

RESEARCH

Open Access

# Research on a stock-matching trading strategy based on bi-objective optimization



Haican Diao, Guoshan Liu\* and Zhuangming Zhu

\* Correspondence: [liuguoshan@mbs.ruc.edu.cn](mailto:liuguoshan@mbs.ruc.edu.cn)  
Business School, Renmin University of China, 59 Zhongguancun Street, Beijing 100872, China

## Abstract

In recent years, with strict domestic financial supervision and other policy-oriented factors, some products are becoming increasingly restricted, including nonstandard products, bank-guaranteed wealth management products, and other products that can provide investors with a more stable income. Pairs trading, a type of stable strategy that has proved efficient in many financial markets worldwide, has become the focus of investors. Based on the traditional Gatev–Goetzmann–Rouwenhorst (GGR, Gatev et al., 2006) strategy, this paper proposes a stock-matching strategy based on bi-objective quadratic programming with quadratic constraints (BQQ) model. Under the condition of ensuring a long-term equilibrium between paired-stock prices, the volatility of stock spreads is increased as much as possible, improving the profitability of the strategy. To verify the effectiveness of the strategy, we use the natural logs of the daily stock market indices in Shanghai. The GGR model and the BQQ model proposed in this paper are back-tested and compared. The results show that the BQQ model can achieve a higher rate of returns.

**Keywords:** Pairs trading, Bi-objective optimization, Minimum distance method, Quadratic programming

## Introduction

Since the A-share margin trading system opened in 2010, there has been a gradual improvement in short sales of stock index futures (Wang and Wang 2013) and investors are again favoring prudent investment strategies, which include pairs-trading strategies. As a kind of statistical arbitrage strategy (Bondarenko 2003), the essence of pairs trading (Gatev et al. 2006) is to discover wrongly priced securities in the market, and to correct the pricing through trading means to earn a profit from the spreads. However, with the increase in statistical trading strategies and the gradual improvement of market efficiency (Hu et al. 2017), profit opportunities using existing trading strategies have become more scarce, driving investors to seek new trading strategies. At present, academic research on pairs trading has mainly concentrated on the construction of pairing models and the optimization design of trading parameters, with a greater focus on the latter. However, merely improving trading parameters does not guarantee a

high return for the strategy, and this drives researchers back to the foundations of the pairs-trading model.

There are three main methods for screening stocks: the minimum distance method, the cointegration pairing method, and the stochastic spread method. The *minimum distance method* was proposed by Gatev et al. (2006)—hence its common name, the GGR model. Gatev et al. (2006) used the distance of a price series to measure the correlation between the price movements of two stocks. When making a specific transaction, the strategy user determines the trading signal by observing the magnitude of the change in the Euclidean distance between the normalized price series of two stocks (the sum of the squared deviations, or SSD). Perlin (2007) promoted GGR as a unitary method rather than a pluralistic one; testing it in the Brazilian financial market, he found that risk can be lessened by increasing the number of pairs and stock. Do and Faff (2010) found that the length of a trading period can affect strategy returns; their study laid the foundation for later research. Jacobs and Weber (2011) found that the GGR model's revenue comes from the difference in the speed of paired-stock information diffusion. Chen et al. (2017) revised the measurement method of the GGR model, changing the original measure (SSD) to the correlation coefficient, and increased the reliability of the multi-pairing strategy. Wu and Cui (2011) first applied the GGR model to the A-share market; conducting a back-test on the stock markets in Shanghai, they found that the GGR model can generate considerable returns, and its profits come from a market's non-validity. Wang and Mai (2014) measured the return on stock markets in Shanghai, Shenzhen, and Hong Kong respectively, and found that improvements to the original approach can bring portfolio construction strategic benefits but can also increase the risk of exploitation of the GGR model.

The *cointegration pairing method* was first used by Vidyamurthy (2004) to find stock pairs with a cointegral relationship. He used cointegrating vectors as the weight of pairs when trading. To solve the problem of single-stock pairing risks, Dunis and Ho (2005) extended the cointegration method from unitarism to pluralism and proposed an enhanced index strategy based on cointegration. By extracting sparse mean–return portfolios from multiple time series, D'Aspremont (2007) found that small portfolios had lower transaction costs and higher portfolio interpretability than the original intensive portfolios. Peters et al. (2010) and Gatarek et al. (2014) applied the Bayesian process to the cointegration test and found that the pairing method can be applied to high-frequency data.

The *stochastic spread* method first appeared in a paper by Elliott et al. (2005), who used the continuous Gauss–Markov model to describe the mean return process of paired-stock spreads, thus theoretically predicting stock spreads. Based on the research by Elliott et al. (2005). Do et al. (2006) first linked the capital asset pricing model (CAPM) with the pairs-trading strategy and achieved a higher strategic benefit than when using the traditional random spread method. Bertram (2010), assuming that the price differences of stock obey the Ornstein–Uhlenbeck process, derived the expression of the mean and variance of the strategic return on the position and found the parameter value when the expected return was maximized.

Based on above approaches, many scholars have begun to study mixed multistage pairing-trading strategies. Miao (2014) added a correlation test to the traditional cointegration method and found that screening stock-correlation analysis improved the

profitability of the strategy. Xu et al. (2012) combined cointegration pairing with the stochastic spread model and conducted a back-test on the stock markets in both Shanghai and Shenzhen; they found that higher returns could be obtained. Following Bertram's (2010) research, Zhang and Liu (2017) examined a pairs-trading strategy based on cointegration and the Ornstein–Uhlenbeck process and found the strategy to be robust and profitable.

In recent years, most scholars have focused on improving the long-term equilibrium of paired-stock prices in the stock-matching process continuously. Few studies have considered the short-term fluctuations of paired-stock spreads, which has led to poor profitability of the strategy. Therefore, this paper focuses on the stock matching of pairs trading and constructs a bi-objective optimized stock-matching strategy based on the traditional GGR model. The strategy introduces weight parameters, conducts long-term stock price volatility spreads, and adjusts the equalizer to match investors' preferences, enhancing the flexibility and practicality of the strategy.

The remainder of this paper is organized as follows. **Basic theory and model** section provides the basic theories and models of pairs-trading strategies and double bi-objective optimization. **Optimized pairing model** section establishes an optimized pairing model. **Pairing strategy empirical analysis** section provides an empirical analysis of the optimal matching strategy proposed in this paper. Finally, **Conclusions** section presents conclusions and suggests future research direction.

## **Basic theory and model**

Based on theories of pairs trading, stock-pairing rules in the minimum distance method, and multi-objective programming, we propose a strategy to improve profits based on the minimum distance method.

### **Pairs trading**

#### ***Pairs-trading parameters***

Using a pairs-trading strategy requires a focus on the following trading parameters:

*Formation period*: the time interval for stock-pair screening using the stock-matching strategy.

*Trading period*: the time interval in which selected stock pairs are used for actual trading.

*Configuration of opening*: the value of the portfolio construction triggered. For example, we can start a transaction by satisfying the following conditions: (1) The user is in the short position state; (2) the degree to which the paired-stock spread deviates from the mean changes; and (3) the degree changes from less than a given standard deviation to more than a given standard deviation.

*Closing threshold*: the value of the position closing triggered. For example, when the strategy user is in position and the paired-stock spread hits the mean.

*Stop-loss threshold*: the value of the stop-loss triggered; that is, when the rules are engaged for exiting an investment after reaching a maximum acceptable threshold of loss or for re-entering after achieving a specified level of gains.

### Minimum distance method

When using the minimum distance method to screen stocks, it is necessary to standardize the stock price series first. Suppose the price sequence of stock  $A$  in period  $T$  is  $P_i^A (i = 1, 2, 3, \dots, T)$ ;  $r_t^A$  is the daily rate of return of stock  $A$ . By compounding  $r$ , we can get the cumulative rate of return of stock  $A$  in period  $T$ , which is recorded as:

$$CP_t^A = \prod_{i=1}^t (1 + r_t^A), \quad (1)$$

where  $t = 1, 2, 3, \dots, T$ . When we record the standardized stock price series as  $SP_t^A$ , the distance SSD of each two-stock normalized price series can be calculated as follows (Krauss 2016):

$$SSD_{AB} = \sum_{t=1}^n (SP_t^A - SP_t^B). \quad (2)$$

### Multi-objective programming

The multi-objective optimization problem was first proposed by economist Vilfredo Pareto (Deb and Sundar 2006). It means that in an actual problem, there are several objective functions that need to be optimized, and they often conflict with each other. In general, the multi-objective optimization problem can be written as a plurality of objective functions, and the constraint equation and the inequality can be expressed as follows:

$$\min F(x) = [f_1(x), f_2(x), \dots, f_v(x)], \quad (3)$$

$$s.t. g_i(x) \leq 0, i = 1, 2, \dots, p, \quad (4)$$

$$h_j(x) = 0, j = 1, 2, \dots, q, \quad (5)$$

where,  $x \in R^u$ ,  $f_i: R^u \rightarrow R (i = 1, 2, \dots, n)$  is the objective function; and  $g_i: R^u \rightarrow R$  and  $h_j: R^u \rightarrow R$  are constraint functions. The feasible domain is given as follows:

$$X = \{x \in R^u | g_i(x) \leq 0, i = 1, 2, \dots, p; h_j(x) = 0, j = 1, 2, \dots, q\}. \quad (6)$$

If there is not an  $x \in X$ , such that

$$f(x) \leq f(x^*), \quad (7)$$

then  $x^* \in X$  is called an effective solution (Bazaraa et al. 2008) to the multi-objective optimization problem.

### Optimized pairing model

Previous studies on the GGR model have mostly focused on similarities in stock trends and have cared less about the volatility of stock spreads. Such studies could not present ways to achieve higher returns. This paper, however, is based on the traditional GGR model, and can thus propose a new pairs-trading model, namely bi-objective quadratic programming with quadratic constraints (BQQ) model. By adjusting the weights between maintaining a long-term equilibrium of paired-stock prices and increasing the volatility of stock spreads (Whistler 2004), we can achieve equilibrium.

**Mean-variance minimization distance model**

Assume that there are  $m$  stocks in the alternative stock pool, and the formation period of the stock pairing is  $n$  days. Take the daily closing price of the stock as the original price series, recorded as  $P_1, P_2, \dots, P_m$ . To make the price sequence smoother, we use the average price series over the past 30 days:  $\overline{P}_1, \overline{P}_2, \dots, \overline{P}_m$  (instead of the original price series), to eliminate short-term fluctuations in stock prices. Then, in the moment,  $t$  can be expressed as follows:

$$\overline{P}_{i,t} = \frac{1}{30} \sum_{j=0}^{29} P_{i,t-j}, t = 1, 2, \dots, n. \tag{8}$$

First we consider  $\sum \alpha_i \overline{P}_i$ .

Let  $\alpha$  be the weight of the stock in the stock pool, and then let

$$\begin{cases} \alpha_i^+ = \alpha_i (\alpha_i > 0, i = 1, 2, \dots, n) \\ -\alpha_i^- = \alpha_i (\alpha_i < 0, i = 1, 2, \dots, n) \end{cases} \tag{9}$$

Then, we divide the stock into two groups according to the positive and negative weights. The stock combination with a positive weight is called  $P_t^+$ , while the stock combination with a negative weight is called  $P_t^-$ , so

$$\begin{cases} P_t^+ = \sum \alpha_i^+ \overline{P}_{i,t} \\ P_t^- = \sum \alpha_i^- \overline{P}_{i,t} \end{cases} \tag{10}$$

According to the GGR method, as long as we are in the formation period  $n$ , we can consider that the groups' prices have to represent a long-term equilibrium relationship. Therefore, we get the bi-objective optimization model as follows:

$$\min \sum_{t=1}^n \left( \sum_{i=1}^m \alpha_i \overline{p}_{i,t} \right)^2 \tag{11}$$

The volatility of the paired-stock spread is a source of revenue for the pairs-trading strategy. Variances are used to describe the volatility of a time series. Therefore, we use the formula below to measure the stock spread:

$$\max \frac{1}{n} \sum_{t=1}^n \left( \sum_{i=1}^m \alpha_i \overline{p}_{i,t} - \frac{1}{n} \sum_{t=1}^n \sum_{i=1}^m \alpha_i \overline{p}_{i,t} \right)^2 \tag{12}$$

Avoiding the case that  $\alpha = 0$ , we increase the regularity constraint; that is, the second-order modulus is 1, so we can obtain the BQQ model as:

$$\min \sum_{t=1}^n \left( \sum_{i=1}^m \alpha_i \overline{p}_{i,t} \right)^2, \max \frac{1}{n} \sum_{t=1}^n \left( \sum_{i=1}^m \alpha_i \overline{p}_{i,t} - \frac{1}{n} \sum_{t=1}^n \sum_{i=1}^m \alpha_i \overline{p}_{i,t} \right)^2 \tag{13}$$

$$s.t. \sum_{i=1}^m \alpha_i^2 = 1. \tag{14}$$

This paper uses a linear weighting method by introducing weight  $\lambda (\lambda > 0)$ , transforming the bi-objective optimization problem into a single-objective optimization problem. The model is denoted as revised quadratic programming with quadratic constraints (RQQ):

$$\min \sum_{t=1}^n \left( \sum_{i=1}^m \alpha_i \bar{p}_{i,t} \right)^2 - \lambda \frac{1}{n} \sum_{t=1}^n \left( \sum_{i=1}^m \alpha_i \bar{p}_{i,t} - \frac{1}{n} \sum_{t=1}^n \sum_{i=1}^m \alpha_i \bar{p}_{i,t} \right)^2. \tag{15}$$

$$s.t. \sum_{i=1}^m \alpha_i^2 = 1. \tag{16}$$

Since users of the matching strategy have different risk preferences,  $\lambda$  can be seen as an important indicator of strategic risk. When  $\lambda$  is large, the model magnifies the volatility of the paired-stock spread sequence, and the strategy may obtain higher returns, but it also raises the risk of divergence in the stock spread. Therefore, users can adjust  $\lambda$  to match their risk preferences, which increases the usefulness of the pairing strategy.

$$\text{Let } \bar{p} = \frac{1}{n} \sum_{t=1}^n \bar{p}_t.$$

To facilitate the model solution, we perform matrix transformation as follows:

$$f(\alpha) = n\alpha^T \left( \frac{1}{n} \sum_{t=1}^n \bar{P}_t (\bar{P}_t)^T \right) \alpha - \lambda \left[ \alpha^T \left( \frac{1}{n} \sum_{t=1}^n \bar{P}_t (\bar{P}_t)^T \right) \alpha - \alpha^T \bar{P} (\bar{P})^T \alpha \right]. \tag{17}$$

That is,

$$f(\alpha) = \alpha^T [(n-\lambda) \left( \frac{1}{n} \sum_{t=1}^n \bar{P}_t (\bar{P}_t)^T \right) + \lambda \bar{P} (\bar{P})^T] \alpha. \tag{18}$$

For a given  $\alpha_k$ , we get the sub-problem of the model as this:

$$\min f_s(\alpha_k) = d^T H \alpha_k + \frac{1}{2} d^T H d. \tag{19}$$

$$s.t. 2d^T \alpha_k + \alpha_k^T \alpha_k - 1 = 0. \tag{20}$$

**The sequential quadratic programming algorithm**

Since the objective and constraints of RQQ are quadratic functions, these are typical nonlinear programming problems. Therefore, the sequential quadratic programming algorithm can solve the original problem by solving a series of quadratic programming sub-problems (Jacobs and Weber 2011; Zhang and Liu 2017). The solution process is as follows:

*Step 1:* Give  $\alpha_1 \in R^m, \varepsilon > 0, \mu > 0, \delta > 0, k = 1, B_1 \in R^{m \times m}$ .

*Step 2:* Solve sub-question  $sub(\alpha_k)$ , and we get its solution  $d_k$  and the Lagrange multiplier  $\mu^k$  in the case of  $|d^k| \leq \varepsilon$ , terminating the iteration; therefore, let  $s_k \in [0, \delta]$  and  $\mu = \max(\mu, \mu^k)$ . By solving this:

$$P(\alpha_k + s_k d_k, \mu) \leq \min_{0 \leq s \leq \delta} P(x_k + \alpha d_k, \mu) + \varepsilon_k, \tag{21}$$

we get  $s_k$ , where  $\varepsilon_k(k = 1, 2, \dots)$  satisfies the non-negative condition and

$$\sum_{k=1}^{\infty} \varepsilon_k < +\infty. \quad (22)$$

Equation (21) is the exact penalty function.

*Step 3:* Let  $\alpha_{k+1} = \alpha_k + s_k d_k$ , and use the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS, Zhu et al. 1997) to find  $B_{k+1}$ , then let  $k = k + 1$  and go back to *Step 2*.

Thus, we find the optimal sub-solution  $d_k$ . Make  $d_k$  the search direction and perform a one-dimensional search in direction  $d_k$  on the exact penalty function of the original problem; we get the next iteration point of the original problem as  $\alpha_{k+1}$ . The iteration is terminated when the iteration point satisfies the given accuracy, obtaining the optimal solution of the original problem.

### Pairing strategy empirical analysis

To verify the profitability of the BQQ strategy, this paper compares the empirical investment effects of the BQQ strategy and the GGR strategy with the same transaction parameters and applies a profit-risk test for the arbitrage results of the two strategies.

#### Data selection and preprocessing

We use SSE 50 Index constituent stocks in the Shanghai stock market as the sample set for this study. We choose this sample set for its high circulation market value and large market capitalizations. Since the stock-pairing method proposed in this paper is based on an improvement of the traditional minimum distance method, this is consistent with the GGR model in the time interval selection of the sample: The paired stocks for trading are selected during the formation period of 12 months, and the stocks are traded in the next 6 months. To verify the effectiveness of the strategy, the paper conducts a strategic back-test from January 2016 to December 2018. Within the period, the broader market experienced a complete set of ups and downs.

Due to the existence of share allotments and share issues by listed companies, and because the suspension of stocks will also lead to a lack of market data, the raw data needs to be preprocessed. By reversing the stock price forward, the stock price changes caused by the allotments and stock offerings are eliminated. In addition, we exclude stocks that have been suspended for more than 10 days in the formation period. These missing data are replaced by the closing price of the nearest trading day.

#### Parameters settings

##### Transaction parameters setting

The implementation of a pairs-trading strategy relies on setting trading parameters. To compare this strategy with the traditional minimum distance method and verify the validity of the BQQ strategy, this paper uses the same parameters used in the GGR model for setting the trading parameters. We set the stop-loss threshold to 3 to prevent excessive losses due to excessive strategy losses and transaction costs. We set the number of paired shares to 10. For convenience, we divide the stocks into groups according to their weights, positive and negative.

**Portfolio construction**

After determining the trading parameters and cost parameters, we also need to determine the stock opening method; assuming that the final selected pair of stocks is  $\{S_1^+, S_2^+, \dots, S_5^+\}, \{S_1^-, S_2^-, \dots, S_5^-\}$  (corresponding to two sets of paired stocks), and the corresponding weight is  $\{\alpha_1^+, \alpha_2^+, \dots, \alpha_3^+\}$  and  $\{\alpha_1^-, \alpha_2^-, \dots, \alpha_3^-\}$ . When the trading strategy issues a trading signal for opening, closing, or stop-loss, the trading begins. The user needs to trade  $\alpha_i/\alpha_1 (i = 2, 3, 4, \dots, 10)$  units of stock  $\{S_1^-, S_2^-, \dots, S_5^-\}$  for each unit of  $\{S_1^+, S_2^+, \dots, S_5^+\}$ . Then, the strategy user has a net position, which is the paired-stock spread.

**Performance evaluation**

To compare the effects of the GGR model and the proposed BQQ model, we verify the effectiveness of the proposed optimization pairing strategy. This paper selects the income coefficient  $\alpha$ , risk coefficient  $\beta$ , and the Sharpe ratio as evaluation indicators, and the two strategies are back-tested and compared on the JoinQuant platform.

**Stock-matching stage**

When adopting the GGR model, we select five groups of stocks with the smallest SSD (two in each group) from each formation period. There is a small distance between these stocks. The stocks are selected from 50 constituent stocks. The matching results are shown in Table 1. When adopting the BQQ model, since the trend of the stock was screened beforehand, we select two sets of stocks (five in each group) for pairing. To explore the impact of  $\lambda$  on strategy performance, we perform a back-test on the optimal matching strategy under different values (when  $\lambda$  is greater than 0.7, the paired-stock spread is relatively poor, resulting in a strategy failure). Therefore, this paper is limited to a  $\lambda$  range from 0 to 0.7. The pairing results are shown in Table 2.

As can be seen in Table 2, when  $\lambda$  changes from 0 to 0.4, the selected stock pairs show a very dramatic change; when  $\lambda$  changes from 0.4 to 0.6, the selected stock pairs are almost identical. At that time, the change of  $\lambda$  cannot significantly affect the return; when  $\lambda$  changes from 0.6 to 0.7, the selected stock pairs change less. However, the positives and negatives of the paired-stock weights have changed. Therefore, compared with the GGR model, the optimized pairing strategy makes better use of stock price information and is more flexible.

**Stock trading stage**

The GGR model and the BQQ model use the same parameters set in the back-test. The trading period is 2016.01–2016.12. The results obtained are shown in Table 3. By comparing the back-test performance of the BQQ strategy with the GGR model, we arrive at five findings:

**Table 1** GGR model stock matching results

Formation period	Group	Stock A	Stock B	SSD
2016.01–2016.12	1	601,818	601,328	0.0954
2016.01–2016.12	2	601,288	600,016	0.1150
2016.01–2016.12	3	601,318	601,166	0.1328
2016.01–2016.12	4	601,998	601,005	0.1615
2016.01–2016.12	5	601,211	600,795	0.1660



**Table 2** BQQ model stock matching results

$\lambda$	Serial No.	Stock A	Stock weight	Serial No.	Stock B	Stock weight
0	1	600,795	0.5028	6	600,016	0.0136
	2	601,398	0.1417	7	601,766	-0.0224
	3	601,328	0.1249	8	600,028	-0.0504
	4	601,390	0.031	9	601,669	-0.0523
	5	601,166	0.0241	10	601,988	-0.8392
0.1	1	600,050	0.7513	6	601,336	-0.0657
	2	600,999	0.4585	7	600,150	-0.1127
	3	600,036	0.1822	8	601,688	-0.1165
	4	600,887	0.1213	9	600,104	-0.201
	5	601,318	-0.0251	10	601,800	-0.3252
0.2	1	601,888	0.3566	6	600,519	0.019
	2	600,606	0.2937	7	600,547	-0.0345
	3	601,601	0.0848	8	600,340	-0.1763
	4	601,336	0.0747	9	601,211	-0.4097
	5	601,688	0.0299	10	600,703	-0.7566
0.3	1	601,628	0.2169	6	600,109	0.0275
	2	600,893	0.1016	7	600,050	-0.009
	3	601,688	0.0875	8	600,150	-0.0099
	4	601,186	0.0665	9	600,999	-0.2757
	5	600,887	0.0381	10	601,857	-0.9232
0.4	1	600,050	0.8067	6	600,887	-0.0156
	2	601,186	0.249	7	601,601	-0.0225
	3	600,999	0.2128	8	601,628	-0.1203
	4	600,893	0.0647	9	600,150	-0.1544
	5	601,688	0.0181	10	601,800	-0.4454
0.5	1	600,050	0.7972	6	600,109	-0.0174
	2	601,186	0.2569	7	601,601	-0.0246
	3	600,999	0.2342	8	601,628	-0.1251
	4	600,893	0.0651	9	600,150	-0.1548
	5	601,688	0.0044	10	601,800	-0.4459
0.6	1	600,050	0.7886	6	601,601	0.0168
	2	601,186	0.2794	7	601,601	-0.1138
	3	600,999	0.1789	8	601,628	-0.1197
	4	601,688	0.0691	9	600,150	-0.1589
	5	600,893	0.065	10	601,800	-0.4542
0.7	1	601,800	0.2475	6	601,601	0.0522
	2	601,688	0.1843	7	600,893	-0.0448
	3	600,150	0.1535	8	600,111	-0.1994
	4	601,601	0.1162	9	600,999	-0.3851
	5	601,186	0.07	10	600,050	-0.8185

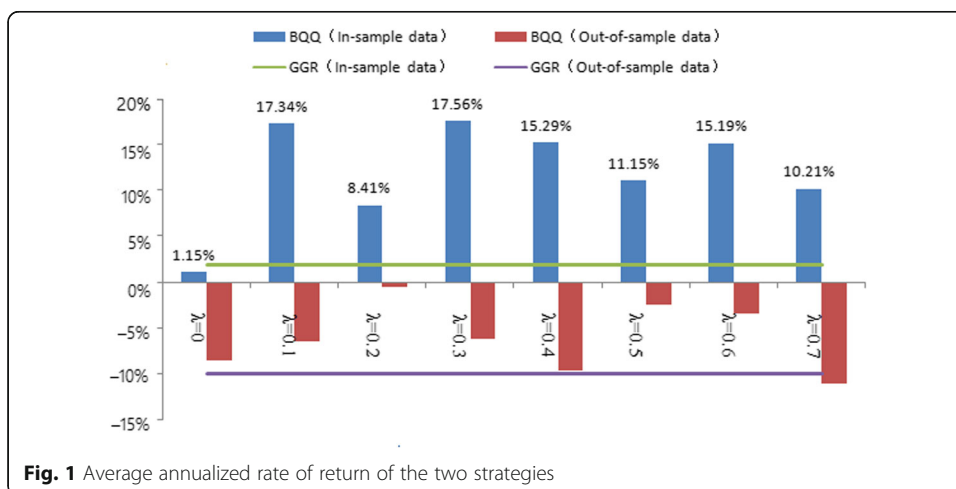
(1) The ability of the BQQ strategy to obtain revenue is significantly stronger than of the GGR model, which shows that the BQQ strategy is effective in increasing the volatility of the spread to improve the profitability of the pairs-trading strategy.

**Table 3** BQQ strategy performance

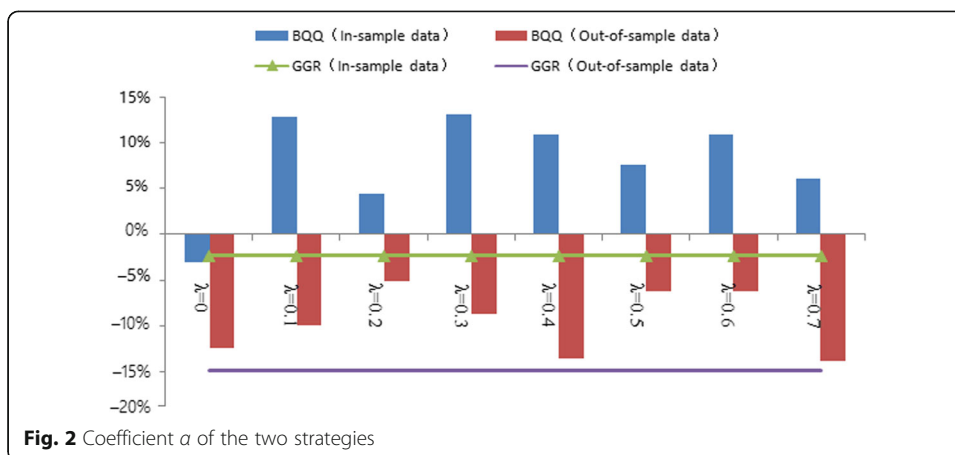
Trading period	$\lambda$	BQQ strategy performance							SSE 50 range ups and downs
		Strategic income	Annualized income	Coefficient $\alpha$	Coefficient $\beta$	Sharpe ratio	Win rate	Maximum withdrawal	
2016.01–2016.12	0	5.11%	5.24%	0.99	-0.035	0.334	0.449	1.35%	-5.53%
	0.1	5.64%	5.78%	0.012	-0.057	0.186	0.520	5.06%	-5.53%
	0.2	11.13%	11.41%	0.068	-0.065	1.028	0.633	2.28%	-5.53%
	0.3	25.49%	26.19%	0.209	-0.131	2.292	0.667	2.17%	-5.53%
	0.4	19.26%	19.78%	0.160	0.025	1.889	0.500	1.51%	-5.53%
	0.5	17.64%	18.11%	0.144	0.029	1.557	0.529	3.03%	-5.53%
	0.6	18.74%	19.24%	0.155	0.026	1.931	0.544	1.69%	-5.53%
	0.7	9.66%	9.91%	0.060	0.010	0.920	0.681	1.81%	-5.53%
GGR		2.67%	2.74%	-0.014	0.010	-0.406	0.438	2.27%	-5.53%

Figure 1 shows the average annualized rate of return of the BQQ strategy and the GGR strategy for different  $\lambda$  values (both in-sample data and out-of-sample data, respectively). For the in-sample rate of return, both strategies were carried out for a total of 32 back-tests, with a total of 31 positive gains. The return of the BQQ strategy is better than that of the GGR strategy in 87.5% of the cases. For the out-of-sample rate of return, the return of the BQQ strategy is better than that of the GGR strategy in 68.8% of the cases. To rule out the deviation of income caused by the different ways of opening a position, we also need to examine the coefficient of the two strategies and the Sharpe ratio.

As shown in Figs. 2 and 3, the BQQ model performs significantly better than the GGR model, both in terms of the coefficient  $\alpha$  and the Sharpe ratio. This result indicates that the BQQ model bears the average return of nonmarket risk during the four trading periods, and the average return on unit risk is higher than with the GGR model. Therefore, the better performance of the BQQ strategy is not from the strategy taking more market risk; rather, it is independent of the way the strategy is opened.



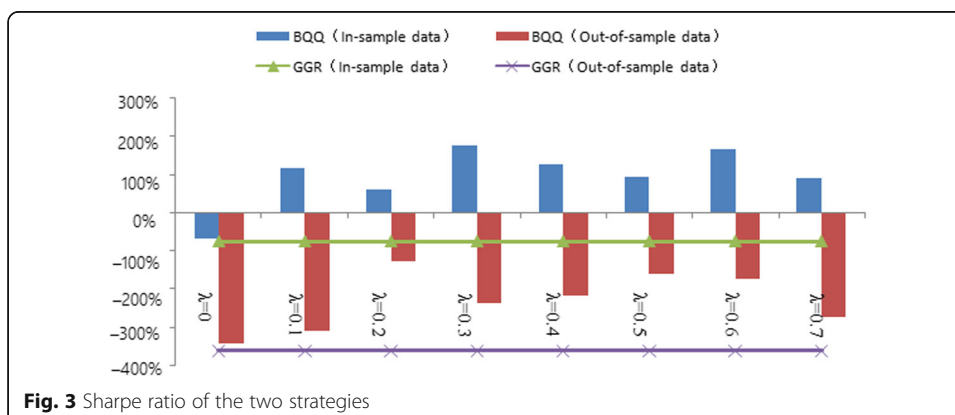
**Fig. 1** Average annualized rate of return of the two strategies



**Fig. 2** Coefficient  $\alpha$  of the two strategies

(2) The BQQ strategy has a strong ability to hedge the market. Table 4 shows the average value of the coefficient  $\beta$  of the BQQ strategy under different values of  $\lambda$ . It can be seen that the absolute value of  $\beta$  is below 0.1, which indicates and proves that the performance of the strategy is not affected by market fluctuations, which in turn proves that the pairs-trading strategy based on the minimum distance method can hedge market risk well. Compared with the GGR model, the coefficient  $\beta$  of the BQQ strategy is magnified because the GGR model uses a capital-neutral approach when in the opening position, while the BQQ strategy uses a coefficient-neutral approach. Due to the existence of the spread, the BQQ strategy cannot guarantee that the market value of the bought stock will be equal to the market value of the sold stock when the position is opened, which is equivalent to the fact that some net positions follow market ups and downs and the coefficient will increase.

(3) Similar to the GGR model, the BQQ strategy performs poorly in out-of-sample data. In the 32 out-of-sample back-tests, the annualized return of the BQQ strategy was positive only six times, and the coefficient  $\alpha$  was positive only eight times. The main reason for this phenomenon is the lack of rationality in the length of the formation period used at the stock-matching stage and the trading parameters used in the stock-trading stage. The yield of the GGR model is affected by trading



**Fig. 3** Sharpe ratio of the two strategies

**Table 4** Coefficient  $\beta$  of the two strategies

	In-sample				Out-of-sample			
	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6	Period 7	Period 8
BQQ	-0.025	-0.014	0.029	0.020	-0.076	-0.040	0.008	-0.039
GGR	0.010	0.006	0.039	-0.003	0.089	0.003	-0.036	-0.013

parameters in many cases, such as the formation period, trading period, and opening threshold. Since this article presents only a methodological improvement for the stock-pairing trading model, it does not provide a more in-depth study of trading parameters.

- (4) Performance of the BQQ strategy is very sensitive to the value of  $\lambda$ . Adjustable  $\lambda$  enhances the practicality of the strategy. In the same trading period, the return of the BQQ strategy does not show a monotonous change with  $\lambda$ . When the value of  $\lambda$  is too large, the stock-matching strategy is invalid because when  $\lambda$  increases, the volatility of the paired-stock spread is increasing, which means that the strategy is likely to obtain higher returns. Conversely, the increase of  $\lambda$  raises the risk of divergence in the spread, making it easier for the strategy to trigger a stop-loss signal and cause losses. Therefore,  $\lambda$  is a significant parameter to adjust the risk of the strategy, and the strategy user can adjust  $\lambda$  to match risk preferences, which enhances the usefulness of the strategy.
- (5) The optimal  $\lambda$  value is time dependent. The benefit of the BQQ strategy are non-monotonic changes in  $\lambda$ . Excessive  $\lambda$  assembly leads to the invalidation of the stock-matching strategy, which means that for a specific trading period, there is an optimized  $\lambda$  that maximizes the strategy’s return. From the perspective of revenue indicators and risk indicators, there are no obvious rules about the performance of the strategy and the change of  $\lambda$ . That is, the optimal  $\lambda$  value varies with the trading period and is time dependent.

Table 5 shows the values of coefficient  $\alpha$  and the Sharpe ratio from four out-of-sample back-tests. When  $\lambda$  is 0.5, coefficient  $\alpha$  and the Sharpe ratio take the maximum value at the same time.

The results show that when  $\lambda$  is 0.5, the average matching revenue of the optimized matching strategy for non-market risk in the four trading periods and the average return from unit risk are the largest, but the value needs to be verified by large-scale data.

**Conclusions**

By introducing multi-objective optimization to the GGR model, this paper considers the long-term equilibrium of stock prices and the volatility of spreads and establishes a BQQ model. This novel pairs-trading model provides a new perspective for pairs-trading strategy research. At the same time, it provides investors with a stock-matching

**Table 5** Values of coefficient  $\alpha$  and the Sharpe ratio for different  $\lambda$

	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$
Coefficient $\alpha$	-0.113	-0.076	-0.070	-0.112	-0.099	-0.062	-0.101	-0.139
Sharpe ratio	-3.268	-2.191	-1.833	-2.294	-1.913	-1.683	-2.243	-2.747

method that effectively improves the profitability of the trading strategy. This paper introduces the weight  $\lambda$  when solving bi-objective optimization problems, and these problems are transformed into single-objective optimization problems and solved by a sequential quadratic programming algorithm. To verify the effectiveness of the optimized pairing strategy, this paper selects the traditional GGR model as model for comparison and conducts back-testing on multiple time intervals on the SSE 50 constituents. We find that the BQQ strategy was able to obtain significantly higher revenue than the GGR model, and the adjustment of the weight  $\lambda$  increases the flexibility and practicality of the strategy.

This paper has some limitations. We used the SSE 50 Index as the research target in our empirical analysis. However, this was subject to the limitation of financing and securities lending; the small number of stocks may have affected the performance of the trading strategy. Additionally, when we performed the validity check of the optimized pairing strategy, there was scarce in-depth research available on the trading parameters and optimal values of the strategy, and this may have affected the profitability of the strategy to some extent. Therefore, subsequent research work should include these aspects. In the future, we will expand the number of stock share pools. In addition, the screening method for the transaction parameter of pairs-trading strategy requires in-depth research to find the right trading parameters for the BQQ strategy. Finally, we will try to establish an optimized pairing strategy by attaining the function of risk indicator  $\lambda$  through extended empirical analysis.

#### Abbreviations

BFGS: The Broyden–Fletcher–Goldfarb–Shanno algorithm; BQQ: Bi-objective quadratic programming with quadratic constraints; CAPM: Capital asset pricing model; GGR: The distance approach proposed by Gatev, Goetzmann and Rouwenhorst in 2006; RQQ: Revised quadratic programming with quadratic constraints; SSD: Sum of squared deviations; SSE: Shanghai Stock Exchange

#### Acknowledgements

Not applicable.

#### Authors' contributions

Diao contributed to the overall writing and the data analysis; Liu conceived the idea; Zhu contributed to the data collection and the data analysis. All authors read and approved the final manuscript.

#### Funding

The research is supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (No. 19XNH089).

#### Availability of data and materials

Shanghai Composite Index  
Please contact authors for data requests.

#### Competing interests

The authors declare that they have no competing interests.

Received: 9 July 2019 Accepted: 25 February 2020

Published online: 25 March 2020

#### References

- Bazaraa, M. S., Sherali, H. D., & Shetty, C. M. (2008). *Nonlinear programming: Theory and algorithms* (3rd ed.). Hoboken: Wiley.
- Bertram, W. K. (2010). Analytic solutions for optimal statistical arbitrage trading. *Physica A: Statistical Mechanics and its Applications*, 389(11), 2234–2243.
- Bondarenko, O. (2003). Statistical arbitrage and securities prices. *Review of Financial Studies*, 16(3), 875–919.
- Chen, H., Chen, S., Chen, Z., & Li, F. (2017). Empirical investigation of an equity pairs trading strategy. *Management Science*, 65(1), 370–389.
- D'Aspremont, A. (2007). Identifying small mean reverting portfolios. *Quantitative Finance*, 11(3), 351–364.
- Deb, K., & Sundar, J. (2006). Reference point based multi-objective optimization using evolutionary algorithms. In *Proceedings of the 8th annual conference on conference on genetic & evolutionary computation* (pp. 635–642).
- Do, B., & Faff, R. (2010). Does simple pairs trading still work? *Financial Analysts Journal*, 66(4), 83–95.

- Do, B., Faff, R., & Hamza, K. (2006). A new approach to modeling and estimation for pairs trading. In *Proceedings of 2006 financial management association European conference* (pp. 87–99).
- Dunis, C. L., & Ho, R. (2005). Cointegration portfolios of European equities for index tracking and market neutral strategies. *Journal of Asset Management*, 6(1), 33–52.
- Elliott, R. J., van der Hoek, J., & Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance*, 5(3), 271–276.
- Gatarek, L. T., Hoogerheide, L. F., & van Dijk, H. K. (2014). Return and risk of pairs trading using a simulation-based Bayesian procedure for predicting stable ratios of stock prices. *Electronic*, 4(1), 14–32 Tinbergen Institute Discussion Paper 14-039/III.
- Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *Social Science Electronic Publishing*, 19(3), 797–827.
- Hu, W., Hu, J., Li, Z., & Zhou, J. (2017). Self-adaptive pairs trading model based on reinforcement learning algorithm. *Journal of Management Science*, 2(2), 148–160.
- Jacobs, H., & Weber, M. (2011). *Losing sight of the trees for the forest? Pairs trading and attention shifts*. Working paper, October 2011. University of Mannheim. Available at <https://efmaefm.org/0efmsymposium/2012/papers/011.pdf>.
- Krauss, C. (2016). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys*, 31(2), 513–545.
- Miao, G. J. (2014). High frequency and dynamic pairs trading based on statistical arbitrage using a two-stage correlation and cointegration approach. *International Journal of Economics and Finance*, 6(3), 96–110.
- Perlin, M. (2007). *M of a kind: A multivariate approach at pairs trading (working paper)*. University Library of Munich, Germany.
- Peters, G. W., Kannan, B., Lasscock, B., Mellen, C., & Godsill, S. (2010). Bayesian cointegrated vector autoregression models incorporating alpha-stable noise for inter-day price movements via approximate Bayesian computation. *Bayesian Analysis*, 6(4), 755–792.
- Vidyamurthy, G. (2004). *Pairs trading: Quantitative methods and analysis*. Hoboken: Wiley.
- Wang, F., & Wang, X. Y. (2013). An empirical analysis of the influence of short selling mechanism on volatility and liquidity of China's stock market. *Economic Management*, 11(3), 118–127.
- Wang, S. S., & Mai, Y. G. (2014). WM-FTBD matching trading improvement strategy and empirical test of Shanghai and Shenzhen ports. *Economy and Finance*, 26(1), 30–40.
- Whistler, M. (2004). *Trading pairs: Capturing profits and hedging risk with statistical arbitrage strategies*. Hoboken: Wiley.
- Wu, L., & Cui, F. D. (2011). Investment strategy of paired trading. *Journal of Statistics and Decision*, 23, 156–159.
- Xu, L. L., Cai, Y., & Wang, L. (2012). Research on paired transaction based on stochastic spread method. *Financial Theory and Practice*, 8, 30–35.
- Zhang, D., & Liu, Y. (2017). Research on paired trading strategy based on cointegration—OU process. *Management Review*, 29(9), 28–36.
- Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4), 550–560.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---