

# The Application of Robust Statistics to Stock Portfolio Problem

Li Xiongying<sup>[1]</sup>, Hong Jiahao<sup>[1]</sup>, Wang Binhui<sup>[1,2]</sup>

(1.School of Economics, Jinan University, Guangzhou 510632,China

2.School of Management, Jinan University, Guangzhou 510632,China)

**Key words:**portfolio; outlier; Fast-MCD; robust regression

**Abstract:**Portfolio theory is used to measure the expected return and risk on the basis of the history data of security return ratio, but in fact there is always excessively high or low return ratio caused by some short-term fundamental good or bad news in the history data of return ratio. We introduce the robust statistic idea into the portfolio theory in this paper, thus reduce outliers' influence on portfolio decision in the history data of return ratios, bring back the portfolio on its long-term investment value track. For the classic Markowitz mean-variance model and Sharp's single index model, we focused on the robust estimate method Fast-MCD and robust regression method, and apply them to solution processing in the portfolio model and obtained good results.

## Introduction

Robust statistic method is the statistic one with robust property including two sides (Wang binhui, 2007, 2015). One is with the characteristic of anti-disturbance, that means the method still keep good statistic performance when actual model differs little from theoretical assumption, the other is the estimate performance can still be acquired and not be destructively influenced when actual model differs much from theoretical assumption (Huber, 1994). The former side means one robust statistic method must perform well in assumed model and around, this guarantee that the statistic model is approximate correct and this approach to the desirable conclusion is the best or nearly when there are few of outliers among the data. The latter side means that some bad cases could be prevented, such as that robust statistic method could not perform poor or lead to completely wrong conclusion when the assumed model differs much from the fact or there are many outliers on database.

In this article, we contribute some minor innovation points. The robust statistic method is introduced into the portfolio field. On basis of the distinctive solution procedures of two portfolio models, we try to find the model solutions by two different portfolio made up of kinds of stocks. Thus it demonstrates the feasibility and necessity of studying portfolio theory on robust statistic method.

## Literature review

In nineteen eighty-one, Huber published another statistic book named as *robust statistic* in which he defined the robust statistic formally, robust statistic theory just grew up until now. Since that, research on robust statistic progressed much further. At home, Chen Xiru (2002) & Zhao Lincheng (2005) from USTC and Li Guoying (1994) & Shi Peide (1994) from CAS made a series of fruitful research achievements at moment estimate with large sample property in the linear, non-linear, partial-linear model and on robust projection pursuit method. Guo Yafan (2007) & Wang binhui (2006, 2007) and the else bore fruits at the respect of constructing robust statistic and

reforming robust statistic method. On board, researchers focused on constructing multivariate location and scatter, high break-down point and high-efficiency estimate in linear regression and test's break-down property. Since robust statistic has extensive field, it could progress further in terms of classic statistic method in case the fact deviates from the assumption, so it become necessary to used robust statistic method to improve classic one.

### Fast-MCD robust estimate model

In nineteen eight-four, Rousseeuw suggested minimum covariance determinant as multivariate robust estimate method, but limited to its complicate algorithm and computing technology, this method did not prevail even though it had strong robustness. After that, Rousseeuw & Van Driessen (1999) improved minimum covariance determinant and suggested Fast-MCD, sped up computing largely. We will estimate robust expected return ratio vector and covariance matrix on basis of Fast-MCD.

Fast-MCD constructs on robust covariance matrix estimator by iteration and mahalanobis distance. This progress can be as follows: on a matrix  $X_{n \times p}$  with  $p$  lines and  $n$  columns, i.e. return ratio data of  $p$  pieces of stocks in  $n$  periods, draw  $h$  samples and compute its mean  $T_1$  and covariance matrix  $S_1$ , then reckon mahalanobis distance from  $n$  samples to their center  $T_1$  by the formula  $d_1(i) = \sqrt{(x_i - T_1)^T S_1^{-1} (x_i - T_1)}$ , choose the smallest  $h$  distances, get the sample mean  $T_1$  and covariance matrix  $S_1$  by these  $h$  samples, it can be proved that there is  $\text{delta}(S_1) \leq \text{delta}(S_2)$ , they will equate if and only if  $T_1 = T_2, S_1 = S_2$ . Likewise, the iteration in this process go on and on until  $\det(S_m) = \det(S_{m-1})$ , this process is called C-step.

### Empirical analysis

We try to select stocks from various industries and regions in this article in order to make sample stocks more representative. Stock sample range from Jan., 2006 to Dec., 2014. The reason why we choose this time interval is that RMB deposit interest rate and deposit reserve ratio were adjusted for times. Altogether with several times of important policy adjustment to stock stamp tax and influence of short-term situation home and abroad, stock market went independently for times in this period. All this lead the date of return ratio to contain some outliers more or less. At a long-term viewpoint, this will misguide us to some extent to select portfolio. This paper's purpose aims at reducing outlier's influence through robust statistic method. On this consideration, we choose six stocks as the researched object in this article, see Table 4-1 in detail:

Table4-1 sample stocks

Item No.	code	Stock name
1	000707	Shuanghuan Science and Technology Stock
2	002024	Suning Commerce Group
3	600383	Gemdale
4	600480	Linyun industry
5	600759	Zhenghe Chemicals
6	600986	Keda Group

#### 4.2 Empirical test

At first, we reckon the skewness and kurtosis of monthly return ratio of six stocks in sample range, see the table 4-2. We find that each stock's skewness and kurtosis deviate apparently from normal distribution such as Zhenghe chemicals has a far deviation with kurtosis up to 37.44 and skewness up to 5.02. So we infer that there must be outliers among stock samples, similarly other stocks perform. As mentioned before, when return ratio data distributes with the property of thin peak and thick tail, we need robust estimate method to reckon expected return and covariance matrix, somewhat avoid the influence of outliers on portfolio. Next we use R software to get expected return and covariance matrices of these six stocks by classic computing method and Fast-MCD method respectively.

Table 4-2 The skewness and kurtosis

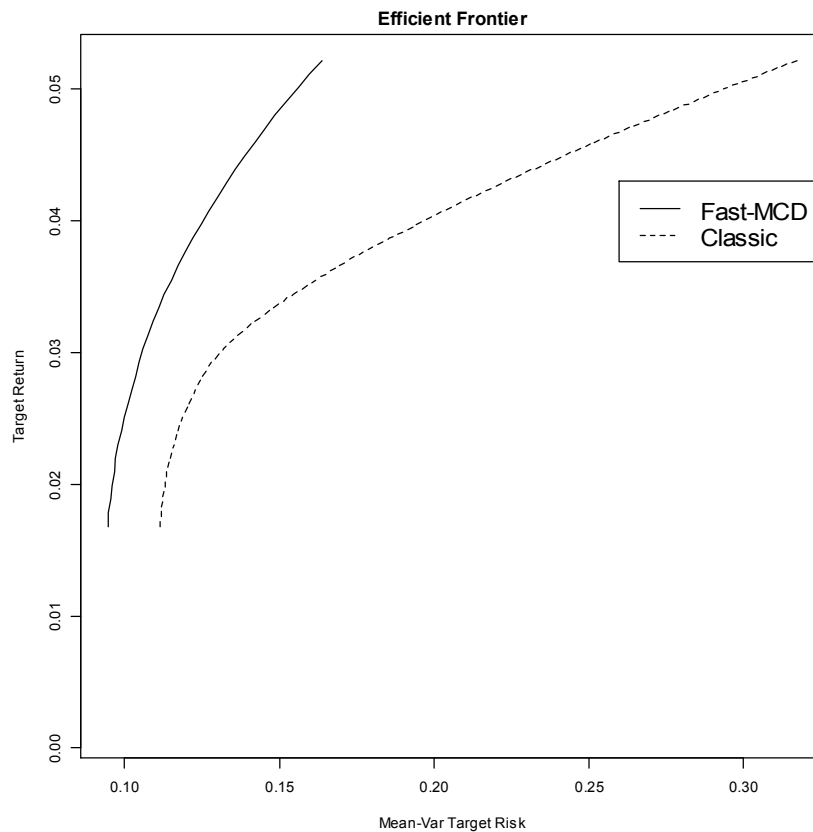
Stock name	skewness	kurtosis
Shuanghuan Science and Technology Stock	0.4728	3.253
Suning Commerce Group	0.9883	3.931
Gemdale	0.506	3.455
Linyun industry	0.7614	3.935
Zhenghe Chemicals	5.02	37.44
Keda Group	0.3408	3.157

We find through observation that a majority of expected returns differs much by Fast-MCD and classic method, sometimes gap between one times and several, even expected return ratios of Shuanghuan Science and Technology Stock, Suning Commerce Group, Linyun Industry become directly negative from positive. There is gap between variance matrix and covariance matrix because we must go on with searching the sample of which  $h$  value make Mahalanobis distance the smallest to iterate when using Fast-MCD method and weight differently the data from robust Mahalanobis distance, thus get robust expected return vector and robust covariance matrix which is desired to reduce the influence of outliers.

We put expected return vectors and covariance matrices by classic statistic method and robust statistic method into Markovits mean-variance portfolio model and find the programming solution by R software, get the efficient frontier of portfolio as shown in chart 4-1, vertical axis represents the expected return of portfolio, horizontal axis represents the standard error of, dotted line represents the portfolio's efficiency frontier of the expected return vector and covariance matrix in the model

on classic statistic method,full line represents portfolio's efficiency frontier of the expected return vector and covariance matrix in the model by Fast-MCD.We can find that with the same expected return it is obvious that optimum portfolio is at less risks by robust statistic method than by classic statistic method.

Chart 4-1 Figure Portfolio efficient frontiers



## Conclusion

In the paper we introduce robust statistic idea into portfolio field,try to reduce the influence on constructed portfolio by overhigh or too low return ratio caused by fundamental good news or bad news in the history data of return ratio,bring the portfolio back on its real value track.Through empirical analysis we can find it obvious that optimum portfolio is at less risk on robust estimate method than on classic method when at the same expected return ratio.This tells that the combination of robust statistic method and portfolio theory is feasible and necessary.

## References

- 1、 Guo yafan. Robust Statistic and Comparative Analysis of the Robustness of the Statistic[J]. Statistical Research,2007(9):82-85.
- 2、 Sun xianhua. Robust statistics in the application of economic indicators and its revelation[J]. Modern finance and economics,2003(12): 36-38.
- 3、 Xie zhenzhong.The single index model robust regression and empirical[J].Statistics and decision,2013(5):27-30.
- 4、 Wang binhui,Chen yifei.A Robust Principal Component Analysis Based on MCD Estimator and Its Empirical Study [J]. Application of Statistics and Management.2006(25):462-468.

- 5、Wang binhui. Robust Principal Component Analysis Method and its Application[J].Statistical Research.2007(24):72-76.
- 6、Xie zhenzhong.Application of Robust Statistical Inference in Mean-variance Model[J].Journal of toaghua normal university,2013(04):3-6.
- 7、Zhao liming. Robust statistics application in samples containing outliers [J]. The environmental monitoring management and technology,1992(4):58-59.