Analyze the Effect of Improving Portfolio Performance by Integrating Science and Technology Innovation Board and Growth Enterprise Market

Based on ARIMA Model and Mean-Variance Portfolio

Yining Gao(✉)

College of Arts and Sciences, Shanghai Maritime University, Shanghai 201306, China
201911010016@stu.shmtu.edu.cn

Abstract. Science and Technology Innovation Board (STIB) and Growth Enterprise Market (GEM) are developed plates in recent years, which makes both sectors full of vitality and have great development prospects in the future. This paper aims to enrich the research contents of STIB and GEM on the return effect of portfolio, and promote the performance and development of GEM and STIB in the investment portfolio. Firstly, the data are made up of two GEM stocks, another two stocks, which are from STIB, and the construction index. The data, selected from March 11th, 2021 to May 30th, 2022, go through data processing and testing prepared for model. Appropriate ARIMA model is selected for data prediction, and then mean-variance model is used to select appropriate portfolio through maximum Sharpe ratio screening. After the addition of stocks from STIB and GEM, the increase of return expectation and the improvement of Sharpe ratio both reflect that the portfolio performs better than the construction index, indicating that the integration of STIB and GEM improves portfolio returns. To a greater extent, it shows that the proposal of STIB and GEM can promote the main board market and improve the performance of the main board market.

Keywords: Science and Technology Innovation Board · Growth Enterprise Market · ARIMA model · Mean-variance portfolio

1 Introduction

Most of the Science and Technology Innovation Board (STIB) is mostly based on those enterprises with the content of innovation. The development trend of the STIB is highly consistent with the general direction of global asset allocation, also Scientific Innovation Board system arrangement is more conducive to the virtuous cycle of the primary and secondary markets and closer to the mature capital market [1]. Growth Enterprise Market (GEM) is mostly based on those enterprises with the business model of innovation. After more than ten years of development, GEM has gradually become mature, but the development in recent years has exposed some aspects of incomplete development. The

© The Author(s) 2023
V. Gaikar et al. (Eds.): FMET 2022, AEBMR 227, pp. 537–548, 2023.
https://doi.org/10.2991/978-94-6463-054-1_58

research of this paper is mainly based on the above environmental background to find out whether the integration of STIB and GEM can improve the portfolio performance.

The selected models and their characteristics are as follows. Difference operation is in order to make non-stationary sequences show the properties of stationary sequences after Difference, so that it can be used in ARIMA model. ARIMA model (Autoregressive Integrated Moving Average model), is generally used to fit the differential stationary series [2], just need the endogenous variables and don't need the exogenous variables, so it is widely used in stocks forecast. Markowitz solved the problem of optimal portfolio by means of mean-variance analysis and quadratic programming, and his portfolio theory is regarded as the cornerstone of modern financial theory. The model mainly assumes that the utility function of investors is determined by the mean and variance of returns, and the variance of returns reflects the investment risk [3]. When portfolio return rate follows normal distribution, mean-variance model can describe portfolio return rate distribution perfectly, which provides a solid foundation for investment risk management [4].

The research data selected in this paper are from March 11th, 2021 to May 30th, 2022, which makes the research results more appropriate to the current market. Moreover, the four stocks, which are sh688317, sh688819, sz300672 and sz300726, are in different fields, so as to avoid the impact of industry development on stock returns. Also the construction industry index is used to replace the selection of several stocks, so as to better reflect its representative for the main market. These data were selected because they are stationary series and are suitable for time series models to make predictions. The research mainly consists of four parts, the first step is checking and screening the data to get basic data, have mentioned, and then the prediction results are obtained by using the basic data. After that is to constructing combination. The last step to do is analyzing combination effect.

It has been proved that the integration of STIB and GEM can effectively improve the performance of portfolio returns, providing theoretical basis for the development of optimized portfolio of STIB and GEM. It is helpful to provide a basis for investors to optimize a variety of stock plates through STIB and GEM.

The framework of this paper has the following steps. The second part is data selection and methods, the third part is results and analysis, and the fourth part is conclusion.

2 Data and Methodology

2.1 Data

2.1.1 Data Sources

The paper uses the closing prices of 4 stocks from March 11th, 2021 to May 30th, 2022, which are obtained from AkShare database by python. The codes of four stocks are sh688317, sh688819, sz300672 and sz300726. Also a plate index, about architectonic and coded 399235, is used in this paper (Fig. 1).

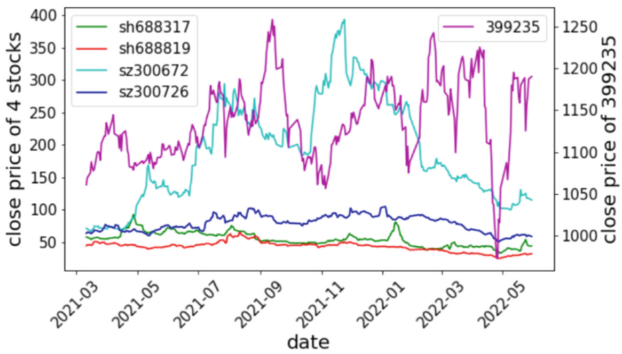

Fig. 1. Close price of 4 stocks and a plate index

2.1.2 Data Processing

Since determining the optimal portfolio based on the stock return, the compounded return is calculated through the closing price. The calculation formula of the compounded return is in Table 1.

$$Ct = ln(\frac{Vt+1}{Vt}) \tag{1}$$

V_t = prices, C_t = Compounded returns.

The main features of data are listed in the Table 1.

The histogram can be used to determine whether the data distribution is normal.

Jarque-Bera test is on skewness and kurtosis of data samples, Jarque-Bera test is used to evaluate whether the given data obey the normal distribution.

Both the histogram and Jarque-Bera test are in order to find out whether the data is available for the model.

As can be seen the results from Jarque-Bera test in Table 1, P value is less than 0.05, and kurtosis of all stocks is greater than 3, suggesting high apex and thickness tail of the return distribution.

The changes of return for four stocks and a plate index are shown in Fig. 2. It shows that the data distribution is normal.

2.2 ARIMA Model Forecast

2.2.1 Tests

To make sure the time series is executable, the stationarity tests are mentioned are to check the stationarity of the sequence.

The most commonly used method to test the stationarity of time series is the ADF (Augmented Dickey-Fuller) test. Hypothesis test H0 means test statistics obey ADF distribution. If H0 is accepted, it means that time series X_t contains the unit root, that is,

X_t is non-stationary. Rejecting H0 means that X_t is stationary [5]. Finding the p-value of Augmented Dickey-Fuller test, if P value is less than 0.05, the sequence is stable.

Ljung-Box is used to test whether the sequence within the range of m-order lag is white noise, and Q statistics obey the Chi-square distribution with m degree of freedom.

Table 1. Data descriptive statistics

	sh688317	sh688819	sz300672	sz300726	399235
Count	294	294	294	294	294
Mean	-0.001	-0.001	0.002	-3.00e-04	4.00e-04
Std	0.041	0.026	0.051	0.031	0.016
Min	-0.136	-0.083	-0.190	-0.099	-0.064
Quartile25%	-0.018	-0.014	-0.024	-0.019	-0.008
Quartile50%	-0.004	-0.002	-0.003	-0.003	0.001
Quartile75%	0.012	0.012	0.021	0.017	0.009
Max	0.165	0.180	0.182	0.114	0.051
Skewness	0.863	1.223	0.749	0.479	-0.421
Kurtosis	6.889	10.634	5.682	4.505	5.101
JB-value	221.739	787.169	115.587	39.008	62.750
P-value	0.0	0.0	0.0	3.38e-09	2.36e-14

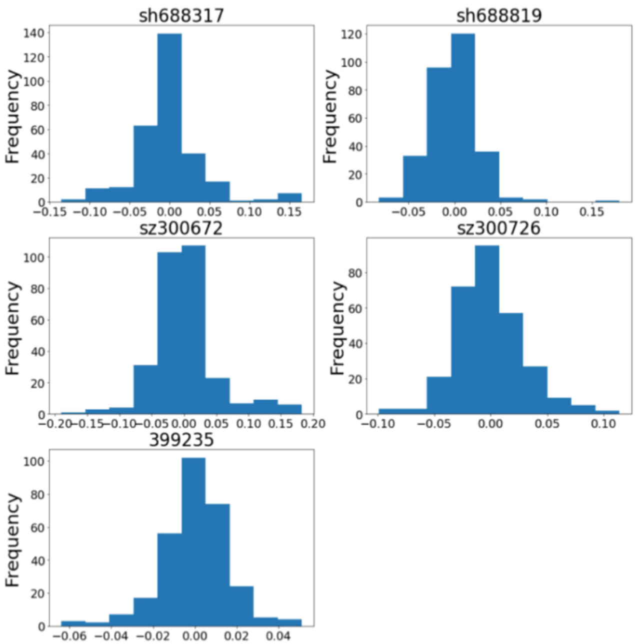

Fig. 2. Histogram of return

If the data is white noise, the data is worthless and there is no need for further analysis. Only the sequence is not white noise can be used in ARIMA model. H0: the sequence is

Table 2. Result of ADF test

	sh688317	sh688819	sz300672	sz300726	399235
Test Statistic	-1.13e + 01	-1.26e + 01	-9.60e + 00	-1.20e + 01	-6.91e + 00
ADF test P-value	1.53e-20	1.80e-23	1.99e-16	3.58e-22	1.19e-09
Num Lags Used	1.00e + 00	1.00e + 00	1.00e + 00	1.00e + 00	1.00e + 01
Num Observations Used	2.92e + 02	2.92e + 02	2.92e + 02	2.92e + 02	2.83e + 02
Critical Value (1%)	-3.45e + 00	-3.45e + 00	-3.45e + 00	-3.45e + 00	-3.45e + 00
Critical Value (5%)	-2.87e + 00	-2.87e + 00	-2.87e + 00	-2.87e + 00	-2.87e + 00
Critical Value (10%)	-2.57e + 00	-2.57e + 00	-2.57e + 00	-2.57e + 00	-2.57e + 00
Inspection result	stable	stable	stable	stable	stable
Ljung-Box test P-value	2.00e-04	0.003	0.013	0.022	0.002
White nose	not	not	not	not	not

white noise. H1: the sequence is not white noise. While P value is less than 0.05, reject H0, and the sequence is not white noise.

2.2.2 Set the Order

The Akaike information criterion (AIC) is one of the most ubiquitous tools in statistical modeling [6]. The minimum AIC criterion is selected as the order determination criterion. From low order to high order, models are established for different values of P and Q, and parameter estimation is carried out to compare the AIC value of each model, so that the model with minimal value can be determined as the best model [7].

AIC encourages good data fitting but tries to avoid over-fitting. Therefore, the preferred model should be the one with the lowest AIC value.

$$\text{AIC} = 2\text{k} + \text{NLN}\ (\text{SSR}/\ \text{N}) \tag{2}$$

N is the observation number, SSR is the SUM SQAURE OF RESIDUE.

2.2.3 Residuals White Noise Test

The probability of white noise associated with residual sequence also shows that the residual of each model satisfies the independence hypothesis [8].

Root mean square error (RMSE), also known as standard error. A smaller value of RMSE, the better fitted the distribution on load profiling data [9].

$$\mathbf{RSS} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

RSS is Residual Sum of Squares.

y_i = Predicted value, $\hat{y}_i$ = true value.

Mean Absolute Deviation (MAE) can avoid the error cancel each other, in order to reflect the actual prediction error accurately. Mean square error (MSE) is to reflect the degree of difference between two different estimators. Using Ljung-Box test to choose the effective models, whose residuals are white noise.

2.3 Portfolio of Mean-Variance Model

Risk is defined as the volatility of return, risk and return are taken as two standards to draw the return-risk chart, and the weight of the point with the largest Sharpe ratio is selected as the optimal weight of the portfolio. Sharpe ratio using the standard deviation as the risk measure, and the corresponding risk adjusted returns, in attachment is fund portfolio and risk-free rate geometric of slope, the value reflects the risk fund unit for more than risk-free interest rate, the size of the excess return advantage is simple, do not need to rely on market index can get the result [10].

According to the best choice in real life, we set the risk-free rate of return as the one-year government load interest rate, which on May 31st, 2022 is 0.019741.

2.4 Result Hypothesis

The integration of STIB and GEM board can improve the investment portfolio to a certain extent.

3 Results

3.1 Forecast by ARIMA Model

3.1.1 Result of Tests

The result of ADF test in Table 2 can clearly tell that the P value is far less than 0.05, indicating the series is stable. Ljung-Box test result in Table 2 shows that the P value is far less than 0.05, indicating the series is not white noise.

3.1.2 ARIMA Model

Based on the minimum value of AIC to choose the model that fit the sequence.

3.1.3 Residuals White Noise Test

Combined with the Ljung-Box test in Table 3, the p value is larger than 0.05, showing that the residuals are basically noise, which means residual test is passed.

In addition, the first 75% of the sequence is taken as training data and the last 25% as testing data. By comparing test values and predictions, three evaluation indexes are got, which lie in the last three lines. It find that the above index values are small in Table 3, so the model is effective.

Table 3. Sample ordering results

ARIMA model	sh688 317	sh688 819	sz300 672	sz300 726	399 235
(p, d, q)	(2,0,3)	(2,0,0)	(4,0,2)	(0,0,1)	(3,0,5)

Table 4. Statistics of residuals

	sh688317	sh688819	sz300672	sz300726	399235
Count	294	294	294	294	294
Mean	-2.97e-07	-7.00e-06	7.90e-05	-1.50e-05	5.87e-04
Std	3.83e-02	0.025	0.048	0.031	0.015
Min	-1.21e-01	-0.066	-0.174	-0.093	-0.061
Quartile25%	-1.83e-02	-0.015	-0.025	-0.020	-0.008
Quartile50%	-3.35e-03	-0.003	-0.006	-0.003	0.001
Quartile75%	1.85e-02	0.014	0.021	0.018	0.011
Max	1.64e-01	0.167	0.162	0.105	0.040
White noise test p-value	0.963	0.912	0.990	0.423	0.439
White nose	yes	yes	yes	yes	yes
MAE	0.037	0.019	0.028	0.021	0.017
MSE	0.003	0.001	0.002	0.001	0.001
RMSE	0.050	0.025	0.039	0.025	0.023

3.1.4 The Result of Forecast

Moreover, models are built for returns from four stock sets and a plate index to get the predictions in the next 22 days. The main features of 22 days forecast are listed in Table 4.

3.2 Portfolio

3.2.1 The Correlation Analysis of the Portfolio

The annualized covariance matrix. The covariance matrix tells us the volatility of the stock and is used for the calculation of volatility in Table 6.

3.2.2 Sharp Ratio Portfolio

Randomly generate a group of weights and repeat the scatter plot for 100000 times to determine the effective boundary portfolio and select the optimal portfolio, which is the portfolio with the maximum Sharpe ratio (Table 5).

Table 5. Main features of forecast

	sh688317	sh688819	sz300672	sz300726	399235
Count	22	22	22	22	22
Mean	-0.001	-0.001	0.001	-0.001	-0.001
Std	0.004	0.001	0.003	0.001	0.008
Min	-0.010	-0.004	-0.008	-0.006	-0.011
Quartile 25%	-0.004	-0.001	-0.002	-3.00e-04	-0.008
Quartile 50%	-0.001	-0.001	0.001	-3.00e-04	-0.002
Quartile 75%	0.003	-0.001	0.004	-3.00e-04	0.006
Max	0.006	0.001	0.005	-3.00e-04	0.011

Table 6. The annualized covariance matrix of forecast

	sh688317	sh688819	sz300672	sz300726	399235
sh688317	4.60e-03	-4.00e-04	1.00e-04	6.00e-04	1.10e-03
sh688819	-4.00e-04	2.00e-04	-1.00e-04	-1.00e-04	-1.00e-04
sz300672	1.00e-04	-1.00e-04	3.00e-03	6.00e-04	6.00e-04
sz300726	6.00e-04	-1.00e-04	6.00e-04	3.00e-04	5.00e-04
399235	1.10e-03	-1.00e-04	6.00e-04	5.00e-04	1.56e-02

Table 7. Each weight of the maximum sharp ratio portfolio

	sh688317	sh688819	sz300672	sz300726	399235
Portfolio weight	0.047	0.029	0.852	0.061	0.010

The sharp ratio is distinguished by color in order to find the maximum sharp ratio point and determine the weight clearly. The red point is the wanted portfolio. The corresponding graphs are shown in Fig. 3.

The optimal portfolio which is based on the weight composition of the maximum sharp ratio, and the return line are shown as follows, and the determined weighted portfolio is presented in Table 7.

As it can be seen from Fig. 4 and Table 8, the average return rate of portfolio is much larger than that of individual stocks.

Table 8. Equal weight Vs maximum sharp ratio VS plate index

	MSR	EW	399235
annual return	0.141	-0.143	-0.195
Volatility	0.048	0.546	0.125
Sharp ratio	0.159	-0.298	-1.720

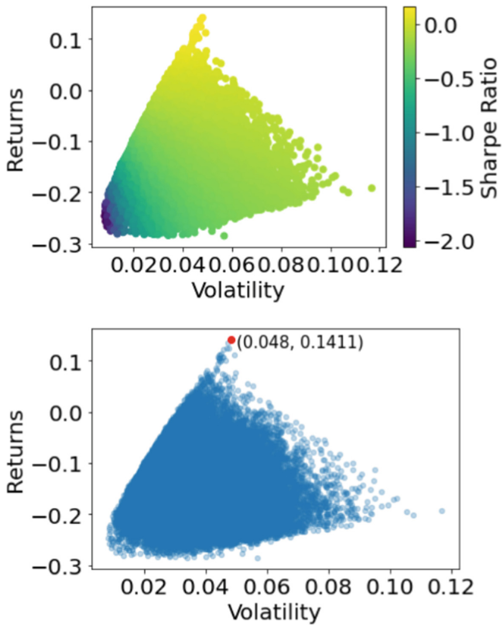

Fig. 3. The 100000 groups of weight plot

The Sharpe ratio of different combinations can well show the performance of combinations. However, it is worth discussing that, due to the limited amount of predicted data, data of different time periods are not selected for each parameter in the calculation process, which makes the value of Sharpe ratio have the possibility of over fitting.

From the perspective of the final cumulative return, the portfolio return of equal proportion portfolio presents a downward trend, while the optimal portfolio presents an upward trend. From the perspective of trend, the optimal portfolio has more effect. Comparing equal weight portfolio, the optimal portfolio and the plate index can be found that the optimal portfolio has the highest income to prove its performance is better than equal weight portfolio and plate index. So the integration of STIB and GEM can improve the portfolio can be proved and help the portfolio to get better revenue.

The portfolio performance is consistent with the previous assumptions, which is the integration of STIB and GEM board can improve the investment portfolio to a

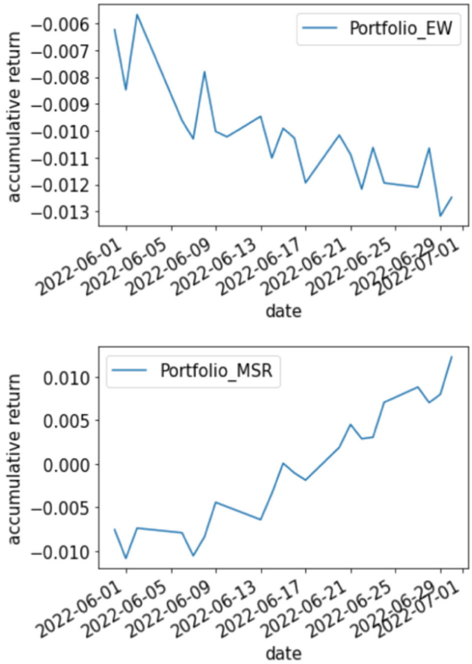

Fig. 4. Accumulative total return: equal weight Vs maximum sharp ratio

certain extent. The conclusion is that the integration of STIB and GEM can be allocated appropriately to improve the portfolio return.

4 Conclusions

This paper studies whether the integration of STIB and GEM can improve portfolio performance, selects four stock data and an industry sector index as research data, and uses time series model and mean-variance model for processing. After data processing and screening, suitable data are selected as the basis of the study. Then through the time series model for prediction and judgment, and check the fitting of the prediction results to find a usable forecast data. Finally, the prediction data are combined to obtain an optimal combination result. From the comparison of portfolio performance results, equal-weight portfolio and construction index itself, it can be seen that the optimal portfolio has the highest income expectation, and the performance of the optimal portfolio is also optimal considering the comprehensive income and risk. The results show that the right amount of stocks in STIB and GEM can improve the performance of investment portfolio to a certain extent. The research conclusion of this paper can be reflected in encouraging the integration of STIB and GEM stock. Although the STIB and GEM have a short development time, they have great development prospects, and it is not difficult to see that they can play a great role in the promotion of the main board market in the future. Investors can optimize the stock portfolio and expand the range of stock selection to obtain higher returns. It is suggested that investors can also consider STIB and GEM for a variety of stock plate optimization. It should be noted that this study has limitations on risk measurement. Standard deviation measures price volatility, which can both creating risk and generating benefits, is not the accurate risk.

References

1. China Securities Journal · China Securities Net. Ti Liu, deputy general manager of Shanghai Stock Exchange, said: The development of science and Technology Innovation Board shows five trends. September 18, 2021 07:45, Retrieved on july 02, 2022. Retrieved from: https://www.cs.com.cn/xwzx/hg/202109/t20210918_6205192.html [In Chinese]
2. Y. Liu, Z. Zhang, Mean-Variance-CVaR portfolio optimization model based on real return distribution. Journal of Systems Management, 2010, 19(04), pp.444-450. [In Chinese]
3. B. Zhang, S. Feng, X. Li, et al., Exchange rates and stock prices interactions in China. An empirical studies after 2005 exchange rate reform. Economic Research journal, 2008, 43(09): pp. 70–81+135. [In Chinese]
4. J.E. Cavanaugh, A.A. Neath, The Akaike information criterion: background, derivation, properties, application, interpretation, and refinements. Wiley Interdisciplinary Reviews: Computational Statistics, 2019, 11(3), pp. n/a-n/a. DOI: https://doi.org/10.1002/wics.1460. [In Chinese]
5. L. Wang, D. Xiao, Analysis of Non- steady Time- series forecast for economy based on ARMA model. Journal of Wuhan University of Technology. (Transportation Science and Engineering), 2004, (01), pp. 133-136. [In Chinese]
6. Y. Liu, X. Wang, ARIMA model in the application of China's energy consumption prediction. Journal of Economic Frame, 2007, (05), pp. 11–13 + 32. DOI: https://doi.org/10.15931/jcarolcarrollnki. 1006–1096.2007.05.002. [In Chinese]
7. A. Ahmad, I. Azmira, S. Ahmad, N.I.A. Apandi, Statistical distributions of load profiling data. 2012 IEEE International Power Engineering and Optimization Conference Melaka, Malaysia, 2012, pp. 199-203, DOI: https://doi.org/10.1109/PEOCO.2012.6230860
8. X. Zhang, S. Du, Can China's securities investment fund beat the market?. Journal of Financial Research, 2002, (01), pp.1-22. [In Chinese]
9. Y. Song, X. Zhao, Normality testing of High-Dimensional data based on principle component and Jarque–Bera statistics. Stats, 2021, 4(1) pp. 216-227. DOI: https://doi.org/10.3390/STATS4010016
10. Y. Liu, X. Li, Y. Ma, et al., Research on stochastic model testing. Journal of Computers, 2015, 38(11), pp. 2145-2162.DOI: https://doi.org/10.11897/SPJ.1016.2015.02145 [In Chinese]
11. G. Sun, Y. Shen, Y. Xu, et al., Time-Series analysis and forecast model for water quality of Yellow River based on Box-Jenkins method. Journal of Agro-environment Science, 2011, 30(09), pp. 1888-1895. [In Chinese]
12. X. Hui, H. Liu, W. Hu, et al., On the GARCH model and the forecast of Renminbi exchange rate. Journal of Financial Research, 2003, (05), pp. 99-105. [In Chinese]
13. P. Feng, X. Cao, Empirical research on stock price analysis and prediction based on ARMA model. Mathematics in Practice and Cognition, 2011, 41(22), pp. 84-90 [In Chinese]

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.