# Problem Set 4

## 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 2017****.*

## Approach –

### Data Source

1. **CRSP Stock Data:** WRDS →CRSP > Annual Update > Stock / Security Files > CRSP Monthly Stock. Took data from Jan 1970 – Dec 2017 (few extra years to avoid data loss during shift command. Output would be exact as required from Jan 1973 to Dec 2017).
   Variables (PERMNO, share code, exchange code, RET, DLRET, PRC, share outstanding)

2. **Compustat financial Data:** WRDS →CRSP →Annual Update→Compustat Merged CRSP / Compustat Merged - Fundamentals Annual
   Screening Variable (Consolidation level =C, Industry Format =INDL (only non-financial companies), Data Format = std, Currency = USD)
   Variables (Fyear, SEQ, CEQ, PSTK, AT, LT, MIB, TXDITC, ITCB, TXDB, PSTKRV, PSTKL) for same data range

3. **Compustat Pension**: WRDS→Compustat - Capital IQ / Other Compustat/North America - Annual Updates / Compustat Annual Updates - Pension Annual
   Variable (gvkey, prba) for same date range

4. **Linking Table**: CRSP / Annual Update / CRSP / Compustat Merged/ CRSP/Compustat Merged Database - Linking Table

5. **Fama French ME (Size) Portfolio**: From FF website →Portfolios_Formed_on_ME

6. **Fama French Book to Market Portfolio**: From FF website → Portfolios_Formed_on_BE-ME

7. **Fama French 3 factor Portfolio**: From FF website → F-F_Research_Data_Factors.CSV

## Data cleaning steps

### ❖ Data cleaning steps for CRSP Stock data:

1. **Universe of stocks**: Following Fama French paper, 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).

2. **Missing returns**: Returns which are marked with characters(non-numeric) due to reasons viz. No valid comparison for an excess return, no listing information, no valid previous price, Off-exchange, Out of Range, no valid price has been marked as NA.

3. **Delisting return calculation**: All non-numeric delisting returns have been converted to NA

4. **Adjusted holding period Return:** Following method has been used to calculated adjusted holding period return from "holding period return" $r^h_{i,t}$ and delisting return $r^d_{i,t.}$

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

5. **Negative Price**: Negative price in CRSP data is bid-ask average. Absolute value of price has been taken.

6. **Lagged Market Capitalization**: price of stock has been multiplied with outstanding shares and it has been lagged by 1 month. If the same PERMCO (firm) has more than one securities in the same Year Month, their adjusted holding return is weighted average sum of their lag market cap.

7. **Year Month:** Date has been formatted as Year Month.

### ❖ Data cleaning steps – Link Table data:

1. **(Primary issue marker for the link) LINKPRIM**: P or C (primary) link data has been filtered.

2. **NA in LINKDT and LINKENDDT:** All NA in these columns has been replaces by first and last dates as per data range. is.na (LINKDT="1970-01-01", LINKENDDT="2017-12-31").

❖ **Data cleaning steps – Compustat Table data:**

1. **Unique:** Unique rows have been taken from the table as some of the rows were repeated with exactly same data in all the columns.

2. **Year Month:** Date has been formatted as Year Month. And also as %Y%m%d in separate columns.

## Data Merging

Compustat and Pension data has been merged using GVKEY and Date.
Compustat_Pension is then merged with Link Table using GVKEY and allow.cartesian = TRUE.

As a trick to directly get book equity data aligned to desired year, a July_Jun_year column has been created in CRSP. This is as required in Fama French paper for portfolio frequency which start with July every year and till Jun next year. Actual year number is not very important if those months are marked same for Jul t-Jun t+1. A lagged year is marked to get the lagged book equity data directly from Compustat.

CRSP Stock is now merged with Compustat_Pension_Link table using PERMNO/LPERMNO, PERMCO/LPERMCO and July_Jun_year/fyear as x and y key's and allow.cartesian = TRUE.

## Cleaning the final merged file

The final merged file has the following issues –

1. Multiple PERMCO in same Year Month:
2. One GVKEY is mapped to multiple PERMCO

In a Year Month, PERMCO should be just one. So, to solve this following approach has been taken in the order –

❖ If LINKTYPE is not LC and PERMCO is repeated – remove that entry

- If LINKPRIM is not 'P' and PERMCO is repeated – remove that entry
- If LIID is not 1 and PERMCO is repeated – remove that entry
- If LINKENDDT is NA and PERMCO is repeated – keeping the no NA entry
- If double entries are still left, keeping the entry, which has been there for longest using LINKENDDT (last date) – LINKDT (first date)
- Keeping the GVKEY that has been around for longest
- Keeping the smallest GVKEY if still duplicate entries
- Removing the entries with GVKEY as NA

- Filter based on date with in first date and last date i.e. date >= LINKDT & date <= LINKENDDT)

  As a final check this command has been run to check if the merged data is accurate -
  nrow(unique(final_DT, by=c('GVKEY', 'date'))) != nrow(final_DT) |
  nrow(unique(final_DT, by = c('PERMCO', 'date'))) != nrow(final_DT))

## Calculation of Book Equity

**Shareholder's Equity (SHE):** Following formula has been used to calculate shareholder's equity (as explained in the paper)-

```
[SHE := seq]
[is.na(SHE), SHE:= ceq + pstk, by = .(DATE, LPERMCO)]
[is.na(SHE), SHE := at - lt - mib, by = .(DATE, LPERMCO)]
[is.na(SHE), SHE := at - lt,  by = .(DATE, LPERMCO)
```

**Deferred taxes and investment tax credit (DT):** Following formula has been used to calculate DT (as explained in the paper)-

```
[DT := ifelse(!is.na(txditc), txditc, itcb+ txdb) ]
```

**Book value of preferred stock (PS):** Following formula has been used to calculate PS (as explained in the paper)-

```
[PS := ifelse(!is.na(pstkrv), pstkrv, ifelse(!is.na(pstkl), pstkl, pstk))]
```

**Book Equity(BE):**

*Book Equity (BE) = Shareholder's Equity(SHE) – Book Value of preferred stock (PS) + Deferred Taxes (DT) - Postretirement Benefit Asset (PRBA)*

**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.*

**Approach for size decile–**

1. **Data Selection:** If a stock doesn't have market equity in Jun t that stock will not be considered for calculation for July t to Jun t+1. This has been achieved by checking lagged market return in July t and keeping it same for July t to Jun t+1.

2. **Decile:** Decile has been created 1 to 10 based on NYSE stock size for each Year Month using lagged market return in July t (this is same as using Jun t actual market return (not lagged) as described in the paper). This decile will be same for July t to Jun t+1.

3. **Value Weighted Return:** Value weighted adjusted return has been calculated for each decile in each Year Month using weights of lagged market cap.

4. **Portfolio frequency:** Decile portfolio is adjusted every year July t for July t to next year Jun t+1.

| Size Decile Portfolio | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dec1 | Dec2 | Dec3 | Dec4 | Dec5 | Dec6 | Dec7 | Dec8 | Dec9 | Dec10 | D1-D10 |
| r-rf | 0.0912 | 0.0916 | 0.0983 | 0.0869 | 0.0882 | 0.0866 | 0.0861 | 0.0820 | 0.0756 | 0.0589 | 0.0323 |
| sigma | 21.8610 | 22.4307 | 20.8452 | 20.2645 | 19.4930 | 18.1328 | 18.1274 | 17.5285 | 16.2020 | 15.2080 | 16.4907 |
| SR | 0.4171 | 0.4085 | 0.4717 | 0.4287 | 0.4522 | 0.4774 | 0.4750 | 0.4676 | 0.4665 | 0.3874 | 0.1956 |
| skew | -0.6714 | -0.6725 | -0.9197 | -0.9301 | -0.9039 | -0.9170 | -0.9070 | -0.8192 | -0.7578 | -0.5883 | 0.8374 |
| Cor | 0.9979 | 0.9983 | 0.9983 | 0.9979 | 0.9977 | 0.9981 | 0.9982 | 0.9977 | 0.9989 | 0.9979 | 0.9943 |

Annualized Mean Excess Return i.e. mean (r – rf) = 12* mean (SizeRet – RF)
Annualized Sigma (standard deviation of return) = sqrt (12) * standard deviation (SizeRet)

Sharpe Ratio (SR) = Annualized Mean Excess Return / Annualized standard deviation
Sk(m) (skewness) = skewness(1+SizeRet)

As we can observe in the data, small size stock portfolio has higher return than the big market equity stocks portfolio. Sharp ratio is also showing a decreasing trend as the size grows. Correlation with FF ME size portfolio has been calculation. Though not exactly 1, it's very close to 1. These minor discrepancies can be attributed to methodologies to calculate annualized returns, actual assumptions taken in data merging and cleaning (which is not detailed in the paper).

**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.*

**Approach Book To Market decile–**

1. **Data Selection:** Only those stocks which have market equity in Dec t-1 and Jun t, book equity in year t-1and book equity is positive are selected for Book to market decile calculation.

2. **Book to Market Ratio:** Book to market ratio has been calculated using last year book equity and Dec t-1 market equity.

3. **Decile:** Decile has been created 1 to 10 based on NYSE stock book to market ratio for each Year Month in July. This decile will be same for July t to Jun t+1.

4. **Value Weighted Return:** Value weighted adjusted return has been calculated for each decile in each Year Month using weights of lagged market cap.

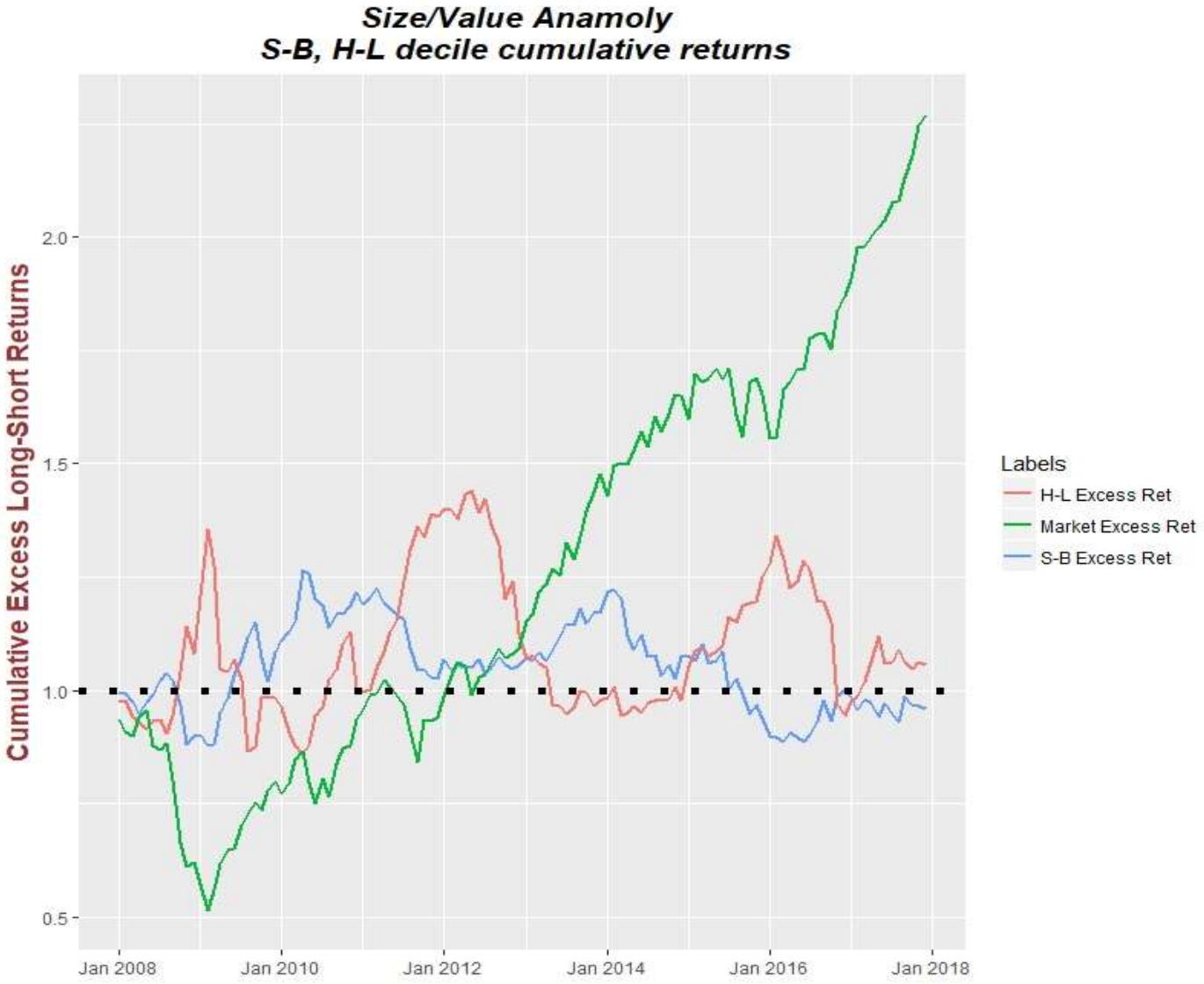
5. **Portfolio frequency:** Decile portfolio is adjusted every year July t for July t to next year Jun t+1.

| BE-ME Decile Portfolio | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Dec1** | **Dec2** | **Dec3** | **Dec4** | **Dec5** | **Dec6** | **Dec7** | **Dec8** | **Dec9** | **Dec10** | **D10-D01** |
| **r-rf** | 0.0525 | 0.0755 | 0.0703 | 0.0818 | 0.0787 | 0.0831 | 0.0860 | 0.0861 | 0.0911 | 0.1201 | 0.0676 |
| **sigma** | 17.5834 | 16.4266 | 16.2977 | 16.4842 | 16.0244 | 15.7318 | 15.5097 | 16.6597 | 16.4897 | 20.2598 | 16.2171 |
| **SR** | 0.2984 | 0.4596 | 0.4316 | 0.4960 | 0.4910 | 0.5281 | 0.5546 | 0.5166 | 0.5527 | 0.5927 | 0.4168 |
| **skew** | -0.4742 | -0.7652 | -0.8113 | -0.9264 | -0.8549 | -0.6918 | -0.5330 | -0.8584 | -0.7560 | -0.7608 | 0.2683 |
| **Cor** | 0.9969 | 0.9872 | 0.9779 | 0.9801 | 0.9697 | 0.9684 | 0.9657 | 0.9553 | 0.9489 | 0.9665 | 0.9414 |

We can observe in the data; mean access return is increasing as we go from Decile 1(low book to market) to Decile 10 (high book to market) i.e. Growth to Value stock. Sharp ratio is also showing an increase as we go from growth to value. The correlation with the Fama French BE ME portfolio is quite high and close to 1. Some discrepancy can be attributed to method of calculating annual returns and data merging and cleaning assumptions.

## Question 4

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




## Small – Big return (Long decile 1 ME size and short decile 10 ME size)

Graph has been plotted to show cumulative return since Jan 2008 if invested in market (excess market return), small – big portfolio (long small stock decile 1 and short big stock decile 10) and long growth stock and short value stock.

As shown in graph, Small – Big (Decile 1- Decile 10), blue line, has given positive return from mid of 2009 (post financial crisis) to mid of year 2012. It was better than market return that time. Post 2012 Small-Big has still given positive return till mid of 2015 though market return has been higher than Small – Big.

So, we can say Small – Big worked during mid of 2010 to mid of 2015. Post mid of 2015 onward it has given negative returns.

**High B/E – Low B/E return (Long decile 10 B/E and short decile 1 B/E )**

From the plotted cumulative return since 2008 in the above graph (red line), H-L characteristic has given positive return during year 2011 till early 2013. This time it has beaten the market excess return. It has also given positive return during 2015 to 2016 but it was less than market return. In the middle of 2017 ($2^{nd}$ and $3^{rd}$ quarter) is has given positive return but not beating the market.

## 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.*

**Approach –**

**SMB (S, B):** – Stock portfolio has been segregated in Small and Big based on median market equity in Jun for NYSE stocks.

**HML (H, M, L):** Stocks portfolio has been segregated in low, medium and high book to market value quantiles based on book to market ratio (book value of last year(t-1) and market value of Dec t-1) in July t based on NYSE stocks only.

**Portfolio frequency:** Portfolio has been rebalanced every July for July t to Jun t+1.

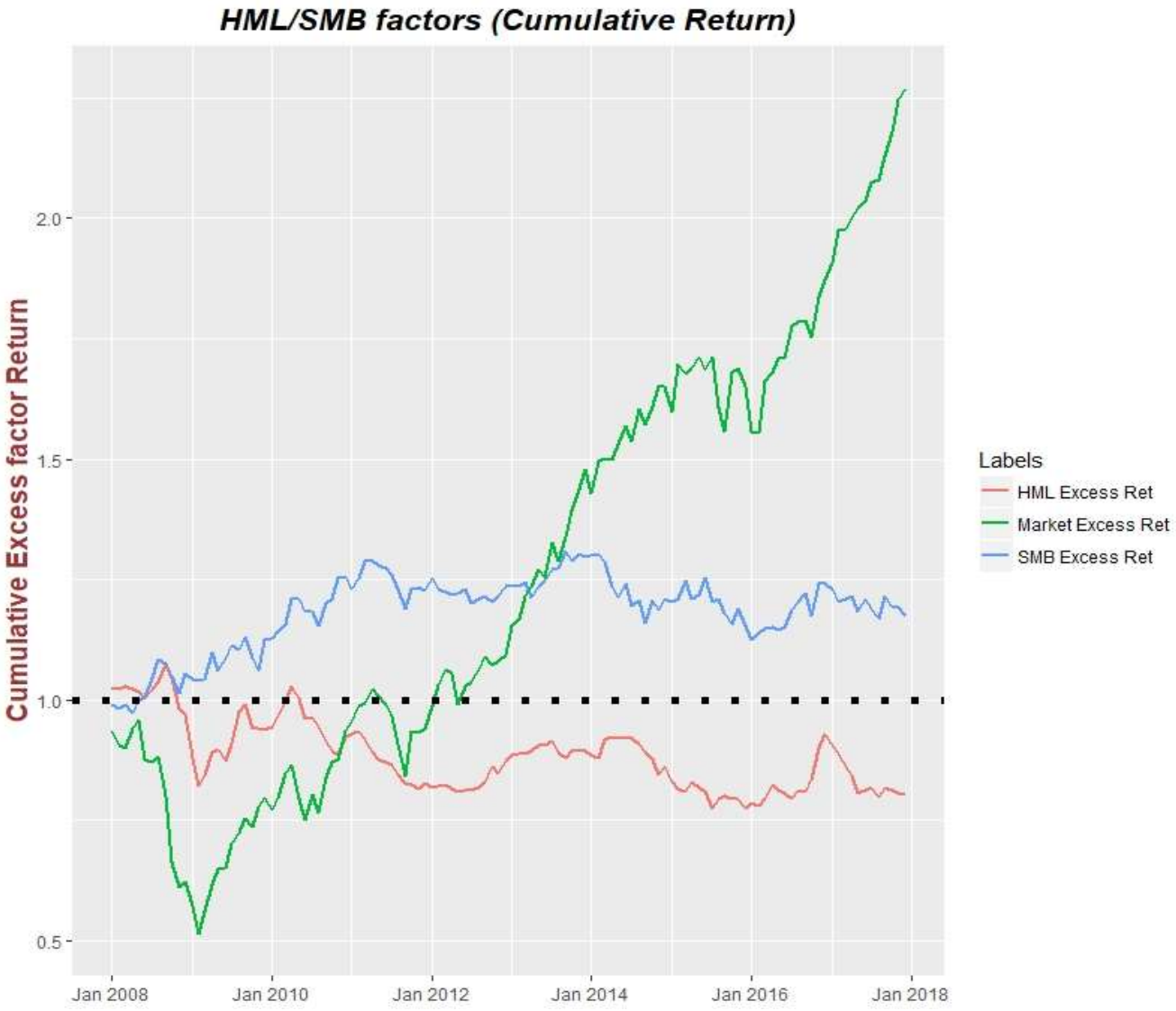
**Value weighted returns**: have been calculated for each double sorted quantiles SL, SM, SH, BL, BM, BH.

**Factor Returns:** Now this double sorted portfolio with quantile – SL, SM, SH, BL, BM, BH has been used to calculate SMB/HML factors returns –

[SMB_Ret:= (1/3)*(SL+SM+SH) - (1/3)*(BL+BM+BH)]
[HML_Ret:= (1/2)*(SH + BH) - (1/2)*(SL + BL)]

| HML/SMB Factor Returns | | |
|---|---|---|
| | HML | SMB |
| r-rf | -0.00863 | -0.02412 |
| sigma | 2.94724 | 3.04732 |
| SR | -0.29269 | -0.79152 |
| skewness | 0.01308 | 0.63452 |
| Correlation | 0.97318 | 0.99237 |

I have calculated the annualized mean excess return for HML and SMB factors.
Also reported the correlation with Fama French 3 factor portfolio. Correlation is quite close to 1.



As obvious from the cumulative return plot since Jan 2008, SMB Factor has given positive return throughout. But it has beaten the market up until middle of 2013, post which it's return is positive but quite less than market.
HML factor has not given cumulative positive return in recent time. Which means, HML factor has not been consistently positive for significant period.

**Question 6**

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

Characteristic portfolio (FF-92) focusses on different returns on small and big stocks by market equity and shows that small stock exhibit higher return than big ME stocks. But it does not mimic the factor which can explain the time series returns for the cross section of stocks.
SMB factor in FF-93 which is formed with difference between simple average of three small stocks (S/L, S/M, S/H) and simple average of three big stocks (B/L, B/M, B/H) mimics the factor. So, it removes the effect of BE/ME with this averaging and purely reflects the size effect.

Characteristic portfolio (FF-92) shows book-to-market equity is related to relative profitability i.e. low BE/ME firms have persistently high earning while high BE/ME firms have persistently poor earnings. In the factor portfolio (FF-93), it's shown that HML, the difference between the returns on high and low BE/ME stocks by weighing the size effect equally as difference between simple average of (S/H, B/H) and simple average (S/L, B/L). This way it mimics the pure value-growth factor without effect of size.

**References-**


- Wharton Research Data Services (WRDS) CRSP data taken on May 12
- http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (FF -3 factor file/ME Size/ BE ME)
- TA help code