlations between the replicated factors and the factor from French's website. Have the factors been consistent across time? Show some empirical evidence.

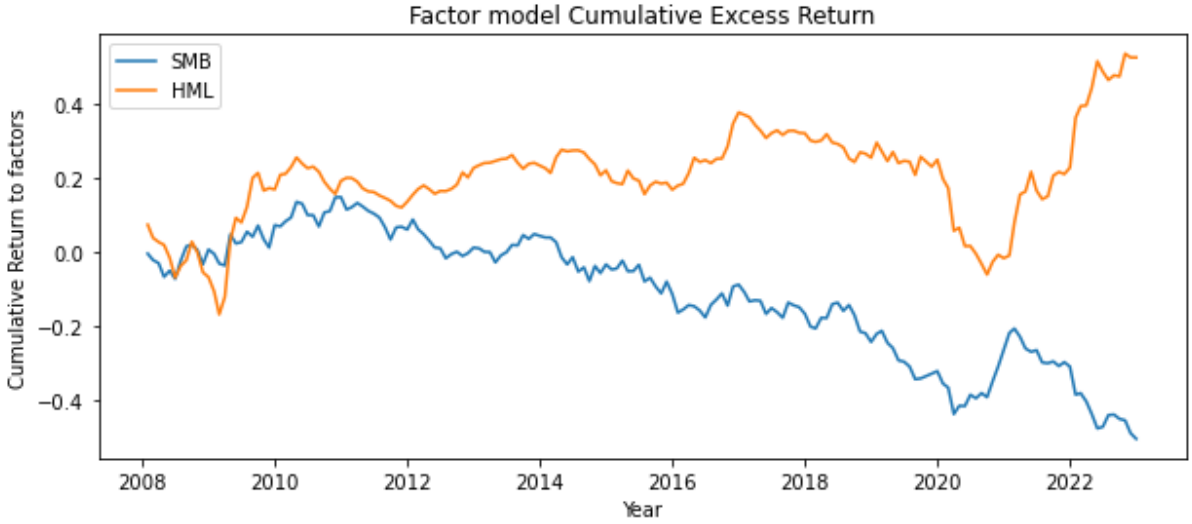
0.0.10 Step description:

Similiar to Question 3, just replace size vwap return with HML and SMB factor return

0.0.11 Result presentation:

Strategy	HML	SMB
Excess return	0.058178	-0.025686
sigma	0.102917	0.094459
SR	0.565285	-0.271927
sk(m)	0.659658	0.044762
Corr	0.86329	0.93498

Table 3



Graph 3

HML factor has similar trends as the WML portfolio in value bins. SMB factor however, is demonstrating negative returns compared to positive WML portfolio in size bins.

Question 6

Compare and contrast using the characteristic portfolios (Fama and French 1992) and the factor portfolios (Fama and French 1993).

In their study, Fama and French (1992) examine the collective influence of market beta, size, earnings/price ratio (E/P), leverage, and book-to-market equity on average stock returns across different stocks. Surprisingly, their findings indicate that market beta has limited explanatory power when it comes to average returns in Fama-Macbeth regressions that incorporate various combinations of beta with size and book-to-market equity.

In their subsequent 1992 paper, Fama and French observe that there is some correlation between size and book-to-market equity when considered individually in relation to average stock returns. Building upon this, their 1993 study delves further into exploring whether the combination of size and book-to-market equity can provide economic explanations for stock returns. Remarkably, their research demonstrates that the three-factor model, encompassing market beta, size, and book-to-market equity, effectively captures the common variation in stock returns, thereby offering a more robust framework for explaining stock market dynamics.