Quantamental Trading: Fundamental and Quantitative Analysis with Multi-factor Regression Model Strategy

Chenling Xie[1] (✉), Yufan Zhang[2], Meng Wang[3], and Ziyue Liu[4]

[1] School of Letters and Science, University of California Los Angles, Los Angeles 90024, USA
joyxie1999@g.ucla.edu

[2] Department of Finance and Management, Shanghai University of International Business and Economics, Shanghai 201600, China

[3] Warwick Business School, The University of Warwick, Coventry CV4 7AL, UK

[4] School of Economics and Management, Nanchang University, Nanchang 330031, China

Abstract. Stock market has always been a risky and changeable field filled with opportunities and traps. Recently, with the burgeoning development of financial theory and statistically tools, investment methods based on quantitative analysis has become immensely predominant aiding accurate stock market predictions. The implementation of effective quantitative stock selection model can reduce risks at a greater level by accurately predicting the future performance of the stocks. The paper develops a multi-factor stock regression model and implements the same of the Chinese market as part of its research object. The proposed model establishes significant relationships between companies' and also discovers the effective financial indicators in association to their expected return. Thus, the multi-factor regression model thus helps to predict the stock's expected return by quantifying the multi-variable linear relationship between a company's financial ratios and its expected return. This enables companies to develop trading strategies considering companies' predicted returns. The results reveal the superiority of the multi-factor regression trading model in achieving an annualized rate of return of 28% during the back test period. The model when implemented on dataset indicates it ability as an effective framework to reduce risk and increase excess return.

Keywords: Multi-factor Regression Model · Quantamental · Chinese A Market

1 Introduction

Our strategy incorporates both fundamental and quantitative analysis to predict a company's future stock performance. Since financial ratios are fundamental factors of stocks and companies that can be used to explain changes in stock prices, we can make predictions to stock's expected return by constructing multi-factor models in order to quantify the multi-variable linear relationship between a company's financial ratio and its expected return [1]. Multi-factor regression models have been popularly used in various domains.

C. T. Dang et al. (Eds.): CONF-BPS 2023, AEPS, pp. 1455–1470, 2023.
https://doi.org/10.1007/978-981-99-6441-3_131

As an example, the study in [2] used multi-factor regression model for the prediction of height and 3D visualization of mining induced water-conducting fracture zone in Western Ordos basin. The study used grey correlation method and fuzzy ordered binary comparison method for determining the comprehensive weight. The weighted improved multiple regression model was computed by combinations and iterations. The relative error of the model was controlled at 10% which generated promising results.

After selecting 800 stocks from the base universe (Chinese A share market) using market capitalization and turnover rate as the filters, we construct the multi-factor model annually using preceding stock ratios and quarterly returns in order to predict the expected return of stocks for a given time. Our base universe is the Chinese A-share market. Each year around 800 stocks will be selected using filters, then we will long the stocks with the 10% highest expected returns and short stocks with the 10% lowest expected returns using signal weighting at the beginning of each quarter. Table 1 shows the estimated performance of our portfolio. The estimated annual portfolio return to reach nearly 10%. Even if the paper mentioned above shows an annualized return of 40%, we expect less than that because our portfolio implements hedging in order to make our portfolio market-neutral, which may reduce volatility at the cost of return. Also, we anticipate the sharpe ratio to be approximately 1.5. IR to be approximately 0.5. Since our strategy is hedged, we expect our portfolio with a low volatility. We expect the standard deviation of our portfolio to be around 0.065. The maximum drawdown is expected to be 10%. To reduce systematic risk and make our strategy not be highly affected by the market performance, we hope by implementing longing and shorting at the same time, we can make our portfolio uncorrelated with the market. But there will certainly be some residual, so we expect the correlation to be −0.005. Quantamental investing is a trading strategy that takes long positions in a company's stock if the company performs financially strong based on their historical fundamentals, and takes short positions in the company's stock if the company has a weak financial performance. Quantamental investing strategy is commonly used in case of hedge funds wherein the human traders use the results which are generated by AI and ML models. These techniques help in better predictions and as a result the performance of the fund improves. Similar approaches have emerged by integrating fundamental and quantitative methods for purchasing financial instruments such as stocks, bonds and derivatives. The source of this strategy comes from a research paper *The Fundamental of Fundamental Factor Models*, done by MSCI Research Center. The paper focused on the fundamental concepts of the factor models being implemented in Barra, Scotland. The authors Rosenberg and Marathe [3]. Developed a theory that which proposed the fact that the effects of macroeconomic events on individual securities could be calculated through the microeconomic characteristics namely the financial structure, industry membership and growth orientation factors. The study basically correlated the macroeconomic events and the microeconomic characteristics which was the primary aspect of the factor model. The study in [4] focused on the intuition behind the fundamental factor model revealing its connectivity with the traditional fundamental analysis and also highlighted the insights these models could provide. The primary objective of the study was to reveal the complimentary role of the fundamental factor model for traditional security analysis.

Table 1. Portfolio Performance Estimate.

Return		Risk		Correlation	
Cumulative Return	50%	Volatility	0.065	Correlation (stock index)	−0.005
Annualized return	10%	Skewness	0.19		
Sharpe Ratio	1.5	Kurtosis	3.56		
IR	0.5	Maximum Drawdown	10%		
Transaction cost	¥50,000	VaR (95%)	1%		
		VaR (90%)	3%		
		Calmar Ratio	3.0		
		Efficacy	40%		

2 Economic Intuition

Our strategy incorporates both fundamental and quantitative analysis to predict a company's future stock performance. Fundamental analysis measures a company's financial and economic well-being by analyzing its fundamentals. Fundamentals include profitability, revenues, assets, liability, financial ratios and etc. For quantitative analysis, it uses mathematical and statistical modeling, measurement, and research to understand behavior [5].

For our strategy, we use financial ratios such as PE ratio, ROE ratio and etc. as our major fundamentals to analyze and gain information about the company's past financial performances. Good financial ratios indicate strong financial performances which can lead to profitability growth and increase in stock price. Weak financial ratios indicate poor financial performances which can lead to potential decrease in stock price [6]. Since ratios are fundamental factors of stocks and companies that can be used to explain changes in stock prices, we can make predictions to stock's expected return by constructing multi-factor models in order to quantify the multi-variable linear relationship between a company's financial ratio and its expected return. The construction of the model can be done by using statistical tools including regression, normalization of data and etc. [7].

3 Quantitative Analysis

3.1 Return: Cumulative Return, Annualized Return, Sharpe Ratio, IR, Transaction Cost

The cumulative return is the total change in the investment price over a certain period of time. The annualized total return is the geometric mean of the amount of money earned from an investment each year over a certain period of time. The Sharp ration compares the return on investment with the associated risk. Sharp ratio was proposed as a part of Capital Asset Pricing Model (CAPM). The numerator is the difference between realized or expected returns and a benchmark such as risk-free rate of return considering the performance of a particular investment category. The denominator is the standard deviation of returns for a period of time measuring the risk. IR is the investors' relations which is a strategy that enables organizations to manage the relationships between executive leaderships and the financial community. The IR value helps investors analyze the affairs of a company and take informed decisions before making investments. The transaction cost represents the cost in making economic trade when participating in an event or cost required for buying or selling goods or services.

3.2 Risk: Volatility, Skewness, Kurtosis, Maximum Drawdown, VaR (95%), VaR (90%), Calmar Ratio, Efficacy

The Skewness is a measure for finding the asymmetry in a distribution. Kurtosis helps in measuring the frequency of occurrence of outliers in a distribution. The maximum drawdown (MDD) is the maximum observed loss from the peak to the trough of a particular portfolio until the peak is achieved. The Calmer ratio helps to compare the average annual compound rate of return with the maximum drawdown work of commodity trading advisors and the hedge funds.

3.3 Correlation: Correlation with Shanghai Composite Index

The Shanghai Composite Index is also known as SSE index which acts as the stock market index for all stock trading in SSE.

3.4 Data

Data Universe

For this research, our base universe is the A-shares of Chinese companies. Selecting stocks from the whole Chinese A-shares can better represent the overall operation of the securities market in China and leave little space for human interference in decision making. However, to get a better trading universe, we need to weed out some undesired stocks from over 4000 stocks in the A-shares market. Therefore, two stock filters: market capitalization and turnover rate are applied in this research. The portfolio was rebalanced every quarter in this research. Thus, the stock screener was implemented on a quarterly basis through using the end of the day data of each quarter. The two stock filters used to construct the trading universe is explained below:

I Market capitalization: Market capitalization is widely used for judging company financial performance and business outlook [8]. Smaller companies have greater operational risk and higher risk compensation. In this paper, we selected 100 companies with the largest market capitalization in each stock market sector categorized by Wind.

II High Turnover Rate: Some stock sectors in the Chinese market have fewer listed companies (less than 100) and small market cap. For those sectors, we select the companies with annual average turnover rate over 1 to ensure enough liquidity on the market to increase the chances of getting into or existing a position [9].

Data Set Stock

(1) Market Sectors: Chinese A-shares are listed into 10 sectors by WIND including: Technology, Consumer Discretionary, Health Care, industries, Consumer staples, Real Estate, Utility, Energy, Materials, and Finance. The stock data in this research is collected by above sectors respectively.
(2) Turn Over rate and Market capitalization: Market capitalization and turnover rate are used as stock screeners in this research. The data of each securities' market capitalization in the base universe were collected on a quarterly basis. Also their average daily turnover rates were calculated for filter stocks.
(3) Fundamental Ratios: Our research intends to construct a multi-factor stocks selection model that employs companies' fundamental information and their correlation with securities return. Therefore, fundamental information of our trading universe was collected on a quarterly basis in this paper such as P/E ratio, P/S ratio, price cash flow ratio, debt to asset ratio, ROE etc.
(4) Stock Prices: In order to conduct a significant test for the fundamental factors selected in this research and construct our portfolio, the daily adjusted closing price of each security in the trading universe is collected from the third quarter of 2013 to the end of 2020.

Data Source

To ensure the integrity for this research, the data used in this paper comes from the database of WIND Python. The Windpowerlib is a library that includes set of functions and classes that calculate the power output of wind turbines. The database was originally part of feedinlib which combined windlib and photovoltaic focusing on wind power models. Additionally, the special treatment stocks and delisting stocks were removed from the trading universe.

Data Range

In this research, to ensure sufficient data were collected in the performance test, the five year in-sample data started from 2015.1.1 to 2019.12.31 while the two year out-sample data range was from 2020.1.1 to 2021.12.31.

Factor Selection

As introduced above, our research is based on the multi-factor quantitative investment models for stock selection. Joseph Piotroski reported in his paper *Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers,* that the use of value multiples or financial criteria to build portfolios that outperform the market [10].

The factors in our research are classified into three categories: 1) Valuation factor: it indicates the expectation of investors on the value of stocks [11]. 2) Growth factor: it estimates the long-term development potential of a company. 3) Financial factors: the ratio reflects the ability of the company to obtain income [12]. Table 2 explains factors applied in this research explicitly.

Table 2. Selected Financial Ratios.

Valuation Factor	Price Earning Ratio (P/E)	Share price/EPS; Valuing a company by its current share price relative to its per-share earnings
	Market Sales Ratio (P/S)	Stock price/Sales per share; Valuing a company's stock price compared to its revenues
	Price to Book Ratio (P/B)	Stock value/Book value per share; Expressing a company's share price in relation to its book value of equity per share
	Price to Cash Flow (P/CF)	Share price*Number of outstanding shares/Operating cash flow; Valuing a stock's price to its operating cash flow per share
	Book to Market Ratio	(Total asset-total liabilities–preferred stock–intangible assets)/; Valuation of a company by comparing its book value to its market value
Growth Factor	Net Profit (YoY)	Comparing a company's growth in profit
	Operating Revenue (YoY)	Comparing a company's growth in revenue
	Return on Equity	Net income/Total assets; The ratio for measuring the profitability of a company in relation to the equity
	Return on Invested Capital	(Net income – dividends)/(Debt + equity); Measuring the percentage return that a company earns on invested capital
Financial Factor	Total Asset Turnover	Net sales/Average total assets; Measuring the revenue a business can create given its asset base
	Return on Total Assets	EBIT/Average total assets; Defining the profitability of a company in relation to its total assets
	Assets Liabilities Ratio	Debt/Asset; Measurement of the company's assets that are financed by debt

3.5 Strategy Detail

Signal Generation

After doing the first two filters with market capitalization and turnover rate, we then need to construct the multi-factor model. Multi-variate regression is a technique that is used to the degree in which the independent variables and dependent variables are linearly related to one another. When the multivariate regression model is applied to a dataset, it enables prediction of behaviour of the response variable considering the corresponding predictor variable. The primary advantage of multiple regression model is the ability to find the relative effect of one or more predictor variables to the criterion value. It also provides the ability to find the outliers or anomalies. The study in [13] performed stock market prediction using data acquired from social media and secondary data sources of financial sites. The data collected from these sources generally tend to be sparse and the selection of predictor variables also is generally poor which degrades the performance of the model. The study thus used a multiple regression model using the R software. The results generated by the model yielded an accuracy of 95% when compared to the traditional approaches. In the preset study, a model was constructed in order to quantify the multi-variable linear relationship between the expected return and the twelve different financial ratios. In order to construct the model, we use ratios as predictors, and the expected return as the outcome variable. Since we use ratios of the preceding quarter to forecast the following quarterly return, for a given quarterly return, we need to correspond it with its last quarter's ratio.

We generate every multi-factor model every year based on the last eight quarters using the selected 800 stocks. For example, when we want to get the model of 2015, we use quarterly returns from 2013Q1 to 2014Q4 and quarterly ratios from 2012Q4 to 2014Q3. Then we make quarterly returns from 2013Q1 to 2014Q4 to be the outcome variable, and ratios from 2012Q4 to 2014Q3 correspondingly to be the predictor.

Since we have 10 sectors, we then convert industry as categorical variable through encoding into the regression, which makes the 11^{th} variable in the model. Then, we need to test each factor's significance by the p-value. The result of significant test is showed in Fig. 1. There may be some factors that are not significant enough for the model that need to be discarded. Thus, we will drop out the insignificant ratios.

	coef	std err	t	$P>\lvert t\rvert$	[0.025	0.975]
const	-0.0013	0.003	-0.391	0.696	-0.008	0.005
DEBTTOASSETS	-0.0056	0.005	-1.155	0.248	-0.015	0.004
PE_LYR	0.0019	0.004	0.455	0.649	-0.006	0.010
PS_LYR	0.0080	0.004	1.949	0.051	-4.64e-05	0.016
PCF_NFLYR	0.0023	0.003	0.673	0.501	-0.004	0.009
PB_LYR	-0.0137	0.005	-2.957	0.003	-0.023	-0.005
YOYPROFIT	-0.0157	0.004	-4.406	0.000	-0.023	-0.009
ROE_BASIC	0.0238	0.009	2.530	0.011	0.005	0.042
YOY_OR	-0.0009	0.004	-0.230	0.818	-0.008	0.006
ROIC	0.0396	0.016	2.491	0.013	0.008	0.071
ROA	-0.0014	0.015	-0.093	0.926	-0.031	0.028
ASSETSTURN1	0.0224	0.004	5.771	0.000	0.015	0.030
BOOK_TO_VALUE	0.0300	0.004	7.007	0.000	0.022	0.038
WGSD_COM_EQ.1	-0.0108	0.004	-3.020	0.003	-0.018	-0.004

Fig. 1. Significant Test for Each Factor.

After we finish constructing the multi-factor model each year, in order to calculate the expected return of a stock for a given quarter, we put its most recent quarter's ratios into the model and receive the expected return of the individual stock for the given quarter, which is our signal. Signal weighting will be used here. We will long the top 10% best performing stocks and short the 10% worst performing stocks. Equal amount of capital will be distributed on both the long and short side. Our signal will be presented as followed:

$$R_e = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

where n is the number of factors remaining in the model.

Portfolio Construction

According to our base universe (the Chinese A-share market), we first divide all stocks into 10 sectors, including Technology, Consumer Discretionary, Health Care, industries, Consumer staples, Real Estate, Utility, Energy, Materials, and Finance. Then, the trading universe was constructed on a quarterly basis by implementing the two stock filters on a quarterly basis, including the market capitalization and the turnover rate.

While selecting stocks for the trading universe for a given quarter, we use its last-quarter's last-day market capitalization for sectors with more than 100 stocks, and rank their market caps from high to low, and choose the top 100 stocks. For choosing stocks from sectors which have lower than 100 stocks for a giver quarter, we use the average of the last-year turnover rate to be the filter. We choose stocks whose turnover rate is more than 1% because 1%-3% is the reasonable floating range of most of the active stocks, while those under 1% may have experienced a period of falling in price and don't receive enough attention. Besides, there are some stocks that are under special treatment by the China Securities Regulatory Commission or have already been delisted. They may also lack part of the ratios we need either because they are newly released or because the company doesn't announce them. So, we need to discard these companies and not include them in the regression. We change the trading universe at the beginning of each quarter.

We involve both long and short positions in our portfolio. When we get the trading universe each quarter, we choose stocks with the top 10% highest expected returns from each sector to constitute the long side, using signal weighting. For the short side, we choose stocks with the worst 10% lowest expected returns and also use signal weighting. The total trading capital is RMB100,000,000 and the trading frequency is quarterly. We allocate half the investment to longing, and half the investment to shorting. For our portfolio, we allocate RMB 50,000,000 to both longing and shorting.

Trade Execution

We can execute the trade at a Chinese trading platform—East Money. For the long side, the transaction cost includes commission and transfer fee, which are 2.5 bps and 0.2 bps respectively while buying. When selling, besides the former two fees, a 10 bps of stamp tax should also be considered. For the short side, the additional fee is the borrowing cost, which is not taken into consideration in our strategy since shorting is not allowed in the Chinese A share stock market. Hence, in this paper, we are conceptualizing the shorting part of the execution with zero borrow cost.

4 Development

Figure 2 shows the cumulative return of our strategy from 2014.12.30 - 2019.12.30. It also includes the Shanghai securities composite index for comparison. In general, a five-year steady growth of returns is shown on the P&L graph regardless of the performance of the market.

Fig. 2. Daily % Cumulative Return of In-sample Data (2014.12.30 - 2019.12.30).

Table 3. Statistics of In-sample Performance of the Constructed Trading Strategy.

Return		Risk		Correlation	
Cumulative Return	46%	Annualized Volatility	6.1%	Correlation (stock index)	0.006
Annualized return	8.4%	Skewness	0.569		
Sharpe Ratio	1.03	Kurtosis	4.156		
IR	0.25	Maximum Drawdown	36.7%		
Transaction cost	¥50,400	VaR (95%)	−1.2%		
		VaR (90%)	10.4%		
		Calmar Ratio	0.17		
		Efficacy	32.4%		

There are three obvious fluctuations in our Fig. 2, including the end of 2015, the beginning of 2018 and the end of 2019. At the end of 2015 and the beginning of 2018, our portfolio experienced two major drop-downs when the A share stock market also tumbled. Although the portfolio is made to be market neutral, its performance can still be impacted by major stock market turbulence.

For the 2018 drop down, the market continued going down for a year while our portfolio quickly recovered. This can be explained by the implementation of the new monetary policy with a low interest rate at the end of 2017 in China. Due to low interest rates, investors crowd into the stock market, pushing the stock price up. Since the companies we choose have large trading volumes, our portfolio also has a faster pace in increase of return, outweighing the shorting side. However, when the market drops significantly in the second half of the year, the shorting position in our portfolio mitigates the decline and makes our portfolio stable enough. The effect of shorting is more obvious throughout 2018 when the market keeps declining, our portfolio can still perform well by showing a slow growth of return.

In addition, there are two hills in our graph throughout 2017. Besides the influence of the market, the portfolio itself has its momentum because we change stocks in our portfolio quarterly, and the good financial data may not lead to good performance directly in the next quarter.

4.1 Quantitative

According to the Table 3, our portfolio performs well to some degree. First, the return of our portfolio experiences a steady increase, with an annualized return reaching 8.4% and a cumulative return reaching 46% in five years. The sharpe ratio is 1.03 and IR is 0.25, which are both decent. Second, the volatility of this portfolio is 0.061, so our strategy can be stated to run a relatively steady trend. The skewness in 2015 and 2016 is less than 0, and the distribution is to the left; The skewness from 2017 to 2019 is greater than 0, and the distribution is to the right. The absolute value of skewness in three years is less than 0.3, so the deviation degree of distribution can be considered not serious. In 2015, the kurtosis was the highest, reaching 11.9, exceeding the kurtosis value 6 of the exponential distribution, which was very steep compared with the normal distribution. However, after 2018, the kurtosis remained in an acceptable range.

4.2 Difference from Expectation

We expect our portfolio to show great risk control. It is able to maintain a positive percentage of return when the stock market tumbles. This is achieved by selecting stocks with strong historical financial performances through the multi-factor model and making the portfolio negatively correlated to market by shorting stocks with relatively weak historical financial performances. Yearly return graphs well capture the risk-control feature of our portfolio. For example, from January 2nd to January 8th 2016, the index experienced a 20% drop in return, while our portfolio maintained a steady 2% increase. This again happened on February 10th, 2017 where the index dropped 8% in the following ten days, while our portfolio increased its return from positive 1% to 9%. The negative correlation between our portfolio and the index is very well captured by the portfolio's 2018 return. The 2018 return graph showed that A share stock market continued to tumble from February to the year end, while our portfolio was able to maintain a steady increase in return since April.

In addition, we expect our portfolio will be less volatile than the index due to its hedging nature, and this expectation is proved to be right as we compare the return curve of the index to our portfolio. The portfolio's return curves are smoother without frequent ups and downs. This is further proved by comparing the annual volatility of the index and the portfolio.

For things that are different from our expectation, even if our portfolio consistently generates positive returns, we did not expect our portfolio is unable to consistently outperform the market index on an annual base. We expect through factor modeling which helps in early identification of trends, our portfolio is able to generate higher average returns than the A share stock market index. However, from in-sample data, 2019 shows a big discrepancy between index and our portfolio's return. SH index's average return appears to be 20% higher than our portfolio return which is only 2%.

In addition, we did not expect there will be big variations between each year return, from 2% to 15%. We expect our portfolio to perform consistently each year due to the modeling and hedging process. However, it turns out our portfolio is performing less stable and consistent than we expect. Lastly, we did not expect our portfolio is unable to catch the opportunities and make great profits when the stock market is blooming. For 2019, the stock market is blooming with an annual return of 23%. However, our portfolio only has an annual return of 2%, being the lowest among all five years. I believe this is caused by false shorting since even companies with weak financials are doing well as the market prospers. Thus, we might need to adjust the weighting for shortening and expanding.

5 Refinement

5.1 Introduction to the Strategy

The refinement of our investing strategies concentrates on how to improve the bias-variance trade-off of our regression model. The original regression model takes 12 ratios as predictors. However, not all of these variables are significant and such a large number of predictors will lead to high variance in prediction. Therefore, we propose to use AIC criteria to reduce the number of predictors while keeping the maximum predicting power [14]. AIC is Akaike Information Criterion, which is a mathematical model. It helps to understand how well a model fits to a data from which it was generated. AIC is thus used for comparative analysis of different models and thereby identify the best fitting one. AIC is calculated considering the number of independent variables that are used to build the model and the maximum likelihood estimate of the model. AIC considers the best fit model as the one having maximum amount of variation using least possible independent variables.

$$AIC = -2\ln(L) + 2k,\ \ where\ L = likelihood,\ \ k = \#\ of\ \ parameters$$

AIC penalizes additional predictors through the term 2 k. The procedure of predictor selection is the following: 1) Start with a regression model with 1 predictor. 2) Adding one more predictor to the model each time and record AIC. 3) Use the model with the smallest AIC value. In this research, we reduced the number of predictors from 12 to 5 [13]. The P&L Graph of in-sample data after refinement is shown in Fig. 3.

Fig. 3. Daily % cumulative return of in-sample data after refinement (2010.12.30 - 2019.12.30).

5.2 Summary of Statistics

See Table 4.

Table 4. Summary of statistics for the refined strategy.

Return		Risk		Correlation	
Cumulative Return	54.3%	Volatility	9.5%	Correlation (SH index)	−0.2
Annualized return	9.1%	Skewness	−0.62		
Sharpe Ratio	1.6	Kurtosis	−0.67		
IR	0.27	Maximum Drawdown	48%		
Transaction cost	¥50,000	VaR (95%)	−2.8%		
		VaR (90%)	3.2%		
		Calmar Ratio	0.08		
		Efficacy	32.88%		

5.3 Analysis

The refinement results shown in Table 4 turns out to have a relatively larger cumulative return but also larger volatility. In general, it outperforms the original strategy 95% of the time. To be more specific, the refinement strategy has a better Sharpe ratio and IR, which are 1.6 and 0.27. The annualized return is 9.5%, better than the original strategy with an annualized return of 8.4%. The annualized volatility of the refined strategy is 9.5% which is higher than the original portfolio which is 6.1%. The efficacy is similar, which means the two portfolios share similar trends. In conclusion, the two strategies appear to have a high correlation with similar trends because the refined strategy does not change the nature of the strategy, which is to quantify the relationship between expected return and historical ratios. The refined strategy turns out to have a higher return with higher volatility since it more accurately predicts the future performance of the stock.

6 Conclusion

The P&L graph of the 2-year out-sample data shown in Fig. 4 is quite different from the in-sample graph. We can notice that the return in 2020 is quite stable, fluctuating around 0%, but has a dramatic increase in the first half of 2021. The volatility is also bigger. However, growth of return in the in-sample period is quite steady and keeps increasing. Referring to the line of the market, it first rose significantly and had a long period of sideways oscillation, which is quite different from our portfolio. It further confirms that the strategy is almost uncorrelated to the market due to hedging. In 2020, since the general market is booming because of the recovery policy after the burst of the pandemic from 2019 [15], it may be the companies we short that drag our return down. Besides, the companies we choose are from various sectors. There may be some correlation among industries that lead to volatility. As for 2021, we use data from 2019 to 2020 to generate the regression, where COVID-19 affects most of the companies in China. Staying on fundamental investing and focusing on companies with good financial performance may bring us a great profit.

Fig. 4. Daily % cumulative return of out-of-sample data (2019.12.31 - 2021.12.31).

Table 5. Statistics of Out-sample Performance of the Constructed Trading Strategy.

Return		Risk		Correlation	
Cumulative Return	28%	Annualized Volatility	6.5%	Correlation (stock index)	−0.061
Annualized return	14%	Skewness	0.19		
Sharpe Ratio	0.27	Kurtosis	3.56		
IR	0.72	Maximum Drawdown	28.2%		
Transaction cost	¥20,160	VaR (95%)	−2%		
		VaR (90%)	5%		
		Calmar Ratio	0.39		
		Efficacy	36.3%		

Table 5 shows our portfolio's statistics of the back test performance over the out-sample period. First of all, the cumulative return is 28%, and the annualized return is 14%, which is higher than that in the in-sample period. The Sharpe ratio is 0.27, which is much lower than that of in-sample data and the IR is 0.72. Furthermore, it's still acceptable that the volatility is 0.065. Next, the skewness and the kurtosis are 0.19 and 3.56, which are both lower than that in the in-sample period. Last, the correlation is −0.061, showing negatively related between the portfolio and the stock index.

7 Additional Concern

- Only three or four factors turn out to be significant throughout the test, this signifies the inefficiency of our model. Also, even if we re-run the regression to get our final multi-factor for expected return calculation, we did not take into consideration that these significant factors might become insignificant after we re-run the test. This can cause inaccurate expected return prediction.
- Due to the inaccessibility of some companies' financial data and the de-listing of companies in our trading universe, we lost data for our data set, including financial ratios and stock prices that are extensively used in model construction. Thus, it might negatively affect the accuracy and predictability of the multi-factor model.
- Beside the accuracy of the test, the global environment is also an indispensable aspect that needs to be taken into consideration. Even if our portfolio is made to be market neutral, it can still be impacted by major stock market events, including covid-19 and the Russia and Ukraine War. The Chinese stock market is extremely turbulent in 2022 due to the uncontrollable spread of virus omicron. The Chinese economy is also experiencing a dark time as global inflation continues.

8 Conclusion

For the conservative investors (risk-averse) who are willing to invest long-term, we recommend our strategy for the following reasons: investors would be able to earn positive returns in the long run return (46% cumulative five-year return) because the portfolio return is steadily growing in the long term. Also, it is highly unlikely for our investors to lose money since our portfolio shows great risk control. It is able to maintain a positive percentage of return when the stock market tumbles.

However, for the aggressive, risk-taking investors who intend to invest short-term and make big fortunes during a small amount of time, we do not recommend him or her to implement this strategy. For using our strategy, investors might be unable to catch the opportunities and make great profits when the stock market is booming due to the market-neutral nature of our portfolio. Another drawback is that the model's annualized return is relatively (8% annualized five-year return) low due to its low-risk nature. Thus, our strategy might not be very attractive to investors seeking high returns with high risk.

Funding. No funding received in this paper.

Data Availability. All data, models, and code generated or used during the study appear in the submitted article.

Appendix

The code built in this research is available from: https://github.com/zyfisacoolboy021/quantamental_trading.

References

1. Zhang, H., Shen, H.l., Liu, Y.C.: Research on multi-factor quantitative stock selection based on self-attention neural network. Math. Stat. Manag. **39**(03), 556–570 (2020)
2. Yin, H.Y., et al.: Height prediction and 3D visualization of mining-induced water-conducting fracture zone in western ordos basin based on a multi-factor regression analysis. Energies **15**(11) (2022)
3. Rosenberg, B., Marathe, V.: Common Factors in Security Returns: Microeconomic Determinants and Macroeconomic Correlates. University of California Institute of Business and Economic Research, Research Program in Finance, Working paper No. 44 (1976)
4. Nielsen, F., Bender, J.: The fundamentals of fundamental factor models. MSCI Barra Research Paper 2010-24 (2010)
5. Xu, J.Z.: Quantitative stock selection analysis based on multi-factor model. Finan. Theory Explor. **4**(03), 30–38 (2017)
6. Hou, X.H., Wang, B.: Quantitative investment based on fundamental analysis: research review and prospects. J. Northeast Normal Univ. (Philos. Soc. Sci. Edn) **4**(01), 124–131+141 (2021)
7. Wang, J.L., Zhou, S.B., Wang, Y.W.: Research on multi-factor model based on Shanghai and Shenzhen 300 index stocks. Nat. Sci. J. Harbin Normal Univ. **35**(02), 19–22 (2019)
8. Reinganum, M.R.: The significance of market capitalization in portfolio management over time. J. Portfolio Manag. **25**(4), 39–50 (1999)

9. Avramov, D., Chordia, T.: Asset pricing models and financial market anomalies. Rev. Finan. Stud. **19**, 1001–1040 (2006)
10. Piotroski, J.D.: Value investing: the use of historical financial statement information to separate winners from losers. J. Acc. Res. **38**, 1–41 (2020)
11. Zhou, M.Y., Yin, L.Q.: Quantitative stock selection strategies based on kernel principal component analysis. J. Finan. Risk Manag. **9**, 23–43 (2020)
12. Gullett, N.S., Kilgore, R.W, Geddie, F.M.: Use of financial ratios to measure the quality of earnings. Acad. Acc. Finan. Stud. J. **22**(2) (2018)
13. Wang, J.J., et al.: Intelligent optimization based multi-factor deep learning stock selection model and quantitative trading strategy. Mathematics **10**(4), 566 (2022)
14. Shyu, S., Jeng, Y., Ton, W.H., Lee, K., Chuang, H.M.: Taiwan multi-factor model construction: equity market neutral strategies application. Manag. Financ. **32**(11), 915–947 (2006)
15. Liu, H.Y., Wang, Y.L., He, D.M., Wang, C.Y.: Short term response of Chinese stock markets to the outbreak of COVID-19. Appl. Econ. **52**(53), 5859–5872 (2020)