Mathematical Statistical Analysis of Financial Investment Industry Combining Resistance Algorithm under the Context of Internet

Yi Zhu*
Shoreline Community College, 98133

*E-mail: Zoe9392@163.com

Abstract. The reason why some mathematical approaches fail to achieve ideal results in social and economic statistics, especially in financial statistics, is that outliers and exceptions in the sample data lead to the fact that the calculation results cannot reflect the internal relationship of variables correctly and thus deviate from the practical economic significance. The resistance algorithm introduced in this paper can be used to solve these problems in many economic fields and achieve an excellent effect, especially in the financial, statistical analysis field. In this paper, the resistance algorithm is used to analyze the relationship of the liquidity preference with the income and interest rate. Based on the analysis of the annual network data in recent years, the effect of interest rate and liquidity preference is analyzed in this paper. The economic theory is combined with the actual situation in China for analysis. It is believed that a proper economic explanation can be obtained based on the application of the resistance algorithm proposed in this paper.

Keywords: Resistance Algorithm, Financial Investment Industry, Mathematical Statistical Analysis, Probability, network data

1. Introduction

The empirical study suggests that the economic significance of some statistical results obtained by mathematical approaches is often puzzling, and some of them are even contradictory and difficult to justify themselves. For example, in the quantitative measurement of scientific and technological progress, to determine the elasticity of capital and labor $\alpha, \beta$, it is usually based on the famous "r.m.so low growth rate equation". The output elasticity coefficient is calculated based on the regression analysis. However\cite{1-2}, it has been pointed out in many relevant data that the results of regression analysis often exceed the interpretation range of the economic significance of elasticity of production (based on many studies at home and abroad, it is reasonable to conclude that $\alpha$ should be between 0.2-0.4; $\alpha$ in China is between 0.2-0.3). Even when the value of the elasticity coefficient is less than zero, it is still unsatisfactory. Hence, what is the cause of this kind of mistake? The social-economic phenomenon is considered extremely complex by some people. It is impossible to try to develop a universal theory by mathematical method. Some people think that the quality of statistical data cannot be guaranteed the main reason why the mathematical method is rarely successful in economic research.
The statistical inference theory and technique of mean regression and quantile regression cannot be modified directly or simply for mode regression, which leads to great difficulties in the statistical inference of the financial development income distribution model. The development of the mode regression statistical inference theory is relatively slow. Currently [5-6], the financial development income distribution model has been concerned by scholars. It is extensively applied in the fields of urban transportation, meteorology, economics, etc. In addition, great progress has been made in the theoretical research on the financial development income distribution model related to nonparametric, semi-parametric model regression estimation, Bayesian model estimation, and mode regression variable selection.

In fact, not only regression analysis often fails in the measurement of scientific and technological progress, but also many other mathematical approaches such as principal component analysis, factor analysis, and cluster analysis are not ideal in economic research. In my opinion, a common and important reason is that they are all non-resistant algorithms. That is, once outliers (outliers) appear in the sample data, the calculation results will not correctly reflect the internal relationship of variables, which will deviate from the practical economic significance, resulting in a variety of difficult to explain situations.

2. Resistance algorithm
In exploratory data analysis, in a certain sense, a mature method should not be affected by individual serious errors, or its response to serious errors should be insensitive. This property is called resistance. To measure the resistance, we introduced the concept of “damage” as follows: for an estimator based on the sample $\xi$ (capacity $n$), it is referred to as $n/K$'s destruction bound. If $K$ is the maximum number of estimators that can keep the estimate bounded when any $k$ data in the sample changes without limit. For the algorithm involving multiple estimators, the minimum value of the destruction bound of each estimator indicates the destruction bound of the algorithm.

The author thinks that the introduction of a resistance algorithm in the mathematical, statistical analysis of the modern financial investment industry, especially in the field of social economy, is very significant and necessary. We analyze the following reasons. For an indicator, in the field of natural science, a group of sample data can be obtained through multiple experiments under the same (or almost the same) conditions, and then various mathematical statistics approaches can be used for processing and analysis. However, in social economy, other than a small part of the statistical indicators through sampling survey, they are obtained by the summary of grass-roots statements, which does not meet the requirement of “similarity” in “multiple human experiments”.

In addition, social and economic statistics are also affected by many human factors, such as the intervention of enterprise leadership, the quality of statisticians, and even the intervention of relevant departments. For the experimental data of natural science, although it is also affected by various factors, only the “experimenter” is a human factor. Therefore, in this sense, the reliability of socio-economic statistical data is worse than that of natural science experimental data, with a higher probability of data “distortion”. Hence, it is insufficient to process and use the social and economic statistics data and completely copy the traditional mathematical statistics method, and the “insensitive to abnormal value data” of resistance algorithm is just a good way to “process” the social and economic statistical data.

To construct the algorithm of resistance, we should first derive the statistics of resistance. Next, we need to analyze which statistics are resistant. For a set of samples as follows

$$M(X) = \begin{cases} X_{k+1}, & n = 2k + 1 \\ \frac{1}{2}(X_i + X_{k+1}), & n = 2k \end{cases}$$

(1)

To explore the dependence of the variable $Y$ on $X$, the scatter plot corresponding to the data $(X_i, Y_i)$ is usually drawn first. If the scatter plot is close to a straight line, fitting the straight line can summarize the dependence of $Y$ on $X$. If the scatterplot is an obvious curve, you can first linearize it with an appropriate transformation, and then perform straight-line fitting. Hence, the following
discussion always assumes that the data \( (X_i, Y_i) \) has or has been converted into a linear relationship.

\( X_i (i = 1, 2, \cdots, n) \) is first sorted in ascending order. It is assumed that \( X_1 \leq X_2 \leq \cdots \leq X_n \) divides all data points into left, center, and right groups on this basis so that the number of data points in each group is as equal as possible. In the case where \( X_i \) has no duplicate values, it can be assigned according to the rules in the following table 1:

**Table 1. Data allocation rules**

|       | n=3k   | n=3k+1 | n=3k+2 |
|-------|--------|--------|--------|
| Left  | k      | k+1    | k+1    |
| in    | k      | k+1    | k      |
| right | k      | k      | k+1    |

When there are duplicate values in \( X_i \), it will hinder the above allocation, because the same \( X_i \) value can only be divided into the same group. After all, the practical operation is more complicated in this case, which will not be discussed herein.

Determine the initial values of the slope and intercept of the fitted line. Note that the fitted straight line is \( \bar{Y} = a + bx \), where \( b \) represents the slope and \( a \) represents the intercept. Based on the three general points that have been obtained, determine the initial value according to the following equation:

\[
\begin{align*}
b_0 &= \frac{(Y_R - Y_L)}{(X_R - X_L)} \\
a_0 &= 1/3[(Y_L - b_0X_L) + (Y_M - b_0X_M) + (Y_R - b_0X_R)](0)
\end{align*}
\]

(2)

### 3. Example and result analysis

In this paper, the mathematical-statistical analysis of financial investment industry based on resistance algorithm is used to study the main influencing factors of the largest proportion of group income in the development of Internet Finance in China, and the results of mode regression empirical analysis are compared with the results of mean regression and quantile regression empirical analysis. The Shanghai stock exchange index is selected as the control variable to avoid the influence of economic factors caused by inadequate filtering on the empirical results. The core density of the two-time series after processing is shown in Figure 1. The results suggest that the mean values of the two-time series are close to 0, which shows that the influence of macroeconomic trend has been basically eliminated.

**Figure 1.** Kernel density estimation of online financial yield after algorithm processing

In fact, not only regression analysis fails in the quantitative measurement of scientific and technological progress, but many other mathematical approaches such as principal component analysis, factor analysis, and cluster analysis are also not ideal in economic studies. In my opinion, one common and important reason is that they are all non-resistant algorithms. In other words, once outliers (outliers) appear in the sample data, the calculation results will not correctly reflect the
internal relationship of variables, which will deviate from the practical economic significance, resulting in a variety of difficult to explain situations.

**4. Conclusion**

Three groups of resistance line algorithms are developed based on technologies in the early stage. These factors involve the grouping of (x, y) data by X and the summary of X and Y in each group. The general least-square regression analysis method is established under the assumption that both the dependent variable y and the independent variable x have observation errors. To make the model more precise and introduce x sampling mechanism, we have to assume that x also has observation errors. For the model under the assumption that both X and Y have observation errors, the least-square regression analysis method is no longer suitable for the straight-line fitting of data. The three-group resistance line fitting method is put forward in this context. In more complex cases, such as the linear fitting of multivariate observation data, experts have also recommended a median smoothing method with excellent resistance.

**References**

[1] António Canelas, Rui Neves, & Nuno Horta. (2013). A sax-ga approach to evolve investment strategies on financial markets based on pattern discovery techniques. Expert Systems with Applications, 40(5), 1579 – 1590.

[2] Ananda Swarup Das, Sameep Mehta, & L.V. Subramaniam. (2017). Annofin – a hybrid algorithm to annotate financial text. Expert Systems with Applications, 88, 270-275.

[3] L.-L. Zhang, H.-X. Wang, & W.-R. Fan. (2011). Image reconstruction algorithm based on 1-norm for electrical resistance tomography. Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/Journal of Tianjin University Science and Technology, 44(9), 786-790.

[4] Seungkwang Lee, Dooho Choi, & Yongje Choi. (2014). Improved shamir’s crt-rsa algorithm: revisit with the modulus chaining method. Etri Journal, 36(3), 469-478.

[5] Xide Cheng, Baiwei Feng, Zuyuan Liu, & Haichao Chang. (2018). Hull surface modification for ship resistance performance optimization based on delaunay triangulation. Ocean Engineering, 153, 333-344.

[6] Travis Miranda Lakshmi, Aruldoss Martin, & Venkatasamy Prasanna Venkatesan. (2016). A genetic bankrupt ratio analysis tool using a genetic algorithm to identify influencing financial ratios. IEEE Transactions on Evolutionary Computation, 20(1), 38-51.