Liquidity and Arbitrage Cost between ETF and Stocks using Agent-Based Model

Takanobu Mizuta

February 25 2019

Vol. 27
This material was compiled based on the results of research and studies by directors, officers, and/or employees of Japan Exchange Group, Inc., its subsidiaries, and affiliates (hereafter collectively the “JPX group”) with the intention of seeking comments from a wide range of persons from academia, research institutions, and market users. The views and opinions in this material are the writer’s own and do not constitute the official view of the JPX group. This material was prepared solely for the purpose of providing information, and was not intended to solicit investment or recommend specific issues or securities companies. The JPX group shall not be responsible or liable for any damages or losses arising from use of this material. This English translation is intended for reference purposes only. In cases where any differences occur between the English version and its Japanese original, the Japanese version shall prevail. This translation includes different values on figures and tables from the Japanese original because methods of analysis and/or calculation are different. This translation does not include some paragraphs which Japanese original includes. This translation is subject to change without notice. The JPX group shall accept no responsibility or liability for damages or losses caused by any error, inaccuracy, misunderstanding, or changes with regard to this translation.
Liquidity and Arbitrage Cost between ETF and Stocks using Agent-Based Model

Takanobu Mizuta

February 25 2019

Abstract

In recent years, exchange-traded funds (ETFs) have been widely spread to individual investors as an easy way to diversify their investments. Some ETFs, however, are not traded with enough volume to discover an adequate price, making them difficult for individual investors to trade. Therefore, to increase the liquidity of low-liquidity ETFs, some stock exchanges have introduced an incentive scheme in which arbitrage orders are charged lower trading fees. The questions, however, of how the liquidity changes depending on arbitrage trading costs and what the mechanism is remain to be answered.

Therefore, in this study I built an artificial market model, which is a kind of agent-based model, containing an ETF, two stocks, and an arbitrage trader. I then investigated the relationship between the liquidity of the ETF and the trading costs.

The results showed that, because the prices of each risk asset fluctuate in their volatility, when the volatility is sufficiently greater than the cost, the arbitrage agent has more chances to make arbitrage trades. Then, when the arbitrage agent trades more, the market price differential becomes lower. In addition, lower cost increases the depth of waiting trades for the ETF, whereas the depth tendency for a stock is the opposite. Furthermore, lower cost increases the trading volume of both. Lower cost reduces the depth and increases the trading volume for a stock because orders for arbitrage trades and the waiting orders for a stock are matched. In real financial markets, however, there are traders who order more when the trading volume increases. My model did not implement this behavior. It is possible that lower cost would increase both the depth and the trading volume because of such behavior, but this remains a future work. I also investigated the case of more liquidity for an ETF and found that the market price differential becomes larger. Even though more arbitrage trades occur because of the greater liquidity, the market price differential rate does not improve. This result implies that, when the trading volume of an ETF increases to near that of a stock, improving the market price differential is more difficult by arbitrage trades like those modeled in this study. Therefore, this study implies that other ways are needed to improve the market price.
1 Introduction

An exchange-traded fund (ETF) is a mutual fund that invests in a diversified portfolio of many stocks or bonds and is also listed and traded at a stock exchange. In recent years, ETFs have been widely spread to individual investors as an easy way to diversify their investments. Some ETFs, however, have not been traded with enough volume to discover an adequate price, making them difficult for individual investors to trade.

An ETF is exchangeable with all stocks held by the ETF. Therefore, when the price of the ETF and the total value of the stocks held by the ETF differ, a trader can buy the cheaper asset, exchange it\(^1\), sell the more expensive asset, and thus earn a profit from the price difference. Such a trade is called an “arbitrage trade.” It is said that increasing the number of traders who perform such arbitrage trades is very important for discovering adequate prices and increasing liquidity.

For example, to increase the liquidity of low-liquidity ETFs, in 2018 the Tokyo Stock Exchange introduced a market-making incentive scheme, in which designated market makers always place orders in return for incentives such as lower fees\(^2\).

The questions, however, of how liquidity changes depending on arbitrage trading costs and of what the mechanism is remain to be answered.

Empirical studies cannot be conducted to investigate situations that have never occurred in actual financial markets, and so many factors affect price formation and liquidity in actual markets that an empirical study cannot be conducted to isolate the direct effect on liquidity. In contrast, artificial market simulation\(^3\) using a kind of agent-based model can isolate the pure contributions of changes to liquidity and simulate changes that have never been observed. These are strong points of artificial market simulation studies. Articles in both Nature (Farmer and Foley (2009)) and Science (Battiston et al. (2016)) have argued that artificial market studies are expected to contribute to greater understanding of actual markets.

Many previous artificial market studies contributed to explaining the nature of financial market phenomena such as bubbles and crashes. Recent artificial market studies have also contributed to discussions of appropriate financial regulations and rules \(^4\). The JPX Working Paper series includes various studies contributing to such discussions.

There have been previous studies using an artificial market model to investigate arbitrage trades between futures, ETFs, and stocks (Xu et al. (2014); Torii et al. (2015)). No previous study, however, has used an artificial market model to investigate the relationship between liquidity and arbitrage.

---

\(^1\) In a real financial market, traders make such exchanges after trading hours.

\(^2\) [https://www.jpx.co.jp/english/equities/products/etfs/market-making/](https://www.jpx.co.jp/english/equities/products/etfs/market-making/)

\(^3\) Excellent reviews include LeBaron (2006); Chen et al. (2012); Todd et al. (2016).

\(^4\) Mizuta (2016) is a great review.
trading costs.

Therefore, in this study, I expanded the artificial market model of Mizuta et al. (2013) to include three risk assets—two stocks and an ETF—as shown in Figure 1. I also added an arbitrage agent to perform arbitrage trading among these risk assets. I then investigated the relationship between the liquidity of an ETF and the trading costs.

2 Artificial Market Model

The model of Chiarella and Iori (2002) is very simple but replicates long-term statistical characteristics observed in actual financial markets: a fat tail and volatility clustering. In contrast, Mizuta et al. (2013) replicates high-frequency micro structures, such as execution rates, cancel rates, and one-tick volatility, that cannot be replicated with the model of Chiarella and Iori (2002).

In this study, I expanded the artificial market model of Mizuta et al. (2013) to include three risk assets, denoted as stock 1, stock 2, and ETF (Figure 1), and I added an arbitrage agent to perform arbitrage trading among these risk assets. The simplicity of the model is very important for this study, because unnecessary replication of macro phenomena leads to models that are overfitted and too complex. Such models prevent understanding and discovery of mechanisms affecting price formation because of the increase in related factors. I explain the basic concept for constructing our artificial market model in the Appendix “Basic Concept for Constructing Model”.

In the expanded model here, the value of ETF is exactly the same as the sum of the values of stocks 1 and 2. The exchange market for each of the three risk assets adopts a continuous double auction to determine the market price. In this auction mechanism, multiple buyers and sellers compete to buy and sell financial assets in the market, and transactions can occur at any time whenever an offer to buy and an offer to sell match TSE (2015). The minimum unit of price change is $\delta P$. The buy-order price is rounded off to the nearest fraction, and the sell-order price is rounded up to the nearest fraction.

For each risk asset, the model includes $n$ normal agents (NAs) that trade only that risk asset, giving a total of $3n$ NAs. The model also includes one arbitrage agent (AA). The quantities of an agent’s holding positions are not limited, so the agents can take any number of shares for long and short positions to infinity.

2.1 Normal Agent (NA)

To replicate the nature of price formation in actual financial markets, I introduced the NA to model a very general investor. Its behavior is as simple as replicating long-term statistical characteristics and very short-term micro structures in real financial markets. First, at time $t = 1$, NA No. 1 places an order to buy or sell its risk asset; then, at $t = 2, 3, \ldots, n$, NAs No. 2, 3, \ldots, $n$ respectively place buy or sell orders. At $t = n + 1$, the model returns to the first NA and repeats
Figure 1  In the artificial market model in this study, one share of ETF can be exchanged for one share each of stocks 1 and 2.

Figure 2  The AA (arbitrage agent) can always place, change, or cancel orders.

Figure 3  Example of an arbitrage trade.
this cycle. An NA always places an order for only one share. For ETF, to investigate various liquidity levels, an NA places orders with a constant probability \( k(0 < k < 1) \). Thus, the trading volume of ETF is smaller than that of stock 1 or stock 2.

An NA determines an order price and buys or sells as follows. It uses a combination of a fundamental value and technical rules to form an expectation on a risk asset’s return. The expected return of agent \( j \) for each risk asset is

\[
r'_{e,j} = \frac{1}{w_{1,j} + w_{2,j} + u_j} \left( w_{1,j} \log \frac{P_f}{P^t} + w_{2,j} r'_{h,j} + w_{3,j} e'_j \right),
\]

where \( w_{i,j} \) is the weight of term \( i \) for agent \( j \) and is independently determined by random variables uniformly distributed on the interval \((0, w_{i,\text{max}})\) at the start of the simulation for each agent. \( P_f \) is a fundamental value for each risk asset and is constant\(^{5}\): \( P_f = P_{f,0} \) for stocks 1 and 2, and \( P_f = 2P_{f,0} \) for ETF. In addition, \( P^t \) is the market price of the risk asset, and \( e'_j \) is determined by random variables from a normal distribution with average 0 and variance \( \sigma_e \). Finally, \( r'_{h,j} \) is a historical price return inside an agent’s time interval \( \tau_j \), where \( r'_{h,j} = \log (P^t_i / P^t_{i-\tau_j}) \), and \( \tau_j \) is independently determined by random variables uniformly distributed on the interval \((1, \tau_{\text{max}})\) at the start of the simulation for each agent\(^{6}\).

The first term of Eq. (1) represents a fundamental strategy: the agent expects a positive return when the market price is lower than the fundamental value, and vice versa. The second term of Eq. (1) represents a technical strategy: the agent expects a positive return when the historical market return is positive, and vice versa.

After the expected return has been determined, the expected price is

\[
P_{e,j}^t = P^t \exp (r'_{e,j}).
\]

An order price \( P_{o,j}^t \) is determined by random variables normally distributed with average \( P_{e,j}^t \) and standard deviation \( P_o \), where \( P_o \) is a constant. Whether to buy or sell is determined by the magnitude relationship between \( P_{e,j}^t \) and \( P_{o,j}^t \):

- when \( P_{e,j}^t > P_{o,j}^t \) the agent places an order to buy one share, but
- when \( P_{e,j}^t < P_{o,j}^t \) the agent places an order to sell one share\(^7\).

### 2.2 Arbitrage Agent (AA)

As described previously, the value of one share of ETF is exactly the same as the sum of the values of one share each of stocks 1 and 2. Therefore, when the AA buys ETF at a lower price than the sum of the prices of stocks 1 and 2, it earns a profit consisting of the price difference, because

---

\(^5\) This enables focusing on phenomena in short time scales, as the fundamental price remains static.

\(^6\) When \( t < \tau_j \), however, \( r'_{h,j} = 0 \).

\(^7\) When \( t < t_c \), however, to generate enough waiting orders, the agent places an order to buy one share when \( P_f > P_{o,j}^t \) or to sell one share when \( P_f < P_{o,j}^t \).
it can exchange ETF for the stocks and then sell the stocks at a higher combined price than the ETF price. The practice of such trades earning the price difference between risk assets is called “arbitrage.” As shown in Figure 2, the AA can always place, change, or cancel orders.

Figure 3 shows examples of order books for each of the risk assets. At this point, the sum of the highest buy-order prices for stocks 1 and 2 is 20000 (=10000+100000). The highest buy-order price for ETF is 19800, and there is no buy order at 19900. In this case, the AA first places an order to buy one share at 19900 (shown in red) and then waits. Once the order is matched and the AA buys ETF, it exchanges the ETF share for stocks 1 and 2 and then sells them each at 10000. Thus, the AA earns a profit of 100 from the price difference. Of course, the AA can also earn a profit in the opposite case, by first selling borrowed ETF at a higher price, buying the stocks at lower prices, exchanging the stocks for ETF, and returning the ETF, again earning the price difference as a profit.

Although the above example assumes that the trading costs are zero, in real financial markets there are several types of costs, and they are never zero. In this study, therefore, I defined \( C = c \times P_{f0} \) as a summation of costs and required a profit for any one trade. Because \( C \) includes the required profit, when the price difference of risk assets is over \( C \), the AA can do an arbitrage trade.

When \( B_1, B_2, \) and \( B_{ETF} \) are the highest respective prices of stock 1, stock 2, and ETF, and \( S_1, S_2, \) and \( S_{ETF} \) are the corresponding lowest prices, the AA places an order to buy one share of ETF at \( B_1 + B_2 - C \) when \( B_{ETF} < B_1 + B_2 - C \), and/or to sell one share of ETF at \( S_1 + S_2 + C \) when \( S_{ETF} > S_1 + S_2 + C \). Note that the AA can place both buy and sell orders at the same time. After that, when the buy order for ETF is matched, the AA immediately sells one share each of stocks 1 and 2 at \( B_1 \) and \( B_2 \), respectively. These sell orders are matched immediately because of waiting buy orders, and an arbitrage trade is completed. Conversely, when the sell order for ETF is matched, the AA immediately buys one share each of stocks 1 and 2 at \( B_1 \) and \( B_2 \), respectively. On the other hand, the AA changes or cancels its orders when one of \( B_1, B_2, B_{ETF}, S_1, S_2, \) and \( S_{ETF} \) changes.

Because the AA places orders only when the price difference is more than \( C \), it never loses money; on the other hand, it is also possible that no profit chance occurs and the AA never places orders.

3 Simulation Results

In this study, I set\(^8\) the same parameters as in Mizuta et al. (2013). Specifically, I set \( n = 1000, w_{1,max} = 1, w_{2,max} = 10, w_{3,max} = 1, \tau_{max} = 10000, \sigma = 0.06, P_o = 30, t_c = 20000, \delta P = 0.01, k = 0.1, \) and \( P_{f0} = 10000 \). In short, the fundamental values for stocks 1 and 2 were both \( P_{f0} = 10000 \), and that for ETF was \( 2P_{f0} = 20000 \). I ran simulations to \( t = t_c = 10000000 \) for \( c = 0\%, 0.005\%, 0.01\%, 0.025\%, 0.05\%, 0.1\%, 0.5\%, \) and for the case without the AA. All these cases not

---

\(^8\) I explain how they verified their model in the Appendix “Verification of the Model”.

only had the other parameters fixed but also used the same random number table. I then simulated these runs 100 times, changing the random number table each time, and I report the results here as the averaged statistical values of the 100 runs.

Figure 4 shows the market price differential rate $M_d$ and the trading volume of the AA for various costs $c$. Here, $M_d$ indicates how much the price of ETF and the sum of the prices of stocks 1 and 2 differ. I defined $M_d$ as

$$M_d = \frac{1}{t_c} \sum_{t=1}^{t_c} \frac{|P_{ETF}^t - (P_1^t + P_2^t)|}{P_1^t + P_2^t},$$

(3)

where $P_{ETF}^t$, $P_1^t$, and $P_2^t$ are the prices of ETF, stock 1, and stock 2 at time $t$, respectively. $\|$ means absolute value. As the figure shows, a lower cost meant more trading volume and a lower price differential. The price differential sharply changed when the cost was near 0.1%. This value of 0.1% was similar to the standard deviation of returns (i.e., the volatility) for 10 time periods, 0.11%. Therefore, whether the cost is higher or lower than the volatility seems to indicate a very important boundary.

Figure 5 shows the relationship between cost and volatility. The red dashed line represents the sum of the highest buy-order prices of stocks 1 and 2, and the black solid line represents the highest buy-order price of ETF. Therefore, the AA can make an arbitrage trade only when the red dashed line is above the black solid line plus the cost. Because the prices of each risk asset fluctuate in their volatility, when the volatility is sufficiently greater than the cost, the AA has more chances for arbitrage trades. It then trades more, and the market price differential becomes
lower.

Figure 6 shows the market inefficiency $M_{ie}$ of ETF and stock 1 for various costs $c$. To directly measure the market efficiency, I defined $M_{ie}$ as

$$M_{ie} = \frac{1}{t_e} \sum_{t=1}^{t_e} \frac{|P^t - P_f|}{P_f}. \quad (4)$$

Here, $M_{ie}$ is always greater than zero, and $M_{ie} = 0$ means that a market is perfectly efficient\(^9\). The larger $M_{ie}$ is, the less efficient the market becomes\(^10\).

With lower cost, the ETF market became more efficient, but that of stock 1 did not. The market inefficiency of ETF also sharply changed when the cost was near 0.1%, but that of stock 1 did not. This means that the reason why the ETF market becomes more efficient is not because it gains efficiency from the stock 1 market.

Figure 7 shows the depths of waiting orders in the ETF and stock 1 markets for various costs $c$. The depth was defined as the average over all time of the number of waiting orders between

---

\(^9\) Even though I calculated the market inefficiency, I did not intend to discuss about efficient market hypothesis. In our model the market is not efficient because of existence of the technical strategy in Eq. (3) and I discussed how efficient more by AA.

\(^10\) This index is sometimes used in experimental financial studies of people, in which this market inefficiency is sometimes called relative absolute deviation (RAD) (Stöckl et al. (2010)). Many indications for measuring market efficiency have been proposed (Verheyden et al. (2013)). A feature of $M_{ie}$ is that it is calculated directly using a fundamental price $P_f$, which is never observed in empirical studies. I can also use $M_{ie}$ in simulation and experimental studies because I can exactly define $P_f$. 

---
±0.1% from the average prices of the highest buy order and the lowest sell order. Lower cost meant a thicker depth for ETF. The depth for ETF also sharply changed when the cost was near 0.1%. On the other hand, the depth for stock 1 had the opposite tendency. Figure 8 shows the trading volumes for ETF and stock 1. In this case, lower cost meant higher trading volume for both. Lower cost thus makes the depth for stock 1 thinner and the trading volume larger, because orders for arbitrage trades and waiting orders for stock 1 are matched.
In real financial markets, however, there are traders who place more orders when the trading volume increases. My model did not implement this behavior. It is possible that lower cost would increase both the depth and the trading volume because of such behavior. This remains for a future work.

Next, I investigated the case with more liquidity for ETF. In the above discussion, I fixed the ETF order ratio to $k = 0.1$. In the following, I considered various $k$ for $c = 0$ (i.e., zero cost). Figure 9 shows the depth and trading volume for stock 1 for various $k$. Larger $k$ meant thinner depth and
more trading volume. It seems that more ETF orders caused more matching of arbitrage trade orders and waiting orders for stock 1. Figure 10 shows the market price differential rate $M_d$ for various $k$. As $k$ increased, so did $M_d$. Even though more arbitrage trades occurred because of the larger $k$, the market price differential rate $M_d$ did not improve.

This result implies that, in the case when the trading volume of ETF increases to near that of a stock, improving the market price differential is more difficult by arbitrage trades like those modeled in this study. Therefore, the result implies that other ways are needed to improve the market price differential in such cases.

4 Summary and Future Works

In this study, I expanded the artificial market model of Mizuta et al. (2013) to include three risk assets, denoted as stock 1, stock 2, and ETF (Figure 1), along with an arbitrage agent (AA) that could perform arbitrage trades among these risk assets. I then investigated the relationship between the liquidity of ETF and the trading costs.

My results showed that, because the prices of each risk asset fluctuate in their volatility, when the volatility is sufficiently greater than the cost, the AA has more chances to make arbitrage trades. As the AA trades more, the market price differential becomes lower. In addition, lower cost means a thicker depth of waiting trades for ETF, whereas the depth tendency of a stock is the opposite. Furthermore, lower cost increases the trading volume of both. Lower cost makes the
depth thinner and the trading volume greater for a stock because the orders for arbitrage trades and the waiting orders for the stock are matched.

Real financial markets, however, include traders who place more orders when the trading volume increases. My model did not implement this behavior. It is possible that lower cost would increase both the depth and the trading volume with such behavior. This remains for a future work.

I also investigated the case with more liquidity for ETF and found that it makes the market price differential larger. Even though more arbitrage trades occur because of the larger ETF liquidity, the market price differential rate does not improve.

This result implies that, when the trading volume of ETF increases to near that of a stock, improving the market price differential is more difficult through arbitrage trades like those modeled in this study. This suggests that other ways are needed to improve the market price differential in such cases.

Appendix

4.1 Basic Concept for Constructing Model

An artificial market model, which is a kind of agent-based model, can be used to discuss situation that have never been realized, can handle regulation changes that have never been made, and it can isolate the pure contribution of these changes to price formation and liquidity LeBaron (2006); Chen et al. (2012); Mizuta (2016); Todd et al. (2016). These are the strong points of the artificial market simulation.
However, the outputs of this simulation would not be accurate or credible forecasts of the actual future. The simulation needs to reveal possible mechanisms that affect price formation through many simulation runs, e.g., searching for parameters or purely comparing the before/after of changes. The possible mechanisms revealed by these runs provide new intelligence and insight into the effects of the changes on price formation in actual financial markets. Other methods of study, e.g., empirical studies, would not reveal such possible mechanisms.

Indeed, artificial markets should replicate macro phenomena existing generally for any asset and any time. Price variation, which is a kind of macro phenomena, is not explicitly modeled in artificial markets. Only micro processes, agents (general investors), and price determination mechanisms (financial exchanges) are explicitly modeled. Macro phenomena emerge as the outcome interactions of micro processes. Therefore, the simulation outputs should replicate existing macro phenomena in order to generally prove that simulation models are probable in actual markets.

However, it is not a primary purpose for an artificial market to replicate specific macro phenomena only for a specific asset or a specific period. An unnecessary replication of macro phenomena leads to models that are over-fitted and too complex. Such models would prevent us from understanding and discovering mechanisms that affect price formation because the number of related factors would increase.

Indeed, artificial market models that are too complex are often criticized because they are very difficult to evaluateChen et al. (2012). A model that is too complex not only would prevent us from understanding mechanisms but also could output arbitrary results by over-fitting too many parameters. It is more difficult for simpler models to obtain arbitrary results, and these models are easier to evaluate.

Therefore, I constructed an artificial market model that is as simple as possible and do not intentionally implement agents to cover all the investors who would exist in actual financial markets.

As Weisberg mentionedWeisberg (2012), “Modeling, (is) the indirect study of real-world systems via the construction and analysis of models.” “Modeling is not always aimed at purely veridical representation. Rather, they worked hard to identify the features of these systems that were most salient to their investigations.” Therefore, under different phenomena to focus on, good models are different. Thus, my model is good only for the purpose of this study and may be not good for other purposes. An aim of my study is to understand how important properties (behaviors, algorithms) affect the investigation of macro phenomena and play a role in the financial system rather than representing actual financial markets precisely.
4.2 Verification of the Model

In many previous artificial market studies, the models were verified to see whether they could explain stylized facts, such as a fat-tail or volatility-clustering LeBaron (2006); Chen et al. (2012); Mizuta (2016); Todd et al. (2016). A fat-tail means that the kurtosis of price returns is positive. Volatility-clustering means that square returns have a positive autocorrelation, and this autocorrelation slowly decays as its lag becomes longer. Many empirical studies, e.g., that of Sewell Sewell (2011), have shown that both stylized facts (fat-tail and volatility-clustering) exist statistically in almost all financial markets. Conversely, they also have shown that only the fat-tail and volatility-clustering are stably observed for any asset and in any period because financial markets are generally unstable.

Indeed, the kurtosis of price returns and the autocorrelation of square returns are stably and significantly positive, but the magnitudes of these values are unstable and very different depending on the asset and/or period. The kurtosis of price returns and the autocorrelation of square returns were observed to have very broad magnitudes of about 1 ∼ 100 and about 0 ∼ 0.2, respectively Sewell (2011).

For the above reasons, an artificial market model should replicate these values as significantly positive and within a reasonable range as I mentioned. It is not essential for the model to replicate specific values of stylized facts because the values of these facts are unstable in actual financial markets.

Table 1 lists the statistics, standard deviation of returns, kurtosis of price returns, and autocorrelation coefficient for square returns of stock 1 when there are no arbitrage agents. This shows that this model replicated the statistical characteristics, fat-tails, and volatility-clustering observed in real financial markets.
References

http://science.sciencemag.org/content/351/6275/818.

http://dx.doi.org/10.1017/S0269888912000136.

https://doi.org/10.1088/1469-7688/2/5/303.

https://www.nature.com/articles/460685a.

http://dx.doi.org/10.1016/S1574-0021(05)02024-1.


https://doi.org/10.1109/SSCI.2016.7850016.

Torii, T., Izumi, K., Yamada, K. 2015. Shock transfer by arbitrage trading: analysis using multi-


