Problem Set 4

Size and Value

Name: Xiangui (Jasmine) Mei

Names of whom I discussed this problem set with: Jiaqi (Johnny) Li, Shuyu(Mia) Dong

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.

Input:

CRSP_Stocks from 1926 to 2018.

Fundamental Annual Updates for North America (CompuStat)

Pension Annual (CompuStat)

Link Table (CRSP CompuStat merged)

To insure the return and decile range from1973 to 2018, I downloaded data from 1960-2019.

Calculation Process:

1.Data Cleanup:

CRSP Cleanup:

i) First, I cleaned the CRSP data) The universe of stocks starts with all firms listed on NYSE, Amex, or Nasdaq (with CRSP exchange code of 1, 2 or 3) for the formation date, moreover, we required using only returns of common shares (with CSRP share code of 10 or 11).

ii) Then, I set all returns where returns equal to -99.0, -88.0, -77.0, -66.0, B, C to NA;

for delisting returns, I set all delisting returns where delisting returns equal to -99.0, -88.0,-77.0, -66.0, P, S, T to NA.

iii) I computed MarketValue for each stocks (PERMNO) by using absolute number of price*shares of outstanding/1000, after doing this, I derived the lagged MV by shift it by one period.

iv) Then, for each firm, year and month combination, I calculated its lagged Market Value, by summing up lagged MV derived from iii).

CompuStat Cleanup:

i) First, I used SIC code from CompuStat as a selection criteria to get non-financial firms (SIC > 6199 or SIC < 6190)

ii) Then, I merged the Computstat and Pension Annual data using the gvkey field which is common between both tables.

iii) To find necessary accounting variables to define BE (Book Value), I followed instruction on CCLE. (i.e. details is in the homework description)

SHE (Shareholders' equity) = SEQ | CEQ + PSTK | AT - LT- MIB | AT - LT

DT (Deferred taxes and investment tax credit) = TXDITC | ITCB+TXDB | ITCB | TXDB

PS (Book value of preferred stock) = PSTKRV | PSTKL | PSTK

BE = SHE – PS + DT – PRBA

PRBA comes from the pension dataset which we have merged earlier on.

Since SHE is necessary to compute BE, then, If SHE is not available, assign BE to NA, otherwise, include other variables in the not missing.

2. Data Merge

Following TA's instruction,

i) First, I derived linkable by using dbConnect, I acquired variables including: gvkey, linkprim, liid, linktype, lpermno, lpermco, USEDFLAG, linkdt, linkenddt and linktype.

i) Then, I merged CompuStat with linkable by gvkey, variables in linkable linkdt and linkenddt are the start and end dates during which the PERMCO is available, so I only selected the PERMCO whose date is within the start and end date range, and named the merged dataset "merged".

ii) In the next step, I merged CRSP and "merged", since CRSP is monthly updated and CompuStats is annually updated, I used by.x(CRSP) = True, which means I fit in CompuStats data into CRSP.

iii) Then I cleaned up the merged data to solve the problem that there might be multiple gvkeys per PERMCO.

step1: for linktype, if there are some types on PERMCO and the linktype is not LC linktype, only keep LC

step2: for linkprim, if there are some types on PERMCO and the linkprim is not P, only keep P

step3: for liid, if there are some types on PERMCO and the liid is not 1, only keep 1

step4: for linkenddt, only use the link that is current

step5: compute the difference between the start and end date of the link, use the link has been around the longest

step6: use the gvkey that has been around the longest

step7: use the smaller gvkey if there are multiple with same PERMCO on the same year and month combination.

Output

```
> output1
  PERMCO year       date PERMNO EXCHCD gvkey      CRET       MV  MKT_Cap    SHE     at DT PS
1      3 1977 1977-08-31  12160      3  2767  0.038961 7.125000 6.857812 18.341 85.893  0  0
2      3 1977 1977-09-30  12160      3  2767  0.000000 7.125000 7.125000 18.341 85.893  0  0
3      3 1977 1977-10-31  12160      3  2767 -0.062500 6.679688 7.125000 18.341 85.893  0  0
4      3 1977 1977-11-30  12160      3  2767  0.040000 6.946875 6.679688 18.341 85.893  0  0
5      3 1977 1977-12-30  12160      3  2767  0.000000 6.946875 6.946875 18.341 85.893  0  0
6      3 1978 1978-01-31  12160      3  2767  0.051282 7.303125 6.946875 19.768 92.555  0  0
      BE linktype linkprim liid     linkdt  linkenddt month     BEME BM_decile keep_BM
1 18.341       LC        C  00X 1977-01-27 1979-01-30     8 2.674468        10       1
2 18.341       LC        C  00X 1977-01-27 1979-01-30     9 2.574175        10       1
3 18.341       LC        C  00X 1977-01-27 1979-01-30    10 2.574175        10       1
4 18.341       LC        C  00X 1977-01-27 1979-01-30    11 2.745787        10       1
5 18.341       LC        C  00X 1977-01-27 1979-01-30    12 2.640180        10       1
6 19.768       LC        C  00X 1977-01-27 1979-01-30     1 2.845596        10       1
```

2. Portfolio Construction:

Size Portfolio

i) to form the size portfolio, I used market equity at the end of December of year t-1 to compute its book to market for t-1, and use its market equity for June of year t to measure its size.

ii) all companies are allocated a rank between 1 to 10 from July of year t to July of year t+1, based on breakpoints calculated using the ME in June of year t of only NYSE stocks. The rebalancing is done at the end of June annually.

iii) after get deciles for each firm, year and month combination, I calculated value-weighted returns for all companies within each decile, here I used lagged market cap for that stock.

iv) then I subtracted the size portfolio outside the full dataset, by using "dcast" function, I transformed it into a matrix. The long-short portfolio is constructed by long the 1st decile and short the 10th decile for every year combination.

v) the data is constrained between 1973 and 2017.

Book to Value Portfolio

i) the BE / ME ratio is computed using the book equity from December of fiscal year t-1 and market equity (market cap) from December of year t-1.

ii) the BE / ME value of all NYSE stocks is used to get BM decile points at end of fiscal year t-1.

iii) the rankings are rebalanced every year at the end of June, using data from year t-1. So deciles of each stock from year t-1 is used for returns from July of year t to June of year t+1 for that stock.

iv) after get deciles for each firm, year and month combination, I calculated value-weighted returns for all companies within each decile, here I used lagged market cap for that stock.

v) then I subtracted the size portfolio outside the full dataset, by using "dcast" function, I transformed it into a matrix. The long-short portfolio is constructed by long the 1st decile and short the 10th decile for every year combination.

v) the data is constrained between 1973 and 2017.

SMB and HML Portfolio

Following the Fama French website, portfolio is sorted by ME *(Market Cap)* and BM *(Book to Market Ratio)*. I followed these steps to achieve the result.

i) For ME, I used annual rebalancing. When calculating for year t, I divided stocks into two groups, the breakpoints is the median of ME for stocks in year t-1. The rank calculated using data from t-1 year, is the rank for returns from July year t+1 to June year t.

ii) For BM, I used annual rebalancing. When calculating for year t, I divided the stocks into three groups based on 30%, 70% and 100% breakpoints of Book to Market Ratio values of NYSE stocks in year t-1. The rank calculated using data from t-1 year, is the rank for returns from July year t+1 to June year t.

iii) Then with quantiles cut by ME and BM, I formed six portfolios:

Small in ME - Low in BM; Small in ME-Medium in BM;

Small in ME-High in BMB; Big in ME – Low in BM;

Big in ME-Medium in BM; Big in ME-High in BMI

Then I marked them with their corresponding ranks (from 1 to 6).

for each of these six portfolios, value weighted return could be calculated for each year and month combination specifically.

iv) In the last step, I formed HML and SMB portfolios.

SMB is the difference between the simple average of all small firms (three portfolios) and all big firms (three portfolios).

$$SMB = \frac{1}{3} * (small\ high + small\ medium + small\ low) - \frac{1}{3} * (big\ high + big\ medium + big\ low)$$

HML is the difference between the simple average of all value firms (two portfolios) and all growth firms (two portfolios).

$$HML = \frac{1}{2} * (small\ high + big\ high) - \frac{1}{2} * (small\ low + big\ low)$$

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.

Input:

the output of Question 1

Calculation Process:

For size portfolio,

i) Mean: mean is calculated for the returns of every year, month and decile combination and then is annualized on an arithmetic basis by *12, then I time the result by 100 to make it into percentage term.

ii) Standard Deviation: Standard deviations of for the returns of every year, month and decile combination and then is annualized on an arithmetic basis by *sqrt (12), then I time the result by 100 to make it into percentage term.

iii) Sharpe Ratio: Sharpe ratio for different are calculated as: excess return/sd

iv) Skewness**:** With the build in function moments, I calculate excess skewness of log returns (log(1+return)), of every year, month and decile combination.

v) Correlation: I downloaded the Fama-French data sorted by single factors: size. To make the correlation calculation process easier, I transformed the data set form to matrix (number of rows=number of combinations for every year and month, number of columns= number of combinations for every year and month). Then I compared correlation for each decile.

Output is below:

```
> output2
        Decile1 Decile2 Decile3 Decile4 Decile5 Decile6 Decile7 Decile8 Decile9 Decile10 Long-Short
Exc_Ret  0.0789  0.0791  0.0711  0.0783  0.0633  0.0747  0.0792  0.0812  0.0753   0.0757     0.0254
sd       0.2093  0.1977  0.2012  0.2027  0.1611  0.1972  0.1583  0.1699  0.1663   0.2090     0.1658
SR       0.3613  0.3774  0.3610  0.4022  0.3975  0.4326  0.4219  0.4478  0.3782   0.3826     0.1779
Skew    -0.0937 -0.2475 -0.5711 -0.5124 -0.4082 -0.3726 -0.4362 -0.3376 -0.0543  -0.0654     0.9544
Cor      0.9672  0.9912  0.9866  0.9773  0.9674  0.9541  0.9531  0.9674  0.9632   0.9799     0.9904
```

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.

Input:

the output of Question 1

Calculation Process:

For Book to Market portfolio,

i) Mean: mean is calculated for the returns of every year, month and decile combination and then is annualized on an arithmetic basis by *12, then I time the result by 100 to make it into percentage term.
ii) Standard Deviation: Standard deviations of for the returns of every year, month and decile combination and then is annualized on an arithmetic basis by *sqrt (12), then I time the result by 100 to make it into percentage term.
iii) Sharpe Ratio: Sharpe ratio for different are calculated as: excess return/sd

iv) Skewness**:** With the build in function moments, I calculate excess skewness of log returns (log(1+return)), of every year, month and decile combination.

v) Correlation: I downloaded the Fama-French data sorted by single factors: book to market. To make the correlation calculation process easier, I transformed the data set form to matrix (number of rows=number of combinations for every year and month, number of columns= number of combinations for every year and month). Then I compared correlation for each decile.

Output is below:

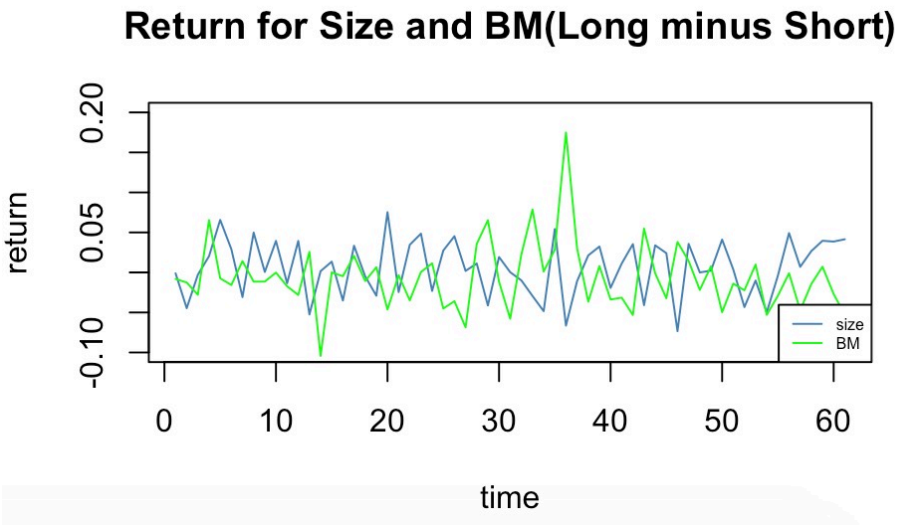
```
> output3
        Decile1 Decile2 Decile3 Decile4 Decile5 Decile6 Decile7 Decile8 Decile9 Decile10 Long-Short
Exc_Ret  0.0485  0.0582  0.0711  0.0783  0.0653  0.0725  0.0712  0.0622  0.1253   0.1157     0.0667
sd       0.1893  0.1677  0.1642  0.1627  0.1611  0.1578  0.1581  0.1624  0.1620   0.2070     0.1634
SR       0.2573  0.3792  0.4629  0.4638  0.3947  0.4782  0.4972  0.5025  0.5782   0.5628     0.4123
Skew    -0.2237 -0.4475 -0.5729 -0.5524 -0.4572 -0.5526 -0.4927 -0.1264 -0.0575  -0.0654     0.5848
Cor      0.9912  0.9843  0.9802  0.9711  0.9718  0.9722  0.9488  0.9457  0.9673   0.9712     0.9323
```

Question 4

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

As we could find from the graph for past 5 years (2013-2018), the return for book to market portfolio is higher than size portfolio, and it is more volatile than size portfolio.

Also, as we could see from output2 and output3, the annual return for long-short portfolio for both size and Book-to-Market portfolio is pretty low (0.0254 for size and 0.0667 for BM), they are even lower than average market return (we derived in the first homework about 11%). This statistic denotes that the size and book to market portfolio don't work very well for current years.



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.

Input:

the output of Question 1

Calculation Process:

For HML and SMB,

i) Mean: mean is calculated for the returns of every year, month and decile combination and then is annualized on an arithmetic basis by *12, then I time the result by 100 to make it into percentage term.

ii) Standard Deviation: Standard deviations of for the returns of every year, month and decile combination and then is annualized on an arithmetic basis by *sqrt (12), then I time the result by 100

to make it into percentage term.
iii) Sharpe Ratio: Sharpe ratio for different are calculated as: excess return/sd

iv) Skewness**:** With the build in function moments, I calculate excess skewness of log returns (log(1+return)), of every year, month and decile combination.

v) Correlation: I downloaded the Fama-French data contained SMB and HML. To make the correlation calculation process easier, I transformed the data set form to matrix (number of rows=number of combinations for every year and month, number of columns= number of combinations for every year and month).

Output:

```
> output5
              HML      SMB
Exc_Ret  0.03782 0.02152
sd       0.13610 0.10380
SR       0.37590 0.19660
Skew    -0.22510 0.63230
Cor      0.95630 0.97190
```

Question 6

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

characteristic portfolios (Fama and French 1992)

The characteristic portfolios concentrate on showing that the Size and Book-to-Market provide an easy and powerful characterization of the cross-section of average stock returns. The paper proves that Size and Book-to-Market are able to capture the cross-sectional variation in average stock returns. The portfolio is based on the Fama Macbeth regression, which regress average returns on variables that are used to explain the returns. Result indicates that: 1) beta doesn't help to explain average stock returns cross-sectionally. 2) leverage is playing roles in E/P

factor portfolios (Fama and French 1993)

The factor portfolios are trying to make the 2 characteristics:

Size , Book-to-Market, and term structure variables for bonds

In this approach, if the assets are priced correctly, some variables that could explain some level of returns must also be able to explain the sensitivity of common risk factors. The paper proves that the mimicking factors of Size and Book-to-Market characteristics capture common variation in stock returns. In addition, the mimicking factors alone cannot explain the average stock returns based on the time-series regression, and the market factor could explain the difference between the average stock returns and average T-bill returns.