上海外国语大学

行为金融与量化投资
Final Report

0191149055 范露元

完 成 日 期 二〇二二 年 六 月 十九

1. Specify the trading strategy

According to existing researches, compared with institutional investors, individual investors always have the gambling behavior in economy activities over-believing in their luck. Since there are many investors having lottery preference when investing stocks, the future stock returns and the pricing of risk by the market are influenced. The lottery stocks are overvalued when irrational investors are pursuing a high return at a low cost, which is absolutely a small probability event. Many scholars have suggested that lottery preference has a significant negative affect on return ratio. The average returns of non-lottery stocks are higher than those of lottery stocks because the lottery stocks which have high history prices are in fact expected to have low future prices.

Therefore, we design a long-short strategy utilizing the investors' behavioral biases to make profit. Before that, we use the max daily return rate of the last month as the identification index of lottery stocks to group them in 10 parts. Since we observe thousands of stocks, the group consisted of 10 percent of stocks are precise enough to avoid the specific risk. We plan to sell the highest MAX group and buy lowest MAX group, which is a zero-investment intended to avoid the exposure to the market risk. The sample period is from January 2007 to January 2017.

In our portfolio, we choose the value-weighted method to buy and sell stocks considering two aspects. Firstly, the equal-weighted method may change the prices of the stocks after our buying or selling stocks since many lottery stocks are too small-sized to vulnerable to the behaviors of investors. Thus, if we used the equal-weighted method, the history prices used in the test would be inaccurate. To ensure the accuracy of our test as much as possible, we choose the value-weighted method. Secondly, we have read some literature in which the equal-weighted method and the value-weighted method exhibit the same character without big difference. And we rebalance our portfolio at a monthly frequency which we think is a moderate frequency during the period of 120 months and also often used in most previous researches.

The table below shows the first ten examples of returns of different quantiles in the value-weighted method. We can see the returns of the first column *q1* are usually larger than those of the tenth column *q10*, which is compatible with the present conclusion that stocks with high prices in the past are expected to have low future prices.

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10
1	-0.0147309138	0.004780667	-0.0249535543	-0.0032638703	-0.0151458367	-0.0748681325	-0.0094646270	-0.0332007268	0.0026350463	-0.0469743813
2	0.0466990023	-0.009862542	-0.0023123621	0.0571727168	-0.0002303599	-0.0036070735	0.0016834689	0.0373460548	-0.0271627445	0.0320632500
3	0.0826978835	0.052338750	0.0557346591	0.0683483920	0.0446124284	0.0647215387	0.0069313348	-0.0791843827	0.0157127559	-0.0243487107
4	0.0293071694	0.039694192	0.0313539032	0.0164641476	0.0565375506	0.0277261570	0.0055108984	0.0538275476	0.0351056030	0.0156569954
5	-0.0307236536	0.009686466	0.0016624514	0.0158127416	0.0406290211	0.0244093937	-0.0527719587	0.0359954755	0.0004556400	0.0206794823
6	-0.0113534874	-0.018864121	0.0116511118	-0.0037446619	-0.0377103495	0.0013142095	-0.0610916053	-0.0860779705	-0.0687483200	-0.0350363680
7	0.0227255825	0.023664629	0.0107398673	0.0154635604	0.0217490695	0.0288098946	-0.0159963267	0.0572450129	0.0366595585	0.0305165221
8	0.0349628319	0.041802414	0.0701861563	-0.0082274965	0.0149785189	0.0524644503	0.0576642649	0.0493920794	0.0349918356	0.0180106723
9	0.1610390590	0.003231446	0.0515840761	-0.0001339736	0.0173509706	-0.0280532980	0.0603953185	-0.0160994502	0.1964984820	0.0005200374
10	-0.0227485316	-0.063688596	-0.0974360882	-0.1200917223	-0.0728182825	-0.0736300792	-0.0505024227	-0.0641215422	-0.1039178154	0.0011499824

Figure-1

2. The raw returns and CAPM-adjusted alphas for strategy

Raw return:

From the figure, if you invest one dollar at the beginning, it can become four dollar after 120 months.

If such investment portfolio is made, the fluctuation curve of return is shown in the figure below. The retracement is relatively large at 18 months, 25 months, 43 months and 65 months respectively.

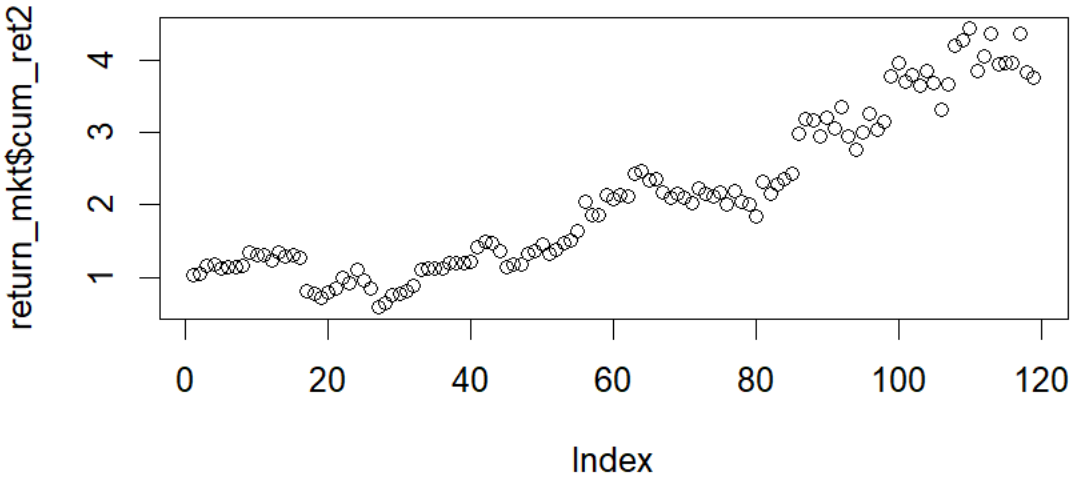

Figure-2

Do t-test of return_equal and return_value. The results are shown in the following figure.

```
        One Sample t-test

data:  return_equal$longshort
t = -0.35051, df = 118, p-value = 0.7266
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.014223686  0.009945682
sample estimates:
   mean of x
-0.002139002
```

Figure-3

Return_ equal has a small t-value and a large p-value. The P-value is 0.7266, obviously greater than 10%, which is not significant. This means that there is no significant difference between the return rate of equal weighted portfolio and 0. Such a launch of strategy is invalid in the selected time period, that is, it is not profitable.

```
        One Sample t-test

data:  return_mkt$longshort
t = 1.7867, df = 118, p-value = 0.07656
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.001743613  0.033929538
sample estimates:
 mean of x
0.01609296
```

Figure-4

The t-value of return_value is relatively large, and the P-value is relatively small. The P-value is 0.07656, which is less than 10%, so we say it is significant in the 90% confidence interval. This means that there is a significant gap between the return rate of the value weighted portfolio and 0. Its average value is 0.01609296, which means that there can be a return of about 1.6% per month, which is significantly greater than zero.

So far, we can see that the return of equal weighted is not significant, and the return of value weighted is significant. We believe that the main reason is that lottery of large stocks has significant benefits, or there are too many stock data.

CAPM - adjusted return:

We adjust raw returns with benchmark and calculate the alphas of the strategy, get the market return at the same time period first. And then, we regress the monthly long-short portfolio return on the contemporaneous monthly market return, the intercept that we get is the alpha of such strategy.

```
Call:
lm(formula = return_mkt$longshort ~ return_mkt$mkt_rf, data = return_mkt)

Residuals:
     Min       1Q   Median       3Q      Max
-0.43137 -0.05213 -0.00379  0.05496  0.26094

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.019732   0.008798   2.243   0.0268 *
return_mkt$mkt_rf -0.005786   0.001922  -3.010   0.0032 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09506 on 117 degrees of freedom
Multiple R-squared:  0.07188,   Adjusted R-squared:  0.06394
F-statistic: 9.061 on 1 and 117 DF,  p-value: 0.003199
```

Figure-5

Through such a regression, we can test whether the long-short portfolio return can be interpreted by the market factory return. If it can be explained, all returns should come from the variables of market factory return. The portfolio return will change as much as the market factory return. If not completely explained by market risk. Then intercept should be significantly greater than zero. Obviously, in the above figure, intercept is 0.019732 and P-value is 0.0268, and intercept is not significantly greater than 0, and it cannot be proved that there is an excess return rate.

The figure above does not add factors, and the figure below adds SMB factors, HML factors and RF factors. The figure below is more professional than the figure above.

Obviously, in the figure below, intercept is 0.019714, P-value is 0.0139, and intercept is significantly greater than 0, which proves that there is an excess return rate alpha=intercept=0.024296. In other words, if you invest in such a portfolio, you can obtain an excess return of about 2.43% per month.

```
Call:
lm(formula = return_mkt$longshort ~ return_mkt$mkt_rf + return_mkt$SMB +
    return_mkt$HML, data = return_mkt)

Residuals:
     Min       1Q   Median       3Q      Max
-0.37549 -0.04439 -0.00942  0.04763  0.21388

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.019714   0.007892   2.498   0.0139 *
return_mkt$mkt_rf -0.001789   0.001928  -0.928   0.3555
return_mkt$SMB    -0.020419   0.003683  -5.545  1.9e-07 ***
return_mkt$HML    -0.001186   0.003018  -0.393   0.6950
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08498 on 115 degrees of freedom
Multiple R-squared:  0.271,     Adjusted R-squared:  0.252
F-statistic: 14.25 on 3 and 115 DF,  p-value: 5.86e-08
```

Figure-6

Intercept and SMB have small p-values, so they are significant.

Intercept is the excess return alpha. SMB factor is the yield of small stocks minus large stocks, which means that small stocks should receive higher compensation because they bear higher risks.

So far, we can return to the different results of equal weighted and value weighted, excluding the possibility that lottery of large stocks has advantages, we believe that the reason why the equal weighted method is not significant is that there are too many stock data.

Excluding the impact of market return, the abnormal return of this investment portfolio is also significantly greater than zero.

3. The influence of sentiment

After analyzing the relationship between ROI of our strategy and market factors, we need to consider the influence of market sentiment on ROI. Investor sentiment, defined broadly, is a belief about future cash flows and investment risks that is not justified by the facts at hand. We know that the higher the market sentiment, the more likely investors are to make mistakes, and the lower the market sentiment, the less likely investors are to make mistakes. This phenomenon indicates that when the market sentiment is high, people are more inclined to gamble and are more willing to invest in the stocks with high lottery. Since investors in the market are not all rational, we import market sentiment indicators from 2007 to 2017 to analyze whether market sentiment significantly affects returns.

We simply divide market sentiment into two grades, high and low, by the median of market sentiment sample. Then use 1 and 0 to represent high and low market sentiment respectively.

First, we used the T-test to test Whether the returns of long-short portfolios are significantly greater than zero when sentiment is high. The result is shown below:

```
        One Sample t-test

data:  return_mkt$longshort[which(sentiment$high == 1)]
t = 1.9412, df = 60, p-value = 0.05694
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.0006252249  0.0416643403
sample estimates:
 mean of x
0.02051956
```

Figure-7

As we can see, the excess return rate here is close to significant at 95% confidence. Mean of X means a monthly return of 2.05 percent in high market sentiment, when you invest in this portfolio.

Then, we also use the T-test to test when market sentiment is low. The results are shown below:

```
        One Sample t-test

data:  return_mkt$longshort[which(sentiment$high == 0)]
t = 0.77087, df = 57, p-value = 0.444
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.01827329  0.04114810
sample estimates:
mean of x
0.0114374
```

Figure-8

It can be seen from the test results that the P value is 0.444, indicating that the excess return rate is not significant when the market sentiment is low. At the same time, mean of X showed a smaller return compared to the high market sentiment. The conclusion is that the average return over the next month is greater in the case of high market sentiment. Long-short portfolios have a stronger ability to invest in high market sentiment.

After a brief analysis, we use linear regression to look at the value of the abnormal return. The results of regression when market sentiment is high is shown below:

```
Call:
lm(formula = test$longshort ~ test$mkt_rf, data = test)

Residuals:
      Min        1Q    Median        3Q       Max
-0.127196 -0.050301 -0.006243  0.036576  0.265572

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.021949   0.010088   2.176   0.0336 *
test$mkt_rf -0.007426   0.002793  -2.659   0.0101 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07868 on 59 degrees of freedom
Multiple R-squared:  0.107,     Adjusted R-squared:  0.09184
F-statistic: 7.068 on 1 and 59 DF,  p-value: 0.01009
```

Figure-9

The first regression is obtained under the condition of only considering the influence of market return. According to the results, the p value is 0.0336, indicating that the excess return rate of the portfolio is significant with 95% confidence. Compared with the previous regression coefficient, we can find that the value of excess return rate increases from 1.97% to 2.19%.

```
Call:
lm(formula = test$longshort ~ test$mkt_rf + test$SMB + test$HML,
    data = test)

Residuals:
     Min       1Q   Median       3Q      Max
-0.12259 -0.04102 -0.02208  0.03815  0.21995

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.020759   0.009358   2.218  0.03053 *
test$mkt_rf -0.003539   0.002824  -1.253  0.21539
test$SMB    -0.018441   0.005045  -3.656  0.00056 ***
test$HML     0.001423   0.004784   0.297  0.76722
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07184 on 57 degrees of freedom
Multiple R-squared:  0.2807,    Adjusted R-squared:  0.2429
F-statistic: 7.416 on 3 and 57 DF,  p-value: 0.000281
```

Figure-10

The second session included two additional variables – SMB and HML. Consistent with the previous results, SMB remains significant. Intercept's p value in the results was 0.0275, also significant at 95% confidence. And the excess return rate increases from 1.97% to 2.07%, which has a relatively smaller increase.

From these two results, we can conclude that the portfolio we use has a relatively high excess return

rate when market sentiment is high. However, since SMB has a significant impact on excess return rate, an increase in market sentiment does not lead to a significant increase in excess return rate when this variable is included.

Following the analysis of high market sentiment, we do the linear regression of what happens when sentiment is low. The results are as follows:

```
Call:
lm(formula = test$longshort ~ test$mkt_rf, data = test)

Residuals:
     Min       1Q   Median       3Q      Max
-0.42115 -0.04673  0.00265  0.06183  0.22260

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.016794   0.014861   1.130   0.2633
test$mkt_rf -0.004924   0.002747  -1.792   0.0785 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1109 on 56 degrees of freedom
  (因为不存在，2个观察量被删除了)
Multiple R-squared:  0.05424,   Adjusted R-squared:  0.03735
F-statistic: 3.212 on 1 and 56 DF,  p-value: 0.07851
```

Figure-11

In this regression, p-value is 0.2633, so the excess return rate is not significant and only 1.67%, which is lower than the previous results, 1.97%.

```
Call:
lm(formula = test$longshort ~ test$mkt_rf + test$SMB + test$HML,
    data = test)

Residuals:
     Min       1Q   Median       3Q      Max
-0.36339 -0.04243  0.01029  0.05873  0.20707

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0190493  0.0132664   1.436 0.156797
test$mkt_rf  0.0002202  0.0030524   0.072 0.942767
test$SMB    -0.0221420  0.0055379  -3.998 0.000195 ***
test$HML    -0.0040208  0.0046274  -0.869 0.388743
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09867 on 54 degrees of freedom
  (因为不存在，2个观察量被删除了)
Multiple R-squared:  0.2776,    Adjusted R-squared:  0.2375
F-statistic: 6.919 on 3 and 54 DF,  p-value: 0.0005024
```

Figure-12

After adding two more variables, we found that SMB remained significant. The p value of the excess return rate was 0.1567, which was not significant, and the excess return rate also decreased slightly compared with the original 1.97%. But there was not as much of a drop as in the previous regression.

To summarize this part, we can find that from the comparison of high and low market emotions, the market with high sentiment will get a better excess return rate on the whole. So, this suggests that our long-short strategy can play a bigger role in high market sentiment. In low-mood markets, however, excess return rate is not significant or even lower. This conclusion perfectly conforms to our hypothesis, that is, when the market mood is high, people are more inclined to gamble and take risks, and are more willing to invest in stocks with high lottery. So, we can make more money on our portfolio.

On the other hand, the inclusion of SMB and HML variables also has a significant impact on the results. The main reason for the influence here is SMB. It can be seen that SMB is always significant in both the previous regression and the regression of high and low market sentiment. This means that SMB largely explains the size of the excess return rate, and at the same time leads to a relatively reduced impact of other variables on the size of the excess return rate. Therefore, when market sentiment changes at this time, due to the limitation of SMB, sentiment no longer has a great impact on the excess return rate.

In conclusion, we recommend implementing our investment strategy in times of high market sentiment.

4. Portfolio evaluation

First, we'll decide which measurement to use to judge the portfolio's performance, based on what was taught in class, the Sharpe ratio doesn't apply to hedge funds, which shows the reward per unit of total risk, nor does the Treynor's ratio apply to hedge funds.

We should use the Information ratio to show the performance of the portfolio. Information ratio can be the appropriate measure while Sharpe ratio can not be used. But in our statics, we have the regression analysis with SMB, HML and another one without those factors. In that case, we should discuss the information ratio in different conditions.

The first one is a regression analysis without SMB and HML. The specific risk is equal to the residual standard error.

```
Call:
lm(formula = return_mkt$longshort ~ return_mkt$mkt_rf, data = return_mkt)

Residuals:
     Min       1Q   Median       3Q      Max
-0.43137 -0.05213 -0.00379  0.05496  0.26094

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.019732   0.008798   2.243   0.0268 *
return_mkt$mkt_rf -0.005786   0.001922  -3.010   0.0032 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09506 on 117 degrees of freedom
Multiple R-squared:  0.07188,   Adjusted R-squared:  0.06394
F-statistic: 9.061 on 1 and 117 DF,  p-value: 0.003199
```

Figure-13

Using those underlined numbers, the Information ratio = 0.019732/0.09506 = 0.20757

The second one is the regression analysis with SMB, HML factors.

```
Call:
lm(formula = return_mkt$longshort ~ return_mkt$mkt_rf + return_mkt$SMB +
    return_mkt$HML, data = return_mkt)

Residuals:
     Min       1Q   Median       3Q      Max
-0.37549 -0.04439 -0.00942  0.04763  0.21388

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.019714   0.007892   2.498   0.0139 *
return_mkt$mkt_rf -0.001789   0.001928  -0.928   0.3555
return_mkt$SMB    -0.020419   0.003683  -5.545  1.9e-07 ***
return_mkt$HML    -0.001186   0.003018  -0.393   0.6950
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08498 on 115 degrees of freedom
Multiple R-squared:  0.271,     Adjusted R-squared:  0.252
F-statistic: 14.25 on 3 and 115 DF,  p-value: 5.86e-08
```

Figure-14

In this case, Information ratio = 0.019714/0.08498 = 0.23198

Compared the two ratios we can find that when we consider those factors, the Information ratio rise and shows that the portfolio can get more market return while the SMB, HML, RF factors are considered.

In high sentiment market, the basic information as bellow.

```
Call:
lm(formula = test$longshort ~ test$mkt_rf, data = test)

Residuals:
      Min        1Q    Median        3Q       Max
-0.127196 -0.050301 -0.006243  0.036576  0.265572

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.021949   0.010088   2.176   0.0336 *
test$mkt_rf -0.007426   0.002793  -2.659   0.0101 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07868 on 59 degrees of freedom
Multiple R-squared:  0.107,     Adjusted R-squared:  0.09184
F-statistic: 7.068 on 1 and 59 DF,  p-value: 0.01009
```

Figure-15

Information ratio =0.27897

```
Call:
lm(formula = test$longshort ~ test$mkt_rf + test$SMB + test$HML,
    data = test)

Residuals:
     Min       1Q   Median       3Q      Max
-0.12259 -0.04102 -0.02208  0.03815  0.21995

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.020759   0.009358   2.218  0.03053 *
test$mkt_rf -0.003539   0.002824  -1.253  0.21539
test$SMB    -0.018441   0.005045  -3.656  0.00056 ***
test$HML     0.001423   0.004784   0.297  0.76722
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07184 on 57 degrees of freedom
Multiple R-squared:  0.2807,    Adjusted R-squared:  0.2429
F-statistic: 7.416 on 3 and 57 DF,  p-value: 0.000281
```

Figure-16

Information ratio = 0.020759/0.07184 = 0.28896

In low sentiment market, the basic information as bellow.

```
Call:
lm(formula = test$longshort ~ test$mkt_rf, data = test)

Residuals:
     Min       1Q   Median       3Q      Max
-0.42115 -0.04673  0.00265  0.06183  0.22260

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.016794   0.014861   1.130   0.2633
test$mkt_rf -0.004924   0.002747  -1.792   0.0785 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1109 on 56 degrees of freedom
  (因为不存在，2个观察量被删除了)
Multiple R-squared:  0.05424,   Adjusted R-squared:  0.03735
F-statistic: 3.212 on 1 and 56 DF,  p-value: 0.07851
```

Figure-17

Information ratio = 0.15143

```
Call:
lm(formula = test$longshort ~ test$mkt_rf + test$SMB + test$HML,
    data = test)

Residuals:
     Min       1Q   Median       3Q      Max
-0.36339 -0.04243  0.01029  0.05873  0.20707

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0190493  0.0132664   1.436 0.156797
test$mkt_rf  0.0002202  0.0030524   0.072 0.942767
test$SMB    -0.0221420  0.0055379  -3.998 0.000195 ***
test$HML    -0.0040208  0.0046274  -0.869 0.388743
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09867 on 54 degrees of freedom
  (因为不存在，2个观察量被删除了)
Multiple R-squared:  0.2776,    Adjusted R-squared:  0.2375
F-statistic: 6.919 on 3 and 54 DF,  p-value: 0.0005024
```

Figure-18

Information ratio = 0.19306

From this analysis we can conclude that in high sentiment market and considered SMB and HML factors, the hedge fund's performance will be better. And in all, while considering the SMB and HML factors, the performance will be better.

Reference:

[1] 赵婉竹. 彩票类股票最大收益效应的存在性及成因研究[D]. 东北财经大学,2020.DOI:10.27006/d.cnki.gdbcu.2020.000105.

[2]张婷婷. 投资者赌博偏好特征及其对股票价格行为的影响研究[D].天津大学,2017.

[3]崔惠颖. 中国博彩型股票的识别、市场表现及其收益成因的研究[D].东北财经大学,2016.