The progress of State-of-art Arbitrage Models Based on Bigdata Analysis and Machine Learning Approaches

Hongpu Chen*

Department of Basic science of Mathematics, Dalian University of Technology, Dalian, China

*Corresponding author: hc302@student.le.ac.uk

Keywords: Arbitrage, Underlying Assets, Big Data, Artificial Intelligence.

Abstract: Starting from the principle of arbitrage, this article introduces the traditional arbitrage models in different forms. Subsequently, the designs of arbitrage model based on big data and neural network in the field of modern innovation are demonstrated, and empirical results are given for evaluation. Afterwards, the drawbacks of the state-of-art models are discussed and the future prospects are proposed accordingly. Overall, these results offer a guideline for future arbitrage models construction.

1. Introduction

As a term in economics and finance, arbitrage refers to the behavior of benefitting from the price variations between multiple underlying assets among (or same assets in different markets) based on the pair transactions. Profit is the difference between market prices for single transactions. When academic circles use arbitrage, arbitrage refers to transactions that do not involve negative cash flow in any probability or time state, but involve positive cash flow in at least one state, e.g., there is the possibility of arbitration where it is possible to buy an item immediately at a lower price and sell it at a higher price [1].

Pure arbitrage refers to an investment strategy that can be used in many fields, that is, investors buy and sell securities in different markets at the same time, and benefit from the price difference in the purchase and sale. On this basis, the terms “arbitrage” and “pure arbitrage” in the financial field are commonly utilized exchangeable. Each time a company's asset is traded in multiple markets, the price may not be synchronized temporarily, i.e., it becomes possible for pure arbitrage [2]. At the same time, it is also possible in situations that exchange ratios result in price differentials, no matter how small. In the final analysis, pure arbitrage is a strategy for investors to take advantage of the inefficiency of the market. With the development of technology and the increasing digitization of transactions, using these scenarios has become more challenging, because the pricing errors of commodities can now be quickly identified and solved by the trading system. This means that the ability to rely on pure arbitrage has become scarce. Overall, there are various of arbitrage types as categorized to 4 types in Refs. [3].

Merger arbitrage is an arbitrage related to the merged entity, such as two listed companies. Generally speaking, it consists of two sides that are the acquisition corporation and its target corporation. If the target company is a listed company, the acquiring company must purchase the issued shares of the company. As a matter of fact, it is a superior approach to determine which stocks are traded, and to announce the profits of shareholders. With the opening of the transaction, traders who want to profit from the transaction buy the shares of the target company to make it closer to the published trading price. The price of the target company rarely matches the transaction price. However, due to the risk that the transaction may fail or fail, it usually trades at a slightly discounted price. There may be several reasons for the failure of the transaction, including changes in market conditions or the refusal of regulators. A common case of it is that investors buy the stocks at a discount price, and then make a profit after the transaction is completed. For example, investors who believe that the transaction may fail will sell (short) the stocks accordingly. Like other bonds, the core

of convertible bonds is corporate bonds that can pay interest to bondholders, which also paves a path for convertible arbitrage. The main difference between convertible bonds and traditional bonds is that for convertible bonds, bondholders have the right to convert them into shares of affiliated companies in the future, usually at the discount rate. Companies issue convertible bonds because it allows them to provide lower interest payments. As for the position investors hold and the proportion of buying and selling, it depends on whether investors believe that the pricing of bonds is fair. When bonds are considered cheap, they usually short on stocks and long on bonds. On the other hand, if the investors think bonds are overpriced or overpriced, they may be long in stocks and short in bonds [4].

This article mainly describes the previous research on the paired trading framework, i.e., the relative value arbitrage strategy involving two or more securities. There are five groups of research contents. First of all, distance method uses nonparametric distance measurement to identify paired trading opportunities. Besides, the cointegration method relies on the formal cointegration test to reveal the stable spread time series.In addition, time series method focuses on finding the best trading rules of mean reversion spread. Moreover, the purpose of stochastic control method is to determine the optimal portfolio holding relative to other available securities in paired trading. The "other methods" category contains more relevant pairs trading frameworks. Finally, the profitability of paired transactions is evaluated according to market friction, and finally the advantages and disadvantages related to further research and implementation are revealed [5].

The rest part of the paper is organized as follows. The Sec. 2 will introduce the basic situation of arbitrage and the principle of arbitrage. The Sec. 3 will list and compare different arbitrage models. The Sec. 4 introduces the arbitrage method under the background of big data and artificial intelligence. Finally, the limitations and future prospects of current research in this field are given.

2. Basic Description of Arbitrage

2.1 The origin of arbitrage

Records of ancient business practices are difficult to find and incomplete. The existing evidence mainly comes from the Middle East, which shows that the commercial trade in the ancient market is very extensive, which provides many ways for risk arbitrage [6]. Potential opportunities are affected by plenty of factors. The history of it can be dated back to around 1760 BC. The Sumerian cuneiform tablet of that era shows that a basic form of bill of exchange transaction was used, that is, payment at one place of the local unit of account, such as barley, in exchange for the agreed amount of local currency to be paid at another place in the future. The date is usually determined by the acceptable time of transportation between two locations. Two weeks to one month is a common time between payment and repayment. The specific place of payment for such transactions is usually in temples. Fig. 1 gives a sketch of the price ratio evolution.

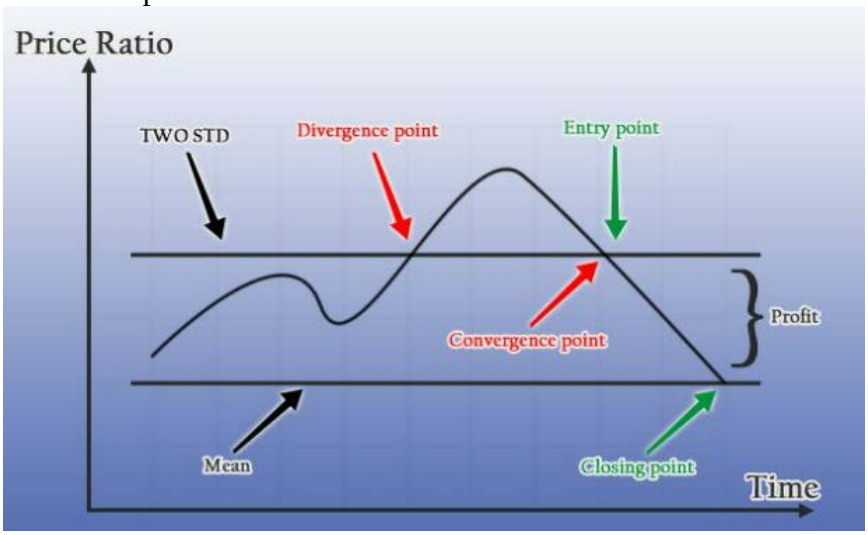

Figure 1. Price Ratio-Time Picture [7].

Table 1. Arbitrage methods of different assets

Arbitrage Modes	Definition	Application Area
Spatial arbitrage	This refers to the practice of buying a commodity at a low price in one market, carrying the same commodity at a higher price in another market, and thus earning the difference between the two markets	All markets
Municipal bond arbitrage	Municipal bond arbitrage refers to the practice of exchanging bonds issued for funds in order to make a profit [9].	Municipal bond market, enterprise bond, local government, urban government, urban construction
Convertible bond arbitrage	Convertible bond arbitrage indicates the profit-making strategy based on the inefficiency of pricing between the bonds and underlying assets [10].	Bond industry
Cross-border arbitrage	Commodity trading in the international market [11].	Cargo trade, hotel accommodation, study abroad education, air tickets, international exhibitions, tourism services, software services, international transportation, international conferences and communication services
Regulatory arbitrage	Regulatory arbitrage captures the disequilibrium in the market, and the key factor determining the attractiveness of arbitrage opportunities is the transaction cost of arbitrage strategy [12].	Transfer from one regulatory body to another; Transfer from one time period to another; From one market subject identity to another; From one business form to another; Transfer from one disclosure method to another
Statistical arbitrage	It refers to the investment process of a model. At this time, without considering the economic meaning, the quantity is directly used for fund portfolio, and then the multiple short positions of the portfolio are built according to the comparison between the securities price and the predicted price of the model, so as to avoid the market and return to stable income [13].	Matching transactions, stock index hedging, securities lending hedging and foreign exchange hedging transactions

Table 2. The backtesting results

Method	ID, TIR+EX, H	Nib of Stocks	Before Transaction Costs											After T.C.		
			Yearly Ret	Mean	Max	Min	Std	t-stat	Skew	Kurt	VAR	Sharpe Ratio	Max DD	Yearly Ret	Daily Ret	Sharpe Ratio
DBN	No, No, H=1	100	55.86	0.18	15.08	-12.68	1.75	8.25	0.79	14.01	-4.80	9.69	39	27.6	0.1	5.12
DBN	No, yes, H=1	100	24.92	0.09	12.73	-17.29	2.00	4.06	-0.39	12.96	-5.85	3.98	57	2.27	0.01	-0.02
DBN	No, No, H=1	300	65.17	0.20	19.41	-17.70	2.33	7.28	0.17	12.52	-6.32	8.14	69	35.22	0.12	4.70
DBN	No, yes, H=1	300	34.11	0.12	15.37	-22.6	2.52	4.37	-0.37	12	-7.62	4.35	73	9.79	0.04	1.17
RF	No, No, H=1	100	52.51	0.17	14.42	-11.2	1.28	10.32	0.82	16.98	-3.43	8	40	16.78	0.06	4.43
RF	No, yes, H=1	100	41.19	0.14	11.84	-16.17	1.81	6.37	-0.76	16.72	-5.58	7.16	50	-5.39	-0.02	-1.68
RF	No, No, H=1	300	92.51	0.26	21.88	-15.52	1.65	12.51	0.76	16.65	-4.24	15.13	32	29	0.1	5.43
RF	No, yes, H=1	300	61.12	0.19	17.82	-19.36	2.14	7.47	-0.12	13.58	-6.23	8.39	47	7.97	0.03	0.92

2.2 Principle of arbitrage

Arbitrage is a transaction that takes advantage of the small price difference between the same assets in two or more markets. Arbitrage traders buy assets in one market and sell assets in another market to make up for the difference between the two prices. “Real" arbitrage requirements do not involve market risk. When securities are traded on multiple exchanges, arbitrage is realized by buying one exchange and selling another exchange at the same time. To give an analytical evaluation, one can has a definition as $P(PV_t \geq 0) = 1$ and $P(PV_t \neq 0) > 0$. Here, PV_0 is set as 0 and PV_t denotes the portfolio value as a function of time t [8].

3. Arbitrage Modes

Table. I list the commonly utilized arbitrage models retrieved from various kinds of literatures. Arbitrage can be understood on the table. Different arbitrage models correspond to different assets. In arbitrage trading, if the price increases for the long one while decreases for the short one, investors will be able to gain from the fluctuation in the price variations. On the contrary, investors have losses.

4. Arbitrage with big data & machine learning

Ref. [14] developed several independent statistical arbitrage strategies based on three methods (random forest, deep belief network and elastic network regression) using a database containing hundreds of predictors. Based on these relevant methods, the key results are shown in Table 2 and basic statistical data. Between 1993 and 2015, all models produced positive (excess) returns before transaction costs. The annualized rate of return is between 7% and 92%, with an average of about 35%, much higher than the market performance in the same period (about 10%). If the transaction cost is included, the average annualized return of 40 models is 11%, and only 6 models have negative returns. Due to frequent rebalancing, the experiment contains a holding period of one or five days,

and these strategies are affected by the transaction costs. Empirically, excluding transaction costs, in the case of the research model, on average, when all predictors are included, the performance is significantly worse than that considering only stock lag return (LR). Compared with random forest, this decline of DBN is more obvious.

5. Limitations and future outlook

First of all, arbitrage has market constraints, some of short-term arbitrage models are impossible for investors to obtain benefits. Secondly, there may be market risk in arbitrage, and some commodities or stocks will be overestimated by the market, resulting in deviation. Thirdly, the arbitrage model may also have deviation, the model analyzed may be incorrect, or two unrelated models. Furthermore, there may be noise trading risks, e.g., external public opinion might lead investors to cognitive bias, resulting in arbitrage failure. In the future, investors may avoid some risks that can be controlled, turn these risks into artificial intelligence control, and judge more accurately when to make profits, so as to reduce the limitations of arbitrage.

6. Conclusion

In summary, this paper introduces the state-of-art arbitrage models based on literature review. Specifically, different arbitrage modes are listed and the innovation of arbitrage in the field of big data and artificial intelligence is evaluated. The purpose of the paper is to make people more rational to carry out arbitrage strategy and choose the right arbitrage mode in different market systems, so as to guide future studies of arbitrage in the field of innovation. This paper mainly shows the application of the current arbitrage investment strategy from the field of big data and machine learning. In the current era, arbitrage needs to be further combined with artificial intelligence and algorithm, so as to find a more market-oriented arbitrage strategy.

References

[1] Dybvig P H, Ross S A. Arbitrage [M]//Finance. Palgrave Macmillan, London, 1989: 57 - 71.

[2] Huberman G. A simple approach to arbitrage pricing theory [M]//Theory of Valuation. 2005: 289 - 308.

[3] Kirchner T. Merger arbitrage: how to profit from global event-driven arbitrage [M]. John Wiley & Sons, 2016.

[4] Calamos N P. Convertible arbitrage: insights and techniques for successful hedging[M]. John Wiley & Sons, 2011.

[5] Krauss C. Statistical arbitrage pairs trading strategies: Review and outlook [J]. Journal of Economic Surveys, 2017, 31 (2): 513 - 545.

[6] Delbaen F, Schachermayer W. The mathematics of arbitrage [M]. Springer Science & Business Media, 2006.

[7] The figure is collected from: https://in.pinterest.com/pin/499899627362010940/.

[8] Deng G, McCann C. Leveraged Municipal Bond Arbitrage: What Went Wrong? [J]. The Journal of Alternative Investments, 2012, 14 (4): 69 - 78.

[9] Agarwal V, Fung W H, Loon Y C, et al. Risk and return in convertible arbitrage: Evidence from the convertible bond market [J]. Journal of Empirical Finance, 2011, 18 (2): 175 - 194.

[10] Rosenbloom H D. Cross-border arbitrage: the good, the bad and the ugly [J]. Taxes, 2007, 85: 115.

[11] Fleischer V. Regulatory arbitrage [J]. Tex. L. Rev., 2010, 89: 227.

[12] Avellaneda M, Lee J H. Statistical arbitrage in the US equities market [J]. Quantitative Finance, 2010, 10 (7): 761 - 782.

[13] Huck N. Large data sets and machine learning: Applications to statistical arbitrage [J]. European Journal of Operational Research, 2019, 278 (1): 330 - 342.

[14] Huck N. Large data sets and machine learning: Applications to statistical arbitrage [J]. European Journal of Operational Research, 2019, 278 (1): 330 - 342.