Yifu-He-Homework6

May 13, 2020

```
In [1]: import pandas as pd
        import numpy as np
        import os
```

1 Question 1

```
In [4]: path = "/Users/yifuhe/Learning/RBS/fin_model/homework6"
        HXZ_factor = pd.read_excel(path+"/Data_q_factors.xlsx")
        five_factor = pd.read_excel(path+"/F-F_Research_Data_5_Factors_2x3.xlsx")
        fama_french = pd.read_excel(path+"/F-F_Research_Data_Factors.xlsx")
        # read data from path
        port_25 = pd.read_excel(path+"/25_Portfolios_5x5.xlsx")
        # 25 Fama-French size and book to market sorted portfolios
        colname = port_25.iloc[14,:].values
        colname[0] = "Date"
        data = port_25.iloc[15:-2,:].reset_index().drop("index",axis = 1)
        data.columns = colname
        for i in range(len(data["Date"])):
            if not isinstance(data["Date"][i],(int,float)):
                print(i,data.Date[i])
        #data["Date"] = pd.to_datetime(data.loc["Date"].apply(str),format = "%Y%m")
```

```
1126   Average Equal Weighted Returns -- Monthly
2254   Average Value Weighted Returns -- Annual
2351   Average Equal Weighted Returns -- Annual
2448   Number of Firms in Portfolios
3576   Average Market Cap
4704   For portfolios formed in June of year t
4705   Value Weight Average of BE/ME Calculated for June of t to June of t+1 as:
4706   Sum[ME(Mth) * BE(Fiscal Year t-1) / ME(Dec t-1)] / Sum[ME(Mth)]
4707   Where Mth is a month from June of t to June of t+1
4708   and BE(Fiscal Year t-1) is adjusted for net stock issuance to Dec t-1
5836   For portfolios formed in June of year t
5837   Value Weight Average of BE_FYt-1/ME_June t Calculated for June of t to June of t+1 as:
5838   Sum[ME(Mth) * BE(Fiscal Year t-1) / ME(Jun t)] / Sum[ME(Mth)]
5839   Where Mth is a month from June of t to June of t+1
5840   and BE(Fiscal Year t-1) is adjusted for net stock issuance to Jun t
```

```
6968   For portfolios formed in June of year t
6969   Value Weight Average of OP Calculated as:
6970   Sum[ME(Mth) * OP(fiscal year t-1) / BE(fiscal year t-1)] / Sum[ME(Mth)]
6971   Where Mth is a month from June of t to June of t+1
7655   For portfolios formed in June of year t
7656   Value Weight Average of investment (rate of growth of assets) Calculated as:
7657   Sum[ME(Mth) * Log(ASSET(t-1) / ASSET(t-2) / Sum[ME(Mth)]
7658   Where Mth is a month from June of t to June of t+1
```

In [7]:
```python
industry_30 = pd.read_excel(path+"/30_Industry_Portfolios.xlsx")
for i in range(len(industry_30.iloc[:,0])):
    if not isinstance(industry_30.iloc[i,0],(int,float)):
        print(i,industry_30.iloc[i,0])

colname = industry_30.iloc[10,:].values
colname[0] = "Date"
AVWR_month = industry_30.iloc[11:1135].reset_index().drop("index",axis=1)
AVWR_month.columns = colname
AVWR_month["Date"] = pd.to_datetime(AVWR_month["Date"],format = "%Y%m")
AVWR_month.head()
```

```
0 It contains value- and equal-weighted returns for 30 industry portfolios.
2 The portfolios are constructed at the end of June.
4 The annual returns are from January to December.
6 Missing data are indicated by -99.99 or -999.
9   Average Value Weighted Returns -- Monthly
1137   Average Equal Weighted Returns -- Monthly
2265   Average Value Weighted Returns -- Annual
2362   Average Equal Weighted Returns -- Annual
2459   Number of Firms in Portfolios
3587   Average Firm Size
4715   Sum of BE / Sum of ME
4813   Value-Weighted Average of BE/ME
4910 Copyright 2020 Kenneth R. French
```

Out[7]:
```
        Date  Food   Beer  Smoke Games  Books Hshld Clths Hlth  Chems  ...  \
0 1926-07-01  0.56  -5.19  1.29  2.93  10.97 -0.48  8.08  1.77  8.14  ...
1 1926-08-01  2.59  27.03   6.5  0.55  10.01 -3.58 -2.51  4.25   5.5  ...
2 1926-09-01  1.16   4.02  1.26  6.58  -0.99  0.73 -0.51  0.69  5.33  ...
3 1926-10-01 -3.06  -3.31  1.06 -4.76   9.47 -4.68  0.12 -0.57 -4.76  ...
4 1926-11-01  6.35   7.29  4.55  1.66   -5.8 -0.54  1.87  5.42   5.2  ...

  Telcm Servs BusEq Paper Trans  Whlsl Rtail Meals Fin   Other
0  0.83  9.22  2.06   7.7  1.93 -23.79  0.07  1.87  0.37   5.2
1  2.17  2.02  4.39 -2.38  4.88   5.39 -0.75 -0.13  4.46  6.76
2  2.41  2.25  0.19 -5.54  0.05  -7.87  0.25 -0.56 -1.23 -3.86
```

```
3 -0.11    -2 -1.09 -5.08 -2.64 -15.38  -2.2 -4.11 -5.16 -8.49
4  1.63  3.77  3.64  3.84   1.6   4.67  6.52  4.33  2.24     4

[5 rows x 31 columns]
```

In [8]:
```python
def get_portfolio_pool(data, J, decile=3):
    res = []
    cumulate = data.rolling(J).sum().dropna()
    slices = [i for i in range(decile)] + [i for i in range(-decile,0)]
    index = cumulate.index
    for i in index:
        sorted_slice = cumulate.sort_values(by = i ,axis = 1, ascending = False)
        res.append(sorted_slice.columns[slices])
    colname = ["long_{}".format(i) for i in range(decile)] + ["short_{}".format(i) for
    return pd.DataFrame(res,columns=colname,index=index)

portfolio_pool_6 = get_portfolio_pool(AVWR_month,6,3)
portfolio_pool_9 = get_portfolio_pool(AVWR_month,9,3)
portfolio_pool_6.head()
```

Out[8]:

	long_0	long_1	long_2	short_3	short_2	short_1
5	Autos	Beer	Chems	Hshld	Paper	Whlsl
6	Beer	Books	Txtls	Other	Paper	Whlsl
7	Servs	Chems	Txtls	Paper	Other	Whlsl
8	Chems	Mines	BusEq	Other	Paper	Whlsl
9	Chems	Mines	Autos	Paper	Oil	Whlsl

In [93]:
```python
def long_short(month_return, portfolio_pool, K, skip=1):
    index, end = portfolio_pool.index[1+skip:], portfolio_pool.index[-1]
    raw_df = pd.DataFrame(np.zeros((len(index),3)),index=index,columns=["long","short'
    for i in index:
        hold_index = list(range(i,min(end,i+K)))
        for i in hold_index:
            long = portfolio_pool.loc[i,["long_0","long_1","long_2"]].values
            short = portfolio_pool.loc[i,["short_3","short_2","short_1"]].values
            #print(company_columns)
            raw_df.loc[i,"long"] += month_return.loc[i,long].sum()
            raw_df.loc[i,"short"] += month_return.loc[i,short].sum()
    raw_df["long"] /= 18
    raw_df["short"] /= 18
    raw_df["arbitrage"] = raw_df["long"] - raw_df["short"]
    return raw_df

p66 = long_short(AVWR_month,portfolio_pool_6,6)
p69 = long_short(AVWR_month,portfolio_pool_6,9)
p96 = long_short(AVWR_month,portfolio_pool_9,6)
p99 = long_short(AVWR_month,portfolio_pool_9,9)
```

Because of the t-value is greater than the 95% tscore, all the excess return are significant.

```
In [97]: print(f"df: {p66.shape[0]}")
         print(f"95%t-score: {1.646}")
         print(f't66: {p66["arbitrage"].mean()/p66["arbitrage"].std()*np.sqrt(p66.shape[0])}')
         print(f't69: {p69["arbitrage"].mean()/p69["arbitrage"].std()*np.sqrt(p66.shape[0])}')
         print(f't96: {p96["arbitrage"].mean()/p96["arbitrage"].std()*np.sqrt(p66.shape[0])}')
         print(f't99: {p99["arbitrage"].mean()/p99["arbitrage"].std()*np.sqrt(p66.shape[0])}')
```

```
df: 1117
95%t-score: 1.646
t66: 30.830085692495242
t69: 30.7784947528333
t96: 24.879829686348785
t99: 24.847067914443098
```

```
In [133]: # get rf
          temp_data = pd.read_excel(path+"/F-F_Research_Data_Factors.xlsx")
          rf = temp_data.iloc[3:1127,4].mean()
          Rm = temp_data.iloc[3:1127,2]
          rf *=10
          print(f"rf: {rf}")

          # compute sharpe ratio
          print(f'sharpe ratio 66: {(p66["long"].mean() - rf)/p66["long"].std()}')
          print(f'sharpe ratio 69: {(p69["long"].mean() - rf)/p69["long"].std()}')
          print(f'sharpe ratio 96: {(p96["long"].mean() - rf)/p96["long"].std()}')
          print(f'sharpe ratio 99: {(p99["long"].mean() - rf)/p99["long"].std()}')
          print(f'market portfolio sharpe ratio: {Rm.mean()/Rm.std()}')
          print("All the sharpe ratio of long portfolio are higher than market portfolio")
```

```
rf: 2.7283807829181446
sharpe ratio 66: 0.15099621770830926
sharpe ratio 69: 0.2752308516025211
sharpe ratio 96: 0.0839389006728528
sharpe ratio 99: 0.21009221254283436
market portfolio sharpe ratio: 0.061386904331229016
All the sharpe ratio of long portfolio are higher than market portfolio
```

2 Question 2

```
In [257]: # import model
          from sklearn.linear_model import LinearRegression
          import statsmodels.api as sm
          from scipy import stats

          # load data
```

```
factor_3 = pd.read_excel(path+"/F-F_Research_Data_Factors.xlsx")
name = [factor_3.iloc[2,1]]
CAPM = pd.DataFrame(factor_3.iloc[3:1127,1].values,columns=name)

name = factor_3.iloc[2,:].to_list()
F3 = pd.DataFrame(factor_3.iloc[3:1127,:].values,columns=name)
factor_3.columns = name


factor_5 = pd.read_excel(path+"/F-F_Research_Data_5_Factors_2x3.xlsx")
name = factor_5.iloc[2,:].to_list()
F5 = pd.DataFrame(factor_5.iloc[3:683,:].values,columns=name)

Q = pd.read_excel(path+"/Data_q_factors.xlsx")
Q["Rm-Rf"] = Q["R_MKT"]-Q["R_F"]
name = ["Rm-Rf","R_ME","R_IA","R_ROE",]
Q = pd.DataFrame(Q.loc[2:635,name].values,columns=name)
Q
```

Out[257]:

```
      Rm-Rf    R_ME    R_IA   R_ROE
0    3.5998  1.9517 -1.6933  1.8876
1    3.4921 -0.7446 -2.9519  1.0983
2   -4.5545  2.9132  2.4686  0.5234
3    2.1555  6.2350 -2.1700  0.2945
4    4.2883  3.0119  2.3713 -0.7125
..      ...     ...     ...     ...
629 -2.4087 -4.2469 -0.9351  1.7673
630  1.2450  1.3935  3.0078  1.6828
631  1.6203  0.8401 -0.5515  0.6286
632  3.2577 -0.0261 -1.0127 -1.1667
633  2.5586  1.2822  1.8360 -1.3200

[634 rows x 4 columns]
```

2.1 66

Almost all the excess return are sinificant according to the p-value

In [328]:

```
# imports
from sklearn.linear_model import LinearRegression
import statsmodels.api as sm
from scipy import stats


# 66 CAPM
length = np.min([p66.shape[0],CAPM.shape[0]])
y = p66["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = CAPM.iloc[-length:,:].reset_index().drop("index",axis=1)
```

```
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 66 CAPM")
print(lm1.summary())


# 66 F3
length = np.min([p66.shape[0],F3.shape[0]])
y = p66["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F3.iloc[-length:,1:4].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 66 F3")
print(lm1.summary())

# 66 F5
length = np.min([p66.shape[0],F5.shape[0]])
y = p66["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F5.iloc[-length:,1:6].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 66 F5")
print(lm1.summary())


# 66 Q
length = np.min([p66.shape[0],Q.shape[0]])
y = p66["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F5.iloc[-length:,1:6].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 66 Q")
print(lm1.summary())
```

```
# 66 CAPM
                                 OLS Regression Results
=======================================================================================
Dep. Variable:              arbitrage   R-squared (uncentered):                   0.018
Model:                            OLS   Adj. R-squared (uncentered):              0.017
Method:                 Least Squares   F-statistic:                              19.90
Date:                Tue, 12 May 2020   Prob (F-statistic):                    9.00e-06
Time:                        22:32:28   Log-Likelihood:                         -3925.6
No. Observations:                1117   AIC:                                      7853.
Df Residuals:                    1116   BIC:                                      7858.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
```

```
==============================================================================
Mkt-RF         0.2019      0.045      4.461      0.000       0.113       0.291
==============================================================================
Omnibus:                      495.243   Durbin-Watson:                   1.060
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5396.868
Skew:                           1.745   Prob(JB):                         0.00
Kurtosis:                      13.187   Cond. No.                         1.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
# 66 F3
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.061
Model:                           OLS   Adj. R-squared (uncentered):              0.058
Method:                Least Squares   F-statistic:                              24.02
Date:               Tue, 12 May 2020   Prob (F-statistic):                    4.58e-15
Time:                       22:32:28   Log-Likelihood:                         -3900.4
No. Observations:               1117   AIC:                                     7807.
Df Residuals:                   1114   BIC:                                     7822.
Df Model:                          3
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.0830      0.048      1.730      0.084      -0.011       0.177
SMB            0.2553      0.079      3.232      0.001       0.100       0.410
HML            0.4352      0.070      6.197      0.000       0.297       0.573
==============================================================================
Omnibus:                      297.371   Durbin-Watson:                   1.080
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1344.735
Skew:                           1.176   Prob(JB):                    9.87e-293
Kurtosis:                       7.834   Cond. No.                         1.91
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
# 66 F5
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.069
Model:                           OLS   Adj. R-squared (uncentered):              0.062
Method:                Least Squares   F-statistic:                              9.966
Date:               Tue, 12 May 2020   Prob (F-statistic):                    3.26e-09
Time:                       22:32:28   Log-Likelihood:                         -2327.6
```

```
No. Observations:                 680   AIC:                              4665.
Df Residuals:                     675   BIC:                              4688.
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.1803      0.073      2.479      0.013       0.038       0.323
SMB            0.2067      0.103      1.998      0.046       0.004       0.410
HML           -0.3838      0.143     -2.686      0.007      -0.664      -0.103
RMW            0.6056      0.143      4.237      0.000       0.325       0.886
CMA            1.2341      0.211      5.858      0.000       0.820       1.648
==============================================================================
Omnibus:                       41.356   Durbin-Watson:                   1.181
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               63.700
Skew:                           0.468   Prob(JB):                     1.47e-14
Kurtosis:                       4.172   Cond. No.                         4.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# 66 Q
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.073
Model:                           OLS   Adj. R-squared (uncentered):              0.066
Method:                Least Squares   F-statistic:                              9.902
Date:               Tue, 12 May 2020   Prob (F-statistic):                    3.95e-09
Time:                       22:32:28   Log-Likelihood:                         -2177.5
No. Observations:                634   AIC:                                     4365.
Df Residuals:                    629   BIC:                                     4387.
Df Model:                          5
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.1691      0.075      2.262      0.024       0.022       0.316
SMB            0.1885      0.108      1.740      0.082      -0.024       0.401
HML           -0.4425      0.148     -2.996      0.003      -0.733      -0.152
RMW            0.6266      0.147      4.250      0.000       0.337       0.916
CMA            1.2905      0.220      5.861      0.000       0.858       1.723
==============================================================================
Omnibus:                       37.356   Durbin-Watson:                   1.197
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               56.116
Skew:                           0.461   Prob(JB):                     6.52e-13
Kurtosis:                       4.128   Cond. No.                         4.06
==============================================================================
```

```
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

2.2 69

Almost all the excess return are sinificant according to the p-value

```
In [330]: # imports
          from sklearn.linear_model import LinearRegression
          import statsmodels.api as sm
          from scipy import stats


          # 69 CAPM
          length = np.min([p96.shape[0],CAPM.shape[0]])
          y = p69["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
          x = CAPM.iloc[-length:,:].reset_index().drop("index",axis=1)
          lm1 = sm.OLS(y,x.astype(float)).fit()
          print("\n# 66 CAPM")
          print(lm1.summary())


          # 69 F3
          length = np.min([p96.shape[0],F3.shape[0]])
          y = p69["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
          x = F3.iloc[-length:,1:4].reset_index().drop("index",axis=1)
          #print(x)
          lm1 = sm.OLS(y,x.astype(float)).fit()
          print("\n# 66 F3")
          print(lm1.summary())

          # 69 F5
          length = np.min([p96.shape[0],F5.shape[0]])
          y = p69["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
          x = F5.iloc[-length:,1:6].reset_index().drop("index",axis=1)
          #print(x)
          lm1 = sm.OLS(y,x.astype(float)).fit()
          print("\n# 66 F5")
          print(lm1.summary())


          # 69 Q
          length = np.min([p96.shape[0],Q.shape[0]])
          y = p69["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
          x = F5.iloc[-length:,1:6].reset_index().drop("index",axis=1)
          #print(x)
```

```python
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 66 Q")
print(lm1.summary())
```

```
# 66 CAPM
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.017
Model:                           OLS   Adj. R-squared (uncentered):              0.017
Method:                Least Squares   F-statistic:                              19.77
Date:               Tue, 12 May 2020   Prob (F-statistic):                    9.61e-06
Time:                       22:33:47   Log-Likelihood:                         -4377.2
No. Observations:               1117   AIC:                                      8756.
Df Residuals:                   1116   BIC:                                      8762.
Df Model:                          1
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.3016      0.068      4.446      0.000       0.168       0.435
==============================================================================
Omnibus:                      497.022   Durbin-Watson:                   1.061
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5439.617
Skew:                           1.752   Prob(JB):                         0.00
Kurtosis:                      13.228   Cond. No.                         1.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# 66 F3
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.061
Model:                           OLS   Adj. R-squared (uncentered):              0.058
Method:                Least Squares   F-statistic:                              24.12
Date:               Tue, 12 May 2020   Prob (F-statistic):                    4.00e-15
Time:                       22:33:47   Log-Likelihood:                         -4351.9
No. Observations:               1117   AIC:                                      8710.
Df Residuals:                   1114   BIC:                                      8725.
Df Model:                          3
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.1225      0.072      1.705      0.088      -0.018       0.264
SMB            0.3851      0.118      3.253      0.001       0.153       0.617
```

```
HML            0.6541      0.105      6.218      0.000       0.448       0.861
==============================================================================
Omnibus:                      298.515   Durbin-Watson:                   1.082
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1354.216
Skew:                           1.180   Prob(JB):                    8.62e-295
Kurtosis:                       7.851   Cond. No.                         1.91
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# 66 F5
                                 OLS Regression Results
=======================================================================================
Dep. Variable:              arbitrage   R-squared (uncentered):                   0.069
Model:                            OLS   Adj. R-squared (uncentered):              0.062
Method:                 Least Squares   F-statistic:                              9.966
Date:                Tue, 12 May 2020   Prob (F-statistic):                    3.26e-09
Time:                        22:33:47   Log-Likelihood:                         -2603.4
No. Observations:                 680   AIC:                                      5217.
Df Residuals:                     675   BIC:                                      5239.
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.2705      0.109      2.479      0.013       0.056       0.485
SMB            0.3100      0.155      1.998      0.046       0.005       0.615
HML           -0.5756      0.214     -2.686      0.007      -0.996      -0.155
RMW            0.9084      0.214      4.237      0.000       0.487       1.329
CMA            1.8511      0.316      5.858      0.000       1.231       2.472
==============================================================================
Omnibus:                       41.356   Durbin-Watson:                   1.181
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               63.700
Skew:                           0.468   Prob(JB):                     1.47e-14
Kurtosis:                       4.172   Cond. No.                         4.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# 66 Q
                                 OLS Regression Results
=======================================================================================
Dep. Variable:              arbitrage   R-squared (uncentered):                   0.073
Model:                            OLS   Adj. R-squared (uncentered):              0.066
Method:                 Least Squares   F-statistic:                              9.902
Date:                Tue, 12 May 2020   Prob (F-statistic):                    3.95e-09
```

```
Time:                        22:33:47   Log-Likelihood:                -2434.6
No. Observations:                 634   AIC:                             4879.
Df Residuals:                     629   BIC:                             4901.
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.2537      0.112      2.262      0.024       0.033       0.474
SMB            0.2828      0.162      1.740      0.082      -0.036       0.602
HML           -0.6638      0.222     -2.996      0.003      -1.099      -0.229
RMW            0.9399      0.221      4.250      0.000       0.506       1.374
CMA            1.9357      0.330      5.861      0.000       1.287       2.584
==============================================================================
Omnibus:                       37.356   Durbin-Watson:                   1.197
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               56.116
Skew:                           0.461   Prob(JB):                     6.52e-13
Kurtosis:                       4.128   Cond. No.                         4.06
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

2.3 96

Almost all the excess return are sinificant according to the p-value. Except Rm-Rf in all the model, and SMB in the last one

```
In [332]: # imports
          from sklearn.linear_model import LinearRegression
          import statsmodels.api as sm
          from scipy import stats


          # 96 CAPM
          length = np.min([p96.shape[0],CAPM.shape[0]])
          y = p96["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
          x = CAPM.iloc[-length:,:].reset_index().drop("index",axis=1)
          lm1 = sm.OLS(y,x.astype(float)).fit()
          print("\n# 66 CAPM")
          print(lm1.summary())


          # 96 F3
          length = np.min([p96.shape[0],F3.shape[0]])
          y = p96["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
          x = F3.iloc[-length:,1:4].reset_index().drop("index",axis=1)
```

```python
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 66 F3")
print(lm1.summary())

# 69 F5
length = np.min([p96.shape[0],F5.shape[0]])
y = p96["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F5.iloc[-length:,1:6].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 66 F5")
print(lm1.summary())


# 96 Q
length = np.min([p96.shape[0],Q.shape[0]])
y = p96["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F5.iloc[-length:,1:6].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 66 Q")
print(lm1.summary())
```

```
# 66 CAPM
                                 OLS Regression Results
=======================================================================================
Dep. Variable:              arbitrage   R-squared (uncentered):                  0.010
Model:                            OLS   Adj. R-squared (uncentered):             0.009
Method:                 Least Squares   F-statistic:                             11.13
Date:                Tue, 12 May 2020   Prob (F-statistic):                   0.000878
Time:                        22:34:54   Log-Likelihood:                        -3835.7
No. Observations:                1114   AIC:                                     7673.
Df Residuals:                    1113   BIC:                                     7678.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.1407      0.042      3.336      0.001       0.058       0.223
==============================================================================
Omnibus:                      417.570   Durbin-Watson:                   1.286
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3746.558
Skew:                           1.470   Prob(JB):                         0.00
Kurtosis:                      11.489   Cond. No.                         1.00
==============================================================================
```

```
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# 66 F3
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.039
Model:                           OLS   Adj. R-squared (uncentered):              0.037
Method:                Least Squares   F-statistic:                              15.07
Date:               Tue, 12 May 2020   Prob (F-statistic):                    1.28e-09
Time:                       22:34:54   Log-Likelihood:                         -3819.0
No. Observations:               1114   AIC:                                      7644.
Df Residuals:                   1111   BIC:                                      7659.
Df Model:                          3
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.0466      0.045      1.036      0.300      -0.042       0.135
SMB            0.2280      0.074      3.075      0.002       0.083       0.374
HML            0.3131      0.066      4.749      0.000       0.184       0.442
==============================================================================
Omnibus:                     252.413   Durbin-Watson:                   1.314
Prob(Omnibus):                 0.000   Jarque-Bera (JB):             1128.400
Skew:                          0.992   Prob(JB):                    9.36e-246
Kurtosis:                      7.514   Cond. No.                         1.91
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# 66 F5
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.061
Model:                           OLS   Adj. R-squared (uncentered):              0.054
Method:                Least Squares   F-statistic:                              8.731
Date:               Tue, 12 May 2020   Prob (F-statistic):                    4.90e-08
Time:                       22:34:54   Log-Likelihood:                         -2287.4
No. Observations:                680   AIC:                                      4585.
Df Residuals:                    675   BIC:                                      4607.
Df Model:                          5
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.1085      0.069      1.583      0.114      -0.026       0.243
SMB            0.1955      0.097      2.005      0.045       0.004       0.387
```

```
HML           -0.4186      0.135     -3.109      0.002      -0.683      -0.154
RMW            0.5852      0.135      4.345      0.000       0.321       0.850
CMA            1.0826      0.199      5.453      0.000       0.693       1.472
==============================================================================
Omnibus:                       45.964   Durbin-Watson:                   1.370
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               88.721
Skew:                           0.435   Prob(JB):                     5.43e-20
Kurtosis:                       4.541   Cond. No.                         4.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# 66 Q
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.067
Model:                           OLS   Adj. R-squared (uncentered):              0.060
Method:                Least Squares   F-statistic:                              9.066
Date:               Tue, 12 May 2020   Prob (F-statistic):                    2.46e-08
Time:                       22:34:54   Log-Likelihood:                         -2140.6
No. Observations:                634   AIC:                                      4291.
Df Residuals:                    629   BIC:                                      4313.
Df Model:                          5
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.1051      0.071      1.490      0.137      -0.033       0.244
SMB            0.1582      0.102      1.548      0.122      -0.043       0.359
HML           -0.4845      0.139     -3.478      0.001      -0.758      -0.211
RMW            0.5822      0.139      4.185      0.000       0.309       0.855
CMA            1.1796      0.208      5.679      0.000       0.772       1.588
==============================================================================
Omnibus:                       41.397   Durbin-Watson:                   1.396
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               77.511
Skew:                           0.427   Prob(JB):                     1.47e-17
Kurtosis:                       4.485   Cond. No.                         4.06
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

2.4 99

Almost all the excess return are sinificant according to the p-value. except the Rm-Rf in all the model

In [333]:
```python
# imports
from sklearn.linear_model import LinearRegression
import statsmodels.api as sm
from scipy import stats


# 99 CAPM
length = np.min([p99.shape[0],CAPM.shape[0]])
y = p99["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = CAPM.iloc[-length:,:].reset_index().drop("index",axis=1)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 99 CAPM")
print(lm1.summary())


# 99 F3
length = np.min([p99.shape[0],F3.shape[0]])
y = p99["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F3.iloc[-length:,1:4].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 99 F3")
print(lm1.summary())

# 99 F5
length = np.min([p99.shape[0],F5.shape[0]])
y = p99["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F5.iloc[-length:,1:6].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 99 F5")
print(lm1.summary())


# 99 Q
length = np.min([p99.shape[0],Q.shape[0]])
y = p99["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F5.iloc[-length:,1:6].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# 99 Q")
print(lm1.summary())
```

```
# 99 CAPM
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.010
```

```
Model:                            OLS   Adj. R-squared (uncentered):              0.009
Method:                 Least Squares   F-statistic:                              10.98
Date:                Tue, 12 May 2020   Prob (F-statistic):                    0.000952
Time:                        22:35:59   Log-Likelihood:                         -4286.7
No. Observations:                1114   AIC:                                      8575.
Df Residuals:                    1113   BIC:                                      8580.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.2094      0.063      3.313      0.001       0.085       0.333
==============================================================================
Omnibus:                      418.791   Durbin-Watson:                   1.286
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3770.311
Skew:                           1.474   Prob(JB):                         0.00
Kurtosis:                      11.517   Cond. No.                         1.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# 99 F3
                                 OLS Regression Results
=======================================================================================
Dep. Variable:              arbitrage   R-squared (uncentered):                   0.039
Model:                            OLS   Adj. R-squared (uncentered):              0.037
Method:                 Least Squares   F-statistic:                              15.10
Date:                Tue, 12 May 2020   Prob (F-statistic):                    1.23e-09
Time:                        22:35:59   Log-Likelihood:                         -4269.9
No. Observations:                1114   AIC:                                      8546.
Df Residuals:                    1111   BIC:                                      8561.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.0681      0.067      1.008      0.313      -0.064       0.200
SMB            0.3425      0.111      3.081      0.002       0.124       0.561
HML            0.4712      0.099      4.768      0.000       0.277       0.665
==============================================================================
Omnibus:                      253.194   Durbin-Watson:                   1.314
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1133.800
Skew:                           0.995   Prob(JB):                    6.29e-247
Kurtosis:                       7.524   Cond. No.                         1.91
==============================================================================

Warnings:
```

```
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# 99 F5
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.061
Model:                           OLS   Adj. R-squared (uncentered):              0.054
Method:                Least Squares   F-statistic:                              8.731
Date:               Tue, 12 May 2020   Prob (F-statistic):                    4.90e-08
Time:                       22:35:59   Log-Likelihood:                         -2563.1
No. Observations:                680   AIC:                                      5136.
Df Residuals:                    675   BIC:                                      5159.
Df Model:                          5
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.1628      0.103      1.583      0.114      -0.039       0.365
SMB            0.2933      0.146      2.005      0.045       0.006       0.580
HML           -0.6279      0.202     -3.109      0.002      -1.024      -0.231
RMW            0.8778      0.202      4.345      0.000       0.481       1.275
CMA            1.6240      0.298      5.453      0.000       1.039       2.209
==============================================================================
Omnibus:                      45.964   Durbin-Watson:                   1.370
Prob(Omnibus):                 0.000   Jarque-Bera (JB):               88.721
Skew:                          0.435   Prob(JB):                     5.43e-20
Kurtosis:                      4.541   Cond. No.                         4.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# 99 Q
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.067
Model:                           OLS   Adj. R-squared (uncentered):              0.060
Method:                Least Squares   F-statistic:                              9.066
Date:               Tue, 12 May 2020   Prob (F-statistic):                    2.46e-08
Time:                       22:35:59   Log-Likelihood:                         -2397.6
No. Observations:                634   AIC:                                      4805.
Df Residuals:                    629   BIC:                                      4828.
Df Model:                          5
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.1576      0.106      1.490      0.137      -0.050       0.365
```

```
SMB            0.2372      0.153      1.548      0.122     -0.064       0.538
HML           -0.7268      0.209     -3.478      0.001     -1.137      -0.316
RMW            0.8732      0.209      4.185      0.000      0.464       1.283
CMA            1.7695      0.312      5.679      0.000      1.158       2.381
==============================================================================
Omnibus:                       41.397   Durbin-Watson:                   1.396
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               77.511
Skew:                           0.427   Prob(JB):                     1.47e-17
Kurtosis:                       4.485   Cond. No.                         4.06
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

3 Question 3

In [380]:
```python
# load data
data = pd.read_excel(path+"/25_Portfolios_5x5.xlsx")
name = data.iloc[14,:].to_list()
new_data = pd.DataFrame(data.iloc[15:1139,:].values,columns=name)
new_data.head()
```

Out[380]:
```
      NaN SMALL LoBM ME1 BM2 ME1 BM3 ME1 BM4 SMALL HiBM ME2 BM1 ME2 BM2  \
0  192607     3.7782 -0.4119 -1.9434   0.353     2.0534  2.1904  2.4192
1  192608    -2.2074 -8.7275  2.4404  0.6086     8.3968  2.1709 -1.1849
2  192609    -6.2113 -0.2989 -6.1982 -1.6368     0.8649  -1.855 -1.2618
3  192610    -8.6241 -3.7532 -5.6719   5.717    -2.5476 -1.7995 -3.2663
4  192611     3.4744  6.6476  2.2634  -4.702     0.5362  2.9051  -2.369

  ME2 BM3 ME2 BM4  ... ME4 BM1 ME4 BM2 ME4 BM3 ME4 BM4 ME4 BM5 BIG LoBM  \
0  0.4926  -1.577  ...  1.5893  1.5278  1.1869  0.2727  2.4678   3.4539
1  4.0084  0.4643  ...  1.3336   3.873  2.0059  2.1706  5.3422   1.0124
2  1.0829 -3.0405  ...  1.0923  -0.525 -1.7314  1.4646   0.873  -1.2906
3 -5.0745  -8.045  ... -3.3361 -2.6559 -2.0316 -3.1051 -5.3525  -2.7413
4  3.0078  4.6649  ...  3.4448  2.3887  3.7403   4.932  1.8213   4.2946

  ME5 BM2 ME5 BM3 ME5 BM4 BIG HiBM
0  6.0902  2.0266  3.1111   0.5623
1  4.1903  2.0131  5.4849   7.7576
2  3.6538   0.095 -0.7487  -2.4284
3 -3.0071 -2.2437 -4.6719  -5.8129
4  2.5326  1.5204  3.6619   2.5636

[5 rows x 26 columns]
```

In [396]:
```python
# use the function from question 1
portfolio_pool_5 = get_portfolio_pool(new_data.iloc[:,1:],5,3)
```

```
portfolio_pool_5

def short_long(month_return, portfolio_pool, K, skip=1):
    index, end = portfolio_pool.index[1+skip:], portfolio_pool.index[-1]
    raw_df = pd.DataFrame(np.zeros((len(index),3)),index=index,columns=["long","shor
    for i in index:
        hold_index = list(range(i,min(end,i+K)))
        for i in hold_index:
            long = portfolio_pool.loc[i,["long_0","long_1","long_2"]].values
            short = portfolio_pool.loc[i,["short_3","short_2","short_1"]].values
            #print(company_columns)
            raw_df.loc[i,"long"] += month_return.loc[i,long].sum()
            raw_df.loc[i,"short"] += month_return.loc[i,short].sum()
    raw_df["long"] /= 60
    raw_df["short"] /= 60
    raw_df["arbitrage"] = raw_df["short"] - raw_df["long"]
    return raw_df

q3 = long_short(new_data,portfolio_pool_5,5)
```

It is significant greater than zero

In [398]:
```
print(f"df: {q3.shape[0]}")
print(f"95%t-score: {1.646}")
print(f't66: {q3["arbitrage"].mean()/q3["arbitrage"].std()*np.sqrt(q3.shape[0])}')
```

```
df: 1118
95%t-score: 1.646
t66: 19.81441972208044
```

According to the t-value Except for Rm-Rf, RMW in 5 factor model, and HXZ 1-factor mode are not significant, all other parameters are significant

In [399]:
```
# imports
from sklearn.linear_model import LinearRegression
import statsmodels.api as sm
from scipy import stats


# q3 CAPM
length = np.min([q3.shape[0],CAPM.shape[0]])
y = q3["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = CAPM.iloc[-length:,:].reset_index().drop("index",axis=1)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# q3 CAPM")
print(lm1.summary())
```

```python
# q3 F3
length = np.min([q3.shape[0],F3.shape[0]])
y = q3["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F3.iloc[-length:,1:4].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# q3 F3")
print(lm1.summary())

# q3 F5
length = np.min([q3.shape[0],F5.shape[0]])
y = q3["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F5.iloc[-length:,1:6].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# q3 F5")
print(lm1.summary())


# q3 Q
length = np.min([q3.shape[0],Q.shape[0]])
y = q3["arbitrage"].iloc[-length:].reset_index().drop("index",axis=1)
x = F5.iloc[-length:,1:6].reset_index().drop("index",axis=1)
#print(x)
lm1 = sm.OLS(y,x.astype(float)).fit()
print("\n# q3 Q")
print(lm1.summary())
```

```
# q3 CAPM
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.049
Model:                           OLS   Adj. R-squared (uncentered):              0.048
Method:                Least Squares   F-statistic:                              57.63
Date:               Tue, 12 May 2020   Prob (F-statistic):                    6.65e-14
Time:                       23:20:30   Log-Likelihood:                         -3613.1
No. Observations:               1118   AIC:                                      7228.
Df Residuals:                   1117   BIC:                                      7233.
Df Model:                          1
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.2590      0.034      7.592      0.000       0.192       0.326
==============================================================================
Omnibus:                     1083.196   Durbin-Watson:                   1.261
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            87348.981
```

```
Skew:                           4.265   Prob(JB):                       0.00
Kurtosis:                      45.454   Cond. No.                       1.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
# q3 F3
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.246
Model:                           OLS   Adj. R-squared (uncentered):              0.244
Method:                Least Squares   F-statistic:                              121.1
Date:               Tue, 12 May 2020   Prob (F-statistic):                    7.02e-68
Time:                       23:20:30   Log-Likelihood:                         -3483.6
No. Observations:               1118   AIC:                                      6973.
Df Residuals:                   1115   BIC:                                      6988.
Df Model:                          3
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.0532      0.033      1.615      0.107      -0.011       0.118
SMB            0.5335      0.054      9.833      0.000       0.427       0.640
HML            0.6432      0.048     13.352      0.000       0.549       0.738
==============================================================================
Omnibus:                      656.093   Durbin-Watson:                   1.336
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            12990.092
Skew:                           2.300   Prob(JB):                         0.00
Kurtosis:                      19.053   Cond. No.                         1.91
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
# q3 F5
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.117
Model:                           OLS   Adj. R-squared (uncentered):              0.110
Method:                Least Squares   F-statistic:                              17.86
Date:               Tue, 12 May 2020   Prob (F-statistic):                    1.18e-16
Time:                       23:20:30   Log-Likelihood:                         -1933.1
No. Observations:                680   AIC:                                      3876.
Df Residuals:                    675   BIC:                                      3899.
Df Model:                          5
Covariance Type:           nonrobust
==============================================================================
```

```
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF         0.0117      0.041      0.286      0.775      -0.068       0.092
SMB            0.3544      0.058      6.120      0.000       0.241       0.468
HML           -0.1734      0.080     -2.168      0.030      -0.330      -0.016
RMW            0.0593      0.080      0.741      0.459      -0.098       0.216
CMA            0.7507      0.118      6.366      0.000       0.519       0.982
==============================================================================
Omnibus:                      180.779   Durbin-Watson:                   1.306
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2628.857
Skew:                           0.760   Prob(JB):                         0.00
Kurtosis:                      12.512   Cond. No.                         4.00
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

# q3 Q
                                 OLS Regression Results
=======================================================================================
Dep. Variable:             arbitrage   R-squared (uncentered):                   0.117
Model:                           OLS   Adj. R-squared (uncentered):              0.110
Method:                Least Squares   F-statistic:                              16.71
Date:               Tue, 12 May 2020   Prob (F-statistic):                    1.63e-15
Time:                       23:20:30   Log-Likelihood:                         -1809.7
No. Observations:                634   AIC:                                     3629.
Df Residuals:                    629   BIC:                                     3652.
Df Model:                          5
Covariance Type:           nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Mkt-RF        -0.0008      0.042     -0.020      0.984      -0.083       0.081
SMB            0.3389      0.061      5.589      0.000       0.220       0.458
HML           -0.2212      0.083     -2.675      0.008      -0.384      -0.059
RMW            0.0622      0.083      0.754      0.451      -0.100       0.224
CMA            0.7938      0.123      6.441      0.000       0.552       1.036
==============================================================================
Omnibus:                      166.971   Durbin-Watson:                   1.312
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2346.004
Skew:                           0.751   Prob(JB):                         0.00
Kurtosis:                      12.303   Cond. No.                         4.06
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

4 Question 4

In [435]:
```
# load data
Q4_data = pd.read_excel(path+"/book_market.xlsx",header =None)
mkt_val = pd.read_csv(path+"/ME.csv")
mkt_val.set_index("Date",inplace=True)
book_val = pd.read_csv(path+"/Book_value.csv")
book_val.set_index("Date",inplace=True)
risk_free_returns = pd.read_csv(path+"/t_bill_returns.csv")
risk_free_returns.set_index("date",inplace=True)
stock_ret = pd.read_csv(path+"/HPR_data.csv")
stock_ret.set_index("date",inplace=True)
mrkt_ret = pd.read_csv(path+"/mrkt_returns.csv")
mrkt_ret.set_index("date",inplace=True)
```

In [436]:
```
import numpy as np
stock_ret = stock_ret.apply(lambda x: np.log(x/100 +1)*100, axis=1,result_type='broac
stock_ret = stock_ret.iloc[-179:]
ln_book = book_val.apply(lambda x: np.log(x))
ln_mkt = mkt_val.apply(lambda y: np.log(y))
bk_mkt_ratio = np.subtract(ln_book,ln_mkt)
bk_mkt_ratio.head()
long_ret = []
short_ret = []
arb_ret = []

for i in range(len(stock_ret)):
    if i > 59 and i+1 < len(stock_ret.index):
        expected_returns = {}

        for n in stock_ret.columns:

            BV_MV_lagged = bk_mkt_ratio[n].iloc[0:i].values
            X = np.column_stack((np.repeat(1, len(BV_MV_lagged)),BV_MV_lagged))
            Y = stock_ret[n].iloc[0:i].values

            estimates =  np.matmul( np.matmul(np.linalg.inv( np.matmul(np.transpose(X
            expected_returns[n] = estimates[0] + estimates[1] * bk_mkt_ratio[n].value

        sorted_data = sorted(expected_returns.items(),key=lambda x:x[1],reverse=True)
        highest = sorted_data[0:3]
        lowest = sorted_data[-3:]
        unzipped_highest = zip(*highest)
        unzipped_lowest = zip(*lowest)
        high_three =list(unzipped_highest)[0]
        low_three = list(unzipped_lowest)[0]

        net_long = 0
```

```
            net_short = 0
            for j in high_three:
                net_long += stock_ret[j].values[i+1]*(1/3)
            for k in low_three:
                net_short += stock_ret[k].values[i+1]*(1/3)

            long_ret.append(net_long)
            short_ret.append(net_short)
            arb_ret.append(net_long - net_short)
```

In [430]:
```
# compute t-status
print("arbitrage:")
print(f"return: {np.array(arb_ret).mean()}")
print(f"t-status: {np.array(arb_ret).mean()/np.array(arb_ret).std()}")

print("short:")
print(f"return: {np.array(short_ret).mean()}")
print(f"t-status: {np.array(short_ret).mean()/np.array(long_ret).std()}")

print("long:")
print(f"return: {np.array(long_ret).mean()}")
print(f"t-status: {np.array(long_ret).mean()/np.array(long_ret).std()}")
```

```
arbitrage:
return: 0.0020941903582484445
t-status: 0.04856800066021095
short:
return: 0.01569530610190985
t-status: 0.3494865082381052
long:
return: 0.0177894964601583
t-status: 0.39611772849835214
```

5 question 5

In [438]:
```
# load data
stock_ret_no_div = pd.read_csv(path+"/HPR_data_wo_div.csv")
stock_ret_no_div.set_index("date",inplace=True)

log_stock_no_div = stock_ret_no_div.apply(lambda x:np.log(x/100 +1)*100, axis=1,resul
log_stock_no_div = log_stock_no_div.iloc[-179:]
dp_ratio=np.subtract(stock_ret , log_stock_no_div)
```

In [439]:
```
long_ret = []
short_ret = []
arb_ret = []
```

```
for i in range(len(stock_ret)):
    if i > 59 and i < len(stock_ret.index)-1:
        expected_returns = {}

        for n in stock_ret.columns:

            DP_ratio = dp_ratio[n].iloc[0:i].values
            X = np.column_stack((np.repeat(1, len(DP_ratio)),DP_ratio))
            Y = stock_ret[n].iloc[0:i].values

            estimates =  np.matmul( np.matmul(np.linalg.inv( np.matmul(np.transpose(X
            expected_returns[n] = estimates[0] + estimates[1] * bk_mkt_ratio[n].value

        sorted_data = sorted(expected_returns.items(),key=lambda x:x[1],reverse=True)
        highest = sorted_data[0:3]
        lowest = sorted_data[-3:]
        unzipped_highest = zip(*highest)
        unzipped_lowest = zip(*lowest)
        high_three =list(unzipped_highest)[0]
        low_three = list(unzipped_lowest)[0]

        net_long = 0
        net_short = 0
        for j in high_three:
            net_long += stock_ret[j].values[i+1]*(1/3)
        for k in low_three:
            net_short += stock_ret[k].values[i+1]*(1/3)

        long_ret.append(net_long)
        short_ret.append(net_short)
        arb_ret.append(net_long - net_short)
```

In [440]:
```
# compute t-status
print("arbitrage:")
print(f"return: {np.array(arb_ret).mean()}")
print(f"t-status: {np.array(arb_ret).mean()/np.array(arb_ret).std()}")

print("short:")
print(f"return: {np.array(short_ret).mean()}")
print(f"t-status: {np.array(short_ret).mean()/np.array(long_ret).std()}")

print("long:")
print(f"return: {np.array(long_ret).mean()}")
print(f"t-status: {np.array(long_ret).mean()/np.array(long_ret).std()}")
```

```
arbitrage:
return: 0.0016691245415136299
t-status: 0.03838523936983855
```

```
short:
return: 0.01635686839374341
t-status: 0.36376643149168636
long:
return: 0.018025992935257036
t-status: 0.40088670803642246
```