

Regression Analysis of Fund Return Rate

Science of Decision Making

Group 5 Report

CONTENT

[1. Abstract 3](#_Toc536204661)

[2. Background 4](#_Toc536204662)

[**3、Problem description** 4](#_Toc536204663)

[**4、RESEARCH APPROACH** 4](#_Toc536204664)

[**4.1 Model & Equation** 5](#_Toc536204665)

[**5、Variables & Data** 5](#_Toc536204666)

[**5.1 Variable Identification** 5](#_Toc536204667)

[**6、Data Source** 7](#_Toc536204668)

[**6.1 Data Limitation** 7](#_Toc536204669)

[**7、RESULT & INTERPRETATION** 7](#_Toc536204670)

[**7.1 Regression Result & Model Evaluation** 7](#_Toc536204671)

[**7.2 Goodness of Fit** 27](#_Toc536204672)

[**7.3 Error Term Assumption Validation** 28](#_Toc536204673)

[**7.4 Interpretation of Coefficients** 28](#_Toc536204674)

[8. Conclusion 28](#_Toc536204675)

**1、Abstract**

With China's rapid economic development, people's income increases, more and more people are aware of the importance of investment. Fund investment is one of the important ways. But many people don't know how to select or evaluate them. This report investigates what elements are most correlated to fund growth rate, also discovers some inapplicability facts. Multiple regressions are the major instruments used in this report. We choose 5 popular funds with 3139 records and collect data from fund website including indicators of 7 aspects: Weekly change of Shanghai composite index, weekly change of M2, Fund-type, standard deviation since the foundation of the fund, Sharpe Ratio, stock picking ability of manager, anti-risk ability. Among these variables, four of them are proved to be insignificant by measurement of testing and are eliminated by backward elimination method.

**2、 Background**

Investors invest in funds, in fact, they entrust funds to fund management companies for investment. Whether for investors or fund management companies, it is necessary to evaluate the operation status of investment funds, that is, fund performance. Modern fund performance evaluation method is based on Capital Asset Pricing Model (CAMP). The single factor overall performance evaluation model represented by Jensen index, Treynor index and Sharpe index is the most popular in the capital markets of developed countries.

At present, most of the research done by securities firms in this area is qualitative research, and the more precise quantitative research based on mathematical statistics analysis is scarce. By establishing regression equation and using regression analysis in mathematical statistics, this paper investigates the fitting degree and saliency of regression equation and the saliency of parameter beta, obtains the estimated value of parameter beta and standard deviation or variance, and then ranks fund performance separately. Finally, the ranking results are briefly analyzed.

**3、Problem description**

According to the above analysis, we know that in order to evaluate the return of a fund, we need to use regression analysis to predict and analyze the future. At the same time, we need some premise assumptions as the basis. For multi-factor linear model, there are usually the following assumptions: there is a linear relationship between strain and independent variables; independent variables are not random; the mathematical expectation of residual terms is zero; the square of residual terms. The difference is a constant; the residual terms of different observations have no correlation; the residual terms obey normal distribution. The following will be verified by the method of mathematical statistics.

**4、RESEARCH APPROACH**

**4.1 Model & Equation**

We have several assumptions and criteria in choosing the model:

1. To use the independent variables (Xi) to estimate dependent variable (Y)

2. To assume linear relationships between Y and Xi. For specific variable, we would assume nonlinear relationship by using both Xi and Xi2.

3. To follow Least Squares Principle: min Σ(𝑌−𝑌̂)2 or Σ(𝑌−𝑌̂)=0

We will have the regression model as below:

Y= β0+Σβ𝑖𝑋𝑖 +ε =β0+β1X1+β2X2+… +βkXk+ε

X1,… ,Xk are the independent variables, Y is the dependent variable

β0 is the Y-intercept

βj is the net change in Y for each unit change in Xj by holding other Xi constant

(i≠j; i and j are both integers)

ε is the model error

And then the regression equation will be:

E(Y) = β0+Σβ𝑖𝑋𝑖 =β0+β1X1+β2X2+… +βkXk

E(Y) is the expected Y.

Now we can come up with the estimated regression equation, the basic formula used in our research:

Ŷ= β0+Σβ𝑖𝑋𝑖 =β0+β1X1+β2X2+… +βkXk

Ŷ is the estimated value of the Y variable for a selected set of X value.

**5、Variables & Data**

**5.1 Variable Identification**

This paper chooses the data of five existing hybrid funds in the market from 2015 to 2017. The time interval is a week. The Shanghai composite index is used to establish the multiple linear regression model. We choose the backward elimination to confirm or eliminate the variables. Finally, the factors that have the greatest impact on the fund return and the related regression equation are obtained. Backward elimination method, which begins with the entire set of variables and eliminates one independent variable at each iteration. All the independent variables are selected into the model, and then the partial F test is carried out on each independent variable. The minimum F value is recorded as FL, which is compared with the pre-defined significant level F0. If FL < F0, the variable is eliminated, the remaining variables are re-fitted to the regression model, and the above steps are repeated until all the independent variables in the model cannot be eliminated. Whether selected or eliminated, backward elimination reduces the difficulty of analysis, improves the computational efficiency, and has better prediction accuracy for the stability of regression equation.

Shanghai composite index∆%. It is an index to measure the trend of the market. Its change affects the net value of funds that related to it.M2∆%. The M2 change also adjusts the total risk. It reflects that when the portfolio is mixed with the corresponding riskless assets to achieve the same level of risk as the market portfolio, the return of the mixed portfolio is higher than that of the market. The purpose is to correct the tendency of investors to only consider the original performance of the fund, and encourage them to pay attention to the risk factors in the performance of the fund at the same time, so as to help investors select the investment funds that can bring the best real performance. Compared with Sharp Ratio, M2 measures all risks as a measure of risk. This risk adjustment method can easily explain why there are different levels of return relative to different market benchmark indices. The order of M2 measurement and Sharp ratio on fund performance is consistent. The bigger the M2 measure, the better the performance of the fund; on the contrary, the worse the performance of the fund.

Fund type 1 and Fund type 2. As we use the hybrid funds, which invest part of their funds in stocks and part of their funds in bonds (the proportion of investments can be adjusted), to research. We need to comminute it into two types of funds for a further calculation. One is Fund type 1 and the other is Liability-type. Here we define hybrid funds as fund type1=0, type2=0; Equity funds as fund type1=1, type2=0; and bonded funds as fund type1=0, type2=1, respectively.

S-deviation. In investment funds, the average person pays more attention to performance, but often after buying the fund with the best recent performance, the performance of the fund is not as expected, because the selected fund is too volatile and has no stable performance. Standard Deviation is one of the tools to measure the volatility of funds. Standard deviation refers to the extent to which the fund may change. The greater the standard deviation, the greater the degree of possible changes in the future net value of the fund, the smaller the stability and the higher the risk.

Sharpe Ratio. That’s one kind of standardized indicators of fund performance evaluation. The research of Sharp Ratio in modern investment theory shows that the size of risk plays a fundamental role in determining the performance of portfolio. Risk-adjusted return rate is a comprehensive index that can consider both return and risk, in order to eliminate the adverse impact of risk factors on performance evaluation. Sharp ratio is one of the three classical indicators that can take into account both return and risk. There is a conventional feature in investment, that is, the higher the expected return of the investment target, the higher the volatility risk that investors can tolerate; on the contrary, the lower the expected return, the lower the volatility risk. Therefore, the main purpose of rational investors to choose investment targets and portfolios is: to pursue the maximum return under the fixed risk, or to pursue the minimum risk under the fixed expected return.

P-ability. Generally speaking, as one of the investment abilities of fund managers, the fund's earnings will be improved with the improvement of fund managers' stock picking ability. The higher the stock picking ability, the more funds can be selected that perform better than the benchmark in the future. Here we define this ability as p-ability.

Anti-risk ability. The anti-risk ability of the fund is mainly compared by its loss frequency and average loss margin. Different loss frequencies and loss ranges reflect the operating style of fund managers to a certain extent. Only a fund with a better balance of loss frequency and loss range can have a strong ability to resist risks and help investors achieve long-term and sustained investment returns.

**6、Data Source**

We are planning to have around 9 variables in regression. In order to ensure the regression accuracy, we need the degree of freedom (sample size - number of variables -1) to be greater than 30. We pick the rolling weekly change data from June 2015 to December 2017 with 629 rows for each fund. So we have 3139 records of data to analysis.

From the official websites of Fund and Tiantian Fund Company, we can find the data required for the regression. The websites are shown as below:

http://www.efunds.com.cn/

http://www.cmfchina.com/

http://www.dfham.com/

http://www.jsfund.cn/

http://fund.eastmoney.com/

**6.1 Data Limitation**

First, due to limited time and data accessibility, we were not able to collect whatever we think is important. For instance, the experience of managers, we cannot use data to evaluate it.

Second, we just select the data from June 2015 to December 2017, in practical application, we will pick data for 3 years.

Third, some funds are established in 2015, some are established in 2012, different time may affect the performance of the fund.

Forth, some variables don't change weekly, such as fund manager's ability to choose time, stock picking ability of manager, anti-risk ability, ect.

**7、RESULT & INTERPRETATION**

**7.1 Regression Result & Model Evaluation**

As mentioned previously, we will set Fund rate change percent as dependent variable Y. All independent variables are listed as following.

|  |  |
| --- | --- |
| **Variables** | **Description** |
| SH composite index∆% | Weekly change percent of Shanghai composite index |
| M2∆% | Weekly change percent of M2 |
| Fund type 1 | Dummy variable, equity fund shown as(Fund type 1=1, type 2=0), bonded fund shown as (Fund type 1=0, type 2=1), Hybrid fund shown as (Fund type 1=0, type 2=0) |
| Fund type 2 |
| S-deviation | Standard deviation, extent to which the fund may change |
| Sharpe Ratio | One kind of standardized indicators of fund performance evaluation |
| P-ability | stock picking ability |
| Anti-risk ability | The ability of the fund compared by its loss frequency and average loss margin |

The sample size is 3139 by selecting 5 funds randomly. And we set the level of significance at 0.05, so we’ll drop the variables with P-value higher than 0.05 in below analysis.

**Step1:** We begin by calculating the correlation matrix shown below. It shows the relationship between each of the independent variables and the dependent variable. It helps to identify which independent variables that are more closely related to the dependent variable (Fund rate∆%)。

The output from minitab is as below:

|  |
| --- |
| **Correlation: Fund rate∆%, SH composite index∆%, M2∆%, Fund type 1, Fund type 2, S-deviation, ...**  Fund rate∆% SH composite index∆% M2∆% Fund type 1  SH composite index∆% 0.669  M2∆% -0.095 -0.100  Fund type 1 0.029 0.001 0.003  Fund type 2 0.003 0.001 0.002 -0.409  S-deviation -0.001 0.009 -0.004 0.601  Sharpe Ratio -0.008 -0.003 0.010 -0.402  P-ability 0.064 -0.002 -0.004 0.082  Anti-risk ability 0.035 -0.001 -0.002 -0.209  Fund type 2 S-deviation Sharpe Ratio P-ability  S-deviation -0.674  Sharpe Ratio **0.928 -0.784**  P-ability 0.050 -0.456 0.244  Anti-risk ability **0.730 -0.882 0.812**  0.571  Cell Contents: Pearson correlation |

If the correlations among independent variables exceeding an absolute value of 0.7, multicollinearity would be a concern. Unfortunately from above output, we can see there exists multicollinearity issue between Fund type 2, P-ability and Anti-risk ability, between S-deviation, Sharpe Ratio and Anti-risk ability, between Sharpe Ratio and Anti-risk ability because their correlation value are more than 0.7 which is marked in red color in above chart.

Even though we know multicollinearity issue exists among Fund type 2, S-deviation, Sharpe Ratio and Anti-risk ability, it is difficult to decide which variable should be dropped at this moment, so we’ll further check on this after the regression model done.

**Step2:** Based on the analysis in Section 5 related to variables and & data, we use backward elimination regression procedures and put all related variables in our regression first, then drop the variables with P-value higher than 0.05 one by one. Below is the result of the first regression model with 8 variables.



Below minitab result for you to get more information.

|  |
| --- |
| **Regression Analysis: Fund rate∆% versus SH Market In, M2∆%, Fund type 1, Liability-ty, ...**  Analysis of Variance  Source DF Adj SS Adj MS F-Value P-Value  **Regression** 8 1.47454 0.18432 330.45 **0.000**  SH Market Index∆% 1 1.40205 1.40205 2513.63 0.000  M2∆% 1 0.00197 0.00197 3.53 0.060  Fund type 1 1 0.00200 0.00200 3.58 0.058  Liability-type 1 0.00060 0.00060 1.07 0.301  S-deviation 1 0.00292 0.00292 5.24 0.022  Sharpe Ratio 1 0.00400 0.00400 7.17 0.007  P-ability 1 0.00493 0.00493 8.84 0.003  Anti-risk 1 0.00316 0.00316 5.67 0.017  Error 3130 1.74585 0.00056  Total 3138 3.22039  Model Summary  S R-sq R-sq(adj) R-sq(pred)  0.0236174  **45.79%**  45.65% 45.30%  Coefficients  Term Coef SE Coef T-Value P-Value VIF  Constant -0.0332 0.0098 -3.39 0.001  SH composite index∆% 0.5891 0.0117 50.14 0.000 1.01  M2∆% -0.1133 0.0603 -1.88 0.060 1.01  Fund type 1 -0.0065 0.0034 -1.89 0.058 15.94  Fund type 2 0.0054 0.0053 1.03 **0.301**  24.83  S-deviation 0.526 0.230 2.29 0.022 65.16  Sharpe Ratio -0.0141 0.0053 -2.68 0.007 15.83  P-ability 0.0001 0.000037 2.97 0.003 3.29  Anti-risk ability 0.0003 0.0001 2.38 0.017 64.90  Regression Equation  Fund rate∆% = -0.03323 + 0.5891 SH composite index∆% - 0.1133 M2∆%  - 0.00650 Fund type 1 + 0.00543 Fund type 2 + 0.526 S-deviation  - 0.01408 Sharpe Ratio + 0.000111 P-ability + 0.000346 Anti-risk ability |

We conduct the global hypothesis test. Here we check to see if any of the regression coefficients are different from 0. We use the .05 significance level.  
H0: β1=β2=β3=β4=β5=β6=β7=β8=0

H1: Not all the βi are 0.  
The p-value is 0.000. Because the p-value is less than the significance level, we reject the null hypothesis and conclude that at least one of the regression coefficients is not equal to zero.

Based on the results of testing P-value of each independent variable, we conclude that linear relationship between Fund type 2, M2∆% and Fund rate∆% is insignificant since the P-value is higher than 0.05, so we’ll drop Fund type 2 first as its P-value is the highest.

**Step3:** After dropping Fund type 2, we re-run the multiple regression as below.

|  |
| --- |
| **Regression Analysis: Fund rate∆% versus SH Market In, M2∆%, Fund type 1, S-deviation**  Model Summary  S R-sq R-sq(adj) R-sq(pred)  0.0236176 45.77% 45.65% 45.33%  Coefficients  Term Coef SE Coef T-Value P-Value VIF  Constant -0.04008 0.00724 -5.54 0.000  SH composite index∆% 0.5887 0.0117 50.13 0.000 1.01  M2∆% -0.1125 0.0603 -1.87  **0.062**  1.01  Fund type 1 -0.00912 0.00232 -3.93 0.000 7.29  S-deviation 0.707 0.150 4.71 0.000 27.72  Sharpe Ratio -0.00947 0.00279 -3.40 0.001 4.44  P-ability 0.000087 0.000029 3.01 0.003 1.94  Anti-risk ability 0.000465 0.000090 5.18 0.000 24.69 |

Based on the above results, we further dropped M2∆% as its P-value is higher than the level of significance, 0.05.

**Step4:** Re-run multiple regression based on conclusion from Step 3, dropping M2∆%, the result is as below.

|  |
| --- |
| **Regression Analysis: Fund rate∆% versus SH Market In, Fund type 1, S-deviation, ...**  ….  Model Summary  S R-sq R-sq(adj) R-sq(pred)  0.0236270 **45.71%** **45.60%** 45.31%  Coefficients  Term Coef SE Coef T-Value P-Value VIF  Constant -0.04149 0.00720 -5.76 0.000  SH composite index∆% 0.5908 0.0117 50.54 0.000 1.00  Fund type 1 -0.00929 0.00232 -4.00 0.000 7.28  S-deviation 0.717 0.150 4.78 0.000 27.67  Sharpe Ratio -0.00963 0.00279 -3.46 0.001 4.44  P-ability 0.000087 0.000029 3.01 0.003 1.94  Anti-risk ability 0.000472 0.000090 5.26 0.000 24.65  Regression Equation  Fund rate∆% = -0.04149 + 0.5908 SH composite index∆% - 0.00929 Fund type 1  + 0.717 S-deviation - 0.00963 Sharpe Ratio + 0.000087 P-ability  + 0.000472 Anti-risk ability |

Now the P-value of each variable is less than 0.05, but the R square is less than 50%, it means the linear relationship between all the independent variables and dependent variable is significant, but the regression model is not good enough to present the relationship between the independent variables and dependent variable, so we’ll do transformation in next step.

**Step5:** Make transformation.

Before making transformation, we make a statistics analysis on the number of unique value of all variables, and we’ll do transformation for the variables with more unique values first. The result is as following.

|  |  |
| --- | --- |
|  | Num of unique values |
| Fund rate∆% | 2852 |
| SH composite index∆% | 630 |
| Fund type 1 | 2 (only 0,1) |
| S-deviation | 14 |
| Sharpe Ratio | 14 |
| P-ability | 4 (only 50,60,80,100) |
| Anti-risk ability | 4 (only 20, 50,70,90) |

So we’ve transformed Fund rate∆% and SH composite index∆% into log absolute value of fund rate∆% (we will call it Log-ABS-Fund rate∆% for short in this paper) and log absolute value of SH composite index∆%(Log-Abs-SH composite index∆% for short).

**Step6:** We run multiple regression after transformation, results as below.

|  |
| --- |
| **Regression Analysis: Log-ABS-Fund rate∆% versus Log-Abs-SH composite index∆%, Fund type 1, S-deviation, ...**  ….  Model Summary  S R-sq R-sq(adj) R-sq(pred)  0.439822 66.39% 66.33% 66.24%  Coefficients  Term Coef SE Coef T-Value P-Value VIF  Constant 1.674 0.134 12.50 0.000  Log-Abs-SH composite index∆% 0.3169 0.0147 21.50 0.000 1.05  Fund type 1 0.8820 0.0440 20.03 0.000 7.55  S-deviation -19.08 2.85 -6.69 0.000 **28.86**  Sharpe Ratio -0.2562 0.0521 -4.92 0.000 4.47  P-ability 0.001773 0.000536 3.31 0.001 1.94  Anti-risk ability -0.02540 0.00170 -14.92 0.000 **25.62**  Regression Equation  Log-ABS-Fund rate∆% = 1.674 + 0.3169 Log-Abs-SH composite index∆% + 0.8820 Fund type 1 - 19.08 S-deviation  - 0.2562 Sharpe Ratio + 0.001773 P-ability - 0.02540 Anti-risk ability |

The adjusted R square is 66.33%, which means 66.33% of Log-Abs-Fund rate **∆%** can be explained by the changes of Log-Abs-SH composite index∆%, Fund type 1, S-deviation and Sharpe Ratio, P-ability and Anti-risk ability.

But we found VIF value for S-deviation and Anti-risk ability is more than 10. It means multicollinearity exists among independent variables.

If we drop S-deviation and Anti-risk ability, and re-run multiple regression, the output result is as below:

|  |
| --- |
| **Regression Analysis: Log-ABS-Fund rate∆% versus Log-Abs-SH composite index∆%, Fund type 1, Sharpe Ratio, P-ability**  Analysis of Variance  Source DF Adj SS Adj MS F-Value P-Value  Regression 4 1112.74 278.186 1263.78 0.000  Log-Abs-SH composite index∆% 1 111.84 111.842 508.09 0.000  Fund type 1 1 198.28 198.280 900.77 0.000  Sharpe Ratio 1 330.99 330.991 1503.67 0.000  P-ability 1 23.20 23.200 105.40 0.000  Error 3134 689.86 0.220  Total 3138 1802.61  Model Summary  S R-sq R-sq(adj) R-sq(pred)  0.469172 61.73% **61.68%** 61.60%  Coefficients  Term Coef SE Coef T-Value P-Value VIF  Constant 0.6294 0.0447 14.07 0.000  Log-Abs-SH composite index∆% 0.3466 0.0154 22.54 0.000 1.00  Fund type 1 0.5719 0.0191 30.01 0.000 1.24  Sharpe Ratio -1.1675 0.0301 -38.78 0.000 1.31  P-ability -0.004439 0.000432 -10.27 0.000 1.11 |

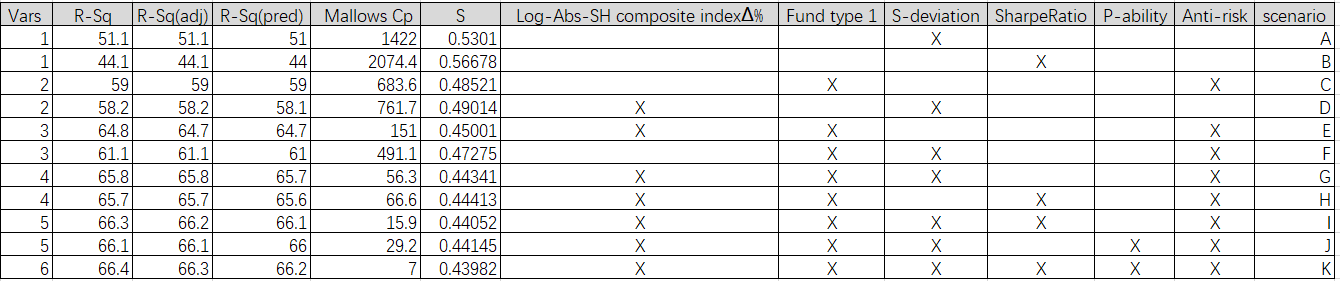
We already got a somewhat satisfying result so far. Because adjusted R square is 61.68% and P value is less than 0.05 and VIF is less than 10. But it cannot be ignored that adjusted R square decreased from previous 66.33% to 61.68%. We may drop too many relevant variables. In order to obtain a bigger adjusted R square, we use best-subset regression instead of removing S-deviation and Anti-risk ability directly.

**Step 7:** Run best-subset regression without removing S-deviation and Anti-risk ability

With best-subset regression, we look at the best model using one independent variable, the best model using two independent variables, the best model with three, and so on. The purpose is to find the model with the largest R2 value, regardless of the number of independent variables. Also, each independent variable does not necessarily have a nonzero regression coefficient. Since each independent variable could either be included or not included, there are 2k-1 possible models, where k equals to the number of independent variables. The advantage of the best-subset method is that we can see more combinations of independent variables thus getting the best one.

Below is the output from Best Subsets Regression with ‘'Log-ABS-Fund rate∆%'’ as response and with 'Log-Abs-SH composite index∆%', 'Fund type 1', 'S-deviation', SharpeRatio ,'P-ability' ,'Anti-risk' as free predictors. (Since original output from minitab has a mess format, we re-format as below table).

Within below table, for each best model from Best Subsets Regression result, we marked as scenario A/B/C/D/E/F/G/H/I/J/K in the last column



Scenarios E/F/G/H/I/J/K become our penitential choice after best-subset regression analysis as their adjusted R square are more than 60%.

VIF checking result for scenarios E/F/G/H/I/J/K as following.

|  |  |  |
| --- | --- | --- |
| Scenario | R-Sq (adj) | VIF checking |
| E | 64.7 | Term Coef SE Coef T-Value P-Value VIF  Constant 1.0559 0.0370 28.56 0.000  Log-Abs-SH composite index∆% 0.3332 0.0147 22.60 0.000 1.00  Fund type 1 0.6868 0.0168 40.98 0.000 1.05  Anti-risk -0.0175 0.0004 -49.86 0.000 1.05 |
| F | 61.1 | Term Coef SE Coef T-Value P-Value VIF  Constant 2.007 0.125 16.00 0.000  Fund type 1 1.1695 0.0411 28.44 0.000 5.70  S-deviation -36.64 2.82 -12.98 0.000 24.52  Anti-risk -0.03592 0.00146 -24.55 0.000 16.37 |
| G | 65.8 | Term Coef SE Coef T-Value P-Value VIF  Constant 2.149 0.118 18.23 0.000  Log-Abs-SH composite index∆% 0.3065 0.0148 20.73 0.000 1.04  Fund type 1 1.0328 0.0391 26.40 0.000 5.87  S-deviation -26.29 2.70 -9.75 0.000 25.39  Anti-risk -0.03073 0.00139 -22.03 0.000 16.92 |
| H | 65.7 | Term Coef SE Coef T-Value P-Value VIF  Constant 0.8935 0.0405 22.04 0.000  Log-Abs-SH composite index∆% 0.3382 0.0146 23.22 0.000 1.00  Fund type 1 0.6192 0.0181 34.20 0.000 1.25  SharpeRatio -0.4293 0.0467 -9.20 0.000 3.52  Anti-risk -0.013085 0.000597 -21.93 0.000 3.09 |
| I | 66.2 | Term Coef SE Coef T-Value P-Value VIF  Constant 1.787 0.130 13.78 0.000  Log-Abs-SH composite index∆% 0.3161 0.0148 21.41 0.000 1.05  Fund type 1 0.9062 0.0435 20.84 0.000 7.35  SharpeRatio -0.3170 0.0488 -6.50 0.000 3.91  S-deviation -20.46 2.82 -7.25 0.000 28.23  Anti-risk -0.02451 0.00168 -14.56 0.000 24.98 |
| J | 66.1 | Term Coef SE Coef T-Value P-Value VIF  Constant 1.871 0.128 14.59 0.000  Log-Abs-SH composite index∆% 0.3105 0.0147 21.07 0.000 1.04  Fund type 1 0.9589 0.0413 23.21 0.000 6.60  P-ability 0.002704 0.000503 5.37 0.000 1.70  S-deviation -22.47 2.78 -8.10 0.000 27.17  Anti-risk -0.03027 0.00139 -21.76 0.000 16.98 |
| K | 66.3 | Term Coef SE Coef T-Value P-Value VIF  Constant 1.674 0.134 12.50 0.000  Log-Abs-SH composite index∆% 0.3169 0.0147 21.50 0.000 1.05  Fund type 1 0.8820 0.0440 20.03 0.000 7.55  SharpeRatio -0.2562 0.0521 -4.92 0.000 4.47  P-ability 0.001773 0.000536 3.31 0.001 1.94  S-deviation -19.08 2.85 -6.69 0.000 28.86  Anti-risk -0.02540 0.00170 -14.92 0.000 25.62 |

But only Scenario E and H passed VIF testing and scenario H has higher R square. So scenario H is the best of all candidate combination from output of best subsets.

**Step 8:** We re-run multiple regression for scenario H based on conclusion from step 7, result as below

|  |
| --- |
| **Regression Analysis: Log-ABS-Fund rate∆% versus Log-Abs-SH composite index∆%, Fund type 1, Sharpe Ratio, Anti-risk ability**  Analysis of Variance  Source DF Adj SS Adj MS F-Value P-Value  Regression 4 1184.43 296.108 1501.19 0.000  Log-Abs-SH  composite index∆% 1 106.38 106.384 539.34 0.000  Fund type 1 1 230.75 230.754 1169.87 0.000  Sharpe Ratio 1 16.70 16.703 84.68 0.000  Anti-risk ability 1 94.89 94.888 481.06 0.000  Error 3134 618.18 0.197  Total 3138 1802.61  Model Summary  S R-sq R-sq(adj) R-sq(pred)  0.444126 65.71% **65.66%**  65.59%  Coefficients  Term Coef SE Coef T-Value P-Value VIF  Constant 0.8935 0.0405 22.04 0.000  Log-Abs-SH  composite index∆% 0.3382 0.0146 23.22 0.000 1.00  Fund type 1 0.6192 0.0181 34.20 0.000 1.25  Sharpe Ratio -0.4293 0.0467 -9.20 0.000 3.52  Anti-risk ability -0.013085 0.000597 -21.93 0.000 3.09  Regression Equation  Log-ABS-Fund rate∆% = 0.8935 + 0.3382 Log-Abs-SH composite index∆%  + 0.6192 Fund type 1 - 0.4293 Sharpe Ratio  - 0.013085 Anti-risk ability |

From above output, P-value is less than 0.05, VIF is less than 10 and also adjusted R square is 65.66%. But we found there exists correlation between Sharpe Ratio and Anti-risk ability as the correlation value is more than 0.7 as below

|  |
| --- |
| **Correlation: Log-ABS-Fund rate∆%, Log-Abs-SH composite index∆%, Fund type 1, Sharpe Ratio, Anti-risk ability**  Log-ABS-Fund rate∆% Log-Abs-SH composite index∆% Fund type 1 Sharpe Ratio  Log-Abs-SH  composite index∆% 0.241  Fund type 1 0.558 0.001  Sharpe Ratio -0.664 0.019 -0.402  Anti-risk ability -0.634 -0.001 -0.209 0.812  Cell Contents: Pearson correlation |

**Step 9:** We run multiple regression equation for scenario E, the best alternative from step 7, result is as below.



To keep more information, we have minitab result as well.

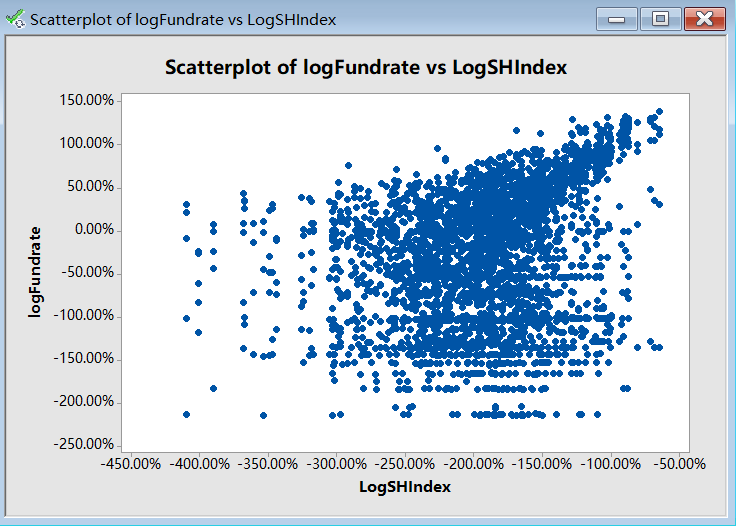
|  |
| --- |
| **Regression Analysis: Log-Abs-fund rate ∆% versus Log-Abs-SH composite index∆%, Find type1, Anti-risk ability**  Analysis of Variance  Source DF Adj SS Adj MS F-Value P-Value  Regression 3 1167.73 389.243 1922.06 0.000  Log-Abs-SH  composite index∆% 1 103.41 103.413 510.65 0.000  Fund type 1 1 340.02 340.024 1679.02 0.000  Anti-risk 1 503.47 503.469 2486.11 0.000  Error 3135 634.88 0.203  Lack-of-Fit 2506 562.54 0.224 1.95 0.000  Pure Error 629 72.33 0.115  Total 3138 1802.61  Model Summary  S R-sq R-sq(adj) R-sq(pred)  0.450015 64.78% 64.75% 64.69%  Coefficients  Term Coef SE Coef T-Value P-Value VIF  Constant 1.0559 0.0370 28.56 0.000  Log-Abs-SH  composite index∆% 0.3332 0.0147 22.60 0.000 1.00  Fund type 1 0.6868 0.0168 40.98 0.000 1.05  Anti-risk ability -0.0175 0.0004 -49.86 0.000 1.05  Regression Equation  Log-Abs-fund rate ∆% = 1.0559 + 0.3332 Log-Abs-SH composite index∆% + 0.6868 Fund type 1 - 0.0175 Anti-risk ability |

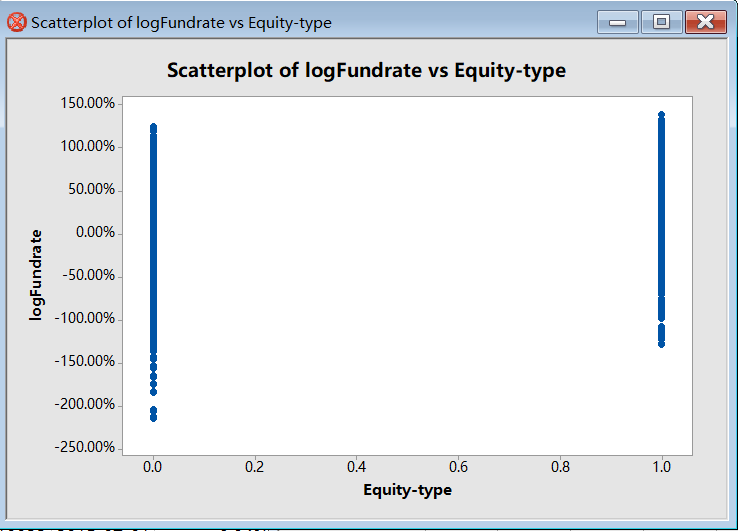
|  |
| --- |
| **Correlation: Log-Abs-fund rate ∆%, Log-Abs-SH composite index∆%, Fund type 1, Anti-risk**  Log-Abs-fund rate ∆% Log-Abs-SH composite index∆% Fund type 1  Log-Abs-SH composite index∆% 0.241  Fund type 1 0.558 0.001  Anti-risk -0.634 -0.001 -0.209  Cell Contents: Pearson correlation |

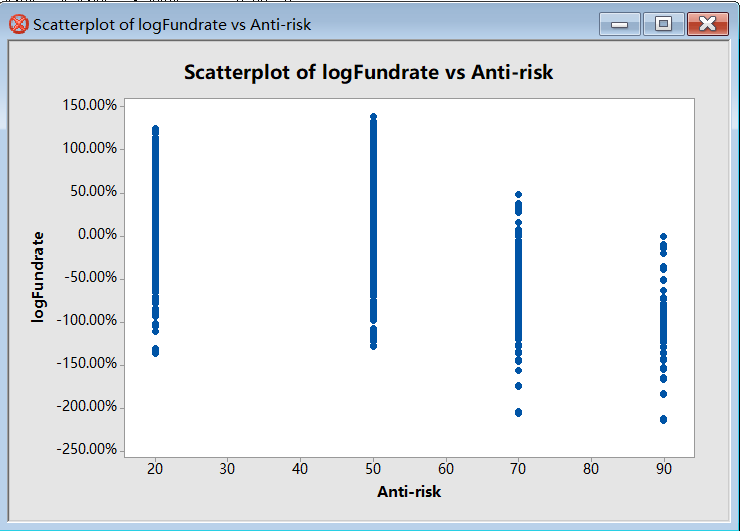
From above output, P-value (for both global test and individual test) is less than 0.05, VIF is less than 10 and also adjusted R square is 64.75%, and correlation values in correlation matrix is between -0.7 and 0.7

This is the final regression model we chose.

**Step 10:** Check scatter plot





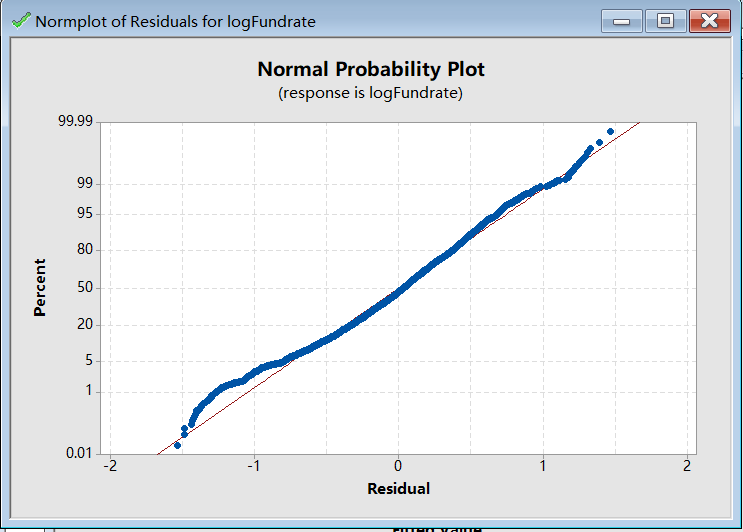


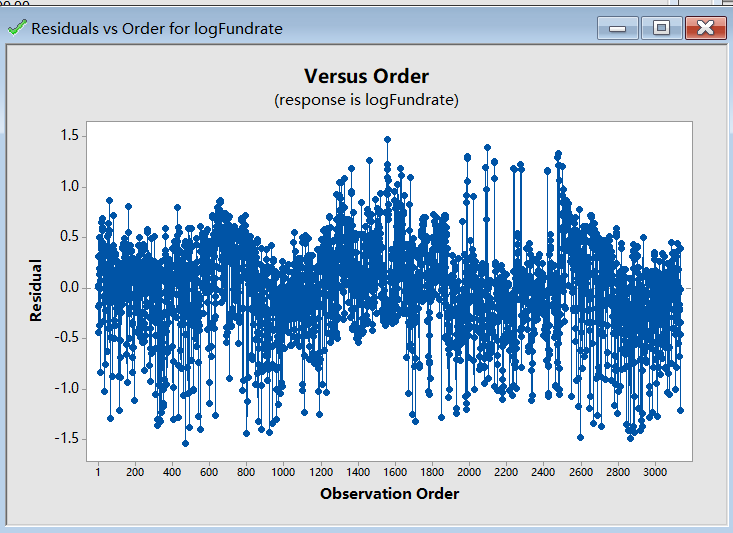
**7.2 Goodness of Fit**

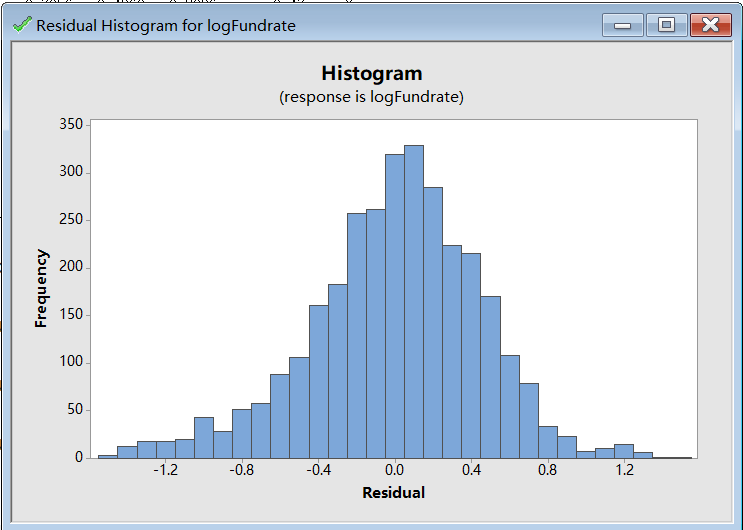
Log-ABS-Fund rate∆% = 1.0559 + 0.3332 Log-Abs-SH composite index∆% + 0.6868 Fund type 1 - 0.0175 Anti-risk ability

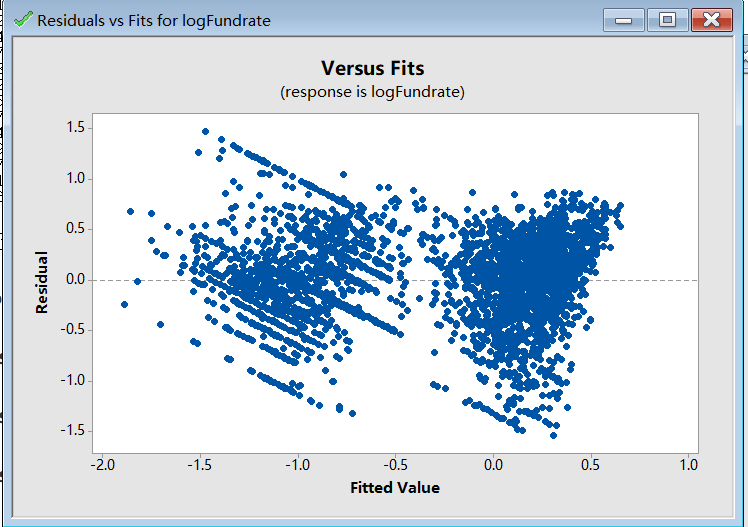
**7.3 Error Term Assumption Validation**

We use two kinds of plots to validate the error term of our regression: normal probability plot for “normal distribution”, and residual plots for “constant variance”.









**7.4 Interpretation of Coefficients**

Based on the final multiple regression equation: Log-ABS-Fund rate∆% = 1.0559 + 0.3332 Log-Abs-SH composite index∆% + 0.6868 Fund type 1 - 0.0175 Anti-risk ability, we can conclude as below.

1. For any additional one unit change in log-Abs- SH composite index∆%, Log-ABS-Fund rate∆% is expected to change by 0.3332 unit;
2. If the fund is equity fund (Fund type 1=1, type 2=0), Log-ABS-Fund rate∆% is expected to change by 0.6868 unit;
3. If the Anti-risk ability change by one unit, Log-ABS-Fund rate∆% is expected to change by 0.0175 unit;

# 8、 Conclusion

After several attempts, our team found that several key factors affecting the fund's rate of return are the Shanghai composite Index, the type of funds, and the anti-risk ability of funds. This is mainly because changes in the Shanghai composite index will affect the overall investment environment and may affect the rate of return of certain funds. However, the relationship between the rate of return of the equity fund and the Shanghai composite index depends mainly on whether the stock held by the fund is consistent with the ups and downs of the broader market.

When studying the type of fund, we found that the rate of return of fund changes of the two types of funds, bonded funds and hybrid funds, are almost indistinguishable. However, the change in the rate of return of equity funds is significantly higher than that of bonded type and hybrid type. Equity funds mainly invest in the stock market. Due to the factors affecting the stock market, the macroeconomic impact, the impact of the listed company's own operating conditions and other aspects of the impact of the equity funds have caused large fluctuations. Investors enjoy high returns and also bear higher investment risks.

We also found that the greater the anti-risk ability of fund, the smaller the fund's rate of return. When investors choose funds, they should also pay attention to the fund's anti-risk ability, mainly through the fund's loss frequency and average loss. Comparison. The different frequency of loss and the extent of loss reflect the operation style of the fund manager to a certain extent. Only funds that balance the loss frequency and the loss rate can have strong anti-risk ability and help investors achieve long-term sustained return on investment.

In real life, we often look at the fund manager's ability to choose time and ability to choose stocks, but from the funds studied by our group, the fund's rate of return has no significant relationship with these two variables.

In this regression analysis, our team relies more on random attempts on the transform of variables. In order to continue the research of regression models in the future, we should continue to learn effective methods of variable transform to fit better regression equations.