# PROBLEM SET 4

*GOURPRABH*

## 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), as well as the HML and SMB factors as defined in Fama and French (1993). 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 2019.*

Datasets used -

**CRSP** -
This dataset was downloaded from WRDS CRSP data. The range of the data is from *January 1970* to *December 2019* with monthly frequency. The dataset contains the following columns -
1. *PERMNO* : Unique security identifier by CRSP.
2. *PERMCO* : Unique firm identifier by CRSP.
3. *SHRCD* : Share code detailing the type of security, for example - common shares or American Depsoitory Receipts etc.
4. *EXCHCD* : Exchange Code detailing the exchange on which the security is listed like NYSE, NASDAQ etc.
5. *PRC* : Closing price of the security on that year and month.
6. *RET* : Holding period return on that security over that month.
7. *DLRET* : Delisting return of that security.
8. *RETX* : Holding period return without dividends on that security over that month.
9. *DLRETX* : Delisting return without dividends of that security.

**COMPUSTAT** -
The dataset was downloaded from WRDS Compustat's North America annual fundamentals data. The range of the data is from *1970* to *2020* with yearly frequency. The dataset contains the following columns -

1. *gvkey* : Unique Global company key assigned to each firm.
2. *datadate* : The date that the data was compiled.
3. *fyear* : The financial year of which the fundamental data is from.
4. *at* : Total assets of the firm.
5. *ceq* : Total Common/Ordinary Equity.
6. *itcb* : Investment Tax Credit.
7. *lt* : Total Liabilities.
8. *mib* : Minority Interest.

9. *pstk* : Total Preferred/Preference Stock(Capital).

10. *pstkl* : Preferred Stock Liquidating Value.

11. *pstkrv* : Preferred Stock Redemption Value.

12. *seq* : Total Stockholder's Equity.

13. *txdb* : Deferred Taxes.

14. *txditc* : Deferred Taxes and Investment Tax Credit.

**COMPUSTAT PENSION** -
The dataset was downloaded from WRDS Compustat's North America annual pension data. The range of the data is from *1973* to *2020* with yearly frequency. The dataset contains the following columns -

1. *gvkey* : Unique Global company key assigned to each firm.

2. *datadate* : The date that the data was compiled.

3. *prba* : Postretirement Benefit Asset.

4. *conm* : Company name

**LINKING TABLE FOR CRSP AND COMPUSTAT** -

This dataset was downloaded from WRDS Linking Suite's COMPUSTAT/CRSP Link data. The linking table provides a way to link unique GVKEY of compustat with PERMCO and PERMNO of CRSP, The range of the dataset is from *1946* to *2020*.The dataset contains the following columns -

1. *gvkey* : Unique Global company key assigned to each firm.

2. *conm* : Company name

3. *LINKPRIM* : Primary Link Marker. This contains four character values - C,J,N,P, which stand for -
   P - Primary, identified by Compustat in monthly security data.
   J - Joiner secondary issue of a company, identified by Compustat in monthly security data.
   C - Primary, assigned by CRSP to resolve ranges of overlapping or missing primary markers from Compustat in order to produce one primary security throughout the company history.
   N - Secondary, assigned by CRSP to override Compustat. Compustat allows a US and Canadian security to both be marked as Primary at the same time. For Purposes of the link, CRSP allows only one primary at a time and marks the others as N.

4. *LIID* : Security Level identifier.
5. *LINKTYPE* : It is a 2 character code providing details about the type of link. We would be primarily focusing on 2 link types:
   LC – Link research complete. Standard connection between databases. LU – Unresearched link to issue by CUSIP.

6. *LPERMNO* : Linked CRSP PERMNO, 0 if no CRSP security link exists.

7. *LPERMCO* : Linked CRSP PERMCO, 0 if no CRSP company link exists.

8. *LINKDT* : First effective calendar date of link record range.

9. *LINKENDDT* : Last effective calendar date of link record range, E if the linking is still valid at the end date of data.

**LINKING CRSP PERMCO TO GVKEY USING LINKING TABLE**

First, I cleaned up the data from CRSP in the below given steps -

*Step 1 SHRCD* - I restricted my dataset to rows in which SHRCD had the values *10,11*. Essentially I restricted my shares to ordinary commons shares of companies incorporated inside the US.
*Step 2 EXCHCD* - I restricted my dataset to rows in which EXHCD had the values *1,2,3*. Essentially I restricted my firms to one listed on New York, American and NASDAQ stock exchange.
*Step 3* To tackle the fact that some of the PRC data had negative values, which indicated that the data was just the average of the bid and ask price at the end of that month, I took the absolute value of the price column(PRC).
*Step 4* The return column (RET) had some rows as 'C',-66.0,-77.0,-88.0 and -99.0 which just denoted that the return data was not available for that month. So, I set the return data of that month to be 0.

**Calculating total returns** -
I converted all *NA* rows in my *RET*,*DLRET*,*RETX* and *DLRETX* data to 0. Then I calculated *total return* AND *total return without dividends* for each row as -

$$tot\ ret = (1 + RET)(1 + DLRET) - 1$$

$$tot\ retX = (1 + RETX)(1 + DLRETX) - 1$$

The above formula covers the following three cases -
1. $DLRET = NA$ - The above formula would give back the return data (as DLRET is set to 0 and 1 + DLRET becomes 1).
2. $RET = NA$ - In this case, the above formula gives back DLRET as outlined above.
3. $RET = NA\ and\ DLRET = NA$ - In this case, the above formula returns 0.

Similarly for RETX and DLRETX.
For all the rows where tot_retX was NA, I substituted data from the corresponding row of tot_ret.

**Lagged market value**
I shifted market equity of each firm, taking care through a valid flag that the previous row contained the data from the same firm, filling *0* in the rows where this operation created NA values.

**Value-weighted return for each firm (PERMCO)** -
I first created market weight for each security(PERMNO) of same firm(PERMCO) each month based on its market value last month. The formula I used is -

$$(Market\ \ weight\ \ of\ \ Security_{ip})_t = \frac{(market\ \ value\ \ of\ \ Security_{ip})_{t-1}}{(Total\ \ lagged\ \ market\ \ value\ \ of\ \ all\ \ Security_p)_{t-1}}$$

Then, I took a weighted sum of the stocks return in each decile portfolio where the weights were from the formula I calcualted above.

$$(Firm\ \ VW\ \ Ret)_{t,p} = sum(Return_{i,t,p} * weight_{i,p})$$

where t stands for the month,
i stands for the specific security
and p stands for the specific firm

I repeated the above process for total return without dividends.

*Step 5* I removed duplicated values of permco and year.

*Step 6* I merged the CRSP data with the CRSP-COMPUSTAT linking table along PERMCO and LPERMCO allowing duplicates at this point.

**Removing Duplicates from the merged data**

*By linking start and end dates* - I restricted the data to the rows which satisifoed the following conditions -

$$Date >= LINKDT(Linking\ \ Start\ \ Date)$$

$$Date <= LINKENDDT(Linking\ \ End\ \ Date)$$

*Using LC* - We know that LC signifies a standard connection between datasets. Hence, for all the duplicate merges that have LC as one of the merges, I removed the rest and only kept the rows with LC as connection type.

*Using LINKPRIM = P* -We know that P signifies the primary link. Hence, for all the duplicate merges, I removed the rest and only kept the rows with LINKPRIM = P.

*Using LINKPRIM = C* - We know that C signifies the primary link assigned by CRSP to resolve dupicates. Hence, for all the remaining duplicate merges, I removed the rest and only kept the rows with LINKPRIM = C.

*Using LIID = 1* - LIID signifies security level. Hence, for all the remaining duplicate merges which have multiple values of LIID, I removed the rest and only kept the rows with LIID = 1.

For removing the rest of the duplicates, I kept only the rows with link that is currently valid. After that, for rest of the duplicates, I kept the link taht has been around the longest. For the rest small number of duplicates, I kept those with the gvkey value that ahs been around the longest.

## MERGING COMPUSTAT WITH PENSION DATA

Here we use GVKEY column to merge the two datasets. The process followed is -
*Step 1* : I set all the rows of the data containing NA value in prba column to 0.
*Step 2* : I extracted the year of the datadate. I will use this column to correspond with datadate column of COMPUSTAT.
*Step 3* : I first removed duplicate rows having the same gvkey, year and prba value. This still left us 40 duplicate rows which I removed by selecting the first of the duplicated rows.
*Step 4* : I merged the COMPUSTAT data with the pension data on gvkey and year column. Since we already took care of the duplicate rows in pension data, the merge is perfect with the same number of rows as the original COMPUSTAT data.

## CALCULATING BOOK VALUE OF EQUITY(BE) OF FIRMS

The Book Value of Equity is given by the following formula -

$$BE = SHE - PS + DT - PRBA$$

Where,
SHE : Shareholder's equity
DT : Deferred Taxes and investment tax credit
PS : Book value of preferred stock
PRBA : Postretirement Benefit Asset

*SHE* - We use Total stockholder's equity (SEQ) as SHE.
If SEQ is not available, we use -

$$Total\ \ Common/Ordinary\ \ Equity(CEQ) + Total\ \ Preferred/Preference\ \ Stock(PSTK)$$

If not available, we use -

$$Total\ \ Assets(AT) - Total\ \ Liabilities(LT) - Minority\ \ Interest(MIB)$$

If not available, we use -

$$Total\ \ Assets(AT) - Total\ \ Liabilities(LT)$$

*DT* : We use the variable Deferred Taxes and Investment Tax Credit(TXDITC).
If not available, we use -

$$Investment\ \ Tax\ \ Credit(ITCB) + Deferred\ \ Taxes\ \ (TXDB)$$

If not available, we sum what is not missing.

*PS* : We use redemption value, which is variable Pre- ferred Stock Redemption Value (PSTKRV). If not available, we use liquidation value, which is Preferred Stock Liquidating Value (PSTKL). If not available, we use par value, which is Total Preferred/Preference Stock (PSTK).

## MERGING CRSP WITH COMPUSTAT

I merged my CRSP linked data to COMPUSTAT using gvkey and financial year(fyear) column of COMPUSTAT and gvkey and date column's year value.
Fama-French form their size and B/M decile portfolios at the end of June.**NYSE** breakpoints are used at the end of June to sort all universe of stocks in size portfolios. The data used is market equity at end of June. For BE/ME portfolios starting from July of t, they are using BE of the last fiscal year of that firm divided by ME of December of t - 1. The breakpoints are **NYSE** breakpoints.

To calculate SMB(Small minus big) and HML(High Medium Low) returns, Fama French first placed the following restrictions on the data -
1. Valid CRSP stock prices for December of t - 1 and June of t.
2. Only those firms are included that have appeared on COMPUSTAT for two years.

### SMB Return

Fama and French use the market equity value of June of year t to sort firms into two portfolio - small and big from July of year t to June of year t + 1. They use NYSE median size as the breakpoint for the split. The value-weighted returns are calculated using the following formula -

$$SMB\ Ret = \frac{1}{3}(S/L + S/M + S/H) - \frac{1}{3}(B/L + B/M + B/H)$$

where,
S = Small
B = Big
H = High
M = Medium
L = Low

### HML Return

Fama and French use the Book to market equity value to sort firms into three portfolio - low, medium and big from July of year t to June of year t + 1. The book common equity is from the fiscal year ending t - 1 which is then divided by market equity at the end of december of t - 1. They use NYSE B/M as the breakpoint for the split, which corresponds to 30% for Low, Middle 40% as Medium and top 30% as High. The value-weighted returns are calculated using the following formula -

$$HML\ Ret = \frac{1}{2}(S/H + B/H) - \frac{1}{2}(S/L + B/L)$$

## 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 port- folios that you have constructed (the 10 portfolios and the long-short portfolio) and those from French's website.*

The required annualized average excess returns, annualized volatility, Sharpe Ratio, skewness and correlation with Fama-French size portfolios are shown below -

| | Annualized Average Excess Return | Annualized Volatility | Sharpe Ratio | Skewness | Correlation |
|---|---|---|---|---|---|
| Decile 1 | 0.0843480 | 0.2170969 | 0.3885270 | -0.2115957 | 0.9985035 |
| Decile 2 | 0.0869189 | 0.2241955 | 0.3876924 | -0.2447023 | 0.9986819 |
| Decile 3 | 0.0945257 | 0.2086320 | 0.4530738 | -0.5297123 | 0.9985074 |
| Decile 4 | 0.0856417 | 0.2034613 | 0.4209239 | -0.5567163 | 0.9983912 |
| Decile 5 | 0.0888891 | 0.1973964 | 0.4503076 | -0.5168147 | 0.9981666 |
| Decile 6 | 0.0849305 | 0.1830225 | 0.4640442 | -0.5885967 | 0.9979015 |
| Decile 7 | 0.0856954 | 0.1827925 | 0.4688128 | -0.5088674 | 0.9981467 |
| Decile 8 | 0.0851676 | 0.1765817 | 0.4823126 | -0.4894341 | 0.9975118 |
| Decile 9 | 0.0755030 | 0.1640481 | 0.4602488 | -0.5025328 | 0.9974077 |
| Decile 10 | 0.0608458 | 0.1523430 | 0.3993998 | -0.3881851 | 0.9977011 |
| Long-short | -0.0688172 | 0.1622101 | -0.4242469 | -0.8769161 | 0.9944965 |

## QUESTION 3

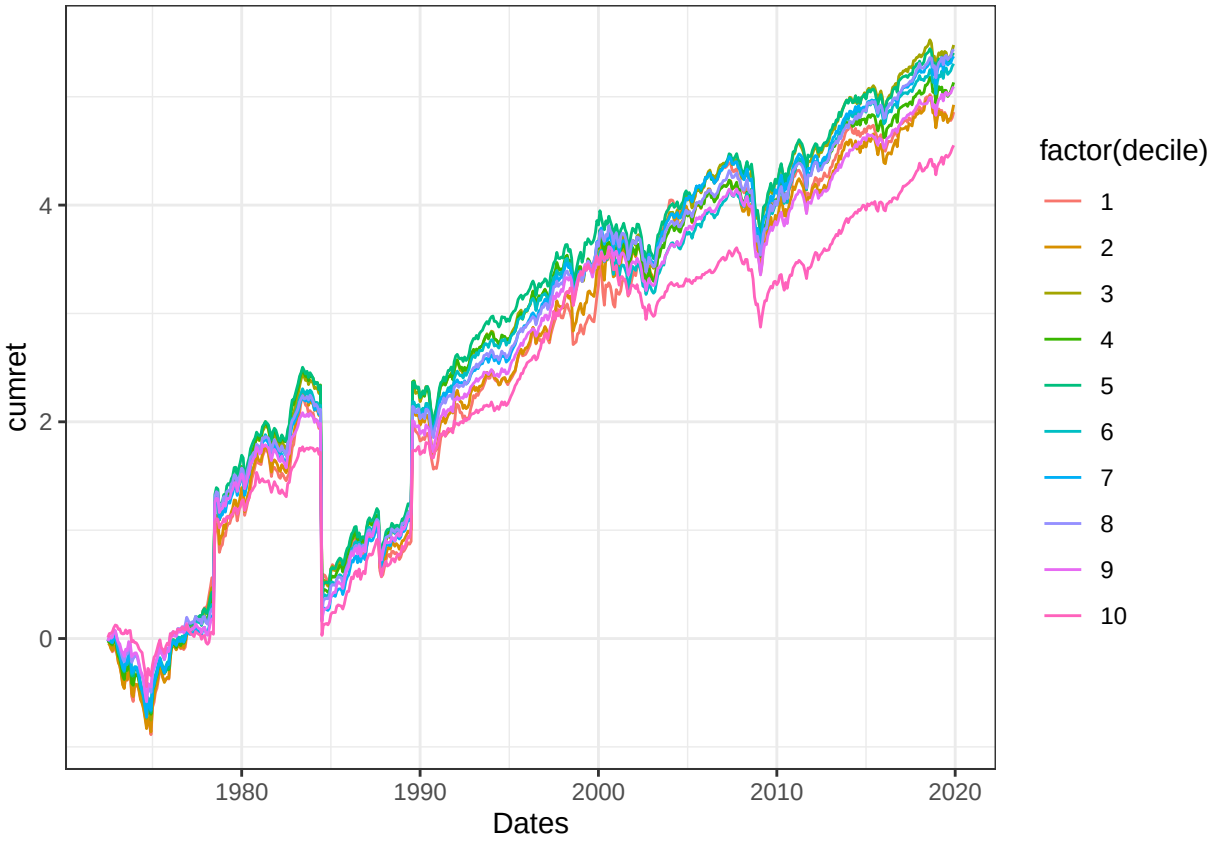
*For each book-to-market decile and the long-short portfolio, report the annualized average excess returns, annualized volatility, Sharpe Ratio, and skewness. Also report the correlation between the portfolios that you have constructed (the 10 portfolios and the long-short portfolio) and those from French's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

The required annualized average excess returns, annualized volatility, Sharpe Ratio, skewness and correlation with Fama-French B/M portfolios are shown below -

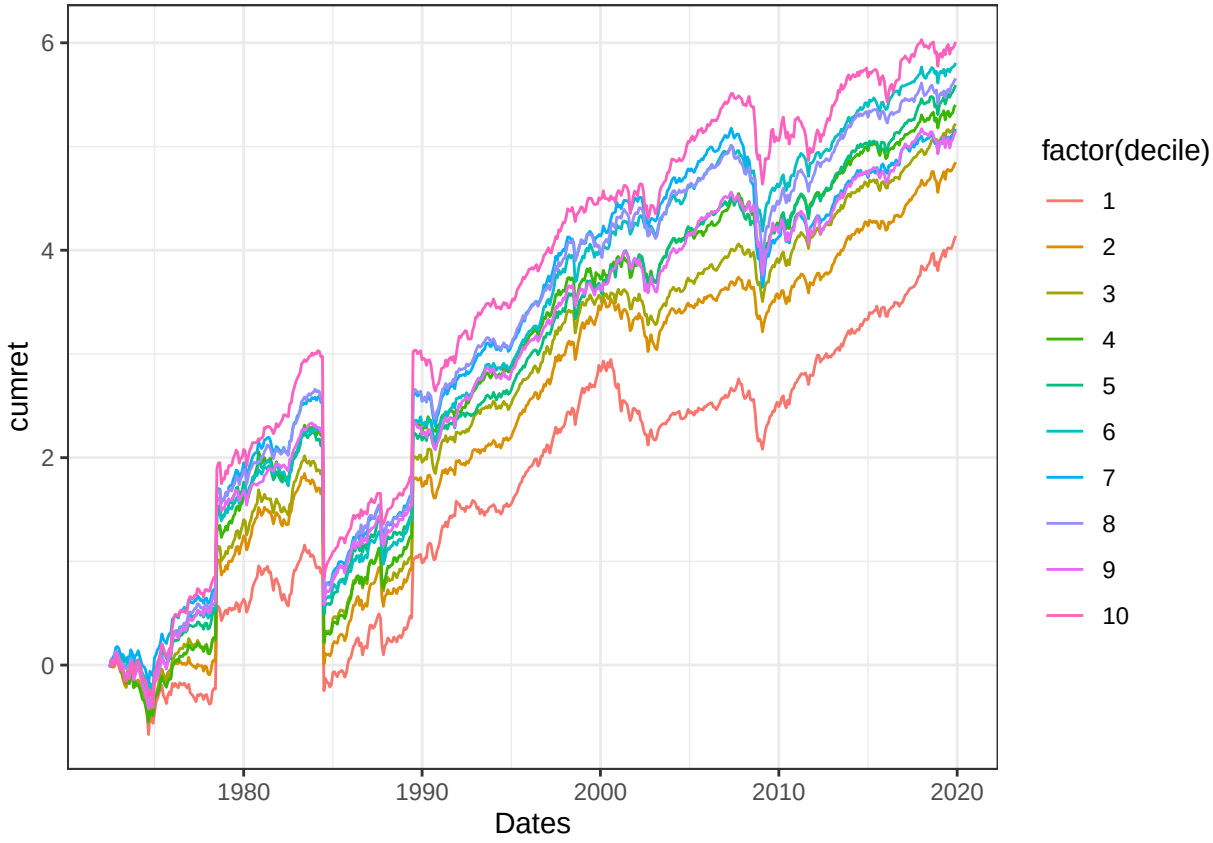
| | Annualized Average Excess Return | Annualized Volatility | Sharpe Ratio | Skewness | Correlation |
|---|---|---|---|---|---|
| Decile 1 | 0.0569256 | 0.1765056 | 0.3225144 | -0.2866180 | 0.9957212 |
| Decile 2 | 0.0709643 | 0.1653719 | 0.4291191 | -0.4913713 | 0.9828217 |
| Decile 3 | 0.0775881 | 0.1608101 | 0.4824829 | -0.4448682 | 0.9711978 |
| Decile 4 | 0.0822237 | 0.1647265 | 0.4991529 | -0.6152009 | 0.9633102 |
| Decile 5 | 0.0858691 | 0.1670797 | 0.5139407 | -0.5448111 | 0.9597973 |
| Decile 6 | 0.0888002 | 0.1582313 | 0.5612051 | -0.4356726 | 0.9431819 |
| Decile 7 | 0.0760812 | 0.1700978 | 0.4472789 | -0.7840278 | 0.9409227 |
| Decile 8 | 0.0869181 | 0.1652048 | 0.5261234 | -0.5564220 | 0.9322098 |
| Decile 9 | 0.0762829 | 0.1660170 | 0.4594885 | -0.2809527 | 0.9397103 |
| Decile 10 | 0.1007889 | 0.1932969 | 0.5214200 | -0.0376269 | 0.9342664 |
| Long-short | -0.0014516 | 0.1520216 | -0.0095486 | 0.5829539 | 0.8862702 |

## QUESTION 4

Below is the plot of cumulative return of all size decile portfolios from 1973 to 2019. We can see that the difference in cumulative return is not strong before 2000. Even after 2000, the divergence is small and hence a long-short portfolio would not earn much. This statement is made taking in consideration the fact that the above returns do not take in factor transaction costs associated with these trades.



Below is the plot of cumulative return of all Book to Market decile portfolios from 1973 to 2019. We can actually see the high book to market firms (value) giving lower returns than low book to market firms in late 1980s. Even after that, the divergence is small and hence a long-short portfolio would not earn much. This statement is made taking in consideration the fact that the above returns do not take in factor transaction costs associated with these trades.



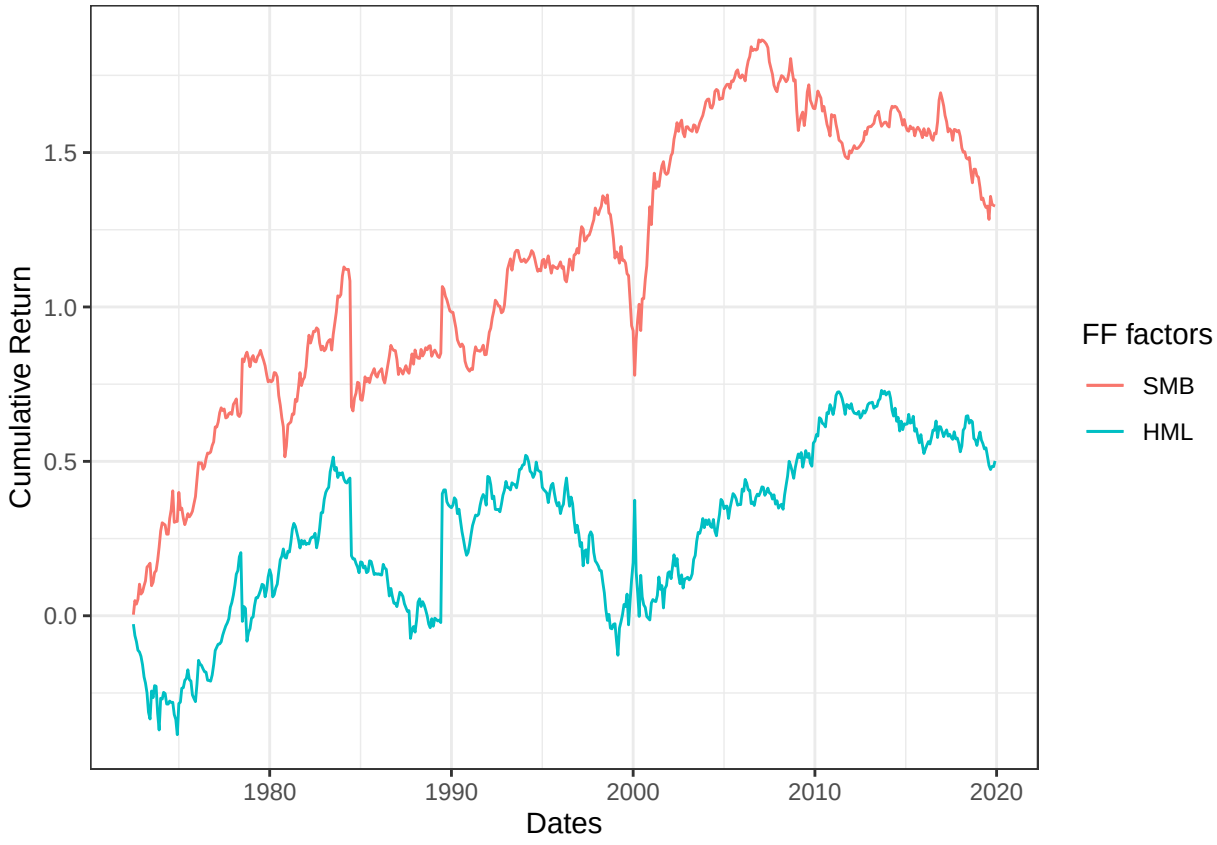
## QUESTION 5

*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.*

The required annualized average excess returns, annualized volatility, Sharpe Ratio, skewness and correlation with Fama-French factor portfolios are shown below -

| | Annualized Average Excess Return | Annualized Volatility | Sharpe Ratio | Skewness | Correlation |
|---|---|---|---|---|---|
| SMB | -0.0256137 | 0.1126667 | -0.2273409 | 0.2830977 | 0.9618795 |
| HML | -0.0134391 | 0.1010088 | -0.1330493 | 0.0823163 | 0.9255694 |

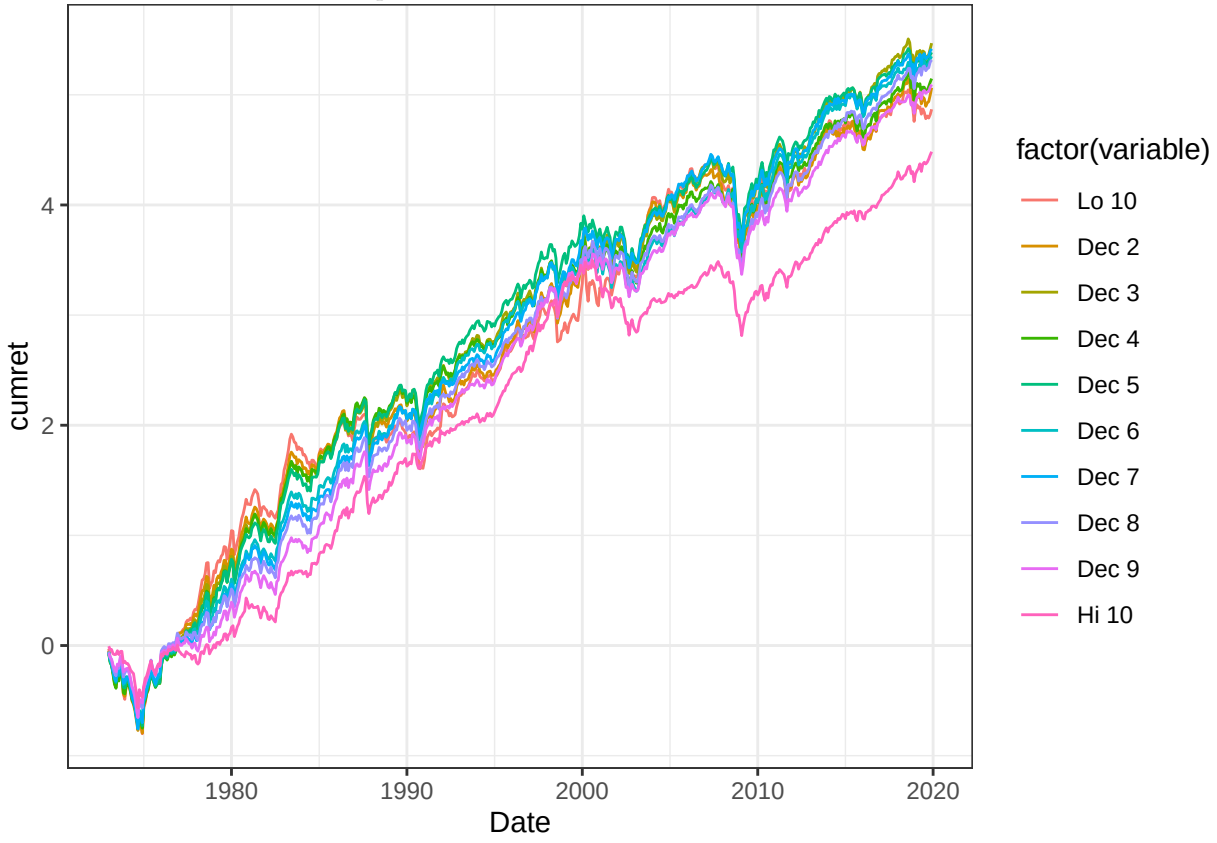
Below, we have plotted the *Cumulative returns* on the SMB and HML portfolios from 1973 to 2019. We can see that SMB factor has been consistent over time with only one major drawdown appearing during the Dot com bubble in 2000. While, HML factor has been inconsistent over the years with the cumulative returns over mid 1980s to 2019 almost 0.



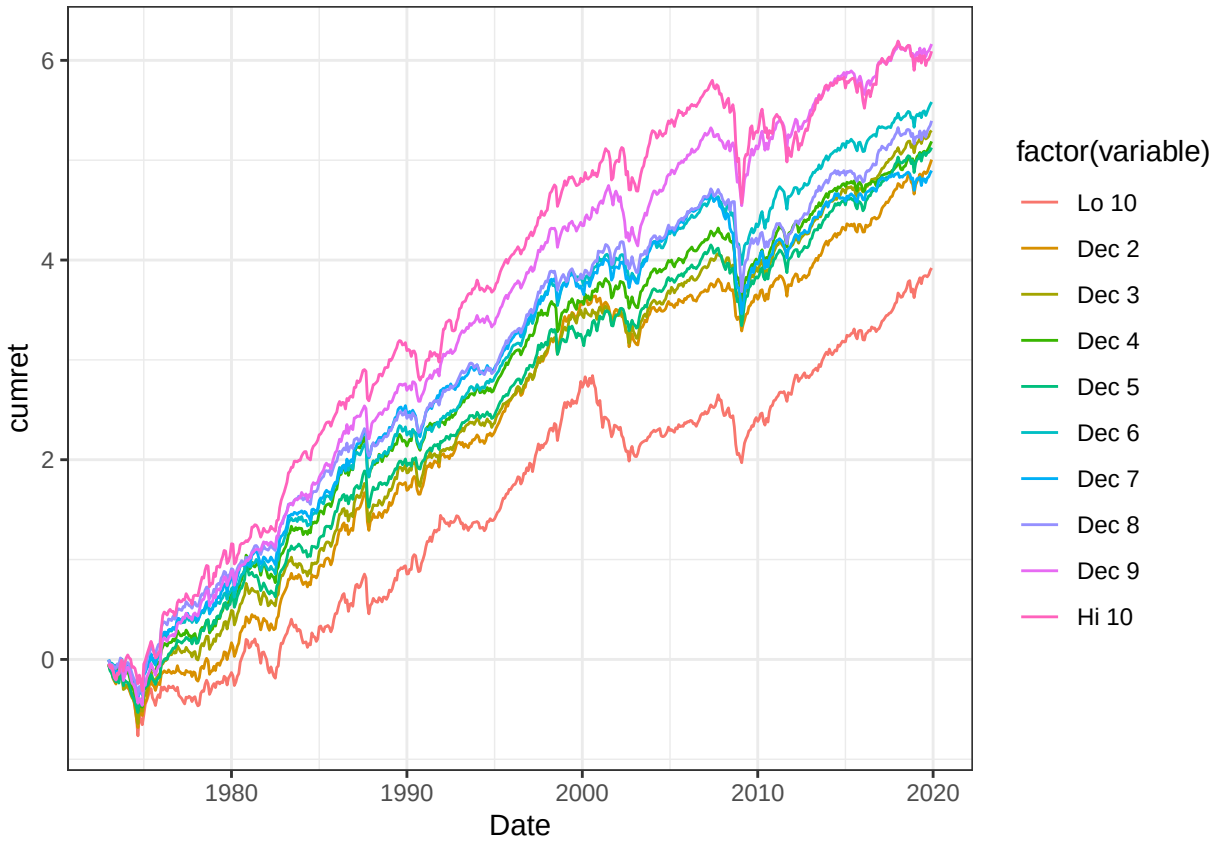
## QUESTION 6

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

We check the empirical data for size, B/M portfolios and SMB and HML factors using the data downloaded from Kenneth French's website.
The cumulative returns for size portfolios are shown below -

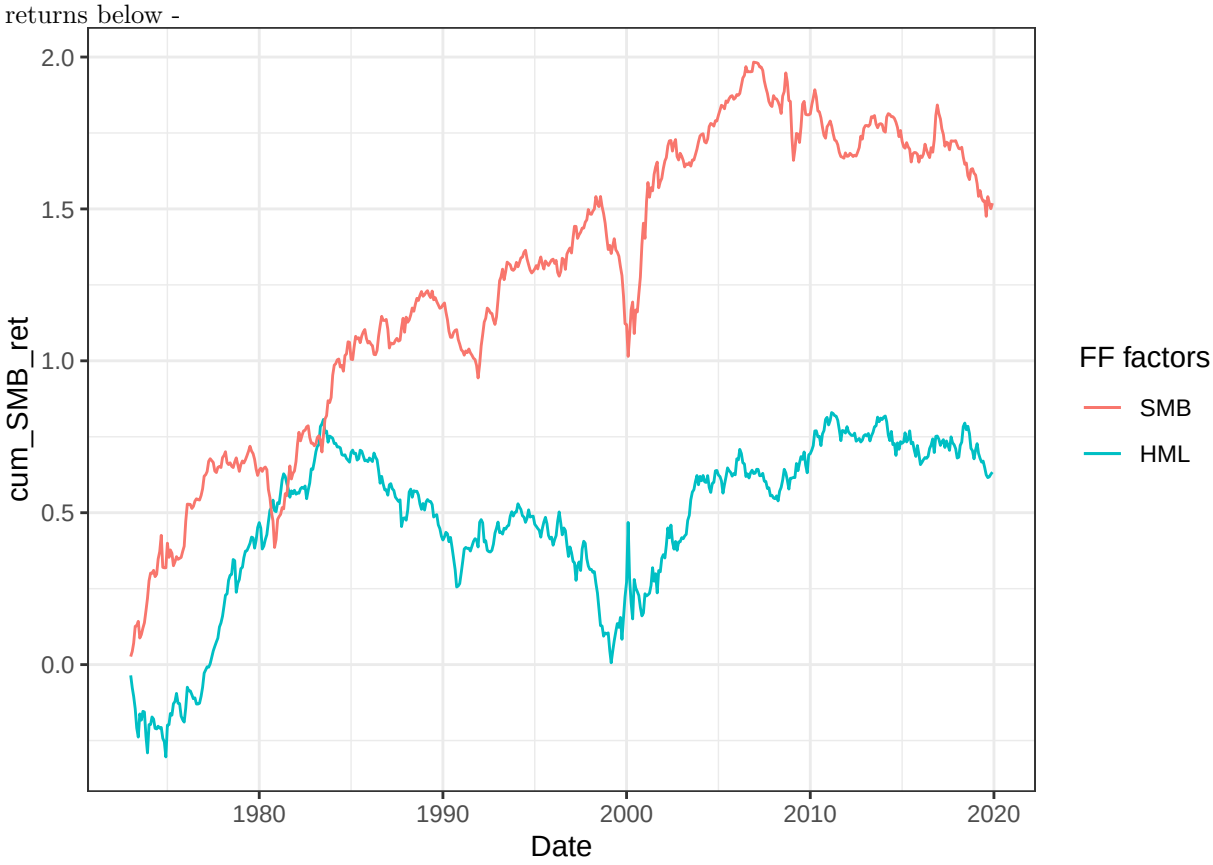



The cumulative returns for the B/M portfolios are shown below -




The size portfolios show the same return structure that we have shown before. But the BM portfolios show consistency in the Fama French's data with no reversal as we saw in my generated returns.

For the Fama French factor data downloaded from Kenneth French's website, we show the empirical cumulative returns below -




We see similar results as in the replicated data. We can see that HML factors hasn't provided significant cumulative returns over a long time, thus the concerns that value factor has died. Though SMB factor has hisotrically fared better, in the last 10 years, even SMB factor has shown negative cumulative returns.