THREE ESSAYS ON MARKET MICROSTRUCTURE

By

Daejin Kim

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Management

May, 2014

Nashville, Tennessee

Approved:

Professor Hans R. Stoll

Professor Nicolas P.B. Bollen

Professor Jacob S. Sagi

Professor Mototsugu Shintani

Professor Robert E. Whaley

To my parents, and my wife Sookyoung

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Professor Hans R. Stoll, for his invaluable guidance, encouragement, and support throughout my doctoral studies. I would also like to thank Professor Robert E. Whaley for his advice and support in my research. Without their guidance and support, this dissertation would not have been possible. I am also grateful to my dissertation committee members, Professors Nicolas P.B. Bollen, Jacob S. Sagi, and Mototsugu Shintani for their invaluable comments and suggestions.

Financial Support from the Financial Markets Research Center at Owen Graduate School of Management and the Owen Graduate School Doctoral Fellowship is gratefully acknowledged.

I am thankful to my loving wife Sookyoung for her love and support to finish my doctoral studies. I also thank my son Wonil and my daughter Wona for giving me a lot of happiness when I was writing my dissertation. Finally, I am very thankful to my parents for their constant love and support throughout my life.

TABLE OF CONTENTS

Page

Chapter

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

The price impact under the risk-averse market maker

1 Introduction

A liquid asset market implies that traders can trade immediately with small transaction costs, and without moving the current market price. While the meaning of liquidity is clear, it is not easy to measure or observe liquidity because it is related to various trading characteristics. In the finance literature, the bid-ask spread and the price impact cost are the most common measures of liquidity. The price impact cost is derived from the relation between order flows and price changes, rather than being an explicit observable measure, while the bid-ask spread is directly observed in the financial markets. The important aspect of both measures is that they reflect the activity of the market makers who provide liquidity to markets. The price impact cost is the market makers' ex post responses to manage incoming order flows, while the bid-ask spread is the ex-ante expected costs of the market makers. Thus, these two measures should reflect the various costs, such as informational costs and inventory holding costs, that the market makers face as they provide liquidity to market participants by managing incoming orders.

The existing market microstructure literature shows that the bid-ask spread is composed of three components: order processing, adverse information, and inventory holding costs. According to the adverse information theory, the bid-ask spread exists to compensate market makers for potential losses from trading with informed traders (Copeland and Galai (1983), Glosten and Milgrom (1985), and Easley and O'hara (1987)). The inventory models suggest that market makers set the bid-ask spread because they bear the risk of holding undesired inventory that deviates from their preferred portfolio (Stoll (1978), Amihud and Mendelson (1980), Ho and Stoll (1981, 1983), Bollen et al. (2004)). Stoll (1978) shows that the bid-ask spread increases with the dealer's risk aversion and asset volatility. The inventory effect is obvious when risk-averse market makers determine the market clearing price because risk-averse market makers consider the inventory position, while risk-neutral market makers do not. As a result, the risk-aversion effects of the market makers should be considered when studying the inventory effect on liquidity.

Unlike bid-ask spread models, the existing price impact models assume that market makers are risk-neutral. Kyle (1985) develops a model in which a single risk-neutral informed trader and a number of noise traders submit orders to competitive risk-neutral market makers who set the market clearing price with zero expected profits. In equilibrium, the monopolistic informed trader strategically chooses his trade size so that it is proportional to the true asset value, and risk-neutral market makers have the linear pricing schedule in the total net order flows. Kyle's model has been extended in subsequent papers by considering different classes of noise traders (Admanti and Pfleiderer (1991)) or multiple informed traders(Foster and Viswanathan (1996), Holden and Subrahmanyam (1992), and Back et al. (2000)). There are very few papers considering risk-averse market makers. One notable exception is Subrahmanyam (1991), who studies multiple informed traders considering competitive risk-averse market makers. However, he does not provide a dynamic model incorporating risk-averse market makers.

The object of this paper is to develop a unified price impact model that incorporates competitive risk-averse market makers and asymmetric information. The price impact cost function, similarly to bid-ask spreads, should reflect both adverse information and inventory holding costs because illiquidity may be caused by both informational and non-informational events. For example, a large block trade by institutional investors may cause large price changes because a risk-averse market maker faces severe imbalances on his trading account. Thus, the risk-averse market maker will consider his inventory position as well as potential adverse information when determining the market clearing price. The inventory and informational effects can be easily observed and derived in the single-period model. However, the static model does not provide answers about how these two effects evolve over time. Thus, this paper presents the first dynamic model in which a single informed trader chooses his trade size strategically by focusing on a linear equilibrium led by Bertrand competition among risk-averse market makers who ultimately earn zero expected utility gains.

The primary results in this paper considering both inventory and asymmetric information effects are consistent with the existing theoretical and empirical literature. The price change consists of the permanent change associated with information and the temporary change related to the inventory effect stemming from the dealer's risk-averse attitude(Hasbrouck (1988), Glosten and Harris (1988)). (2) The price change is reversed because of the temporary price impact caused by the dealer's inventory costs(Grossman and Miller (1988), Campbell et al. (1993), Andrade et al. (2008)). (3) The first-order serial covariance of price changes is negative and is associated with the illiquidity measure(Roll (1984), Stoll (1989)).

First, the price impact is composed of a permanent and a transient impact. The permanent price impact is associated with the fundamental value of a risky asset, and this effect updates a current price to a new equilibrium market price. However, the inventory effect is temporary. The model in this paper predicts that risk-averse market makers want to be compensated for the potential loss due to informed traders and the inventory holding costs. In other words, risk-averse market makers are paid for providing liquidity to market participants even though there is no information asymmetry. This inventory effect is proportional to the conditional variance of a risky asset and the dealer's risk attitude.

Second, the price impact due to the dealer's risk aversion is temporary and causes a price reversal. Market participants rationally expect that order imbalances result in an increase in the dealer's inventory holding costs, and price changes, according to this temporal inventory effect, should return to the original price level. In other words, market makers would like to maintain their preferred inventory positions by taking positions opposite to their unwanted inventory positions. For example, if risk-averse market makers purchase unwanted shares, they will be willing to resell these unwanted imbalances to the market at a discount. The decision will ultimately be executed and a price reversal will be observed.

Finally, the first-order serial covariance of price changes based on the price schedule developed in this paper is shown to be negative. This covariance measure is associated with the market depth, the risky asset volatility, and the volatility of noise traders. That is, this serial covariance will be zero when market makers are risk-neutral while the measure is proportional to the dealer's risk-aversion coefficient. In addition, Roll (1984) and successive studies show that the first-order serial covariance has a negative relation with the bid-ask spread, which is the common measure for illiquidity. Stoll (1989) shows how the serial covariance depends on both inventories and information. This dependence implies that the market depth, which is the inverse of the price impact coefficient, is closely related to the bid-ask spread. In the numerical illustration, the Roll spread, measured by the square root of negative serial covariance, decreases through time. This relation implies that the private information of the informed trader is incorporated into the price; thus the asymmetric information cost decreases over time.

In sum, the dynamic model in this paper theoretically shows that the price change can be decomposed into permanent change due to the information regarding the fundamental asset value and temporal change caused by the inventory effects of risk-averse market makers. This temporary price impact caused by the dealer's risk aversion is the evidence of price reversal. Moreover, the model suggests that the first-order serial covariance measure is closely related to the price impact. This suggestion is consistent with the existing literature, which finds that there exists a significant positive relation between a bid-ask spread and the price impact measured by Kyle's lambda.(Hasbrouck (2009), Korajczyk and Sadka (2008)).

The rest of the paper is organized as follows. The next section presents a single and a dynamic auction model in which a single risk-neutral informed trader and a number of noise traders submit orders to competitive risk-averse market makers. Section 3 discusses properties of the linear equilibrium and derives the first-order serial covariance measure of price changes. Section 4 presents the empirical results to provide evidence to support the theoretical arguments. In section 5, I conclude and summarize the paper.

2 Model

2.1 Single period model

In this section, I extend Kyle (1985) model to incorporate the pricing strategy of competitive risk-averse market makers with identical risk aversion. Consider a market with a single informed trader, a number of noise traders, and competitive risk-averse market makers. Each trader submits a market order to the market maker, who sets a price, denoted by $\tilde{p}$, competitively to clear the market by absorbing all of the remaining order imbalances. When determining the price, the market makers observe the total net order imbalances, denoted by $\tilde{y}$, which are the sum of the quantity from the informed trader, denoted by $\tilde{x}$, and the quantity from noise traders, denoted by $\tilde{u}$. All of the market participants observe the last traded price, denoted by p_0. Conditional on the last trading quantities, the ex-post liquidation value, denoted by $\tilde{v}$, of the risky asset is normally distributed with mean μ_0 and variance Σ_0 [1].The

[1]Kyle (1985) assumes that the mean of the risky asset is equal to the most recent observed market price p_0. This should be true when market makers are risk-neutral. The mean of the risky asset, however, is not necessary to be equal to the trade price when market makers are risk-averse. To be account for these concerns, I relax this assumption where the mean of the risky asset is some

quantity submitted by the noise traders, $\tilde{u}$, is normally distributed with mean zero and variance σ_u^2. The quantity traded by noise traders and the ex-post liquidation value of the risky asset are independently distributed.

The informed trader knows the post-liquidation value of the risky asset and maximizes his expected profit by choosing a quantity $\tilde{x}$ where the profit function, denoted by $\tilde{\pi}$, is simply $\tilde{x}(\tilde{v} - \tilde{p})$. The optimization problem of the informed trader is

$$\max_x E[\tilde{x}(\tilde{v} - \tilde{p})|\tilde{v} = v] \tag{1}$$

Consider a risk-averse market maker who has a negative exponential utility function of the form $U(\tilde{W}) = -e^{-\gamma\tilde{W}}$, where γ is an absolute risk aversion coefficient. Assume that the market makers hold zero inventory at the initial time for convenience and can borrow or lend with a zero interest rate. Because the market makers cannot distinguish the quantity of the informed trader from that of the noise traders, they need to make a conjecture about the trading quantity of the informed trader. As in Kyle, the quantity of the informed trader is supposed to be a linear form in the asset value so that $\tilde{x} = \alpha + \beta\tilde{v}$, where α and β are some constants. Based on this structure, market makers have the prior distribution for the total net order flow $\tilde{y} = \tilde{u} + \tilde{x}$. The positive (negative) value of the total order flows implies that the market buys (sells) and the dealer sells (buys), and thus the dealer should give (receive) the $\tilde{y}$ shares of stocks and receive (pay) the corresponding cash amount $p\tilde{y}$. Thus, his terminal wealth, denoted by $\tilde{W}$, can be written

$$\tilde{W} = -\tilde{y}(\tilde{v} - p) \tag{2}$$

The market clearing price, denoted by $\tilde{p}$,is assumed to be a linear function of the

constant parameter μ_0.

total net order flow such that

$$\tilde{p} = \mu + \lambda \tilde{y} \tag{3}$$

where λ is the sensitivity of total net order flow $\tilde{y}$ and μ is some constant. The linear pricing rule is proven to be true in Kyle (1985) under competitive risk-neutral market makers. In addition, Huberman and Stanzl (2004) argue that the price-impact function should be linear when there is no price manipulation. Moreover, Subrahmanyam (1991) shows that the linear pricing rule holds under competitive risk-averse market makers. So, this linear pricing rule can be applied to the world of competitive risk-averse market makers.

Following Subrahmanyam (1991), I require that the single market maker absorbs entire order imbalances and earns zero-expected utility gain. That is, the expected utility of not market making is equal to -1. This approach is also introduced in Stoll (1978), who argues that a risk-averse market maker should be compensated for market making because his trading account deviates from his optimal portfolio to supply the immediacy. Stoll solves the equilibrium cost function where the expected utility of not making market is equal to the expected utility under making market. Assuming that a dealer holds nothing initially, the dealer's problem in this paper is identical to that of Stoll (1978). Moreover, Ho and Stoll (1983) analyze the competition of risk-averse market makers and show that prices are equal to the reservation price of the second best dealer, who eventually determines the market bid-ask spread. As a result, the Bertrand competition narrows the market spread, which eventually converges to the reservation spread of any dealer who earns zero expected profits. Therefore, the expected utility of the risk-averse market maker satisfies

$$E[U(\tilde{W})|\tilde{y} = y] = -e^{-\gamma E[\tilde{W}|y] + \frac{\gamma^2}{2} Var(\tilde{W}|y)} = -1 \tag{4}$$

where γ is an absolute risk aversion coefficient. Combining the zero-profit condition of the competitive risk-averse market makers in equation (4) with the profit maximization of the informed trader in equation (1), I obtain the following equilibrium.

Proposition 1. *For competitive risk-averse market makers and a risk-neutral informed trader, there is a unique equilibrium for the trading strategy of the informed trader and the pricing rule of the competitive risk-averse market makers such that*

$$\tilde{x} = \beta(\tilde{v} - \mu_0) \tag{5}$$

$$\tilde{p} = \mu_0 + \lambda(\tilde{x} + \tilde{u}) \tag{6}$$

where $\mu_0 = E[\tilde{v}|p_0]$ and p_0 is the recent trade price. Then, the equilibrium value of λ and β is given by

$$\lambda = \sqrt{\frac{\Sigma_0}{\sigma_u^2}}\left(\frac{\gamma\sqrt{\Sigma_0\sigma_u^2} + \sqrt{4 + \gamma^2\Sigma_0\sigma_u^2}}{4}\right) \tag{7}$$

and

$$\beta = \sqrt{\frac{\sigma_u^2}{\Sigma_0}}\left(\frac{-\gamma\sqrt{\Sigma_0\sigma_u^2} + \sqrt{4 + \gamma^2\Sigma_0\sigma_u^2}}{2}\right) \tag{8}$$

where γ is the absolute risk aversion coefficient of the risk-averse market maker.

Proof. See Appendix A1. □

From proposition 1, the linear price rule of the market makers and the linear trading strategy of the informed trader still hold even if the market makers are risk-averse. When the risk aversion coefficient γ is equal to zero, the term inside the bracket of the equilibrium λ in (7) is equal to one half, and the equilibrium λ is equal to the Kyle's lambda. The proposition also shows that the price impact costs λ should reflect both information and inventory effects. This additional term from the original Kyle's lambda is incurred due to the risk-aversion of market makers, and can be interpreted as the dealer's compensation for holding inventory or providing liquidity.

From equation (A.4) in the appendix, it is easy to show that the price impact costs should still be positive even though there is no informed trader ($\beta = 0$). This result means that the risk-averse market makers want to be compensated not only for the potential losses due to the informed trader but also for providing liquidity to clear the current order imbalance. In consequence, the magnitude of the price impact under risk-averse market makers is larger than that under risk-neutral market makers.

However, the increase in the price impact cost due to inventory costs should vanish in the long run because the inventory effect is not related to the fundamental asset value. The single-period model does not appear to provide a reasonable explanation for the evolution of information and inventory effects on the price change over time. Addressing these issues, I derive the dynamic model in the following section.

2.2 Comparison with Kyle (1985)

It is of interest to compare some features of Kyle's(1985) model to those of my model. From equations in (7) and (8), the values of λ and β appear to increase and decrease respectively when market makers are risk-averse. To compare the parameters of risk-averse market makers with those of risk neutral market makers, let define θ as

$$\theta = \frac{\gamma}{2}\sqrt{\Sigma_0 \sigma_u^2} \tag{9}$$

The value of θ is proportional to the dealer's attitude toward the risk(γ), the volatility of the risky asset(Σ_0), and the volatility of noise traders(σ_u). If market makers are risk averse ($\gamma > 0$), then the value of θ is always positive. That is, the value of θ increases when market makers are more risk-averse, the risky-asset is more volatile, or quantity from noise traders is more volatile. Thus, the defined parameter θ can be interpreted as the inventory cost of competitive market makers. With this measure θ, I discuss how the market depth, the price informativeness, and the profit of the informed trader are changed under risk-averse market makers.

First, the price impact coefficient λ increases when market makers are risk averse. The price impact of risk-averse markets markers(λ^A) can be expressed as $\lambda^A = \lambda^N(\sqrt{1+\theta^2}+\theta)$ where λ^N is the price impact of risk-neutral market makers. Kyle defines the inverse of this price impact coefficient λ as the market depth which is the necessary order flow to change prices by one dollar. The market depth under risk-averse market makers is simply $1/\lambda^A = 1/\lambda^N(\sqrt{1+\theta^2}-\theta)$. Because the term in the bracket $((\sqrt{1+\theta^2}-\theta))$ ranges in $(0,1]$, the market depth will be close to zero when market makers are extremely risk-averse.

Second, the price informativeness is measured as the conditional asset variance. Kyle argues that half of the insider's price information is reflected into asset prices and the volatility of noise trading doesn't affect the volatility of prices. The ratio of the conditional variance, $\Sigma_1^2/\Sigma_0 = 1/2(1+\theta/\sqrt{1+\theta^2})$, is greater than one-half when market makers are risk-averse. This result implies that less than one-half of the informed trader's price information is incorporated into prices. When market makers are risk-averse, the informed trader appears to be reluctant to reveal his information because his marginal costs to exploit the private information with risk-averse market makers. As a result, the informed trader needs more auctions to reveal his private information through the market.

Finally, the informed trader's profit decreases when market makers are risk averse. The informed trader's profit can be expressed as $E[\pi^A] = E[\pi^N](\sqrt{1+\theta^2}-\theta)$. This expression implies that the informed trader's expected profit is affected by the magnitude of the dealers' risk aversion. The informed trader can still earn the positive profit at the expense of noise traders. This expected profit, however, converges to zero when market makers are extremely risk-averse. The decrease in the informed trader's profit is mainly because the informed trader pays some fees to market makers who want to be compensated from the arrival of the informed trader.

2.3 Dynamic model

In this section, I consider the sequential auction framework to derive the equilibrium condition and the price movement when market makers are competitive and risk-averse. Suppose there are N rounds of trade occurring in a given trading day, and t_n denotes the time at which the nth auction takes place. For simplicity, assume that the time interval between the nth and the $n-1$th auction is equally distributed, i.e., $\Delta t_1 = \Delta t_2 = \cdots = \Delta t_N = \Delta t$. At the beginning stage, given the last trade price p_0, the mean and variance of the risky asset are μ_0 and Σ_0, respectively.

At each auction, there are noise traders and a single informed trader. The quantity traded by the noise traders, denoted by $\Delta\tilde{u}_n$, is normally distributed with mean zero and variance $\sigma_u^2 \Delta t$. Moreover, $\Delta\tilde{u}_n$ is serially uncorrelated and independent of $\tilde{v}$. The informed trader chooses $\Delta\tilde{x}_n$ at the nth auction to maximize his entire profits given his guesses about the pricing rule set by a market maker. The informed trader's profits for each auction n is given by

$$\tilde{\pi}_n = \sum_{k=n}^{N} (\tilde{v} - p_k)\Delta\tilde{x}_k \tag{10}$$

For risk-averse market makers, let $\tilde{W}_n$ denote the aggregate wealth of the market makers after the nth auction, so that $\Delta\tilde{W}_n$ denotes the wealth increment of the market makers at the nth auction.

$$\Delta\tilde{W}_n = -\Delta\tilde{y}_n(\tilde{v} - \tilde{p}_n) \tag{11}$$

For simplicity, the initial wealth and its increase at time 1 is 0, i.e., $\Delta W_0 = W_0 = 0$. This assumption means that each market maker initially has no inventory in his account so that his initial expected utility is equal to -1 under the CARA utility function. As in the single-period model, I also impose the restrictions that the single risk-averse market maker absorbs the entire order imbalance by competition and earns zero expected profit. At each time, the market makers observe the total net order flow $\Delta\tilde{y}_n = \Delta\tilde{x}_n + \Delta\tilde{u}_n$, and they cannot discriminate the order flow from the informed trader from the quantity from noise traders. Similar to the single-period model, the zero profit condition is imposed so that the market maker earns zero-expected utility gain. As a result, the expected utility of the wealth change at each auction n should be

$$E_n\left[-e^{-\gamma\Delta\tilde{W}_n}|\Delta\tilde{y}_n = \Delta y_n\right] = -1 \qquad \forall n = 1, \cdots, N \tag{12}$$

where γ is an absolute risk aversion coefficient. This equation implies that the expected utility obtained by not making the market should be equal to that obtained by making the market. Combining the problem of the informed trader and the condition of expected utility from the competitive risk-averse market makers leads to the following proposition.

Proposition 2. *For a risk-neutral informed trader and a risk-averse market maker, there is a unique equilibrium in the model such that for*

$$\begin{aligned}
\Delta\tilde{x}_n &= \beta_n(\tilde{v} - \mu_{n-1})\Delta t & (13)\\
\tilde{p}_n &= \mu_{n-1} + \lambda_n(\Delta\tilde{x}_n + \Delta\tilde{u}_n) & (14)\\
\Sigma_n &= Var(\tilde{v}|\Delta\tilde{x}_1 + \Delta\tilde{u}_1, \cdots, \Delta\tilde{x}_n + \Delta\tilde{u}_n) & (15)\\
\mu_n &= E(\tilde{v}|\Delta\tilde{x}_1 + \Delta\tilde{u}_1, \cdots, \Delta\tilde{x}_n + \Delta\tilde{u}_n) & (16)\\
E[\tilde{\pi}_n|p_1, \cdots, p_{n-1}, v] &= \alpha_{n-1}(v - \mu_{n-1})^2 + \delta_{n-1} & (17)
\end{aligned}$$

for all $n = 1, \cdots, N$. The constants $\alpha_n, \beta_n, \delta_n$, and Σ_n are the unique solution to the *difference equation system*

$$\alpha_{n-1} = \frac{1+2\gamma\alpha_n\Sigma_n}{4\left(\lambda_n - \alpha_n\left(\lambda_n - \frac{\gamma\Sigma_n}{2}\right)^2\right)} \tag{18}$$

$$\delta_{n-1} = \delta_n + \alpha_n\left(\lambda_n - \frac{\gamma}{2}\Sigma_n\right)^2 \sigma_u^2 \Delta t \tag{19}$$

$$\beta_n \Delta t = \frac{1-2\alpha_n\left(\lambda_n - \frac{\gamma\Sigma_n}{2}\right)}{2\left(\lambda_n - \alpha_n\left(\lambda_n - \frac{\gamma\Sigma_n}{2}\right)^2\right)} \tag{20}$$

$$\lambda_n = \frac{\gamma\Sigma_n}{2} + \frac{\beta_n\Sigma_n}{\sigma_u^2} \tag{21}$$

$$\Sigma_n = \left(1 - \Delta t \beta_n\left(\lambda_n - \frac{\gamma\Sigma_n}{2}\right)\right)\Sigma_{n-1} \tag{22}$$

for all $n = 1, \cdots, N$ *subject to the boundary condition* $\alpha_N = \delta_N = 0$ *and the second order condition is.*

$$\lambda_n - \alpha_n\left(\lambda_n - \frac{\gamma}{2}\Sigma_n\right)^2 > 0 \tag{23}$$

Proof. See Appendix A2. □

The value of β_n represents the intensity with which the informed trader trades based on his private information. Note that even if there is no informed trader ($\beta_n = 0$), the equilibrium liquidity parameter λ_n is still positive in (21) because the market maker still needs compensation for the risk associated with the noise trades. Thus, the price impact coefficient in my model reflects not only the informational cost but also the inventory cost.

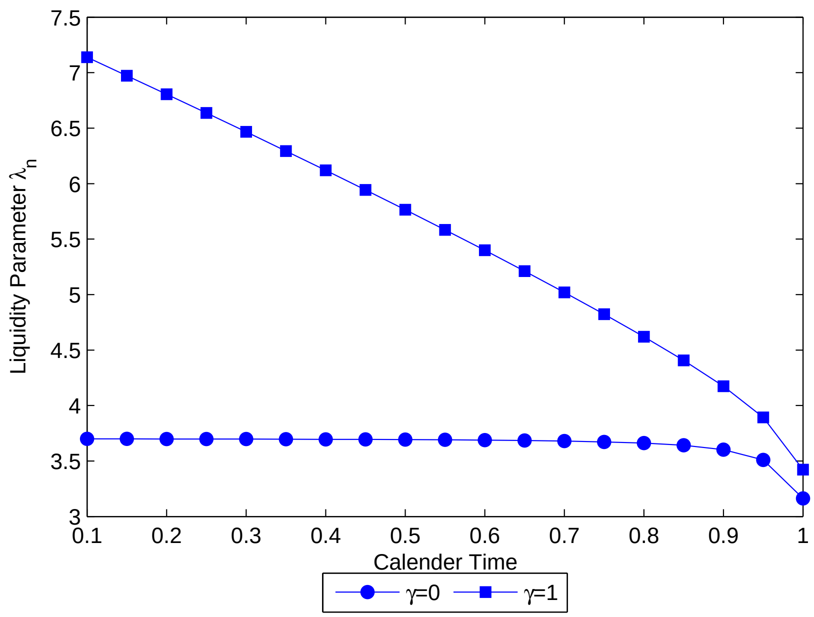
3 Properties of equilibrium in dynamic model

3.1 Numerical illustration

In this section, I present numerical examples based on the linear equilibrium derived in the previous section. In all of the numerical examples, assume that $\Sigma_N = 1$, $\sigma_u^2 = 1$, and $\Delta t = 1/N$. Figure 1 plots λ_n for the particular cases of $\gamma = 0$ and $\gamma = 1$ when the number of auctions is fixed at $N = 20$. As seen from Figure 1, the λ_n of the

Figure 1: Liquidity parameter over time for different values of risk aversion coefficient

This figure plots the liquidity parameter at each auction for different values of risk aversion coefficient $\gamma = 0, 1$ when the number of auction is fixed at $N = 20$. This figure compares the liquidity parameter, λ_n, when a market maker is risk-neutral ($\gamma = 0$) and is risk-averse($\gamma = 1$). The variance of noise trading per unit time, σ_u^2, and the variance of the risky asset at the end of time, Σ_N, are equal to 1. Each auction occurs at equally spaced interval, $\Delta t = 1/N$, over $[0, 1]$.

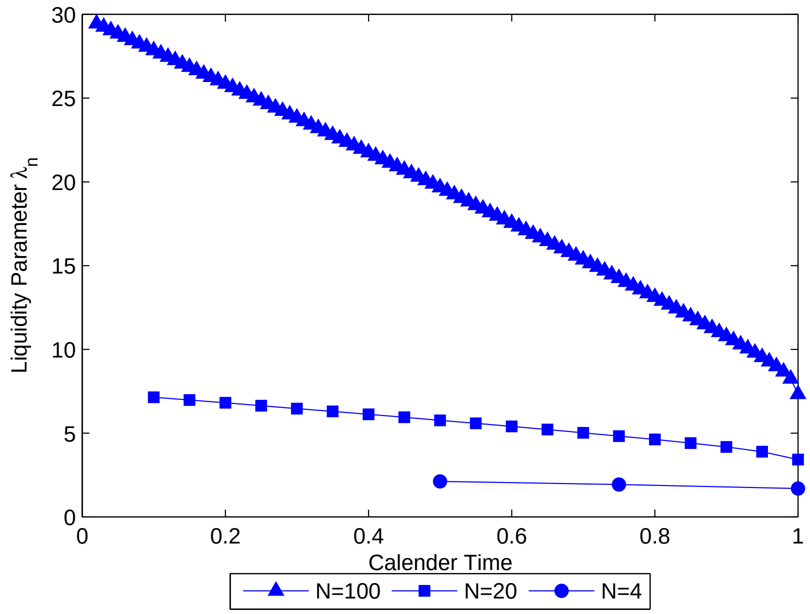


risk-averse market makers is larger than that of the risk-neutral market makers. This result implies that a risk-averse market maker bears the cost of managing his inventory position and sets the price impact cost function to be moved up. In addition, the difference between the price impact function of the risk-averse market makers and the risk-neutral market makers is larger at the beginning of the auction than at the end of the auction.

Figure 2 plots λ_n for the particular cases of $N = 4, N = 20$, and $N = 100$ when the risk aversion coefficient is at $\gamma = 1$. This figure is quite similar to the risk-neutral market maker cases presented in Kyle's original paper. The illiquidity measure λ_n monotonically decreases through time. The figure shows that the price

Figure 2: Liquidity parameter over time for different values of N

This figure plots the liquidity parameter at each auction for different values of number of auctions $N = 4, 20$, and 100 when the risk aversion coefficient is fixed at $\gamma = 1$. The variance of noise trading per unit time, σ_u^2, and the variance of the risky asset at the end of time, Σ_N, are equal to 1. Each auction occurs at equally spaced interval, $\Delta t = 1/N$, over $[0, 1]$.



impact function declines over time regardless of the magnitude of the market maker's risk-aversion.

3.2 Price behavior over time

Given the equilibrium in the previous section, it is easy to derive the closed form of the price movement. The constant term of price movement in (14) can be replaced with the known term using the conditional mean of the multivariate normal distribution. The closed form of price movement is shown in the following Lemma.

Lemma 1. *For a risk-neutral informed trader and a risk-averse market maker, the* *pricing rule set by competitive risk-averse market maker is*

$$\Delta p_n = \lambda_n \Delta y_n - \frac{\gamma}{2}\Sigma_{n-1}\Delta y_{n-1} \tag{24}$$

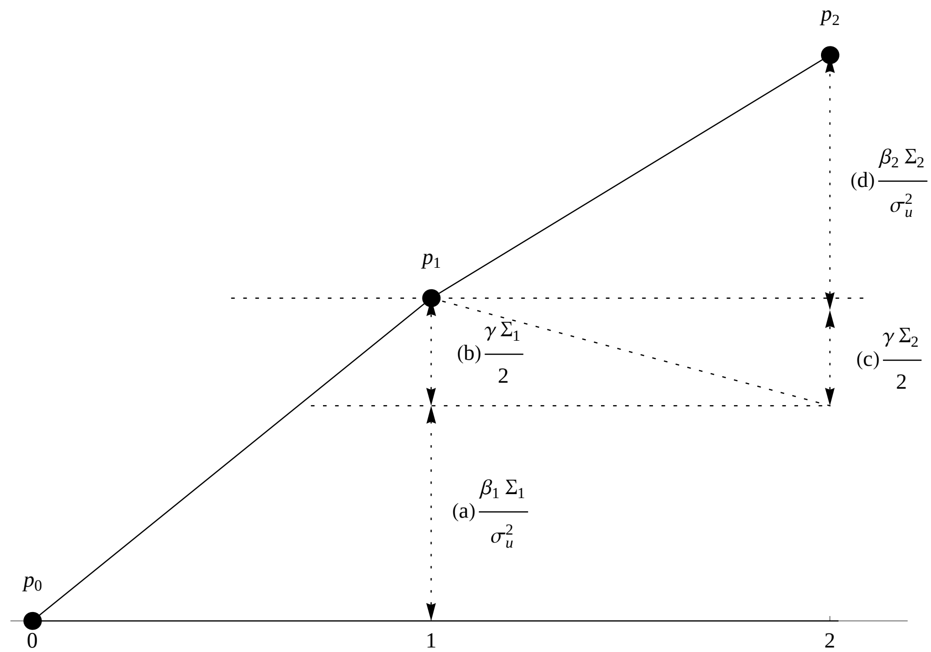
for all $n = 1, \cdots, N$.

Proof. See Appendix A3. □

This price formation in (24) is consistent with the previous empirical studies(Glosten and Harris (1988), Madhavan et al. (1997),Huang and Stoll (1997), Sadka (2006)). There are some implications from lemma 1. First, the price change is linear both in the contemporary net order flow and in the lagged one. This result suggests that the risk-averse market maker sets the market clearing price from the current order flow to cover the losses from the informed trader and to adjust his trading account to his optimal portfolio. Second, the coefficient of the contemporary order flow reflects both informational and inventory effects while the coefficient of the lagged order imbalance only reflects the inventory information. This result suggests that the price impact occurring in the current day will reverse in the following day due to the inventory cost effect. These results are consistent with the two-period model developed in Chordia and Subrahmanyam (2004). It appears that the coefficient of the lagged order flow can provide a more accurate measure of illiquidity because it contains the risk attitude of the market makers and the asset risk, but not the informational content(Campbell et al. (1993), Pastor and Stambaugh (2003)). The intensity of the informed trader, measured by the β_n, should ultimately be impounded in the price. But the non-informational part should disappear from the price if it is priced in the previous period. If the market makers are highly risk-averse due to managing high risk assets, then the immediate price change will be larger than the degree of the price change, reflecting the information content, and the magnitude of the price reversal could be large. Thus, the product of the asset risk and the market maker's degree of

Figure 3: Price dynamics from time 0 through time 2

This figure illustrates the price dynamics when a market maker is risk-averse. The order imbalance at time $n = 0, 1$, and 2 is assumed to be $\Delta y_0 = 0$, $\Delta y_1 = 1$ and $\Delta y_2 = 1$.

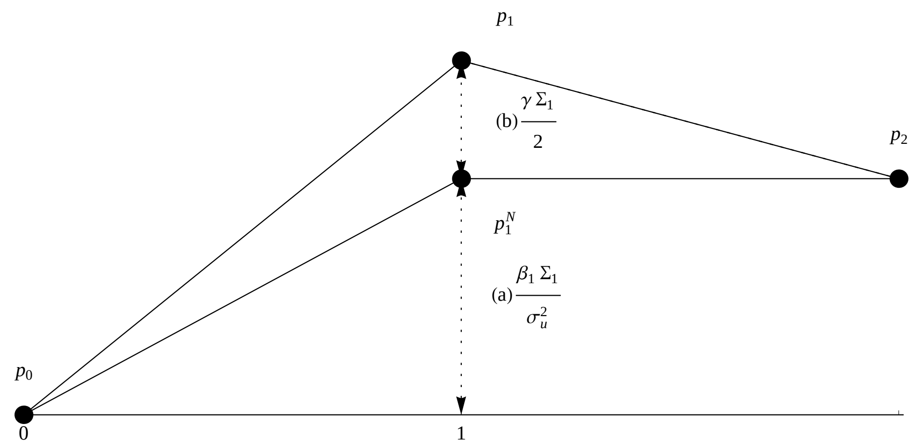


risk-averseness could be the appropriate illiquidity measure.

Figure 3 illustrates the price movement up to time 2 when the competitive market makers are risk-averse. For convenience, suppose that there was no order imbalance at time 0 ($\Delta y_0 = 0$). The net order imbalances at times 1 and 2 are assumed to be one unit($\Delta y_1 = \Delta y_2 = 1$). From the price rule in equation (24), the price at time 1 is equal to $p_0 + \lambda_1$ and the price impact coefficient has two parts: (a) the informational price impact, $\frac{\beta_1}{\sigma_u^2}\Sigma_1$, and (b)the non-informational price impact, $\frac{\gamma}{2}\Sigma_1$. These two impacts imply that given orders, the risk-averse market maker sets the price to reflect both informational and non-informational components. However, the non-informational impact, (b) in the illustration, should vanish because it is unrelated to the fundamental value of the risky asset. Thus, the price at time 2 first reverses to a price level that reflects (a) the informational component and contemporaneous price impact, which is the sum of (c) the informational price impact, $\frac{\beta_2}{\sigma_u^2}\Sigma_2$, and (d)

Figure 4: Comparisons risk-averse market maker price setting with risk-neutral market maker price setting from time 0 through time 2

This figure compares price settings between risk-averse and risk-neutral market maker. The order imbalance at time $n = 0, 1$, and 2 is assumed to be $\Delta y_0 = 0$, $\Delta y_1 = 1$ and $\Delta y_2 = 0$. The price set by risk-averse market maker is p_1 while the price set by risk-neutral market maker is p_1^N at time $n = 2$



the non-informational price impact, $\frac{\gamma}{2}\Sigma_2$ at time 2. Thus, the price change from time 0 to time 2 is $\frac{\beta_1}{\sigma_u^2}\Sigma_1 + \frac{\beta_2}{\sigma_u^2}\Sigma_2 + \frac{\gamma}{2}\Sigma_2$.

These examples suggest that the informational component at each time is impounded into the price. The inventory effect appears to be transient, but it may be long-lived unless the order imbalances are equally distributed over time. That is, the price movement up to time n can be expressed as

$$p_n - p_0 = \sum_{i=1}^{n}\left(\frac{\beta_i\Sigma_i}{\sigma_u^2}\Delta y_i\right) + \frac{1}{2}\gamma\left(\Sigma_n\Delta y_n - \Sigma_0\Delta y_0\right) \tag{25}$$

The first term represents the sum of the informational impact over time, and the second term indicates the inventory impact. This inventory effect shown in my model is the primary difference from Kyle's model. Equation (25) suggests that the price at time n reflects all of the prior information up to time n, while only the current inventory cost is impounded into the price. In other words, all of the historic information should be related to the current price permanently, but the previous inventory costs are not related to the current asset price.

Figure 4 illustrates the difference in price setting between a risk neutral market maker and a risk-averse market maker. Assume the initial conditional mean of the risky asset is equal to the last trade price($\mu_0 = p_0$), and the order imbalances at time 0 and at time 2 are zero($\Delta y_0 = \Delta y_2 = 0$). At time 1, there is one unit of order imbalance($\Delta y_1 = 1$). The price set by the risk neutral market maker, denoted by p_1^N, reflects only the informational component ((a)). However, the price set by the risk-averse market maker, denoted by p_1, reflects both the informational effect and the inventory effect, so that the price of the risk-averse market maker is higher than that of the risk neutral market maker. At time 2, the risk neutral market maker will set the price at time 2 equal to the price at time 1 because there are no incoming orders at time 2. But, the risk-averse market maker will set the price at time 2 equal to a price that reflects only the informational component at time 1. If the risk-averse market maker manages his inventory position toward zero, then the final price at the end of day will converge to the price that is equal to the price of risk neutral market maker.

3.3 Serial covariance and the Roll(1984) spread

Roll (1984) shows that the effective spread can be estimated from the first order serial covariance of the transaction price changes. Based on the pricing rule, it is easy to calculate the serial covariance of price differences at each auction n.

Lemma 2. *At each auction n, the serial covariance of price change is*

$$Cov(\Delta p_{n+1}, \Delta p_n) = -\frac{1}{2}\gamma \Delta t \sigma_u^2 \lambda_n \Sigma_{n-1} \tag{26}$$

Proof. See Appendix A4. □

The first order serial covariance measure of price change is simply the product of the coefficient of the contemporary order flow, the coefficient of the lagged order flow, and the variance of noise traders. Thus, the first-order serial covariance of price change is affected by both information and inventory effects. These results are consistent with the recent study by Vayanos and Wang (2012), who argue that the negative value of the first-order serial covariance measure of price change is related to the price reversal, and is higher when the risk-aversion coefficient increases. However, my result indicates that the covariance measure represents not only the price impact (λ_n) but also the price reversal ($\gamma/2\Sigma_{n-1}$). This finding implies that the covariance measure could provide a noise measure of price reversal The coefficient of the lagged order flow is more appropriate to measure price reversal.

Based on the covariance measure calculation, the Roll's spread measure can be easily obtained. From equation (26), the Roll's spread is simply

$$S_n^{Roll} = \sqrt{-Cov(\Delta p_{n+1}, \Delta p_n)} = \sigma_u \sqrt{\Sigma_{n-1}} \sqrt{\frac{\gamma}{2}\lambda_n \Delta t} \tag{27}$$

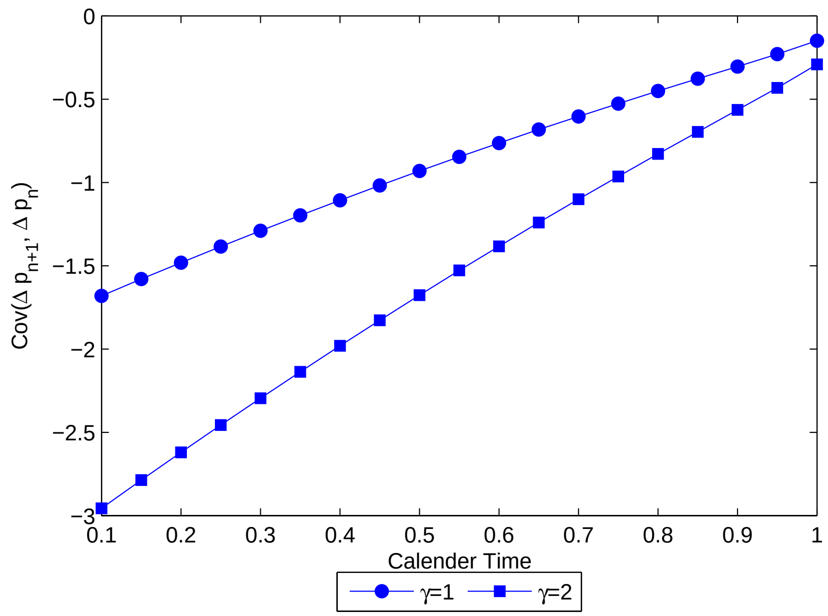
The Roll's spread will be zero if the market makers are risk-neutral($\gamma = 0$). The Roll's spread is widely used for estimating the bid-ask spread when the bid-ask spread cannot be observed in the market. Based on the calculation, the serial covariance of the price difference can be thought as the expected costs of the risk-averse market maker when he makes the market to provide liquidity. So, the potential price impact costs will be considered in the risk-averse market maker's cost function and will show that the Roll's spread is proportional to the price impact parameter λ_n.

If there is no informed trader at time n ($\beta_n = 0$), then the equilibrium λ_n is equal to $\frac{\gamma}{2}\Sigma_{n-1}$ from (B.10).). In this case, the market maker will consider only his inventory when he makes the market. The Roll's spread is simply

$$S_n^{No-inform} = \frac{1}{2}\gamma\sigma_u\Sigma_{n-1}\sqrt{\Delta t} \tag{28}$$

Figure 5: The serial covariance of price difference over time for different values of risk aversion coefficient

This figure plots the first serial covariance of price change, $Cov(\Delta p_n, \Delta p_{n-1}) = -1/2\gamma\Delta t\sigma_u^2\lambda_n\Sigma_{n-1}$, at each auction for different values of risk aversion coefficient $\gamma = 1, 2$ when the number of auction is fixed at $N = 20$. The variance of noise trading per unit time, σ_u^2, and the variance of the risky asset at the end of time, Σ_N, are equal to 1. Each auction occurs at equally spaced interval, $\Delta t = 1/N$, over $[0, 1]$.

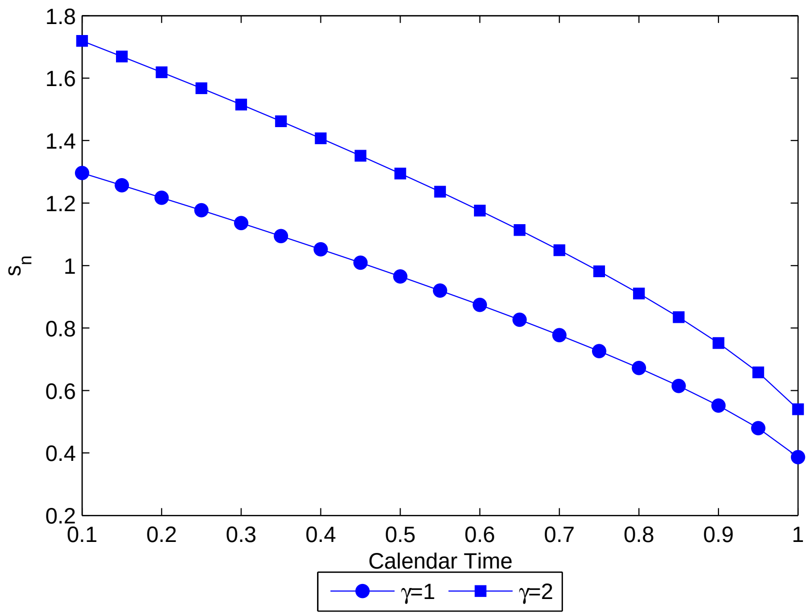


This equation is quite similar to the bid-ask spread derived in Stoll (1978). The equation implies that the bid-ask spread still exists even though there is no information asymmetry among traders because the market maker will want to be compensated for deviating from his preferred portfolio.

Figure 5 and 6 plot the first order serial covariance of the price difference and the Roll's spread, respectively, for different degrees of the risk aversion coefficient, $\gamma = 1$ and $\gamma = 2$ when the number of auctions is fixed at $N = 20$. As in the previous section, all parameters are estimated by assuming $\Sigma_N = \sigma_u^2 = 1$. Figure 5 shows that the covariance measure increases over time, and moves toward zero. The covariance moves faster toward zero when the market maker is more risk-averse. Figure 6illustrates the

Figure 6: The Roll's spread over time for different values of risk aversion coefficient

This figure plots the Roll's spread, $S_n^{Roll} = \sigma_u \sqrt{\Sigma_{n-1}\gamma/2\lambda_n \Delta t}$, at each auction for different values of risk aversion coefficient $\gamma = 1, 2$ when the number of auction is fixed at $N = 20$. The variance of noise trading per unit time, σ_u^2, and the variance of the risky asset at the end of time, Σ_N, are equal to 1. Each auction occurs at equally spaced interval, $\Delta t = 1/N$, over $[0, 1]$.



movement of Roll's spread over time. Roll's spread decreases over time because both the asset variance Σ_n and the price impact coefficient λ_n decrease over time.

4 Empirical Tests

4.1 Data and methodologies

The data used in this paper are obtained from the New York Stock Exchange(NYSE) Trade and Quotation(TAQ) and Center for Research in Security Prices(CRSP) data set [2]. I use only the decimal tick size period for data consistency. The sample period spans from Jan 1, 2002 to Dec 31, 2010 and price and quote data in TAQ must occur during regular trading hours (9:30 a.m.-4:00 p.m.). I first select stocks

[2]Each firm is identified by CRSP permno not by ticker because ticker is used duplicate. I match CRSP permno with TAQ ticker symbol using cusip number.

whose CRSP share codes are either 10 or 11 so that securities such as ADRs, REITS, certificates, shares of beneficial interest, units, closed-end funds, and preferred stocks are excluded in my sample. The trading venues of each stock should be unchanged during the sample period. Each stock should not have stock splits, stock dividends, repurchases, or secondary offerings during the sample period. The monthly average prices must be between \$3 and \$1000. I also require a minimum number of 100 trades in a particular stock over the sample period. The final sample consists of 2,645 US common stocks.

Following Lee and Ready (1991), each trade is classified as buyer-initiated or seller-initiated using the tick rule. That is, each trade is matched with the first quote at least five seconds prior to the transaction. As in Chordia and Subrahmanyam (2004) and Chordia et al. (2002), I construct daily time series for six different measures of order imbalance for each stock:

- Δy_t^{Num} : the daily number of buyer-initiated trades less the daily number of seller-initiated trades on day t
- $\Delta y_t^{\%Num}$: Δy_t^{Num} divided by total number of trades on day t
- Δy_t^{Vol}: the daily buyer-initiated dollars less the daily seller-initiated dollars on day t
- $\Delta y_t^{\%Vol}$: Δy_t^{Vol} divided by total dollar volume on day t
- Δy_t^{Sh} : the daily number of buyer-initiated shares purchased minus the daily number of seller-initiated shares sold on day t
- $\Delta y_t^{\%Sh}$: Δy_t^{Sh} divided by total share volume on day t

For each security i, the following multiple regression model is used to estimate the price impact and the price reversal coefficients:

$$\Delta p_{i,t} = \alpha_i + \lambda_{i,1}\Delta y_{i,t} + \lambda_{i,2}\Delta y_{i,t-1} + \epsilon_t \tag{29}$$

where $\Delta p_{i,t}$ is open-to-close log price change and $\Delta y_{i,t}$ is daily order imbalance for firm i at time t [3]. From Lemma 1, the price impact coefficient λ_1 is expected to be positive, and the price reversal coefficient λ_2 is expected to be negative. For the comparison purposes, the simple time-series regressions are also estimated. The comparison regression models are

$$\Delta p_{i,t} = \alpha_i + \lambda^S_{i,1}\Delta y_{i,t} + \xi_t \tag{30a}$$

$$\Delta p_{i,t} = \alpha'_i + \lambda^S_{i,2}\Delta y_{i,t-1} + \xi'_t \tag{30b}$$

If both order imbalances should be included but one of them is missed in the estimation of the regression model, the residual ξ_t (ξ'_t) can be correlated with $\Delta y_{i,t}$ ($\Delta y_{i,t-1}$) [4]. As a result, the specifications in (30) could violate the fundamental assumption of OLS estimates and lead to the biased estimates [5].

4.2 Summary statistics

Table 1 provides the summary statistics for the cross-sectional averages of the daily time-series means of the bid-ask spreads, the order imbalances, and the trading activities. In Panel A, the average effective spread is \$0.058 while the average quoted spread is \$0.068, reflecting the within-quote trading. Both proportional spreads, RQS and RES, have similar statistics with mean 0.4% and median 0.2%. As expected, the

[3] I also perform the time series regression using close-to-close price change. The results were quite similar to results reported in this paper.

[4] If order imbalances are not autocorrelated, equations (30) have the unbiased estimates. However, the data shows that daily order imbalances are exhibited to be positively autocorrelated.

[5] Previous empirical papers didn't consider both order flows when estimating illiquidity measures (See Pastor and Stambaugh (2003), Breen et al. (2002), Hasbrouck (2009)). Chordia and Subrahmanyam (2004) regress excess returns on order imbalances including lagged order imbalances, find that up to 5 days order imbalances are significant.

Table 1: Descriptive Statistics

This table contains the cross-sectional averages of the time-series means of the variables for 2645 US stocks from January 2002 through December 2010. All the variables are computed from TAQ data. In panel A, each daily spread is calculated by averaging the intraday observations. The quoted half spread, QS, is half of the difference between the ask price and the bid price of the quote. The relative half-spread, RQS, is the absolute value of the trade price and midpoint of quoted ask and bid divided by the midpoint. The effective spread, ES, is the absolute value of the difference between the trade price and the quote midpoint just prior to the trade. The relative effective spread, RES, is calculated by dividing the effective spread by the trade price. In Panel B, for each firm daily order imbalance is the buyer-initiated order minus the seller-initiated order. The unscaled order imbalance $\Delta y = \Delta y^{Num}, \Delta y^{Vol}$ and Δy^{Sh} is the order imbalance in number of transaction, in dollar volume, and in number of shares traded. The scaled order imbalance $\Delta y = \Delta y^{\%Num}, \Delta y^{\%Vol}$ and $\Delta y^{\%Sh}$ is the order imbalance in number of transaction divided by total number of trades, in total dollar volume divided by total dollar volume, and in number of shares traded divided by total share volume on a given day. In panel C, Sh is the daily number of transaction, Vol is the daily dollar volume, and Sh is the daily number of shares traded.

Variable	Mean	Std. Dev.	Median	Skewness	Kurtosis
Panel A. Spreads					
QS	0.068	0.112	0.030	5.649	53.897
RQS	0.004	0.006	0.002	3.058	13.343
ES	0.058	0.090	0.026	5.178	42.595
RES	0.004	0.005	0.002	3.118	14.164
$Cov(\Delta p_t, \Delta p_{t-1})$	−0.013	0.104	0.000	−15.489	289.102
Panel B. Order Imbalance					
Δy^{Num} (Thousands)	−0.034	0.191	−0.003	−10.618	161.288
$\Delta y^{\%Num}(\%)$	−0.018	0.081	−0.010	−1.626	5.781
Δy^{Vol} (Millions)	0.138	1.956	−0.018	−0.893	79.111
$\Delta y^{\%Vol}(\%)$	−0.025	0.086	−0.017	−1.351	4.571
Δy^{Sh} (Thousands)	0.068	75.675	−1.208	−5.401	125.484
$\Delta y^{\%Sh}(\%)$	−0.026	0.086	−0.018	−1.337	4.465
Panel C. Trading Activity					
Number of trades (Thousands)	2.429	6.222	0.696	7.685	87.444
Number of shares (Thousands)	739.865	2598.977	179.698	13.267	241.273
Dollar volume (Millions)	22.841	89.147	3.355	12.917	248.504

average of serial covariance has a negative mean of -0.013 and is negatively skewed.

Panel B of Table 1 presents the cross-sectional averages of the daily time-series means of the order imbalances. The daily mean value of the order imbalance in terms of the number of transactions is -34 transactions per day during the sample period, indicating that more selling transactions occurred during the sample period. On the contrary, other unscaled average values for the order imbalance in total dollar volume or in the number of shares have positive means and medians. These results suggest that small or medium investors with small quantities submit their orders more frequently than institutional investors with large quantities (Chan and Fong (2000)[6]). Finally, all other scaled order imbalances have negative means but positive medians, indicating that days with large selling pressure are more frequently observed than days with large buying pressure.

Panel C reports the cross-sectional averages of the daily trading activities for each firm. The mean value of the total number of trades is 2,429 transactions, the mean value of the number of shares traded is 0.7 million shares, and the mean value of the total dollar volume is 22 million dollars per day. Different values for the mean and median suggest that trading activities are not symmetric; rather, they demonstrate a positive skew.

Table 2 presents the cross-sectional averages of the time-series correlation between order imbalance measures, the price difference, and trading activities. Panel A shows the cross-sectional averages of the daily time-series correlations among the scaled and unscaled order imbalances, the price differences, the total number of trades, the dollar trading volume and the trading volume. First, the correlation between the dollar trading volume and the trading volume is very high for both the scaled and the unscaled measures (0.959 and 1.00). In fact, the correlation between the scaled measures is close to one. Second, the correlation measures between one order imbalance and another order imbalance are generally positive and very high, ranging from 0.378 to 1. Finally, the correlations between the price difference and the order imbalance measures are positive for all different measures of order imbalance. The correlation between the price difference and the order imbalance in terms of the number of transactions is higher than between the price difference and other trading

[6]Chan and Fong (2000) classify the trades into five categories based on trading volume, and report that the average daily number of trades decreases from small size category to the large size category.

Table 2: Correlations and Autocorrelations

This table presents the cross-sectional averages of the correlations and the autocorrelation up to lag 10 for each stock. The unscaled order imbalance $\Delta y = \Delta y^{Num}, \Delta y^{Vol}$ and Δy^{Sh} denotes the order imbalance in number of transaction, in dollar volume, and in number of shares traded. The scaled order imbalance $\Delta y = \Delta y^{\%Num}, \Delta y^{\%Vol}$ and $\Delta y^{\%Sh}$ represents the order imbalance in number of transaction divided by total number of trades, in total dollar volume divided by total dollar volume, and in number of shares traded divided by total share volume on a given day. Sh is the daily number of transaction, Vol is the daily dollar volume, and Sh is the daily number of shares traded.

Panel A. Correlations

	Δy^{Num}	$\Delta y^{\%Num}$	Δy^{Vol}	$\Delta y^{\%Vol}$	Δy^{Sh}	$\Delta y^{\%Sh}$	Num	Vol	Sh
Δp	0.315	0.284	0.188	0.226	0.185	0.226	0.015	0.014	0.010
Δy^{Num}		0.659	0.579	0.517	0.583	0.516	−0.096	−0.075	−0.081
$\Delta y^{\%Num}$			0.378	0.792	0.397	0.792	−0.052	−0.023	−0.024
Δy^{Vol}				0.544	0.959	0.544	−0.122	−0.154	−0.158
$\Delta y^{\%Vol}$					0.576	1.000	−0.051	−0.031	−0.034
Δy^{Sh}						0.576	−0.130	−0.154	−0.168
$\Delta y^{\%Sh}$							−0.051	−0.030	−0.034
Num								0.755	0.782
Vol									0.938

Panel B. Daily Autocorrelation

Lag	Δy^{Num}	$\Delta y^{\%Num}$	Δy^{Vol}	$\Delta y^{\%Vol}$	Δy^{Sh}	$\Delta y^{\%Sh}$
1	0.222	0.232	0.162	0.200	0.159	0.200
2	0.139	0.165	0.106	0.142	0.104	0.142
3	0.109	0.140	0.086	0.118	0.085	0.118
4	0.094	0.120	0.078	0.103	0.077	0.103
5	0.082	0.111	0.069	0.096	0.068	0.096
6	0.072	0.101	0.062	0.087	0.061	0.087
7	0.065	0.093	0.058	0.084	0.058	0.084
8	0.063	0.093	0.056	0.081	0.055	0.081
9	0.059	0.087	0.054	0.077	0.054	0.077
10	0.056	0.085	0.052	0.076	0.052	0.076

volume based imbalance measures.

In Panel B of Table 2, the cross-sectional averages of the time series autocorrelations for different measures of order imbalances are reported. All of the order imbalances are positively autocorrelated. For example, the first-lag autocorrelation ranges from 16% to 23%. The autocorrelations for the order imbalance in terms of the number of transactions is slightly higher than for the order imbalance in terms of the number of shares or in terms of the dollar trading volume. The autocorrelations for the scaled order imbalances are higher than for the unscaled ones. The existence of autocorrelations suggests that both order imbalances should be included when estimating either the price impact or the price reversal measure. Otherwise, the exclusion of one of those variables cause biased regression estimates as a result of a missing variable problem.

4.3 Empirical results

Table 3 reports the summary for the regression estimation results of equations (29) and (30). In each line, reported are cross-sectional average coefficients of the regression estimates for each stock and the percentages of the significance at the 5% level for each coefficient. Column 1 indicates the use of order imbalances for independent variables when estimating the regression lines. Columns 2 and 3 are the averages of price impact and price reversal coefficients estimated from the equation (29). Columns 4 and 5 report results of (30a) and (30b) respectively. The last two columns present the test results whether estimated coefficients from equation (29) are different from the estimated one from equation (30a) or (30b). Odd lines report the results of from using unscaled order imbalances, and even lines report the results from using scaled versions.

Throughout the lines, column 2 shows that the average coefficients for the current order imbalance are positive and significant for most firms. More than 95 of estimated price impact coefficients are positive and significant. Furthermore, column 3 shows that the averages of price reversal coefficients for most firms are negative with approximately 50% of coefficients being negative and significant. Although there are some cases with positive price reversal coefficients, the significance level of these positive coefficients are negligible (approximately less than 1%). The empirical results support the theoretical predictions such that the price coefficient is positive and the price reversal coefficient is negative.

For the comparison, the average coefficients λ_1^S without the lagged order imbalance are also reported in column 4. The magnitude of the coefficient without lagged

Table 3: Daily regressions of open-to-close price differences on order imbalances

This table presents the cross-sectional averages of coefficients for individual daily time series regressions. For each stock, I estimate the time series regression $\Delta p_{i,t} = \alpha_i + \lambda_{i,1}\Delta y_{i,t} + \lambda_{i,2}\Delta y_{i,t-1}$ where $\Delta p_{i,t}$ is the open-to-close log price difference of stock i on day t, and $\Delta y = \Delta y^{Num}, \Delta y^{Vol}, \Delta y^{Sh}, \Delta y^{\%Num}, \Delta y^{\%Vol}$, and $\Delta y^{\%Sh}$. $\Delta y^{Num}_{i,t}, \Delta y^{Vol}_{i,t}$ and $\Delta y^{Sh}_{i,t}$ is the unscaled order imbalance in number of transaction, in dollar volume, and in number of shares traded for stock i on day t. $\Delta y = \Delta y^{\%Num}_{i,t}, \Delta y^{\%Vol}_{i,t}$ and $\Delta y^{\%Sh}_{i,t}$ is the scaled order imbalance in number of transaction divided by total number of trades, in total dollar volume divided by total dollar volume, and in number of shares traded divided by total share volume for stock i on day t. λ^S_i denotes the coefficient when the regression model is estimated with missing one of independent variables. λ^S_1 is estimated with excluding lagged variable and λ^S_2 is estimated with excluding contemporary variable. The significance of coefficients in each stock's time series is tested at 5% level.

Δy	Multiple		Single		Difference	
	λ_1 (%+, %+Sig)	λ_2 (%-, %-Sig)	λ^S_1 (%+, %+Sig)	λ^S_2 (%-, %-Sig)	$\lambda_1 - \lambda^S_1$ (t-stat)	$\lambda_2 - \lambda^S_2$ (t-stat)
Δy^{Num}	0.2316	−0.0287	0.2257	0.0069	0.0059	−0.0356
	(99.09, 95.73)	(89.41, 59.47)	(98.98, 95.58)	(55.80, 9.60)	(13.45)	(−21.75)
$\Delta y^{\%Num}$	0.0426	−0.0109	0.0394	−0.0006	0.0032	−0.0103
	(99.74, 96.90)	(86.96, 54.48)	(99.70, 96.86)	(52.36, 7.49)	(32.94)	(−53.81)
Δy^{Vol}	0.0194	−0.0021	0.0189	0.0001	0.0005	−0.0022
	(97.13, 83.14)	(77.96, 29.11)	(97.20, 82.76)	(55.84, 5.82)	(3.61)	(−6.24)
$\Delta y^{\%Vol}$	0.0277	−0.0067	0.0258	−0.0007	0.0019	−0.0060
	(99.66, 95.58)	(82.61, 41.55)	(99.62, 95.58)	(53.65, 6.28)	(26.52)	(−45.42)
Δy^{Sh}	0.2849	−0.0203	0.2806	0.0102	0.0043	−0.0305
	(96.90, 85.52)	(78.03, 32.78)	(96.98, 85.41)	(55.50, 7.37)	(5.37)	(−12.24)
$\Delta y^{\%Sh}$	0.0277	−0.0067	0.0258	−0.0007	0.0019	−0.0060
	(99.66, 95.58)	(82.57, 41.55)	(99.62, 95.54)	(53.61, 6.35)	(26.52)	(−45.42)

imbalance decreases, implying the existence of a correlation between the current order imbalance and the error term. The estimated price impact coefficients are shown to be underestimated when excluding the lagged order imbalances. Column 5 reports the estimated reversal coefficient when excluding current order imbalances. The estimation results are disappointed. The percentages of negative values are approximately less than 55% and those of negative and significant values are less than 10%. Comparing with the results in column 3, the percentages of negative values in column 5 are significantly lower than those in column 3. Moreover, some average values in column 5 are positive suggesting that excluding the contemporaneous order flows could cause the serious biased estimates when estimating the price reversal measures.

The last two columns present the test results whether there exist any differences between coefficients from multiple regression and from simple regression. The column 6 reports the test results of comparing λ_1 from (29) with λ_1^S from (30a). The test results show that λ_1^S without lagged variable is significantly different from λ_1 from full regression model. The column 7 also provides the test results of λ_2 from (29) versus λ_2^S from (30b). Similarly, the test results show the estimated coefficients from the simple regression are overestimated and are significantly different from the estimated price reversal coefficients from the full regression. These results confirm that both order imbalances variables should be included in the model specification when researchers need to correctly estimate the price impact and the price reversal coefficients.

Note that the results are much stronger when scaled order imbalances are used than when unscaled order imbalances are used. For example, the percentage of significant price impact coefficients in terms of the number of shares increases from 95.73% for an unscaled order imbalance to 96.90% for a scaled order imbalance as shown in lines 1 an 2. Among the different measures for the order imbalance, the order imbalance in terms of the number of transactions has much stronger results than the order imbalance in terms of the dollar trading volume or the order imbalance in terms of the number of shares traded.

Table 4 summarizes the cross-sectional averages of the correlation matrix between the estimated parameters and other observed illiquidity measures like the serial covariance of price changes and bid-ask spreads. First, the price impact coefficients λ_1 estimated from the unscaled order imbalances are positively correlated with the bid-ask spread measures and are negatively correlated with the first-order serial covariance of price difference. However, the correlations between the price impact when the scaled imbalances are used and the bid-ask spreads are not highly correlated and are negative in most cases. Second, the price reversal coefficients λ_2 estimated from the unscaled order imbalances are negatively correlated with the bid-ask spread. This result suggests that the absolute value of the price reversal coefficient, which is the

Table 4: Correlation of open-to-close price difference regression

This table shows the cross-sectional averages of correlations among liquidity variables. λ_1 and λ_2 are estimated from the daily time series regression of open-to-close log price difference on both contemporary and lagged order imbalances. $Cov(\Delta p_{t+1}, \Delta p_t)$, QS, and ES are averages of daily measures from TAQ data. QS is the half quoted spread and ES is the half effective spread. $Cov(\Delta p_{t+1}, \Delta p_t)$ is the daily serial covariance of price difference.

Variables	QS	RQS	ES	RES	$Cov(\Delta p_t, \Delta p_{t-1})$
Panel A. Order imbalance in number of shares					
λ_1	0.470	0.533	0.470	0.531	−0.016
λ_2	0.068	−0.081	0.094	−0.077	0.063
$\lambda_1\lambda_1$	−0.037	−0.205	−0.031	−0.213	0.079
Panel B. Scaled order imbalance in number of shares					
λ_1	−0.330	−0.383	−0.329	−0.377	−0.051
λ_2	0.302	0.360	0.303	0.355	0.169
$\lambda_1\lambda_1$	0.186	0.227	0.182	0.222	0.124
Panel C. Order imbalance in number of dollars					
λ_1	0.250	0.521	0.256	0.530	−0.047
λ_2	−0.054	−0.259	−0.055	−0.268	0.078
$\lambda_1\lambda_1$	−0.081	−0.271	−0.087	−0.286	0.072
Panel D. Scaled order imbalance in number of dollars					
λ_1	−0.330	−0.383	−0.329	−0.377	−0.051
λ_2	0.302	0.360	0.303	0.355	0.170
$\lambda_1\lambda_1$	0.186	0.227	0.182	0.222	0.124
Panel E. Order imbalance in number of transactions					
λ_1	0.471	0.666	0.495	0.686	0.013
λ_2	−0.130	−0.328	−0.139	−0.345	0.104
$\lambda_1\lambda_1$	−0.096	−0.242	−0.104	−0.252	0.077
Panel F. Scaled order imbalance in number of transactions					
λ_1	−0.374	−0.444	−0.371	−0.433	−0.035
λ_2	0.365	0.439	0.367	0.434	0.192
$\lambda_1\lambda_1$	0.241	0.298	0.236	0.291	0.136

inventory effect, is positively correlated with the bid-ask spread. However, the correlation between the reversal coefficient in the scaled version and the first-order serial covariance of price difference is negative although it should be expected to be positive.

The product of these two coefficients represents the covariance measure as in (26) and is shown to be positively correlated with the serial covariance measure in both the scaled and the unscaled versions. Although the estimated coefficients have different signs when using the scaled or the unscaled order imbalance, the product of the two coefficients is a meaningful parameter to estimate the liquidity. Overall, the correlation table confirms that the theoretical predictions are consistent with the actual market data.

5 Conclusion

This paper has presented a theory regarding the price impact costs when competitive market makers are risk-averse in markets with asymmetric information. I extend the Kyle's model to incorporate the inventory effect incurred by the dealer's risk aversion.

First, the single-period model shows that both information and inventory effects are reflected in the price impact function. The inclusion of these effects implies that the non-informational price impact cost, compensation to risk-averse market makers for providing liquidity, can exist although there are no informed traders in the market. Second, the dynamic model shows that the information effect is permanent while the inventory effect is temporary. The current market price reflects all of the available information as well as temporal non-informational events due to the arrivals of large orders. The inventory effect, however, is temporary and disappears in the subsequent trade. The predictions of the dynamic model are more consistent with previous bid-ask spread decomposition models. Based on the derived price in the dynamic model, the first order serial covariance measure is proportional to both the price impact and the price reversal.

In empirical tests, the estimated coefficients show that the average price impact coefficients are positive and the price reversal coefficients are negative. Most coefficients are quite significant, and the signs of the coefficients are consistent with those predicted in the model. The estimated coefficients are compared with those when one of order imbalances are excluded from regression specification. Both price impact and reversal coefficients from simple regressions tend to be biased and be underestimated. This result suggests that both order imbalances should be included when estimating the price impact and reversal coefficients correctly. Moreover, the estimated coefficients are highly correlated with the existing illiquidity measures such as the bid-ask spread or the first order serial covariance of the price difference. This finding confirms that my theoretical results are well supported by the empirical findings.

This paper presents a unified model that incorporates both inventory and information effects. Although the existing microstructure studies separately incorporate either the inventory or the information effect, there are very few papers that theoretically combine these two effects. The important contribution of this paper is that both the inventory and the information effects are shown in one unified model. Moreover, the derived price behavior in this paper reflects empirical regularities.

Appendix

A Proof of Proposition 1

From (1), the optimal quantity traded by the informed trader is $x = \frac{1}{2\lambda}(v - \mu)$ which should be equal to the market maker's conjecture $\tilde{x} = \alpha + \beta\tilde{v}$. Thus, we have

$$\alpha = -\mu/2\lambda \qquad \text{and} \qquad \beta = 1/2\lambda \tag{A.1}$$

Substituting for the linear pricing rule in (3) into (4), the zero profit condition of competitive risk-averse market makers implies

$$-yE[\tilde{v}|y] + \mu y + \lambda y^2 - \frac{\gamma}{2} y^2 Var(\tilde{v}|y) = 0 \tag{A.2}$$

Then, the above equation can be written

$$\mu + \lambda y \ = \ E[\tilde{v}|y] + \frac{\gamma}{2} y Var(\tilde{v}|y) \tag{A.3}$$

From the properties of conditional normal distribution, we have get the following equations

$$\mu = \frac{\sigma_u^2 \mu_0 - \alpha\beta\Sigma_0}{\sigma_u^2 + \beta^2\Sigma_0} \qquad \text{and} \qquad \lambda = \frac{\left(\beta + \frac{\gamma}{2}\sigma_u^2\right)\Sigma_0}{\sigma_u^2 + \beta^2\Sigma_0} \tag{A.4}$$

Solving (A.1) and (A.4), we have $\mu = \mu_0$ and

$$\lambda = \sqrt{\frac{\Sigma_0}{\sigma_u^2}} \left(\frac{\gamma\sqrt{\Sigma_0\sigma_u^2} + \sqrt{4 + \gamma^2\Sigma_0\sigma_u^2}}{4} \right) \tag{A.5}$$

because the second order condition in the problem of the informed trader in (1) implies $\lambda > 0$ such that only the positive value of λ is valid.

B Proof of Proposition 2

At each auction n, given the price p_{n-1} the mean and variance of the risky asset are μ_{n-1} and Σ_{n-1}, that is $\tilde{v}|p_{n-1} \sim (\mu_{n-1}, \Sigma_{n-1})$. As in Kyle (1985), the optimization problem of the informed trader can be written

$$E[\tilde{\pi}_{n+1}|p_1, \cdots, p_n, v] = \alpha_n(v - \mu_n)^2 + \delta_n \tag{B.1}$$

for some constants α_n and δ_n. Since $\tilde{\pi}_n = (\tilde{v} - \tilde{p}_n)\Delta\tilde{x}_n + \tilde{\pi}_{n+1}$, we have

$$E[\tilde{\pi}_n|p_1, \cdots, p_{n-1}, v] = \max_{\Delta x_n} E[(\tilde{v} - \tilde{p}_n)\Delta\tilde{x}_n + \alpha_n(v - \mu_n)^2 + \delta_n|p_1, \cdots, p_{n-1}, v] \tag{B.2}$$

The market pricing rule is given by

$$\tilde{p}_n = \mu_{n-1} + \lambda_n(\Delta\tilde{x}_n + \Delta\tilde{u}_n) \tag{B.3}$$

Plugging (B.3) and the formula of μ_n into (B.2), the conditional expectation yields

$$\max_{\Delta x_n}\left\{(v-\mu_{n-1}-\lambda_n\Delta x_n)\Delta x_n+\alpha_n\left(v-\mu_{n-1}-\frac{\beta_n\Sigma_n}{\sigma_u^2}\Delta x_n\right)^2+\alpha_n\left(\frac{\beta_n\Sigma_n}{\sigma_u^2}\right)^2\sigma_u^2+\delta_n\right\} \tag{B.4}$$

Solving the optimization problem, we have

$$\Delta x_n=\frac{1-2\alpha_n\left(\frac{\beta_n\Sigma_n}{\sigma_u^2}\right)}{2\left(\lambda_n-\alpha_n\left(\frac{\beta_n\Sigma_n}{\sigma_u^2}\right)^2\right)}(v-\mu_{n-1})=\beta_n(v-\mu_{n-1})\Delta t \tag{B.5}$$

And the second order condition is

$$\lambda_n-\alpha_n\left(\frac{\beta_n\Sigma_n}{\sigma_u^2}\right)^2>0 \tag{B.6}$$

Plugging (B.5) into (B.4), we have

$$\frac{1+4\alpha_n\left(\lambda_n-\frac{\beta_n\Sigma_n}{\sigma_u^2}\right)}{4\left(\lambda_n-\alpha_n\left(\frac{\beta_n\Sigma_n}{\sigma_u^2}\right)^2\right)}(v-\mu_{n-1})^2+\alpha_n\left(\frac{\beta_n\Sigma_n}{\sigma_u^2}\right)^2\sigma_u^2\Delta t+\delta_n \tag{B.7}$$

Now for the risk-averse market maker, the conditional expected utility in a mean-variance fashion with zero profit condition is

$$-\Delta y_n E[\tilde{v}|\Delta y_n]+\mu_{n-1}\Delta y_n+\lambda_n\Delta y_n^2-\frac{\gamma}{2}\Delta y_n^2 Var(\tilde{v}|\Delta y_n)=0 \tag{B.8}$$

The above equation implies that

$$\mu_{n-1}+\lambda_n\Delta y_n=E[\tilde{v}|\Delta y_n]+\frac{\gamma}{2}\Delta y_n\Sigma_n \tag{B.9}$$

where $\Sigma_n=Var(\tilde{v}|\Delta y_n)$. Using the properties of the conditional normal distribution, the solutions are

$$\lambda_n = \frac{\beta_n\Sigma_{n-1}+\frac{\gamma}{2}\sigma_u^2\Sigma_{n-1}}{\sigma_u^2+\Delta t\beta_n^2\Sigma_{n-1}} \tag{B.10}$$

$$\Sigma_n = \frac{\sigma_u^2\Sigma_{n-1}}{\sigma_u^2+\Delta t\beta_n^2\Sigma_{n-1}} \tag{B.11}$$

These are equivalent to (21) and (22). Plugging (21) and (22) into (B.5) and (B.7), we can have (18) , (19), and (20). Finally, combine (20) with (21) to obtain

$$\left(1-\left(\lambda_n-\frac{\gamma\Sigma_n}{2}\right)^2\frac{\sigma_u^2}{\Sigma_n}\Delta t\right)\left(1-\alpha_n\left(\lambda_n-\frac{\gamma\Sigma_n}{2}\right)\right)=\frac{1}{2}+\frac{\gamma}{2}\Delta t\sigma_u^2\left(\lambda_n-\frac{\gamma\Sigma_n}{2}\right) \tag{B.12}$$

This is a cubic equation in λ_n and has three real roots. The middle root of (B.12) satisfies the second order condition.

C Proof of Lemma 1

Assume that the risky asset has the conditional normal distribution with mean μ_{n-1} and variance Σ_{n-1}, that is $\tilde{v}|p_{n-1}\sim N(\mu_{n-1},\Sigma_{n-1})$. From the definition of (13), we know that

$$v|p_n\sim\mathcal{N}\left(\mu_{n-1}+\frac{\beta_n\left(p_n-\mu_{n-1}\right)\Sigma_{n-1}}{\lambda_n\left(\sigma_u^2+\Delta\mathrm{t}\beta_n^2\Sigma_{n-1}\right)},\frac{\sigma_u^2\Sigma_{n-1}}{\sigma_u^2+\Delta\mathrm{t}\beta_n^2\Sigma_{n-1}}\right) \tag{C.1}$$

Thus, $E[\tilde{v}|p_n]=\mu_n$ can be written as

$$\mu_n=\mu_{n-1}+\frac{\left(p_n-\mu_{n-1}\right)\beta_n\Sigma_n}{\lambda_n\sigma_u^2} \tag{C.2}$$

Moreover, since $\mu_{n-1}=p_n-\lambda_n\Delta y_n$, we have

$$\mu_n=p_n-\left(\lambda_n-\frac{\beta_n\Sigma_n}{\sigma_u^2}\right)\Delta y_n \tag{C.3}$$

Combining (21) with (C.3), we get the (24).

D Proof of Lemma 2

Given a recent traded price p_{n-1} at the end of $n-1$ and at the beginning of n, the risk asset is normally distributed with mean μ_{n-1} and variance Σ_{n-1}. To get the serial covariance measure, we need to rearrange the price difference Δp_n and Δp_{n+1}. From the (1) and (13), the price difference at time n and $n+1$ can be written

$$\Delta p_n=\left(\Delta u_n+\Delta t\left(v-\mu_{n-1}\right)\beta_n\right)\lambda_n-\frac{1}{2}\gamma\Sigma_{n-1}\Delta y_{n-1} \tag{D.1}$$

and

$$\Delta p_{n+1}=\left(\Delta u_{n+1}+\Delta t\left(v-\mu_n\right)\beta_{n+1}\right)\lambda_{n+1}-\frac{1}{2}\gamma\left(\Delta u_n+\Delta t\left(v-\mu_{n-1}\right)\beta_n\right)\Sigma_n \tag{D.2}$$

where μ_n is

$$p_{n-1}+\left(\Delta u_n+\Delta t\left(v-\mu_{n-1}\right)\beta_n\right)\left(\lambda_n-\frac{1}{2}\gamma\Sigma_n\right)-\frac{1}{2}\gamma\Sigma_{n-1}\Delta y_{n-1} \tag{D.3}$$

Then, the covariance measure $Cov[\Delta p_{n+1}, \Delta p_n]$ is

$$-\frac{1}{2}\Delta t\lambda_n\left(\left(\sigma_u^2+\Delta t\beta_n^2\Sigma_{n-1}\right)\left(\gamma\Sigma_n+\Delta t\beta_{n+1}\lambda_{n+1}\left(2\lambda_n-\gamma\Sigma_n\right)\right)-2\Delta t\beta_n\beta_{n+1}\lambda_{n+1}\Sigma_{n-1}\right) \tag{D.4}$$

From (21) and (B.11)

$$2\lambda_n-\gamma\Sigma_n = \frac{2\beta_n\Sigma_n}{\sigma_u^2} \tag{D.5}$$

$$\sigma_u^2+\Delta t\beta_n^2\Sigma_{n-1} = \frac{\sigma_u^2\Sigma_{n-1}}{\Sigma_n} \tag{D.6}$$

Plugging (D.5) and (D.6) into (D.4) and rearranging, we have

$$Cov(\Delta p_{n+1}, \Delta p_n) = -\frac{1}{2}\gamma\Delta t\sigma_u^2\lambda_n\Sigma_{n-1} \tag{D.7}$$

which is the serial covariance measure at time n.

REFERENCES

Admanti, A. R., Pfleiderer, P., 1991. Sunshine trading and financial market equilibrium. Review of Financial Studies 4, 443.

Amihud, Y., Mendelson, H., 1980. Dealership market: Market-making with inventory. Journal of Financial Economics 8, 31–53.

Andrade, S. C., Chang, C., Seasholes, M. S., 2008. Trading imbalances, predictable reversals, and cross-stock price pressure. Journal of Financial Economics 88, 406–423.

Back, K., Cao, C. H., Willard, G. A., 2000. Imperfect competition among informed traders. Journal of Finance 55, 2117–2155.

Bollen, N. P. B., Smith, T., Whaley, R. E., 2004. Modeling the bid/ask spread: measuring the inventory-holding premium. Journal of Financial Economics 72, 97–141.

Breen, W. J., Hodrick, L. S., Korajczyk, R. A., 2002. Predicting equity liquidity. Management Science, 470–483.

Campbell, J. Y., Grossman, S. J., Wang, J., 1993. Trading volume and serial correlation in stock returns. The Quarterly Journal of Economics 108 (4), 905–939.

Chan, K., Fong, W. M., 2000. Trade size, order imbalance, and the volatility-volume relation. Journal of Financial Economics 57, 247–273.

Chordia, T., Roll, R., Subrahmanyam, A., 2002. Order imbalance, liquidity, and market returns. Journal of Financial Economics 65, 111–130.

Chordia, T., Subrahmanyam, A., 2004. Order imbalance and individual stock returns: Theory and evidence. Journal of Financial Economics 72, 485–518.

Copeland, T., Galai, D., 1983. Information effects on the bid-ask spread. Journal of Finance 38, 1457–1469.

Easley, D., O'hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69–90.

Foster, F. D., Viswanathan, S., 1996. Strategic trading when agents forecast the forecasts of others. The Journal of Finance 51 (4), 1437–1478.

Glosten, L., Milgrom, P., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.

Glosten, L. R., Harris, L., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123–142.

Grossman, S. J., Miller, M. H., 1988. Liquidity and market structure. The Journal of

Finance 43 (3), 617–633.

Hasbrouck, J., 1988. Trades, quotes, inventories, and information. Journal of Financial Economics 22, 229–252.

Hasbrouck, J., 2009. Trading costs and returns for u.s. equities: Estimating effective costs from daily data. Journal of Finance 64, 1445–1477.

Ho, T., Stoll, H., 1983. The dynamics of dealer markets under competition. Journal of Finance 38, 1053–1074.

Ho, T., Stoll, H. R., 1981. Optimal dealer pricing under transactions and return uncertainty. Journal of Financial Economics 9, 47–73.

Holden, C. W., Subrahmanyam, A., 1992. Long-lived private information and imperfect competition. The Journal of Finance 47 (1), 247–270.

Huang, R. D., Stoll, H. R., 1997. The components of the bid-ask spread: a general approach. Review of Financial Studies 10, 995–1034.

Huberman, G., Stanzl, W., 2004. Price manipulation and quasi-arbitrage. Econometrica 72, 1247–1275.

Korajczyk, R. A., Sadka, R., 2008. Pricing the commonality across alternative measures of liquidity. Journal of Financial Economics 87, 45–72.

Kyle, A. S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–754.

Madhavan, A., Richardson, M., Roomans, M., 1997. Why do security prices change? a transaction-level analysis of nyse stocks. Review of Financial Studies 10, 1035–1064.

Pastor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Sadka, R., 2006. Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. Journal of Financial Economics 80 (2), 309–349.

Stoll, H., 1978. The supply of dealer services in securities markets. Journal of Finance 33, 1133–1151.

Stoll, H. R., 1989. Inferring the components of the bid-ask spread: Theory and empirical tests. Journal of Finance 44, 115–134.

Subrahmanyam, A., 1991. Risk aversion, market liquidity, and price efficiency. Review of Financial Studies 4 (3), 417.

Vayanos, D., Wang, J., 2012. Liquidity and asset returns under asymmetric information and imperfect competition. Review of Financial Studies 25 (5), 1339–1365.

CHAPTER II

The behavior of merger arbitrage investors: the role of market making and price discovery

1 Introduction

Merger arbitrage or risk arbitrage is a specialized investment strategy in which arbitrageurs purchase the target company's stock and profit from the arbitrage spread, the difference between the purchase price and the offer price. The merger arbitrage spread can be caused by uncertainty regarding the merger, the investors' limited capital, or transaction costs. The relation between the merger arbitrage spread and liquidity will be examined by investigating the trading behavior around the merger announcement. In other words, the investigation will determine whether those two spread variables affect each other through the market microstructure model. Moreover, it is important to understand how the incorporation of new information occurs during a merger period because new information associated with the target stock will result in a change in the value of the merged firm, which will be reflected in the current bidder stock price, and the change in the bidder's stock price may affect the target stock price that the target shareholders receive when the deal completes successfully. Thus, the time series behavior of both bidder and target stocks and the price discovery after the merger announcement must be discovered by considering the prices of the two stocks simultaneously.

The existing studies of the microstructure effect on mergers study the behavior of the bid-ask spread of target stocks around a merger announcement by focusing on changes in the bid-ask spread and its components. (Conrad and Niden (1992), Jennings (1994), Foster and Viswanathan (1995)) [1]. However, the bid-ask spread

[1] Foster and Viswanathan (1995) find that spreads increase before announcement, and decrease after announcement using daily date. On an intraday analysis, Jennings (1994) finds that the

of target stocks may be affected by the activities of merger arbitrage; thus, only focusing on the time series for the target stock may result in a biased estimate. For example, when investors purchase a target stock and sell a bidder stock at the same time, they may postpone buying or adjust the bidding price because they may be exposed to greater deal failure risk if it is difficult to short-sell the target stock. This postponement or adjustment can potentially affect the liquidity of the target stock. Moreover, if a bidder stock lacks liquidity, the holders of the target shares face the risk of holding an exchanged bidder stock because they may have trouble selling the exchanged stock after the deal completes. Thus, a structural model is developed to estimate the components of a bid-ask spread using multiple time series to account for the interactions between the bidder and the target stock.

In this paper, I study how investors trading merger stocks are exposed to risks and are compensated, using the Huang and Stoll (1994) spread decomposition model. Thus, each price change can be expressed as the function of the bid-ask spread components, and the arbitrage spread can also be expressed as the sum of the informational component and the inventory component of the bid-ask spread. This analogy suggests that the arbitrage opportunity can be associated with liquidity risk(Kumar and Seppi (1994), Roll et al. (2007)). Roll et al. (2007) argue that deviations from the no-arbitrage relations should be related to liquidity because liquidity facilitates arbitrage. They investigate the relation between stock market liquidity and the index futures basis and show that arbitrageurs affect liquidity with their trading and that liquidity plays an important role in moving markets toward an efficient outcome. Similarly, if the positive excess merger arbitrage returns are thought to be compensation to merger arbitrageurs for providing immediacy to the target shareholders, merger arbitrage returns and liquidity should be related.

quoted bid-ask spread of target company shares is abnormally high immediately after the takeover announcements but falls quickly thereafter. In a recent study, Lipson and Mortal (2007) show that spreads drop and that quoted depth increases for bidder firms, and these changes can be explained by firm characteristics.

The existing merger arbitrage literature shows that the merger arbitrage strategy appears to generate excess risk-adjusted returns to arbitrageurs[2]. This positive excess return appears to be valid even after considering the market impact costs and broker fees. In addition to general transaction costs, the existence of these positive excess returns may be interpreted as compensation for various risks borne by the merger arbitrage investors trading target stock. The most compelling reason for this interpretation is that arbitrageurs could be compensated for bearing the deal failure risks. Larcker and Lys (1987) assume that risk arbitrageurs are better informed than the market about the likelihood of takeover success and argue that the excess return could be compensation to arbitrageurs for acquiring costly information related to the deal outcome. The second reason is that investors may be compensated due to limits to arbitrage caused by limited capital. Shleifer and Vishny (1997) argue that arbitrageurs may not have enough resources to handle the selling pressure caused by the existing target shareholders, and the consequence of this limited capital keeps the target stock price below the offer price. Moreover, investors who invest in a merger arbitrage strategy may be less likely to be diversified because merger arbitrage is a very specialized strategy that only trades one or two stocks, and so they care about both systemic and idiosyncratic volatility. As a result, the investors specializing in merger arbitrage will require some compensation for holding an undiversified portfolio with limited capital.

Thus, the compensations requested by merger arbitrage investors are similar to those required by market makers who provide liquidity. A recent survey by Moore

[2]For example, using a sample of 94 SEC 13D filings from 1977 to 1983, Larcker and Lys (1987) show that arbitrageurs generate excess returns of 5.3% on their portfolio positions from the transaction date to the resolution date. Jindra and Walkling (2004) examine 362 cash tender offers of publicly traded US targets during the period from 1971 to 1985 and report annualized returns to arbitrageurs of 46.5%. Using the calendar-time portfolio construction, Baker and Savasoglu (2002) report an average annualized abnormal return of 7.2% to 10.8% for a sample of 1901 cash and stock deals during the periods from 1981 to 1996 after controlling for the capital asset pricing model and Fama-French three-factor.

et al. (2006) reports some stylized facts about the behaviors of merger arbitrageurs[3]. From the survey, merger arbitrageurs have their own private information related to the announced deals, are limited in their capital, and act as if they are market makers by providing liquidity during the merger period. So, the compensation for the deal failure risk can be thought of as an asymmetric information cost, and holding the undiversified portfolio with limited capital can be regarded as an inventory cost. Therefore, the compensation requested by merger arbitrage investors can be inferred from the analysis of the bid-ask spread components.

The estimation results from the state space models with a Kalman filter provide evidence that each component of the merger stocks appears to have different movement around the merger announcement. First, the estimated bid-ask spreads of both stocks decrease after the announcement, and are similar to the average effective spread of each stock. Second, the information components in the bid-ask spreads of both stocks decrease after the announcement. This finding implies that merger arbitrage investors may be afraid of the arrival of more informed traders before the announcement. Moreover, the decrease in informational components can be interpreted as the reduction of informational asymmetry after the announcement. Finally, the inventory components for the target stocks significantly increase after the merger announcement while those for the bidder stocks are unchanged. The increase in the inventory com-

[3] The authors sent surveys to 28 arbitrageurs who worked at brokerage and money management firms who operate an arbitrage operation within large firms. They received responses from 21 of 28. The sample covers not only the firms who are required to file 13F to SEC but also hedge funds that are not required to file the 13F. Even though their sample size is quite small, their survey results are quite consistent with academic literature. First, most arbitrageurs take their positions within two weeks after the merger announcement while they do not unwind their positions immediately when the deal is canceled. Second, arbitrageurs typically use leverage to take arbitrage positions. This implies that they might face the limited capital problem. Third, only a small number of arbitragers invest in unannounced deals such as rumored deals or anticipated deals. Moreover, they use outside consultants such as antitrust attorneys and tax and accounting specialists implying that many arbitrageurs are more informed traders than the general public investors. Fourth, they use the position limit rule to control risks in their arbitrage portfolios. For example, they limit their individual positions based on a percentage of the overall portfolio or the total amount of capital that can be lost when the deal is canceled. Fifth, they use derivatives in their positions and use the reverse positions to manage portfolio risks. Finally, they carefully use limit orders to execute their trades.

ponents of the target stocks suggests that merger arbitrage investors may play the role of market makers by providing liquidity to the existing target shareholders who do not want to hold the shares. The increase in inventory components also suggests the limit to arbitrage because the limited capital of the arbitrage investors requires more compensation for providing liquidity to the existing target shareholders.

Next, I investigate the dynamic relation between the bid-ask spread and the merger arbitrage spread using vector autoregression. The Granger-causality tests show that the bid-ask spread of each stock helps to predict the future movement of the merger arbitrage spread. Moreover, the unexpected shock of the bid-ask spread causes the increase in the merger arbitrage spread. Finally, the unexpected shock to the merger arbitrage spread appears to affect each bid-ask spread permanently.

Finally, I estimate the informational share proposed by Hasbrouck (1995) to investigate how each stock reflects the arrival of new information in the merger market. To do so, I use the vector error correction model because a cointegration relation exists between the two merger stocks. The estimated results show that the informational share of each stock after the merger announcement is quite similar to the proportion of the market value of each stock to the sum of the two firm's values prior to the announcement. This relation implies that merger arbitrage investors help to form the efficient price. In other words, the merger arbitrage investors eliminate the arbitrage opportunity so that the target stock price reaches the agreed price between the merger parties. As a result, the positive excess return from merger arbitrage is thought of as the compensation associated with forming the effective price.

The remainder of this paper is organized as follows. The next section provides the market microstructure model for merger stocks during the merger period and shows the relation between the merger arbitrage spread and the bid-ask spread. Section 3 describes the data and the variables used in this paper. Section 4 shows the estimation results derived in section 2 by using the state space model with a Kalman filter.

Section 5 investigates the relation between the merger arbitrage spread and the bid-ask spread by using both cross sectional and time series data. Section 6 investigates the price discovery and the informational share. Section 7 concludes the paper.

2 Market microstructure model for merger arbitrage

2.1 Model for bidder and target stocks

The analysis in this paper is based on the Huang and Stoll (1997) spread decomposition model. The model provides a three-way decomposition of the bid-ask spread into adverse selection, inventory holding, and order processing cost components. Suppose there is an unobservable stand-alone fundamental value for each firm $i = 1, 2$, $V_{i,t}^{Alone}$, which is equivalent to the firm value when the merger deal does not exist or fails. Assume that the stand-alone fundamental firm share value evolves as

$$V_{i,t}^{Alone} = V_{i,t-1}^{Alone} + \alpha_i \frac{s_i}{2} v_{i,t-1} + u_{i,t} \qquad i = 1, 2 \tag{31}$$

where $u_{i,t}$ is the unexpected public information shock of firm i at time t and $v_{i,t}$ is the unexpected quote shock or the private information shock incorporated in the quote of firm i at time t. Suppose the $u_{i,t}$s are uncorrelated white-noise processes with mean zero, $E[u_{1,t}u_{2,t}] = 0$ and $E[u_{i,t}^2] = \sigma_u^2$. The quote innovation $v_{i,t}$s are correlated with each other and have mean zero and variance $E[v_{i,t}^2] = \sigma_{v_i}^2$. The quote innovation $v_{i,t}$ can be written as the difference between the observed quote direction and the information in $q_{i,t-1}$ that is not a surprise, that is

$$v_{i,t} = q_{i,t} - E[q_{i,t}|q_{i,t-1}] \tag{32}$$

If $E[q_{i,t}|q_{i,t-1}] = 0$, the system in equation (31) is equal to the original assumption in Huang and Stoll (1997). Alternatively, if $E[q_{i,t}|q_{i,t-1}] = \eta_i q_{i,t-1}$ where η_i is the transformation of the probability of trade reversal[4], then equation (31) will be equivalent to the extended model with induced serial correlation in trade flows, which is ultimately the $AR(1)$ process for quote direction. But I do not have any specific assumption about the expected value of the quote direction so that the quote innovation $v_{i,t}$ could have any functions of $q_{i,t-1}$.

Now, suppose the bidder offers the cash C and the stock with the exchange ratio γ for each share of target stock. Let $V_{i,t}^{Success}$ and n_i be the firm i's share value and the total shares outstanding when the merger succeeds. Following Houston and Ryngaert (1997), the value of the offer to the target and the bidder of the combined firm will be

$$n_1 V_{1,t}^{Success} = (1-\kappa)[V_t(n_1+\gamma n_2) - Cn_2] \tag{33a}$$

$$n_2 V_{2,t}^{Success} = Cn_2 + \kappa[V_t(n_1+\gamma n_2) - Cn_2] \tag{33b}$$

where $V_t = (n_1 V_{1,t}^{Alone} + n_2 V_{2,t}^{Alone} + S)/(n_1+\gamma n_2)$ is the fundamental share value of the combined firm , $\kappa = \gamma n_2/(n_1+\gamma n_2) \in [0,1]$ is the proportion of the combined firm that the target shareholders will receive when the deal completes successfully, and S is any synergies created by combining two firms. Then, the fundamental share value of each firm i can be written as

$$V_{1,t} = \pi(1-\kappa)\left[V_t\left(\frac{n_1+\gamma n_2}{n_1}\right) - C\frac{n_2}{n_1}\right] + (1-\pi)V_{1,t}^{Alone} \tag{34a}$$

$$V_{2,t} = \pi\left[(1-\kappa)C + \kappa V_t\left(\frac{n_1+\gamma n_2}{n_2}\right)\right] + (1-\pi)V_{2,t}^{Alone} \tag{34b}$$

where $\pi \in [0,1]$ is the probability of success of the merger. It is then easy to show that

[4]Assuming that $q_{i,t} = \pm 1$ equally likely, then η_i is equivalent to $1-2\phi_i$ where ϕ_i is the reversal probability of a trade flow i. Suppose that $P(q_{i,t}=+1|q_{i,t-1}=-1) = P(q_{i,t}=-1|q_{i,t-1}=+1) = \phi_i$ and $P(q_{i,t}=+1|q_{i,t-1}=+1) = P(q_{i,t}=-1|q_{i,t-1}=-1) = 1-\phi_i$. Since the reversal probability ϕ_i is less than one and greater than zero, this condition satisfies the stationarity assumption of $AR(1)$ process: $0<\phi_i<1$ implies $|\eta_i|<1$. If $\eta_i = \frac{1}{2}$, then η_i will be equal to zero.

the change of the fundamental value of the bidder firm is $\Delta V_{1,t} = (1-\pi\kappa)\Delta V_{1,t}^{Alone} + \pi(1-\kappa)\frac{n_2}{n_1}\Delta V_{2,t}^{Alone}$, and that of the target firm is $\Delta V_{2,t} = \pi\kappa\frac{n_1}{n_2}\Delta V_{1,t}^{Alone} + (1-\pi+\pi\kappa)\Delta V_{2,t}^{Alone}$.

The midpoints of both stocks should be related to the fundamental value of the combined firm so that the quote midpoints reflect the information from both the stand-alone bidder and the stand-alone target. That is, the change in the fundamental value of each stock should be embedded in the midpoint of each stock. But arbitrageurs will believe that the deal will complete successfully and set the midpoint of the target stock to reflect the midpoint of the offer price, which can be inferred from the recent trading price of the bidder stock. Moreover, the midpoint of each stock is set to reflect the inventory component of its stock. Moreover, assume that the mid-quote point of the target stock is set to follow the offer price at best so that the expected value of the offer is the same as the current mid-point of the target stock. Let $m_{i,t}$ be the mid-quote price of firm i at time t where firm 1 is the bidder and firm 2 is the target company. Thus, the midpoint of each stock is assumed to evolve as

$$m_{1,t} = m_{1,t-1} + \Delta V_{1,t} + \beta_1 \frac{s_1}{2} q_{1,t-1} \tag{35a}$$

$$m_{2,t} = C + \gamma m_{1,t-1} + \Delta V_{2,t} + \beta_2 \frac{s_2}{2} q_{2,t-1} \tag{35b}$$

where β_i is the proportion of the half-spread attributable to inventory holding costs [5]. The econometric specification of the midpoint in equation (35) implies that the bidder stock and the target stock share a common component based on cointegration. Because the target stock's efficient price typically depends on the bidder's efficient price except for the cash merger, the cointegrated price system in equation (35) appears to be appealing. Finally, the observed trade price for each firm can be

[5] Here, sizes of past trades are assumed to be of a normal one so that $-q_t = I_t - I_{t-1} = \sum_{i=1}^{t} q_i - \sum_{i=1}^{t-1} q_i$ where I_t is the cumulative inventory from the market open until time t. So, the inventory at time t is $I_t = I_{t-1} - q_t$. If $q_t = -1$ which transaction is seller initiated so that dealer buys, then $I_t = I_{t-1} + 1$. If $q_t = +1$ which transaction is a buyer initiated so that dealer sells, then $I_t = I_{t-1} - 1$

written as

$$p_{1,t} = m_{1,t} + \frac{s_1}{2} q_{1,t} + \delta q_{2,t-1} \tag{36a}$$

$$p_{2,t} = m_{2,t} + \frac{s_2}{2} q_{2,t} \tag{36b}$$

where $\delta q_{2,t-1}$ reflects the hedging activity by arbitrage investors. The coefficient of q_2 in equation (36a) measures the change in the bidder stock due to the quote behavior of the target stock. Under these price systems, it is easy to show that the bid-ask spread will be s_1 and s_2 for the bidder and the target, respectively.

2.2 Merger arbitrage spread

A direct calculation shows that the arbitrage position, $A_t = p_{2,t} - (C + \gamma p_{1,t})$, is

$$\begin{aligned} A_t \; = \; & \frac{s_2}{2} q_{2,t} + \beta_2 \frac{s_2}{2} q_{2,t-1} + (1-\pi)\alpha_2 \frac{s_2}{2} v_{2,t-1} + (1-\pi) u_{2,t} \\ & -\gamma \left[\frac{s_1}{2} q_{1,t} + \beta_1 \frac{s_1}{2} q_{1,t-1} + (1-\pi)\alpha_1 \frac{s_1}{2} v_{1,t-1} + (1-\pi) u_{1,t} + \delta q_{2,t-1} \right] \end{aligned} \tag{37}$$

That is, the arbitrage positions can be expressed as the compensation for the asymmetric information costs, the inventory holding costs, and the transaction costs. If an investor buys a target stock ($q_2 = -1$) and sells a bidder stock($q_1 = 1$), then he should pay A_t to the arbitrageurs. The first terms in each line reflect the compensation for the simple transaction costs, which are the same as half of the bid-ask spread. The second terms in each line reflect the inventory holding costs charged by the arbitrageurs because they manage their portfolio until the deal consummates. The third and the fourth components in each line are the compensation for the deal completion risks. The last term in the second line bracket is the compensation that arises due to the short-selling costs. All of the costs arising from the bidder stock are proportional to the exchange ratio. So, if the merger structure is a pure cash merger ($\gamma = 0$), there are no costs related to the bidder stock.

2.2.1 Merger arbitrage spread and information asymmetry

If the deal completes successfully, i.e. $\pi = 1$, then the arbitrage position is

$$\frac{s_2}{2}q_{2,t} + \beta_2\frac{s_2}{2}q_{2,t-1} - \gamma\left[\frac{s_1}{2}q_{1,t} + \beta_1\frac{s_1}{2}q_{1,t-1} + \delta q_{2,t-1}\right] \tag{38}$$

which implies that the arbitrageurs will be compensated for their physical transaction costs and inventory holding costs even if the deal completes successfully. Moreover, there is no asymmetric component in the arbitrage positions unless a deal failure risk exists. If the deal fails, i.e., $\pi = 0$, then the arbitrage position is

$$\begin{aligned}&\frac{s_2}{2}q_{2,t} + \beta_2\frac{s_2}{2}q_{2,t-1} + \alpha_2\frac{s_2}{2}v_{2,t-1} + u_{2,t}\\ &-\gamma\left[\frac{s_1}{2}q_{1,t} + \beta_1\frac{s_1}{2}q_{1,t-1} + \alpha_1\frac{s_1}{2}v_{1,t-1} + u_{1,t} + \delta q_{2,t-1}\right]\end{aligned} \tag{39}$$

so that the arbitrageurs bear the additional informational risk, which is the difference between the equations (38) and (39).

$$\alpha_2\frac{s_2}{2}v_{2,t-1} + u_{2,t} - \gamma\left[\alpha_1\frac{s_1}{2}v_{1,t-1} + u_{1,t}\right] \tag{40}$$

which is the combination of the public and private information of the bidder and target stocks, implying that the adverse selection components of the bid-ask spread accurately reflect the deal failure risk.

2.2.2 Merger arbitrage spread and inventory costs

Baker and Savasoglu (2002) provide a simple stylized model assuming that the target shareholders receive $1 + p$ with probability π and receive 1 with probability $1 - \pi$. Then, p can be thought of as the premium to the target shareholders offered by the bidder. Moreover, a number of target shareholders exist who do not want to bear the deal completion risk and sell a total of X shares. In the market, there is a limited number A of arbitrageurs who have the mean-variance utility with absolute risk aversion z^A. Under these assumptions, the mean and variance of the offer will be $\mu = 1 + \pi p$ and $\sigma^2 = \pi(1-\pi)p^2$, respectively. The authors demonstrate that the offering price by the target shareholders should be

$$p^T = 1 + \pi p - \frac{X}{A} z^A \pi(1-\pi)p^2 = \mu - \frac{X}{A} z^A \sigma^2 \tag{41}$$

where z^A is the absolute risk aversion coefficient. The second term is then the additional compensation to the arbitrageurs offered by the target shareholders. This premium occurs because the target shareholders sell to a limited number of capital-constrained arbitrageurs. These authors define the arbitrageurs' capital as the total institutional equity holdings, and show that the capital inversely relates to the subsequent equity returns. Within the inventory theory of market microstructure, a market maker should be compensated for providing liquidity service because he bears a level of risk that is inconsistent with his optimal preferences. The arbitrageurs' positions during the merger period are reflected in their portfolios, which are different from their own preferences. They then want to be compensated for the arbitrage positions that deviate from their optimal portfolios.

Analytically, the second term in equation (41) is equivalent to the bid-ask spread derived under the inventory holding cost argument. Suppose only inventory holding costs exist, i.e., $\beta_2 = 1$, but no informational costs $u_{i,t} = v_{i,t} = 0$. Moreover, the mid-quote price can correctly reflect the payoff of the target share so that $\mu = C + \gamma m_{1,t-1}$. The buying price of the target share becomes $p^T = \mu - s_2$ under the situation when trades occurs at the bid price($q_{2,t} = q_{2,t-1} = -1$). Then, the second term in equation (41) and the bid-ask spread due to inventory holdings, s_2 are equivalent[6]. Stoll (1978)

[6]For complete derivations, see the Appendix

shows that the proportional bid-ask spread is

$$\frac{s_2}{p^T} = z_A \sigma_R^2 |Q| = \frac{X}{A} z_A \sigma^2 \frac{1}{p^T} \tag{42}$$

because $\sigma_R^2 = \frac{\sigma^2}{(p^T)^2}$ and $Q = \frac{X}{A} p^T$ where σ_R^2 is the variance of the return and σ^2 is the variance of the stock price. So, the target price can be written

$$p^T = \mu - s_2 = \mu - \frac{X}{A} z_A \sigma^2 \tag{43}$$

which is the same as the equation (41). Equation (43) shows that the bid-ask spread due to the inventory is equivalent to the premium that is offered by the existing target shareholders. Note that the price in Stoll (1978) is the buying price submitted by the market makers, while the price in Baker and Savasoglu (2002) is the offer price provided by the existing shareholders. Thus, the price in equation (43) should be the market clearing price.

2.3 Estimation

The state-space representation of the dynamics of $\Delta \mathbf{p}_t$ is given by the following system of equations:

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}\boldsymbol{\xi}_t + \mathbf{e}_{t+1} \tag{44a}$$

$$\Delta \mathbf{p}_t = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\boldsymbol{\xi}_t \tag{44b}$$

where $\mathbf{F}, \mathbf{A}'$, and $\mathbf{H}'$ are matrices of parameters of dimension $(r \times r), (2 \times k)$, and $(2 \times r)$, respectively, and $\mathbf{x}_t$ is a $(k \times 1)$ vector of exogenous variables. Equation (44a) is known as the state equation, and equation eqrefobserve1 is known as the observation equation. Here, I use the Kalman filter algorithm to estimate the system of equations (44) by using the maximum likelihood estimation. See the Appendix for details.

One advantage of the structure in equation (44) is that I can estimate the inventory component and the information component without further assumptions about the unobserved vectors of $\mathbf{v}_t$. Here, I assume that the quote direction is an exogenous observable variable. Although the quote direction cannot be observed directly from the market data, previous studies show that the algorithm of Lee and Ready (1991) well approximates the actual quote direction.

Based on the assumptions in the previous section, the price difference $\Delta\mathbf{p}_t = (\Delta p_{1,t}, \Delta p_{2,t})'$ can be written as

$$\begin{aligned}\Delta\mathbf{p}_t &= A(L)\mathbf{q}_t + B(L)\mathbf{u}_t + C(L)\mathbf{v}_t \\ &= \left[A_0 + A_1 L + A_2 L^2\right]\mathbf{q}_t + \left[B_0 + B_1 L\right]\mathbf{u}_t + \left[C_1 L + C_2 L^2\right]\mathbf{v}_t \qquad (45)\end{aligned}$$

where $\mathbf{u}_t = (u_{1,t}, u_{2,t})'$, $\mathbf{q}_t = (q_{1,t}, q_{2,t})'$ and L is the lag operator with $L^k x_t = x_{t-k}$. Here, I assume that a trade direction vector $\mathbf{q}_t$ is the exogenous variable; that is, $\mathbf{x}_t \equiv \mathbf{q}_t$. The coefficients A_i, B_i and C_is are

$$A_0 = \begin{pmatrix} \frac{s_1}{2} & 0 \\ 0 & \frac{s_2}{2} \end{pmatrix} \quad A_1 = \begin{pmatrix} -(1-\beta_1)\frac{s_1}{2} & \delta \\ 0 & -(1-\beta_2)\frac{s_2}{2} \end{pmatrix} \quad A_2 = \begin{pmatrix} 0 & -\delta \\ \gamma\beta_1\frac{s_1}{2} & -\beta_2\frac{s_2}{2} \end{pmatrix}$$

$$B_0 = \begin{pmatrix} 1-\pi\kappa & \pi(1-\kappa)\frac{n_2}{n_1} \\ \pi\kappa\frac{n_1}{n_2} & 1-\pi(1-\kappa) \end{pmatrix} \quad B_1 = \begin{pmatrix} 0 & 0 \\ \gamma(1-\pi) & -(1-\pi) \end{pmatrix}$$

$$C_1 = \begin{pmatrix} (1-\pi\kappa)\alpha_1\frac{s_1}{2} & \pi(1-\kappa)\alpha_2\frac{s_2}{2}\frac{n_2}{n_1} \\ \pi\kappa\alpha_1\frac{s_1}{2}\frac{n_1}{n_2} & [1-\pi(1-\kappa)]\alpha_2\frac{s_2}{2} \end{pmatrix} \quad C_2 = \begin{pmatrix} 0 & 0 \\ \gamma(1-\pi)\alpha_1\frac{s_1}{2} & -(1-\pi)\alpha_2\frac{s_2}{2} \end{pmatrix}$$

$$
\boldsymbol{\xi}_t = \begin{pmatrix} \mathbf{u}_t \\ \mathbf{u}_{t-1} \\ \mathbf{v}_t \\ \mathbf{v}_{t-1} \\ \mathbf{v}_{t-2} \end{pmatrix} \quad \mathbf{F} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \end{pmatrix} \quad \mathbf{e}_t = \begin{pmatrix} \mathbf{u}_t \\ 0 \\ \mathbf{v}_t \\ 0 \\ 0 \end{pmatrix}
$$

and

$$
\mathbf{H'} = \begin{pmatrix} B_0 & B_1 & C_0 & C_1 & C_2 \end{pmatrix} \quad \text{and } \mathbf{A'} = \begin{pmatrix} A_0 & A_1 & A_2 \end{pmatrix}
$$

3 Data and variables

3.1 Sample Construction

The initial sample of mergers and acquisitions is drawn from the Thomson Financial SDC mergers and acquisitions database. A total of 2,518 US merger and acquisition deals between January 1, 1994, and December, 31, 2008, meet the following screen criteria. First, the deal status must be completed or withdrawn. That is, uncertain or rumored deals are deleted from the sample. Second, the forms of the deals are all mergers (SDC deal form code: M), acquisition of majority interest (AM), acquisition of partial interest (AP), and acquisition of remaining interest (AR). Thus, I exclude all deals classified as exchange offers, acquisition of assets, acquisition of particular assets, privatization, buybacks, recapitalization, and acquisition (of stock). Third, the method of payments should be combinations of cash, common stock, or cash equivalents. Thus, if the deal price is determined based on earnings, sales, or cash flow after the deal completion, then the corresponding deals is deleted. Moreover, I keep only the firms whose CRSP share codes are either 10 or 11 (common stock). Fourth, the bidder must hold less than 50% of shares and be seeking control at the time of the announcement date. Fifth, both merger parties are U.S. firms publicly traded on the NYSE, AMEX, or Nasdaq and have 60 days of price data prior to the merger announcement. Finally, the deal duration should be more than 15 days after the announcement to obtain sufficient stock price data. Basic deal information such as the announcement date, the deal completion/withdrawal date, and the method of payment is collected from the SDC. Because the original SDC data have some errors, I use Lexis-Nexis or EDGAR to correct wrong information by inspecting the original 8-F statement reported to the SEC. I found that there are some errors and missing data in the SDC database [7].

The merger arbitrage spread can be related to the structure of the payment to the target shareholders because the wealth of the target shareholders changes in the case of floating value or fixed exchange ratio deals. The payment method used in this paper is either cash or stock. When a bidder uses the stock financed mergers, he could use the fixed exchange ratio or the floating exchange ratio for the merger consideration. That is, the bidder could offer a fixed number of shares for each target share or the fixed value of the bidder's stock to be exchanged for each target share at the announcement date. A mixed offer is the sum of the fixed cash and fixed exchange ratio, the fixed cash and floating exchange ratio, or sometimes a fixed value offer so that the proportion of cash and stock is determined later by the shareholders. Another interesting merger technique is the collar offer, in which the number of bidder shares depends on the bidder stock price range over some given period. The collar offer can be thought of as a mixture of the fixed exchange ratio offer and the floating exchange ratio offer.

To analyze the effect of the deal structure, I first group the merger deals into three categories: pure cash deals, pure stock deals, or a mixture of cash and stock. The pure stock deals are divided into three groups: fixed exchange ratio stock swap, fixed pricing (or floating exchange) stock swap, and collar. Similarly, the mixed deals

[7] Overall, the SDC seems to record the final deal condition rather keep the initial contract. For the time series analysis, however, I need contents both of initial contract and of amended contract. The SDC also record the exchange ratio of the fixed value deals even though it is not determined at the announcement date. This may mislead users to misclassify the fixed value deal as the fixed-exchange ratio deal. Moreover, some deals classify as the fixed value deals even though they are collar type deal.

are divided into three groups: cash plus fixed exchange ratio stock, cash plus fixed value pricing, and collar with a cash payment. Therefore, depending on the payment structure, I have seven groups.

To analyze the merger arbitrage, I use the TAQ database to extract the trade and quote data, and the CRSP database to identify the stock information and the dividend. The entire stock price is adjusted for dividends and for stock splits using the CRSP cumulative factor (CFACPR). I classify the trade direction by using the Lee and Ready (1991) algorithm that compares the transaction price to the posted bid and ask quotes. I first identify the quote, which must be at least five seconds old. The algorithm is that if a trade is closer to the bid (ask) of the quote, it is classified as seller (buyer) initiated. If the transaction occurs at the midpoint of the quote, a tick test is performed. A tick test means that if the last price change prior to the trade is negative (positive), the trade is classified as seller (buyer) initiated.

3.2 Sample statistics

In Table 5, I provide a summary of the mergers used in this study. Over the sample period, approximately 28% of the deals was pure-cash deals, and approximately 54% of the deals was pure stock deals. Among the pure stock exchange deals, the most frequent deal type is the fixed exchange ratio deal, while the floating exchange deal type is relatively small. Moreover, collar deals were approximately 16% of the sample. Consistent with the merger wave literature, the aggregate takeover activities appear to have been relatively high in the late 1990s. (See Harford (2005)).

Table 6 reports the deal characteristics of my takeover sample. First, the average deal value is higher when stocks are considered than when a simple cash-like (pure cash and fixed value stock) deal is considered. When the deal value increases, the bidders are more likely to use a stock swap rather than a pure cash payment. Second, the market average value of the bidder firms is always greater than that of the target

Table 5: Yearly distribution of Mergers and Acquisitions announcements from 1994 to 2008

This table provides the yearly distribution of mergers and acquisitions samples in SDC database from 1994 to 2008. The sample includes only deals where the deal status must be completed or withdrawn, and exclude all deals classified as exchange offers, acquisition of assets, acquisition of certain assets, privatization, buybacks, recapitalization, and acquisition of stock. Moreover, both target and acquirer firms are listed in the CRSP database and are US public firms. We keep only firms whose CRSP share codes are 10 or 11. That is, if there exists a stock swap offer, the exchanged stock must be a common stock. "Cash" means a pure cash deal, and "Stock" represents a pure stock deal. Of the "Stock" deals, "Stock(FL)" represents the pure stock deals with fixed value (floating exchange ratio) while "Stock(FX)" means the pure stock deals with a fixed exchange ratio. "Mix" represents a type of deals that a merger consideration consists of stock and cash. "Mix(FL)" and "Mix(FX)" represent a mixed deal with floating exchange ratio and fixed exchange ratio respectively. "Collar" means that there exists a contingent claim in merger deal at announcement date. Of the collar deals, "Stock" means that a target shareholder will receive a bidder stock when a deal completes, and "Mix" means a target shareholder will receive an additional cash amount as well as a bidder stock.

Year	No Collar						Collar			Total
	Cash	Stk(FL)	Mix(FL)	Stk(FX)	Mix(FX)	Sum	Stk	Mix	sum	
1994	38	6	5	51	6	106	25	2	27	133
1995	47	9	1	93	3	153	25	5	30	183
1996	44	7	1	88	9	149	41	6	47	196
1997	54	9	6	140	11	220	55	6	61	281
1998	66	13	6	155	14	254	49	4	53	307
1999	83	5	6	133	16	243	44	18	62	305
2000	60	7	7	124	13	211	21	8	29	240
2001	43	0	1	50	23	117	17	9	26	143
2002	30	1	4	20	15	70	9	7	16	86
2003	30	2	5	29	22	88	11	9	20	108
2004	35	3	13	35	27	113	3	5	8	121
2005	44	1	2	21	31	99	6	3	9	108
2006	62	1	1	20	23	107	5	4	9	116
2007	56	0	3	19	31	109	3	7	10	119
2008	33	0	2	17	18	70	1	1	2	72
Total	725	64	63	995	262	2,109	315	94	409	2,518

firms. Third, the ratio of the target firm value to the deal value is less than one, implying the existence of synergy effects. Fourth, the deal duration, the difference between the deal consummation date and the announcement date, is 134 days on average. The duration is typically longer when considering a stock swap than when considering only cash. Finally, the average cumulative abnormal return for target firms is typically high. The average CAR for bidder firms are close to zero or negative as expected.

Table 6: The summary of deal characteristics

This table provides the summary of deal characteristics of our samples. The number of observations, deal value (V_{Deal}), the market values of bidder(V_1) and target firms(V_2), the ratio of target firm market value to deal value(V_2/V_{Deal}), a deal duration(Dur), and cumulative abnormal return over a 5-day window of each firm are reported (CAR_i where i = 1 for bidder and i = 2 for target). The deal value is extracted from SDC database. The market values of each firm are calculated by averaging the three day market values prior to the merger announcement. The deal duration is the difference between the announcement date and the deal completion date or withdrawn date. The cumulative abnormal return(CAR) of each firm is measured over a 5-day window around the merger announcement date using a market model. The parameters of market model are estimated over a 180-day window from 240 to 60 days prior to the announcement date.

	No Collar					Collar		Total
	Cash	Stock(FL)	Mix(FL)	Stock(FX)	Mix(FX)	Stock	Mix	
Obs.	725	64	63	995	262	315	94	2,518
V_{Deal}(Mill. $)	799	299	1,259	2,250	1,976	700	1,154	1,494
Dur(Day)	99	143	139	147	160	142	167	134
V_1 (Mill. $)	19,287	27,729	11,772	11,615	7,989	8,084	3,596	13,119
V_2 (Mill. $)	506	206	829	1,583	1,426	428	649	1,023
V_2/V_{Deal}(%)	67.66	68.52	67.57	71.96	71.75	70.19	62.89	69.94
V_1 (%)	0.58	-0.74	-0.28	-4.71	-3.62	-1.46	-1.28	-2.32
V_2 (%)	32.42	24.47	27.05	16.31	19.53	19.99	22.46	22.45

The various average bid-ask spread measures before and after the announcement are reported in Table 7. The quoted half spread usually measures the total transaction costs. The effective spread is the absolute value of the difference between the trade price and the quote midpoint just prior to the trade. The daily average values of these spread measures are calculated by weighting each spread by the number of trades. If the trades occur only at the ask or bid quote price, then the quoted spread and the effective spread should be the same. From table 7, it is obvious that the effective spread is less than the quoted half spread across all subsamples. The average spread measures are lower for bidder stocks than for target stocks. This result implies that the target stocks are usually more illiquid than the bidder stocks. Moreover, both spread measures decrease after the announcement, consistent with the previous literature. Before the announcement, it is typical for the spread measures of the target shares to be greater than those of the bidder firms. After the announcement, however, the magnitude of the bid-ask spread difference between the bidder and the target is

Table 7: The summary of bid-ask spread around the merger announcement

This table provides the comparison of the bid-ask spread of bidder and target before and after the merger announcement. The spread measures presented in this table is daily average of spread measures using TAQ database. The daily spread is calculated by weighting each observed spread by the number of trades. The pre-announcement period for calculating the average spread is 100 days prior to the announcement, and the post-announcement period is from the announcement date to the deal completion date or the withdrawn date. The quoted bid-ask spread is the difference between the ask price and the bid price. The effective spread is the absolute value of the difference between the trade price and the quote midpoint just prior to the trade. "*", "**", "***" indicate significance at the 10%, 5%, and 1% levels respectively.

Method Of	Bidder				Target			
Payment	Before	After	Diff	T-Value	Before	After	Diff	T-Value
Panel A. Quoted Half Spread								
Cash	0.063	0.063	−0.001	−0.78	0.116	0.057	−0.059	−16.63***
Stock(FxVal)	0.085	0.080	−0.005	−3.47***	0.202	0.133	−0.069	−4.21***
Mix(FxVal)	0.108	0.072	−0.035	−1.14	0.122	0.080	−0.041	−6.89***
Stock(FxRatio)	0.099	0.082	−0.017	−13.03***	0.150	0.140	−0.010	−3.41***
Mix(FxRatio)	0.086	0.076	−0.010	−2.53**	0.099	0.093	−0.006	−1.82*
Collar	0.102	0.091	−0.011	−5.71***	0.162	0.137	−0.025	−6.19***
Mixed Collar	0.075	0.066	−0.009	−3.74***	0.127	0.089	−0.038	−6.54***
Panel B. Effective Spread								
Cash	0.048	0.048	0.000	−0.24	0.092	0.045	−0.047	−16.15***
Stock(FxVal)	0.062	0.057	−0.005	−3.37***	0.157	0.101	−0.055	−4.10***
Mix(FxVal)	0.078	0.054	−0.024	−1.16	0.100	0.061	−0.039	−6.75***
Stock(FxRatio)	0.081	0.067	−0.013	−11.09***	0.120	0.111	−0.009	−3.88***
Mix(FxRatio)	0.068	0.061	−0.007	−2.86***	0.083	0.072	−0.011	−4.75***
Collar	0.079	0.071	−0.008	−5.28***	0.128	0.105	−0.023	−7.12***
Mixed Collar	0.059	0.051	−0.008	−3.28***	0.102	0.068	−0.034	−6.35***

reduced.

Table 8 shows that the trading volume and the dollar trading volume increase significantly for both stocks after the announcement. The number of trades for the bidder stocks increases significantly across the deal payment methods. However, the number of trades for the target shares is similar before and after the announcement. This result implies that the trading volume per trade of a target share increases significantly after the merger announcement. Easley and O'hara (1987) argue that informed traders prefer to trade large amounts, which cause an adverse selection problem. The increase in the trading volume in the target stock may be related to

Table 8: The summary of trading activities around the merger announcement

This table provides the summary statistics of trading activities around the merger announcement. The trading activity variables presented in this table is average value of daily measures from the TAQ database. The pre-announcement period for calculating each variable is 100 days prior to the merger announcement, and the post-announcement period is from the announcement date to the deal completion date or the withdrawn date. "*", "**", "***" indicate significance at the 10%, 5%, and 1% levels respectively.

Method Of	Bidder				Target			
Payment	Before	After	Diff	T-Value	Before	After	Diff	T-Value
Panel A. Trading Volume (Million Shares)								
Cash	2.507	2.622	0.115	2.50**	0.233	0.486	0.253	9.33***
Stock(FxVal)	1.206	1.295	0.089	2.10**	0.087	0.135	0.048	2.46**
Mix(FxVal)	2.240	2.513	0.273	1.51	0.754	0.914	0.160	1.70*
Stock(FxRatio)	1.304	1.746	0.442	8.64***	0.449	0.549	0.100	4.19***
Mix(FxRatio)	1.064	1.463	0.399	6.05***	0.403	0.498	0.095	3.22***
Collar	0.915	1.222	0.307	4.11***	0.137	0.196	0.058	6.26***
Mixed Collar	0.513	0.691	0.178	3.70***	0.130	0.259	0.130	5.34***
Panel B. Dollar Trading Volume (Million Dollars)								
Cash	94.363	95.631	1.268	0.62	4.572	12.311	7.740	9.18***
Stock(FxVal)	67.218	69.122	1.904	0.54	1.802	3.569	1.767	3.08***
Mix(FxVal)	69.609	68.113	−1.496	−0.20	16.100	21.802	5.701	1.38
Stock(FxRatio)	70.029	86.049	16.020	4.71***	12.851	16.997	4.146	3.71***
Mix(FxRatio)	45.361	59.212	13.851	5.96***	14.437	20.653	6.216	5.35***
Collar	43.214	54.850	11.636	2.74***	3.461	6.135	2.674	7.46***
Mixed Collar	19.220	25.167	5.947	3.73***	3.307	7.824	4.517	4.95***
Panel C. The number of Trades (Thousand)								
Cash	4.441	4.883	0.442	3.25***	0.622	0.640	0.018	0.71
Stock(FxVal)	1.233	1.315	0.083	1.93*	0.088	0.097	0.009	0.56
Mix(FxVal)	3.976	4.491	0.515	2.02**	2.005	2.007	0.002	0.01
Stock(FxRatio)	2.125	2.920	0.796	4.23***	0.870	0.895	0.025	0.40
Mix(FxRatio)	2.191	3.191	1.001	6.53***	1.022	1.158	0.136	2.14**
Collar	0.953	1.236	0.284	2.97***	0.147	0.144	−0.003	−0.31
Mixed Collar	0.880	1.241	0.361	2.25**	0.250	0.304	0.054	2.37**
Panel D. Order Imbalance (Thousand Shares)								
Cash	61.777	43.135	−18.642	−2.24**	0.116	−89.329	−89.445	−17.23***
Stock(FxVal)	68.593	77.969	9.376	0.79	1.393	−26.416	−27.809	−5.26***
Mix(FxVal)	−26.257	−22.681	3.576	0.17	−22.640	−63.293	−40.653	−2.09**
Stock(FxRatio)	50.982	105.750	54.768	8.76***	4.767	−1.649	−6.415	−2.97***
Mix(FxRatio)	48.551	89.841	41.290	3.92***	13.239	0.987	−12.253	−3.04***
Collar	46.224	71.288	25.064	3.67***	2.072	−16.425	−18.497	−10.13***
Mixed Collar	34.658	69.416	34.757	3.70***	4.758	−17.608	−22.365	−4.98***

the increase in the asymmetric component in the bid-ask spread. The order imbalance before and after the announcement of each firm is reported in panel (D). The order imbalance measure in this study is the trading volume at the ask price (buying shares) less the trading volume at the bid price (selling shares). A positive order imbalance implies a net buying pressure, and a negative value implies a net selling pressure. Before the announcement, the order imbalances of both stocks are positive, indicating that buying pressure exists. The average order imbalance of the bidder firms increases after the announcement, implying that there are more buyers for the bidder stock after the announcement. An interesting feature is that the average order imbalance of the target shares becomes negative after the announcement. This result implies that a selling pressure from the existing target shareholders exists, and the arbitrageurs use limit orders during the merger period.

4 Estimation results

Table 9 reports the bid-ask spread estimates and their components when the trade direction is assumed to be observable. I estimate parameters in each deal and report the average of the estimated component by the methods of payment. All of the parameter estimates satisfy the convergence criteria. The focus here is on the change in the bid-ask spread components before and after the merger announcement.

First, the bid-ask spread measures of both stocks drop significantly after the merger announcement, and the difference is significant. Although the estimated bid-ask spreads in panel (A) are slightly larger than the actual quoted half spread measures, the estimated spread measures have quite similar patterns to the actual data. The decrease in the bid-ask spread measures is consistent with the previous literature [8]. The decrease in the estimated bid-ask spreads is more pronounced in the target shares than in the bidder shares. Moreover, the decrease in the bid-ask spread implies an increase in liquidity, which can be related to an increase in trading volumes.

[8] See Conrad and Niden (1992), Foster and Viswanathan (1995)

Table 9: Estimation of parameters of state space model by using Kalman filter algorithm

This table provides the estimation results of price system in the state space model of (45). The estimated bid-ask spreads for bidder and target stock are s_1 and s_2 respectively. The informational component of bid-ask spread for bidder and target stock are α_1 and α_2 respectively. The inventory component of bid-ask spread for bidder and target stock are β_1 and β_2 respectively. "*", "**", "***" indicate significance at the 10%, 5%, and 1% levels respectively.

Method	# of	Bidder				Target			
Of Payment	Obs	Before	After	Diff	T-Value	Before	After	Diff	T-Value
Panel A. Estimated Bid-Ask spread (s_i)									
Cash	725	0.079	0.080	0.002	0.97	0.119	0.048	−0.070	−15.43***
Stock(FxVal)	64	0.110	0.106	−0.004	−0.77	0.221	0.111	−0.110	−5.09***
Mix(FxVal)	63	0.091	0.078	−0.013	−1.22	0.126	0.073	−0.053	−4.48***
Stock(FxRatio)	995	0.149	0.121	−0.028	−9.59***	0.185	0.121	−0.065	−12.14***
Mix(FxRatio)	262	0.086	0.071	−0.015	−3.51***	0.087	0.061	−0.027	−4.66***
Collar	315	0.147	0.128	−0.019	−4.23***	0.181	0.095	−0.086	−10.78***
Mixed Collar	94	0.090	0.073	−0.017	−3.34***	0.116	0.055	−0.062	−5.18***
Panel B. Information Component (α_i)									
Cash	725	0.540	0.496	−0.044	−2.29**	0.525	0.524	−0.002	−0.08
Stock(FxVal)	64	0.560	0.553	−0.007	−0.13	0.550	0.598	0.048	0.66
Mix(FxVal)	63	0.619	0.500	−0.119	−2.38**	0.534	0.488	−0.047	−0.81
Stock(FxRatio)	995	0.557	0.517	−0.040	−2.52**	0.564	0.535	−0.028	−1.80*
Mix(FxRatio)	262	0.554	0.564	0.010	0.33	0.548	0.574	0.026	0.88
Collar	315	0.550	0.540	−0.010	−0.34	0.553	0.513	−0.041	−1.44
Mixed Collar	94	0.535	0.473	−0.062	−1.23	0.492	0.541	0.048	0.96
Panel C. Inventory Compoenet (β_i)									
Cash	725	0.426	0.442	0.016	0.98	0.297	0.320	0.023	1.61
Stock(FxVal)	64	0.414	0.436	0.023	0.52	0.214	0.299	0.085	1.77*
Mix(FxVal)	63	0.375	0.445	0.071	1.50	0.278	0.392	0.115	2.39**
Stock(FxRatio)	995	0.396	0.383	−0.013	−1.32	0.307	0.324	0.017	1.49
Mix(FxRatio)	262	0.487	0.452	−0.034	−1.41	0.405	0.402	−0.004	−0.14
Collar	315	0.393	0.391	−0.002	−0.13	0.251	0.314	0.063	3.19***
Mixed Collar	94	0.417	0.373	−0.044	−1.08	0.302	0.413	0.110	2.45**

Second, the information components in the bid-ask spreads of both stocks decrease after the announcement, although some are insignificant. However, the adverse information components are still the largest portion of the bid-ask spread, suggesting that merger arbitrage investors have concerns about the deal failure risks. Copeland and Galai (1983) and Glosten and Harris (1988) predict that market makers may adversely increase the bid-ask spread when they expect informed trades. The large proportion of adverse information components prior to the merger announcement suggests that market makers recognize the likelihood of informed trading from the market order flows and widen the bid-ask spread prior to the announcement. However, these adverse information problems can be resolved by the public announcement, which results in a decrease in the bid-ask spread.

Third, the inventory components of the bid-ask spreads appear to increase after the announcement even though the difference in the proportion of inventory components for the bidder stocks is unchanged. However, the inventory components for the target stocks increase significantly after the merger announcement. As reported in Table 8, a substantial selling pressure exists as a result of selling demands from the existing target shareholders. This finding supports the limits of arbitrage theory.

Moreover, a significant portion of the inventory costs reflects the carrying costs of the merger arbitrage portfolios until the deal consummation date. The increase in the inventory components of the target shares suggests that arbitrageurs can play the role of a market maker. If arbitrageurs are buyers of target shares in takeover markets, they will handle these negative imbalances and increase their inventory holding, which is shown in the increase in the inventory component. This result implies that arbitrageurs can provide liquidity to sellers, and require compensation by quoting their best buying price.

The estimated components of the bid-ask spread reported in Table 9 are the proportional elements of the bid-ask spread. The proportion of the bid-ask spread may not be adequate to explain the change in each component of the bid-ask spread. If the total bid-ask spread decreases due to pure order processing costs after the announcement and other components still remain, each proportion for the information and inventory components in the bid-ask spread could increase. To address these issues, I calculate the dollar spread components. Table 10 reports the dollar information spread and the dollar inventory spread of each stock around the announcement. The

Table 10: Estimated dollar spread of state space model by using Kalman filter algorithm

This table provides the estimation results of price system in the state space model of (45). The dollar bid-ask spread related with asymmetric information for bidder and target stock are $\alpha_1 s_1$ and $\alpha_2 s_2$ respectively. The dollar bid-ask spread related with inventory component for bidder and target stock are $\beta_1 s_1$ and $\beta_2 s_2$ respectively. "*", "**", "***" indicate significance at the 10%, 5%, and 1% levels respectively.

Method	# of	Bidder				Target			
Of Payment	Obs	Before	After	Diff	T-Value	Before	After	Diff	T-Value
Panel A. Dollar Spread (Information portion)									
Cash	725	0.043	0.040	−0.003	−1.27	0.061	0.024	−0.036	−12.09***
Stock(FxVal)	64	0.061	0.060	−0.001	−0.17	0.124	0.067	−0.057	−3.11***
Mix(FxVal)	63	0.047	0.039	−0.008	−0.96	0.065	0.028	−0.037	−3.82***
Stock(FxRatio)	995	0.080	0.063	−0.017	−5.29***	0.105	0.063	−0.042	−9.20***
Mix(FxRatio)	262	0.045	0.039	−0.005	−1.40	0.044	0.037	−0.007	−1.59
Collar	315	0.081	0.069	−0.012	−2.17**	0.096	0.051	−0.045	−6.56***
Mixed Collar	94	0.049	0.030	−0.019	−3.47***	0.054	0.031	−0.023	−2.93***
Panel B. Dollar Spread (Inventory spread)									
Cash	725	0.032	0.034	0.003	1.94*	0.025	0.018	−0.006	−3.55***
Stock(FxVal)	64	0.036	0.042	0.005	0.96	0.035	0.045	0.011	0.95
Mix(FxVal)	63	0.028	0.028	0.001	0.25	0.027	0.036	0.010	1.15
Stock(FxRatio)	995	0.058	0.046	−0.012	−5.70***	0.050	0.047	−0.002	−0.63
Mix(FxRatio)	262	0.037	0.029	−0.008	−2.00**	0.026	0.028	0.002	0.59
Collar	315	0.057	0.047	−0.010	−3.54***	0.037	0.036	−0.001	−0.12
Mixed Collar	94	0.034	0.027	−0.007	−2.00**	0.026	0.028	0.002	0.33

dollar components in the bid-ask spreads appear to display similar patterns to the proportions of the components. The decrease in the bid-ask spreads is more pronounced in the target stocks. The bid-ask spread due to an adverse information decrease after the announcement implies that most of the decrease in the bid-ask spread after the announcement is due to the decrease in information asymmetry. For target stocks, the inventory bid-ask spread after the announcement appears to be unchanged or to slightly increase after the announcement, although the inventory bid-ask spread of cash deals decreases significantly. These results are similar in terms of the results of the change in the percentage inventory component around the event. However, the dollar inventory component of the bidder stocks decreases after the announcement, especially in fixed exchange stock deals. This decrease may be due to the hedging demand of arbitrageurs who buy target shares and sell bidder shares simultaneously.

5 Merger arbitrage spread and bid-ask spread

5.1 Cross-sectional relation between merger arbitrage and bid-ask spread

As seen in the previous section, the arbitrage spread can be expressed as a function of the bid-ask spread components. In this section, I investigate the relation between the merger arbitrage spread and the bid-ask spread cross-sectionally. The dependent variable is the daily merger arbitrage spread, which is calculated based on the information from the financial statement reported by the merger parties. If there are other events such as stock splits, dividend payments, or a change in deal conditions, then I adjust the price for each day. The primary independent variables are the quoted half spread of each firm. I include the Fama-French 3 factors to control the market wide premium for each date. Moreover, I include the interaction term, which is the product of the market excess return and the indicator variable, which is 1 if the market excess return is positive, and zero otherwise. This interaction variable can control the non-linear characteristics of the merger arbitrage returns (Mitchell and Pulvino (2001)). If the market is volatile, the merger arbitrage spread is likely to widen due to market uncertainty. I include the volatility index (VIX) to control market uncertainty. Moreover, I include other market microstructure variables related to trading activities. Finally, because the merger arbitrage spread is affected by time and firm size, I include the time from the announcement date and each firm's market value. Table 11 reports the cross-sectional regression for subsamples of deals where the dependent variable is the merger arbitrage spread.

First, the quoted bid-ask spread of the target firm is positively related to the merger arbitrage spread across all subsamples. But, the quoted spread of bidder firms is negative for the fixed value deals while it is positive when the merger deal is varying.

Table 11: Regressions of merger arbitrage spread for subsamples of deals

This table shows pooled OLS estimates of merger arbitrage spread, defined as $a_t = p_{2,t} - \gamma p_{1,t} - C$ where $p_{1,t}$ and $p_{2,t}$ are bitter stock price and target stock price at time t, and γ is the fixed exchange ratio. Each column is the regression result of subsample: pure cash deal(I), pure stock deal with fixed value(II), mixed deal with fixed value(III), pure stock deal with fixed exchange ratio(IV), mixed deal with fixed exchange ratio(V), pure stock deal with collar(VI), and mixed deal with collar(VII). $QS_{i,t}$ is the daily quoted half-spread of firm i at time t. R_t^M and R_t^F are daily market return and risk-free rate at time t. $\boldsymbol{I}$ is an indicator variable with one if $R_t^M - R_t^F > 0$, and zero otherwise. SMB_t and SML_t denote the Fama-French factors at time t. VIX_t denotes the CBOE volatility index. $Comp_t$ is an indicator variable taking one if a deal completes successfully, and zero otherwise. $\$Vol_{i,t}$ denotes the daily dollar volume for stock i at time t. $MV_{i,t}$ is the daily market value of stock i at time t. $Trades_{i,t}$ denotes the daily number of transaction for stock i at time t. Δ_t is the duration from announcement date to day t. The subscript i is a bidder firm for $i = 1$ and a target firm for $i = 2$. The numbers in parentheses are standard errors. "*", "**", "***" represents significance at 10%, 5%, and 1% level respectively.

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
$QS_{1,t}$	−0.439**	−0.624	−0.222***	3.106***	0.022	1.132***	1.549***
	(0.203)	(0.628)	(0.079)	(0.149)	(0.044)	(0.165)	(0.523)
$QS_{2,t}$	3.727***	2.123***	1.873***	1.892***	1.467***	2.053***	2.845***
	(0.242)	(0.172)	(0.195)	(0.064)	(0.089)	(0.092)	(0.236)
$R_t^M - R_t^F$	−5.407**	−10.980***	−7.590**	1.319	−3.310*	−6.194***	0.376
	(2.112)	(4.073)	(3.281)	(1.363)	(1.820)	(2.303)	(3.791)
$(R_t^M - R_t^F) \cdot \boldsymbol{I}$	7.061**	13.717**	11.570**	−0.929	11.305***	23.036***	−7.044
	(3.496)	(5.837)	(5.130)	(2.135)	(3.046)	(3.404)	(5.756)
SMB_t	1.236	−2.260	−6.177**	−0.347	3.906**	3.232	−0.351
	(2.179)	(3.665)	(3.135)	(1.301)	(1.896)	(2.150)	(3.283)
HML_t	−3.133	−9.254*	−1.257	−3.318**	2.137	6.963**	−18.648***
	(2.188)	(4.884)	(3.488)	(1.503)	(1.926)	(2.731)	(4.026)
VIX_t	0.047***	0.043***	0.027***	0.014***	0.046***	0.049***	0.075***
	(0.002)	(0.004)	(0.003)	(0.001)	(0.002)	(0.002)	(0.004)
$Comp_t$	−1.548***	−0.253*	0.135*	−2.012***	−1.313***	−0.698***	−1.736***
	(0.044)	(0.134)	(0.079)	(0.029)	(0.043)	(0.064)	(0.053)
$\log(\$Vol_{1,t})$	0.224***	−0.092**	−0.085***	0.065***	0.030*	−0.210***	−0.164***
	(0.019)	(0.038)	(0.025)	(0.013)	(0.017)	(0.018)	(0.028)
$\log(\$Vol_{2,t})$	−0.078***	0.036	−0.183***	0.302***	0.165***	0.182***	−0.123***
	(0.014)	(0.025)	(0.021)	(0.011)	(0.016)	(0.015)	(0.026)
$\log(MV_{1,t})$	−0.254***	−0.114***	0.319***	0.239***	0.032**	0.034**	−0.321***
	(0.016)	(0.028)	(0.028)	(0.010)	(0.016)	(0.015)	(0.030)
$\log(MV_{2,t})$	0.445***	0.418***	0.141***	−0.090***	0.188***	0.204***	0.959***
	(0.015)	(0.030)	(0.028)	(0.010)	(0.017)	(0.015)	(0.030)
$\log(Trade_{1,t})$	−0.182***	−0.190***	−0.215***	−0.127***	−0.062***	0.168***	0.069**
	(0.019)	(0.039)	(0.023)	(0.014)	(0.017)	(0.020)	(0.031)
$\log(Trades_{2,t})$	0.165***	0.003	0.285***	−0.127***	−0.104***	−0.007	0.068**
	(0.018)	(0.033)	(0.025)	(0.014)	(0.018)	(0.020)	(0.033)
Δ_t	0.003***	0.001***	0.000	0.001***	−0.001***	−0.001***	−0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	14.56	19.23	10.80	14.13	13.47	11.87	25.42
Obs.	40,384	4,370	5,316	89,720	26,474	25,938	9,597

Second, the non-linear relation appears to exist between the merger arbitrage spread and the market excess return, supporting Mitchell and Pulvino (2001). The positive coefficient for the interaction term implies that the merger arbitrage spread has call option-like features. Other asset pricing factors are not likely to relate to the merger arbitrage spread. Third, market volatility is positively related to the merger arbitrage spread. This result suggests that arbitrageurs may require more compensation when the market is more volatile. Fourth, when the merger deal is successfully completed, the merger arbitrage spread should be narrower than when the merger fails. The negative coefficient of $Comp_t$ in the regression reflects these predictions. Fifth, heavy trading volume in a particular stock may narrow the arbitrage spread. Finally, as time moves far away from the announcement date, the merger arbitrage spread appears to increase for the fixed value deals, but to decrease for the floating value deals. The merger arbitrage spread converges to zero when the deal is close to the deal consummation date. But, for the fixed deals, the merger arbitrage spread appears to widen as time moves far away from the announcement date, implying that merger arbitrageurs expect that the deal duration for the fixed value deals ended sooner than the floating value deals. Finally, it is clear that the merger success and the merger arbitrage spread are negatively and significantly related across all subsamples.

Overall, this cross-sectional regression confirms that there are clear relations between the merger arbitrage spread and the bid-ask spread of a target stock. In next section, I examine the relation between the merger arbitrage spread and the bid-ask spread using dynamic time series analysis.

5.2 Time-Series analysis

5.2.1 Vector autoregressive regression analysis

This section examines the dynamic behavior of merger arbitrage spreads and bid-ask spreads. As noted in the previous section, the merger arbitrage spread can be expressed as the bid-ask spread components. Each component of the bid-ask spread can be matched to the different types of compensation to the arbitrageurs. This ability suggests that the total bid-ask spread should be related with the merger arbitrage spreads during the merger period. I use the vector autoregression analysis (VAR) to more precisely analyze the predictability of a bid-ask spread to a merger arbitrage spread. Let $\mathbf{y}_t = [A_t, s_{1,t}, s_{2,t}]'$ be the vector of the time series where A_t is the arbitrage spread at time t, $s_{1,t}$ is the bid-ask spread of a bidder firm at time t, and $s_{2,t}$ is the bid-ask spread of a target firm at time t. Then, the reduced-form vector autoregressive regression(VAR) of order p has the following standard representation:

$$\mathbf{y}_t = B_0 + \sum_{i=1}^{p} B_i \mathbf{y}_{t-i} + \boldsymbol{\epsilon}_t \tag{46}$$

where B_i is the (3×3) coefficient matrix and $\boldsymbol{\epsilon}_t$ is an (3×1) unobservable zero mean white noise vector process. The covariance of the vector of reduced-form residuals $\boldsymbol{\epsilon}_t$ is denoted as Σ. I choose $p = 3$ to estimate the model in equation (46) based on AIC criteria.

5.2.2 Granger-causality

Table 12: Granger Causality between merger arbitrage spread and bid-ask spread

This table reports the results of the linear Granger-causality test. A_t is arbitrage spread at time t calculated as $p_{2,t} - \gamma p_{1,t} - C$ where $p_{1,t}$ is the price of bidder stock , $p_{2,t}$ is the price of target stock at time t, γ is the stock exchange ratio per share, and C is the per share cash amount paid to target shareholders. $s_{1,t}$ and $s_{2,t}$ is the daily quoted half spread of bidder and target at time t respectively. Each test is conducted by deal by deal. The reported results are percentage of the significance on overall subsamples. The significance of test in each deal is tested at 5% level.

Method Of Payment	Total Observations	Percent Sig $A_t \rightarrow s_{1,t}$	Percent Sig $A_t \rightarrow s_{2,t}$	Percent Sig $s_{1,t} \rightarrow A_t$	Percent Sig $s_{2,t} \rightarrow A_t$
Cash	725	33.24	50.34	13.52	28.14
Stock(FxVal)	64	31.25	43.75	6.25	21.88
Mix(FxVal)	63	55.56	50.79	9.52	23.81
Stock(FxRatio)	995	50.35	57.19	19.50	37.99
Mix(FxRatio)	262	54.58	59.16	22.52	34.73
Collar	315	45.08	50.48	11.75	28.89
Mixed Collar	94	47.87	57.45	12.77	19.15

In this section, I investigate whether each spread for the merger firms is useful in forecasting merger arbitrage or vice versa by using the Granger-causality tests. The Granger-causality tests examine whether the lagged value of one variable helps to predict another variable. If an independent variable does not help to predict the dependent variable, then all of the coefficients of the lagged independent variables are jointly zero. Table 12 summarizes the Granger-causality test results for the VAR model in equation (46). For each deal, I examine whether one variable Granger-causes another variable by testing whether the relevant sets of coefficients are zero. I then count the number of significant samples and calculate the significant percentage in each set of subsamples based on the payment method.

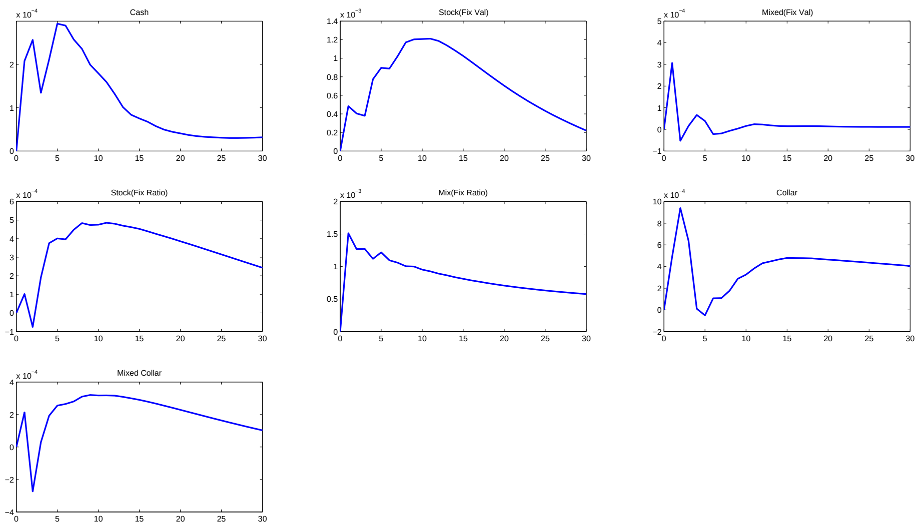
The results indicate that the merger arbitrage spreads appear to help predict a bid-ask spread for each stock in approximately half of the samples as shown in the third and fourth column of Table 12. The results are stronger in the target stocks than in the bidder stocks. Moreover, the results are much stronger in stock swap mergers than in cash mergers. These results imply that both the bidder's and the target's bid-ask spread are necessary to predict the merger arbitrage spreads in approximately half of the sample. The third and fourth column presents the statistics regarding whether each spread helps to predict the merger arbitrage spreads. The results are also stronger in the target stocks, and stock-swap deals have more significant results.

5.2.3 Impulse response function analysis

A simple way of capturing the net effects for all the coefficients in the VAR analysis is to form identified impulse response functions. The VAR model in equation (46) can be written in the following VMA(∞) manner:

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=1}^{\infty} \Theta_i \boldsymbol{\epsilon}_{t-i} \tag{47}$$

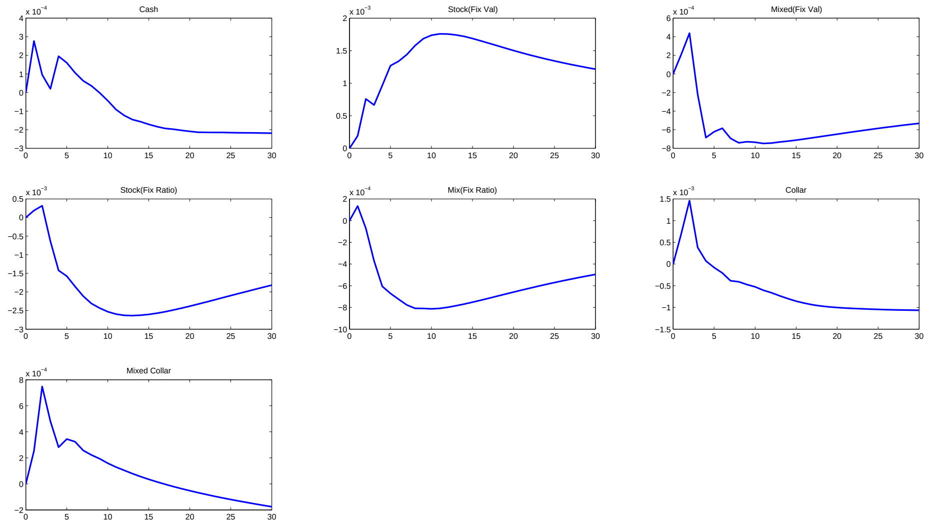
Figure 7: The impulse response functions of merger arbitrage to bid-ask spread of bidder



where Θ_i is the unidentified impulse response function which has the interpretation $\Theta_s = \partial \mathbf{y}_{t+s} / \partial \boldsymbol{\epsilon}'_t$. That is, the row i and column j element of Θ_s identifies the consequences of a one-unit increase in the jth variable's innovation at time t for the value of the ith variable at time $t + s$, holding all other innovations at all dates constant. Because the bid-ask spread shock is correlated with the arbitrage spread shock, it is uncertain whether the response is the response of the arbitrage spread to the bid-ask spread, or to a technology shock that happens to occur at the same time as the bid-ask spread shock. Therefore, the identified impulse response function C_i satisfying $C_i \boldsymbol{\nu}_{t-i} = \Theta_i \boldsymbol{\epsilon}_{t-i}$ where $E[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_t] = PP' = \Sigma$ and $\boldsymbol{\nu}_{t-i} = P^{-1} \boldsymbol{\epsilon}_{t-i}$ with $E[\boldsymbol{\nu}_t \boldsymbol{\nu}'_t] = I_3$ for each i must be calculated.

Figures 7 and 8 depict the responses of a merger arbitrage spread to a unit innovation in the bid-ask spread for each stock up to horizon 30 by method of payments. First, the immediate shock in the bid-ask spread follows the increase in the merger arbitrage spread. Second, the unexpected rise in the bid-ask spread of bidder stocks

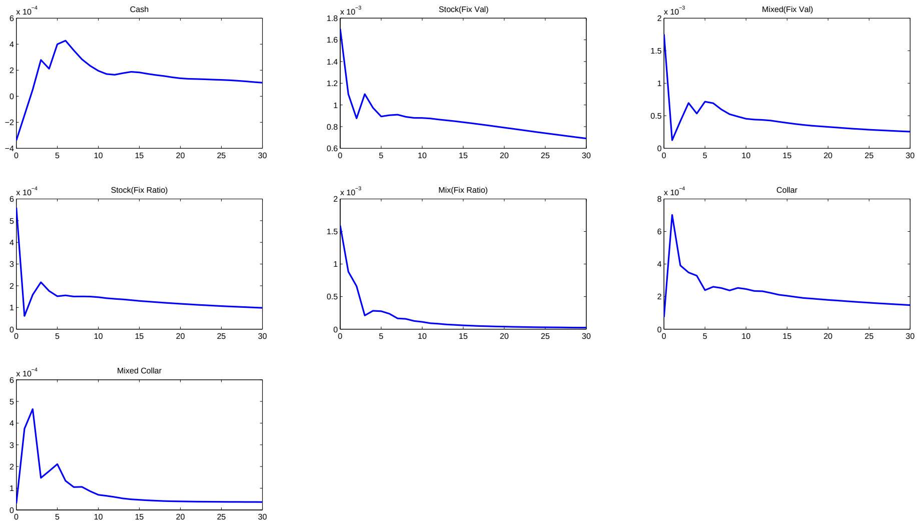
Figure 8: The impulse response functions of merger arbitrage to bid-ask spread of target



decays over time for the fixed value deals (first rows), and is associated with a persistent increase in the arbitrage spread for the floating value deals. The positive effect of the target bid-ask spread on the merger arbitrage spread appears to be stronger than that of the bidder bid-ask spread.

Figures 9 and 10 depict the responses of the bid-ask spread of each stock to a unit shock in a merger arbitrage spread up to horizon 30 by method of payments. The impulse response function of the bid-ask spreads to the merger arbitrage spreads has patterns consistent with the results of the Granger-causality test. The increases in the merger arbitrage spread have a persistent and positive effect on both the bidder and the target bid-ask spread. Roll et al. (2007) argue that the increase in the arbitrage spread induces an increase in incoming orders and eventually widens the bid-ask spread. The effect of the impulse responses of the bid-ask spreads on the merger arbitrage spread is stronger in target stocks than in bidder stocks.

Figure 9: The impulse response functions of bid-ask spread of bidder to merger arbitrage

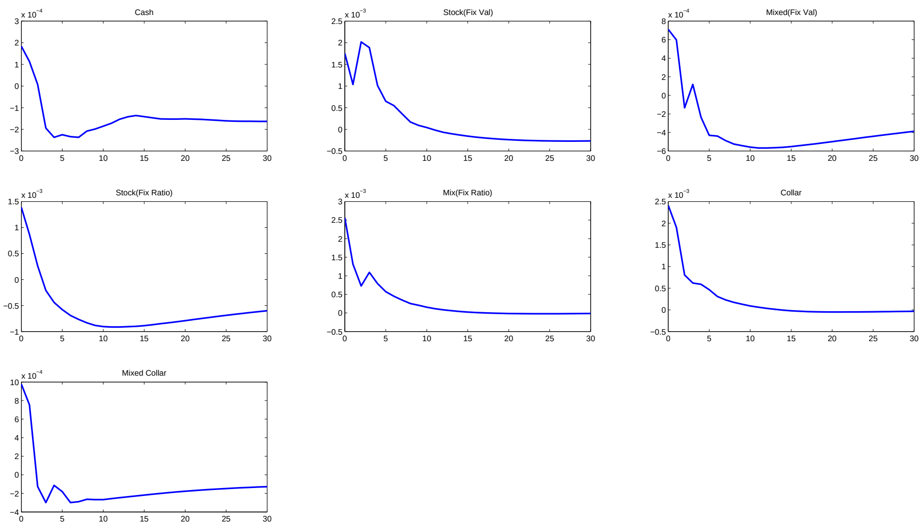


6 Price discovery and information share

6.1 Cointegration and vector error correction model(VECM)

In this section, I examine the price discovery between two stocks during the merger period. After the merger, it is typical for the bidder stock company to exist as a combination of the target and the original bidder companies. As a result, the current bidder stock price should reflect not only the new information related to the original bidder company but also the new information related to the target company. When a merger deal completes successfully, the target shareholders receive the bidder stock in the case of a stock swap merger. In this case, the target stock price will be affected by the new information of the target stock itself as well as the information of the bidder stock. The question is how the new information of the target stock is transmitted to the bidder stock or vice versa. During the merger period, the target stock price is likely to be cointegrated with the bidder stock because the current target stock price is linked to the bidder stock price based on the exchange ratio.

Figure 10: The impulse response functions of bid-ask spread of target to merger arbitrage



I use the vector error correction model to accommodate this cointegrating system and can identify the price discovery process between the bidder and the target stock. Suppose a trade direction follows an autoregressive process, that is $\mathbf{q}_t = \Upsilon(L)\mathbf{v}_t$ where $\Upsilon(L)$ is some autoregressive function. One possible specification for a trade direction is assumed to follow the $AR(1)$ process of $q_{i,t} = \eta_i q_{i,t-1} + v_{i,t}$ to admit the serial correlation in trade flows as in Huang and Stoll (1997) and Madhavan et al. (1997). Then, the autoregressive coefficient η_i is related to the reversal probability [9].

Assuming that a trade direction follows some autoregressive process, the price

[9] Assuming that $q_{i,t} = \pm 1$ equally likely, then η_i is equivalent to $1 - 2\phi_i$. Suppose that $P(q_{i,t} = +1 | q_{i,t-1} = -1) = P(q_{i,t} = -1 | q_{i,t-1} = +1) = \phi_i$ and $P(q_{i,t} = +1 | q_{i,t-1} = +1) = P(q_{i,t} = -1 | q_{i,t-1} = -1) = 1 - \phi_i$ where ϕ_i is the reversal probability of a trade flow i. Since the reversal probability ϕ_i is less than one and greater than zero, this condition satisfies the stationarity assumption of $AR(1)$ process: $0 < \phi_i < 1$ implies $|\eta_i| < 1$. If $\eta_i = \frac{1}{2}$, then η_i will be equal to zero. Moreover, there is a correlation between two order flows, i.e., $E[q_{1,t}q_{2,t}] = \rho$. The innovation $v_{i,t}$ has mean zero with variance $E[v_{i,t}^2] = 1 - \eta_i^2$, and $E[v_{1,t}v_{2,t}] = (1 - \eta_1\eta_2)\rho$.

system in equation (45) can be written in the following $VMA(\infty)$ manner:

$$\begin{aligned}\Delta\mathbf{p}_t &= B(L)\mathbf{u}_t + [A(L)\Upsilon(L) + C(L)]\mathbf{v}_t \\ &\equiv \Psi(L)\boldsymbol{\epsilon}_t \quad \text{where } \Psi(L) = \sum_{j=0}^{\infty}\Psi_j L^j \text{ and } \Psi_0 = I_2 \end{aligned} \tag{48}$$

and new innovation vector $\boldsymbol{\epsilon}_t$ is uncorrelated white-noise with $E[\boldsymbol{\epsilon}_t] = 0$ and $E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'] = \Omega_\epsilon$. In general, equation (48) can be estimated from some appropriate $VAR(p)$ model . But the polynomial $\Psi(L)$ may not be invertible when $L = 1$, and no finite order vector autoregressive model could describe $\Delta\mathbf{p}_t$. These problems can be resolved by employing the vector error correction model proposed by Engle and Granger (1987) and used in the microstructure application by Hasbrouck (1995).

Under the price structure defined in the previous section, the price vector $\mathbf{p}_t$ and the difference in the price vector $\boldsymbol{\Delta}\mathbf{p}_t$ can be written as

$$\mathbf{p}_t = \Psi(1)\sum_{s=0}^{t}\boldsymbol{\epsilon}_s + \Psi^*(L)\boldsymbol{\epsilon}_t \tag{49}$$

$$\Delta\mathbf{p}_t = \Psi(1)\boldsymbol{\epsilon}_t + (1-L)\Psi^*(L)\boldsymbol{\epsilon}_t, \tag{50}$$

where $\Psi^*(L)\boldsymbol{\epsilon}_t$ is covariance stationary and $(1-L)\Psi^*(L)\boldsymbol{\epsilon}_t$ is a stationary, noninvertible moving average. Now, consider the merger arbitrage strategy $A_t = p_{2,t} - \gamma p_{1,t} - C = \alpha'\mathbf{p}_t - C$ where $\alpha = [-\gamma \quad 1]'$ is the cointegrating vector. Premultiplying equation (49) by α' and rearranging the equations result in

$$z_t = \alpha'\Psi(1)\sum_{s=0}^{t}\boldsymbol{\epsilon}_s + \alpha'\Psi^*(L)\boldsymbol{\epsilon}_t \tag{51}$$

where $z_t = \alpha'\mathbf{p}_t \equiv A_t + C$ which implies that it must $\alpha'\Psi(1) = 0$ where $\Psi(1) = I_2 + \Psi_1 + \Psi_2 + \cdots$ for the requirement of $\alpha'\mathbf{p}_t$ to be stationary. Thus, one possible long-run impact matrix can be written as $\Psi(1) = \alpha_\perp\psi$ where $\psi = (\psi_1, \psi_2)$ denote the common row vector in $\Psi(1)$ and $\alpha_\perp$ satisfies $\alpha'\alpha_\perp = 0$. Then, the price system in equation (49) becomes

$$\mathbf{p}_t = \alpha_\perp \psi \sum_{s=0}^{t} \boldsymbol{\epsilon}_s + \Psi^*(L)\boldsymbol{\epsilon}_t \tag{52}$$

Hasbrouck (1995) argues that the term $\psi\boldsymbol{\epsilon}_t$ represents the random-walk component impounded permanently in the price due to new information. To derive the vector error correction form, suppose that the price level $\mathbf{p}_t$ can be represented as a non-stationary pth-order autoregression:

$$\Phi(L)\mathbf{p}_t = \boldsymbol{\epsilon}_t \tag{53}$$

where $\Phi(L) = I_2 - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p$. It is easy to show that $\Phi(1)\Psi(1) = 0$, thus it follows that there exists a (2×1) vector β such that $\Phi(1) = \beta\alpha'$. Then, by the Granger representation theorem(Engle and Granger (1987)) there exist (2×2) matrices $\Gamma_1, \Gamma_2, \cdots, \Gamma_{p-1}$ such that

$$\Delta\mathbf{p}_t = \Gamma_1 \Delta\mathbf{p}_{t-1} + \Gamma_2 \Delta\mathbf{p}_{t-2} + \cdots + \Gamma_{p-1}\Delta\mathbf{p}_{t-p+1} - \beta(\alpha'\mathbf{p}_{t-1} - E\left[\alpha'\mathbf{p}_{t-1}\right]) + \boldsymbol{\epsilon}_t \tag{54}$$

where $\Gamma_s = -[\Phi_{s+1} + \Phi_{s+2} + \cdots + \Phi_p]$ for $s = 1, 2, \cdots, p-1$ and β is error correction vector.

I need to examine whether there exists a cointegration relation between a bidder stock and a target stock to justify the vector error correction model specification in equation (54). I employ the methodologies in Johansen (1991) and Stock and Watson (1988) to test the null hypothesis that two time series are not cointegrated.

Table 13 shows the results of the cointegration tests between two stocks. After performing the cointegration tests for each deal, I count the number of significant cases at the 5% level. Both test results reveal similar patterns across the method of payments. First, the successful deals appear to be more cointegrated than the failed deals, suggesting that both stock prices efficiently reflect the relevant information.

Table 13: Cointegration Test results

This table shows the cointegration test results. Two tests are performed: Stock-Watson common trend test and Johansen cointegration test. Each test is conducted by deal by deal. The reported results are percentage of the significance on overall subsamples.

Method	Total Observations		Stock-Watson Test(%)		Johansen Test(%)	
Of Payment	Fail	Success	Fail	Success	Fail	Success
Cash	61	664	26.23	52.11	27.87	58.58
Stock(FxVal)	8	56	25.00	44.64	62.50	50.00
Mix(FxVal)	9	54	22.22	46.30	55.56	51.85
Stock(FxRatio)	86	909	59.30	90.21	63.95	93.62
Mix(FxRatio)	17	245	64.71	81.22	70.59	85.71
Collar	10	305	40.00	65.25	60.00	74.10
Mixed Collar	10	84	20.00	54.76	40.00	61.90
Total	201	2317	36.78	62.07	54.35	67.97

Second, both test statistics show that there is a substantial cointegration relation between two stocks when the merger parties consider the fixed exchange ratio stock deals. Among the successful deals, approximately 90% of fixed exchanged stock deals show that there exists a strong cointegration relation between two stocks. There is lower cointegration relation in the case of fixed value deals, but approximately half of the sample appears to be cointegrated. Although the fixed valued deal with stock payment (Stock(FxVal)) is the same structure as the pure cash payment, there is a stronger cointegration relation in stock deals than in pure cash deals.

Overall, over half of the successful deals are shown to be cointegrated, while there are few cointegrated relations in failed deals. Thus, the vector error correction specification may be inappropriate for the failed deals. However, this model is well suited for the successful deals. In the next section, I analyze price discovery and the information share using the vector error correction model discussed in this section.

6.2 Price discovery and information share

Hasbrouck (1995) introduces the information share measure to explain where the price discovery occurs. The information share of one security can be defined as the proportional contribution of the security's innovations to the common innovations. If the price innovations are correlated across securities, Hasbrouck (1995) suggests using the Cholesky factorization of $\Omega_\epsilon = FF'$ to remove the contemporaneous correlation. Then, the information shares are provided by

$$S_j = \frac{([\psi F]_j)^2}{\psi \Omega_\epsilon \psi'} \tag{55}$$

where $[\psi F]_j$ is the jth element of the matrix ψF and F is the lower triangular matrix. According to Hasbrouck (1995), this information share is a relative measure allocating information to the different securities and simply measures the informational transmission speeds in the process of price adjustment.

It is natural that the target stock price should reflect the information of its own shares as well as the information of the bidder stocks when there are stock exchange mergers. Regardless of the payment method, the bidder stock price should reflect the new information for the target shares because all of the information related target firms could affect the acquirer firm values and the shareholder values of the bidder firms. For example, the target stock price does not depend on the bidder stock price when the method of payment is pure cash. However, paying cash to the target firm shareholders forces the bidder firms to use internal funds or to finance additional capital to meet the deal conditions. Moreover, unexpected changes in the target shares may affect the future cash flows of the target firms, and ultimately affect the future cash flows of the bidder firms. In this sense, the information share of the efficient stock price is similar to the ratio of each firm's value to the total firm value, which is the sum of the bidder stock and the target stock.

Suppose each party derives the deal condition to reflect each firm value at its best; then, the stock price of the bidder after the merger announcement should reflect each firm's value proportionally well through the trading activities. I calculate the

Table 14: Information share of efficient price and the ratio of firm value

This table provides the estimation results of information shares and the ratio of firm value. The information share is calculated from the fitted residuals of vector error correction model. The ratios of each firm are calculate to divide each firm value by the sum of target and bidder firm values. Each firm equity value is the average of the product of the market closing price and the shares outstanding 30days prior to the merger announcement.

Method	Total	Ratio		Std Dev (Ratio)	
Of Payment	Observations	Bidder	Target	Bidder	Target
Panel A. Information Share of the efficient price					
Cash	725	87.20	12.80	0.196	0.196
Stock(FxVal)	64	80.46	19.54	0.259	0.259
Mix(FxVal)	63	76.97	23.03	0.270	0.270
Stock(FxRatio)	995	89.28	10.72	0.158	0.158
Mix(FxRatio)	262	71.30	28.70	0.317	0.317
Collar	315	89.81	10.19	0.168	0.168
Mixed Collar	94	78.12	21.88	0.267	0.267
Panel B. The ratio of firm value to total firm values					
Cash	725	89.10	10.90	0.133	0.133
Stock(FxVal)	64	93.82	6.18	0.109	0.109
Mix(FxVal)	63	84.44	15.56	0.167	0.167
Stock(FxRatio)	995	80.65	19.35	0.160	0.160
Mix(FxRatio)	262	78.47	21.53	0.166	0.166
Collar	315	87.72	12.28	0.126	0.126
Mixed Collar	94	80.27	19.73	0.149	0.149

proportion of firm value by dividing each firm's total stock value by the sum of the stock values of the bidder and target firms. I use the -45 days to -30 days prior to the merger announcement to avoid the effect of the target price run-up before the merger announcement.

Table 14 provides the estimated information share of the efficient price by using the vector error correction model and the ratio of firm value. Panel A reports the information share of the efficient price after the announcement and Panel B calculates the ratio of firm value before the announcement. Most estimated numbers for the information share of the efficient price are quite similar to the ratio of firm value except for the fixed value deals using stocks as a method of payment. Although the information share is estimated from the post-merger stock price data, it reflects the information content of the bidder stock. The difference between the information share and the ratio of firm value is the smallest in the pure cash deal. This result may be due to the minimal asymmetric information between the bidder and target shareholders. While the difference between the two measures in the bidder share is 7.71% in the fixed exchange deals, the difference is reduced to 0.22% when merger parties use the collar deals.

7 Conclusion

In this paper, I analyze the activities of arbitrageurs playing the role of market maker in the takeover markets. The cost function for the market makers is reflected in the quoted bid-ask spread. Merger arbitrageurs typically use limit orders to exploit the arbitrage opportunity. The merger arbitrage strategy is promising, as reported in the previous literature.

I show that the merger arbitrage spread is closely related to the bid-ask spread. The existence of merger arbitrage is due to transaction costs, deal failure risk, and the limits of arbitrage. These three arguments are similar to the bid-ask components in market microstructure theory: order handling costs, adverse information costs, and inventory holding costs. Using the spread decomposition model in Huang and Stoll (1997), I provide evidence that arbitrageurs play the market maker role during the mergers and acquisition period because their inventory costs for target shares increase after the merger announcement. Moreover, the information components of both stocks decrease after the announcement.

Next, I investigate the relation between the merger arbitrage spread and the bid-ask spread by both cross sectional regression and time series analysis. I find that there is a positive and significant relation between the merger arbitrage spread and the bid-ask spread. This result is also confirmed in a time series vector autoregressive analysis. The impulse response function analysis shows that the effect of one variable on another appears to be persistent.

Finally, the cointegration analysis suggests that there exists a cointegration relation between the bidder stock and the target stock. The cointegration relation is much stronger in fixed exchange ratio deals than in fixed value deals. In addition, I calculate the information share to test whether the information of each stock is appropriately reflected in the efficient price series using the vector error correction model. If the information of each stock transmits to the efficient price effectively, the information share should be similar to the ratio of each stock to the sum of each stock's market value. The results are consistent with the hypothesis that the market is efficient even though some arbitrage opportunity exists. Therefore, the arbitrage opportunity is the reasonable compensation to the arbitrageurs who provide liquidity to the existing target shareholders.

Appendix

A Inventory holding costs and the limits of arbitrage

Suppose there exist only inventory costs, i.e. $\beta_2 = 1$, but no information contents: $u_{i,t} = v_{i,t} = 0$. Then, the price of target stock becomes

$$p_{2,t} = C + \gamma m_{1,t-1} + \frac{s_2}{2}(q_{2,t} + q_{2,t-1}) \tag{A.1}$$

Now, suppose the offer price, $C + \gamma m_{1,t-1}$, can correctly reflect the expected payoff of the target price so that $\mu = E[p_{2,t}] = C + \gamma m_{1,t-1}$ because $E[q_{2,t}] = 0$. The buying price from the arbitrageurs or the selling price from the existing target shareholders ($q_{2,t} = q_{2,t-1} = -1$) becomes

$$p^T = \mu - s_2 \tag{A.2}$$

under the assumption that trades occurs at bid price. Further, assume that arbitrageurs have the mean-variance utility function

$$U(\tilde{W}) = E[\tilde{W}] - \frac{z_A}{2} Var(\tilde{W})$$

Stoll (1978) argues that the dealer must be compensated to offset the expected utility loss by deviating from his initial portfolio. Let Q denote the dollar value of a transaction in target stock having return R with $E[R] = \mu_R$ and $Var[R] = \sigma_R^2$. Let R_e be the efficient portfolio return with $E[R_e] = \mu_e$ and $Var[R_e] = \sigma_e^2$. Let W_1 and W_2 be the terminal wealth of the initial portfolio and the new portfolio after the transaction. Suppose that the dealer has initially the optimal portfolio, i.e., the dollar value of stocks in trading account is zero. Then,

$$W_1 = W_0\left[1 + kR_e + (1-k)R_f\right] = W_0(1 + R^*) \tag{A.3}$$

$$W_2 = W_0(1 + R^*) + (1 + R)Q - (1 + R_f)(Q - C) \tag{A.4}$$

where k is the optimal fraction of the dealer's wealth in optimal portfolio R_e and R_f is the risk-free rate. First, the optimal fraction k can be obtained by solving

$$\frac{\partial EU[W_1]}{\partial k} = W_0(\mu_e - R_f) - kW_0^2 z_A \sigma_e^2 \equiv 0 \Rightarrow k^* = \frac{\mu_e - R_f}{W_0 z_A \sigma_e^2}$$

Next, the dollar cost to the dealer must $EU[W_1] \equiv EU[W_2]$, that is

$$C = \frac{\frac{1}{2} z_A \sigma_R^2 Q^2 - Q\left[(R - R_f) - (R_e - R_f)\frac{\sigma_{ie}}{\sigma_e^2}\right]}{1 + R_f} \tag{A.5}$$

The capital asset pricing model states $R - R_f = (R_e - R_f)\frac{\sigma_{ie}}{\sigma_e^2}$ and assume the zero risk free rate($R_f = 0$), we have

$$C = \frac{1}{2} z_A \sigma_R^2 Q^2 \tag{A.6}$$

and the the unit cost $c = C/Q$ will be $\frac{1}{2}z_A\sigma_R^2 Q$. Then, the proportional bid-ask spread can be expressed as

$$2c = \frac{p^{T,a} - p^T}{p^T} + \frac{p^T - p^{T,b}}{p^T} = \frac{s_2}{p^T} = z_A\sigma_R^2|Q| \tag{A.7}$$

where $p^{T,b}$ and $p^{T,b}$ are the bid and ask price of the target share. Here, since $\sigma_R^2 = \frac{\sigma^2}{(p^T)^2}$ and $Q = \frac{X}{A}p^T$, the proportional spread can be written as

$$\frac{s_2}{p^T} = z_A \frac{\sigma^2}{(p^T)^2}\frac{X}{A}p^T = \frac{X}{A}z_A\sigma^2\frac{1}{p^T}$$

So, equation (A.2) can be written

$$p^T = \mu - s_2 = \mu - \frac{X}{A}z_A\sigma^2$$

which is equivalent to the equation (41)

B State space representation and Kalman filter

The state-space representation of the dynamics of $\Delta\mathbf{p}$ is given by the following system of equations:

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}\boldsymbol{\xi}_t + \mathbf{e}_{t+1} \tag{B.1a}$$

$$\Delta\mathbf{p}_t = \mathbf{A'}\mathbf{x}_t + \mathbf{H'}\boldsymbol{\xi}_t \tag{B.1b}$$

where $\mathbf{F}, \mathbf{A'}$, and $\mathbf{H'}$ are matrices of parameters of dimension $(r \times r), (2 \times k)$, and $(2 \times r)$, respectively, and $\mathbf{x}_t$ is a $(k \times 1)$ vector of exogenous variables. Equation (B.1a) is known as the state equation, and (B.1b) is known as the observation equation. The $(r \times 1)$ vector $\mathbf{e}_t$ is vector white noise:

$$E[\mathbf{e}_t\mathbf{e}_s'] = \begin{cases} \mathbf{Q} & \text{for } t = s \\ \mathbf{0} & \text{otherwise.} \end{cases} \tag{B.2}$$

where $\mathbf{Q}$ is $(r \times r)$ matrix. Assume that $\boldsymbol{\xi}_1$ is uncorrelated with any realizations of $\mathbf{e}_t$: $E[\mathbf{e}_t\boldsymbol{\xi}_1'] = 0$.

Consider the linear projection of $\boldsymbol{\xi}_{t+1}$ on $\mathcal{P}_t$ and a constant:

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \hat{E}[\boldsymbol{\xi}_{t+1}|\mathcal{P}_t]$$

where $\mathcal{P}_t = (\Delta\mathbf{p}_t', \Delta\mathbf{p}_{t-1}', \cdots, \Delta\mathbf{p}_1')'$. The Kalman filter calculates these forecasts recursively, generating, $\hat{\boldsymbol{\xi}}_{1|0}, \hat{\boldsymbol{\xi}}_{2|1}, \cdots, \hat{\boldsymbol{\xi}}_{T|T-1}$ in succession. Associated with each of these forecasts is a mean squared error matrix, represented by the following $(r \times r)$ matrix

$$\mathbf{P}_{t+1|t} = E\left[\left(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t}\right)\left(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t}\right)'\right] \tag{B.3}$$

Given starting values $\hat{\boldsymbol{\xi}}_{1|0}$ and $\mathbf{P}_{1|0}$, the state vector can be written

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{K}_t \left(\Delta\mathbf{p}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1} \right) \tag{B.4}$$

where $\mathbf{K}_t = \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H})^{-1}$ is known as the gain matrix and the mean squared error matrix is

$$\mathbf{P}_{t+1|t} = \mathbf{F}[\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H})^{-1}\mathbf{H}'\mathbf{P}_{t|t-1}]\mathbf{F}' + \mathbf{Q} \tag{B.5}$$

Finally, all the parameters can be estimated using the maximum likelihood estimation where the density function for $\Delta\mathbf{p}_t$ is

$$\begin{aligned} f(\Delta\mathbf{p}_t|\mathcal{P}_{t-1}) &= (2\pi)^{-1}\left|\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H}\right|^{-1/2} \\ &\times \exp\left\{-\frac{1}{2}\left(\Delta\mathbf{p}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}\right)'\left(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H}\right)^{-1}\left(\Delta\mathbf{p}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\hat{\boldsymbol{\xi}}_{t|t-1}\right)\right\} \end{aligned} \tag{B.6}$$

REFERENCES

Baker, M., Savasoglu, S., 2002. Limited arbitrage in mergers and acquisitions. Journal of Financial Economics 64, 91–115.

Conrad, J., Niden, C. M., 1992. Order flow, trading costs and corporate acquisition announcements. Financial Management, 22–31.

Copeland, T., Galai, D., 1983. Information effects on the bid-ask spread. Journal of Finance 38, 1457–1469.

Easley, D., O'hara, M., 1987. Price, trade size, and information in securities markets. Journal of Financial Economics 19, 69–90.

Engle, R., Granger, C., 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica 55 (2), 251–276.

Foster, F. D., Viswanathan, S., 1995. Trading costs for target firms in corporate takeovers. Advances in Financial Economics 1, 37–57.

Glosten, L. R., Harris, L., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123–142.

Harford, J., 2005. What drives merger waves? Journal of Financial Economics 77, 529–560.

Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. The Journal of Finance 50 (4), 1175–1199.

Houston, J. F., Ryngaert, M. D., 1997. Equity issuance and adverse selection: A direct test using conditional stock offers. Journal of Finance 52, 197–219.

Huang, R. D., Stoll, H. R., 1994. Market microstructure and stock return predictions. Review of Financial Studies 7, 179–213.

Huang, R. D., Stoll, H. R., 1997. The components of the bid-ask spread: a general approach. Review of Financial Studies 10, 995–1034.

Jennings, R., 1994. Intraday changes in target firms' share price and bid-ask quotes around takeover announcements. Journal of Financial Research 17, 255.

Jindra, J., Walkling, R. A., 2004. Speculation spreads and the market pricing of proposed acquisitions. Journal of Corporate Finance 10, 495–526.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. Econometrica 59 (6), 1551–1580.

Kumar, P., Seppi, D. J., 1994. Information and index arbitrage. The Journal of Business 67 (4), 481–509.

Larcker, D. F., Lys, T., 1987. An empirical analysis of the incentives to engage in costly information acquisition: The case of risk arbitrage. Journal of Financial Economics 18, 111–126.

Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–754.

Lipson, M. L., Mortal, S., 2007. Liquidity and firm characteristics: evidence from mergers and acquisitions. Journal of Financial Markets 10, 342–361.

Madhavan, A., Richardson, M., Roomans, M., 1997. Why do security prices change? a transaction-level analysis of nyse stocks. Review of Financial Studies 10, 1035–1064.

Mitchell, M., Pulvino, T., 2001. Characteristics of risk and return in risk arbitrage. The Journal of Finance 56, 2135–2175.

Moore, K. M., Lai, G. C., Oppenheimer, H. R., 2006. The behavior of risk arbitrageurs in mergers and acquisitions. The Journal of Alternative Investments 9, 19–27.

Roll, R., Schwartz, E., Subrahmanyam, A., 2007. Liquidity and the law of one price: The case of the futures-cash basis. The Journal of Finance 62, 2201–2234.

Shleifer, A., Vishny, R. W., 1997. The limits of arbitrage. Journal of Finance 52, 35–55.

Stock, J., Watson, M., 1988. Testing for common trends. Journal of the American Statistical Association 83 (404), 1097–1107.

Stoll, H., 1978. The supply of dealer services in securities markets. Journal of Finance 33, 1133–1151.

CHAPTER III

Liquidity risk and Exchange-traded-fund returns, variances, and tracking errors

1 Introduction

Since its introduction in 1993, the US exchange-traded fund (ETF) market has grown explosively. ETFs are designed to provide an alternative investment opportunity for particular markets, countries, or sectors by following a specific representative index. With index-based ETFs, investors can benefit from access to foreign markets or different asset categories with low costs. The fundamental risk of ETFs is the market risk associated with the underlying index. In addition, each ETF also has its own idiosyncratic risk, i.e. tracking error risk, after removing the market risk. ETFs trade on the stock exchange but their shares are created and redeemed on the primary market. This structure results in the existence of two prices for a single asset; one is market ETF prices determined on stock exchanges and the other is the fund's net-asset value (NAV) calculated based on the value of underlying securities. Intuitively, no arbitrage condition implies that the daily market closing price and the daily closing net asset value of an ETF must be same. However, various factors can widen the gaps between the NAV prices and the ETF prices. Among various factors, this paper particularly focuses on the role of illiquidity in the ETF market.

ETFs are fundamentally same as the open-ended mutual funds[1]. They are structured, managed, and regulated just like traditional mutual funds [2]. One difference

[1]SEC defines ETFs as "Exchange-traded funds, or ETFs, are investment companies that are legally classified as open-end companies or Unit Investment Trusts (UITs), ...". For more details, see http://www.sec.gov/answers/etf.htm

[2]There are similar products with ETFs. These products include exchanged-traded-notes (ETN) and exchanged-traded-commodity (ETC). The exchanged-traded-note is a senior unsecured debt obligation designed to track the total return of an underlying index or benchmark. The ETNs are exposed both to the market risk of the linked indexes and the credit risk of the issuer. The exchange-traded-commodity (ETC) is similar to ETFs, but it holds physical commodities or currencies. Both ETNs and ETCs are not registered under the Investment company act of 1940 but are regulated

from conventional mutual funds is that ETFs are traded continuously in the stock markets like closed-end funds. That is, ETFs are designed to combine the creation and redemption process of open-end funds with the continuous exchange trading of the closed-end funds. The creation/redemption process in ETFs is the crucial mechanism that enables ETF prices to stay close to their NAVs [3]. The ETF price deviation from its NAV can be eliminated by the arbitrage activity of authorized participants who have the responsibility of creating/redeeming ETF shares or constructing the underlying ETF portfolios. However, this arbitrage mechanism can be limited if either ETFs or the underlying securities are illiquid. The lack of liquidity in the underlying securities may result in a tracking error of the NAV on the index. Illiquid ETFs may also have a mispricing problem with respect to their NAVs. Thus, illiquidity and ETF tracking errors are quite related. Moreover, illiquidity risk is another risk factor to determine the asset returns.

There are numerous studies investigating the effect of liquidity on asset returns and suggesting that systemic liquidity risk is priced in the asset return[4] . Acharya and Pedersen (2005) develop a liquidity adjusted capital asset pricing model(LCAPM) and find that the individual asset return is affected by the liquidity risk. They argue that their LCAPM explains the asset return better than the standard CAPM. Similarly, Pastor and Stambaugh (2003) empirically find that the individual stock return is affected by the aggregate market liquidity, which is a cross-sectional average of the individual return reversal estimates. In addition to studying the relation between asset returns and liquidity in the US equity market, there are also numerous studies investigating the effect of liquidity on asset returns in different markets or assets, for

under Securities Act of 1933. In this paper, I exclude both ETNs and ETCs for the consistent analysis.

[3]The closed-end funds are also listed in the major stock exchanges and are traded like common stocks. Unlike ETFs, however, the closed-end funds typically trade at a discount to the portfolio value. This is called "closed-end fund discount puzzle". See Pontiff (1996) for more details.

[4]For example, see Amihud and Mendelson (1986), Huang and Stoll (1994), Brennan and Subrahmanyam (1996), Chordia et al. (2001), Amihud (2002)

example, emerging markets (Bekaert et al. (2007)) and global markets(Lee (2011)), hedge funds (Getmansky et al. (2004), Sadka (2010)), IPO markets (Eckbo and Norli (2005)), and closed-end-funds(Cherkes et al. (2009)).

However, there are only few papers that study the effect of liquidity on the ETF return and variance. The related existing studies include the ETF pricing for the Flash Crash of May 6, 2010 (Borkovec et al. (2010), Madhavan (2011)), the interaction between the ETF market and the underlying securities market (Cespa and Foucault (2012), Ben-David et al. (2011)), and whether ETFs are priced efficiently with respect to the NAV or the underlying index (Engle and Sarkar (2006), Elton et al. (2002)). Regarding the lack of liquidity in the market, Borkovec et al. (2010) report that a sharp increase in the bid-ask the spread leads to a failure in ETF price discovery during the Flash Crash. Studies have also investigated the interaction between the ETF and its underlying securities. For example, Cespa and Foucault (2012) argue that the lack of liquidity in ETFs may lead to an increase in the uncertainty of the underlying securities, which results in a decrease in the liquidity of the corresponding ETFs. However, there are no existing studies that cover the effects of liquidity on ETF returns and tracking errors comprehensively.

In the first part of this paper, I begin by analyzing the relations between ETF tracking errors and market illiquidity. I present evidence that tracking errors and ETF illiquidity are positively related [5]. That is, the cross-sectional analysis shows that the level of illiquidity is positively related to the ETF-NAV or ETF-index tracking errors. Moreover, equity-type or domestic ETFs tend to have the smaller tracking errors. US equity markets are the most liquid markets in the world. Thus, the negative relation between tracking errors and such funds can be consistent with the fact that illiquidity and tracking errors are positively related.

[5] ETF tracking errors are similar to the relation between futures and underlying asset prices. Roll et al. (2007) study the interactions between illiquidity and the futures basis in the S&P 500 futures markets. They conclude that the contemporaneous shocks to the futures basis and bid-ask spreads are positively correlated.

Next, I investigate whether liquidity shocks in ETFs are priced based on the LCAPM of Acharya and Pedersen (2005). Using every ETF ever listed in the US markets, I first construct 10 portfolios sorted on liquidity and 10 portfolios sorted on tracking errrors. The sorted portfolios provide evidence that illiquidity is positively related to ETF returns and to ETF tracking errors. In addition, the portfolios show that the illiquidity and the tracking errors of ETFs are shown to be persistent over time, suggesting that there exist positive common shocks affecting both ETF illiquidity and tracking errors. I also estimate the portfolio betas in Acharya and Pedersen (2005) to investigate whether there exist any systemic risk factors associated with illiquidity. The results of estimated betas show that illiquid ETFs tend to have large absolute liquidity betas and positive liquidity risk premium. That is, illiquid ETFs tend to be more sensitive to either market liquidity or the market return. Using pre-estimated betas, I also estimate the liquidity premium using GMM method. The annualized return due to the liquidity risk is approximately 0.36%, suggesting that there exists a positive liquidity premium in the US ETF market.

Finally, I investigate whether infrequent trading affects ETF variances with respect to NAV variances. Lo and MacKinlay (1990) develop an econometric model to show that the asset variance increases when the asset is not traded frequently. Extending the Lo and MacKinlay (1990) econometric model to consider the autocorrelation of the index return, I provide evidence that the nontrading probability is positively related to an increase in the ETF variance with respect to the NAV variance. The derived equation shows that the ETF return variance can be expressed as the sum of NAV return variance and positive terms associated with nontrading probability. Moreover, the existence of autocorrelation in the index return could increase the gap between the ETF variance and the NAV variance when the ETFs are illiquid. The panel regression also confirms that nontrading probability and the difference between ETF and NAV return variances is positively related, suggesting that illiquid ETFs have more risk when the ETFs are traded actively in the market.

The bottom line of this paper is that a lack of liquidity is related to the expected return and the variance of ETFs. Illiquid ETFs tend to have large tracking errors with respect to their underlying index or their NAV returns. These findings imply that illiquid ETFs may be vulnerable to a sudden change in the liquidity level such as the Flash Crash on May 6, 2010.

The remainder of the paper is organized as follows. Section 2 provides the data sources, sample construction procedure, and variable constructions. In section 3, I investigate whether the level of illiquidity is related to the ETF tracking errors. In section 4, I provide the estimation results of the LCAPM and the cross-sectional evidence for the liquidity risk in the ETF market. Section 5 derives the closed form of the ETF variance when ETFs are not traded frequently and compares the variance of ETFs with the variance of the NAV. The cross-sectional regression analysis is also provided for whether the nontrading probability is related to the difference between the ETF variance and the NAV variance. Section 6 concludes the paper.

2 ETF data and variables

2.1 Exchange traded fund data

The ETF sample used in this paper includes all of the ETFs that have ever been listed and traded in the major US stock exchanges from 1993 through 2012. The country of domicile for each ETF must be the US at the inception date. The initial data also include all of the delisted ETFs that were traded in the US market during the sample period. All ETF data are extracted from the Bloomberg database. The Bloomberg provides all daily historical prices for ETFs, NAVs, and the underlying indexes as well as detailed information about the ETFs[6]. When the necessary data from Bloomberg are not available, the data are collected from the ETF product webpage when available.

[6]Petajisto (2011) reports that the Bloomberg data cover up to 90% of all ETFs

Table 15: ETF sample construction

This table presents the process of constructing the sample used in this paper. Initial Exchange-traded-fund(ETF) data are extracted from the Bloomberg database for all of the ETFs that have ever been listed and traded in the US from 1993 through 2012.

Description		Number of ETFs
Initial Sample	(1)	1495
ETFs on the BATs	(2)	17
Actively Managed Funds	(3)	57
Underlying Index data is missing		
- Barclays Capital Bond Index	(4)	68
- The combination of commondity prices	(5)	11
- Index level data is not available	(6)	20
NAV data is missing	(7)	6
Price data is missing	(8)	9
Total number of samples deleted ((2) (8))	(9)	188
Final Sample	(1) - (9)	1307

To effectively investigate the effect of liquidity on the ETF returns and variances, actively managed funds are excluded from the sample. Actively managed funds first appeared in 2008 and are managed to achieve excess return on the typical benchmark index by frequently buying or selling assets in the portfolio rather than passively following the index. As a result, actively managed funds are more likely to deviate from a particular underlying index return because their portfolio composition weights change frequently. Because the tracking error in actively managed funds could be caused by the management style, it is not easy to separate the effect of liquidity from the effect of the management style on the return and the variance. Therefore, excluding actively managed funds from the sample is reasonable for an analysis of liquidity's effect on the return and variance. That is, the final sample in this paper only includes index-based ETFs.

In addition, if an ETF has no available information about the traded price, the NAV, or the underlying index during the sample period, then it is excluded from the sample. Table 15 shows the sample construction procedure for this paper. The final sample consists of 1307 US listed exchange traded funds.

Table 16: ETF trends

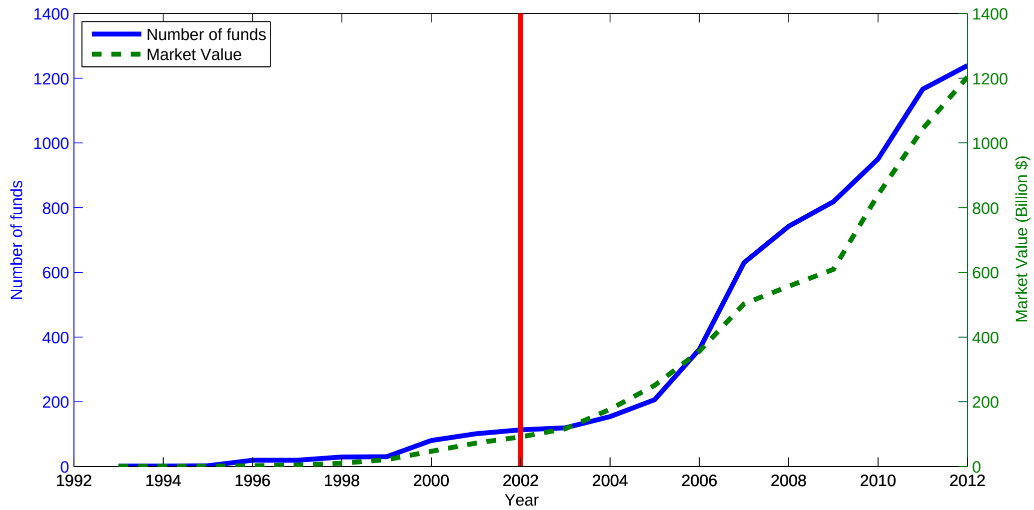
This table reports the annual breakdown of the sample by number of funds incepted, number of funds delisted, number of funds available at the end of year, average market value, average trading volume and average dollar trading volume. The sample includes all the US ETFs which were listed on the US exchange during 1993-2012.

Year	Incept	Delist	N	MV (Bill. $)	Volume (Mill. Shares)	Dollar Volume (Bill. $)
1993	1	0	1	0.26	0.2	0.01
1994	0	0	1	0.47	0.4	0.02
1995	1	0	2	0.67	0.3	0.02
1996	17	0	19	1.67	1.2	0.07
1997	0	0	19	4.01	3.7	0.30
1998	10	0	29	10.01	9.9	0.90
1999	1	0	30	20.63	25.4	2.05
2000	50	0	80	46.42	45.9	4.19
2001	21	0	101	72.07	98.0	5.41
2002	15	3	113	91.31	151.7	7.27
2003	12	6	119	116.41	155.3	8.02
2004	35	0	154	176.38	206.3	11.56
2005	52	0	206	250.49	272.8	16.59
2006	157	1	362	355.95	390.1	24.41
2007	268	0	630	502.33	701.6	54.70
2008	162	50	742	556.80	1449.0	93.68
2009	127	51	818	608.84	1409.9	69.92
2010	180	48	950	839.38	1162.8	68.23
2011	231	15	1166	1043.69	1259.2	77.48
2012	155	82	1239	1203.74	923.2	56.19

Constructing 10 liquidity and tracking error portfolios, I restrict the data to after 2002 for the following two reasons. First, a sufficient number of ETFs is needed to construct 10 portfolios. At the end of 2001, 101 ETFs are listed in the US market, which enables each portfolio to contain more than 10 ETFs each year. Table 16 reports the annual breakdown of the sample by number of funds initiated, number of funds delisted, number of funds available at the end of year, average market value, average trading volume and average dollar trading volume. As seen in Table 16 and Figure 11, the number of funds and trading volume increases sharply after the early 2000s, thus the number of ETFs traded in the US increases to 1440 by the end of 2012. Consequently, each portfolio in 2012 should have more than 100 ETFs. Second, the minimum tick size of the bid-ask spread reduces from 1/16 to 1/100 in

Figure 11: The number of funds and the market value of ETFs

This figure illustrates the number of ETFs and market value of US ETFs at the end of year from 1993 to 2012.



2001. The change in the minimum tick size is related to the exogenous shock to the liquidity. Moreover, figure 11 shows that there appears to be a significant increase in the trading volume of the ETF market after 2002 [7]. The increase in trading volume and the decrease in the bid-ask spread imply an important change in the liquidity measure, so I use the data from 2002.

2.2 Liquidity measure

The daily liquidity of each ETF is measured by using the daily relative effective spread calculated from the trading and quote data of the NYSE TAQ database. The daily relative effective half spread is defined as the ratio of the effective half spread to the trade price. The effective half spread is defined as the difference between the quote

[7] In the middle of 2001, the NYSE began trading three unlisted ETFs(DIA, SPY, and QQQ) which are listed on the American Stock Exchange. Another 27 ETFs are allowed to trade on the NYSE on April 15, 2002. Boehmer and Boehmer (2003) report that these events lead to a large improvement in liquidity due to the competition for order flow among market centers.

midpoint and the corresponding trade price. That is,

$$c_t^i = \frac{1}{n_t^i} \sum_{k=1}^{n_t^i} \frac{|p_{k,t}^i - m_{k,t}^i|}{p_{k,t}^i} \tag{56}$$

where $p_{k,t}$ is the traded price, $m_{k,t}$ is the quote midpoint, and n_t is the number of trades at time k on day t for security i. The relative effective spread is close to the liquidity measure of 'dollar cost per dollar invested' used in Acharya and Pedersen (2005). Their empirical studies use the normalized illiquidity measure by transforming the Amihud Illiquidity measure where the cross-sectional mean and variance are equal to the effective half spread reported in Chalmers and Kadlec (1998). As a result, their liquidity measure is ultimately similar to the relative effective half spread, which can be obtained from the TAQ data directly. One advantage of the use of the illiquidity measure from the TAQ data is that the relative effective spread can be observed on a daily basis. Daily illiquidity measures will provide considerably large data set containing more stock information. Moreover, the daily liquidity measure is suitable for the levered or inversed ETFs because the use of the monthly measure may cause the difference between the monthly realized return and the monthly holding return of levered funds.

2.3 Tracking errors

I use two measures of tracking errors. The first definition is calculated from the regression analysis. This tracking error is the absolute difference between one and the coefficient of the regression of two return series; ETF return vs NAV return, NAV return vs index return, or ETF return vs index return. The second definition is measured by calculating the standard deviation of return difference between two return series. Table 17 provides the summary statistics and cross-sectional correlations of estimated tracking errors for ETFs from the inception date to the end of 2012 or the delisting date. This table exhibits some interesting features. First,

Table 17: Summary statistics and correlations of ETF tracking errors

This table provides the summary statistics and correlations of estimated tracking errors for ETFs from the inception date to the end of 2012 or the delisting date. Two tracking errors are defined. $\theta(Y-X)$ is the tracking error by taking the absolute value of the difference between one and the coefficient of X from regression of Y on X. $\sigma(Y-X)$ is the standard deviation of the return difference between Y and X. The six tracking errors are estimated for each ETF using all daily return. r_t, v_t, and f_t denote the daily ETF, NAV, and index returns, respectively.

Variables	$\sigma(r_t-f_t)$	$\sigma(v_t-f_t)$	$\sigma(r_t-v_t)$	$\theta(r_t-f_t)$	$\theta(v_t-f_t)$	$\theta(r_t-v_t)$
Panel A. Summary Statistics for estimated tracking errors						
Mean	1.200%	0.427%	1.153%	15.884%	4.169%	16.526%
Std. Dev.	1.221%	0.952%	1.002%	17.675%	9.305%	17.548%
Panel B. Tracking error correlations for individual ETFs						
$\sigma(r_t-f_t)$	1.000					
$\sigma(v_t-f_t)$	0.712	1.000				
$\sigma(r_t-v_t)$	0.810	0.317	1.000			
$\theta(r_t-f_t)$	0.436	0.133	0.399	1.000		
$\theta(v_t-f_t)$	0.301	0.524	0.083	0.326	1.000	
$\theta(r_t-v_t)$	0.307	−0.010	0.444	0.805	0.094	1.000

panel A shows that NAV-index tracking errors are smaller than ETF-NAV tracking errors. In addition, ETF-index return tracking errors are similar to the ETF-NAV tracking errors suggesting that ETF-index tracking errors seem to be explained by the ETF-NAV tracking errors. Second, panel B shows that tracking errors from regression are highly correlated with those calculated from standard deviations. Third, the correlations between ETF-NAV and NAV-index tracking errors are lower than other correlation numbers (0.317 for standard deviation tracking errors and 0.094 for regression tracking errors). This result suggests that there are some factors, for e.g., ETF market conditions, that could be related to the ETF-NAV tracking errors but not be associated with NAV-index tracking errors.

3 The effect of illiquidity on ETF tracking errors

3.1 Arbitrage activity of authorized participants

Three observed return series (ETF, NAV, and index returns) must be same on the daily basis due to the ETF structures. However, the market data shows that there exist gaps among these return series. Those return differences can be caused by various factors such as trading activity, product structures, illiquidity of underlying securities, or ETF market conditions. This section investigates whether illiquidity is related to ETF tracking errors by using the panel regression analysis.

Authorized participants in the ETF market play the important role to keep the return series close to each other. The return differences are typically removed away by the arbitrage activity of authorized participants. Authorized participants or market makers keep ETF prices in line with the value of their underlying portfolios by trading both ETFs and underlying securities simultaneously, so called creation-redemption process. For instance, if an ETF price is lower(higher) than its NAV, APs buy(sell) ETF shares and sell(buy) the basket of securities. More precisely, if the current market prices of an ETF become higher than its NAVs, APs buy underlying securities to form a creation unit and deliver it to the ETF provider. After receiving the ETF shares from the ETF issuer, APs sell these ETF shares to the market [8].

The arbitrage activity in the ETF market would be possible when authorized participants are able to trade ETFs or underlying securities immediately and limitlessly. However, APs may get into trouble with constructing the basket of securities or with trading ETFs if underlying securities or ETF markets suffer from the lack of liquidity. Because each ETF has its own way of portfolio construction, the ETF provider could choose the appropriate method to replicate the underlying index return precisely [9]. Depending on the ETF prospectuses, authorized participants or market makers can borrow the underlying securities or use derivatives to construct the basket of portfolios so that ETF portfolios (i.e. NAVs) can achieve the promised returns. So, there are many alternatives to tie the NAV returns to the underlying index returns. Depending

[8] It is obvious that there exist other types of arbitrage opportunities. For example, investors can use both S&P 500 futures contracts and S&P 500 index based ETFs to achieve the arbitrage profits

[9] There are broadly two ways of creation and redemption process. The "in-kind" method is that authorized participants create a basket of securities which is exchange for ETF shares. Another technique is "Cash" method. The "Cash" creation/redemption process is allowed for some ETFs so that authorized participants deliver cash to the issuer and receive the ETF shares.

on the market conditions, however, observed ETF prices are frequently different from their announced NAVs. That is, the lack of liquidity or low trading volume in the ETF market could lead to the large price impact or the presence of stale prices thus can cause the price gap between the ETF and its NAV. As a result, APs may bear unwanted costs related to borrowing underlying securities or holding inventories to make the ETF market if ETFs or underlying securities markets are not fully liquid. This situation implies that the lack of liquidity in the ETF market causes the increase of trading costs as well as ETF tracking errors.

3.2 ETF tracking errors and illiquidity

Figure 12 shows the relations between ETF and NAV returns (left) and between NAV and index returns (right) for all the US ETFs in the sample. Each point in the figure represents the average daily return of each ETF from the inception date to the end of 2012 or the delisting date. The solid lines indicate the fitted regression lines between two return series, and the dotted lines are the 45 degree lines. From the right side of Figure 12, most US ETFs appear to be managed correctly to track the underlying indexes although some ETFs are shown to have some tracking errors. The coefficient of NAV-index cross-sectional regression is 0.95 which is close to one, suggesting that US ETF portfolios are managed to track the underling indexes precisely. On the contrary, the left panel shows that there exist relatively large tracking errors between ETFs and their NAVs. The fitted coefficient of ETF-NAV regression is 1.34, implying that ETF returns appear to more frequently deviate from NAV returns than NAV returns do from index returns. Thus, figure 12 suggests that ETF returns can deviate from their NAV returns although ETF portfolios are managed precisely to mimic underlying indexes.

Figure 13 and 14 also provide the same illustrations for the individual ETFs. Figure 13 depicts the return relations for SPDR S&P 500 ETF Trust (SPY) and

Figure 12: Return distributions among ETF, NAV, and index returns

This figure illustrates relations between ETF returns and NAV returns(left figure), and between NAV returns and underlying index returns(right figure). Each point represents the daily average returns for the entire sample period.

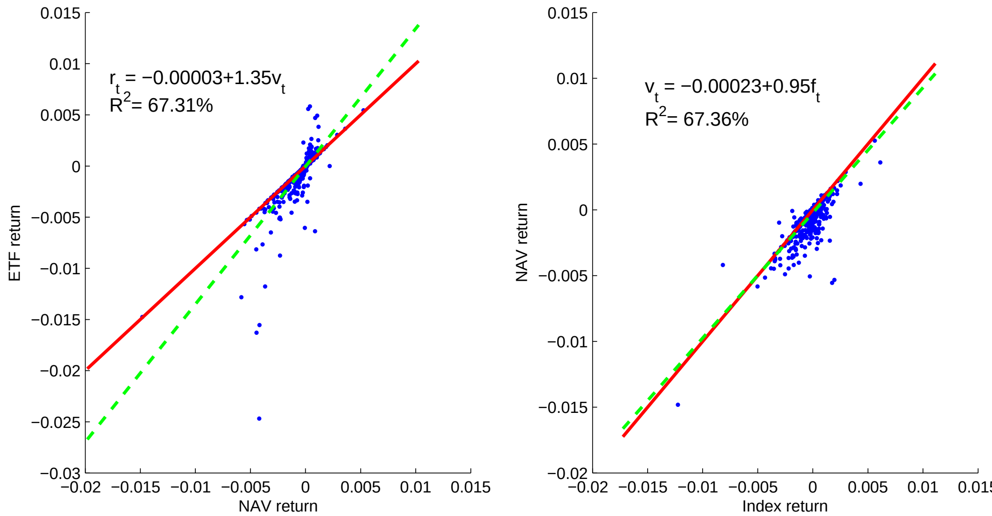


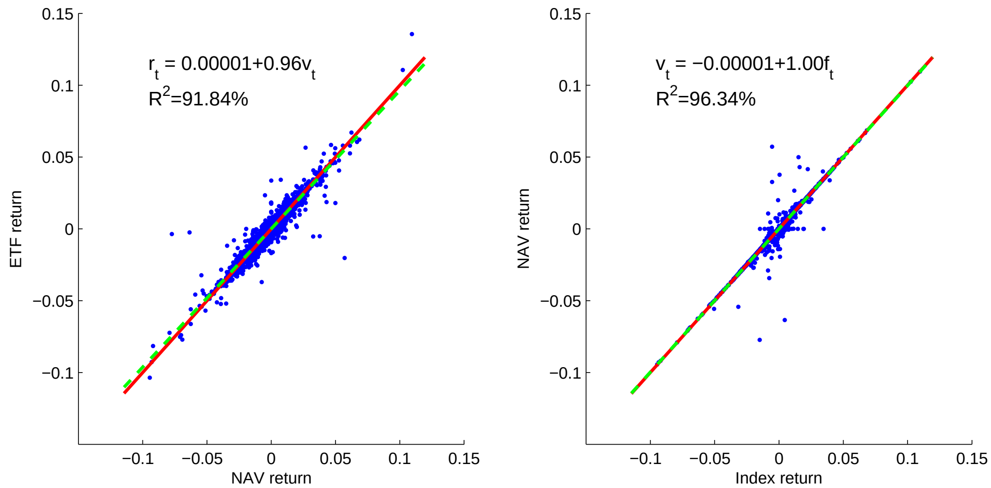
figure 14 is for iShares MSCI Emerging Markets Index (EEM). Both ETFs are very liquid assets in the US markets. Incepted in 1993, the SPY, the oldest and the largest ETF in the US, tracks the price and yield performance of the S&P 500 index. From figure 13, the fitted coefficient of ETF-NAV regression is 0.96 with R^2= 92% and that of NAV-index regression is 0.99 with $R^2 = 96\%$. These results suggest that the SPY ETF tracks the S&P 500 index correctly and its market prices are formed close to its NAVs.

On the contrary, the EEM in figure 14 appears to have relatively larger tracking errors than the SPY does[10]. The EEM, one of the most popular international ETFs in the US, is designed to track the price and yield performance of MSCI emerging market

[10]In 2012, the average trading volume for the SPY is 136 million shares and that for the EEM is 53 million shares. The average relative bid-ask spread for SPY is 0.01% and that for the EEM is 0.017%. The average turnover for the SPY is 18.96% and that for the EEM is 5.45%. For the comparison, the average trading volume is 0.70 million shares, the average relative bid-ask spread is 0.24%, the average turnover is 3.6% for entire ETFs in 2012.

Figure 13: Return distributions among ETF, NAV, and index returns (SPY)

This figure illustrates ETF, NAV, and index returns of SPY. SPY is SPDR S&P 500 ETF issued by State Street Global Advisors. The left panel depicts the relation between daily ETF and NAV returns, the right panel depicts the relation between daily NAV and index returns from Jan 22,1993 to Dec 31, 2012.

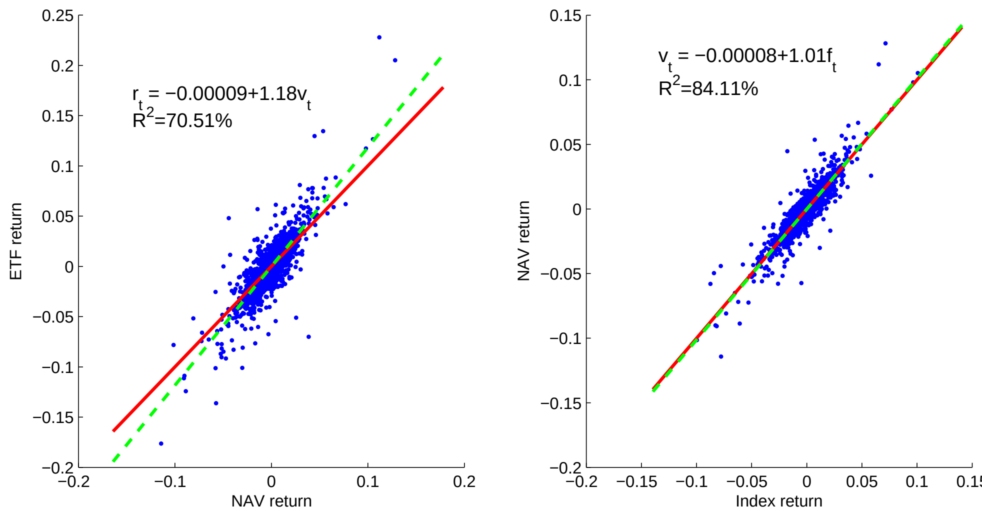


index. Because the EEM holds the emerging market stocks directly, the EEM market prices may not reflect the movements of underlying securities immediately, thus they have some tracking errors although the EEM is a quite liquid asset. Although the EEM is managed correctly to track the underlying index ($R^2 = 84\%$ and coefficient is 1.01), its market prices are shown to more frequently deviate from its NAVs ($R^2 = 70\%$ and coefficient is 1.18). In sum, both figures suggest that tracking errors in the ETF-NAV returns appear to be more severe than those in NAV-index returns.

The time series relation between return differences and illiquidity are illustrated in figure 15. The first line depicts the average of the daily relative bid-ask spread. The bottom two lines are the average absolute daily return differences between ETFs and NAVs, and NAVs and indexes. First, both return differences and illiquidity co-moved over time. That is, there appears to be common factors affecting both illiquidity and return differences. Second, the ETF-NAV return differences are generally higher

Figure 14: Return distributions among ETF, NAV, and index returns (EEM)

This figure illustrates ETF, NAV, and index returns of EEM. EEM is iShares MSCI Emerging Markets Index ETF issued by iShares. The left panel depicts the relation between daily ETF and NAV returns, the right panel depicts the relation between daily NAV and index returns from Apr 11,2003 to Dec 31, 2012.



than NAV-index return differences over the sample period. This result confirms that ETF-NAV tracking errors are larger than NAV-index tracking errors. Finally, both ETF market illiquidity and tracking errors increase during the financial crisis period after 2008, suggesting the illiquidity measure reflects the recent liquidity crisis well.

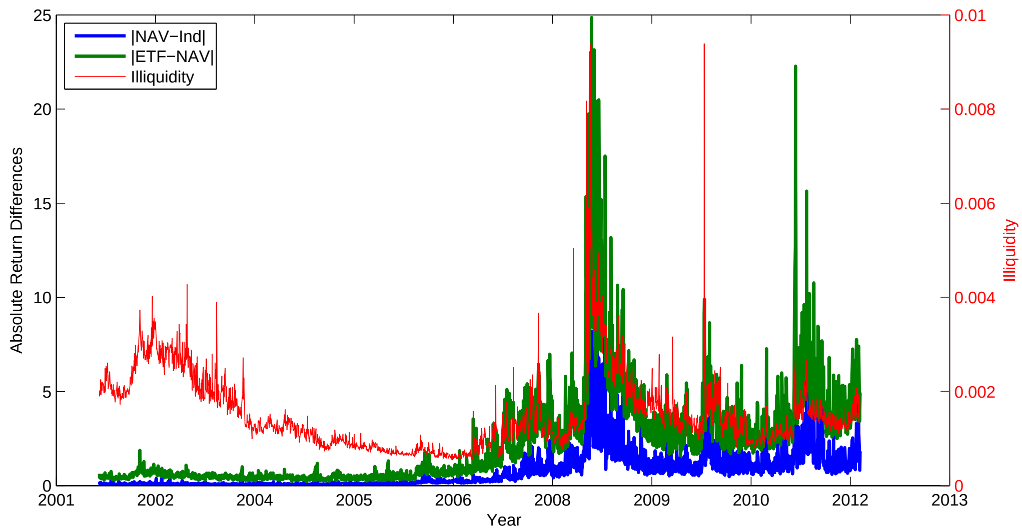
3.3 Panel regression

Three dimensions of tracking errors are used to investigate the effect of liquidity on tracking errors: ETFs vs NAVs, ETFs vs Indexes, or NAVs vs Indexes. As defined in the above, I use two types of tracking errors to investigate the effect of liquidity on tracking errors. The main variable of interest is illiquidity which is defined as the average of the relative effective half spread during the year. Thus, high values of relative spread imply the high illiquidity.

It is obvious that ETF prices are affected by both product structures and ETF

Figure 15: Daily time series for illiquidity and return differences

This figure illustrate historical time series of illiquidity and absolute values of return differences from 2002 to 2012. The average illiquidity is the equal-weighted daily average of daily relative effective spreads of all the US ETFs. The absolute return differences are absolute values of daily return differences between NAV and index returns or ETF and NAV returns.



market conditions. That is, the ETF-NAV tracking errors could exist when ETF market is illiquid although the underlying portfolios are constructed to track the underlying index correctly and perfectly. On the other hand, the NAV-index return tracking error could be affected by ETF structures as well as by the ETF trading activity. For instance, ETFs replicating the US market indexes are less likely to deviate from the underlying indexes than ETF investing in other countries. Furthermore, the NAV-index return tracking error could be caused by the way of replicating the index such as holding underlying securities directly or creating the return using either futures or swaps. Thus, I include different types of variables to capture both market conditions and fund characteristics.

Controlling the ETF market conditions, I include dollar trading volume, underlying index return volatility, shares outstanding, and volatility of shares growth. If the underlying index return is very volatile, the ETF price may not reflect the underlying index movement promptly because the market makers or ETF investors need to trade the ETFs more frequently. The annual underlying index return volatility is added to control this effect. Moreover, the log of the average dollar trading volume during the year is included. The ETFs with a large dollar trading volume may cause tracking error because a heavy trading volume is related to unnecessary price pressure. Finally, ETF shares are easily created based on the market demand. The number of shares represents the size of the ETF or the cash flows into funds. Further, the volatility of the shares growth rate indicates how active the ETF is in the market. A frequent change in outstanding shares implies the active management by authorized participants to reduce tracking error or the volatile fund cash flows. To control these effects, I include the log of the average number of shares and the standard deviation of the shares growth rate during the year.

Aside from the variables associated with market conditions, the ETF characteristics variables are also included to capture any additional effects caused by fund structures. Those are (1) US Based: whether the underlying securities in the ETF baskets invest in US assets, (2) Derivatives Based: whether an ETF uses derivatives to replicate the underlying index return, (2) Swap Based: whether an ETF uses swaps to replicate the underlying index return, (3) Futures Available : whether an ETF has a futures product based on it, (4) Options available: whether an ETF has options based on it, (5) Levered Fund: whether an ETF is levered or inversed, (6) Expense Ratio: the annual expense ratio of the ETF.

3.4 Empirical results

Tables 18 reports results from the pooled panel regression of yearly tracking errors on illiquidity as well as other control variables. Because the dependent variables are tracking errors, positive coefficient signs imply larger tracking errors. The tracking errors are calculated by taking the absolute value of difference between one and the regression coefficient, which is estimated from regressing one return series on another return series for each ETF every year. From left to right, the dependent variable in each column represents the tracking error between the ETF return and the underlying index return, between the ETF return and the NAV return, and the NAV return and the underlying index return respectively. All regression specifications include year fixed effects and standard errors are clustered at the fund level.

Columns 1 and 2 show that the coefficients on average illiquidity are positive and significant at the 1% level. These results suggest that illiquid ETFs tend to be more likely to deviate from their NAVs or underlying index returns. The coefficient of 13.98 indicates that if an ETF's average relative spread increases by 1%, then the tracking error in ETF-index return increases by 14% when holding the other characteristics constant. This illiquidity measure also affects similarly in the ETF-NAV return tracking error in column 2. Interesting thing is that both magnitudes of coefficients on illiquidity are similar (13.98 and 13.80). However, the results in the third column (NAV-index tracking error) show that ETF illiquidity is not associated with the tracking error between NAV and underlying index returns. In other words, the ETF market conditions don't account for the tracking error between NAV and index returns.

Large trading volume doesn't seem to widen the ETF tracking errors. Heavy trading volume can increase the efficiency of the asset price because large trading volume indicates the presence of informed traders. The test results show that the coefficients on dollar trading volume are shown be positive for ETF-NAV and be negative for ETF-index and NAV-index tracking errors, albeit insignificantly. The coefficients on the number of shares, which could measure the size of the fund, are shown to be negative for all tracking errors but only significant for ETF-NAV tracking errors. These results imply that large ETFs tend to have smaller tracking errors.

Table 18: Illiquidity and tracking errors from regrssion

Dependent variables are tracking errors, calculated by taking the absolute difference between one and regression coefficients of ETF returns on underlying index returns(I), of ETF returns on NAV returns(II), or of NAV returns on underlying index returns(III). Independent variables are the following: the average of daily relative bid-ask spread(Average liquidity), the average of daily dollar trading volume(Dollar Trading Volume), the standard deviation of the underlying index return(Index Volatility), the log of the average shares outstanding (Shares Outstanding), the standard deviation of shares' growth rate(Shares Volatility), a dummy being equal to 1 if an ETF uses derivatives(Derivatives Based), a dummy being equal to 1 if an ETF uses swaps(Swap Based), a dummy being equal to 1 if underlying securities in the ETF baskets invest in US assets (Invested in US assets),a dummy being equal to 1 if an ETF has a futures or options based on it (Futures Available , Options Available), a dummy being equal to 1 if an ETF is levered(Levered Fund), the annual expense ratio(Expense Ratio).The numbers in parentheses are t-statistics. Year fixed effects are included, and standard errors are clustered at the fund level. "*", "**", "***" represents significance at 10%, 5%, and 1% level respectively.

	θ(ETF-IND)	θ(ETF-NAV)	θ(NAV-IND)
Intercept	37.15***	49.44***	14.73***
	(3.42)	(7.41)	(3.55)
Average illiquidity	13.98***	13.80***	−0.43
	(2.65)	(2.64)	(−0.45)
Dollar Trading Volume	−0.02	0.51	−0.07
	(−0.04)	(1.63)	(−0.33)
Index Volatility	0.10	−0.95*	0.77
	(0.17)	(−1.88)	(1.20)
Shares Outstanding	−0.79	−1.48***	−0.38
	(−1.39)	(−3.54)	(−1.44)
Shares Volatility	−17.29**	−17.36***	3.19
	(−2.14)	(−3.30)	(1.03)
Derivatives Based	−7.93***	−10.59***	−1.46
	(−2.79)	(−3.04)	(−0.94)
Swap-Based	9.78*	2.52	8.89*
	(1.84)	(0.79)	(1.87)
Equity-type ETF	−12.95***	−14.49***	−5.99***
	(−4.74)	(−7.62)	(−5.02)
Invested in US assets	−4.83***	−3.73***	−3.96***
	(−4.35)	(−3.93)	(−6.07)
Futures Available	−0.91	−0.14	−1.35*
	(−0.74)	(−0.12)	(−1.79)
Options Available	0.10	−0.67	1.38
	(0.09)	(−0.69)	(1.65)
Levered Fund	−3.89	3.46*	−6.97
	(−0.64)	(1.74)	(−1.18)
Expense Ratio	−1.32	−0.55	−0.11
	(−0.81)	(−0.42)	(−0.08)
Year Fixed Effects	Yes	Yes	Yes
Observations	5,459	5,472	5,595
Adjusted R^2	8.99	19.84	10.76

The coefficients on the shares growth volatility are negative and significant for ETF-NAV or ETF-index tracking errors. The share growth volatility is measured as the standard deviation of the shares growth rate. The large value of shares volatility implies that the cash flows through ETFs are volatile. Thus, the high volatility in the shares growth rate implies the attractiveness of ETFs in the market. Alternatively, the frequent adjustment share of the ETF can be interpreted as the active management of market makers to reduce the tracking error between the ETF and the NAV or the underlying index.

Regarding the fund characteristics, the tracking error decreases when the ETF uses derivatives to replicate the underlying index return, when the underlying securities in the ETF baskets invest in US assets, and when futures are available for the ETF. Market makers are able to manage the ETF shares more easily when they use futures to replicate the index return because they need to manage only one or two assets. In addition, because the index futures are created based on the index, managing the ETF using futures can reduce the gap between the ETF and the index. Consistent with previous studies, ETFs replicating the US based indexes tend to have small tracking errors [11]. This finding is not surprising because the US market is one of the most liquid markets in the world, and traders can trade both ETF and underlying securities without a time lag. In addition, the equity-type ETFs tend to have smaller tracking errors than non-equity ETFs. This result implies that the arbitrage activity of authorized participants for equity type-ETFs can easily manage their portfolios because equity or equity-based derivatives markets are more liquid than other asset markets.

Table 19 confirms the results in Table 18. The dependent variables used in Table 19 is the yearly standard deviation of daily return differences to measure tracking

[11] Engle and Sarkar (2006) investigate the premiums (discounts), which is the same as the ETF-NAV tracking errors, for 21 domestic and 16 international ETFs. They find that the tracking errors for domestic ETFs are generally small and temporary but those for international ETFs are large and persistent.

Table 19: Illiquidity and tracking errors from standard deviation of return difference

Dependent variables are tracking errors, calculated by taking the standard deviation of return series between ETF returns and underlying index returns(I), between ETF returns and NAV returns(II), or between NAV returns and underlying index returns(III). Independent variables are the following: the average of daily relative bid-ask spread(Average liquidity), the average of daily dollar trading volume(Dollar Trading Volume), the standard deviation of the underlying index return(Index Volatility), the log of the average shares outstanding (Shares Outstanding), the standard deviation of shares' growth rate(Shares Volatility), a dummy being equal to 1 if an ETF uses derivatives(Derivatives Based), a dummy being equal to 1 if an ETF uses swaps(Swap Based), a dummy being equal to 1 if underlying securities in the ETF baskets invest in US assets (Invested in US assets),a dummy being equal to 1 if an ETF has a futures or options based on it (Futures Available , Options Available), a dummy being equal to 1 if an ETF is levered(Levered Fund), the annual expense ratio(Expense Ratio).The numbers in parentheses are t-statistics. Year fixed effects are included, and standard errors are clustered at the fund level. "*", "**", "***" represents significance at 10%, 5%, and 1% level respectively.

	σ(ETF-IND)	σ(ETF-NAV)	σ(NAV-IND)
Intercept	1.07***	1.23***	0.31
	(3.92)	(5.13)	(0.89)
Average illiquidity	1.37***	1.52***	0.10
	(13.91)	(16.66)	(1.62)
Dollar Trading Volume	−0.01	0.00	−0.01
	(−0.91)	(−0.17)	(−0.54)
Index Volatility	0.28***	0.24***	0.34***
	(6.33)	(4.96)	(2.66)
Shares Outstanding	−0.04**	−0.06***	0.00
	(−2.08)	(−3.42)	(−0.07)
Shares Volatility	−0.18	−0.42*	−0.05
	(−0.78)	(−1.85)	(−0.22)
Derivatives Based	−0.09	−0.21	−0.13
	(−0.44)	(−1.15)	(−0.95)
Swap-Based	0.02	−0.46*	0.55*
	(0.05)	(−1.92)	(1.86)
Equity-type ETF	−0.16*	−0.16*	−0.26**
	(−1.86)	(−1.93)	(−2.22)
Invested in US assets	−0.54***	−0.34***	−0.38***
	(−12.90)	(−10.57)	(−9.39)
Futures Available	−0.01	0.05	−0.07
	(−0.17)	(0.96)	(−1.35)
Options Available	0.01	−0.04	0.00
	(0.22)	(−1.10)	(0.02)
Levered Fund	0.06	0.37	−0.36
	(0.16)	(1.51)	(−0.94)
Expense Ratio	0.07	0.15**	−0.10
	(1.02)	(2.41)	(−1.00)
Year Fixed Effects	Yes	Yes	Yes
Observations	5,459	5,472	5,595
Adjusted R^2	52.58	48.87	30.89

errors. Results are very similar with this alternative tracking error measure. The coefficients on illiquidity are still positive and significant for ETF-index and ETF-NAV tracking errors. ETFs with investing in the US assets or with equity-based tend to have small tracking errors. One different feature is that coefficients on index return volatility is positive and significant, suggesting that authorized participants can have trouble in tracking underlying indexes or constructing portfolios when underlying indexes are volatile.

Overall, ETF tracking errors are severe when ETFs are not actively traded in the market. Given the negative relation between the number of shares and tracking errors, increase in the number of shares can lead to the liquidity increase of the ETFs because large ETFs could attract more investors and be easily traded in the markets. Investors can also avoid high transaction costs when investing in index-based ETFs that trace inaccessible markets. Illiquid ETFs, however, may be riskier than investing directly in underlying assets due to the high tracking errors. In these situations, investors face systemic liquidity risk, which results in different outcomes.

4 The effect of liquidity on ETF returns

4.1 Liquidity adjusted asset pricing model

This section investigates the effect of liquidity on the expected ETF return and on the ETF tracking error with respect to the index using the liquidity adjusted capital asset pricing model. Acharya and Pedersen (2005) developed this model, which leads to three different types of risk premium associated with liquidity risk as well as market risk, and they argue that the asset price reflects these risk premiums. That is, the cost adjusted net asset return has a linear relation with the market return considering the market transaction cost. They show that the individual net return can be expressed as

$$E(r_t^i - r_t^f) = E(c_t^i) + (\beta^{1i} + \beta^{2i} - \beta^{3i} - \beta^{4i})E(r_t^M - c_t^M - r^f) \tag{57}$$

where four betas are defined as

$$\begin{aligned} \beta^{1i} &= \frac{cov(r_t^i, r_t^M)}{var(r_t^M - c_t^M)} & \beta^{2i} &= \frac{cov(c_t^i, c_t^M)}{var(r_t^M - c_t^M)} \\ \beta^{3i} &= \frac{cov(r_t^i, c_t^M)}{var(r_t^M - c_t^M)} & \beta^{4i} &= \frac{cov(c_t^i, r_t^M)}{var(r_t^M - c_t^M)} \end{aligned} \tag{58}$$

From equation (57), the net beta consists of four different betas. In addition to the conventional market beta (β^{1i}), there additionally appear three liquidity betas that represent the relation between market liquidity and individual asset liquidity (β^{2i}), between market liquidity and the individual asset return (β^{3i}), and between individual asset liquidity and the market return (β^{4i}).

β^{2i},representing the relation between individual liquidity and market liquidity, is expected to be positive. Illiquid stocks tend to have large values for β^{2i}, implying that they are significantly affected by the lack of liquidity when the market is illiquid. β^{3i}, which shows the relation between the individual asset return and market liquidity, is expected to be negative. The expected return on the illiquid stock decreases further because the illiquid assets should be sold at a lower price than expected when the market is illiquid. Finally, β^{4i} also has a negative value and measures the relation between the market return and the individual stock liquidity. This negative value implies that the expected return on the illiquid asset decreases when the market declines.

Like general common stocks, the liquidity of the ETF market also affects the expected return of individual ETFs, which replicate the specific index return. An ETF with lower liquidity than the market liquidity may not be able to correctly reflect the level of the underlying index. In other words, the price of an ETF with high liquidity immediately reflects the movement of the underlying index when the underlying index changes. However, insufficient trading may cause the illiquid ETF to fail to trace the underlying index accurately. As a result, a tracking error occurs if an ETF suffers from a lack of liquidity due to insufficient trading activity.

Another issue related to the tracking error is that the tracking error could be caused by the illiquidity of the underlying securities in the ETF baskets. That is, the NAV may not fully reflect the current value of the underlying index due to the illiquidity of the underlying securities in the case of in-kind ETFs. If the underlying securities in the ETF are not traded actively in the market, the market makers fail to properly create or redeem the ETF unit. In this case, the liquidity problem in the underlying securities may cause the difference between the NAV return and the underlying index return. Thus, this type of liquidity problem is not related to ETF market liquidity. Even though the effect of the liquidity of the underlying securities is also an important issue, in this paper, I focus on ETF market liquidity and investigate the effect of ETF liquidity on the return and the variance.

To investigate the liquidity effect, I first estimate the portfolio betas of LCAPM by using 10 liquidity portfolios and 10 tracking error portfolios. However, calculating portfolio betas may lose important information regarding the ETF characteristics because each ETF has its own benchmark index and traces that index rather than the entire ETF market. To mitigate these concerns, I calculate the betas for each individual ETF and report the average of the betas within each portfolio by assuming that the corresponding underlying index return for each ETF is treated as the market return.

4.2 Portfolio construction

I construct 10 liquidity portfolios and 10 tracking error portfolios to investigate the effect of liquidity on the ETF return. All of the ETFs are equally weighted within each portfolio. The 10 liquidity portfolios are constructed for each month t by ranking all ETFs with their liquidity measures at the end of month $t-1$. The liquidity for each month is the average of the daily relative effective half spread of each ETF having at least 15 observations in each month. Similarly, 10 tracking error portfolios are formed for each year y by sorting the ETFs having at least 60 observations in the previous year with tracking error. The tracking error is defined as the absolute difference between one and the estimated coefficient from the regression of the ETF return on the underlying index return. The daily return of each portfolio is simply the average daily return of the ETFs included in each portfolio.

$$r_t^M = \sum_{i=1}^{N_t} w_t^i f_t^i \tag{59}$$

The daily market return is computed as the average of the underlying index return for each ETF used in constructing portfolios. The underlying index return traced by each ETF is not actually traded in the market. The use of the underlying index return to calculate the market return avoids potential measurement error due to trading effects such as the bid-ask bounce or price reversal.

The daily portfolio liquidity is the average of the relative effective bid-ask spreads of the securities included in each portfolio. That is,

$$c_t^p = \sum_{i \in p} w_t^p c_t^i \tag{60}$$

where p is either portfolio or market.

Similarly, the daily market liquidity is calculated by taking the average of the relative effective bid-ask spreads of all ETFs included in the portfolio's construction. Given the persistence of liquidity, it is desirable to use liquidity innovation rather than the observed relative effective bid-ask spread. The liquidity innovation of each security is obtained from the fitted residual of the following $AR(2)$ specification.

$$c_t^i = a_0 + a_1 c_{t-1}^i + a_2 c_{t-2}^i + u_t^i \tag{61}$$

The portfolio liquidity innovation and the market liquidity innovation are calculated in the same way.

4.3 Liquidity risk

Table 20 shows the characteristics of the liquidity portfolios (Panel A) and the tracking error sorted portfolios (Panels B and C). As seen in each panel, the liquidity and tracking error portfolios show similar patterns. That is, the transaction cost and the tracking error increase as liquidity decreases even if the portfolios are constructed based on the past illiquidity and the past tracking error of the ETF. This result implies that both illiquidity and the tracking error of the ETF are persistent.

Panel A shows that the expected transaction cost($E[c]$) is shown to monotonically increase from portfolio 1 through portfolio 10. For instance, the expected transaction cost for portfolio 1, which is the most liquid portfolio, is only 0.032% while that of portfolio 10 is 0.438%. Although the liquidity cost differences are reduced for the tracking error portfolios in panels B or C, the increasing pattern in the liquidity cost through the portfolios is similar to the liquidity portfolios. Moreover, the turnover rate, which measures how an ETF is actively traded in the market, is shown to be lower in the low liquidity portfolio than in the high liquidity portfolio. Next, the portfolio volatility in column 9 shows that that there is no big difference between the portfolios. However, the volatility of the difference between the ETF and the underlying index returns, which is another definition of the tracking errors, increases as liquidity decreases. This result implies that an ETF with low liquidity cannot perfectly follow the underlying index return and that liquidity and the tracking errors are positively related.

Table 20: Properties of sorted portfolios

This table presents the characteristics of 10 equal-weighted liquidity and tracking error portfolios. The 10 liquidity portfolios are constructed for each month t by ranking all ETFs with their liquidity measures at the end of month $t-1$. The liquidity(c) for each month is the average of the daily relative effective half spread of each ETF having at least 15 observations in each month. The 10-tracking error portfolios are formed for each month t by sorting the ETFs having at least 15 observations in the previous year with tracking error. The tracking error ($|1-\theta|$) is defined as the absolute difference between one and the estimated coefficient (θ) from the regression of the ETF return on the underlying index return. $Prem$ is the ETF premium or discount defined as the difference between the ETF price and the NAV divided by the NAV. trn denotes the daily ETF turnover defined as the trading volume divided by the ETF shares outstanding. $\sigma(r^p)$ is the standard deviation of the daily portfolio return. $\sigma(r^{e,p})$ is the standard deviation of the daily portfolio excess return on the underlying index return. The numbers in parentheses are t-statistics.

	β^{1p} (.10)	β^{2p} (.10)	β^{3p} (.10)	β^{4p} (.10)	$E(c^p)$ (%)	$\|1-\theta\|$ (%)	Prem (%)	trn (%)	$\sigma(r^p)$ (%)	$\sigma(r^{e,p})$ (%)
					Panel A. Illiquidity portfolios					
1	11.043	0.002	−0.051	−0.010	0.032	8.202	0.032	7.989	1.200	0.207
	(185.32)	(54.94)	(−7.22)	(−5.51)						
2	11.259	0.005	−0.054	−0.018	0.054	9.327	0.046	6.493	1.224	0.184
	(183.43)	(49.72)	(−7.48)	(−4.20)						
3	9.943	0.005	−0.048	−0.028	0.067	9.716	0.043	6.485	1.087	0.178
	(170.27)	(76.43)	(−7.32)	(−7.39)						
4	10.001	0.007	−0.049	−0.024	0.082	9.851	0.039	6.826	1.088	0.170
	(181.60)	(66.00)	(−7.52)	(−4.25)						
5	10.027	0.008	−0.053	−0.034	0.098	11.365	0.034	4.766	1.100	0.215
	(163.78)	(77.98)	(−8.11)	(−5.65)						
6	10.448	0.009	−0.054	−0.052	0.117	14.723	0.026	3.765	1.142	0.222
	(171.58)	(75.63)	(−7.96)	(−8.19)						
7	11.101	0.010	−0.054	−0.064	0.142	14.917	0.030	3.078	1.212	0.257
	(174.50)	(70.03)	(−7.59)	(−8.58)						
8	11.595	0.014	−0.064	−0.074	0.182	18.791	0.030	2.516	1.259	0.343
	(186.78)	(76.33)	(−8.54)	(−7.45)						
9	11.012	0.015	−0.069	−0.097	0.252	20.994	0.048	2.123	1.223	0.492
	(144.91)	(61.98)	(−9.54)	(−8.44)						
10	11.352	0.026	−0.081	−0.147	0.438	24.890	0.055	1.944	1.282	0.489
	(126.61)	(58.56)	(−10.68)	(−7.19)						
					Panel B. Tracking Error Portfolios (Regression)					
1	10.910	0.006	−0.050	−0.052	0.099	6.164	0.044	6.188	1.218	0.171
	(140.93)	(47.52)	(−7.04)	(−8.96)						
2	11.198	0.009	−0.058	−0.034	0.103	6.570	0.047	6.251	1.227	0.179
	(171.12)	(55.77)	(−7.94)	(−4.75)						
3	10.822	0.008	−0.050	−0.055	0.106	6.868	0.046	6.407	1.180	0.166
	(183.30)	(60.63)	(−7.17)	(−8.45)						
4	11.081	0.010	−0.058	−0.043	0.114	7.714	0.047	5.865	1.215	0.183
	(169.97)	(61.88)	(−8.14)	(−5.35)						
5	10.336	0.009	−0.055	−0.060	0.123	8.716	0.049	5.586	1.136	0.188
	(164.20)	(60.95)	(−8.19)	(−8.41)						
6	10.644	0.012	−0.056	−0.053	0.137	10.632	0.058	4.712	1.161	0.195
	(179.22)	(64.63)	(−8.33)	(−5.91)						

Table 20: Properties of sorted portfolios - Continued

	β^{1p} (.10)	β^{2p} (.10)	β^{3p} (.10)	β^{4p} (.10)	$E(c^p)$ (%)	$\|1-\theta\|$ (%)	Prem (%)	trn (%)	$\sigma(r^p)$ (%)	$\sigma(r^{e,p})$ (%)
7	11.003	0.010	−0.058	−0.061	0.157	12.982	0.086	3.971	1.197	0.262
	(188.54)	(57.12)	(−8.23)	(−7.39)						
8	11.336	0.010	−0.060	−0.056	0.179	16.626	0.101	3.571	1.234	0.348
	(186.05)	(48.21)	(−8.20)	(−6.35)						
9	10.624	0.014	−0.067	−0.066	0.198	22.466	0.158	2.794	1.177	0.470
	(151.01)	(53.63)	(−9.87)	(−5.50)						
10	8.897	0.011	−0.056	−0.063	0.219	39.775	0.141	2.445	1.015	0.609
	(120.10)	(33.39)	(−9.40)	(−5.28)						
Panel C. Tracking Error Portfolios (Excess Return Volatility)										
1	10.699	0.005	−0.048	−0.023	0.057	4.262	0.029	3.931	1.171	0.085
	(173.24)	(57.73)	(−7.07)	(−5.56)						
2	10.828	0.006	−0.051	−0.027	0.068	5.398	0.031	4.859	1.184	0.098
	(175.24)	(75.03)	(−7.32)	(−5.92)						
3	10.175	0.007	−0.048	−0.030	0.083	7.530	0.039	5.673	1.123	0.157
	(156.91)	(72.92)	(−7.21)	(−5.89)						
4	9.647	0.008	−0.042	−0.042	0.096	9.409	0.043	6.167	1.068	0.161
	(152.84)	(87.67)	(−6.70)	(−7.08)						
5	10.084	0.010	−0.051	−0.050	0.115	10.556	0.045	6.810	1.102	0.165
	(175.52)	(78.50)	(−7.82)	(−7.02)						
6	10.631	0.015	−0.058	−0.050	0.145	12.250	0.060	5.727	1.163	0.198
	(173.86)	(57.50)	(−8.43)	(−4.18)						
7	10.975	0.014	−0.058	−0.078	0.186	15.745	0.085	4.436	1.213	0.301
	(155.66)	(71.07)	(−8.17)	(−7.50)						
8	11.913	0.011	−0.082	−0.059	0.209	19.011	0.141	3.057	1.315	0.495
	(157.98)	(44.26)	(−10.78)	(−5.89)						
9	12.135	0.011	−0.078	−0.112	0.228	23.690	0.140	2.654	1.352	0.715
	(143.91)	(45.54)	(−9.78)	(−10.99)						
10	9.841	0.012	−0.052	−0.085	0.255	30.500	0.165	4.573	1.202	0.714
	(85.98)	(31.95)	(−7.35)	(−6.29)						

The relation between illiquidity and the tracking error is well illustrated from the distribution of the tracking error in column 6. The tracking error for the low liquidity portfolio appears to be larger than that for the high liquidity portfolio. That is, the price of the ETF with low liquidity tends to deviate more frequently from its underlying index. Moreover, the average premium of the ETF relative to the NAV is positive and increases as the liquidity decreases. The ETF-NAV return differences are affected by both the illiquidity of the underlying securities and the ETF liquidity. Arbitrageurs in the ETF market try to trade the ETF close to the publicly announced NAV price and the underlying index. However, if an ETF does not have enough liquidity so that traders cannot immediately trade the ETF to respond to the movement of the underlying index, the ETF market illiquidity causes the disparity between the ETF price and the underlying index.

The estimated betas are also reported by multiplying by 10 for convenience. It is not surprising that β^{1p}, measuring the market risk, is close to one, which implies that the ETFs in the US stock exchange trace the underlying index well on average. Moreover, three liquidity betas reflect the characteristics of the liquidity well even though the magnitude is small. The portfolio β^{2p}s, indicating the relation between market liquidity and individual liquidity, are positive, implying that individual liquidity decreases when market liquidity decreases. Illiquid ETFs have large values for β^{2p} and are more sensitive to market liquidity shocks. As expected, both the β^{3p}s and the β^{4p}s have negative values. Moreover, the ETFs in the low-liquidity or high-tracking error portfolios tend to have large absolute values for β^{3p} and β^{4p}. This result suggests that illiquid ETFs are more likely to deviate from the underlying index return and are more sensitive to a change in the market return or the market liquidity.

The discussion above relies on the portfolio betas rather than the individual ETF betas. However, it is desirable to calculate the individual ETF betas because each ETF is designed to follow the specific underlying index. Considering the underlying index return as the ETF's market return, the calculated market and liquidity betas provide more reliable variables for measuring the market and liquidity risk. To account for these concerns, I provide the average betas in each portfolio after estimating the individual ETF betas. The yearly portfolios are formed using the same method as the previous portfolio beta calculation. Table 21 reports the average betas of the liquidity and tracking error portfolios. Overall, the results are quite similar to the patterns in the portfolio betas. The illiquid ETFs tend to be more sensitive to market liquidity or the market return. Moreover, the illiquid ETFs are more likely to deviate from their underlying index return.

Table 21: Properties of sorted portfolios based on individual ETFs

This table presents the characteristics of 10 equal-weighted liquidity and tracking error portfolios. The method for constructing portfolios and variable definitions are the same as in Table 20. After estimating yearly betas for each ETF, equal weighted averages within portfolios are reported. Other statistics follow the same procedure. The numbers in parentheses are t-statistics.

	β^{1p} (.10)	β^{2p} (.10)	β^{3p} (.10)	β^{4p} (.10)	$E(c^p)$ (%)	$\|1-\theta\|$ (%)	Prem (%)	trn (%)	$\sigma(r^p)$ (%)	$\sigma(r^{e,p})$ (%)
Panel A. Illiquidity portfolios										
1	9.407	0.001	−0.030	−0.008	0.033	5.606	0.024	7.774	1.325	0.369
	(142.74)	(1.64)	(−5.08)	(−4.56)						
2	9.398	0.002	−0.029	−0.009	0.057	6.386	0.049	7.562	1.467	0.412
	(87.11)	(1.84)	(−4.73)	(−2.72)						
3	9.304	0.015	−0.048	−0.012	0.071	7.328	0.037	7.424	1.545	0.475
	(136.70)	(1.11)	(−2.33)	(−4.96)						
4	9.379	0.002	−0.021	−0.017	0.085	6.662	0.019	5.339	1.625	0.507
	(77.14)	(1.97)	(−2.11)	(−3.85)						
5	9.332	0.003	−0.030	−0.021	0.098	7.458	0.030	4.497	1.654	0.561
	(66.55)	(1.96)	(−4.37)	(−3.73)						
6	9.226	0.004	−0.029	−0.025	0.116	8.195	0.016	3.152	1.622	0.579
	(80.61)	(1.82)	(−3.85)	(−2.66)						
7	9.103	0.005	−0.043	−0.029	0.139	11.187	0.071	3.251	1.659	0.687
	(96.20)	(2.01)	(−3.48)	(−3.29)						
8	8.866	0.010	−0.042	−0.053	0.187	14.038	0.085	2.321	1.797	0.914
	(77.28)	(2.02)	(−2.92)	(−4.02)						
9	8.636	0.007	−0.054	−0.063	0.240	15.848	0.113	2.234	1.847	1.067
	(80.05)	(1.33)	(−3.11)	(−4.69)						
10	8.740	0.015	−0.044	−0.156	0.377	14.865	0.107	1.846	1.803	1.186
	(59.85)	(2.67)	(−3.87)	(−6.14)						
Panel B. Tracking Error Portfolios (Regression)										
1	9.783	0.003	−0.034	−0.019	0.087	2.941	0.021	5.337	1.649	0.354
	(213.05)	(1.52)	(−2.87)	(−2.43)						
2	9.755	0.002	−0.027	−0.019	0.086	2.845	0.045	7.546	1.592	0.355
	(188.40)	(1.70)	(−4.30)	(−3.03)						
3	9.685	0.003	−0.028	−0.023	0.100	3.776	0.061	5.230	1.589	0.385
	(174.55)	(1.85)	(−4.31)	(−3.43)						
4	9.594	0.002	−0.025	−0.025	0.107	4.711	0.023	5.988	1.656	0.438
	(119.58)	(2.25)	(−5.03)	(−4.98)						
5	9.510	0.004	−0.028	−0.031	0.124	5.426	0.062	6.285	1.714	0.547
	(144.59)	(2.63)	(−4.14)	(−3.40)						
6	9.487	0.003	−0.029	−0.038	0.140	6.651	0.054	5.377	1.671	0.637
	(137.21)	(1.97)	(−3.75)	(−6.98)						
7	9.339	0.008	−0.036	−0.039	0.173	8.696	0.061	3.404	1.669	0.768
	(100.93)	(2.33)	(−5.10)	(−3.21)						
8	8.993	0.006	−0.033	−0.045	0.168	12.363	0.096	2.039	1.552	0.848
	(63.43)	(2.68)	(−3.74)	(−4.06)						
9	8.283	0.010	−0.042	−0.084	0.208	17.700	0.094	1.836	1.626	1.084
	(88.80)	(2.37)	(−3.10)	(−5.37)						
10	6.916	0.025	−0.090	−0.073	0.203	32.958	0.043	2.490	1.630	1.329
	(27.25)	(1.64)	(−3.44)	(−3.76)						

Table 21: Properties of sorted portfolios based on individual ETFs - Continued

	β^{1p} (.10)	β^{2p} (.10)	β^{3p} (.10)	β^{4p} (.10)	$E(c^p)$ (%)	$\lvert 1-\theta \rvert$ (%)	Prem (%)	trn (%)	$\sigma(r^p)$ (%)	$\sigma(r^{e,p})$ (%)
	Panel C. Tracking Error Portfolios (Excess Return Volatility)									
1	9.737	0.002	−0.028	−0.013	0.050	2.294	0.017	4.664	1.260	0.182
	(211.39)	(1.58)	(−5.91)	(−2.83)						
2	9.658	0.004	−0.032	−0.015	0.062	3.203	0.027	4.590	1.334	0.232
	(172.06)	(2.04)	(−2.43)	(−2.39)						
3	9.592	0.029	−0.058	−0.020	0.081	5.437	0.017	5.579	1.438	0.301
	(126.97)	(1.91)	(−3.04)	(−2.73)						
4	9.485	0.008	−0.033	−0.026	0.093	5.621	0.056	5.028	1.504	0.374
	(98.42)	(2.42)	(−4.79)	(−2.58)						
5	9.489	−0.001	−0.041	−0.025	0.108	6.298	0.023	5.262	1.614	0.448
	(112.62)	(−0.25)	(−3.49)	(−3.49)						
6	9.415	0.005	−0.033	−0.040	0.151	8.561	0.050	6.799	1.681	0.619
	(94.27)	(1.85)	(−3.84)	(−3.46)						
7	9.133	0.006	−0.038	−0.059	0.183	10.903	0.062	4.163	1.745	0.780
	(71.81)	(2.32)	(−3.16)	(−3.54)						
8	8.885	0.004	−0.047	−0.060	0.194	13.518	0.142	2.830	1.693	0.974
	(56.16)	(2.47)	(−2.81)	(−5.93)						
9	8.223	0.005	−0.034	−0.058	0.227	18.492	0.118	2.864	1.822	1.194
	(63.66)	(2.69)	(−3.94)	(−4.78)						
10	7.751	0.003	−0.028	−0.080	0.251	23.455	0.051	3.849	2.271	1.650
	(33.94)	(3.51)	(−4.01)	(−4.80)						

In sum, the above results support that liquidity is the important factor in determining the ETF return, and it causes the tracking error of the ETF with respect to the underlying index or the NAV returns. The estimated portfolio betas suggest that the liquidity risk is the undiversified systemic risk even if constructing the portfolio. Moreover, the liquidity risk is closely related to the ETF tracking errors. In general, the return difference between the ETF and the underlying index or the NAV can be removed through arbitrage activity. However, a lack of liquidity in the ETF can cause an unexpected loss to the arbitrageurs when they fail to trade the ETF on their target price. Thus, liquidity plays an important role in eliminating the arbitrage opportunity in the ETF market. ETF investments provide a valuable opportunity to indirectly invest in inaccessible markets. However, if there is a tracking error due to a lack of liquidity, investing in ETFs brings a different result from the direct investment in the particular markets. Moreover, if the liquidity risk is the systemic risk that exists even after constructing the portfolio from ETFs, the merit of investing in the ETF is less attractive. If liquidity risk from investing in the ETF exists, the ETF investors must be compensated for bearing this liquidity risk. Thus, the next section investigates how the liquidity risk affects the expected ETF return.

4.4 Liquidity premium

This section investigates the effect of liquidity on the expected return of the ETF from a cross-sectional regression by using pre-estimated betas. The regression is estimated by using the GMM method. As did in Acharya and Pedersen (2005), the standard error is calculated by using Newey and West (1987) with lag 2. The following three equations are used to estimate parameters:

$$E(r_t^p) = \alpha + \kappa E(c_t^p) + \lambda \beta^{net,p} \tag{62}$$

$$E(r_t^p) = \alpha + \kappa E(c_t^p) + \lambda_1 \beta^{1p} + \lambda \beta^{net,p} \tag{63}$$

$$E(r_t^p) = \alpha + \kappa E(c_t^p) + \lambda_1 \beta^{1p} + \lambda_2 \beta^{2p} + \lambda_3 \beta^{3p} + \lambda_4 \beta^{4p} \tag{64}$$

The above models are estimated either when the coefficient on the expected trading cost, κ, is fixed as the average turnover rate or when it is considered to be the free parameter. The equations are estimated by either using pre-estimated portfolio betas or pre-estimated individual ETF betas. The estimated parameters using portfolio betas are reported in Table 22 and those using individual betas are reported in Table 23. Panel A of each table reports the estimated results from the liquidity portfolios and panels B and C do the same for the tracking error portfolios. The odd and even lines of each panel report the estimation results when κ is fixed as the average daily turnover rate and treated as the free parameter, respectively.

Table 22: Illiquidity and tracking error portfolios: portfolio betas

This table presents the estimated coefficients from cross-sectional regressions of the liquidity-adjusted CAPM for 10 equal-weighted portfolios using daily data during 2002-2012. The odd and even lines of each panel report the estimation results when κ is fixed as the average daily turnover rate and treated as the free parameter, respectively. The numbers in parentheses are t-statistics

	Constant	$E(c^p)$	β^{1p}	β^{2p}	β^{3p}	β^{4p}	$\beta^{net,p}$
	(.10)	(.10)	(.10)	(.10)	(.10)	(.10)	(.10)
			Panel A. Illiquidity portfolios				
1	−0.01***	4.60					0.05***
	(−8.44)						(8.90)
2	−0.02***	−119.99***					0.05***
	(−10.32)	(−10.81)					(8.97)
3	−0.07***	4.60	−0.81				0.89
	(−10.75)		(−1.02)				(1.14)
4	−0.07***	18.92***	−0.89				0.97
	(−10.77)	(3.66)	(−1.12)				(1.24)
5	−0.01***	4.60	0.02	2.63	−5.85	1.32	
	(−8.30)		(0.15)	(0.04)	(−0.19)	(0.16)	
6	−0.02***	−112.21***	−0.03	−0.10	−16.88	2.93	
	(−9.96)	(−10.63)	(−0.23)	(0.00)	(−0.56)	(0.34)	
			Panel B. Tracking Error Portfolios(Regression)				
1	−0.01***	4.78					0.04***
	(−8.17)						(7.14)
2	−0.02***	−94.23***					0.04***
	(−10.04)	(−9.93)					(7.37)
3	−0.07***	4.78	−6.93***				6.95***
	(−10.90)		(−3.03)				(3.07)
4	−0.07***	32.78***	−6.90***				6.92***
	(−10.82)	(5.92)	(−3.00)				(3.04)
5	−0.01***	4.78	0.00	−21.69	−5.75	−5.56	
	(−7.66)		(0.02)	(−0.25)	(−0.17)	(−0.71)	
6	−0.02***	−97.64***	0.01	−6.11	−2.69	−5.23	
	(−9.55)	(−10.16)	(0.06)	(−0.07)	(−0.08)	(−0.67)	
			Panel C. Tracking Error Portfolios(Standard Deviation)				
1	−0.01***	4.79					0.04***
	(−7.08)						(6.05)
2	−0.02***	−157.90***					0.04***
	(−9.17)	(−9.92)					(5.99)
3	−0.06***	4.79	1.57				−1.46
	(−10.26)		(1.23)				(−1.16)
4	−0.07***	10.20	1.64				−1.53
	(−10.32)	(1.46)	(1.29)				(−1.22)
5	−0.01***	4.79	0.05	4.22	3.13	−0.18	
	(−7.03)		(0.90)	(0.14)	(0.31)	(−0.04)	
6	−0.02***	−148.14***	0.05	−4.74	2.80	−0.27	
	(−9.24)	(−10.10)	(1.05)	(−0.15)	(0.28)	(−0.06)	

The first line of each panel is the GMM estimation result of equation (62). The risk premium is positively significant at the 1% level and quite similar in either the liquidity or the tracking error portfolios (0.05%, 0.04%, and 0.04%) when using the fixed κ. The results are unchanged even when κ is estimated as the free estimator (line 2). The negative coefficient for the expected cost κ can be interpreted as a result of the managerial fees for the ETFs. Finally, the alpha is negatively significant, which is due to the fixed costs including the managerial fees from the ETF.

In lines 3 and 4, the risk premium is estimated to separate the liquidity risk from the market risk by using equation (63). As pointed out in Acharya and Pedersen (2005), a substantial multicollinearity problem exists even when using the ETF data. The $\beta^{net,p}$ is shown to be positive and significant in panel B and insignificant in panel A as well as C. The coefficient for β^{1p} is negative and significant in panel B, which is not necessarily true because the market premium cannot be negative and the net beta also contains the value of β^{1p}. For example, the estimated market premium using the liquidity portfolio is still positive, i.e., $-0.093\beta^{1p} + 0.095\beta^{net,p} = 0.002\beta^{1p} + 0.095(\beta^{2p} - \beta^{3p} - \beta^{4p})$ [12]. This result implies that both the market risk and the liquidity risk are positively related to the expected ETF return.

Finally, lines 5 and 6 report the estimation results of equation (63) when each beta is considered as a separate variable. All of the estimated coefficients are not significant, suggesting the existence of the severe multicollinearity problem.

The economic significance can be found in the investment performance by calculating the return difference between portfolios 1 and 10. The effect of β^{2p}, β^{3p}, and β^{4p}on the annualized return difference between liquidity portfolio 1 and 10 is 0.04%, 0.06%, and 0.26%, respectively. Thus, the annualized return due to the liquidity risk is approximately 0.36%. Consistent with Acharya and Pedersen (2005), the effect of

[12] The market premium is also positive in panel A because $-0.081\beta^{1p} + 0.089\beta^{net,p} = 0.008\beta^{1p} + 0.089(\beta^{2p} - \beta^{3p} - \beta^{4p})$. In panel C, the market premium is 0.011. But, both market premiums in panels A and C are insignificant.

Table 23: Illiquidity and tracking error portfolios: individual ETF betas

This table presents the estimated coefficients from cross-sectional regressions of the liquidity-adjusted CAPM for individual securities using daily data during 2002-2012. The odd and even lines of each panel report the estimation results when κ is fixed as the average daily turnover rate and treated as the free parameter, respectively. The numbers in parentheses are t-statistics

	Constant (.10)	$E(c^i)$ (.10)	β^{1i} (.10)	β^{2i} (.10)	β^{3i} (.10)	β^{4i} (.10)	$\beta^{net,i}$ (.10)
1	−0.01∗∗∗	5.72					0.03∗∗∗
	(−11.71)						(14.08)
2	−0.01∗∗∗	−1.28					0.03∗∗∗
	(−11.71)	(−0.60)					(14.08)
3	−0.04∗∗∗	5.72	0.01				0.06∗∗∗
	(−23.24)		(0.39)				(2.74)
4	−0.04∗∗∗	16.06∗∗∗	0.00				0.07∗∗∗
	(−22.86)	(6.74)	(−0.18)				(3.23)
5	−0.01∗∗∗	5.72	0.02∗∗∗	−0.04∗∗∗	−0.11∗∗∗	−0.03	
	(−9.48)		(12.80)	(−3.04)	(−6.65)	(−0.79)	
6	−0.01∗∗∗	−4.96∗∗	0.02∗∗∗	−0.04∗∗∗	−0.11∗∗∗	−0.03	
	(−11.18)	(−2.35)	(12.90)	(−3.00)	(−6.62)	(−0.92)	

the covariance of an ETF's illiquidity to market returns seems to have the largest impact on the expected returns. Furthermore, this liquidity risk is still an important factor even when investing in the tracking error portfolio. The portfolio with a large tracking error gains more excess return than that with a small tracking error. For instance, the total annualized return difference between regression (standard deviation) tracking portfolios 1 and 10 is 0.03% (0.11%), consisting of 0.01%(0.01%) for β^{2p}, 0.01% (0.01%)for β^{3p}, and 0.03% (0.09%) for β^{4p}. These results imply that the tracking error is related to the liquidity risk, which is a non-negligible risk in ETF investment.

In sum, the liquidity of the ETF market is also an important factor for determining the expected return of the ETF because the ETF is traded like a common stock in the market even if the ETF is designed to strongly replicate the particular index return. Moreover, the liquid ETF tends to track its underlying index better than the illiquid one.

5 The liquidity effect on volatility

5.1 Nontrading probability and variance difference

In this section, the effect of liquidity on the ETF variance is investigated using Lo and MacKinlay (1990)'s econometric model. In particular, I investigate whether a lack of liquidity could cause the difference between the NAV return variance and the ETF return variance. Lo and MacKinlay (1990) develop an econometric model to explain the effect of infrequent trading and shows that nontrading increases the return variance and causes negative serial correlation. If an individual security trades very frequently without any time delays, then the variance of the observed return must be same as the variance of the true asset return. However, the increase in the expected nontrading days can cause a gap between the observed return and the true return.

It is not easy to evaluate whether infrequent trading can increase the asset return variance with respect to the true return variance because the true asset return cannot be observed in general. However, the NAV return can be regarded as the ETF's true return, which is publicly announced in the market. Given the NAV return, it is attractive to test whether nontrading causes the increase in the ETF return variance or in the gap between the ETF return variance and the NAV return variance. Moreover, the NAV return can be easily modeled using a single linear factor model because each ETF is designed to trace its particular index. For the NAV return series, assume the following linear relation between the NAV return and the underlying index return.

$$v_t = \alpha + \beta f_t + \epsilon_t \tag{65}$$

where v_t is the NAV return and f_t is the underlying index return on day t. If an ETF replicates the underlying index perfectly, then the β should be close to one and the α should be close to the fund's expense ratio. While Lo and MacKinlay (1990) assume that the factor return is serially uncorrelated, it is more realistic to assume that a serial correlation exists in the factor return series. The following autoregressive process is suitable to account for the serial correlation of the factor return series:

$$f_t = \phi_0 + \phi f_{t-1} + \xi_t \tag{66}$$

where ξ_t is zero mean noise with variance σ^2. The coefficient of the lagged return is the well-known autocorrelation function of the $AR(1)$ process and is equal to the autocorrelation of lag 1.

As introduced in Lo and MacKinlay (1990), the following two random variables are defined to explain the ETF return process with the nontrading effect. First, the indicator variable δ_t is defined as having the value one if ETF does not trade at the particular date t with probability p. Second, the indicator variable $X_t(k)$ is defined as being one if ETF trades at time t but has not traded in the k previous periods. The indicator variable $X_t(k)$ can be expressed as

$$\begin{aligned} X_t(k) &= (1-\delta_t)\delta_{t-1}\delta_{t-2}\cdots\delta_{t-k}, \qquad k > 0 \\ &= \begin{cases} 1, & \text{with probability}(1-p)p^k \\ 0, & \text{with probability}1-(1-p)p^k \end{cases} \end{aligned} \tag{67}$$

Given the definition of the indicator variable $X_t(k)$, the ETF return can be written as

$$r_t = \sum_{k=0}^{\infty} X_t(k) v_{t-k} \tag{68}$$

From equation (68), the daily ETF return and the daily NAV return should be same if the ETF is traded every day. So, equation (68) means that the ETF return at time t can be expressed as the sum of the NAV returns from time $t-k$ to time t if the ETF has not been traded during the previous k period. Given the definition of the ETF return in equation (68),the variance of the ETF return can be expressed as

$$Var(r_t) = Var(v_t) + \frac{2p}{1-p}(\alpha + \beta\mu)^2 + \frac{2\phi p}{1-\phi p}\beta^2 Var(f_t) \tag{69}$$

Equation (69) shows that the ETF return variance is composed of the NAV return variance and the terms associated with the nontrading and the autocorrelation effects. If the ETF trades every day, which means that the nontrading probability is close to zero, the ETF return variance should be the same as the NAV return variance. The third term, which is related to the product of the nontrading probability and the serial correlation in the underlying index return, is not shown in Lo and MacKinlay (1990). The important aspect in equation (69) is that the nontrading probability plays a critical role in increasing the ETF return variance. The no trading effect does still exist even though there is no serial correlation in the underlying index return. Moreover, the expected return for the ETF is always same as the NAV return; the nontrading probability does not cause any difference between the ETF return and the NAV return. An increase in the nontrading probability could cause the increase in the ETF return variance but not cause any change in the expected return for the ETF. Thus, if an ETF has a high probability of nontrading due to the lack of trading volume, the risk of investing in the ETF could also increase.

Table 24 reports the variance of each return series and the difference between return series. The reported variance is the annualized cross-sectional average of each ETF's variance calculated from the daily return series during the sample period. Panel A shows the average of variances by asset category. First, there exist significant differences between the ETF return variance and the NAV return variance except for the domestic/sector and the real estate ETFs. In particular, the commodity type ETFs tend to have the largest variance difference. Although the differences in the currency ETFs or debt ETFs are small, they are still significant. For the equity-type

Table 24: Variance comparison

This table reports the variance differences among ETF return, NAV return, and underlying index return. The variance is calculated for individual ETF from the inception date to the end of year 2012 or the delisted date. σ_r^2, σ_v^2, and σ_f^2 denote the annual variance of ETF returns, NAV returns, and underlying index returns. The underlying index returns are adjusted for the leverage factor. "*", "**", "***" represents significance at 10%, 5%, and 1% level respectively.

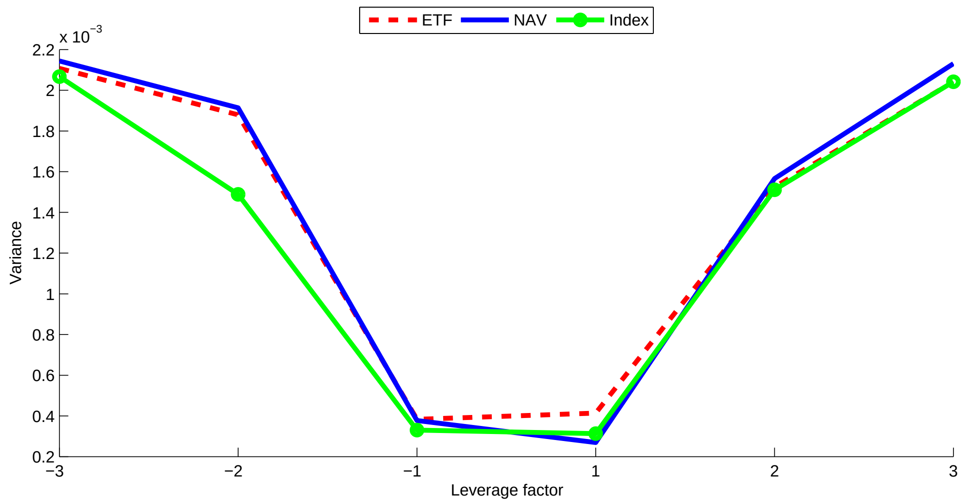
Category	N	σ_r^2 (%)	σ_v^2 (%)	σ_f^2 (%)	$\sigma_r^2 - \sigma_v^2$ (%)	$\sigma_r^2 - \sigma_f^2$ (%)	$\sigma_v^2 - \sigma_f^2$ (%)
Panel A. Asset Category							
Asset Allocation	40	13.27	9.23	8.13	4.04***	5.15***	1.11
Commodity	42	21.02	16.99	17.41	4.03	3.61	−0.43
Currency	20	2.58	2.14	2.14	0.44**	0.44**	0.00
Debt	109	2.53	1.23	1.08	1.29**	1.44**	0.15***
Domestic/Equity	318	12.21	11.07	10.83	1.14***	1.38***	0.24***
Domestic/Sector	259	18.82	18.67	17.14	0.15	1.68*	1.53
Global/Equity	329	18.32	11.39	14.63	6.92***	3.69	−3.23
Global/Sector	147	15.28	10.91	10.51	4.37***	4.77***	0.40
Real Estate	43	20.52	18.82	18.31	1.70	2.21*	0.51
Panel B. Levered or Inversed							
Non-levered	1115	10.43	6.78	7.89	3.65***	2.54*	−1.11
Levered	192	41.78	42.81	38.79	−1.02*	2.99**	4.01***

ETFs, the variance difference for the ETFs based on US equity is smaller than that based on international equity. Next, the difference between the NAV return variance and the index return variance is also significant, although it is smaller than that between the ETF return and the NAV return. This result implies that the price movement of the ETF is more volatile than that of the Index, suggesting that there could be a substantial tracking error in the ETF with respect to the Index.

Panel B reports the averages of the variance difference based on whether the ETF is levered or non-levered. On average, the levered ETFs have higher variances than the non-levered ETFs. Moreover, the ETF variances are higher than the NAV return variance whether the ETFs are levered or not. The return difference between the ETF return and the NAV return is bigger in the non-levered products than in the levered products. Moreover, there is no significant difference between the NAV return variance and the index return variance for the non-levered ETFs, but a significant

Figure 16: Variances of ETF return, NAV return, and Index return by leverage

This figure illustrates the averages of ETF, NAV, and index return variances by leverage factors. The return variances are calculated from ETF daily returns.



difference exists for the levered products. This result suggests that leverage may cause an increase in the true variance, but it does not necessarily cause an increase in the variance relative to the true variance. Figure 16 illustrates the variances of the different return series by the degree of leverage. The variance of the index return is calculated after considering the leverage factor. Figure 16 shows that the variance also increases when the degree of leverage increases. The plot also suggests that the leverage does not necessarily and directly cause the increase in the variance of the trading asset ETFs. In any case, there exists a difference between the ETF return variance and the NAV return variance regardless of whether the ETFs are levered or not. In sum, the return variance or volatility increases when the ETF is not actively traded in the market.

The return variance of the illiquid ETF increases because the price of the illiquid ETF cannot immediately reflect the price of the index; thus, it should reflect all of the past fluctuations of the index, which are not involved in the price due to the nontrading of the asset. As a result, the lack of liquidity could cause an increase in the risk of investing in the ETF because it increases the return variance and decreases the ETF performance even if the expected return is independent of the illiquidity.

5.2 Nontrading probability and the variance

In the previous section, the Lo and MacKinlay (1990) model shows how infrequent trading could affect the ETF return variance compared to the NAV return variance. This section provides empirical evidence to support the previous econometric model. The variance of each ETF is calculated based on the daily data from 2002 to 2012. In the case of an ETF incepted after 2002, the variance is calculated from the inception date. The probability of nontrading is simply defined as the proportion of nontrading days to the actual trading days during the sample period.

Table 25 provides average values for ETF and NAV return variances classified by the nontrading probability. An average of each variance is calculated for each stock from the sample period. Category 1 includes only the ETFs that have been traded every day during the sample period. That is, the nontrading probability of category 1 is zero. The remaining categories are constructed by sorting ETFs that have at least one nontrading day during the sample period with the nontrading probability. The reported variances are annualized for convenience. Equation (69) shows that the difference between the ETF variance and the NAV variance is related to not only the nontrading probability but also to the autocorrelation of the underlying index return. The average autocorrelations for the underlying index return are also reported in the table.

Table 25 shows that the nontrading probability is related to the difference between the ETF variance and the NAV variance. First, as the nontrading probability increases, the difference between the ETF variance and the NAV variance also increases. For instance, ETFs included in category 10 were not traded for 70% of

Table 25: Nontrading probability and variances

This table presents the summary statistics of variances, no trading probability, the expected no trading day of ETFs. In addition, the first-order autocorrelation, the $AR(1)$ coefficient, and the sum of the autocorrelations from lag 1 to lag 10 for the underlying index returns are reported. All of the statistics are calculated from the daily return series for the entire sample period. σ_r^2 and σ_v^2 denote the variance of ETF returns and the variance of NAV returns, respectively. No trading probability,p, is the ratio of observations to total trading days. $E(k)$ is calculated by $p/(1-p)$. ρ_i denotes the lag i autocorrelation. ϕ denotes the coefficient of the $AR(1)$ for the underlying index return.

	σ_r^2 (%)	σ_v^2 (%)	$\sigma_r^2 - \sigma_v^2$ (%)	p (%)	$E(k)$ day	ρ_1 (%)	ϕ (%)	$\sum_i \rho_i$ (%)
1	10.86	10.51	0.35	0.00	0.00	1.82	1.83	−1.38
2	20.87	21.66	−0.79	0.12	0.00	−1.57	−1.57	−5.35
3	19.18	19.60	−0.42	0.28	0.00	1.61	1.60	0.27
4	14.18	14.28	−0.1	0.76	0.01	1.46	1.52	0.46
5	13.72	13.17	0.55	1.92	0.02	1.33	1.34	−0.59
6	11.17	10.73	0.44	4.48	0.05	−0.36	−0.37	−1.58
7	11.75	10.35	1.4	9.12	0.10	1.46	1.44	3.08
8	14.71	9.42	5.29	19.89	0.25	1.99	2.08	−6.55
9	15.63	7.77	7.86	40.01	0.69	−0.27	−0.95	−8.02
10	25.70	6.00	19.7	66.91	2.40	0.19	0.09	−9.16

the trading days and the annual variance difference for those ETFs is 28%. Second, the number of expected nontrading days also increases when the nontrading probability increases. The ETFs included in category 10, which show the least trading activity, have not been traded during three consecutive days on average. Third, the autocorrelation with lag 1 and the $AR(1)$ coefficient for the underlying index return are reported in column 6 and column 7. Moreover, the sum of the autocorrelations from lag 1 to lag 10 for the underlying index return is reported in the last column. Columns 6 and 7 show that there is no clear relation between the autocorrelation of the underlying index return and the nontrading probability of the ETF.

5.3 Cross-sectional regression of variance difference

Equation (69) shows that the variance difference appears to be closely related to the nontrading probability and the autocorrelation. In particular, the nontrading probability plays the role of increasing the ETF volatility relative to the NAV volatility.

The regression analysis is performed to investigate whether the nontrading probability, the autocorrelation of the index return, and the interaction between the two are related to the difference between the ETF return variance and the NAV return variance. In this regression analysis, I use annual variables to control the seasonal effect and to obtain more observations. In each year, the variables are calculated from the daily data for each ETF that has more than 60 observations. The annual nontrading probability is calculated from the proportion of the observed data to the actual market trading days. The primary dependent variable in this regression analysis is the difference between the ETF return variance and the NAV return variance.

Table 26 reports the panel regression results. As seen in column 1 of the table, the coefficient for the nontrading probability is positive and significant. This evidence suggests that the ETF risk can increase when the ETF is not traded actively in the market. This coefficient value is quite stable even if other control variables are included in column 4.

Columns 2 and 3 investigate the effect of the autocorrelation of the underlying index return on the ETF variance. First, both estimated coefficients in columns 2 and 3 are quite stable (0.017 and 0.016). Second, the existence of autocorrelation in the underlying index return also is positively related to the difference between ETF and NAV variances.

Column 4 considers all three variables: the nontrading probability, the autocorrelation, and the interaction term between the nontrading probability of the ETFs and the index autocorrelation. Consistent with the previous results, the nontrading probability and the autocorrelation are positively related to the difference between the ETF variance and the NAV variance.

Finally, column 5 reports the regression result when the related variables are transformed to follow the form in equation (69). After including the transformed variables in the regression, the nontrading probability term is still positive and significant but

Table 26: Nontrading probability and difference between ETF and NAV return variances

The dependent variable of this regression is the annual variance difference between ETF and NAV returns. p is the ratio of total observations to trading days. ϕ is the coefficient of the $AR(1)$ for the index. Other variables are the following: average dollar trading volume(Dollar Trading Volume), std. dev. of the index return(Index Volatility), log of average shares(Shares Outstanding), the std. dev. of shares' growth(Shares Volatility), a dummy of 1 if an ETF uses derivatives or swaps(Derivatives Based, Swap Based), a dummy of 1 if ETF underlying assets invest in the US(Invested in US assets),a dummy of 1 if an ETF has futures or options based on it (Futures Available , Options Available), a dummy of 1 if an ETF is levered(Levered Fund). The numbers in parentheses are t-statistics. "*", "**", "***" represents significance at 10%, 5%, and 1% level respectively.

	(I)	(II)	(III)	(IV)	(V)
Intercept	−0.057***	0.019	0.019	−0.061***	−0.028**
	(−4.280)	(1.460)	(1.450)	(−4.560)	(−2.090)
Notrading Prob(p)	0.103***			0.104***	
	(10.320)			(10.470)	
$AR(1)$ coefficient(ϕ)		0.017***	0.016***	0.023***	
		(3.210)	(3.020)	(5.420)	
$p * \phi$			0.020	0.008	
			(0.290)	(0.130)	
$p/(1-p)$					0.023***
					(6.830)
$p\phi/(1-p\phi)$					0.061
					(1.170)
Dollar Trading Volume	−0.003***	−0.004***	−0.004***	−0.003***	−0.004***
	(−4.690)	(−4.940)	(−4.930)	(−4.570)	(−5.060)
Shares Outstanding	0.004***	−0.001	−0.001	0.004***	0.002**
	(4.430)	(−0.840)	(−0.830)	(4.400)	(2.460)
Shares Volatility	−0.007	−0.034***	−0.034***	−0.008	−0.013
	(−0.530)	(−2.640)	(−2.640)	(−0.670)	(−1.030)
Derivatives Based	0.000	0.001	0.001	0.001	0.001
	(−0.040)	(0.180)	(0.170)	(0.170)	(0.090)
Swap-Based	0.008	0.002	0.002	0.007	0.006
	(0.900)	(0.210)	(0.220)	(0.780)	(0.630)
Equity-type ETF	0.004*	0.005**	0.005**	0.005**	0.004*
	(1.750)	(2.120)	(2.110)	(2.430)	(1.810)
Invested in US assets	−0.013***	−0.010***	−0.010***	−0.010***	−0.012***
	(−12.290)	(−8.150)	(−8.180)	(−9.170)	(−11.710)
Futures Available	−0.001	0.007***	0.007***	0.000	0.002
	(−0.520)	(4.340)	(4.340)	(−0.290)	(1.560)
Options Available	0.002*	0.002*	0.002*	0.002	0.002
	(1.670)	(1.710)	(1.720)	(1.490)	(1.470)
Levered Fund	−0.011	−0.015*	−0.015*	−0.011	−0.013
	(−1.480)	(−1.810)	(−1.810)	(−1.400)	(−1.570)
Year Fixed Effects	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	6061	6061	6061	6061	6061
Adjusted R^2	20.99	10.36	10.36	21.42	19.23

the interaction term is not significant.

In sum, these empirical findings support the relation shown in equation (69). The ETF variance can increase relative to the NAV variance when the ETF is not actively traded in the market. These results suggest that the lack of liquidity due to infrequent trading could increase the risk of ETF investment.

6 Conclusion

The ETF market has grown tremendously over the last two decades. ETFs are considered as being more transparent, less expensive, and more tax-efficient than traditional mutual funds. Moreover, ETFs provide investment opportunity to access other inaccessible markets or asset categories.

This paper investigates the effect of liquidity on the ETF return and variance. Similar to general common stocks, liquidity is an important risk factor affecting the ETF return and variance. Illiquid ETFs are more sensitive to the market return and market liquidity. The liquidity risk explains approximately 0.31% of the ETF returns annually. The level of liquidity is also related to the tracking error of the ETF with respect to the underlying index or the NAV.

Moreover, the lack of liquidity increases the ETF variance with respect to the NAV variance. Extending the Lo and MacKinlay (1990) econometric model to consider the autocorrelation of the underlying index return, the ETF variance can be decomposed into the NAV variance and the terms related to the nontrading probability. This finding implies that the variance of the ETF can increase when the ETF is traded infrequently. The calculated ETF variances are shown to be larger than the NAV variance. Moreover, the cross-sectional regression shows that the ETF variance is positively related to the nontrading probability.

ETFs are recognized as effective investment vehicles that provide the opportunity to access new markets. ETFs are designed to trace a particular index representing a particular market or sector. Therefore, an ETF must provide the same expected return as the return of the particular index. However, if ETFs have liquidity risks and thus substantial tracking error, investors may bear another risk in addition to the market risk. Therefore, investors must be cautious when investing in illiquid ETFs.

Appendix

A Proof of the variance

Under the $AR(1)$ process, the autocovariance of f_t is

$$Cov(f_t, f_{t-k}) = \phi^k Var(f_t) \tag{A.1}$$

For $l > k$,

$$\begin{aligned}
E[v_{t-k}v_{t-l}] &= E[(\alpha + \beta f_{t-k} + \xi_{t-k})(\alpha + \beta f_{t-l} + \xi_{t-l})] \\
&= \alpha^2 + 2\alpha\beta E[f_t] + \beta^2 E[f_{t-k}f_{t-l}] \\
&= \alpha^2 + 2\alpha\beta E[f_t] + \beta^2(E[f_t]^2 + \phi^{l-k}Var(f_t)) \\
&= (\alpha + \beta E[f_t])^2 + \beta^2\phi^{l-k}Var(f_t) \\
&= E[v_t]^2 + \beta^2\phi^{l-k}Var(f_t)
\end{aligned}$$

The second moment of the r_t is

$$\begin{aligned}
E[r_t^2] &= E\left[\sum_{k=0}^{\infty} X_t(k)v_{t-k} \sum_{l=0}^{\infty} X_t(l)v_{t-l}\right] \\
&= \sum_{k=0}^{\infty} E\left[X_t^2(k)v_{t-k}^2\right] + 2\sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty} E\left[X_t(k)X_t(l)\right] E\left[v_{t-k}v_{t-l}\right] \\
&= \left(Var(v_t) + E[v_t]^2\right)\sum_{k=0}^{\infty}(1-p)p^k + 2\sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty}(1-p)p^l\left(E[v_t]^2 + \phi^{l-k}\beta^2 Var(f_t)\right) \\
&= Var(v_t) + E[v_t]^2 + 2E[v_t]^2\sum_{k=0}^{\infty}p^{k+1} + 2\beta^2 Var(f_t)(1-p)\frac{\phi}{1-\phi p}\sum_{k=0}^{\infty}p^{k+1} \\
&= Var(v_t) + E[v_t]^2 + \frac{2p}{1-p}E[v_t]^2 + \frac{2\phi p}{1-\phi p}\beta^2 Var(f_t)
\end{aligned}$$

The expected return of the ETF return is simply

$$\begin{aligned}
E[r_t] &= E\left[\sum_{k=0}^{\infty} X_t(k)v_{t-k}\right] \\
&= \sum_{k=0}^{\infty} E[X_t(k)]E[v_{t-k}] \\
&= E[v_t]\sum_{k=0}^{\infty}(1-p)p^k = E[v_t]
\end{aligned}$$

So, the variance of the ETF return is

$$
\begin{aligned}
Var(r_t) &= E[r_t^2] - E[r_t]^2 \\
&= Var(v_t) + \frac{2p}{1-p}E[v_t]^2 + \frac{2\phi p}{1-\phi p}\beta^2 Var(f_t)
\end{aligned}
$$

REFERENCES

Acharya, V. V., Pedersen, L. H., 2005. Asset pricing with liquidity risk. Journal of Financial Economics 77, 375–410.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. Journal of Financial Economics 17, 223–249.

Bekaert, G., Harvey, C. R., Lundblad, C., 2007. Liquidity and expected returns: Lessons from emerging markets. Review of Financial Studies 20 (6), 1783–1831.

Ben-David, I., Franzoni, F., Moussawi, R., 2011. ETFs, arbitrage, and shock propagation. Working paper.

Boehmer, B., Boehmer, E., 2003. Trading your neighbors ETFs: Competition or fragmentation? Journal of Banking & Finance 27 (9), 1667–1703.

Borkovec, M., Domowitz, I., Serbin, V., Yegerman, H., 2010. Liquidity and price discovery in exchange-traded funds: One of several possible lessons from the flash crash. The Journal of Index Investing 1 (2), 24–42.

Brennan, M. J., Subrahmanyam, A., 1996. Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. Journal of Financial Economics 41, 441–464.

Cespa, G., Foucault, T., 2012. Illiquidity contagion and liquidity crashes. Working paper.

Chalmers, J. M. R., Kadlec, G. B., 1998. An empirical examination of the amortized spread. Journal of Financial Economics 48 (2), 159–188.

Cherkes, M., Sagi, J., Stanton, R., 2009. A liquidity-based theory of closed-end funds. Review of Financial Studies 22 (1), 257.

Chordia, T., Roll, R., Subrahmanyam, A., 2001. Market liquidity and trading activity. Journal of Finance 56, 501–530.

Eckbo, B. E., Norli, Ø., 2005. Liquidity risk, leverage and long-run IPO returns. Journal of Corporate Finance 11 (1), 1–35.

Elton, E., Gruber, M., Comer, G., Li, K., 2002. Spiders: Where are the bugs? The Journal of Business 75 (3), 453–472.

Engle, R., Sarkar, D., 2006. Premiums-discounts and exchange traded funds. The Journal of Derivatives 13 (4), 27–45.

Getmansky, M., Lo, A. W., Makarov, I., 2004. An econometric model of serial correlation and illiquidity in hedge fund returns. Journal of Financial Economics 74 (3), 529–609.

Huang, R. D., Stoll, H. R., 1994. Market microstructure and stock return predictions. Review of Financial Studies 7, 179–213.

Lee, K.-H., 2011. The world price of liquidity risk. Journal of Financial Economics 99 (1), 136–161.

Lo, A. W., MacKinlay, C. A., 1990. An econometric analysis of nonsynchronous trading. Journal of Econometrics 45, 181–211.

Madhavan, A., 2011. Exchange-traded funds, market structure and the flash crash. Working paper.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–08.

Pastor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

Petajisto, A., 2011. Inefficiencies in the pricing of exchange-traded funds. Working paper.

Pontiff, J., 1996. Costly arbitrage: Evidence from closed-end funds. The Quarterly Journal of Economics 111 (4), 1135–1151.

Roll, R., Schwartz, E., Subrahmanyam, A., 2007. Liquidity and the law of one price: The case of the futures-cash basis. The Journal of Finance 62, 2201–2234.

Sadka, R., 2010. Liquidity risk and the cross-section of hedge-fund returns. Journal of Financial Economics 98 (1), 54–71.