Optimizing Statistical Arbitrage Models via Bayesian Method

Guo Liang
Institute of Statistics and Big Data
Renmin University of China

May 8, 2023

1 Introduction

1.1 Motivation

Speculators, hedgers, and arbitrageurs are the three groups of traders. Speculators need to take a lot of risk. Hedgers are risk-averse individuals who are mostly hedged by banks and market makers. Arbitrageurs seek to reduce as much risk as possible while locking in profits. Statistical arbitrage [1, 2] is using statistical analysis of historical data to guide arbitrage trading. It estimates the probability distribution of relevant variables and integrates fundamental data to guide arbitrage trading.

In general, we assume that stocks obey geometric Brownian motion (GBM)

$$\frac{dS}{S} = \mu_S dt + \sigma_S dW, \tag{1}$$

where μ_S is the drift, σ_S is the volatility and dW is the standard Brownian motion. The prices of stocks are not stationary. Co-integration analysis allows the construction of a stationary portfolio using multiple non-stationary stocks. Based on the mean-reversion property of the stationary series, positions are opened when the portfolio's price deviates and closed for gain after the price reverts. Our goals are as follows

- Heuristically select the stock pairs.
- Construct arbitrage model Using Bayesian method.
- Backtest with historical data and compare the results with the conventional method.

1.2 Data

Our data is from https://www.kaggle.com/datasets/borismarjanovic/price-volume-data-for-all-us-stocks-etfs. The dataset provide the full historical daily price and volume data for all US-based stocks and ETFs trading on the New York stock exchange (NYSE) and national association of securities dealers automated quotations (NASDAQ). The data is presented in CSV format as follows: Date, Open, High, Low, Close, Volume, OpenInt. The prices have been adjusted for dividends and splits.

1.2.1 Data Preprocess

The dataset includes data on 7,195 stocks in total, but many of these stocks have data that is incomplete or only available for a short period of time, making them useless for research. During the pre-processing step, we remove stocks whose sample size is less than 600. Among the remaining stocks, we randomly select 100 stocks and try to find pairs of stocks with co-integration relationships among these 100 stocks.

2 Exploratory Data Analysis

We first test for first-order single integer of the stock, i.e., the stock series is not smooth and the stock series is smooth after differencing. The single integer of a particular stock is shown in Figure 1. Similar tests can be done for the rest of the stocks. When all these stocks are homogeneous single integer (e.g., all are first-order single integer), we can look for pairs of stocks with co-integration relationship among these stocks.

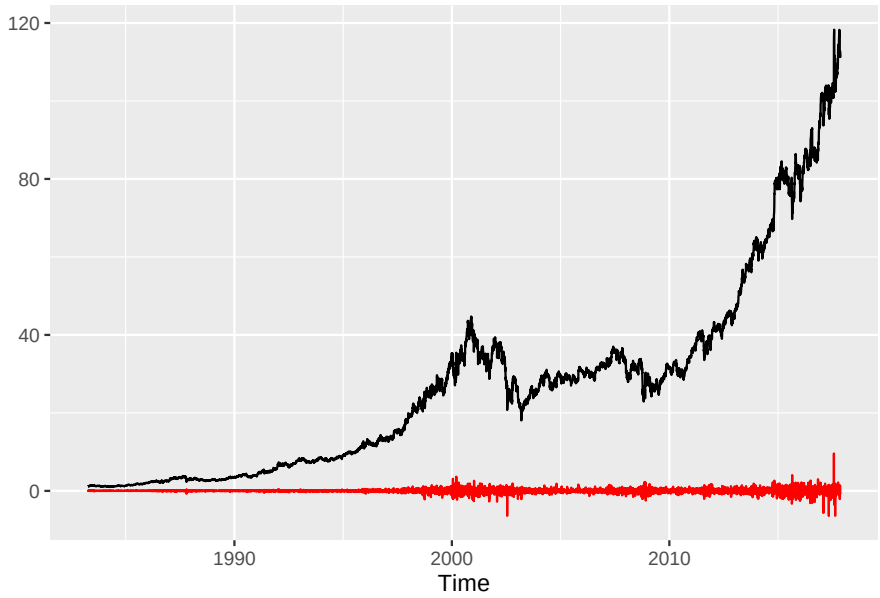


Figure 1: Stationary test.

Although we select only 100 stocks, it is still time consuming to select the co-integrated pairs directly among the 100 stocks. The correlated stocks are not necessarily co-integrated, such as Coca-Cola and Pepsi.[1] However, if the price movements of the two stocks are very similar, the probability of their co-integration is relatively higher. From (1), it is clear that the price movements of stocks are fully determined by the drift term μ_S and the volatility term σ_S. Therefore, for these 100 stocks, we estimate μ_S and σ_S separately using historical data [3]. The estimation method is as follows:

- Take observations $S_0, S_1, \ldots, S_n$ at intervals of T years (e.g. for daily data $T = 1/250$).

[1]The author thanks Prof. Ma Wei for pointing out my mistake in my slides, because correlation is not equivalent to co-integration, and in fact Coca-Cola and Pepsi are only correlated but not co-integrated.

- Calculate the continuously compounded return in each interval as:

$$u_i = \ln\left(\frac{S_i}{S_{i-1}}\right).$$

- Calculate the mean and standard deviation of the u_i's, denoted by $\bar{u}$ and s.
- The historical drift and volatility estimate are:

$$\frac{\bar{u} - s^2/2}{T}, \quad \frac{s}{\sqrt{T}},$$

where T is just a constant that affects the unit.

Then we cluster these 100 stocks with respect to these two variables and select the pairs of stocks with co-integration in each class. The clustering results are shown in Figure 2.

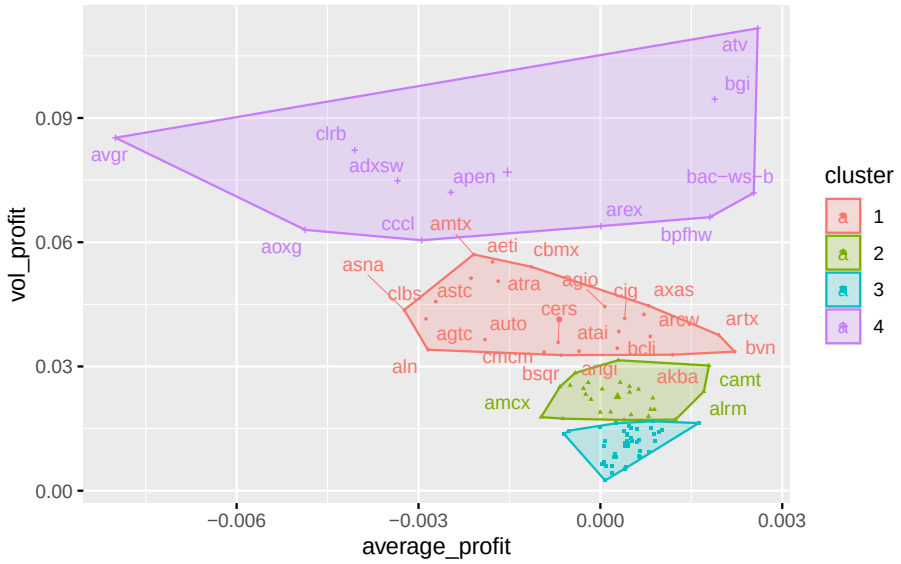


Figure 2: The clustering results.

In each cluster, we can perform the Engle-Granger (EG) co-integration test for each pair of stocks and make a heat map of the p-value of the test. Figure 3 shows the operation for the fourth cluster. We finally choose two stocks, bbk.us and bfy.us, which pass the co-integration test with the p-value equal to 0.02, and the co-integration property is maintained over a long period of time. The price trends of the two stocks are shown in Figure 4.

3 Method

Assuming that we have chosen stocks S_1 and S_2, a linear regression of S_1 on S_2 yields

$$S_1 = S_2\gamma + \mu + \epsilon, \tag{2}$$

Figure 3: Heat map of Clus 4.

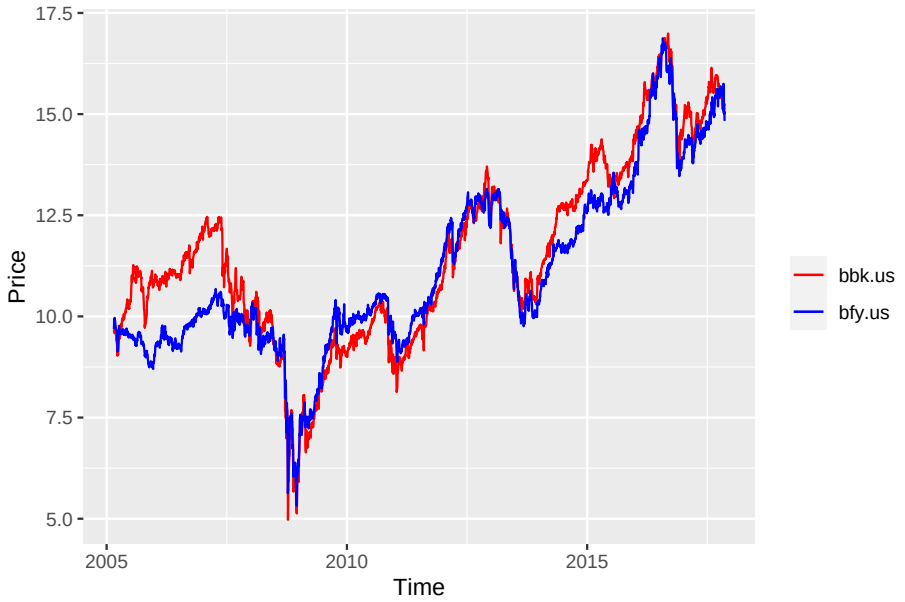


Figure 4: The price of bbk.us and bfy.us.

where the slope γ represents the hedge ratio and ϵ is a mean-reverting process with zero mean. Here we set the variance of ϵ as σ^2. We can estimate γ using ordinary least square (OLS) method and set the estimator as $\widehat{\gamma}$, then $S_1 - S_2\widehat{\gamma}$ is a mean-reverting process with mean μ and variance σ^2. If we know μ and σ for any spread s precisely, a pairs trading strategy can be implemented as follows: when the latest spread s exceeds $\mu(s) + \delta\sigma(s)$, S_1 is over-valued, and S_2 is under-valued, therefore we open 1 unit of short position for S_1 , and γ unit of long position for S_2; when the latest s is less than $\mu(s) - \delta\sigma(s)$, S_1 is under-valued, and S_2 is over-valued, as a result we open 1 unit of long position for S_1, and γ unit of short position for S_2. The parameter δ here is a threshold number to open trades.

To control the risk, we also need to apply the stop loss rule to the strategy: if we are long S_1, and short S_2, when s does not mean-revert to its historical mean $\mu(s)$, but instead deviates further to be even smaller than $\mu(s) - \Delta\sigma(s)$, where Δ is a multiplier which is usually larger than entry threshold δ, we will close the position and realize the loss. Similarly, when we are short S_1, long S_2, and s does not mean-revert to its historical mean $\mu(s)$, but moves further to be even larger than $\mu(s) + \Delta\sigma(s)$, the position will be closed and a loss will be realized. Similar to δ, the exit threshold Δ is also a number to be used for closing trades.

However, μ and σ are unknown and we should estimate them using the data. We use the Bayesian method to estimate the values of μ and σ. We use a conjugate prior [4]

$$(\mu, \sigma^2) \sim N - Inv - \chi^2(\mu_0, \sigma_0^2/v_0; v_0, \sigma_0^2),$$

i.e.,

$$\mu|\sigma^2 \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right), \quad \sigma^2 \sim Inv - \chi^2(v_0, \sigma_0^2).$$

Given the data y, the posterior distributions of μ and σ are

$$p(\mu, \sigma^2|y) \propto p(y|\mu, \sigma^2)p(\mu, \sigma^2) = ... = N - Inv - \chi^2(\mu_n, \sigma_n^2/\kappa_n; v_n, \sigma_n^2), \tag{3}$$

where

$$\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y}, \\
\kappa_n &= \kappa_0 + n, \\
v_n &= v_0 + n, \\
v_n\sigma_n^2 &= v_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}\left(\bar{y} - \mu_0\right)^2.
\end{aligned}$$

We then use the mean of the posterior distribution as the estimates of μ and σ^2.

4 Results

In this section, we compare the results of our method with the conventional method. First we need to divide the samples into a training set and a testing set. We use the stock price data of the last 1000 days as the testing set and use the samples within the remaining time as the training set. The training set contains 2197 days of stock data. We compare the following two methods.

- **Non-Bayes:**
  - *Step 1:* Estimate the hedge ratio γ using OLS and set the estimate as $\widehat{\gamma}$.
  - *Step 2:* Set

$$\widehat{\mu} = \frac{1}{2197}\sum_{s=1}^{2197} S_1(s) - \widehat{\gamma}S_2(s), \widehat{\sigma}^2 = \frac{1}{2197}\sum_{s=1}^{2197}\left(S_1(s) - \widehat{\gamma}S_2(s) - \widehat{\mu}\right)^2.$$

– *Step 3:* Trading according Section 3 with $\delta = 0.5$ and $\Delta = 3$.[2]

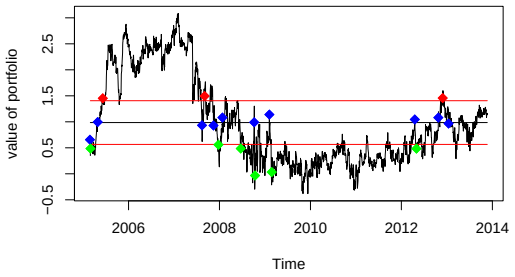
- **Bayes:**
  - *Step 1:* Estimate the hedge ratio γ using OLS and set the estimate as $\widehat{\gamma}$.
  - *Step 2:* Set the prior parameter μ_0, κ_0, s_0^2 and ν_0, Estimate $\widehat{\mu}$ and $\widehat{\sigma}$ using (3) and Markov Chain Monte Carlo (MCMC) method. During the implementation, we set $\mu_0 = \widehat{\mu}$, $v_0 = 2\widehat{\sigma}^2/(\widehat{\sigma}^2 - 1)$, $\kappa_0 = 0.01$ and $s_0^2 = 1$, where $\widehat{\mu}$ and $\widehat{\sigma}^2$ are the same as **Non-Bayes**. Besides, we set the number of MCMC as 10000 and the number of burn-in as 1000.
  - *Step 3:* Trading according Section 3 with $\delta = 0.5$ and $\Delta = 3$. The estimates of the parameters μ and σ are gradually updated as time goes on and we delay the opening of positions to reduce the probability of stop loss.

Figure 5 shows the performance of our method and the conventional method for in-sample and out-of-sample data. Figure 5a and Figure 5b show the performance of the conventional method for in-sample and out-of-sample data, and Figure 5c and Figure 5d show the performance of our method for in-sample and out-of-sample data, respectively. The horizontal axis in the figure represents time and the vertical axis represents the value of the portfolio. The horizontal line in the middle represents the mean value at the current time, i.e., μ. We add the upper and lower 0.5σ lines as a hint to open a position. The red dots in the graph represent the time when we sell the portfolio, the green dots represent the time when we buy the portfolio, and the blue dots represent the time when we close the position. As can be seen from the figure, the estimates of μ and σ obtained by our method are significantly different compared to the conventional method. In the in-sample performance, the number of trades and the location of the trading points do not differ much between the two methods, however, in the out-of-sample performance, the conventional method underestimates the mean value, while the Bayesian method does a good job of pulling the estimates in the right direction.

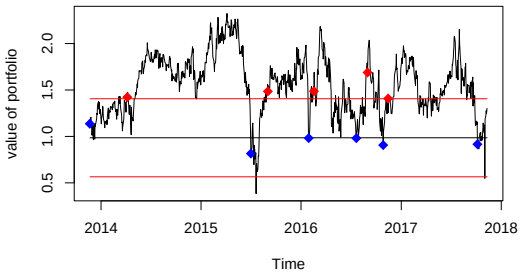
To more visually demonstrate the advantages of Bayesian methods in statistical arbitrage, we summarize the annualized return, and maximum drawdown of both methods for in-sample and out-of-sample performance in Table 1. As can be seen in Table 1, for the in-sample data, although the use of the Bayesian approach results in some decrease in annualized returns, the maximum drawdown also decreases by approximately 18%. For the out-of-sample data, the Bayesian method increases the annualized return by approximately 4% and decreases the maximum drawdown by approximately 26%. This shows that Bayesian method can help us to get more accurate parameter estimates and thus establish positions and close them at better times to obtain more stable and substantial returns.

Figure 6 illustrates the excess returns of our approach using the S&P 500 as a benchmark. As can be seen from the figure, while our arbitrage does not deliver excess returns until June 2015, statistical arbitrage can deliver substantial excess returns as the asset holding time grows. The annualized excess return from using the Bayes method is calculated to be approximately 28%. However, the relative volatility of our method's return is also greater than that of the S&P 500 as seen in the graph.

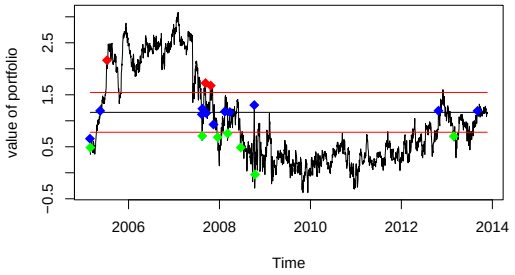
[2]Although [5] considered how to optimize the threshold, since we mainly consider the advantages of Bayesian method in statistical arbitrage, we directly set the threshold as $\delta = 0.5$ and $\Delta = 3$.



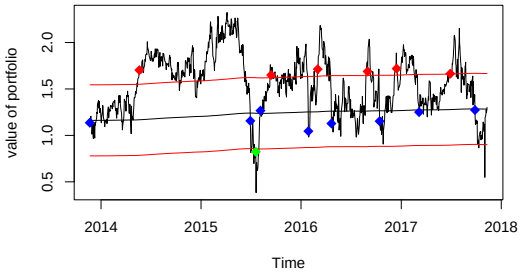
(a) Non-Bayes for in-sample



(b) Non-Bayes for out-of-sample



(c) Bayes for in-sample



(d) Bayes for out-of-sample

Figure 5: Comparison of the results for in-sample and out-of-sample.

Table 1: Comparison of the results for in-sample and out-of-sample.

In-sample	*Method*	*Annualized Return*	*Maximum Drawdown*
	Non-Bayes	31.51%	59.72%
	Bayes	29.51%	41.13%
Out-of-sample	*Method*	*Annualized Return*	*Maximum Drawdown*
	Non-Bayes	26.25%	71.14%
	Bayes	30.33%	45.82%

5 Discussion

We first selected stock pairs with co-integration relationships from the original list of stocks by methods such as clustering and co-integration analysis. This step is not the focus of this report, but it is crucial. Only when two stocks have a co-integration relationship, we can construct a portfolio and arbitrage. This is what the author wants to investigate further afterwards.

After selecting the appropriate stocks, we construct the portfolio and estimate the value level and volatility level of the portfolio through linear regression and Bayesian methods to construct a trading strategy. The experiments show that the Bayesian approach has better estimation of the value level and volatility level of the portfolio and can generate more stable excess returns compared to the conventional approach.

Some additional experimental results can verify the superiority of our approach over

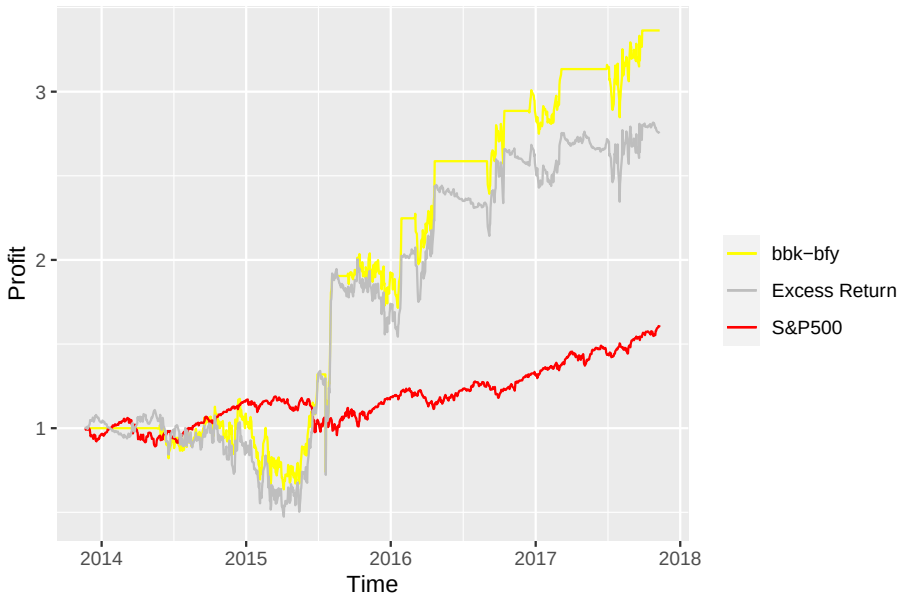


Figure 6: Portfolio excess return on S&P 500 index.

conventional methods. For example: We choose the data of Coca-Cola and Pepsi from 2008 to 2014 as the training set and the data from 2014 to 2018 as the testing set. The p-value obtained from the EG co-integration test within the training set is 0.70, which means that the two stocks are not co-integrated.[3] Table 2 shows the out-of-sample annualized return, Sharpe ratio and maximum drawdown for the Coca-Cola Pepsi portfolio. We can find that the conventional statistical arbitrage may incur losses when there is no co-integration relationship between the two stocks, and the use of Bayesian approach can reduce the losses to some extent. (We do not take into account transaction fees, so the return should be lower).

Table 2: Comparison of the results for out-of-sample of Coca-Cola and Pepsi.

Method	*Annualized Return*	*Sharpe Ratio*	*Maximum Drawdown*
Non-Bayes	-90.52%	-0.48	75.37%
Bayes	5.62%	1.06	48.70%

In the future, I will consider how to select stocks more efficiently while better monitoring whether the co-integration of stock pairs is maintained. Also, I plan to set adaptive thresholds to get more stable returns.

Acknowledgement

The author thanks Prof. Ma Wei for his insightful comments and valuable suggestions that improve my final report. The author also thanks the provider of the dataset that allows my analysis to proceed.

[3]In my slides, I mistyped the stock code of Coca-Cola. In the report, we also show the results of these two stocks.

References

[1] Ganapathy Vidyamurthy. *Pairs Trading: quantitative methods and analysis*, volume 217. John Wiley & Sons, 2004.

[2] Simão Moraes Sarmento and Nuno Horta. Enhancing a pairs trading strategy with the application of machine learning. *Expert Systems with Applications*, 158:113490, 2020.

[3] John C Hull. *Options futures and other derivatives*. Pearson Education India, 2003.

[4] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. CRC press, 2013.

[5] Zhengqin Zeng and Chi-Guhn Lee. Pairs trading: optimal thresholds and profitability. *Quantitative Finance*, 14(11):1881–1893, 2014.