Composite Score of China's A-share market

1 Abstract

This study aims to complete the construction of the Quality Minus Junk (QMJ) factor and explore the significance of this factor in China's A-share market. The study first investigated the contribution of different indicators to factors by performing Principal Component Analysis (PCA) dimensionality reduction operations on multiple company indicators in the Chinese A-share market. The z-scores for three indicators — Profitability, Growth, and Safety—are calculated, and combined to compute the final Quality score. After grouping by market cap, Conditional Double Sorting was used to construct value-weighted and equal-weighted portfolios for Quality scores, forming the QMJ factor, which was then evaluated for significance using t-tests. This study also performed an H-L (High-Low) analysis on Profitability, Growth, Safety, and Quality scores to assess their predictive power on stock returns by evaluating differences between extreme groups. Meanwhile, the QMJ factor's performance is evaluated against the common benchmark model of the FF3, FF5, and HXZ factor models. By testing the alpha of three regression, whether the returns of the new factor are independent of existing factors and can provide additional information is examined. The research results show that the QMJ factor is still significant in China's A-share market. Besides, the equally weighted QMJ factor shows higher average returns and statistical significance, and its returns mainly come from the excess performance of small-cap stocks. This study highlights the significant impact of weighting schemes on portfolio returns and volatility, expanding the understanding of factor model construction methods. The findings offer strategic guidance for investors, such as selecting equal-weighted factors for higher returns or value-weighted factors for lower volatility, while also suggesting future research directions on optimizing weighting schemes and adapting to varying market conditions.

2 Introduction

Introduced by Asness et al. (2018), Quality Minus Junk (QMJ) factor was constructed by the classification of size and quality and its return was defined as subtracting the average return of the junk portfolio of large size and small size from the average return of the high-quality portfolio of large size and small size. While QMJ has been well-documented in developed markets, its applicability to China's A-share market remains unexplored. It seems that stock return predictability in China is less robust than in the U.S., with only a limited number of firm-specific variables showing significant predictive power (Chen et al., 2010).

The main objective of this project is to investigate the potential informational value of compan indicators in China’s stock market and assess the significance of the QMJ factor in this context. Specifically, this study compares value-weighted and equal-weighted portfolios, which provides a deeper insight than previous studies (Asness et al., 2018) that have solely relied on value-weighted portfolios.

The key empirical findings of this study demonstrate that the QMJ factor remains significant in the Chinese A-share market, with high-quality companies consistently outperforming low-quality ones. Notably, the analysis reveals that equal-weighted portfolios exhibit better performance compared to value-weighted portfolios, capturing small-cap performance more effectively in the Chinese market.

The report is structured as follows: section three reviews relevant literature on factor investing and its application to the Chinese market. The methodology section outlines the models and techniques used to analyze the data, followed by a presentation of the empirical results. In the discussion and conclusions section, we analyze the implications of the findings and their practical applications in the Chinese stock market. Finally, the report concludes with suggestions for future research.

3 Literature Review

According to Fama and French (2015), the five-factor asset pricing model increases profitability and investment factors on the basis of the original three-factor asset pricing model of Fama and French (1993). Each factor is constructed using a double-sort grouping method, where a sample of stocks is first sorted based on a specific financial indicator or market capitalization and then divided into different groups to calculate the difference in returns between the high-feature and low-feature groups. Then, the factor returns are calculated by constructing a long-short combination. An extension to Fama and French’s model is the Quality Minus Junk (QMJ) factor, introduced by Asness et al. (2018). Asness et al. (2018) first consider a dynamic model of firm quality and find the value affected by profitability, growth, and safety. Then the QMJ factor constructs the factor in the same way but with a new quality score defined to distinguish between high and low feature groups by calculating the quality scores based on the accounting-based variables.

Although the QMJ factor proves effective in many markets, it still lacks validation in China's A-share market. There are a large number of listed companies in China's stock market, which has very important research significance. Chen et al. (2010) consider that stock price changes in the Chinese market whose R-square is 0.46 are less reflected in company-specific information such as ROE, while in the U.S. market, stock prices are driven more by company-specific information with an R-square equal to 0.12. This indicates that the market maturity, investor composition, and information disclosure system of the two markets may be different, and QMJ's performance in China's A-share market may also be affected.

Pearson's 1901 paper first introduces PCA for finding best-fitting lines in multivariate space. In 1936, Fisher applies it to reduce variables for group distinction. PCA is useful in factor investment for dimensionality reduction. Chamberlain and Rothschild (1983) utilize PCA to extract factors from a large set of financial ratios with crucial information remaining at the same time. When constructing QMJ factors, we use relevant company operating indicators such as GPOA, ROE, and so on to calculate profitability. Therefore, we can consider using PCA dimensionality reduction to extract potentially important information. Ma, Y. (2021) selects listed companies in the transportation industry of China's A-share market and successfully reduces the dimension of 12 financial indicators to 5 principal components after PCA reduction operation, which explains 73.939% of the total information. This not only shows that most of the key information is preserved in the process of dimensionality reduction but also provides evidence of the feasibility of this method in China's A-share market.

4 Methodology

4.1 Principal Component Analysis (PCA)

According to Asness et al. (2018), the Quality score is made up of the Profitability score, the Growth score, and the Safety score. However, the indicators that build these three scores are highly likely to be strongly correlated, such as ROE (Return on Equity) and ROA (Return on Assets) showing a positive correlation. Therefore, we used the following steps for PCA (Principal Component Analysis) dimensionality reduction in order to extract underlying information during the construction of the three Scores (take computational Profitability scores as an example):

Step 1: On each date, all stocks are sorted against an indicator to get an $r_i = rank(i)$. Calculate the z-score of each stock, defined as Asness et al. (2018) mention in the paper is $z_x = z(x) = \frac{r_i - \overline{r_i}}{\sigma(r_i)}$, where $\overline{r_i}$ is the mean rank of all stocks on the same date and $\sigma(r_i)$ is the standard deviation of ranks across all stocks on the same date.

Step 2: Once the z-scores for each indicator (e.g., GPOA, ROE, ROA, CFOA, GMAR, and ACC) are calculated, we combine them into a matrix for each date, with each row representing a stock and each column corresponding to the z-score of an indicator. This matrix serves as the input for PCA.

Step 3: Since I need to ensure that the explained variance ratio is around 70% which is sufficient to explain the majority of the variance in the data, I choose the dimension of two and get two indicators PCA1 and PCA2 for each date.

Step 4: Investigate the contribution of each company's indicator to the explained variance of the principal components. After obtaining the loadings L_{ij} (i.e., the contribution of each feature to the principal components), I square the loadings to determine the relative contribution of each feature to each principal component. Then, I multiply each feature's contribution by the explained variance ratio of the corresponding principal component λ_j to calculate the feature's actual contribution to the total explained variance which is $C_{ij} = (L_{ij})^2 \cdot \lambda_j$.

Step 5: In the process of PCA dimensionality reduction, I store the maximum contribution features of each principal component at each time point. Take the two or three indicators with the greatest impact frequency (approximately more than 15% of the time with the greatest variance explanation). Again, calculating the z-score of selected indicators and adding them together. Then Profitability score is the z-score of the sum of z-scores. Growth and Safety scores build in the same way.

Besides, the $Quality\ score = z\ (Profitability +\ Growth +\ Safety\ )$

4.2 Quality-Based Stock Grouping

Performing Quality-Based Stock Grouping before constructing QMJ helps confirm whether the Quality score has predictive power for returns. The following steps are involved:

Step 1: For each month, the stocks are ranked based on their quality score, which divides the stocks into 10 quantile-based groups (deciles) where 1 represents the lowest quality and 10 represents the highest quality (ignore NA when ranking).

Step 2: Once the stocks are grouped, the mean return for each group is calculated. The highest-quality group and the lowest-quality group are specifically analyzed to calculate the return difference between these two groups.

Step 3: The difference between the mean return of the high-quality group (H) and the low-quality group (L) is computed for each time period. This H-L difference provides a measure of the predictive power of the quality factor for stock returns.

Step 4: T-statistic and p-value are calculated based on the H-L differences to assess if the return difference between high- and low-quality groups is statistically significant. The null hypothesis assumes no difference, while the alternative suggests a significant difference.

4.3 Conditional Double Sorting

In order to calculate the factor’s return, based on the assumption that there may be a correlation between market cap and quality score, I adopted the same model in Fama and French’s paper published in 1993, that is, conditional double sorting. The steps are as follows:

Step 1: For each time point, all stocks in the A-share market were divided into large-cap stocks and small-cap stocks using a 50th quantile (ignore NA when ranking).

Step 2: For stocks in the large-cap category, for another characteristic (quality, growth, or otherwise), they are divided into three categories using the 30th and 70th quantiles, with less than the 30th quantile being low and more than the 70th quantile being high. The same goes for stocks in the small-cap category. There are $2 \times 3 = 6$ categories.

Step 3: When calculating factor return, I need to use the formula used by Asness et al. (2018) to calculate QMJ factor return. Value-weighted (where the weight of each stock is the market cap of that stock divided by the total market cap of that category) and equal-weighted are used to calculate the return for each category. Subtract the average return of the group with the characteristic high (large cap and small cap) from the average return of the group with the characteristic low:

$$Factor\ Return = \frac{1}{2}(Small\ High + Big\ High) - \frac{1}{2}(Small\ Low + Big\ Low)$$

Step 4: In order to judge whether the excess return generated by the portfolio composed of this composite factor is significant, it is necessary to conduct a T-test on its return and calculate the t-statistic and p-value.

4.4 Factor Regression Analysis

After obtaining the composite factor return of the monthly frequency trading data, I conduct the factor regression analysis, which uses the benchmark model to verify the performance of the new factor and evaluate its incremental value. Here the Fama-French Three Factors Model, the Fama-French Five Factors Model, and the Hou-Xue-Zhang Four Factors Model are chosen. The regression model is as follows (Take the Fama-French Three Factors as an example):

$$Composite\ Factor\ Return = \alpha + \beta_{MKT}MKT + \beta_{SMB}SMB + \beta_{HML}HML + \varepsilon$$

This analysis assumes that there is a linear relationship between composite factor return and tested factors market factor (MKT), size factor (SMB), and book-to-market factor (HML) and that the error terms have independence and homoscedasticity.

α is the intercept term in the regression and a significantly positive α means the composite factor can generate excess returns that cannot be fully explained by existing market, size, and value factors. Therefore, the t-statistic and p-value of α will be calculated, based on the null hypothesis that $\alpha = 0$, if the p-value is less than 0.1, it is concluded that the composite factor also performs well in China's A-share market.

In addition, a higher R-squared indicates that a larger proportion of the variability in the composite factor return is explained by the factors (MKT, SMB, HML, and so on). While there may be other information from the excess return of the composite factor that the benchmark model does not capture if the R-squared is low.

5 Empirical Results

5.1 Results for Principal Component Analysis

When constructing profitability, growth, and safety scores, since there is a strong correlation between the indicators, in order to extract the potential information in indicators, we adopt PCA dimension reduction processing to ensure that the variance interpretation is maintained at a high level. Figure 1, Figure 2, and Figure 3 respectively show the explanatory ratios retained by the indicators used in the construction of profitability, growth, and safety after dimensionality reduction into two dimensions.

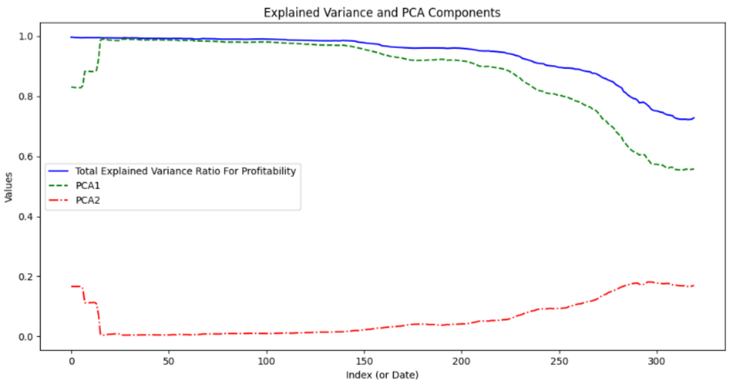

Figure 1: Explained Variance Ratio for Profitability

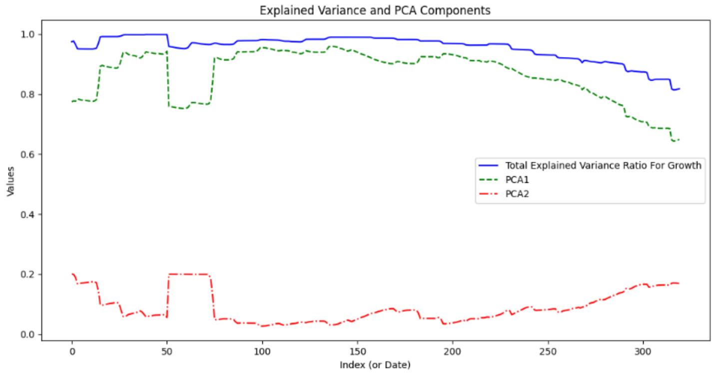

Figure 2: Explained Variance Ratio for Growth

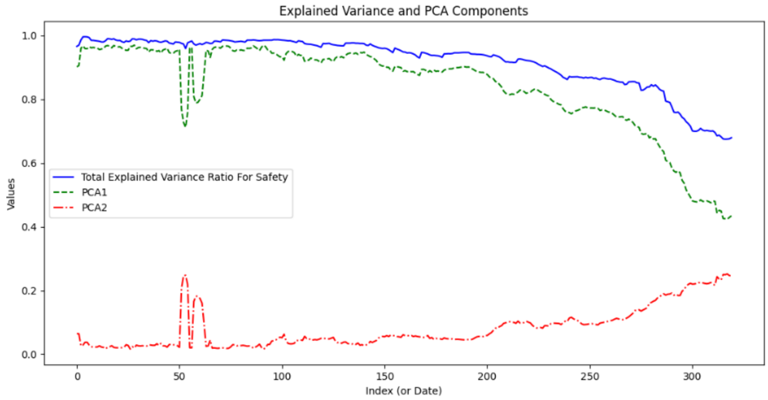

Figure 3: Explained Variance Ratio for Safety

The three figures show that the indicators after PCA dimensionality reduction still retain about 70% of the interpretation ratio, which is acceptable. Besides, it can be observed that with the change of time, the explanation degree of the most significant core factor decreases, while the explanation degree of the second significant factor increases.

After ensuring that the PCA reduction to two dimensions explains at least 70% of the variance, I visualized the contribution rates of various financial indicators to the profitability, growth, and safety scores. The results are shown in the figure below:

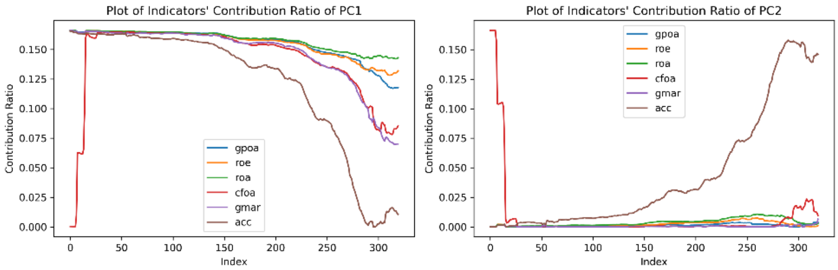

Figure 4: Indicators' Contribution Rate for Profitability

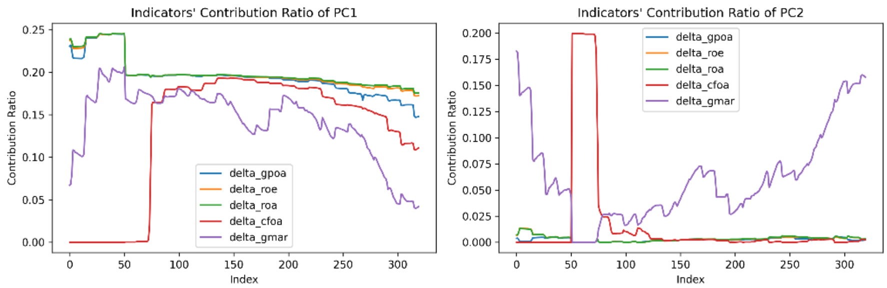

Figure 5: Indicators' Contribution Rate for Growth

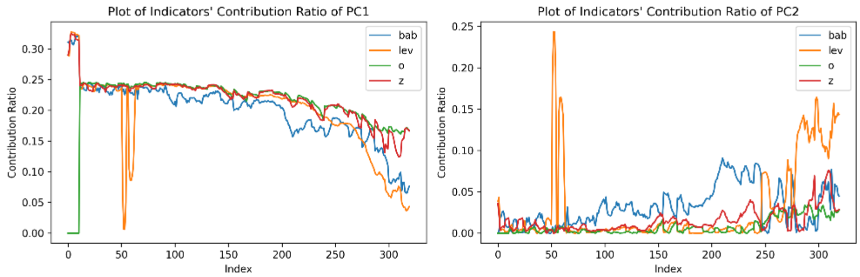

Figure 6: Indicators' Contribution Rate for Safety

From these figures, we can see that the highest contribution rates to Profitability Score are ROA and ACC, the highest contribution rates to Growth Score are ΔROA and ΔGMAR, and the highest contribution rates to Safety Score are Ohlson's O and BAB.

To find the indicators that contribute the most to the three scores, I counted the two indicators that contribute the most to the two principal components during PCA dimensionality reduction. Finally, I found the number of times each indicator appeared (two for each date). The larger contributing indicators calculate the z-score and are used to construct the score. The occurrence times of different indicators are shown in Table 1.

The Score Calculated	1st Contribution Features Counts	2nd Contribution Features Counts	3rd Contribution Features Counts	Total Features
Profitability	ACC: 293	ROA: 223	GPOA: 67	640
Growth	ΔROA:295	ΔGMAR: 292	ΔCFOA: 28	640
Safety	Ohlson's O: 236	BAB: 226	LEV: 119	640

Table 1: The Contribution of Each Feature to the Principal Component

Judging from the results of PCA dimensionality reduction, accruals (ACC) and return on assets (ROA) should be used to build Profitability. Change of return on assets

(ΔROA) and change of gross margin (ΔGMAR) should be used to build Growth. Ohlson's O, beta adjusted beta (BAB), and leverage (LEV) should be used to build Safety.

5.2 Results for Quality-Based Stock Grouping

Quality-Based Stock Grouping is used to verify whether quality scores have significant predictive power on stock returns, which provides theoretical support for the subsequent construction of QMJ quality factor portfolios. The result is shown in Table 2:

	Quality	Profitability	Growth	Safety
T-statistic	2.3477	3.3062	-0.1940	3.3066
p-value	1.9502e-02	1.0536e-03	8.4628e-01	1.0522e-03

Table 2: Results of T-test for High-Low

It can be seen from the results that the Profitability score and Safety score have strong explanatory power and the highest group return rate is significantly higher than the lowest group return, while there is no significant difference between the highest and lowest group Growth score returns. Quality score is significant, proving that it can be used and constructed QMJ factors in China's A-share market.

Since the Growth score is not significant in interpreting return, I tried to use only the Profitability score and Safety score to construct a new quality score by calculating the z-score (I will call it the new factor later).

5.3 Results for Conditional Double Sorting

I have completed the conditional double sorting by market size and quality factor and constructed both value-weighted and equal-weighted portfolios to calculate the excess returns on each date. The results are presented in Figure 7, Figure 8 and Figure 9.

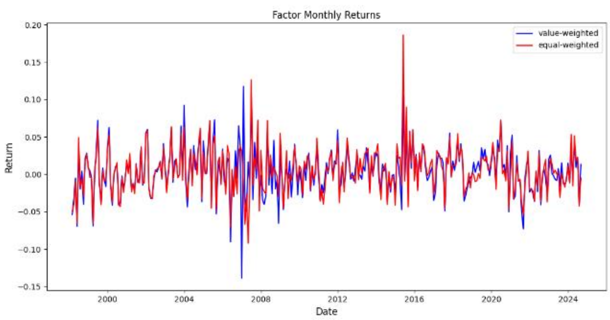

Figure 7: Factor Monthly Return without PCA

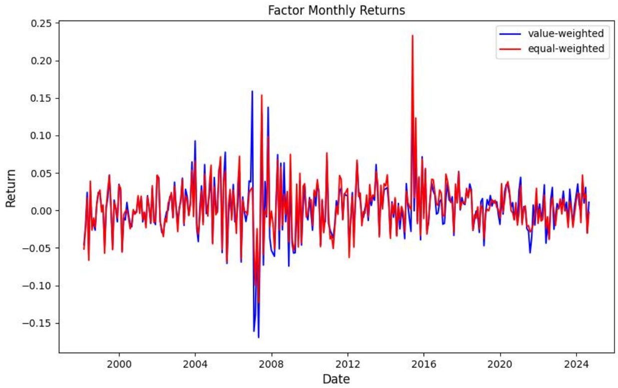

Figure 8: Factor Monthly Return with PCA

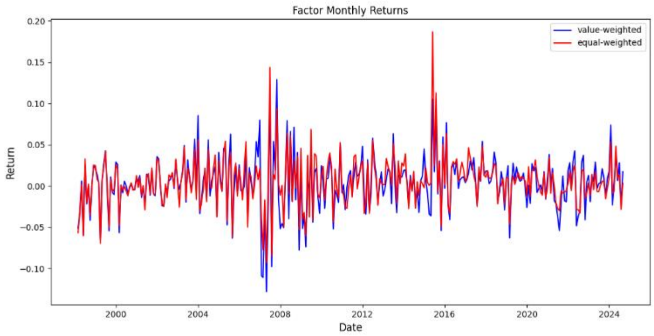

Figure 9: Factor Monthly Return with new composite factor constructed without Growth

In addition, I also completed the T-test for excess returns, and the results are shown in Table 3.

	value-weighted without PCA	equal-weighted without PCA	value-weighted with PCA	equal-weighted with PCA	value-weighted for new factor	equal-weighted for new factor
Average excess return	0.004021	0.004894	0.003316	0.004963	0.003763	0.005309
T-statistic	2.301407	2.825720	1.605531	2.57866	2.040352	3.154159
p-value	0.022015	0.005016	0.109370	0.010368	0.042141	0.001763

Table 3: Results of T-test for 4 Different Portfolios

The results suggest that the equal-weighted portfolio consistently outperformed the value-weighted portfolio. The application of PCA, while reducing dimensionality, led to a weaker performance in the value-weighted portfolio, though it still produced significant results in the equal-weighted portfolio. The return of the new factor constructed without Growth score is significantly different from zero in both investment portfolios, and the equal-weighted portfolio performs better than the original QMJ factor. It seems that the QMJ factor still performs well in China's A-share market.

5.4 Result of Factor Regression Analysis

In this part, I have used the excess monthly returns of the equal-weighted portfolio and value-weighted portfolio from the last part and fit an ordinary least squares (OLS) regression model with heteroskedasticity- and autocorrelation-robust standard errors, using the HAC (Heteroskedasticity and Autocorrelation Consistent) estimator. I conducted t test on the alpha obtained after regression, and the results are shown in Table 4 and Table 5.

Benchmark	FF3	FF5	HXZ
Alpha	0.009429	0.005282	0.002506
T-statistic	5.081780	3.406913	1.542974
p-value	3.739123e-07	0.000657	0.122837
Sharpe Ratio	0.158458	0.158459	0.201027
Adj. R-squared	0.398	0.529	0.371

Table 4: Test Result for Return of Equal-weighted Portfolio

Benchmark	FF3	FF5	HXZ
Alpha	0.008515	0.004477	0.002020
T-statistic	5.471502	3.362779	1.418706
p-value	4.46233e-08	0.000771	0.155984
Sharpe Ratio	0.129056	0.129056	0.157188
Adj. R-squared	0.331	0.427	0.294

Table 5:Test Result for Return of Value-weighted Portfolio

The results of the regression of the QMJ factor's excess returns on the three benchmark models (FF3, FF5, and HXZ) show that based on the FF3 and FF5 models, alpha is significantly non-zero, demonstrating that the QMJ factor generates additional benefits based on FF3 or FF5 that the benchmark models cannot explain.

However, for HXZ model as the benchmark, it is not significant that alpha is not 0, which may be because ROE in HXZ model is also used in the construction of QMJ factor, so the basis for the construction of the two is relatively similar.

6 Discussion and Conclusions

This study provides valuable insights into factor investing in the Chinese stock market. Specifically, the findings demonstrate the significance of the Quality Minus Junk (QMJ) factor in the Chinese A-share market. Furthermore, the results highlight the importance of portfolio weighting methods, the introduction of Principal Component Analysis (PCA), and the potential of new factors in enhancing investment strategies.

The average excess returns and corresponding T-statistics show that equal-weighted portfolios consistently outperform value-weighted portfolios, regardless of whether PCA or new factors are included. This may result in better diversification and higher exposure to small-cap and mid-cap stocks, which often outperform large-cap stocks in the Chinese market.

From a practical perspective, these findings have important implications for composite factor-based investment strategies. The effectiveness of the QMJ factor in the Chinese market indicates that quality-based factor investing strategies can help investors better identify and invest in high-quality, low-risk stocks, thereby enhancing long-term returns. Besides, equal-weighted strategies show greater robustness and higher returns, making them a compelling option for practitioners in the Chinese stock market.

7 Future Work

Future research in this area could focus on several key aspects to refine and enhance the models. First, incorporating higher-frequency data, such as intraday or weekly data, would help capture short-term market volatility and trends, enhancing the responsiveness of factor models. Additionally, exploring new factors, such as ESG scores and company innovation, could offer deeper insights into a company's future performance. Finally, classifying stocks by industry before ranking them based on their quality scores to construct the QMJ factor could be valuable, as industries may exhibit distinct trends during different economic development periods.

References

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2018). Quality minus junk. *Review of Accounting Studies*. https://doi.org/10.1007/s11142-018-9470-2

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica, 51*(5), 1281–1304.

Chen, X., Kim, K. A., Yao, T., & Yu, T. (2010). On the predictability of Chinese stock returns. *Pacific-Basin Finance Journal, 18*(4), 403-425. https://doi.org/10.1016/j.pacfin.2010.04.003

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116*(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179-188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Guo, B., Zhang, W., Zhang, Y., & Zhang, H. (2017). The five-factor asset pricing model tests for the Chinese stock market. *Pacific-Basin Finance Journal*, *43*, 84-106. https://doi.org/10.1016/j.pacfin.2017.02.001

Ma, Y. (2021). The Research on Influencing Factors of Stock Price Fluctuation of Listed Companies in China Based on PCA-Multiple Regression. Open Journal of Social Sciences, 9, 305-315. https://doi.org/10.4236/jss.2021.93020

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559-572. https://doi.org/10.1080/14786440109462720