Based on Diversification Measure: Application in Portfolio Strategy of Chinese Biopharmaceutical Industry

Chaoquan Jiang[1], Lizhang Chen[1], Sipei Qin[1] and Tao Huang[1]

[1] School of Science, Beijing Jiaotong University, Haidian District, Beijing, 100044, China

First author's e-mail: cqjiang@bjtu.edu.cn

Abstract. We study and explore the application the latest diversification measure of portfolio based on Rao's Quadratic Entropy. We improve the previous model and propose an algorithm that can help to apply diversification measure to practice and improve the return performance of the portfolio model. In addition, our experiment focuses on the application of diversification measure to the portfolio strategy research in China's biopharmaceutical industry. The COVID-19 in 2020 is undoubtedly a global and declared health emergency. Currently, a lot of work is devoted to studying the impact of COVID-19 on the stock market. Our research on investment strategy from the perspective of diversification measure and portfolio optimization is innovative to some extent.

1. Introduction

Diversity Measure is a kind of description of diversity and it is widely used in various fields. In 1952, Markowitz proposed the principle of portfolio diversification and Mean-Variance Model, which is known as the beginning of modern portfolio theory[4]. In 1960, Whittaker first recognized three different components of species diversity[12]. In 2005, Diversity Measure is also used in ensemble learning among the base classifiers[11]. In 1982, Rao proposed a widely applicable measure of diversification, the Rao-quadratic entropy, which has been widely used in energy management[10] and income inequality analysis[5]. Our main goal is to explore the application of Rao's Quadratic Entropy and compare its actual performance after improved by us.

In portfolio theory, a well-known saying is "Don't put all your eggs in one basket", diversification has become a consensus. Portfolio diversification is a risk management strategy in portfolio management. Portfolio theory has developed several classical portfolio theories based on the idea of diversification. Choueifaty[1] proposed the model of maximizing diversification (DR) in 2008, Markowitz's mean-variance model (MV) is also considered as a diversification measure to maximum the return of a portfolio for a given of uncertainty level.

It is not surprising to study portfolio theory and practice from the perspective of diversification, but there is a lack of adequate framework: how to measure portfolio diversification quantitatively, clearly with low computational cost? Markowitz asked a similar question in 1999, arguing that the problem is the lack of an adequate measure of diversification in relation to risk, the ability to distinguish between efficient and inefficient portfolios, and the ability to balance benefits and risks[4]. Koumou was inspired by Rao-quadratic entropy, and in his 2017 doctoral dissertation initially proposed a unified and consistent framework for diversification measure. Koumou describes the properties of *diversification measure* in a more comprehensive way, Rao's Quadratic Entropy has been used to measure the diversification of the portfolio [3], and a lot of researches have been done in theory and some classical portfolio diversification measures have been covered.

Diversification measure of portfolio is a popular research problem. However, portfolio diversification can be measured in theory, that does not mean the performance in practice. The goal of portfolio optimization is to obtain stable and high returns and to avoid risks at the same time. We propose to take the estimated return of stocks into account in the practical application, which is the weakness of diversification measure, to trade-off the return of portfolio and the diversification of portfolio. We improve the measure of diversification simply taking expected returns into account and add norm penalty wisely during optimization, which can select the sparse portfolio and promote the practical application of the diversification measure.

We selected 31 stock assets in the biopharmaceutical industry to conduct the experiment and also proved that our improvement has practical significance. In addition, we give some suggestion about how to apply diversification measure based on the portfolio optimization of China's biopharmaceutical industry during the new coronavirus pneumonia (COVID-19) in 2020. This paper can provide some reference for the research of diversification measure and investment strategy.

2. Diversification measure

2.1. Rao's Quadratic Entropy

Rao-quadratic entropy (RQE) is a diversification measure proposed by Rao in 1982[7]. The RQE taking into account the degree of dissimilarity within population. It is formulated by assuming that there are N classes of population P, and the probability of each class is P_i. In ecology, a quadratic function measure of population diversity is defined, which means the average dissimilarity：

$$H(P) = \sum_{i,j} d_{ij} P_i P_j \quad (1)$$

, where the d_{ij} measures the differences between class i and j. The larger the H, the greater the diversification of P. This measure has been studied by many scholars and has found a good application in various fields[8; 9; 13]. In addition to quadratic entropy, Shannon entropy, Simpson index can also measure dispersion.

2.2. RQE and Portfolio

The Rao-quadratic entropy is introduced by KOUMO into the portfolio field. Under certain conditions, the property of diversification measure can be satisfied and the degree of dissimilarity can be measured. We briefly introduce the RQE-based portfolio diversification framework: [5]

Suppose that there are N stock assets, the allocation proportion(weights) among the assets is: $w = (w_1, w_2, \ldots, w_N)^T$, and the function measure of the difference between the assetsi, j is denoted by d_{ij}, then a $N \times N$ matrix $D = \left(d_{ij}\right)_{i,j=1}^{N}$ is constructed. The portfolio diversification measure of N assets is obtained as follows:

$$Max\ H_D(w) = \frac{1}{2} w^T D w \quad (2)$$

$$st. w^T 1_N = 1$$

, where the diversification matrix D is Conditional Negative Definite and the $H_D(w)$ is concave, the model has the optimal solution： $w^* = \frac{D^{-1}1}{1^T D 1}$. We call the model MDM (Maximize Diversification measure).

MDM model explain the average dissimilarity of portfolio w and in order to promote its practical application, we consider the following simple, intuitive, basic measure of asset differences as a representative, to explore the actual performance.

$$d_{ij} = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}^2 \quad (3)$$

, where σ_i represents the variance of the return of stock assets i and σ_{ij}represents the covariance of the return rate of stock assets i and j .[2]

2.3. Markowitz Mean-Variance（MV）

Markowitz is known for mean-variance model, which minimizes the risk of an asset portfolio at a certain level of return. Markowitz assumes that investors are rational, and his model is based on the following assumptions: the return of stocks conforms to a certain probability distribution; the risk of stocks can be estimated by the variance of historical returns; In a certain income level, hope to obtain the lowest risk, and in a certain risk level, hope to obtain the highest income.

$$Min\ w^T \Sigma w \tag{4}$$

$$st. w^T 1 = 1, w^T \mu \geq \mu_k$$

, where μ_k is specified by the investor in accordance with his or her willingness and ability to assume uncertainty.[2]

Markowitz is also a benchmark model for our follow-up experiments, the formula(4) can be modified to the utility function $U(w)$ as follows：

$$Max\ \ U(w) = \tau w^T \mu - w^T \Sigma w \tag{5}$$

$$st. w^T 1 = 1$$

, where the $\mu = (\mu_1, \mu_2, \dots, \mu_N)^T$ represents the respect return of N stock assets estimated by historical daily return rate ; $\Sigma = (\sigma_{ij})_{i,j=1}^{N}$, σ_{ij} represents the covariance between the return on the assets i and j , estimated by historical daily return rate too; $w^T \Sigma w$ measure the risk of portfolio; $w^T \mu$ measures respect return; $\tau \geq 0$ is the degree of investor's aversion to risk , which is determined by the investor himself.

The significance of the model is the trade-off between risk and return. This model also serves as one of the benchmark models to be compared.

2.4. Weakness and solution of portfolio diversification

We focus on the practical application of this uniform dispersion metric. In our research, we find that this consistent dissimilarity metric D although clearly provides a framework for dissimilarity metrics, has two weakness, we've come up with a solution as follows.

2.4.1. A Weakness. The portfolio diversification model MDM is a good measure of portfolio diversification, but the number of assets considered is too large to allow investors to focus on a small number of stocks. Because of their limited energy to manage stocks, managing many stocks at the same time is unrealistic, which greatly hinders the practical application of diversification measure. Generally investors are advised to hold no more than 20 assets, and if some stock is from the same industry, they should hold fewer. In the language of optimization, the number of non-zero elements in the optimal solution is constrained by the model of optimal dispersion*[6]*：

$$w^* = \operatorname{argmax} H_D(w) = \frac{1}{2} w^T D w \tag{6}$$

$$st. \begin{cases} w^T 1_N = 1 \\ w_i \geq 0 \\ \sum_{i=1}^{N} 1_{\{w_i > 0\}} \leq n_w \end{cases}$$

The solution of the above-mentioned problem is a non-convex problem with great difficulties. Solution: we incorporate the respected return of stock assets into our model, and proposing a widely adaptable, computation-easy diversification measure of sparse weights. Referring to Yu-min Yen's work, we use the l_1 norm penalty method [13; 6] to solve the following problem and obtain the optimal weight sparse optimal weight w^*. We suggest that the norm penalty can be applied to other portfolio diversification measures and portfolio models in practice.[13]

$$w^* = argmin - \frac{1}{2} w^T D w + \lambda \left\| w \right\|_1 \tag{7}$$

$$st.w^T 1 = 1$$

, where λ represents the force of penalty, which is a common method in optimization theory. This method greatly reduces the computational complexity, at the same time getting a non-short (W > 0) portfolio.

2.4.2. Another Weakness.

The model only considers the diversification of the portfolio and ignores the expected return of the portfolio. In practice, when the correlation of stock returns is high in same industry, only focusing on the dissimilarity of stocks is inadequate.

We propose a trade-off between return and diversification to arrive at an eventual Maximize Diversify-Return model (MDR). We improved the MDM model based on the mean-variance model of Markowitz. According to Markowitz's hypothesis, the return rate of assets follows the normal distribution. We can estimate the expected return as $w^T\mu$, an improved MDR model is obtained：

$$Min\ -\frac{1}{2}w^T Dw - \phi w^T \mu + \lambda\left\|w\right\|_1 \quad (8)$$
$$st.w^T 1 = 1$$

, where ϕ means the level of chasing return, the meaning is similar to the level of aversion risk τ(5).

This norm-penalized equality constraint problem can be transformed into an unconstrained optimization problem in Joseph-Louis Lagrange's form：

$$w^* = \arg\min_{w,\gamma} L(w,\gamma) = -\frac{1}{2}w^T Dw - \phi w^T \mu + \lambda\left\|w\right\|_1 - \gamma(w^T 1 - 1) \quad (9)$$

, where $\gamma \in R$ is Lagrange multiplier.

Because the dissimilarity D between assets can calculate from different properties of assets according to the need, this model has great adaptability, and solves a part of weakness applying the diversification measure to practice.

2.5. Solution Algorithm

Considering that the MDR (7) is a quadratic programming problem with equality constraints and penalty term, we develop a coordinate descent algorithm to solve the problem. When $d_{ij} = \sigma_i^2 + \sigma_i^2 - 2\sigma_{ij}$, we can't make sure D is Conditional Negative Definite, the problem is transformed into the form as follows：

$$w^* = \arg\min_{w,\gamma} L(w,\gamma) = \frac{1}{2}w^T \Sigma w - w^T(\phi\mu + \sigma^2) + \lambda\left\|w\right\|_1 - \gamma(w^T 1 - 1) \quad (10)$$

We refer to CCD coordinate descent algorithm, Lasso linear regression algorithm, and improved coordinate descent algorithm [14; 6], and design a changed coordinate descent algorithm to solve this kind of problem, greatly improve the efficiency of the algorithm.

$$0 = \frac{\partial L}{\partial w_i} = \sigma_i^2 w_i + \sum_{j \neq i} \sigma_{ij} w_j - \left(\phi\mu_i + \sigma_i^2\right) - \gamma + \lambda\partial|w_i| \quad (11)$$

$$0 = \frac{\partial L}{\partial \gamma} = w^T 1 - 1$$

We introduce a *soft-threshold function*:

$$SH(u; v) = sign(u) \cdot \max(|u| - v, 0)$$

Then the above problems can be solved by our improved CCD coordinate descent algorithm, the similar details of algorithm please reference[13; 6]：

$$w_i = \frac{SH\left(\gamma + \phi\mu_i + \sigma_i^2 - \sum_{j \neq i} \sigma_{ij} w_j\, ; \lambda\right)}{\sigma_i^2} \quad (12)$$

At the same time, using w to solve out γ satisfying: $w^T 1 = 1$. The algorithm format for solving the problem is as follows:

Table 1. The Algorithm of CCD Descent

1	Chose Initial Value $\lambda, \gamma^0 \gg \lambda$
2	Random Initial $w_i^0 \in [0,1]$
3	$k \leftarrow 0$
	Repeat:
	$$w_i^{k+1} = \frac{SH\left(\gamma + \phi\mu_i + \sigma_i^2 - \sum_{j<i}\sigma_{ij}w_j^k - \sum_{j>i}\sigma_{ij}w_j^{k+1}; \lambda\right)}{\sigma_i^2}$$ $$\eta_i = \sum_{j\neq i}\sigma_{ij}w_j^{k+1} + \phi\mu_i + \sigma_i^2 - \lambda\cdot sign(w_i^{k+1})$$ $$\gamma^{k+1} = \left(\sum_{i:w_i\neq 0}\frac{1}{\sigma_i^2}\right)^{-1}\left(1 + \sum_{i:w_i\neq 0}\frac{\eta_i}{\sigma_i^2}\right)$$
	Until convergence.
4	Get best w^*

The coordinate descent algorithm can greatly improve the speed of the solution, and has a high application value for solving similar problems.

3. Evaluate model performance and asset preferences

To evaluate the performance of a portfolio, we use the rolling-window strategy [13] and set the time interval δ for rebalancing the portfolio weights, take the length $period$ of historical estimated window before the time t of each update weight. We calculate the dissimilarity matrix D between the assets in each parameter-estimate window. We get the number T of windows in the overall rolling-window strategy. Based on biopharmaceutical industry of *the China's Shenzhen Stock Exchange (SZSE)*, we consider stocks' daily closing price from January 1,2018 to January 1,2021 as our dataset. We calculate the optimal portfolio weights for each time period (window t) and rebalance the portfolio weights. The return of all candidate assets in window is vector R_t, and the return of the portfolio can be expressed as the weighted average of the return of the assets r_t:

$$r_t = \sum_i w_{t,i}R_{t+1,i} = w_t^T R_{t+1} \tag{13}$$

, where R_{t+1} is the return rate when the weights are rebalanced.

To measure portfolio outperformance, we use several parameters to measure model performance and the preference of model to assets, as shown in Table 2, taking into account multiple returns.

3.1. Performance of model

After performing the roll-window strategy over all time periods, the average return (*MR*), cumulative return (*CR*), the Sharpe ratio (*SR*) and the turnover ratio (*TOR*) are used to measure the performance of the model compared with two benchmark portfolio optimization model. Mean return (MR) and cumulative return (*CR*) are the most direct indicators to evaluating models; Sharpe ratio (SR) can measure the stability of the return, the larger the Sharpe ratio is, the more stable the return is; The turnover rate (*TOR*) of the whole strategy can evaluate the transaction cost, and we tend to lower the transaction cost under the same income.

In addition, we calculate the average of the number of nonzero weight (*PNA*) through the roll-window strategy. Suppose there are many stocks to choose, we only get several stocks ($w_i > 0$), which is our goal. If there are the same number of candidate stocks, the more the PNA is, the better the ability to select stocks is.

We propose to use the weight of holding stock i in all periods to measure the degree of preference for that asset in the whole strategy, we choose the ratio of the weight($0 \leqslant PDA \leqslant 1$) of holding stock i in a sliding window strategy. The larger and more frequently the investor holds the asset, the greater the preference for the stock. We supplement the average number of non-zero weights per window in the optimal portfolio to measure the ability of the model to select a small number of assets.

Table 2. The parameter of performance of model and preference of stock assets.

Parameter	calculate
Mean Return (MR)	$\bar{r} = T^{-1} \sum_t r_t$
Standard Deviation (SD)	$SD = \left((T-1)^{-1} \sum_t (r_t - MR)^2 \right)^{\frac{1}{2}}$
Sharp Ratio (SR)	$\mathrm{SR} = MR \times SD^{-1}$
End Cumulative Return (CR)	$ECR = \prod_t (1 + r_t)$
Turnover Rate of asset i (TOR_i)	$TOR_i = T^{-1} \sum_t \lvert w_{t+1,i} - w_{t,i} \rvert$
Whole Turnover Rate (TOR)	$TOR = \sum_i TOR_i$
Number of Nonzero Asset(PNA[13])	$PNA = T^{-1} \sum_{t,i} I_{\{w_{t,i} > 0\}}$
Preference Degree to Asset i (PDA)	$PDA_i = \left(\sum_t w_{t,i} \right)^{-1} \sum_t w_{t,i}$

4. Experiments and results

4.1. Background

The stock market has always been regarded as the barometer of the national economy, and the impact of unexpected events is often quickly transmitted to the stock market. The COVID-19 of 2020 is undoubtedly a global public health emergency. In the early days of the epidemic, the sudden outbreak of the epidemic increased investors' expectations of the pharmaceutical industry. The new pneumonia epidemic had a significant positive impact on the pharmaceutical industry, causing the rise in the stock prices of concepts such as face masks and medical devices, but the effect is short-term [15]. In the second half of 2020, the epidemic situation in China is basically under control, but due to the recurrence of the epidemic situation, the performance of the vaccine-related concept stocks is strong, and the stock price of the bio-pharmaceutical industry fluctuates obviously.

The epidemic situation of COID-19 in China has caused the short-term fluctuation of our country's economic situation, there are many researches on the impact of new coronavirus epidemic situation on the industry, but from the point of view of decentralization, the empirical analysis of the impact of epidemic situation is relatively few.

Maximize diversification measure (MDM) can clearly and properly quantify the diversification of portfolio, and the improved *Maximize diversification-return measure* (MDR) solves the weakness in practical application, and considers the optimal solution of return and sparsity. Then, the improved model is applied to the portfolio optimization of biopharmaceutical industry during the epidemic period. We have conducted an empirical study on the performance of the corresponding enterprises during the epidemic period, and the data is from the https://tushare.pro/ 。

4.2. Results

First, we study the impact of diversification on portfolio return. Assuming that there is no-short sell, we set the risk aversion level $\gamma = 20$ in this experiment, and the model has a good selection ability, the weight of all portfolios are nonnegative. After a large number of numerical experiments, we determine the optimal length of the history estimation window: *period = 210*, the optimal adjustment of asset weights, the time interval $\delta = 21$, the model has a higher performance. To study the effectiveness the *Maximize diversification-return measure* (MDR) in China's biopharmaceutical industry, the benchmark portfolio model is *the uniform portfolio* (1/N, every weight equals) and the *mean-variance model* (MV) of Markowitz, the performance of MDM and MDR and two benchmark models (MV, 1/N) are as follows.(There are all 31 stock assets)

Table 3. Model performance and time 20190101-20201229

	MDM. ($\phi = 0, \lambda = 50$)	MDR. ($\phi = 10, \lambda = 50$)	1/N.	MV ($\tau = 10, \lambda = 50$)
MR	0.013	0.015	0.034	0.0327
SR	0.158	0.150	0.452	0.371
CR	1.092	1.099	1.315	1.292
TOR	0.320	0.406	0	0.65
PNA	4.67	5.0	31	13.78

, where we consider the MDM is a special example of MDR model when $\phi = 0$.

The results show that the two models (MDM, MDR) both can select several stock assets well. On average, PNA < 5 means that both can select no more than 5 stocks more effectively. However, the direct use of the MDM model performed poorly and the cumulative return (CR) and the return Sharpe ratio (SR) is significantly poor than those of the two benchmark portfolio models (1/N, MV).

Even when expected return is taken into account ($\phi = 10$), there is few improvements in the performance of the return, the results appear to be very poor. The reason for our analysis is that historical data don't adequate only from January 1,2019 to December 29,2020 and the investment strategy was implemented for only one year, monthly ($\delta = 21$) updating the portfolio weight, which means only adjusting the portfolio weight 11 times, there is contingency. In addition, last year the price of related stocks increased greatly in the biopharmaceutical industry, stock returns within the industry are more relevant. So using maximum diversification measure is not conducive to earnings performance.

We further exaggerated the scope of the study, select the biomedical industry from 2008-01-01 to 2019-01-01 data, a total of 10 years, the cumulative implementation of portfolio strategy, the results are as follows.

Table 4. Model performance and time 20080101-20190101

	MDM ($\phi = 0, \lambda = 50$)	MDR ($\phi = 10, \lambda = 50$)	1/N	MV ($\tau = 10，\lambda = 50$)
MR	0.004	0.015	0.008	0.009
SR	0.041	0.132	0.098	0.078
CR	0.819	2.727	1.697	1.330
TOR	0.238	0.627	0	0.262
PNA	3.8	2.34	8	3.5

Through experiments, we find that the MDM model performs poorly still, and the cumulative return (CR) and the Sharpe ratio (SR) of returns are significantly lower than the two benchmark portfolio optimization models (1/N, MV). Although the MDM is good in theory, we prove that the diversification measure without improvement is not a good guide to practice. With the same level $\phi = \tau = 10$ of chasing return, the optimal *diversification-return model (MDR)* has the best effect, which shows that our improvement has a certain significance. Similar model evaluation can be extended to other types of diversification measures to find better performance portfolio diversification measures, evaluate model performance and study its effectiveness.

5. Analysis stocks assets

From 2018 to 2020, China's biopharmaceutical industry has a total of 31 stocks, corresponding to 31 companies. After we trade-off portfolio diversification and return in the MDR model, the three stocks (number: stock code) that were most preferred by the model during the COVID-19 epidemic in 2020 are *Saisheng Medicine (300485)* , *Dongbao Medicine (300239)* and *Weiming Medicine (002581)* .

5.1. Analysis of experimental results

We calculated the degree of preference of the three stocks in the model and calculated the average daily rate of return for 2019-2020 from the daily rate of return data, and obtained its rank of mean daily return rate in the biopharmaceutical industry of 31 stocks, as follows:

Table 5. Information of 3 stocks

stocks	Preference（PDA）	The average daily rate rate*252 ,the rank of mean daily return rate
300485(SM)	0.08	0.22，rank 15
300239(DM)	0.12	0.25，rank 13
002581(WM)	0.51	0.78，rank 3

Analysis: the model has the biggest preference to *Weiming Medicine*, the average holding ratio PDA is more than 50%, and the holding preference to it and enterprise is extremely high. According to the model, choosing these three stocks in biopharmaceutical industry is the best combination to balance the diversification and return. When we calculate the correlation coefficient of these three stocks, we find that *Saisheng Medicine* has the highest correlation with *Weiming Medicine* and *Dongbao Medicine* has the lowest correlation with *Weimin Medicine*. We think the result is diversity enough.

Table 6. Correlation Coefficient of Daily return (Pearson)

stocks	ρ	P-value<0.01
(WM, SM)	0.45	1.5e-25
(WM, DM)	0.22	2.5e-6
(DM, SM)	0.26	2.9e-9

The results show that the model selects 5 or so stocks from 31 stocks, improving performance, and includes one stock with high return rate – *Weiming Medicine*. At the same time, the selected portfolio has a high diversification as much as possible.

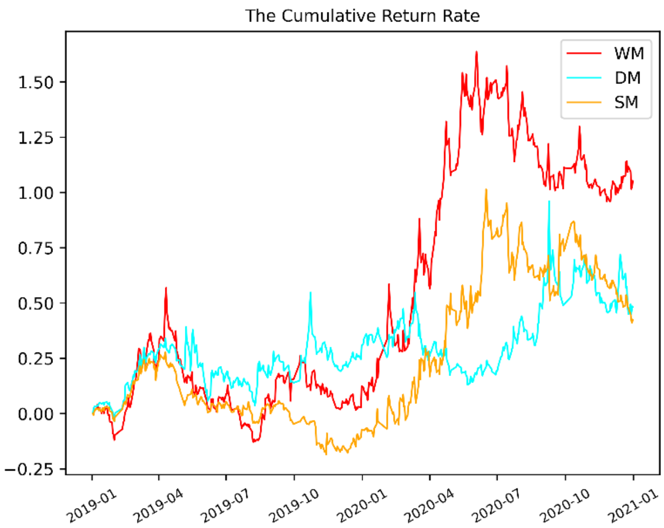

Figure 1. 2018-2020 Cumulative returns of WM, DM, SM.
Weiming Medicine was strong in the first half of 2020, and *Dongbao Medicine* were very different from the other two stocks. In addition, all three companies ended up with large increases in return rate.

5.2. The impact of the epidemic on stocks

Since January 20,2020, when high-level experts informed China of the confirmed “Human-to-human" outbreak, the number of people affected by the outbreak has increased rapidly, and the country has been strengthening control measures to combat the epidemic. Biopharmaceuticals have become an important weapon in the fight against new pneumonia. Global anti-epidemic measures continue to move forward. Vaccines and diagnostic reagents in biopharmaceuticals will have a greater opportunity for development. According to the China Biopharmaceutical Industry Development Report 2020, in the first half of 2020, the number of domestic biopharmaceutical investment events and the amount of investment were 91 and 32.516 billion yuan respectively, both rising year-on-year, of which the number of investment events increased by 28.17 percent year-on-year, investment increased by 39.62% year on year. The impact of the epidemic, domestic bio-pharmaceutical enterprises by increasing attention, domestic bio-pharmaceutical investment market active.[15; 16]

We have collated the three stocks of the main business and the relationship with the new crown epidemic, as follows:

- *Weiming Medicine (WM)*:

Enterprises belong to the pharmaceutical manufacturing industry, the main business is pesticide, pharmaceutical research and sales. In the industry, the health care reform policy will be continued and improved on the basis of the policy in 2019. With the introduction of relevant policies, the health care reform will continue to advance, the market competition will continue to intensify, and there are higher requirements for enterprises; The biopharmaceutical industry is influenced by many factors, such as raw materials, research and development investment, and so on. The research and development of products has many unpredictable risks, vaccine-related concept stocks may be affected by the news, anti-epidemic-related news, product development may affect the company’s share price.

- *Dongbao Medicine (DM)*:

The main business of the company is the research and development, production and sale of gelatin series products and low molecular weight collagen, which belong to the concept of medical cosmetology and artificial meat. In the first half of 2020, affected by the epidemic, the business performance of enterprises has been affected to a certain extent; although it is a biopharmaceutical concept, the research and development and sales of products related to the fight against the epidemic are not close enough.

- *Saisheng Medicine (SM)*:

The company is mainly engaged in the research and development, production and sales of bio-chemical drugs, involving cardiovascular and cerebrovascular diseases, immune diseases and neurological disorder, belonging to the concept of bio-vaccine and bio-medicine.

During 2020, while all in the biopharmaceutical industry the three companies that are the model's top preference will have very different main businesses. Among them, *WM* is the most relevant to epidemic prevention materials and vaccines, and its revenue is the leading level in the same industry. Meanwhile, *SM* which is similar to its business, has a lower preference of model, and is more suitable for our purpose of diversifying investment and ensuring a higher return. In addition, the main business of *DM* preferred by the model too is cosmetic medicine, which is quite different from the business of other two companies. According to the data, it can also be found that there is a big difference of return rates between *WM* and *SM,* the model fits our expectations well.

At the same time, investors should consider the uncertainty of the stock market as far as possible and analyse the influence of industry, market and policy enterprises in order to understand more information of stocks invested by them and evaluate the risk of the portfolio. The model can only be used as a reference for portfolio evaluation, not as a basis for purchasing stocks. In addition, the same industry is not recommended to choose too many stocks, in the model the biopharmaceutical industry only needs a few stocks to achieve a high diversification, investors are advised to choose in multiple industries.

6. Summary

The diversification measure can effectively evaluate the diversification of portfolio and can evaluate the dissimilarity of different stocks. This measure has great adaptability. Compared with two benchmark models (See Table 3, 4), the improved model considering return and the norm penalty can avoid the weakness of the diversification measure framework and increase the potential ability of the practical application. The performance of the improved model has better performance than previous models, which gives a highlight approach to think about how investors can measure portfolio diversification in different ways and trade-off between return and diversification.

Our model was applied to China's biopharmaceutical industry during the COVID-19, which gave us an added bonus. The model selects the most diversified portfolio from the biopharmaceutical industry 31 stocks assets. By collecting and analysing the data of these stock assets, we can judge the response trend of enterprises to the market news for investors' reference.

Acknowledgments

This project supported by Beijing Jiaotong University Training Program of Innovation and Entrepreneurship for Undergraduates, Beijing Training Program of Innovation and Entrepreneurship for Undergraduates and National Training Program of Innovation and Entrepreneurship for Undergraduates.

References

[1] CHOUEIFATY, Y. C. (2008). "FORMAT Diversification Toward Maximum." ournal of Portfolio Management.

[2] Koumou, G. B. (2020). "Mean-variance model and investors' diversification attitude: A theoretical revisit." Finance Research Letters 37: 101360.

[3] Koumou, N. B. G. (2017). "Rao' s Quadratic Entropy, Risk Management and Portfolio Theory.".

[4] Markowitz, H. (1982). "Portfolio selection." ournal of the Institute of Actuaries.

[5] Nayak, T. K. and J. L. Gastwirth (1989). "The Use of Diversity Analysis to Assess the Relative Influence of Factors Affecting the Income Distribution." Journal of Business & Economic Statistics 7 (4): 453-460.

[6] Perrin, S. and T. Roncalli (2019). "Machine Learning Optimization Algorithms & Portfolio Allocation.".

[7] Rao, C. R. (2010). "Quadratic Entropy and Analysis of Diversity." Sankhya. Series. A 72 (1): 70-80.

[8] Ricotta, C. and L. Szeidl (2009). "Diversity partitioning of Rao’s quadratic entropy." Theoretical Population Biology 76 (4): 299-302.
[9] Ricotta, C. and M. Moretti (2011). "CWM and Rao’s quadratic diversity: a unified framework for functional ecology." Oecologia 167 (1): 181-188.
[10] Stirling, A. (2010). "Multicriteria diversity analysis." Energy Policy 38 (4): 1622-1634.
[11]Tang, E. K. and P. N. Suganthan, et al. (2006). "An analysis of diversity measures." Machine Learning 65 (1): 247-271.
[12] Whittaker, R. H. (1960). "Vegetation of the Siskiyou mountains." Oregon and California Ecol. Monogr(30): 279-338.
[13] Yen, Y. (2016). "Sparse Weighted-Norm Minimum Variance Portfolios." Review of Finance 20 (3): 1259-1287.
[14] Yen, Y. and T. Yen (2014). "Solving norm constrained portfolio optimization via coordinate-wise descent algorithms." Computational Statistics & Data Analysis 76: 737-759.
[15] 段又源 (2020). "新冠肺炎疫情对我国股市的影响— 基于医药行业的实证分析." 中国商论 18 (817): 34-36.
[16] 何诚颖与闻岳春等 (2020). "新冠病毒肺炎疫情对中国经济影响的测度分析." 数量经济技术经济研究 37 (05): 3-22.