Exchange-Traded Funds and Information Asymmetry

Keebong Park

Doctor of Philosophy

Aston University

September 2008

ASTON UNIVERSITY
THESIS SUMMARY
Exchange-Traded Funds and Information Asymmetry
Keebong Park
Doctor of Philosophy
2008

This thesis focuses on three main questions. The first uses Exchange-Traded Funds (ETFs) to evaluate estimated adverse selection costs obtained spread decomposition models. The second compares the Probability of Informed Trading (PIN) in Exchange-Traded Funds to control securities. The third examines the intra-day ETF trading patterns. These spread decomposition models evaluated are Glosten and Harris (1988); George, Kaul, and Nimalendran (1991); Lin, Sanger, and Booth (1995); Madhavan, Richardson, and Roomans (1997); Huang and Stoll (1997). Using the characteristics of ETFs it is shown that only the Glosten and Harris (1988) and Madhavan, et al (1997) models provide theoretically consistent results. When the PIN measure is employed ETFs are shown to have greater PINs than control securities. The investigation of the intra-day trading patterns shows that return volatility and trading volume have a U-shaped intra-day pattern. A study of trading systems shows that ETFs on the American Stock Exchange (AMEX) have a U-shaped intra-day pattern of bid-ask spreads, while ETFs on NASDAQ do not. Specifically, ETFs on NASDAQ have higher bid-ask spreads at the market opening, then the lowest bid-ask spread in the middle of the day. At the close of the market, the bid-ask spread of ETFs on NASDAQ slightly elevated when compared to mid-day.

Key Words: Exchange-Traded Funds (ETF), Information-Asymmetry, Adverse-Selection, PIN, Intra-day

I would like to dedicate this thesis to my loving parents: Taeseob Park and Myungja Oh.

Acknowledgements

And I would like to acknowledge Prof. Patricia Chelley-Steeley for her supervision and guidance throughout this research.

I also would like to thank the examiners: external examiner Dr. Abimbola Adedeji and internal examiner Prof. Nathan Joseph, for their critical comments on the thesis. Additionally I am greatly indebted to Prof. Jim Steeley.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis is composed of two parts: the first half concerns the bid-ask spread; the second half contains the empirical results of the thesis. The first component is contained in Chapter Two to Chapter Four and provides a review of the bid-ask spread. The second component is contained in Chapter Five to Chapter Seven and highlights the empirical contribution of the thesis. Chapter Eight provides a summary of the contribution contained in the thesis and offers some concluding comments.

1.1 The Review of The Bid-Ask Spread

Chapter Two reviews the measure of the bid-ask spread because it is a key concept in the thesis. In economic theory, there is only one trade price, i.e. the price when quantity supplied is matched with quantity demanded. In practice, however, an ask price and a bid price exist. The ask price is the price which an agent wants to receive for the sale of the asset and the bid price is the price which an agent wants to pay for the purchase of the asset. When ask and bid prices exist, these prices generate three types of bid-ask spread. The quoted bid-ask spread stands for the difference between the bid price and the ask price. The effective spread represents the difference between the trade price and the mid-price of the quoted spread. The Roll implied spread is the effective spread and is developed by Roll (1984) to measure the fluctuation at transaction prices from the bid to ask and ask to bid. The realised spread computes the profits realised by dealers after a trade had taken place. All these bid-ask spreads represent transaction costs that agents face when they want to trade in an asset.

Chapter Three discusses why the bid-ask spread exists. Three reasons are suggested: the order processing cost, the inventory holding cost and the adverse selection cost. The order processing cost is due to the order-handling fee borne by the dealer. When the dealer receives buy/sell orders from the public, the dealer bears some costs such as administrative costs that reflect on the bid-ask spread as the order processing cost. The inventory holding cost arises due to the imbalanced inventory of the dealer. When the dealer provides the service of immediate trading, the dealer accumulates the unwanted inventory after trading. In the end, this unwanted inventory becomes the imbalanced inventory and the dealer runs out of cash. Thus, the dealer allows for the default risk caused by the imbalanced inventory. The adverse selection cost is due to the information asymmetry between the dealer and informed traders. This means that when informed traders know the intrinsic price of the asset, the dealer will face a trading loss due to the superior knowledge of the informed traders. To avoid this trading loss, the dealer needs to earn a bid-ask spread.

Chapter Four introduces spread decomposition models when the order processing, the inventory holding and the adverse selection costs exist. I focus on the five most popular spread decomposition models: Glosten and Harris (1988); George, Kaul, and Nimalendran (1991); Lin, Sanger, and Booth (1995); Madhavan, Richardson, and Roomans (1997); and Huang and Stoll (1997). Four models (not the George et al. (1991) model) use trade direction to estimate three components of the bid-ask spread; thus these models can be considered as trading indicator models. The George et al. (1991) model is based on the covariance of trades.

These five different spread decomposition models are employed in a broad area. One example is the impact of regulation change. i.e. how the changed regulation affects traders behaviour. When US Securities and Exchange Commission adopt a Regulation Fair Disclosure, there are some disputes over the impact of the Regulation Fair Disclosure. Thus, academic researchers employ spread decomposition models to investigate what impact Regulation Fair Disclosure have on traders. Academic researchers find that spread decomposition models do not provide consistent results about the impact of Regulation Fair Disclosure. This possibly suggests that spread decomposition models do not perform effectively.

Straser (2002) using the Huang and Stoll (1997) model show the adverse selection cost rises. Chiyachantana et al (2004) shows the adverse selection costs decline using the Lin, et al (1995) model, the Glosten and Harris (1988) model and the George, et al (1991) model.

Another conflicting result comes from the relation between institutional ownership and adverse selection cost. Sarin (2000) shows the adverse selection cost increases with institutional ownership using the Glosten and Harris (1988) model, but the adverse selection cost declines when the George et al (1991) model is used. Denis and Weston (2001) use the Huang and Stoll (1997) model to show the adverse selection costs increase but Jiang and Kim (2005) use the Glosten and Harris (1988) model and the Lin, et al (1995) model to show these costs decline.

All these conflicting results can be clear if we know how well the spread decomposition models estimate the adverse selection cost. i.e. chapter four reviews the use of spread decomposition models in financial research. This chapter shows that the estimated adverse selection cost varies with different spread decomposition models. This motivates the evaluation of spread decomposition models such as that undertaken by Neal and Wheatley (1998); Van Ness, Van Ness, and War (2001); De Winne and Majois (2003). However, none of these papers are able to suggest which models provide the most effective estimates of the adverse selection cost.

Therefore, my thesis attempts to evaluate these five spread decomposition models (Chapter Five). Secondly, I employ the probability of informed trading (Chapter Six). Final analysis is related with the intra-day behaviour of bid-ask spread. I also look into the intra-day behaviour of return volatility and trading volume (Chapter Seven). The intra-day behaviours are related with several factors; information asymmetry, inventory risk and difference in trading mechanism. I investigate what happens in ETFs in terms of these factors.

1.2 The Empirical Studies

In Chapter Five, I will employ Exchange-Traded Funds (ETFs) to evaluate spread decomposition models. Exchange-Traded Funds have unique features when compared to ordinary stocks. An ETF is a security that comprises ordinary stocks and the payoff of an ETF is identical to the weighted average payoffs of constituent stocks. The ETF is created/redeemed with the exchange of constituent stocks. Any traders (including a dealer) can know the constituent stocks of the ETF at any time. Moreover, the Net Asset Value (NAV) of ETFs is easily computed and every trader, including the dealer, uses the NAV to trade ETFs on the stock exchange. Thus, the lower information asymmetry of ETFs should lead to lower adverse selection costs of the spread.

Subrahmanyam (1991) and Gorton and Pennacchi (1993) support the hypothesis that is introduced later. Subrahmanyam (1991) shows that a basket security has lower information asymmetry than the weighted average of its constituent securities. This happens because the information asymmetry between constituent securities is diversified. Although the framework used by Subrahmanyam (1991) is one in which traders are risk neutral, Gorton and Pennacchi (1993) show similar results when traders are risk averse.

ETFs are best tools to employ the results of Subrahmanyam (1991) and Gorton and Pennacchi (1993). A reason is that the definition of ETFs is the same as the basket security in the Subrahmanyam (1991) and Gorton and Pennacchi (1993). This leads to the lower information asymmetry that ETFs have. In addition, the features of ETFs imply that ETFs have lower information asymmetry.

Using the results of Subrahmanyam (1991) and Gorton and Pennacchi (1993) and the features of ETFs, I evaluate spread decomposition models. My criterion for evaluation is that ETF should have lower adverse selection costs than control securities.

In Chapter Six, the Probability of Informed Trading (PIN) will be employed to see whether there are differences in the information asymmetry of ETFs and matched control securities. The PIN is different from the adverse selection cost of the spread in that the PIN is a probability measure. When the dealers face information asymmetry due to informed traders, the dealers need to update their probability of trading with informed traders. This is captured by the probability of informed trading which causes dealers to adjust their bid-ask spreads. After the dealer changes the bid-ask spread, the dealer possibly trades with uninformed traders. In this case, the dealer earns a trading profit that is the same amount as the adverse selection cost. However, when the dealer changes the probability of informed trading, after trading with informed traders, the dealer updates his/her probability.

The adverse selection cost and the PIN are complementary in that both measures are about information asymmetry between informed traders and a dealer. I used the same hypothesis in Chapter Six, i.e. I compare ETFs with matched control securities. If ETFs have lower information asymmetry than control securities I would anticipate that ETFs also have lower PINs.

Chapter Seven will be concerned with the intraday ETF bid-ask spreads, return volatility, and trading volume. The bid-ask spread varies as trading hours pass. On the New York Stock Exchange (NYSE)/the American Stock Exchange (AMEX), the intraday bid-ask spreads have been found to be U-shaped. Research on individual securities tends to show that at the market open or close the bid-ask spread is greater, but in the afternoon the bid-ask spread is smaller. Moreover, the intraday bid-ask spread on the NASDAQ (the National Association of Securities Dealers Automated Quotation) has different patterns. Studies on this market have shown that when the market is first open bid-ask spreads are greater, thereafter spreads start to decrease. When this market closes, the bid-ask spreads tend to decrease. Return volatility and trading volume tend to display U-shaped, being larger at the open and close than during periods of the day.

To date intraday trading pattern of ETFs have not been studied. I therefore study whether ETFs have the same pattern as common stocks. Then I go on to examine whether trading patterns are related to the exchange as seems to be the core for individual stocks.

In Chapter Eight I summarise the findings of my thesis and provide some conclusion to my study.

Chapter 2

Bid-Ask Spread

2.1 Bid-Ask Spread

The Bid-ask spread is the difference between ask and bid prices, as Figure 2.1 shows. The bid is the price at which a dealer[1] wants to purchase the underlying asset from the public while the ask is the price at which a dealer wants to sell the underlying asset to the public. Figure 2.1 provides an example. The ask price is 11.50, the bid price 11.00 giving rise to a bid-ask spread of 0.50, which is called the quoted spread.

The bid-ask spread[2] is usually interpreted as a transaction cost. In a complete world, the bid-ask spread need not exist. The reason is that in a complete market any buyer (seller) can purchase (sell) any amounts of the asset at the market price, which is determined by demand and supply schedules, at any time. However, in a real world, buyers (sellers) may not find proper sellers (buyers) due to the discordance of location or trading time, therefore any agent (buyer or seller) who has an urgent need to trade may be satisfied with the service of a dealer who transacts with the agent even though he/she may pay some charges for the service of a dealer. It is also possible in the real world that the exchange charges

[1]Dealer, market-maker, and specialist all have same meaning in this paper in that the main fundamental role is to facilitate the trading of a given asset. In a dealer market, the dealer sets bid and ask prices, absorbs the imbalance between supply and demand of an asset, as well as prevent the price of the asset from fluctuating widely.

[2]There are several types of bid-ask spread; the quoted spread, the effective spread, and the realised spread. All these types of bid-ask spread will be discussed in the later section.

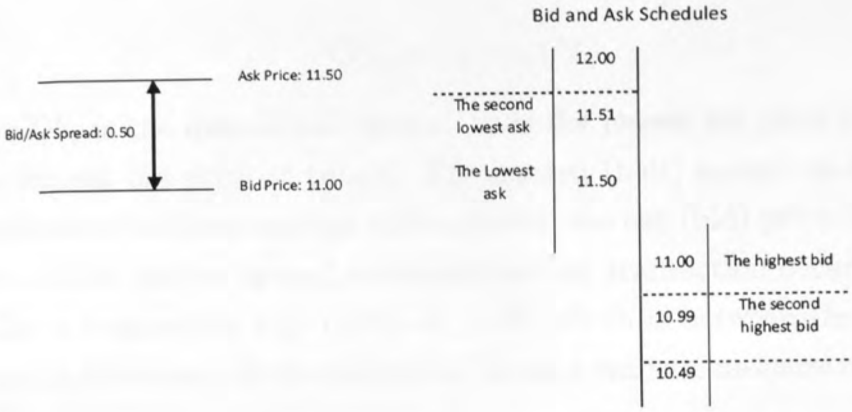


Figure 2.1: Bid-ask Spread

the dealer for using the exchange's facility and the dealer will need to recover the fee by posting the bid-ask spread to any buyer/seller. Academic researchers have questioned why the bid-ask spread arises in a real world and, if the bid-ask spread is an unavoidable transaction cost, then what causes the bid-ask spread to exist. Furthermore academic research focuses on how to measure the bid-ask spread as a transaction cost.

2.2 The Quoted Spread

The quoted spread is the difference between the highest posted bid price and the lowest posted ask price. In Figure 2.1, the quoted spread is 0.5. The quoted spread represents the worst transaction price in the view of an agent (buyer or seller). When an agent buys an asset and sells it immediately, the agent will have to pay the ask price for a buy and receive the bid price for a sale. Thus transacting provides a loss equal to the quoted spread for the agent. Therefore, the quoted spread is sometimes called "the cost of immediacy" and implies a two trade (buy and sell). The quoted half spread is widely used to reflect the cost of one trade. The quoted half spread is computed as

$$QS_t = (a_t - b_t)/2 \tag{2.1}$$

where QS_t is the quoted half spread, a_t is the lowest ask price at time t and b_t is the highest bid price at time t. The quoted (half) spread as a transaction cost implies that each transaction takes place at the ask (bid) price. For instance, in Figure 2.1 the quoted spread assumes that any transaction occurs at 11.50 or 11.00. But a transaction may occur at 11.25, which is between the ask and the bid prices. In this case, the quoted spread is not a realistic measure of transaction costs.

The quoted spread is broadly used as a representative transaction cost. For example, Huang and Stoll (1996), Battalio, Hatch, and Jennings (2004) and Barclay and Hendershott (2004) employ the quoted spread to measure transaction costs. Huang and Stoll (1996) use the quoted spread to compare execution costs[1] on the New York Stock Exchange (NYSE) and NASDAQ[2].

Battalio, Hatch, and Jennings (2004) use the quoted spread to examine how each equity option market integrates a national option market system after US Securities and Exchange Commision (SEC) mandated each option markets to electronically link other option markets[3]. They find that market quality is improved between two sample periods. Barclay and Hendershott (2004) use the quoted spread to analyze trading costs after regular trading hours, and find that the reduced trading activity after regular trading hours leads to higher trading costs.

[1]Execution costs are measured by the quoted spread, the effective spread, the realised spread, the Roll (1984) implied spread, and the post-trade variability.

[2]NADSAQ is symbol of National Association of Securities Dealers Automated Quotations. NADSAQ is a US electronic stock market.

[3]Thus most actively traded equity options are listed on more than one options exchange, but previously they are not.

2.3 The Effective Spread

The effective half spread is the absolute difference between the transaction price and the quote mid-point. This is expressed by

$$ES_t = |P_t - M_t| \tag{2.2}$$

where ES_t is the effective half spread at time t, P_t is the transaction price at time t and M_t is the quote midpoint at time t, computed as $(a_t + b_t)/2$. The effective half spread considers the fact that transactions occur not only at the quoted ask (bid) price but also within the quoted spread. Thus the effective half spread is more likely to reflect the true cost incurred by a trader. The effective half spread seems to represent the extent that transaction price deviates from the true value of the asset when the quote midpoint M_t is regarded as the true value of the asset.

The effective spread is deemed to measure better the cost of immediacy and more practically the cost of transaction. The argument that the effective spread is a better measure than the quoted spread is empirically supported. Blume and Goldstein (1992) find that 12% to 31% of the trades in their NYSE and Non-NYSE samples occur inside the quoted spread. Petersen and Fialkowski (1994) show that the effective half spread is almost always less than the quoted half spread, suggesting that the quoted spread overstates trading costs. Thus the effective half spread is expected to be less than the quoted half spread and seems to reflect better the true costs borne by traders. It is common that studies of market frictions examine both the quoted half spread and the effective half spread.

Huang and Stoll (1996) , Battalio, Hatch, and Jennings (2004) , and Barclay and Hendershott (2004)[1] and so on employ the effective spread. For example, Huang and Stoll use the effective spread as one example of execution costs. Huang and Stoll suggest that NASDAQ stocks have double the execution costs of NYSE stocks. Battalio, Hatch, and Jennings (2004) compare effective option spreads to evaluate liquidity costs across option exchanges. Battalio, Hatch, and Jennings

[1]These papers used the quoted spread also. See the quoted spread section.

(2004) find that the effective spreads of options decrease by over 60% from June 2000 to January 2002 and that differences in effective spreads across exchanges in June 2000 disappear by January 2002. Barclay and Hendershott test liquidity externalities by focusing on trading costs after trading hours. They find the effective spreads are three to four times larger after trading hours than during regular trading hours.

2.3.1 The Roll Implied Spread

Roll (1984) shows how the effective spread can be estimated from transaction level data. When bid and ask quotes are not available, the Roll implied spread can be used to estimate the spread.

The asset is assumed to be traded in an informationally-efficient market. Thus only new information can move transaction prices and the arrival of new information is randomly distributed. Thus, bid and ask quotes are not changed when trades do not have any new information content.

$$P_t - P_{t-1} = e_t \tag{2.3}$$

Roll assumes that the probability distribution of observed price changes is stationary at least over short intervals. In addition, with no new information about the asset and with a random order arrival rate, the probability of arriving a buy (sell) trade is equally likely. The probability of a buy order is independent of the previous trade, while the probability that the trader is a buyer is independent of e_t.

2.3.1.1 The model

Figure 2.2 and Figure 2.3 illustrate the possible paths of observed market prices in two successive time periods, given that no new information arrives in the market. Figure 2.2 highlights the case that a buy order arrives at time $t-1$ and a transaction occurs at the ask price. Figure 2.3 illustrate the case that a sell order arrives at time $t-1$ and the transaction occurs at the bid price. In the figure, s represents the bid-ask spread. Each path has a probability of one half,

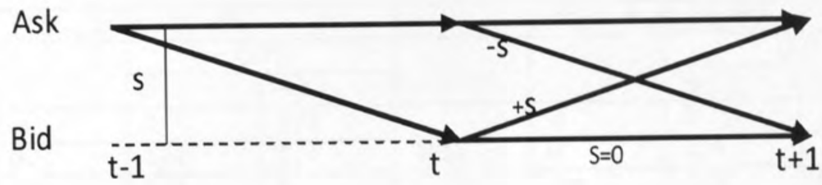


Figure 2.2: The Movement of Trade from Ask price

implying that after a transaction at time $t-1$ at the bid price, a transaction at t is equally likely to occur at the ask or bid price. While a transaction at time $t+1$ also occurs equally at the ask or at the bid. Thus if transactions both at time $t-1$ and at time t occur at the bid, the transaction price change at t is zero.

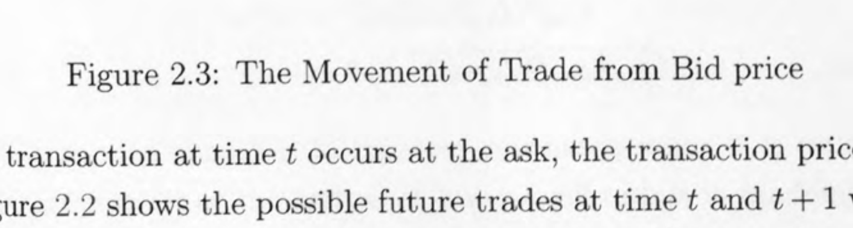


Figure 2.3: The Movement of Trade from Bid price

When transaction at time t occurs at the ask, the transaction price change at t is s. Figure 2.2 shows the possible future trades at time t and $t+1$ when a buy order arrives at time $t-1$ and a transaction occurs at the ask price.

The possible transaction paths in Figure 2.2 and Figure 2.3 are simplified in Table 2.1, which provides the joint probability of successive price changes ($\Delta P_t \equiv P_t - P_{t-1}$ and $\Delta P_{t+1} \equiv P_{t+1} - P_t$).

When a trade at $t-1$ is equally likely at the bid or the ask, Table 2.2 shows the combined joint distribution of successive price changes (ΔP_t and ΔP_{t+1})

Table 2.1: The Probability between P_t and P_{t+1} Depending on Ask or Bid

P_{t-1} is at the bid				P_{t-1} is at the ask			
		ΔP_t				ΔP_t	
		0	+s			-s	0
	-s	0	1/4		-s	0	1/4
P_{t+1}	0	1/4	1/4	P_{t+1}	0	1/4	1/4
	+s	1/4	0		+s	1/4	0

Table 2.2: The Combined Probability between ΔP_t and ΔP_{t+1}

		ΔP_t		
		-s	0	+s
	-s	0	1/8	1/8
ΔP_{t+1}	0	1/8	1/4	1/8
	+s	1/8	1/8	0

If one computes the covariance between successive price changes, then

$$Cov(\Delta P_t, \Delta P_{t+1}) = 1/8(-s^2 - s^2) = -s^2/4 \tag{2.4}$$

The middle row and middle column can be ignored[1]. Additionally, the variance of ΔP_t is $1/2 * s^2$ and the autocorrelation coefficient is -1/2.

The Roll implied spread is computed from equation (2.4).

$$\begin{aligned} -s^2/4 &= Cov(\Delta P_t, \Delta P_{t+1}) \\ s &= 2\sqrt{-Cov(\Delta P_t, \Delta P_{t+1})} \end{aligned} \tag{2.5}$$

The Roll implied spread is interpreted cautiously when it is estimated. The Roll model assumes that the arrival of buy or sell orders is random so that the probability of an order being a buy or sell is one half. However, this assumption is not supported by the empirical finding of Garbade and Lieber (1977). Garbade and Lieber found that buy orders seem to follow buy orders and sell orders follow sell orders over short intervals of time, suggesting positive serial correlation of

[1]The reason is that the middle row of ΔP_t and the middle column of ΔP_{t+1} in Table 2.2 are zero.

transaction types. Therefore Choi, Salandro, and Shastri (1988) extend the Roll (1984) model by relaxing the assumption of serial independence.

2.3.2 The Extension of the Roll Model by Choi, Salandro, and Shastri(1988)

Choi, Salandro, and Shastri (1988) denote δ as the conditional probability that successive transactions at time $t+1$ and at time t take place at the same bid or ask prices. When δ is defined as the conditional probability of the successive transactions, Choi, et al. (1988) show that Table 2.1 changes into Table 2.3, which represents the probability distribution of successive price changes.

Table 2.3: The Probability between ΔP_t and ΔP_{t+1} in Choi, Salandro, and Shastri (1988)

P_{t-1} is at the bid				P_{t-1} is at the ask			
		ΔP_t				ΔP_t	
		0	+s			-s	0
ΔP_{t+1}	-s	0	$(1-\delta)^2$	ΔP_{t+1}	-s	0	$(1-\delta)\delta$
	0	δ^2	$(1-\delta)\delta$		0	$(1-\delta)\delta$	δ^2
	+s	$(1-\delta)\delta$	0		+s	$(1-\delta)^2$	0

Table 2.3 gives rise to a combined joint distribution of price changes as shown in Table 2.4.

Table 2.4: The Combined Probability in Choi, Salandro, and Shastri (1988)

		ΔP_t		
		-s	0	+s
ΔP_{t+1}	-s	0	$(1-\delta)\delta/2$	$(1-\delta)^2/2$
	0	$(1-\delta)\delta/2$	$\delta 2$	$(1-\delta)\delta/2$
	+s	$(1-\delta)^2/2$	$(1-\delta)\delta/2$	0

Given the joint distribution in the Table 2.4, the covariance between successive price changes is computed as

$$\begin{aligned} Cov(\Delta P_t, \Delta P_{t+1}) &= -s^2 * (1-\delta)^2/2 - s^2 * (1-\delta)^2/2 \\ &= -s^2 * (1-\delta)^2 \end{aligned} \tag{2.6}$$

Since the middle column of ΔP_t and the middle row of ΔP_{t+1} are still zero, the middle row and column in the Table 2.4 are excluded.

The covariance between successive price changes becomes

$$\begin{aligned} -s^2 * (1-\delta)^2 &= Cov(\Delta P_t, \Delta P_{t+1}) \\ s &= \sqrt{-Cov(\Delta P_t, \Delta P_{t+1})}/(1-\delta) \end{aligned} \tag{2.7}$$

Equation 2.7 reduces to Equation 2.5 when one assume δ =0.5. Thus the Roll implied spread reformulated by Choi, et al. (1988) suggests that when positive serial correlation in transactions exists, s in Equation 2.5 is smaller than s in Equation 2.7, implying that the Roll measure is a downward-biased estimate of the spread.

Though Choi, et al. (1988) improves the original Roll measure of the spread, both spread models do not assume new information arrival in the transaction process whereas in a reality new information arrives at the market continuously. This may cause these models to understate the true cost of trading.

The Roll measure, however, is widely used in the finance to measure a trading cost and is particularly useful when only transaction data is available. Huang and Stoll (1996) use the Roll implied spread as one measure of execution costs when they compare execution costs of NYSE and those of NASDAQ.

2.4 The Realised Spread

The realised spread is defined as the signed difference between the transaction price at time t and the transaction price observed after τ[1] minutes following the

[1] It is possible to use other time interval such as 30 minute or 10 minute or some other interval. Huang and Stoll(1996) use different time periods such as five minute, five to ten minutes, 30 minutes, or 35 minutes.

trade at t, as is shown in the Equation 2.8. Put another way, the realised spread is defined as the difference between the transaction price and the quote mid-point of the asset after trade, and this definition is expressed in Equation 2.9.

$$\begin{aligned} (RS_\tau|b_t) &= [(P_{t+\tau} - P_t)|(P_t = b_t)] \quad \text{for trades at the bid} \\ (RS_\tau|a_t) &= [-(P_{t+\tau} - P_t)|(P_t = a_t)] \quad \text{for trades at the ask} \end{aligned} \tag{2.8}$$

or

$$\begin{aligned} (RS_\tau|b_t) &= (M_{t+\tau} - b_t) \quad \text{for trades at the bid} \\ (RS_\tau|a_t) &= (a_t - M_{t+\tau}) \quad \text{for trades at the ask} \end{aligned} \tag{2.9}$$

where RS_t is the realised half spread, t is the time of the transaction and τ is the length of time after the transaction, a_t is the ask price at time t, b_t is the bid price at time t, and M_t is the quote midpoint at time t.

The realised spread provides an ex post measure of profitability for the dealer since it measures how much transaction prices change after a buy (sell) transaction occurs. The difference between the effective half spread and the realized spread is the estimated loss of the dealer to informed traders. If an informed trader does not initiate trading, then transaction prices fluctuate within or at the spreads, causing the realised spread to be greater than the effective spread. As a result, the difference between the effective half spread and the realised spread has been used to measure the information content of a trade. Barclay and Hendershott (2004) employ the realised spread to show that there is more information asymmetry after hours than during the trading day. Hendershott and Jones (2005) use the effective half spread and the realised spread to show that the change in transparency on Island Exchange caused an increase in the effective half spread and the realised spread.

To summarise, the spread can be defined in different ways to measure transaction costs. Figure 2.4 details a range of spread measures and shows the difference between them. QS_1 means the Quoted spread at time t_1: ES_2 the effective half spread at time t_2: $RS_3$the realised half spread at time t_3. Mid Point is the quote mid point, TP_2 and TP_3 represent transaction prices at time t_2 and

t_3 respectively. $\updownarrow$ symbolizes the quoted spread, dotted $\uparrow$ the realised half spread, and half-dashed $\uparrow$ the effective half spread. All these spreads attempt to estimate transaction cost that the dealer incurs to provide liquidity in the market.

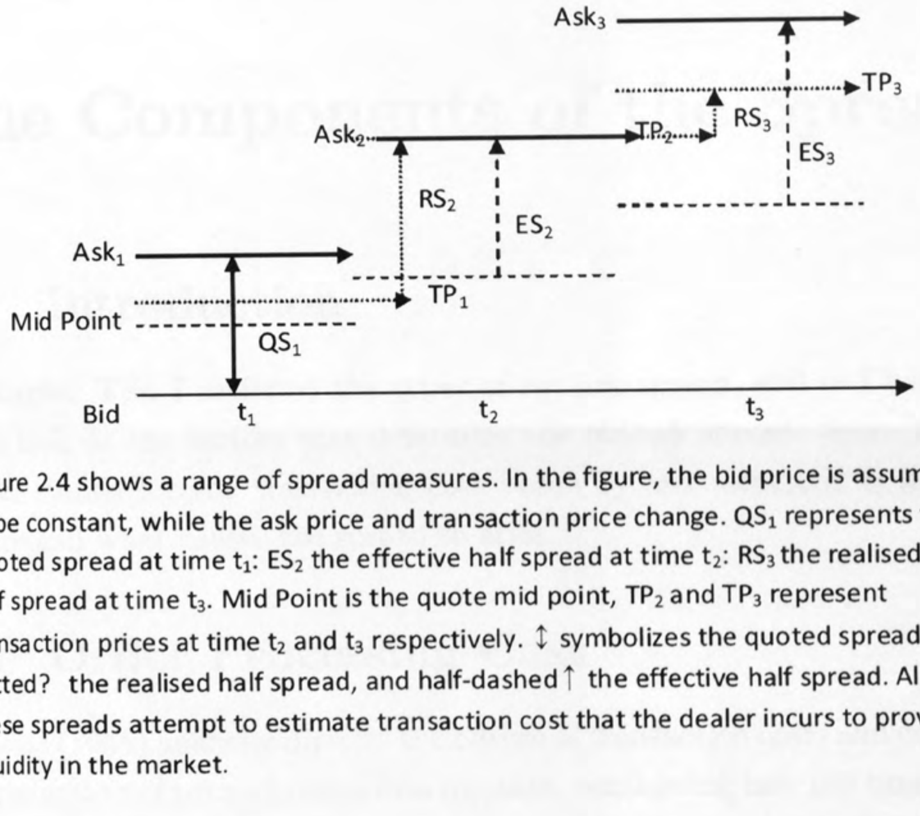


Figure 2.4: Types of Spread

Chapter 3

The Components of the Spread

3.1 Introduction

In Chapter Two I reviewed the types of bid-ask spread, and in Chapter Three I will look at the factors that determine the bid-ask spread. Since the bid-ask spread stands for the transaction cost borne by any trader, it is necessary to understand what causes the spread to arise.

3.2 Order Processing Cost

Demsetz (1968) analyzes directly the nature of transaction costs and examines the determination of prices in securities markets, considering how the time dimension of supply and demand affects market prices. Demsetz (1968) believes that trading activity involves two types of transaction costs, explicit and implicit costs. Explicit costs include brokerage fees or exchange membership fees. These costs are regarded as order processing costs. These explicit costs are borne by the dealer and then the dealer transforms these explicit costs into the bid-ask spread borne by traders. The implicit cost is caused by the need for immediate execution of transactions. i.e. any traders who want immediate execution may bear the cost of immediacy, which is the bid-ask spread.

Figure 3.1 illustrates why the bid-ask spread is called the cost of immediacy and regarded as an order processing cost. In Figure 3.1, D and S illustrate the

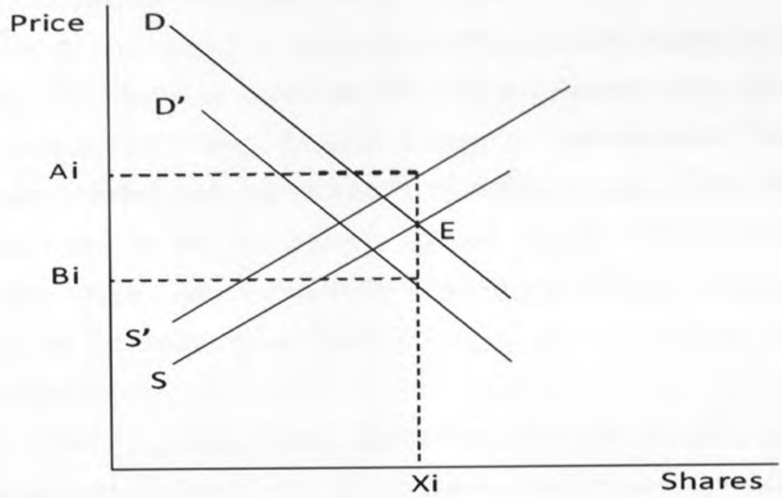


Figure 3.1: The Demand and Supply of Shares

demand and supply flows per unit of time for security i. Intersection E shows the equilibrium price at which public traders transact their assets when sell orders match buy orders or vice versa. However, if buyers/sellers who want to trade X shares of security i do not come to the market simultaneously, then buyers/sellers must wait until the counterpart sellers/buyers arrive at the market. There can be no transaction in cases where the amounts of shares which sellers want to sell do not match the amounts of shares which buyers want to buy. These imbalances in terms of time and amount can lead to a trading-time delay of minutes, hours, or even days.

This delay is resolved by the presence of a dealer, who executes immediately trades which public buyers/sellers want to commit to. The dealer stands ready to sell/buy at posted prices immediately upon receipt of a customer order to buy/sell. To cover the cost of executing a customer order immediately, the dealer is willing to buy X at a price that is slightly below E, Curve D', and to sell X at a price that is slightly above E, Curve S' in the Figure 3.1. The difference between these two prices, Ai and Bi, becomes the bid-ask spread. Thus, the bid-ask spread arises in the executing process of trades.

There are two major contributions of Demsetz (1968). First, Demsetz shows that the bid-ask spread occurs in the process of executing immediately a customer buy/sell order. So Demsetz considers the bid-ask spread as a transaction cost necessary to execute an order. Second, Demsetz demonstrates that the specific structure of the market can affect the transaction price. Since the number of traders is important to set the bid-ask spread, higher volume leads to smaller spread and vice versa. More generally the actual trading mechanism used to set prices such as the exchange is likely to affect the equilibrium price itself and should be analyzed.

Therefore Demsetz (1968) views the order executing procedure as one cause of the bid-ask spread. Later Stoll (1978a) breaks the bid-ask spread up into three sources: order processing, inventory holding, and adverse selection costs. In Stoll (1978a), the order processing cost is confined to costs reflecting the nature of the trading mechanism. An example of the order processing cost is the exchange membership fee, the cost of labour and capital to provide quote information, order routing, execution and clearing etc. Stoll, like Demsetz (1968), argues that if the dealer has to pay the order processing cost then the dealer will be compensated by the bid-ask spread since the dealer buys an asset at a lower value or sells it at a higher value.

3.3 Inventory Holding Cost

3.3.1 Garman(1976)

Garman (1976) posed the question of how a dealer can survive in the long run when he/she faces short-run quantity imbalances between customer buy and sell orders. Garman (1976) concludes that the dealer needs to set bid and ask prices in order to avoid potential failure. However, he also shows that posting the bid and ask prices does not guarantee the survival of the dealer. In his context, the dealer's failure happens when the dealer runs out of either cash or inventory.

The Garman (1976) model has six main assumptions:

(1) Arrivals of buy and sell orders to the market are Poisson distributed in time, with stationary rate functions $\lambda_A(p)$ and $\lambda_B(p)$, and q(order quantity) is equal to one.

(2) All transactions are made through a single central dealer.

(3) the dealer sets a price p_A at which he will fill buy orders and correspondingly a price p_B for sell orders, yielding the reluctant order rates $\lambda_A(p)$ and $\lambda_B(p)$, respectively.

(4) At time 0, the dealer has cash and stock inventories of $I_c(0)$ and $I_s(0)$, respectively.

(5) The dealer seeks to maximize expected profit per unit time, subject to the avoidance of certain ultimate failure.

(6) There are no transaction costs for the dealer.

A key feature of the Garman (1976) model is that the order flow from individual traders is viewed independently from the trading desires of them. This feature enables the interaction between the dealer and the behaviour of the order flow to be analysed. The first assumption means that uncertainties in the Garman (1976) model arise from the arrival of the buy and sell orders and, by the same token, the order arrival is stationary but not identically stochastic. Finally, the dealer is confronted with the significant problem of balancing the level of stock/cash inventory to avoid running out of it(see Appendix 3. A).

The Garman (1976) model shows that, to avoid certain failure, the dealer must set an ask price p_a and a bid price p_b. When the ask price and the bid price simultaneously satisfy two conditions which are both $p_a{*}\lambda_a(p_a) > p_b{*}\lambda_b(p_b)$ and $\lambda_b(p_b) > \lambda_a(p_a)$ (to see this result, see Appendix 3.A), the dealer avoids certain failure. These price requirements lead to the bid-ask spread, which means that the dealer sets a lower price when he/she buys stock and a higher price when he/she sells. This implies that the spread is an inherent property of trading.

While Garman (1976) argues that the inventory of the dealer determines his/her viability, inventory per se in the Garman (1976) model does not play any role in the dealer's decision problems. This limitation leads to an extension of Garman (1976) by Amihud and Mendelson (1980).

3.3.2 Amihud and Mendelson(1980)

Amihud and Mendelson (1980) extend the Garman (1976) model and approach the dealer's problem with the question of how the dealer's prices do change as his/her inventory position varies over time.

To answer this question, Amihud and Mendelson (1980) rewrite one of the assumptions of the Garman (1976) model. The first assumption[1] in the Garman (1976) model is rewritten as follows:

(1') For a given pair of prices, P_a and P_b, the next incoming order will be a buy order with probability $\mathrm{D}(P_a)/(\mathrm{D}(P_a) + \mathrm{S}(P_b))$ [2], or a sell order with probability $\mathrm{S}(P_b)/(\mathrm{D}(P_a)+\mathrm{S}(P_b))$. The time till the next arriving order has an exponential distribution with mean $1/(\mathrm{D}(P_a)+\mathrm{S}(P_b))$.

This modified assumption implies that the process of inventory development due to the discrepancy between buy and sell orders becomes a birth-and-death process[3]. The birth-and-death process is transformed into a semi-Markov decision process where the state variable is the inventory of the dealer, and the decision made for a given inventory level J is a pair of prices, P_{aj} and P_{bj}, which determine the respective buy and sell order rates. In the Amihud and Mendelson model, the bid and ask prices set by the dealer depend on the level of the dealer's inventory position and are adjusted over time.

Additionally, the inventory is assumed to be bounded above and below by exogenous parameters, L and -K, respectively. This assumption is required to avoid the possible failure of the dealer. Second, the dealer's revenue and cost functions, $R(\cdot)$ and $C(\cdot)$, respectively, are twice continuously differentiable when (i) $R(\cdot)$ is strictly concave, (ii) $C(\cdot)$ is strictly convex, and (iii) $R'(0) > C'(0)$, $R'(\infty) < C'(\infty)$.

[1]Arrivals of buy and sell orders to the market are Poisson distributed in time, with stationary rate functions $\lambda A(p)$ and $\lambda B(p)$, and q(order quantity) is equal to one.

[2]$\mathrm{D}(P_a)$ and $\mathrm{S}(P_b)$ represent buy and sell order arrival rates, respectively, and are in the forms of λA(p) and λB(p) in the Garman model.

[3]The birth and death process is one of Markov process. In the birth and death process, the transitions are limited to the nearest neighbours. For a population process, birth is the transition toward increasing the population by 1 while death is the transition towards decreasing the population size by 1.

The Amihud and Mendelson (1980) model has three main results; First, the optimal bid and ask prices decrease monotonically with the dealer's inventory position. Second, regardless of what is expected to happen to the value of the stock, the dealer has a preferred inventory position. Finally, the optimal bid and ask prices display a positive spread. The Amihud and Mendelson (1980) model suggests clearly that the bid-ask spread arises due to the inventory of the dealer and the bid-ask spread varies with the inventory level. Furthermore Amihud and Mendelson (1980) show that if there is competition among dealers, competition forces the market-making service to be priced in the form of cost. If the cost of providing the market-making service is zero, the competition among dealers leads to a zero spread. This prediction, though, is contrary to the findings of Laux (1995) when competition among dealers exists. Laux (1995) investigates the linkages amongst the number of dealers in a security, the extent of outside competition (non-dealer[1]), and the market bid-ask spread for NASDAQ stocks. He finds that the stocks with little outside competition have large spread even though the number of dealers is larger than for comparable stocks. This finding suggests that even though competition amongst dealers exists, the market spread is not zero, implying that the cost providing market-making service is not zero with the presence of competition among dealers.

The focuses of both Garman (1976) and Amihud and Mendelson (1980) are on the dealer's decision problem, in which the dealer chooses the optimal pricing strategy to maximize his/her profit and the outcome of his/her optimizing behaviour is security price. However, both models do not recognise how the dealer is rewarded for providing his/her services, which is the topic of Stoll (1978b).

3.3.3 Stoll(1978)

Stoll (1978b) focuses on how the dealer determines appropriate compensation to recover the costs he/she faces in providing market-making services, which is the notion of the dealer as a supplier of immediacy.

The assumptions Stoll (1978b) used are as follows:

[1]In Laux paper, outside competitors means the one who competes with dealers for providing liquidity in a security. The example of outside competitors is institutional investors.

(1) Dealer inventory positions acquired in the process of providing immediacy are financed solely at the risk free rate of interest, R_f.

(2) The dealer has a utility function for terminal wealth.

(3) The dealer estimates the true price and the true rate of return that would exist in the absence of transaction costs.

(4) The dealer makes one transaction per trading interval during which the stock's price does not change.

With the assumptions above, the dealer has a preferred portfolio which is composed of cash and stocks. The dealer hopes that the expected utility of terminal wealth of the initial and new portfolios be the same, which is represented as

$$EU(\tilde{W}^*) = EU(\tilde{W}) \tag{3.1}$$

where $\tilde{W}^*$= terminal wealth of the initial portfolio and $\tilde{W}$= terminal wealth of the new portfolio after the transaction

Simply $\tilde{W}^*$=W_0(1+$\tilde{R}^*$), W_0 is initial wealth and $\tilde{R}^*$ is the rate of return on the initial portfolio.

Terminal wealth under the new portfolio is given by:

$$\tilde{W} = W_0(1 + \tilde{R}^*) + (1 + \tilde{R}_i)Q_i - (1 + R_f)(Q_i - C_i) \tag{3.2}$$

where

Q_i = true dollar value of the transaction in stock i, the stock in which immediacy is being provided. Negative values indicate a sale; positive values a purchase.

$\tilde{R}_i$= rate of return on stock i.

C_i = present dollar cost to the dealer of trading the amount Q_i. C_i is positive or negative according as the transaction in stock i raises or lowers the costs of holding the inventory Q_p.

Meanwhile, Equation (3.1) becomes

$$E[U(W_0(1+\tilde{R}^*))] = E[U(\tilde{W})]. \tag{3.3}$$

Using a Taylor series around the respective means, dropping terms of order higher than two and setting $R_f = 0$ yields the dollar holding cost function(for further proofs, see Appendix 3.B).

$$C_i = \frac{z}{W_0}\sigma_{ip}Q_pQ_i + \frac{1}{2}\frac{z}{W_0}\sigma_i^2Q_i^2 \tag{3.4}$$

where z is the dealer's coefficient of relative risk aversion, Q_p is the true dollar value of stocks held in the dealer's trading account, σ_{ip} is the correlation between the rate of return on stock i and the rate of return on the optimal efficient portfolio, and σ_i^2 is the variance of stock i's return.

The dollar holding cost function refers to the cost of a single additional transaction undertaken in the trading interval.

The percentage cost function is easily found in the form of Equation (3.5).

$$\frac{C_i}{Q_i} = \frac{z}{W_0}\sigma_{ip}Q_p + \frac{1}{2}\frac{z}{W_0}\sigma_i^2Q_i = c_i(Q_i) \tag{3.5}$$

These two cost functions ((3.4) and (3.5)) suggest factors on which the cost of providing immediacy (C_i or C_i/Q_i) depends. The first factor is dealer characteristics, that is, the dealer's wealth and risk preferences. Greater initial wealth of the dealer W_0 reduces the cost of immediacy, while greater risk aversion z increases it. The second factor is the transaction size in stock i Q_i. The percentage cost increases linearly with Q_i and the dollar holding cost does with the square of Q_i. The third factor is the characteristics of the stock, that is, the variance of stock return σ_i^2 and the correlation between the return on stock i and the return on the initial trading account portfolio σ_{ip}. The fourth factor is the size of the initial position in the trading account Q_p. If Q_p is positive, the cost of buying stock i is larger than if there were no initial position. Conversely the cost of selling stock i is smaller than if there were non initial position.

Stoll (1978b) also looks into the determinants of bid and ask prices. In the Stoll (1978b) model, if the market is competitive, bid and ask prices are the only way to compensate a dealer for providing immediacy. i.e. the dealer is compensated by purchasing stocks at the bid price which is below the true price and by selling stocks at the ask price which is above the true price.

If the bid price is expressed in percentage terms relative to the true price P_i^*, then the optimal bid price for a transaction with true value Q_i^b is $\frac{(P_i^*-P_b)}{P_i^*} = c_i(Q_i^b)$. The corresponding optimal ask price for Q_i^a is $\frac{(P_i^*-P_a)}{P_i^*} = c_i(Q_i^a)$.

The resultant percentage spread is the percentage difference between bid and ask prices. The spread is

$$\frac{(P_a - P_b)}{P_i^*} = c_i(Q_i^b) - c_i(Q_i^a) = \frac{z}{W_0}\sigma_i^2|Q_i| \quad \text{for} \quad Q_i^a = Q_i^b = |Q_i|. \tag{3.6}$$

The resultant percentage spread suggests interesting features. The first is that the spread rises as trade size Q_i does. That the spread function 3.6 does not include dealer's inventory except for initial wealth W_0 means that the transaction done by the dealer causes a change in the dealer's inventory, but does not affect the spread size. The second is that the inventory of the dealer after transaction Q_p would affect bid and ask prices(see Equation (3.5)). This means that a large positive inventory leads to higher costs for providing more immediacy, which lowers both bid and ask prices[1]

The Stoll (1978b) model does not consider several things. First, Stoll (1978b) assumes that the stock is liquidated at time 2, so the dealer is not confronted with any uncertainty over how long he/she must hold inventory position. Ho and Stoll (1981) attempt to resolve this question by extending the Stoll (1978b) model to a multi-period framework in which both order flow and portfolio returns are stochastic. Another limitation is the assumption about risk aversion of the dealer. If the dealer were risk-neutral or able to diversify, then the cost of providing immediacy will fall. If there are differences in spreads between markets, the Stoll (1978b) model explains these differences by their risk-bearing ability. However,

[1]In the Equation 3.5, greater Q_p leads to higher cost and higher $c_i(Q_i^a)$ and $c_i(Q_i^b)$ cause the optimal bid and asks to lower through $\frac{(P_i^*-P_b)}{P_i^*} = c_i(Q_i^b)$ and $\frac{(P_i^*-P_a)}{P_i^*} = c_i(Q_i^a)$.

risk-bearing ability can not provide the entire explanation. In addition, the Stoll (1978b) model does not explain why spread for a stock varies during trading hour in a day.

3.3.4 Ho and Stoll(1981)

Ho and Stoll (1981) have fundamentally the same economic setting as the Stoll (1978b) model except that a multi-period framework is taken into consideration. In the Ho and Stoll (1981) model, the dealer maximises his/her expected utility of terminal wealth by adjusting bid and ask prices through time while he/she faces stochastic returns and stochastic transactions.

The assumptions used in the Ho and Stoll (1981) model are the following:

(1) Transactions evolve as a stationary continuous time stochastic jump process which is a Poisson Process. Buy order process and sell order process are independent. The arrival rate of buy orders(dealer sell) λ_a or sell orders(dealer buy) λ_b depends on the dealer's ask price or bid price, respectively.

(2) A time horizon is finite. At the end of time horizon, the inventory of the dealer is liquidated at the fixed price p, which is the true value of the stock.

(3) The dealer trades only with the public in a passive way. This leads to uncertainty about the timing of subsequent transactions.

(4) The dealer determines immediacy cost b if a customer sale order arrives, and immediacy cost a if a customer buy order arrives. Thus, the quoted ask price p_a is $p + a$ and the quoted bid price p_b is $p + b$[1].

(5) The dealer faces uncertainty over the future return on his existing portfolio X. In the absence of transactions, portfolio growth dX is represented by $dX = r_x X dt + X dZ_x$, where r_x is the mean return per unit time; dZ_x is a non-standard Wiener process with mean zero and instantaneous variance σ_x^2.

As a goal, the dealer maximises the expected utility of his/her total wealth $EU(W_T)$ at time T, where $W_T = F_T + I_T + Y_T$ by choosing bid and ask prices.

[1] In the Ho and Stoll model, the bid-ask spread is defined as the sum of immediacy price like $a + b$.

Changes in cash account occur when the dealer sells or buys securities. The change in the value of the cash account dF is

$$dF = r^*F^*dt - (p - b) * dq_b + (p + a) * dq_a. \tag{3.7}$$

The dealer's inventory consists of shares of stock in which he/she makes a market. The change in the value of the inventory dI is

$$dI = r_I{}^*I^*dt + p^*dq_b - p^*dq_a + I^*dZ_I. \tag{3.8}$$

Note that the stock is evaluated at p in case of changes in inventory whereas $(p - b)$ and $(p + a)$ are used to evaluate the change in the cash.

The dealer has base wealth Y. The change in base wealth dY is

$$dY = r_Y{}^*Y^*dt + Y^*dZ_y. \tag{3.9}$$

The three variables, cash F, inventory I and base wealth Y, are state variables which describe any state of world at time t.

The optimal performance function $J(\bullet)$ is defined as

$$J(t,F,I,Y) = \max_{a,b}\left[EU(W_T)|t,F,I,Y\right]. \tag{3.10}$$

The function $J(\bullet)$ is the solution to the maximisation problem for an optimal bid and ask from t_0 to T. The maximisation problem of Equation (3.10) results in a dynamic programming problem. Since there is no intermediate consumption before T, the fundamental recurrence relation implied by the principle of optimality of dynamic programming requires that

$$\max_{a,b} dJ(t,F,I,Y) = 0 \quad and \quad J(T,F,I,Y) = U(W_T). \tag{3.11}$$

Ho and Stoll (1981) do not solve directly the maximisation problem above. The reason is that solving Equation (3.11) requires solving explicitly for the $J(\bullet)$ function(see appendix 3.C), while the $J(\bullet)$ function is not clearly known. Thus, Ho and Stoll (1981) transform and simplify the maximisation problem (3.11).

They consider the problem only at the period when the time remaining τ is equal to zero and take a first-order approximation of the Taylor's series expansion of $J(\bullet)$. (See Appendix 3.C for (3.12))

After simplifications and substitutions, the dealer's problem is restated as

$$J_\tau = LJ + \max_{a,b}\{\lambda(a)aQSJ_F - \lambda(a)[J(\bullet) - SJ] + \lambda(b)bQBJ_F - \lambda(b)[J(\bullet) - BJ]\} \tag{3.12}$$

Equation (3.12) suggests that the change with respect to time remaining τ of derived utility J_x depends on two parts: the LJ term and the Max term. The LJ term represents the return and risk of the dealer's current wealth and the dealer cannot have control over the risk and return. The Max term represents the net contribution to the derived utility of the dealer from his/her sales and purchases. Since the dealer can control the variables in the max term, the dealer can find the optimal pricing strategies from the max term.(See Appendix 3.C for (3.13) and (3.14))

In the case of the bid fee, the dealer's optimal price which is the solution to the problem (3.12) yields

$$b^* = \frac{\alpha}{2\beta} + \frac{J(\bullet) - BJ(\bullet)}{2BQJ_F}. \tag{3.13}$$

In the case of ask fee, the corresponding optimal price yields

$$a^* = \frac{\alpha}{2\beta} + \frac{J(\bullet) - SJ(\bullet)}{2SQJ_F} \tag{3.14}$$

The optimal buying fee $b*$ and the optimal selling fee $a*$ are set at a level which maximises expected profits. The optimal buying (selling) fee is composed of the stochastic arrival term of public seller(buyer) and a risk premium term.

When the dealer has positive inventory, the optimal buying fee should be greater than the optimal selling fee, i.e. $b^* > a^*$, while if the dealer's inventory is zero, $b^* = a^*$. This suggests that if the dealer has positive inventory, b increases and a decreases, while if the dealer has negative inventory, b decreases and a increases. As inventory changes, a changes in the opposite direction to b. The reason is that when the dealer has positive inventory, he/she prefers a sale (public buy) to a purchase (public sale) and when the dealer has the negative inventory, they prefer the opposite.

From a^* and b^*, the dealer's spread s is

$$s = a + b = \frac{\alpha}{\beta} + \frac{(J - SJ)}{2SJ_FQ} + \frac{(J - BJ)}{2BJ_FQ} \tag{3.15}$$

If the solution of $J(\bullet)$ is substituted into the spread equation above, then the optimal dealer spread is

$$s = \frac{\alpha}{\beta} - \frac{1}{2}Z\frac{Q}{W}\sigma_I^2\tau + \frac{1}{2}Z\frac{Q}{W}\sigma_I^2\left[(r_I - r + G_I) + Z(r_W + \frac{2\Pi}{W})\right]\tau^2 \tag{3.16}$$

where $Z = U''W/U'$, the coefficient of relative risk aversion, Q is transaction size, $G_I = r_I + \frac{1}{2}\sigma_I^2$, which is the instantaneous growth in the variance of $I, r_W = r(F/W) + r_I(I/W) + r_Y(Y/W)$, $\tau$is remaining time, and $\Pi = \alpha^2 Q/4\beta$.

The spread in the Ho and Stoll (1981) model has three indications for the dealer's optimal pricing behaviour. The first is that the spread depends on the remaining time τ of the dealer. The second is that the spread is decomposed into three parts: a risk neutral spread $\frac{\alpha}{\beta}$, and an adjustment for uncertainty, the remaining terms of the spread. The adjustment depends on the dealer's coefficient of relative risk aversion Z, the size of the transaction Q, and the risk of the stock G_I. The third is that the spread is independent of the inventory level. This means that the dealer's inventory position does not affect the spread size. The second and third properties are the same as in Stoll (1978b).

Ho and Stoll (1981), however, do not explain three things. The first is that the inventory of the dealer is liquidated at a future time T. This finite time horizon reduces the dealer's risk, otherwise the dealer would face a change in the value of inventory. The second limitation concerns a fixed true price for the stock. The assumption of the fixed intrinsic value of the stock is justified when the time horizon is short enough. When one reason to use a multi-period framework in Ho and Stoll (1981) is to reflect the possible fluctuations in the stock value which have an impact on the dealer, it is not clear that the assumption of a fixed price for the stock captures this effectively. The third limitation concerns the order process assumptions. Even though the assumption that the stochastic order flows follow the Poisson distribution seems to be realistic, the requirement for using that assumption is that no trader knows more about the future movement of prices. If any one trader has better information then the presence of that trader may affect order flow behaviour. However, Ho and Stoll (1981) focus on the effect of the dealer's inventory on the spread and the presence of the trader who is better informed is not important in the Ho and Stoll (1981) model.

3.3.5 O'Hara and Oldfield (1986)

Whereas Garman (1976) and Ho and Stoll (1981) employ the assumption of a Poisson order arrival rate, O'Hare and Oldfield (1986) consider market order[1] and limit order[2]. The Poisson order flow assumption employed by Garman (1976) and Ho and Stoll (1981) restricts order types adopted by traders to market orders. Limit orders, however, are extensively used in the trading procedure on the stock exchange. Since limit orders are more price-contingent, the stochastic Poisson order flow assumption does not capture the price-contingent property of limit orders. Thus, O'Hare and Oldfield (1986) address this issue by modelling the dynamic pricing policy of a risk averse dealer who receives both limit and market orders, and who faces order flow and inventory value uncertainty in the discrete-time multi-period framework.

In their model, they assume that a given trading day contains n trading intervals and the dealer maximises the expected utility of trading profits by choosing bid and ask prices over an infinite number of trading days.

$$\max_{a,b} E\left[\sum_{j=0}^{\infty} \alpha^j U\left(\sum_{t=1}^{n} (\tilde{\pi}_{jt})\right)\right], \tag{3.17}$$

where U is a Von-Neumann Morganstern utility function[3], α is a discount rate $0 <= \alpha < 1$, $\tilde{\pi}_{jt}$ is the trading profit in period t of day j.

[1]Market orders are the ones to buy and sell for immediate execution.

[2]Limit orders are the ones to buy or sell at some pre-specified price

[3]The Von-Neumann Morganstern utility function is increasing, concave, bounded, and twice differentiable(U"<0)

Transaction arises as follows; at the start of each period, the dealer knows the limit orders submitted for that trading period and has expectations over the market orders the dealer will receive at each possible price quote. After the dealer sets bid and ask prices, the dealer executes at the quoted prices all proper limit orders and any market orders which arrive during the period. The final inventory of the dealer is computed at the price cleared orders.

A dealer confronts two types of orders: limit orders and market orders. The limit orders to buy from the dealer in period t A^L are assumed to be linear functions of the price and are represented cumulative order functions.

$$A_t^L \; = \alpha^L - a_t\gamma^L \; = \int_{a_t}^{\bar{a}} q_a(a_t)da_t \tag{3.18}$$

The corresponding limit orders to sell to the dealer is

$$B_t^L \; = \beta^L - a_t\phi^L \; = \int_{\bar{b}}^{b_t} q_b(b_t)db_t \tag{3.19}$$

where the superscripted L represents the limit order; α, γ, β, and ϕ are parameters of the limit order flow; the q functions are the incremental orders at each price; and $\bar{a}$ and $\bar{b}$ are the highest ask and lowest bid price.

The market order flow is assumed to be composed of both price-dependent orders and liquidity-induced orders.

$$\begin{aligned} \tilde{A}_t^M &= \alpha^M - a_t\gamma^M + \tilde{w}_t \\ \tilde{B}_t^M &= \beta^M - b_t\phi^M + \tilde{\varepsilon}_t \end{aligned} \tag{3.20}$$

where the superscripted M represents market order; α, γ, β, and φ, are parameters. $\tilde{w}_t$ and $\tilde{\varepsilon}_t$ are random variables, which are independently identically distributed over time.

Then, a period's total buy and sell orders are $\tilde{A}_t$ and $\tilde{B}_t$.

$$\begin{aligned} \tilde{A}_t = \alpha - a_t\gamma + \tilde{w}_t \quad & \text{if } \alpha^L - a_t\gamma^L \geq 0 \\ \alpha^M - a_t\gamma^M + \tilde{w}_t \quad & \text{otherwise} \end{aligned} \tag{3.21}$$

$$\begin{aligned} \tilde{B}_t = \beta + b_t\phi + \tilde{\varepsilon}_t \quad & \text{if } \beta^L + b_t\phi^L \geq 0 \\ \beta^M + b_t\phi^M + \tilde{\varepsilon}_t \quad & \text{otherwise} \end{aligned} \tag{3.22}$$

where $\alpha = \alpha^L + \alpha^M$; $\gamma = \gamma^L + \gamma^M$; $\beta = \beta^L + \beta^M$; and $\phi = \phi^L + \phi^M$.

Since limit orders are positive, the optimisation problem of the dealer can be restated as

$$\max_{\{a_n,b_n\}} E\left[U\sum_{t=1}^{n}\tilde{\pi}_t + V(I_n)\right] \tag{3.23}$$

$$\max_{\{a_n,b_n\}} E\left[\begin{array}{l} U\left(\begin{array}{l}\sum_{t=1}^{n-1}\pi_t + a_n(\alpha - a_n\gamma + \tilde{w}_n) - b_n(\beta + b_n\phi + \tilde{\varepsilon}_n) + \\ r\tilde{p}(I_{n-1} + \beta + b_n\phi + \tilde{\varepsilon}_n - \alpha + a_n\gamma - \tilde{w}_n)\end{array}\right) \\ +V(I_{n-1} + \beta + b_n\phi + \tilde{\varepsilon}_n - \alpha + a_n\gamma - \tilde{w}_n)\end{array}\right] \tag{3.24}$$

$$\text{subject to } \alpha^L - a_t\gamma^L \geq 0;\ \beta^L + b_t\phi^L \geq 0$$

$V(I_n)$ is the dealer's derived value function. In the O'Hare and Oldfield (1986) model, daily settlement takes place and this means that the dealer's inventory affects the current cash flow and the dealer's future operation. Thus, inventory can become the state variable of the optimisation problem and is incorporated in the form of $V(I_n)$.

In the utility function U, the second term to the fourth term represents the dealer's profit in period n. The second term and the third represent direct cash flow from trading, and the fourth term represents the cash flow from financing or lending the ending inventory to other dealers.

The first-order conditions can be solved for the optimal bid and ask prices for period n. The optimal bid and ask prices are

$$a_n = \frac{\alpha}{2\gamma} + \frac{E(U' * \tilde{w}_n)}{E(U') * 2\gamma} + \frac{rE(U' * \tilde{p})}{2E(U')} + \frac{E(V')}{2E(U')} \tag{3.25}$$

$$b_n = -\frac{\beta}{2\phi} - \frac{E(U' * \tilde{\varepsilon}_n)}{E(U') * 2\phi} + \frac{rE(U' * \tilde{p})}{2E(U')} + \frac{E(V')}{2E(U')} \tag{3.26}$$

where $U' = \frac{\partial U}{\partial \pi}$ and $V' = \frac{\partial V}{\partial I_n}$

The first term in a_n and b_n comes from the known limit orders and expected market orders. The second term represents the effect of market order flow, i.e. the risk adjustments for the variability of market orders. The last two terms consider the effect of inventory, i.e. the direct price level adjustment induced by inventory change.

The optimal spread of the dealer is

$$a_n - b_n = \frac{\alpha\phi + \beta\gamma}{2\phi\gamma} + \frac{\phi E(\tilde{w}_n) + \gamma E(\tilde{\varepsilon}_n)}{2\phi\gamma} + \frac{\phi cov(U', \tilde{w}_n) + \gamma cov(U', \tilde{\varepsilon}_n)}{2\phi\gamma E(U')} \tag{3.27}$$

The first two terms show the expected order flow of the dealer and the last term reflects the effects of market order uncertainty. In case of risk neutral dealers, the third term will be zero whereas in case of risk averse dealers the third term can be negative or positive. Overall, the optimal spread has a similar structure to the spread in Ho and Stoll (1981).

To examine when the dealer faces order and price uncertainties simultaneously, O'Hare and Oldfield (1986) assume that the dealer has a negative exponential utility, in which the preference of the dealer exhibits constant absolute risk aversion. Then the dealer's optimisation problem (3.23) can be restated as

$$\max_{\{a_n, b_n\}} E\left[-\exp\left(c\sum_{t=1}^{n}\tilde{\pi}_t\right) - \exp\left(-d\tilde{p}\tilde{I}_n\right)\right] \tag{3.28}$$

The dealer's final period problem (3.24) is transformed into

$$\max_{\{a_n,b_n\}} \sum_{t=1}^{n-1}(\pi_t) + E(\tilde{\pi}_n) - \frac{c}{2}Var(\tilde{\pi}_n) + E(\tilde{p}\tilde{I}_n) - \frac{d}{2}Var(\tilde{p}\tilde{I}_n) \tag{3.29}$$

$$\text{subject to } \alpha^L - a_t\gamma^L \geq 0;\ \beta^L + b_t\phi^L \geq 0.$$

To get (3.29), O'Hare and Oldfield (1986) need another assumption: trading profits and inventory value are jointly normally distributed. This assumption allows the expected utility maximising problem to change into a function that is linear in expected values and variances. Market order flows also allow for this transformation.

When the dealer faces only price variability, the optimal spread is

$$a_n - b_n = \frac{\alpha\phi + \beta\gamma}{2\phi\gamma} + \frac{\phi E(\tilde{w}_n) + \gamma E(\tilde{\varepsilon}_n)}{2\phi\gamma}. \tag{3.30}$$

The optimal spread (3.30) is just the spread (3.27) in case of risk neutral dealers.

When the dealer faces only order variability, the optimal spread is

$$a_n - b_n = \frac{\alpha\phi+\beta\gamma}{2\phi\gamma} + \frac{\phi E(\tilde{w}_n)+\gamma E(\tilde{\varepsilon}_n)}{2\phi\gamma} + \frac{c}{2\phi\gamma}\left[\begin{array}{l} \gamma Var(\tilde{\varepsilon}_n)\left(\frac{-\beta-E(\tilde{\varepsilon}_n)+(1+r)\bar{p}\phi-cr\bar{p}Var(\tilde{\varepsilon}_n)}{2\phi+cVar(\tilde{\varepsilon}_n)} + r\,\bar{p}\right) \\ -\phi Var(\tilde{w}_n)\left(\frac{\alpha+E(\tilde{w}_n)+(1+r)\bar{p}\gamma-cr\bar{p}Var(\tilde{w}_n)}{2\gamma+cVar(\tilde{w}_n)} + r\,\bar{p}\right) \end{array}\right] \tag{3.31}$$

where $\bar{p}$ is fixed inventory price.

The optimal spread (3.31) suggests that when supply and demand are known the dealer can determine his optimal inventory level by moving symmetrically the placements of bid and ask prices. The optimal spread (3.31) suggests that the spread contains a risk-adjustment term, which is the third term in (3.31), and the level of inventory does not affect the spread.

When the dealer faces simultaneously both market order variability and price variability, the optimal solution is more complicated. The inventory level the dealer holds affects the placements of bid and ask prices. In addition, the dealer's inventory has an impact on the magnitude of the spread. In other words, when there are multiple uncertainties, the dealer can not control his risk enough just by moving his bid and ask prices. Thus, when the dealer can change the size of the spread, he is more flexible to offset his risk. This result suggests that the spread is not independent of the dealer's inventory.

O'Hare and Oldfield (1986) show that when market order uncertainty and inventory price[1] uncertainty coexist, the optimal spread set by the dealer depends on the placements and size of bid and ask prices. They, however, focus on the optimization problem in period n. Though they have the optimal solution for the spread, the computation process is too complex to be tractable. In addition, they do not consider the interaction between the dealer's current bid and ask prices and future limit orders. In their model, after limit orders are submitted, the dealer decides current bid and ask prices.

3.3.6 Summary

So far I have looked into a range of papers that consider the price-decision process of the dealer facing inventory risk, which is incorporated into the inventory holding component of the spread. These papers show that when the dealer is confronted with unwanted inventory holding risk, the dealer tries to find a method to reduce or compensate for this inventory risk so that the dealer adjusts bid and ask prices.

3.4 Adverse Selection Cost

The previous sections 3.2 and 3.3 provide theoretical works showing that the spread arises due to the order-processing cost and the inventory holding cost. Meanwhile there is another reason why a bid-ask spread appears. This approach assumes an information asymmetry between dealer and informed trader.

[1] The price is used to compute the inventory value.

3.4.1 Bagehot(1971)

The first paper to consider this issue was Bagehot (1971) who distinguishes trading gains from market gains. Market gains arise when a portfolio is completely diversified and the only investment risk remaining is market risk, investors play a fair game and receive a neutral market rate of return over time whether prices tend to rise or fall. When market prices go up in general, most investors gain. Trading gains arise due to the existence of traders who have superior information, these are called informed traders. Bagehot (1971) distinguishes between informed traders and liquidity-motivated traders who have no superior information but want to convert securities into cash or cash into securities. The dealer always gains in his transactions with liquidity-motivated traders and loses with informed traders. Thus, the bid-ask spread set by the dealer reflects a balancing of losses to the informed traders with gains from liquidity-motivated traders. Bagehot (1971) argues that information costs also affect the spread as a result of the balancing behaviour of the dealer.

3.4.2 Copeland and Galai(1983)

Copeland and Galai (1983) formalise the information cost as adverse selection cost[1]. The adverse selection cost arises as the dealer trades with informed traders, i.e. agents who have better information about the underlying assets.

In the Copeland and Galai (1983) model, a single risk-neutral dealer exists. Informed traders convey information about the realizations of the true underlying asset value P. Uninformed traders and the dealer do not know the realisation of P. The generating process of information, transformation between informed and uninformed, and the reason why uninformed traders[2] want to trade are exogenous events in their model. The dealer knows that any given trade comes from an informed trader with probability π_I and from an uninformed trader with probability $1 - \pi_I$.

[1]The reason we call adverse selection cost is that the information asymmetry between the dealer and informed traders is ascribed to superior information of informed trader, which leads to adverse selection problem the dealer faces.

[2]Uninformed traders are sometimes called as liquidity traders.

Copeland and Galai (1983) use an instantaneous quote framework in which the dealer sets bid and ask prices and trades occur with no time passing. The expected loss of the dealer to an informed trader is

$$\pi_I \left\{ \int_{P_A}^{\infty} (P - P_A) f(P) dP + \int_{0}^{P_B} (P_B - P) f(P) dP \right\} \tag{3.32}$$

where $f(P)$ represents the stochastic process governing price changes, and P_A and P_B are ask and bid prices.

The expected gain of the dealer from uninformed traders is

$$(1 - \pi_I) \{ \pi_{BL}(P_A - P) + \pi_{SL}(P - P_B) + \pi_{NL}(0) \} \tag{3.33}$$

where π_{BL} is the probability that an uninformed trader will buy, π_{SL} is the probability that the uninformed trader will sell, π_{NL} is that the uninformed trader will not trade at all.

Then the dealer chooses bid and ask prices to maximise her expected profit. This is expressed as

$$\max_{P_A, P_B} \left[\begin{array}{l} (1 - \pi_I) \{ \pi_{BL}(P_A - P) + \pi_{SL}(P - P_B) + \pi_{NL}(0) \} \\ -\pi_I \left\{ \int_{P_A}^{\infty} (P - P_A) f(P) dP + \int_{0}^{P_B} (P_B - P) f(P) dP \right\} \end{array} \right] \tag{3.34}$$

The dealer's objective function (3.34) explicitly incorporates bid and ask prices to maximise his/her expected profits. This model captures the feature of information costs described by Bagehot(1971). The objective function (3.34) suggests that, to compute the dealer's expected profits, one needs to know the probabilities of informed and uninformed traders, the probabilities of trading intention of the liquidity traders, and the stochastic behaviour of the stock price. The objective function (3.34) also suggests that the size and placement of the spread affect the expected profits of the dealer. However, Copeland and Galai (1983) do not solve explicitly the dealer's optimisation problem (3.34).

One important implication of the objective function (3.34) is that the bid-ask spread will not disappear even when risk-neutral competitive dealers exist in the market. The size of the spread depends on several factors including the population parameters of the informed and uninformed. As long as some informed traders exist the spread is never zero. Thus Copeland and Galai (1983) show that information alone is sufficient to induce market spreads.

However, there are some limitations to Copeland and Galai (1983). The first limitation is that their results are based on a static one-trade framework. The second limitation is about order flow. Copeland and Galai (1983) do not consider that the trade itself may convey information. With asymmetric information, the dealer knows the existence of informed traders, and the dealer also knows that the informed traders place orders in the order flow the dealer receives even though the dealer does not discern which order relates with informed traders. Thus the dealer will use trade as a signal about the true value of the underlying asset. Moreover, the dealer may infer some information from the continued trading with the informed trader.

3.4.3 Glosten and Milgrom (1985)

Glosten and Milgrom (1985) focus on the second limitation of Copeland and Galai (1983). Glosten and Milgrom (1985) argue that informed traders are likely to place their orders on one-side of the market since they know the true asset value, and thus informed traders' trading direction reveals their information about the underlying asset to the dealer. The dealer updates his/her belief about the underlying asset through the one-side accumulated transactions with informed traders. Thus in the Glosten and Milgrom (1985) model, trades reveal information about the underlying information structure.

The model assumes that (1) only unit transactions take place. (2) There are informed traders and pure liquidity traders. (3) All participants, traders and dealers are risk-neutral. (4) Investors arrive one by one, randomly and anonymously at the specialist's post. This assumption implies that trade takes place sequentially while one trader is allowed to transact at any point in time. (5) The dealer knows the probabilistic structure of the arrival process. This assumption requires that the population of traders the dealer faces is always the same as the population of potential traders. (6) Investors, upon arriving at the market and hearing the bid and ask, maximise expected utility given their information to date.

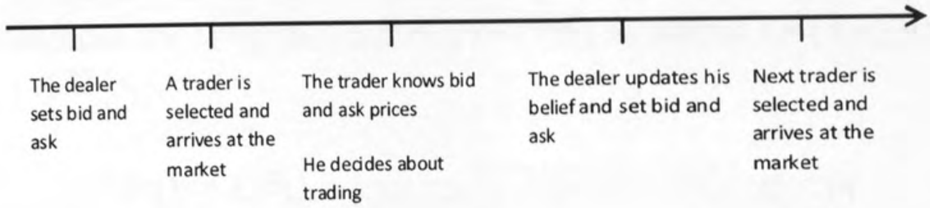


Figure 3.2: Tradig Procedure

The trading process is described in Figure 3.2; firstly, the dealer sets bid and ask prices, and a trader is selected from the population and goes to the market. Secondly after the trader finds out bid and ask prices, he/she chooses one of three options: to buy one unit at the ask price or to sell one unit at the bid price or to leave the market. Thirdly, the dealer is free to change his/her bid and ask prices at any time after an arriving trader made a decision and before the next arrival of another trader. In the trading process, the dealer posts bid and ask prices such that the expected profit on any trade is zero. This zero profitability arises due to competition in the dealers and the risk-neutrality of the dealer.

The bid and ask prices the dealer sets are a conditional expectation of the asset value given the type of trade.

$$a_1 = E[V|B_1] = \underline{V} * \Pr\{V = \underline{V}|B_1\} + \overline{V} * \Pr\{V = \overline{V}|B_1\} \tag{3.35}$$

$$b_1 = E[V|S_1] = \underline{V} * \Pr\{V = \underline{V}|S_1\} + \overline{V} * \Pr\{V = \overline{V}|S_1\} \tag{3.36}$$

where B_1 and S_1 are bid and ask prices. The true value of the asset V will be either low $\underline{V}$, or high $\overline{V}$.

The high or low true values depend on the information set. Ask price (3.35) reflects the conditional expectation of the true asset value V given that a trader wants to buy from the dealer. The bid price (3.36) reflects the corresponding expectation.

To compute the ask price a_1 and the bid price b_1, the dealer must know four probabilities: $Pr\ \{V{=}\underline{V}\ |B_1\}$ and $Pr\ \{V{=}\overline{V}\ |B_1\}$ as well as $Pr\ \{V{=}\underline{V}\ |S_1\}$ and $Pr\ \{V{=}\overline{V}\ |S_1\}$.

$$\Pr\{V=\underline{V}|B_1\} = \frac{\Pr\{V=\underline{V}\}*\Pr\{B_1|V=\underline{V}\}}{\Pr\{V=\underline{V}\}*\Pr\{B_1|V=\underline{V}\}+\Pr\{V=\overline{V}\}*\Pr\{B_1|V=\overline{V}\}} \tag{3.37}$$

$$\Pr\{V=\overline{V}|B_1\} = \frac{\Pr\{V=\overline{V}\}*\Pr\{B_1|V=\overline{V}\}}{\Pr\{V=\underline{V}\}*\Pr\{B_1|V=\underline{V}\}+\Pr\{V=\overline{V}\}*\Pr\{B_1|V=\overline{V}\}} \tag{3.38}$$

$$\Pr\{V=\underline{V}|S_1\} = \frac{\Pr\{V=\underline{V}\}*\Pr\{S_1|V=\underline{V}\}}{\Pr\{V=\underline{V}\}*\Pr\{S_1|V=\underline{V}\}+\Pr\{V=\overline{V}\}*\Pr\{S_1|V=\overline{V}\}} \tag{3.39}$$

$$\Pr\{V=\overline{V}|S_1\} = \frac{\Pr\{V=\overline{V}\}*\Pr\{S_1|V=\overline{V}\}}{\Pr\{V=\underline{V}\}*\Pr\{S_1|V=\underline{V}\}+\Pr\{V=\overline{V}\}*\Pr\{S_1|V=\overline{V}\}} \tag{3.40}$$

Constructing a simple tree diagram (see Figure 3.3) allows the dealer to calculate easily these probabilities above ((3.37) to (3.40)). The symbols beside each node represent the probability. In the first node, nature chooses information type. The probability that the signal is good news is represented by θ. In the second node the type of trader is selected with the probability of being informed μ. In the third node each trader will decide to buy an asset or sell. r^B is the probability that an uninformed trader buys and r^S is the probability that an uninformed trader sells. In case of informed trader, informed trader tends to buy asset with probability one and to sell asset with probability zero when the signal is good news. When the signal is bad news, the informed trader takes the opposite behaviour.

To compute four probabilities above ((3.37) to (3.40)), the dealer should know $Pr\{B_1\ |V{=}\underline{V}\}$, $Pr\{B_1\ |V{=}\overline{V}\}$, $Pr\{S_1\ |V{=}\underline{V}\}$, and $Pr\{S_1\ |V{=}\overline{V}\}$. For example, to compute the value $Pr\{V{=}\underline{V}|B_1\}$, the dealer can calculate easily

$Pr\{B_1 \mid V=\underline{V}\}$ and $Pr\{B_1 \mid V=\overline{V}\}$ in the Figure 3.3. $Pr\{B_1 \mid V=\underline{V}\}$ is computed by $\frac{0+(1-\theta)*(1-\mu)*r^B}{(1-\theta)} = (1-\mu)*r^B$. The corresponding $Pr\{B_1 \mid V=\overline{V}\}$ is computed by $\frac{\theta*\mu+\theta*(1-\mu)*r^B}{\theta} = \mu + (1-\mu)*r^B$. Thus the dealer can find the initial bid and ask prices. After transaction in period 1, the dealer updates his/her beliefs by using the same procedure above. The dealer uses all available information over time in a Bayesian manner.

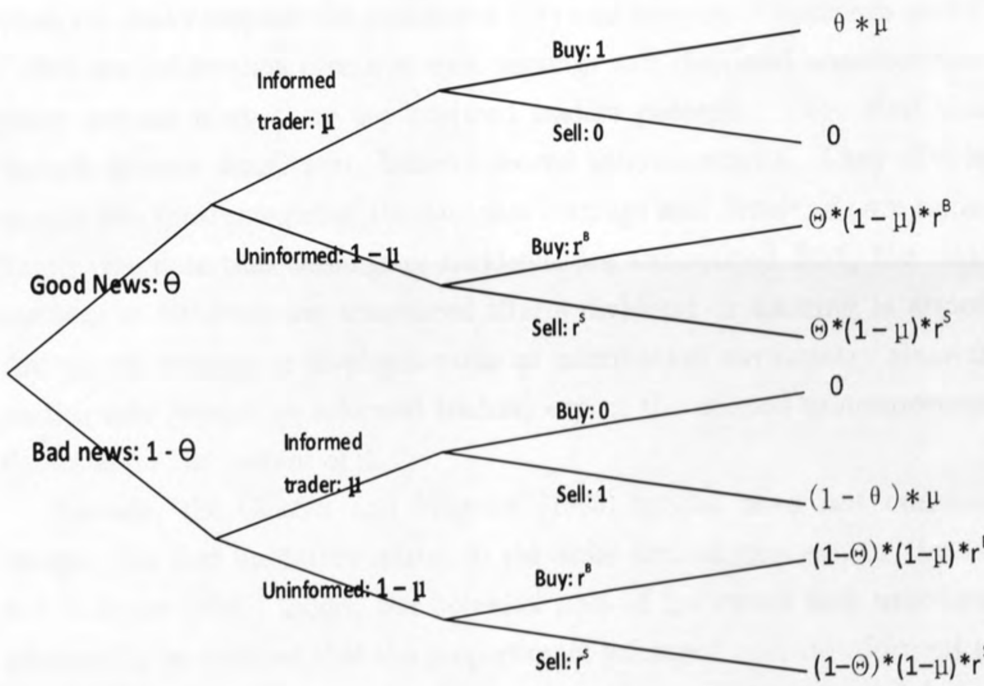


Figure 3.3: Probability of Trading Type

Glosten and Milgrom (1985) suggest two important implications. The first implication is that the bid-ask spread arises due to the change in the dealer's beliefs even when any order-processing cost or inventory-holding risk do not exist. The bid and ask prices are expected values given publicly available information.

In analysing the spread, they show that the spread depends on several factors, which are the nature of underlying information, the number of informed traders, and traders' elasticity. Secondly, in the Glosten and Milgrom (1985) model the transaction prices follow a Martingale process with respect to the dealer's information. This means that if any trader did not have better information than the dealer the trader cannot do better in predicting the future price than by simply using the current price.

Venkatesh and Chiang (1986) test the prediction of both Copeland and Galai (1983) and Glosten and Milgrom (1985) that the dealer widens the bid-ask spread when the dealer suspects the presence of informed traders. Venkatesh and Chiang (1986) use information events as such earnings and dividend announcements to proxy periods when there are informed traders present. They find that raw spreads increase significantly before a second announcement. They divide their sample into three categories: the date that earnings and dividends are announced jointly, the date that earnings or dividends are announced first, the date that earnings or dividends are announced after a dividend or earning is announced. Announced earnings or dividends cause an information asymmetry since market participants (except for informed traders) expect the second announcement but do not know the content of it.

However, the Glosten and Milgrom (1985) model does not consider two things. The first limitation relates to the order arrival process, in the Glosten and Milgrom (1985) model, the potential pool of informed and uninformed is assumed to be static so that the proportion of informed and uninformed traders does not change. Secondly, the Glosten and Milgrom (1985) model does not consider the strategic behaviour of the informed. If informed traders have better information than the dealer, then the informed traders sometimes disguise their trading intention or use larger transaction size to maximise their expected profit. Since informed traders in the Glosten and Milgrom (1985) model transact only one unit of asset, the informed traders cannot behave strategically. Supporting evidence of this criticism is provided by Barclay and Warner (1993). They find that informed traders break their order into smaller trade sizes to avoid revealing their trading intention and to extend their informational advantage. The impact of trade size is greatest for medium sizes.

3.4.4 Easley and O'Hara(1987)

Easley and O'Hara (1987) add two discerning dimensions to Copeland and Galai (1983) and Glosten and Milgrom (1985). Easley and O'Hara (1987) focus on the adverse selection problem in trade size and on the nature of the information uncertainty.

In the Easley and O'Hara (1987) model, an information event is defined as the occurrence of a signal s about the value of the asset V. A signal s occurs before the trading day and takes one of two possible values, low L and high H. The probability that the signal s indicates low value L is δ and the probability that the signal s indicates high value H is $(1-\delta)$. $\underline{V}$ is defined as $E\{V|s=L\}$ and $\overline{V}$ is $E\{V|s=H\}$. The probability that a signal s occurs is α. When a signal s occurs, some fraction μ of the traders receive the signal and become informed. Thus the dealer is uncertain of the existence of new information since it is possible that no signal occurs.

In the Easley and O'Hara (1987) model, transactions occur sequentially. The individual trader trades based on the probabilities of trader types in the population. In other words, the fraction of trades coming from informed traders, not the number (or fraction) of informed traders, is important for the price of the asset. The individual trader can choose trade quantity, that is, the trader may buy a large or a small quantity (B_1 and B_2) or sell a large or a small quantity (S_1 and S_2). A liquidity trader facing a large liquidity need does not divide one large quantity into several small quantities. The liquidity trader prefers to trade once at a large quantity. Meanwhile informed traders are risk-neutral and competitive.

Within this framework, Easley and O'Hara (1987) demonstrate that there are two possible equilibria: a separating equilibrium and a pooling equilibrium. In the separating equilibrium, the informed trader transacts only the large quantity since informed traders want to maximise their profits, while in the pooling equilibrium the informed trader transacts large and small quantities.

When the informed trader transacts only large quantities and this trading activity is known to the dealer, a separating equilibrium or pooling equilibrium may exist. Whether a separating equilibrium exists or a pooling equilibrium exists depends on the condition that Easley and O'Hara (1987) show. In the separating equilibrium, the dealer does not need to post the spread for small quantity trades, since small transactions are not based on information. The dealer does not need to update his/her probability after a small transaction. As a result, the initial bid and ask prices for small quantities equal $V^*= \delta^*\underline{V}+ (1\text{-}\delta)^*\overline{V}$, that is, there is no spread for small quantity trade in the separating equilibrium.

The dealer, however, sets the spread for large transactions. The initial bid b^* and ask a^* prices for large quantities are

$$b^* = V^* - \frac{\sigma_v^2}{\overline{V} - \underline{V}} \left[\frac{\alpha\mu}{X_S^2(1-\alpha\mu) + \delta\alpha\mu} \right] \tag{3.41}$$

$$a^* = V^* + \frac{\sigma_v^2}{\overline{V} - \underline{V}} \left[\frac{\alpha\mu}{X_B^2(1-\alpha\mu) + (1-\delta)\alpha\mu} \right], \tag{3.42}$$

where X denotes the fraction of uninformed traders who trade the large quantity(subscripted to indicate buy and sale), σ_v^2 is the variance of V, $\alpha\mu$ represents the probability of an informed trading and δ is the probability that V is equal to $\underline{V}$.

For the bid (3.41) and the ask (3.42) to constitute a separating equilibrium, the following conditions must hold:

$$\frac{S^2}{S^1} \geq 1 + \frac{\alpha\mu\delta}{X_S^2(1-\alpha\mu)}, \tag{3.43}$$

$$\frac{B^2}{B^1} \geq 1 + \frac{\alpha\mu(1-\delta)}{X_B^2(1-\alpha\mu)}, \tag{3.44}$$

where $S^2(B^2)$ denotes the larger sell(buy) size and $S^1(B^1)$ is the smaller sell(buy) size. These conditions guarantee that the informed trader's profit is higher trading the larger quantity at the worse price than it is trading the smaller quantity at a better price.

If these two conditions ((3.43) and (3.44)) do not hold, then a pooling equilibrium exists. In this case, the informed traders do no stick to order large quantity. i.e. in a pooling equilibrium, the informed traders places both large and small quantities of order. Thus, the dealer sets two spreads: one for the large trade and the other for the small trade. Easley and O'Hara (1987) do not show the exact formula for the bid and ask price, but it must be the case that the informed traders make a larger profit by pooling in large quantity than separating small from large quantities.

For the informed traders to be pooling, three conditions must be satisfied. First, an informed trader must be indifferent between trading a large and a small quantity. Second, the dealers must anticipate zero expected profit from each trade. Third, it must be possible to choose a ψ_1^S [1] with $0 < \psi_1^S \leq 1$ and simultaneously satisfy the equal-profit condition for any informed traders and the zero-profit condition for the dealers. These three conditions are summarized into following inequalities:

$$\frac{S^2}{S^1} < 1 + \frac{\alpha\mu\delta}{X_S^2(1-\alpha\mu)} \tag{3.45}$$

$$\frac{B^2}{B^1} < 1 + \frac{\alpha\mu(1-\delta)}{X_B^2(1-\alpha\mu)} \tag{3.46}$$

If the two inequalities above are satisfied, there are two different ask (bid) prices: one for a large trade and the other for a small trade. The inequalities (3.45) and (3.46) are exactly the opposite of the inequalities (3.43) and (3.44). The results dictate that the market is always either in a separating or a pooling equilibrium, depending on the market parameters.

In the Easley and O'Hara (1987) model, three market parameters such as $\alpha\mu$, X_S^2, and X_B^2 determine whether the market is in a separating or pooling equilibrium. If $\alpha\mu$ <1, then in markets where large amounts can be traded in a single transaction, a separating equilibrium will prevail. In markets where large trades rarely occur(i.e. X_S^2, and X_B^2 are small), a pooling equilibrium will prevail. In markets where a low probability of information-based trading(i.e. $\alpha\mu$ is low), a separating equilibrium prevails.

Easley and O'Hara (1987) show that with regard to the informed trading, the spread has the following characteristics:

(1) If $\alpha\mu$ =0, there is no spread at either large or small quantities.

[1] ψ_1^S is the probability that an informed trader trades the smaller quantity S^1 given that the informed trader knows a low signal $s = L$

(2) If $\alpha\mu > 0$, the spread increases with trade size.

(3) If $1 > \alpha\mu > 0$, the spread at large quantity in the separating equilibrium decreases with increased depth and with reduction in $\alpha\mu$ and increases with increased variance of the value of the asset.

These characteristics suggest that the spread arises due to information effect, which is due to the uncertainty surrounding the information event, α, and to the uncertainty regarding any individual's trading motivations, μ.

As in predicted in the Easley and O'Hara (1987), Lee, Mucklow, and Ready (1993) find that spreads widen and depths drop in response to an increase in volume. They also find that spreads widen and depths fall before earnings announcements. Their findings suggest that dealer adjusts both spread and depth to manage information asymmetry due to informed traders.

3.4.5 Summary

So far I have examined four papers that consider adverse selection costs but there are other papers[1] using alternative approaches to model information asymmetry. Four papers[2] covered in above sections, however, are enough to show that the bid-ask spread incorporates adverse selection cost.

Overall, the theoretical works about the sources of the bid-ask spread dictate that the order processing cost is in the spread because of the costs that the dealer bears to process receiving orders. The inventory holding cost arises when the dealer endures some risks to provide immediate liquidity in the market that causes the suboptimal inventory positions for the dealer. Additionally, there is a source of the spread ascribed to the information asymmetry between the dealer and informed traders. The dealer needs to set an optimal spread to make a balance between the gains from liquidity traders and the losses from informed traders. While I review briefly three major components of spread, in the next

[1] The Copeland and Galai model and the Glosten and Milgrom model use a sequential trading model while another class of model use strategic informed trader. The example is Kyle(1985).

[2] Bagehot(1971), Copeland and Galai(1983), Glosten and Milgrom(1985), and Easley and O'Hara(1987)

chapter I look into how these components can be empirically extracted from spreads.

A Appendix 3.A

At time 0, the dealer is assumed to have $I_c(0)$ units of cash and $I_s(0)$ units of stock. At time t, $I_c(t)$ and $I_s(0)$ are

$$I_c(t) = I_c(t) + p_a N_a(t) - p_b N_b(t) \tag{A.1.1}$$

and

$$I_s(t) = I_s(t) + N_b(t) - N_a(t). \tag{A.1.2}$$

When $Q_k(t)$ is the probability that $I_c(t) = k$ and $R_k(t)$ is the probability that $I_s(t) = k$, the dealer have exactly k units at time t from following three cases.

1. the dealer held exactly k − 1 units at time t − Δt and in the Δt an order to sell one unit to him arrives; or
2. the dealer held exactly k units at time t − Δt and in the Δt no order arrives; or
3. the dealer held exactly k + 1 units at time t − Δt and in the Δt an order to buy one unit to him arrives.

When a unit of cash arrives[1] with rate $\lambda_a p_a$ and departs at rate $\lambda_b p_b$, the probability that the dealer exactly k units of cash at time t, $Q_k(t)$, is composed of the probabilities of three cases.

1. the probability the dealer had $k - 1$ units of cash and in the interval Δt receives a cash inflow is $Q_{k-1}(t\text{-} \Delta t)[\lambda_a(p_a)p_a \ \Delta t][1\text{-}\lambda_b(p_b)p_b\Delta t]$
2. the probability the dealer had k units of cash and in the interval Δt no cash flow occurs is $Q_k(t\text{-} \Delta t)[1\text{-} \ \lambda_a(p_a)p_a \ \Delta t][1\text{-}\lambda_b(p_b)p_b\Delta t]$
3. the probability the dealer had $k+$ 1units of cash and in the interval Δt receives a cash inflow is $Q_{k+1}(t\text{-} \Delta t)[\lambda_b(p_b)p_b \ \Delta t][1\text{-}\lambda_a(p_a)p_a\Delta t]$

Thus,

[1]use the expression 'arrive' and 'depart' since the order process is a Poisson process. Simply $\lambda_a p_a$ means cash-in (or stock-out) rate and $\lambda_b p_b$, means cash-out(or stock-in) rate in view of the dealer.

$$Q_k(t) = Q_{k-1}(t-\Delta t)[\lambda_a(p_a) * p_a \Delta t][1 - \lambda_b(p_b) * p_b \Delta t]$$
$$+ Q_k(t-\Delta t)[1 - \lambda_a(p_a) * p_a \Delta t][1 - \lambda_b(p_b) * p_b \Delta t]$$
$$+ Q_{k+1}(t-\Delta t)[\lambda_b(p_b) * p_b \Delta t][1 - \lambda_a(p_a) * p_a \Delta t].$$

(A.1.3)

To calculate the time derivative of the probability $Q_k(t)$, take the limit as $\Delta t \to 0$ of $[Q_k(t)$ - Q_k $(t\text{-}\Delta t)]/\Delta t$.

$$\frac{\partial Q_k(t)}{\partial t} = Q_{k-1}(t)[\lambda_a(p_a) * p_a]$$
$$+ Q_k(t)[\lambda_a(p_a) * p_a][\lambda_b(p_b) * p_b]$$
$$+ Q_{k+1}(t)[\lambda_b(p_b) * p_b].$$

(A.1.4)

Before one get the solution of (A.1.4), the Gamler's Ruin Problem shows how to solve for (A. 1.4). In a typical gambler's ruin problem, the gambler has some initial wealth and wagers until he either reaches a certain threshold level or loses all his money. The failure probability = f {the odds of winning, the odds of losing, the threshold level, the initial wealth }

In case of the dealer's problem, the dealer does not bound its maximum gain, but it stops only when it loses its money(or stock). Provided the odds of winning exceed the odds of losing, the ultimate failure probability is

$$\left(\frac{\text{odds of losing} \times \text{amount of loss}}{\text{odds of winning} \times \text{amount of gain}}\right)^{\text{initial wealth position}}$$

Suppose that a dealer gains a unit of stock with probability p and loses a unit of stock with probability q, where $p > q$. In addition, the dealer has an initially S_0 units of stock. The probability that he runs out of stock at time t given the dealer currently has S units is $Pr\{F|S\}$.

$$\text{Pr}\{F|S\}=q^*\text{Pr}\{F|S-1\}+p^*\text{Pr}\{F|S+1\}. \quad \text{(A.1.5)}$$

The solution of (A.1.5) yields the general expected failure probability

$$\Pr\{\mathrm{F}|\mathrm{S}_0\} = \left(\tfrac{q}{p}\right)^{S_0}$$

The solution of (A.1.4) yields the failure probability from running out of cash, which is

$$\begin{aligned}\lim_{t\to\infty} Q_0(t) &\approx \left(\frac{\lambda_b(p_b) * p_b}{\lambda_a(p_a) * p_a}\right)^{I_c(0)/\bar{p}} && \text{if } \lambda_a(p_a) * p_a > \lambda_b(p_b) * p_b \\ &= 1 && \text{otherwise}\end{aligned} \tag{A.1.6}$$

where $\bar{p}$ is defined to be the mean of the bid and ask prices.

The corresponding stock failure probability is

$$\begin{aligned}\lim_{t\to\infty} R_0(t) &= \left(\frac{\lambda_a(p_a)}{\lambda_b(p_b)}\right)^{I_s(0)} && \text{if } \lambda_a(p_a) < \lambda_b(p_b), \\ &= 1 && \text{otherwise}\end{aligned} \tag{A.1.7}$$

B Appendix 3.B

$$EU(\tilde{W}^*)=EU(\tilde{W}) \qquad \text{(B 1.1)}$$

where $\tilde{W}^*$ = terminal wealth of the initial portfolio $\tilde{W}$ =terminal wealth of the new portfolio after the transaction

For right side of (B 1.1),

$$\tilde{W}^*=W_0\left[1+k\tilde{R}_e+\frac{Q_p}{W_0}\tilde{R}_p+\left(1-k-\frac{Q_p}{W_0}\right)R_f\right] \qquad \text{(B 1.2.)}$$

$$= W0(1+\tilde{R}^*) \qquad \text{(B 1.3)}$$

W_0 = initial wealth

$\tilde{R}^*$ = the rate of return on the initial portfolio.

k = optimal fraction of the dealer's wealth invested in the optimal efficient portfolio of risky assets, E, and assumed to be constant.

Q_p = true dollar value of stocks in trading account. If Q_p =0, the initial portfolio is the desired one. Q_p ><0 according to whether the dealer has long or short position in the trading account.

$\tilde{R}_e$ = return on portfolio E

$\tilde{R}_p$ = rate of return on the trading account

R_f = risk free rate

For the left side of (B 1.1),

Terminal wealth under the new portfolio is given by:

$$\tilde{W} = W_0(1+\tilde{R}^*) +(1+\tilde{R}_i)Q_i - (1+R_f)(Q_i - C_i) \qquad \text{(B 1.4)}$$

Q_i = true dollar value of the transaction in stock i, the stock in which immediacy is being provided. Negative values indicate a sale; positive values, a purchase.

$\tilde{R}_i$ = rate of return on stock i.

C_i = present dollar cost to the dealer of trading the amount Q_i. C_i is positive or negative according as the transaction in stock i raises or lowers the costs of holding the inventory Q_p.

Then,

$$E[U(W_0(1+\tilde{R}^*))]=E[U(\tilde{W})]. \qquad \text{(B 1.5) or (3.3)}$$

Using a Taylor series around the respective means and dropping terms of order higher than two and setting R_f=0 yields the dollar holding cost function.

$$E\left[U(\bar{W}^*)+U'(\bar{W}^*)(W^*-\bar{W}^*)+\frac{1}{2}U''(\bar{W}^*)(W^*-\bar{W}^*)^2\right]$$

$$=E\left[U(\bar{W})+U'(\bar{W})(W-\bar{W})+\tfrac{1}{2}U''(\bar{W})(W-\bar{W})^2\right] \quad \text{(B 1.6)}$$

where the bar(-) over a variable indicates expected value.

Using (B 1.3) and (B 1.4), (B 1.6) is re-expressed by

$$U(\bar{W}^*)+0+\frac{1}{2}U''(\bar{W}^*)W_0^2\sigma_*^2$$

$$=U(\bar{W})+0+\tfrac{1}{2}U''(\bar{W})\left[W_0^2\sigma_*^2+2W_0Q_i cov(R^*,R_i)+Q_i^2\sigma_i^2\right] \quad \text{(B 1.7)}$$

σ_*^2 and σ_i^2 are the variance of rate of return of the initial portfolio and of stock i.

By approximating $U''(\bar{W}^*)=U''(\bar{W})$ and $\frac{U(\bar{W})-U(\bar{W}^*)}{U'(\bar{W}^*)}=\bar{W}-\bar{W}^*$, (B 1.7) is written as

$$\tfrac{1}{2}\tfrac{z}{W_0}\left[Q_i^2\sigma_i^2+2W_0Q_i cov(R^*,R_i)\right]-\left[\bar{W}-\bar{W}^*\right]=0 \quad \text{(B 1.8)}$$

Meanwhile, z is the Pratt index of relative risk aversion $-\frac{U''(\bar{W}^*)}{U'(\bar{W}^*)}W_0$.

$$\bar{W}-\bar{W}^*=Q_i(\bar{R}_i-R_f)+C_i(1+R_f) \quad \text{(B 1.9)}$$

comes from (B 1.3) and (B 1.4) after taking expectation.

$$cov(R^*,R_i)=k\sigma_{ie}+\tfrac{Q_p\sigma_{ip}}{W_0} \quad \text{(B 1.10)}$$

is computed from the definition of $\tilde{R}^*$.

When the slope of the dealer's indifference curve (B 1.12) is equal to the slope of the desired opportunity set (B 1.13),

$$k=\tfrac{\bar{R}_e-R_f}{z\sigma_e^2}. \quad \text{(B 1.11)}$$

To get the slope of the dealer's indifference curve, if Q_p=0, terminal wealth is

$$\tilde{W}^*=W_0\left[1+k\tilde{R}_e+(1-k)R_f\right]=W0(1+\tilde{R}^*)$$

Taking differential of $E[U(W_0(1+\tilde{R}^*))]=U(\bar{W}^*)+0+\frac{1}{2}U''(\bar{W}^*)\,W_0^2\sigma_*^2$ yields

$$dEU(W^*) = \frac{\partial U(W^*)}{\partial \bar{W}^*} \cdot \frac{\partial(\bar{W}^*)}{\partial \bar{R}^*} d\bar{R}^* + U''(\bar{W}^*)W_0^2\sigma_* d\sigma_*$$

note that ${}^{\partial \bar{W}^*}/_{\partial \bar{R}^*} = W_0$ and $\sigma_* = k\sigma_e$.

Set the differential equal to zero and solve for the slope of the indifference curve:

$$\frac{d\bar{R}^*}{d\sigma_*} = -\frac{U''(\bar{W}^*)}{U'(\bar{W}^*)} W_0 k\sigma_e. \quad \text{(B 1.12)}$$

The slope of the opportunity set is

$$\frac{(\bar{R}_e - R_f)}{\sigma_e} \quad \text{(B 1.13)}$$

Using (B 1.8), (B 1.9), (B 1.10), and (B 1.11) yields

$$\frac{1}{2}\frac{z}{W_0}Q_i^2\sigma_i^2 + \frac{z}{W_0}Q_iQ_p\sigma_{ip} - C_i(1+R_f) - Q_i\left[(\bar{R}_i - R_f) - (\bar{R}_e - R_f)\frac{\sigma_{ie}}{\sigma_e^2}\right] = 0 \quad \text{(B 1.14)}$$

(B 1.14) is re-written as

$$C_i = \frac{\frac{1}{2}\frac{z}{W_0}Q_i^2\sigma_i^2 + \frac{z}{W_0}Q_iQ_p\sigma_{ip} - Q_i\left[(\bar{R}_i - R_f) - (\bar{R}_e - R_f)\frac{\sigma_{ie}}{\sigma_e^2}\right]}{(1+R_f)} \quad \text{(B 1.15)}$$

In the portfolio equilibrium for the dealer,

$$(\bar{R}_i - R_f) = (\bar{R}_e - R_f)\frac{\sigma_{ie}}{\sigma_e^2} \quad \text{(B 1.16)}$$

If one assume R_f=0, then (B 1.15) with using (B 1.16), leads to

$$C_i = \frac{z}{W_0}\sigma_{ip}Q_pQ_i + \frac{1}{2}\frac{z}{W_0}\sigma_i^2Q_i^2 \quad \text{(3.4) or (B 1.17)}$$

The percentage cost function is

$$\frac{C_i}{Q_i} = \frac{z}{W_0}\sigma_{ip}Q_p + \frac{1}{2}\frac{z}{W_0}\sigma_i^2Q_i = c_i(Q_i) \quad \text{(3.5)}$$

C Appendix 3.C

Writing out the partial differential equation implied in the (3.11) yields

$$\max_{a,b}(dJ/dt) = J_t + LJ + \max_{a,b}\left\{ \begin{array}{l} \lambda_a[J(F+(p+a)*Q, I-pQ, Y) - J(F,I,Y)] \\ +\lambda_b[J(F-(p-b)*Q, I+pQ, Y) - J(F,I,Y)] \end{array} \right\} = 0$$

(C.1.1)

where Q is transaction size, J_t is the time derivative and L is the operator defined as

$$LJ = J_F rF + J_I r_I I + J_Y r_Y Y + \tfrac{1}{2}J_{YY}\sigma_Y^2 Y^2 + \tfrac{1}{2}J_{II}\sigma_i^2 I^2 + J_{IY}\sigma_{IY} IY \tag{C.1.2}$$

To get (C.1.2) Ito's Lemma is employed. Iro's Lemma is;

Suppose there is a smooth function $Y = f(x,t)$, where t is time and x is well-defined Ito process, $dx = \mu dt + \sigma dz$. If one wants to maximise Y by choosing x, then one need to take the first order derivative of Y, and Ito's Lemma provides this.

$$dY = \tfrac{\partial f}{\partial t}dt + \tfrac{\partial f}{\partial x}dx + \tfrac{1}{2}\tfrac{\partial^2 f}{\partial x^2}(dx)^2$$
$$= \tfrac{\partial f}{\partial t}dt + \tfrac{\partial f}{\partial x}[\mu dt + \sigma dz] + \tfrac{1}{2}\tfrac{\partial^2 f}{\partial x^2}\sigma^2 dt$$
$$dY = \left[\tfrac{\partial f}{\partial t} + \tfrac{\partial f}{\partial x}\mu + \tfrac{1}{2}\tfrac{\partial^2 f}{\partial x^2}\sigma^2\right]dt + \tfrac{1}{2}\tfrac{\partial f}{\partial x}\sigma dz$$

which is Ito's lemma and gives the formula for finding the derivative of a function that depends on time and a stochastic process.

For better understanding, compare (C.1.2) and Ito's lemma.

$$\frac{\partial f}{\partial t} \Leftrightarrow J_t, \frac{\partial f}{\partial x}\mu \Leftrightarrow J_F rF + J_I r_I I + J_Y r_Y Y,$$

$$\frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma^2 \Leftrightarrow \frac{1}{2}J_{YY}\sigma_Y^2 Y^2 + \frac{1}{2}J_{II}\sigma_i^2 I^2 + J_{IY}\sigma_{IY} IY.$$

Since the dealer maximises expected utility, the expectation of the dZ in the (C.1.2) is equal to zero at optimum.

In the (C.1.1), the final term, the max term, appears to reflect the effect on the dealer's utility when transactions happen at the bid and ask prices.

Even though (C.1.1) including (C.1.2) will yield the solution, the $J(\cdot)$ function should have an explicit form to find the actual solution.

To get (3.12), when τ is the amount of time remaining to the horizon date T, replace t in (C.1.1) with τ. When τ is zero,

$$J(0, F, I, Y) = U(W). \tag{C.1.3}$$
$$J_t = -J_\tau \tag{C.1.4}$$

due to the transformation of the time variable.

Since aQ and bQ is small relative to Q, using first-order approximation and assuming that p=1 yield

$$\begin{aligned} J(F{+}Q{+}aQ,\ I - Q,\ Y) &= SJ + SJ_F aQ \\ J(F{-}Q{+}bQ,\ I + Q,\ Y) &= BJ + BJ_F bQ \end{aligned} \tag{C.1.5}$$

Where S is the sell operator which is defined as

$$\begin{aligned} SJ{=}S[J(F, I, Y)]&{=}J(F{+}Q{,}I - Q, Y) \\ SJ_F = S[J_F(F, I, Y)]&{=}J_F(F{+}Q{,}I - Q, Y) \end{aligned} \tag{C.1.6}$$

And B is the buy operator which is defined as

$$\begin{aligned} BJ{=}\ B[J(F, I, Y)]&{=}J(F{-}Q,\ I{+}Q,\ Y) \\ BJ_F = B[J_F(F, I, Y)]&{=}J_F(F - Q,\ I + Q,\ Y) \end{aligned}$$

The assumption that the same linear demand relation for dealer sales and purchases is given by

$$\lambda_a{=}\lambda(a) = \alpha - \beta{*}a \text{ and } \lambda_b = \lambda(b) = \alpha - \beta{*}b. \tag{C.1.7}$$

Substitution of (C.1.4), (C.1.5), and (C.1.7) into (C.1.1) yields (3.12),

$$J_\tau = LJ + \max_{a,b}\left\{\lambda(a)aQSJ_F - \lambda(a)[J - SJ] + \lambda(b)bQBJ_F - \lambda(b)[J - BJ]\right\}.$$

To get (3.13) and (3.14), find the solution for max term in the (3.12). The max term in the (3.12) can be restated as

$$\max_{a,b}\left\{(\alpha - \beta a)aQSJ_F - (\alpha - \beta a)[J - SJ] + (\alpha - \beta b)bQBJ_F - (\alpha - \beta b)[J - BJ]\right\}$$
$$\max_{a,b}\left\{\begin{array}{l} (-a^2\beta QSJ_F + a\alpha QSJ_F + a\beta[J - SJ] - \alpha[J - SJ]) \\ +(-b^2\beta QBJ_F + b\alpha QBJ_F + b\beta[J - BJ] - \alpha[J - BJ]) \end{array}\right\} \tag{C.1.8}$$

Performing the maximisation of (C.1.8) with respect to a and b yields
(3.13), $b^* = \frac{\alpha}{2\beta} + \frac{J(\bullet) - BJ(\bullet)}{2BQJ_F}$,
and
(3.14), $a^* = \frac{\alpha}{2\beta} + \frac{J(\bullet) - SJ(\bullet)}{2SQJ_F}$.

Chapter 4

Spread Decomposition Models

4.1 Spread Decomposition Model

Academic researchers have developed spread decomposition models to measure empirically each component of the bid-ask spread. This effort has led to the development of two types of spread decomposition models: the covariance-based models and the trade indicator models. George, Kaul, and Nimalenddran (1991) developed spread decomposition model based on a covariance within transaction returns while trade indicator models have been developed by Glosten and Harris (1988), Lin, Sanger, and Booth (1995), and Madhavan, Richardson, and Roomans (1997). Huang and Stoll (1997) develop a model that nests both type of models. Unlike other models, Huang and Stoll (1997) provide a model that decomposes all three components of the bid-ask spread while other spread decomposition models decompose the adverse selection component and other two components.

This chapter explains how spread decomposition models measure or estimate empirically the adverse selection component. The five most widely used spread decomposition models will be discussed: Glosten and Harris (1988), Madhavan, Richardson, and Roomans (1997), George, Kaul, and Nimalenddran (1991), Lin, Sanger, and Booth (1995), and Huang and Stoll (1997).

4.1.1 Glosten and Harris (1988)

Glosten and Harris (1988) assume that the spread incorporates two components: a transitory component[1] and an adverse selection component that is dependent on trade size. An advantage of Glosten and Harris (1988) is that it is consistent with Easley and O'Hara (1987)'s result that the adverse selection component of the bid-ask spread can increase with trade size[2].

The Glosten and Harris (1988) model starts with four relationships:

$$P_t = m_t + Q_t C_t, \tag{4.1}$$

$$m_t = m_{t-1} + Q_t Z_t + U_t, \tag{4.2}$$

$$C_t = c_0 + c_1 V_t, \tag{4.3}$$

$$Z_t = z_0 + z_1 V_t, \tag{4.4}$$

where the adverse selection component of a transaction at t is Z_t and the transitory component is C_t. P_t is the observed transaction price at time t. V_t is the number of shares traded in the transaction and m_t represents the intrinsic price. U_t is a white noise error process that captures public information arrival and a rounding error. Q_t is a trade indicator that is $+1$ if a transaction at t is initiated by a buyer and -1 if a transaction at t is initiated by a seller.

When Z_t is positive, a one unit buy order causes the intrinsic price m_t to move up by Z_t while a one unit sell order induces m_t to move down by $-Z_t$. Thus Z_t is expected to be a positive function of transaction size V_t and results in a permanent price change. On the contrary, the transitory component C_t is expected to be a linear function of V_t and leads to temporary price changes, which are negatively correlated and bounce within the quoted spread.

[1]Transitory component includes inventory holding and order processing components.

[2]See Chapter Three Section 3.4.4 Easley and O'Hara(1987)

Using Equations from (4.1) to (4.4), the observed price change ΔP_t can be expressed as

$$\begin{aligned}\Delta P_t &= P_t - P_{t-1} \\ &= m_t + Q_t C_t - (m_{t-1} + Q_{t-1} C_{t-1}) \\ &= c_0(Q_t - Q_{t-1}) + c_1(Q_t V_t - Q_{t-1} V_{t-1}) \\ &+ z_0 Q_t + z_1 Q_t V_t + U_t\end{aligned} \tag{4.5}$$

More concisely,

$$\Delta P_t = c_0 \Delta Q_t + c_1 \Delta Q_t V_t + z_0 Q_t + z_1 Q_t V_t + U_t \tag{4.6}$$

Equation (4.6) in the Glosten and Harris (1988) model can be seen to estimate the adverse selection component. The bid-ask spread is measured as the sum of the transitory and the adverse selection components, which is

$$s = 2(c_0 + c_1 V_t) + 2(z_0 + z_1 V_t). \tag{4.7}$$

Since the transitory component includes the inventory holding component and the order processing component, Glosten and Harris (1988) cannot isolate these two components. However the model does isolate the adverse selection component. Equation (4.8) captures percentage the adverse selection component, which is

$$\begin{matrix}\%Adverse \\ selection \\ \cos t\end{matrix} = 100 * \frac{Adverse\,Selection}{Spread} = 100 * \frac{2(z_0+z_1V_t)}{2(c_0+c_1V_t)+2(z_0+z_1V_t)} \ . \tag{4.8}$$

The weakness of the Glosten and Harris (1988) model is that when Equation (4.6) is estimated using Ordinary Least Squares (OLS), the coefficients may be inefficiently estimated. The reason for inefficient coefficients is that the variance of the error term U_t depends on the time interval between transaction t and $t-1$, the variance of the error term when the time interval is minutes is different from the variance of the error term when time interval is hours. An alternative to estimate the coefficients is Maximum Likelihood(ML). However, coefficients using ML may be biased since the Glosten and Harris (1988) model is dependent upon the specification of price discreteness. The reason is that ML is based on logarithm transformation and price discreteness is an obstacle for the logarithm transformation of price.

Glosten and Harris (1988), notwithstanding the weakness above, shows the presence of the positive adverse selection component in the bid-ask spread and considers that transaction size conveys private information, which is consistent with Easley and O'Hara (1987). In addition, Glosten and Harris (1988) indicates that the spread is linearly related to trade size.

4.1.2 George, Kaul, and Nimalendran (1991)

George, Kaul, and Nimalenddran (1991) argue that positively auto-correlated time-varying expected returns lead to biases in estimates of the spread and its components. Conrad and Kaul (1988)[1] find that a stationary first-order autoregressive process explains well the variation over time in expected returns and the extracted expected returns explain up to 26% fraction of the variance in realised returns. George, et al. (1991) utilises the return difference to remove the time-varying expected returns.

George, et al. (1991) assumes that all trades are of unit size or that the bid-ask spread is independent of trade size, despite the theoretical work of Easley and O'Hara (1987) that shows the importance of trade size. The ex ante probabilities of transacting at the bid or ask are assumed to be equal. In the George, et al. (1991) model, P_t is the observed transaction price at t, Q_t is the unobservable indicator for the bid-ask classification of P_t, in which $Q_t = +1$ if transaction at t is at the ask or $Q_t = -1$ if transaction at t is at the bid, M_t is the unobservable true price that reflects all publicly available information immediately following the transaction at t, E_t is the unobservable expected return for the period between transactions $t-1$ and t, conditional on all public information at $t-1$, U_t is the unobservable innovation in true prices due to the arrival of public information

[1] In addition, they find that the proportion magnitude of the variance in realised returns has a monotonic inverse relation with size and the degree of variation in expected returns also changes systematically over time.

between transactions $t-1$ and t and is a white noise. Similar to the Glosten and Harris (1988) model, George, et al. (1991) do not distinguish the inventory holding component and the order processing component.

When s_q represents the quoted spread of the market-maker, π is the unobservable proportion of the quoted spread due to order-processing costs and $1-\pi$ is the unobservable proportion of the quoted spread due to adverse selection George, et al. (1991) sets up the relations (4.9) and (4.10).

$$P_t = M_t + \pi(\frac{s_q}{2})Q_t \tag{4.9}$$

$$M_t = E_t + M_{t-1} + (1-\pi)\frac{s_q}{2}Q_t + U_t, \tag{4.10}$$

Equation (4.9) shows that the observed transaction price P_t is composed of the unobservable true price M_t and the quoted spread due to order-processing cost. Equation (4.10) shows that the unobservable true price M_t is the sum of the unobservable expected return at t E_t, the unobservable true price at $t-1$ M_{t-1}, the quoted spread due to adverse selection cost and the unobservable innovation U_t.

R_{iTt}[1], the continuously compounded transaction return of security i between time $t-1$ and t, is expressed as

$$R_{iTt} = E_{it} + \pi_i\frac{s_{qi}}{2}(Q_{it} - Q_{it-1}) + (1-\pi_i)\frac{s_{qi}}{2}Q_{it} + U_{it}. \tag{4.11}$$

The compounded transaction return R_{iTt} is composed of four terms: the unobservable expected return, the returns due to order-processing component, the returns due to the adverse selection component, and the error term which captures public information.

1

$$\Delta P_t = P_t - P_{t-1} = M_t + \pi\frac{s_q}{2}Q_t - M_{t-1} - \pi\frac{s_q}{2}Q_{t-1}$$
$$= E_t + (1-\pi)\frac{s_q}{2}Q_t + U_t + \pi\frac{s_q}{2}(Q_t - Q_{t-1}) = R_{iTt}$$

The bid quote return of security i R_{iBt} [1] is computed as

$$R_{iBt} = E_{it} + (1 - \pi_i)\frac{s_{qi}}{2}Q_{it} + U_{it}. \tag{4.12}$$

Equation (4.12) shows that the bid return for security i comprises the unobservable expected return, the return due to the adverse selection, and the error term. To calculate the bid quote return, George, et al. (1991) uses the bid quote posted simultaneously or immediately after a transaction at time t occurs.

From Equations (4.11) and (4.12), RD_{it}, the difference between R_{iTt} and R_{iBt}, is computed as

$$RD_{it} = R_{iTt} - R_{iBt} = \pi_i\frac{s_{qi}}{2}\left[Q_{it} - Q_{it-1}\right]. \tag{4.13}$$

The return differential RD_{it} eliminates the unobservable expected return at t E_t and does not depend on its variation. When the serial covariance of RD_{it} is computed as

$$Cov(RD_{it}, RD_{it-1}) = -\pi_i^2\frac{s_{qi}^2}{4}, \tag{4.14}$$

the serial covariance of RD_{it} provides a spread measure and is expressed as

$$\pi_i s_{qit} = 2\sqrt{-Cov(RD_{it}, RD_{it-1})} = \hat{s}_{it}. \tag{4.15}$$

Equation (4.15) is similar to the Roll implied spread in Roll (1984)[2]. The difference is that the transaction return TR_t in the Roll model is substituted with the return difference RD_t.

Simply when the ordinary least squared method is applied to estimate the components of the bid-ask spread, the following regression equation can be used.

$$\hat{s}_{it} = \alpha_i + \beta_i s_{qit} + \varepsilon_{it} \tag{4.16}$$

[1] Since the constant bid-ask spread s_q is assumed, the return of bid quotes is the same as that of M_t.

$$R_{Bt} = \left(M_t - \frac{s_q}{2}\right) - \left(M_{t-1} - \frac{s_q}{2}\right) = \Delta M_t = E_t + (1-\pi)\frac{s_q}{2}Q_t + U_t$$

[2] See Section 2.3.1. the Roll implied spread in the Chapter Two

In the equation (4.16), $\hat{s}_{it}$ is the computed bid-ask spread based on equation (4.15) and s_{qit} the quoted bid-ask spread. The regression equation (4.16) implies that the estimator of the slope coefficient $\hat{\beta}_i$ is an unbiased measure of ϕ_i and that the expected value of $\hat{\alpha}_i$ is equal to α_i which is zero. Since OLS is employed, $\hat{\alpha}_i$ and $\hat{\beta}_i$ are efficient estimators.

George, et al. (1991) shows that the Roll implied spread may overstate the bid-ask spread. George, et al. (1991) shows that the continuously compounded transaction returns of security i R_{iTt} includes the time variation in expected return E_{it}, suggesting that when transaction returns of security i are used for estimating the Roll implied spread, the expected return E_{it} will cause the Roll implied spread to be overstated. Whereas using the return difference RD_{it} provides less biased adverse selection estimators. Additional contribution is that when bid and ask quote data are available, the George, et al. (1991) model can easily estimate the proportion of the bid-ask spread due to order-processing costs (Equation (4.13)).

The George, et al. (1991) model, however, has some weaknesses. For example, the George, et al. (1991) model does not consider trade size. If transaction size conveys the private information of an informed trader, then the adverse selection cost based on unit size trading which George, et al. (1991) assumes can be biased. In other words, the unit size trading assumption can be problematic, since theoretically Easley and O'Hara (1987) demonstrates that trade size can convey information. Moreover, Barclay and Warner (1993) find that price changes are greater for medium size trades when they classify trade size into small, medium, and large categories. In depth quantities, Barclay and Warner find that informed traders prefer to transact medium-size, suggesting that informed traders act strategically and their behaviour causes the adverse selection component of the spread to depend on and increase with trade size.

An additional limitation of the George, et al. (1991) model is the assumption that transactions at the bid and ask prices are equally likely. Moreover, if buy and sell orders do not arrive equally, the return based on bid quote prices R_{iBt} is biased, which results in a biased return differential RD_{it}, again leading to a biased adverse selection component. Finally the George, et al. (1991) model does not consider the inventory holding component of the spread which, as shown in the section (3.3) of Chapter Three, is an important component of the spread.

4.1.2.1 The Extension of George, et al. (1991) by Neal and Wheatley (1998)

Neal and Wheatley (1998) adjust the George, et al. (1991) model to allow the proportional spread to vary through time and to allow the probability of a buy or sell order arriving to deviate from 0.5. With this modification, the George, et al. (1991) model changes from the covariance based model to the trade indicator model.

Thus the George, et al. (1991) model including Equations (4.9) and (4.10) is represented as

$$P_t = M_t + \pi(\tfrac{s_q}{2})Q_t \tag{4.9}$$

$$M_t = E_t + M_{t-1} + (1-\pi)\tfrac{s_q}{2}Q_t + U_t \tag{4.10}$$

$$MP_T = M_T \tag{4.17}$$

$$M_T = M_{T-1} + E_T + (1-\pi)\frac{s_{qit}}{2}Q_t + U_T \tag{4.18}$$

From Equations (4.9) to (4.18), P_t is the log of the transaction price and M_t is the log of the true price. Q_t represents $+1/-1$ buy-sell indicator and s_q is the percentage bid-ask spread. MP_T represents the midpoint of the bid and ask prices at time T and E_t represents the expected return from time t to $t-1$. The upper case T subscript represents the time for the quote midpoint and the lower case t subscript represents the time for the transaction. George, et al. (1991) assumes that the quote is computed immediately following the transaction at time t. U_t captures public information innovation.

The transaction return comes from Equation (4.9) and (4.10) and is computed as

$$TR_t = E_t + \pi\frac{s_q}{2}(Q_t - Q_{t-1}) + (1-\pi)\frac{s_q}{2}Q_t + U_t \tag{4.19}$$

The quote midpoint return from $T-1$ to T is computed from Equations (4.17) and (4.18) and is

$$MR_T = E_T + (1-\pi)\frac{s_q}{2}Q_t + U_T \tag{4.20}$$

Subtracting the quote midpoint return (4.20) from the transaction return (4.19) yields a return differential like Equation (4.21).

$$RD_{it} = \frac{\pi}{2}\left(Q_{it} - Q_{it-1}\right) * s_{qit} + V_t. \tag{4.21}$$

where $V_t = (E_t - E_T) + (U_t - U_T)$ is an error term.

When the constant spread assumption is relaxed, Equation (4.21) is transformed into

$$RD_{it} = \alpha_i + \frac{\beta_i}{2}\left(Q_{it} * s_{qit} - Q_{it-1} * s_{qit-1}\right) + V_t \tag{4.22}$$

The regression equation (4.22) implies that the estimator of the slope coefficient, $\hat{\beta}_i$, is an unbiased measure of ϕ_i and that the expected value of $\hat{\alpha}_i$ is equal to α_i which is zero. In addition, $\hat{\alpha}_i$ and $\hat{\beta}_i$ are efficient estimators. Neal and Wheatley do not employ the serial covariance of return differential between t and $t-1$ which is used in the George, et al. (1991) model.

4.1.3 Lin, Sanger, and Booth (1995)

Lin, Sanger, and Booth (1995) developed a simple method for estimating the adverse selection component of the bid-ask spread. Lin, et al. (1995) model the transaction process which incorporates informed trader's information: The superior information of the informed trader is revealed by a trade at time t. The revelation of information leads to quote revisions after the trade at t. When quote revisions occur, possible information asymmetry between market-maker and the informed trader is reflected as a function of the signed half effective spread[1]. For instance, the market-maker adjusts both bid and ask prices downward after informed trader's sell order. Thus the market-maker's behaviour to informed

[1] $B_{t+1} = B_t + \lambda z_t$ and $A_{t+1} = A_t + \lambda z_t$

trading results in mid-quote revision as a function of the one-half signed effective spread.

This transaction process is expressed as Equation (4.23) and (4.24). The quote change that reflects the adverse selection component is expressed as

$$Q_{t+1} - Q_t = \lambda z_t + \varepsilon_{t+1} \tag{4.23}$$

$$z_{t+1} = \theta z_t + \eta_{t+1} \tag{4.24}$$

where $Q_t = (A_t + B_t)/2$ is the quote mid point at t, $z_t = P_t - Q_t$ is the signed half effective spread (if a sell order arrives, then $z_t < 0$, but if a buy order arrives, then $z_t > 0$). λ is the proportion of the spread due to information asymmetry between market-maker and informed trader, namely, the adverse selection component. θ is the amount of order persistence, namely, the tendency of buy(sell) orders to follow buy(sell) orders. ε_t and η_t are uncorrelated disturbance terms and are a white noise. Lin, et al. (1995) adopts the logarithm transaction price and logarithm quote midpoint. This logarithm price provides an easy cross-sectional comparison and moderates the price discreteness problem since the logarithm price produces a continuously compounded rate of return for the dependent variable and a relative spread for the independent variable.

In the equations (4.23) and (4.24), two things should be noted. Since Lin, et al. (1995) assume a constant effective spread, if a buy(sell) trade occurs at time t, then $(1+\theta)/2$ is the probability that the next trade will be a buy(sell) order whereas $(1-\theta)/2$ is the probability of order reversal. Second, the disturbance term ε_t may incorporate the arrival of public information and market frictions such as price discreteness and the lag in price adjustment due to limit orders. Thus it is possible that these market frictions induce serial dependence in ε_t. However, Lin, et al. (1995) do not find evidence that autocorrelation of the disturbance term is a serious concern.

4.1.4 Madhavan, Richardson, and Roomans (1997)

Madhavan, Richardson, and Roomans (1997) developed a structural model for estimating the adverse selection component of the spread. To do so, they decompose intraday volatility into two components: public information and trading frictions. The Madhavan et al. model does not discern inventory holding and order processing costs, as was the case in Glosten and Harris (1988), George, Kaul, and Nimalenddran (1991), and Lin, Sanger, and Booth (1995).

Notations and relevant assumptions in the Madhavan et al. (1997) model are as follows:

x_t = an indicator variable for trade initiation, where $x_t = +1$ if a buyer initiates a trade at t and $x_t = -1$ if a seller initiates a trade at t. When both buyer and seller initiate trade t, $x_t = 0$.

$\lambda = Pr[x_t{=}0]$, the unconditional probability that the transaction occurs within the quoted spread

Buys and sells are assumed to be equally likely to occur, thus $E[x_t]{=}0$ and $Var[x_t]{=}(1{-}\lambda)$.

Changes in beliefs are assumed to arise from two sources: (1) new information announcements ε_t which cause innovation in beliefs without any trading volume and (2) order flow, θ (x_t - E [x_t | x_{t-1}]) , which positively correlates with private information about fundamental asset value.

ε_t= the innovation in beliefs between $t-1$ and t, where ε_t is assumed to be an independent and identically distributed random variable with mean zero and variance σ_t^2.

$\theta(x_t{-}E[x_t|x_{t-1}])$ = the change in beliefs due to order flow, where $(x_t - E[x_t|x_{t-1}])$ is the surprise in order flow and θ is the degree of information asymmetry.

γ= Pr$[x_t = x_{t-1}|x_{t-1} \neq 0]$ is the probability that a transaction at the ask(bid) follows a transaction at the ask(bid).

$\rho = \frac{E[x_t, x_{t-1}}{var[x_{t-1}]}$ indicates the first-order autocorrelation of the trade initiation variable and implies $\rho = 2\gamma-(1-\lambda)$. Then the conditional expectation $E[x_t|x_{t-1}]$ $= \rho x_{t-1}$ [1].

[1] If $x_{t-1} = 0$, $E[x_t|x_{t-1}] = 0$.
If $x_{t-1} = 1$, $E[x_t|x_{t-1}] = Pr[x_t = 1|x_{t-1} = 1] - Pr[x_t = -1|x_{t-1} = 1] = \gamma - (1-\gamma-\lambda) = \rho$.

μ_t= the post-trade expected value of the stock conditional upon public information and the trade initiation variable.

$$\mu_t = \mu_{t-1} + \theta(x_t - E[x_t|x_{t-1}]) + \varepsilon_t \tag{4.25}$$

P_t^a and P_t^b = the pre-trade ask and bid prices at time t.

$$P_t^a = \mu_{t-1} + \theta(1 - E[x_t|x_{t-1}]) + \phi + \varepsilon_t \text{ since } x_t = +1 \tag{4.26}$$

$$P_t^b = \mu_{t-1} + \theta(-1 - E[x_t|x_{t-1}]) - \phi + \varepsilon_t \text{ since } x_t = -1 \tag{4.27}$$

where ϕ is the dealer's compensation for transaction costs, inventory holding costs, and risk bearing costs so that ϕ captures the temporary effect of order flow on prices.

P_t = the transaction price of the security at time t. When transactions are assumed to be executed at the midquote like $(P_t^a + P_t^b)/2$,

$$P_t = \mu_t + \phi x_t + \xi_t, \tag{4.28}$$

where ξ_t is an independent and identically distributed random variable with mean zero and captures the effect of stochastic rounding errors.

It is simple to derive the model for estimating the adverse selection component. When order size is assumed to be fixed, combining Equation (4.26) with Equations (4.25) yields

$$P_t = \mu_{t-1} + \theta(x_t - E[x_t|x_{t-1}]) + \phi x_t + \varepsilon_t + \xi_t \tag{4.29}$$

Equation (4.29) shows that if one knows the unobservable μ_{t-1}then one may estimate easily the adverse selection component. However, one does not know μ_{t-1} so Madhavan et al. (1997) transform Equation (4.29) into the differential form of Equation (4.30). With using Equation (4.28) at time $t-1$ and the

If $x_{t-1} = -1$, $E[x_t|x_{t-1}] = Pr[x_t = 1|x_{t-1} = -1] - Pr[x_t = -1|x_{t-1} = -1] = -(\gamma - (1-\gamma-\lambda)) = -\rho$

conditional expectation $E[x_t|x_{t-1}] = \rho x_{t-1}$, Equation (4.29) is transformed into a price differential form,

$$P_t - P_{t-1} = (\phi + \theta)x_t - (\phi + \rho\theta)x_{t-1} + \varepsilon_t + \xi_t - \xi_{t-1} \tag{4.30}$$

Equation (4.30) is a basic model of Madhavan et al. (1997) and an empirically estimable equation since it does not include the unobservable prior belief μ_{t-1}.

Madhavan et al. (1997) adopts the Generalised Method of Moments (GMM)[1] to estimate four parameters (θ, π, λ, ρ). Four parameters are selected to minimise a criterion function based on the orthogonality restrictions implied by Madhavan et al. (1997).

In the case of $m_t = P_t - P_{t-1} - (\pi + \theta)x_t + (\pi + \rho\theta)x_{t-1}$, following population moments help to identify four parameters and a constant drift α:

$$E\begin{pmatrix} x_t x_{t-1} - x_t^2 \rho \\ |x_t| - (1 - \lambda) \\ m_t - \alpha \\ (m_t - \alpha)x_t \\ (m_t - \alpha)x_{t-1} \end{pmatrix} = 0. \tag{4.31}$$

The first term in Equation (4.31) is the autocorrelation in trade initiation, the second term defines the crossing probability, the third term defines constant drift term α as the average pricing error, the fourth and the final terms are the OLS normal equations.

After four parameters are estimated, the estimated implied spread is calculated by

$$S = 2(\theta + \phi) \tag{4.32}$$

When the percentage adverse selection and the percentage temporary components are computed as spread proportions, two components are expressed

[1]GMM is preferred to Maximum Likelihood method or Ordinary Least Square method since GMM does not need to make strong distributional assumptions, that is, the variables of interest can be serially correlated and conditionally heteroscedastic. Thus, if the distribution about variable of interest is known, ML method can be recommended. If not, however, GMM method will be the best chosen.

as

$$\%\,\theta = \frac{2\theta}{2(\theta+\phi)} \tag{4.33}$$

$$\%\,\phi = \frac{2\phi}{2(\theta+\phi)} \tag{4.34}$$

4.1.5 Huang and Stoll (1997)

Huang and Stoll (1997) proposed a model which decomposes the bid-ask spread into the adverse selection, inventory holding, and order processing components. This is an important progression as it is the first paper to model simultaneously all three components. Huang and Stoll (1997) argue that their model subsumes the previously explained four models[1] and that their model is a more general approach decomposing the bid-ask spread since the Huang and Stoll (1997) model discerns the inventory holding component and the order processing component[2].

In the Huang and Stoll (1997) model V_t is the unobservable fundamental value of the stock when transaction costs do not exist. M_t is the quote midpoint which is computed from the bid/ask quotes prevailing before a transaction. P_t represents the transaction price at time t. Q_t is the buy-sell trade indicator variable for the transaction price P_t, where $Q_t = +1$ if the transaction is initiated by a buyer and occurs above M_t; $Q_t = -1$ if the transaction is initiated by a seller and occurs below M_t; $Q_t = 0$ if the transaction occurs at M_t

Then V_t , M_t and P_t evolve as follows:

$$V_t = V_{t-1} + \alpha\frac{S}{2}Q_{t-1} + \varepsilon_t, \tag{4.35}$$

where S is the constant spread, α is the percentage of the half spread attributable to the adverse selection cost, and ε_t is the serially uncorrelated public

[1] Glosten and Harris(1988), George, Kaul, and Nimalenran(1991), Lin, Sanger, and Booth(1995), Madhavan, Richardson, and Roomans(1997)

[2] Huang and Stoll(1997) also provide the Huang and Stoll (1997) two-way decomposition model where discerns order processing component with other components (adverse selection and inventory holding). Since the Huang and Stoll (1997) two-way decomposition model does not distinguish the adverse selection and inventory holding components, the Huang and Stoll (1997) two-way decomposition model is not addressed here

information shock.

$$M_t = V_t + \beta \frac{S}{2} \sum_{i=1}^{t-1} Q_i, \tag{4.36}$$

where β is the proportion of the half spread attributable to inventory holding costs, $\sum_{i=1}^{t-1} Q_i$ represents the cumulated inventory from the market open till time $t-1$, and Q_1 represents the initial inventory for the day.

$$P_t = M_t + \frac{S}{2} Q_t + \eta_t, \tag{4.37}$$

where the error term η_t captures the deviation of the observed half spread $P_t - M_t$ from the half constant spread $S/2$.

The Huang and Stoll (1997) model starts with the conditional expectation of the trade indicator at time $t-1$ given Q_{t-2}. The conditional expectation of quote at $t-1$ is computed as

$$E(Q_{t-1}|Q_{t-2}) = (1 - 2\pi) Q_{t-2}, \tag{4.38}$$

since $Q_{t-1} = Q_{t-2}$ with probability $(1-\pi)$and $Q_{t-1} = -Q_{t-2}$ with probability π. π represents the probability that the trade at t is opposite in sign to the trade at $t-1$.

When the asset market is assumed to know the conditional expectation of (4.38), the change in the fundamental value ΔV_t is expected to be

$$\Delta V_t = \alpha \frac{S}{2} \left(Q_{t-1} - (1 - 2\pi) Q_{t-2} \right) + \varepsilon_t, \tag{4.39}$$

Note that the expectation of ΔV_t conditional on the information after Q_{t-2} is observed but before Q_{t-1} is observed should be zero. i.e.

$$E(\Delta V_t | V_{t-1}, Q_{t-2}) = 0 \tag{4.40}$$

Combining Equations (4.36) and (4.39) yields the change in the mid-point quote ΔM_t as

$$\begin{aligned}\Delta M_t &= \alpha\frac{S}{2}\left(Q_{t-1} - (1-2\pi)Q_{t-2}\right) + \beta\frac{S}{2}Q_{t-1} + \varepsilon_t \\ &= \frac{S}{2}(\alpha+\beta)Q_{t-1} - \frac{S}{2}\alpha(1-2\pi)Q_{t-2} + \varepsilon_t\end{aligned} \tag{4.41}$$

Taking an expectation of ΔM_t conditional on the information that is available after M_{t-1} is observed but before Q_{t-1} and M_t are observed is

$$E(\Delta M_t | M_{t-1}, Q_{t-2}) = \beta\frac{S}{2}(1-2\pi)Q_{t-2}. \tag{4.42}$$

Equation (4.42) shows that the expected quote midpoint change does not depend on the adverse selection component.

The Huang and Stoll (1997) model has two estimable equations (4.38) and (4.41) since the unobservable fundamental value of the stock is not included. To estimate the spread components, the constant spread of Equation (4.41) is replaced with the observable posted quote. Equation (4.41) is changed into the following variant.

$$\Delta M_t = (\alpha+\beta)\frac{S_{t-1}}{2}Q_{t-1} - \alpha(1-2\pi)\frac{S_{t-2}}{2}Q_{t-2} + \varepsilon_t \tag{4.43}$$

The final estimable equations of Huang and Stoll (1997) are

$$E(Q_{t-1}|Q_{t-2}) = (1-2\pi)Q_{t-2} \tag{4.44}$$

$$\Delta M_t = (\alpha+\beta)\frac{S_{t-1}}{2}Q_{t-1} - \alpha(1-2\pi)\frac{S_{t-2}}{2}Q_{t-2} + \varepsilon_t, \tag{4.45}$$

where Q_t is the buy-sell indicator for the transaction price P_t, and π is the probability that the trade at time t is opposite in sign to the trade at $t-1$. M_t is the midpoint of the quote that prevails just before the transaction at time t. S_t is the posted spread just prior to the transaction. α is the percentage of the half spread attributable to the adverse selection costs and β is to the inventory cost. The order processing cost is equivalent to $(1-\alpha-\beta)$. Huang and Stoll

(1997) employs Generalized Method of Moments[1] to estimate the parameters of Equations (4.44) and (4.45).

The advantage of Huang and Stoll (1997) is that all three theoretical components of the bid-ask spread are estimated. However, the concern with Huang and Stoll (1997) is that the estimated adverse selection component frequently shows a negative sign, which is economically unreasonable. Even Huang and Stoll (1997) present negative adverse selection components in their paper. This economically unreasonable negative adverse selection cost makes Hatch and Johnson (2002) to choose the Madhavan et al. (1997) model instead of the Huang and Stoll (1997) model.

4.1.6 Discussion

Extensive research has been undertaken based on the empirical results of the spread decomposition models. Academic research, moreover, has selected spread decomposition models without knowing which spread decomposition model provide the best estimates of adverse selection component. Lack of knowledge about the validity of spread decomposition models has caused some academic research to depend on two or more spread decomposition models.

For example, when the debate about the impact of new regulation is associated with the findings based on spread decomposition models, the validity of spread decomposition models is crucial. According to the chosen spread decomposition model, findings about the impact of new regulation are sometimes contrary. In this case, the policy makers can not be certain of whether the new regulation comes into an effect that policy makers want. Thus, knowing the effectiveness of spread decomposition models provides policy maker with the properly testing tool for the impact of new regulation.

[1]GMM is preferred to Maximum Likelihood method or Ordinary Least Square method since GMM does not need to make strong distributional assumptions, that is, the variables of interest can be serially correlated and conditionally heteroscedastic. Thus, if the distribution about variable of interest is known, ML method can be recommended. If not, however, GMM will be the best chosen.

4.2 The Paper Used Spread Decomposition Models

Table 4.1 summarises a range of papers that use spread decomposition models. Table 4.1 shows that extensive financial research uses spread decomposition models for empirical studies. In addition, the arguments or findings of financial research rest on the empirical results of spread decomposition models. Table 4.1 does not include Flannery, Kwan, and Nimalendran (2004), in which spread decomposition models are used in the appendix for robustness[1].

For the sake of brevity, I choose some papers from Table 4.1 to show how spread decomposition models have been used.

Hatch and Johnson (2002) investigate whether acquisitions among NYSE specialist firms can affect the market quality of the stock specialist firms manage. Hatch and Johnson find that, when using the Madhavan, et al. (1997) model, stocks traded by acquiring specialist firms and targeted specialist firms show significant decreases in spread components such as the adverse selection and temporary components. Hatch and Johnson, however, find that the matched control stocks also show a decrease in trading costs and market quality.

Greene and Smart (1999) use the Madhavan, et al. (1997) model to address the question of how increased noise trading affects market liquidity and trading costs. They find an increase in liquidity and a decrease in the adverse selection component of the bid-ask spread when analysts recommend stocks in "The Wall Street Journal's Investment Dartboard" column.

Hegde and McDermott (2003) investigate the effect of market liquidity on a recent sample of stocks added to and deleted from the Standard and Poor's (S&P) 500 index. Using the Madhavan, et al. (1997) model and the Glosten and Harris (1988) model, they find that the decrease in the direct cost of transacting and a

[1]Flannery, Kwan and Nimalenran(2004) use the Glosten and Harris (1988) model and the Lin, et al. (1995) model not in their main context, but in Appendix A. They examine whether the market microstructure properties of U.S. banking firms' equity incorporate the opaqueness of bank assets. i.e. when market microstructure properties such as spread, adverse selection component, and trading volume etc exhibit the characteristics of shares, they examine whether banking firms' equity market features show that bank assets are relatively more opaque during the period 1990 through 1997. To do so, they use Glosten and Harris(1988) and Lin, et al (1995) model as one of market microstructure properties in the Appendix A.

Table 4.1: Papers used Spread Decomposition Models

* means that the authors change a little the basic model.
JFR is Journal of Financial Review, JCF Journal of Corporate Finance, JFM* Journal of Futures Markets, JF Journal of Finance, AFE Applied Financial Economics, JFE Journal of Financial Economics, JBFA Journal of Business Finance & Accounting, JREFE Journal of Real Estate Finance and Economics, JFM** Journal of Financial Markets, JEF Journal of Empirical Finance, JFI Journal of Financial Intermediation, JIFM Journal of International Financial Markets Institutions and Money, PBFJ Pacific-Basin Financial Journal.
Papers are selected randomly and there are other papers which used one of spread decomposition models.
GH stands for Glosten and Harris (1988), GKN for George, et al. (1991), LSB for Lin, et al. (1995), MRR for Madhavan, et al. (1997), HS for Huang and Stoll (1997)

	GH	GKN	LSB	MRR	HS
Ahn, Cai, Hamao, and Ho (2002)(JEF)	Y			Y	
Alexander, Ors, Peterson, and Seguin (2004)(JCF)	Y			Y	
Brockman and Chung (2001)(JFE)	Y		Y		Y
Chou and Chung (2006)(JFM*)	Y	Y		Y	Y
Chakravarty, Van Ness, and Van Ness (2005)(JBFA)		Y	Y		
Clarke, Fee, and Thomas (2004)(JCF)			Y		
Domowitz, Glen, and Madhavan (1998)(JF)		Y*			
Glascock, Hughes, and Varshney (1998)(JREFE)			Y		
Gibson, Singh, and Yerramilli (2003)(JFI*)					Y
Greene and Smart (1999)(JF)				Y	
Hatch and Johnson (2002)(JFE)				Y	
Hegde and McDermott (2003)(JFM**)	Y			Y	
Kim, Ko, and Noh (2002)(PBFJ)					Y
Kofman and Moser (1997)(AFE)		Y			
Kumar, Sarin, and Shastri (1998)(JF)		Y			
McDonald, Nixon, and Slawson (2000)(JREFE)					Y*
McInish, Van Ness, and Van Ness (2001)(JIFMIM)					Y
Wei (1992)(JFR)	Y				
Weston (2000)(JF)					Y

smaller decrease in the adverse selection component induce an improvement in liquidity of added stocks.

Gibson, Singh, and Yerramilli (2003) ask how market-makers and traders are affected by price discreteness caused by a change from one-sixteenth or one-eighth quotes to decimalisation. They find that the change to decimal pricing results in significantly tighter spreads, but does not have an impact on dollar spreads attributable to the adverse selection and the inventory holding costs. The Huang and Stoll (1997) model is used to compute each component of the spreads and is used to compare those components before and after decimalization.

Weston (2000) investigates whether the competitive structure of NASDAQ is improved after the Securities and Exchange Commission (SEC) in US implements the NASDAQ Order-Handling rule. Weston finds a decrease in spreads after new regulation is imposed. Moreover, Weston finds that the drop in spreads is not due to changes in the adverse selection costs and inventory holding components when Weston uses the Huang and Stoll (1997) model.

Wei (1992) examines intraday variations in trading activity and the bid-ask spread. He finds a U-shaped pattern for the adverse selection component and an inverted U-shaped pattern for transitory component of the Glosten and Harris (1988) model.

Kumar, Sarin, and Shastri (1998) investigate whether stock options have a beneficial impact on the market for the underlying securities. By using the George, et al. (1991) model, they find that the adverse selection component of the underlying stocks drops after option listings.

Spread decomposition models are also employed in corporate finance, since the adverse selection cost reflects information asymmetry. Brockman and Chung (2001) examine how managerial trading affects corporate liquidity. After controlling for concurrent changes in price, volume, and volatility variables, Brockman and Chung find that market depths decrease. They check this finding by decomposing spreads into the adverse selection component and the other components. They use the Lin, et al. (1995) model, the Glosten and Harris (1988) model, and the Huang and Stoll (1997) model. They find that managers' behaviours are similar to informed traders when repurchasing company shares takes place in open market.

Even though I refer to a small number of papers, Table 4.1 clearly shows that an extensive range of studies have employed spread decomposition models. Meanwhile, financial researchers employ different spread decomposition models for the same research topic and they find different results depending on spread decomposition models. This is addressed in the next section.

4.2.1 Conflicting Results

Even though each spread decomposition model has its own theoretical background, financial researchers are confronted with contradictory results based on different spread decomposition models. This means that empirical results vary depending on the model used. i.e. only one true adverse selection cost exists in the trading process of an asset, while different spread decomposition models provide differently estimated adverse selection costs. These raise question about the effectiveness of spread decomposition models.

Table 4.2 suggests a summary of papers providing conflicting results. Panel A of Table 4.2 summarises those papers that examined the effect of Regulation Fair Disclosure (Regulation FD)[1]. Regulation FD changes the informational environment in the trading procedure and triggers debates[2] about the impact of Regulation FD on market participants. Some support for Regulation FD (see for example Chiyachantana, Jiang, Taechapiroontong, and Wood (2004)).

[1]The Securities and Exchange Commission(SEC) in US adopts Regulation Fair Disclosure rule (Regulation FD) and makes Regulation FD effective on October 23 2000. The objective of FD is to deteriorate information asymmetry among investors. Before Regulation FD is imposed, companies do not disclose relevant information to securities analysts and institutional investors as well as to the general public at the same time. This conventional practice causes informational disadvantage to the general public such as individual investors in terms of the timing and content of information. Thus when companies unintentionally disclose material non-public information to selective audience, SEC requires companies to issue a press report or to file an 8-K form within 24 hours or before the opening of the next day of trading on the NYSE. When companies intentionally release their material information, companies have to disclose simultaneously to securities analysts, institutional investors, investing public as well.

[2]The group supporting Regulation FD argues that the asymmetric information cost faced by the general investing public is reduced by Regulation FD since Regulation FD requires companies to guarantee equal access to material non-public information. The group opposing Regulation FD argues that Regulation FD reduces information flow from companies, and thus less information flow leads to less informative prices, finally those who can access information through other channels will have a great benefit from their private information.

Table 4.2: Papers Have Conflicting Results Depending on Spread Decomposition Models

*1. Straser use the probability of i nformed trading(Easley, Kiefer, O'Hara and Paperman(1996 EKOP))
*2 HS are symbol for Huang and Stoll(1997), GH Glosten and Harris(1988), GKN George, Kaul, and Nimalendran(1991), LSB Lin, Sanger, and Booth(1995).
*3 AS represents Adverse Selection component of the bid-ask spread.
*4 Dennis and Weston uses other method: the price impact of a trade (Hasbrouck 1991) and the probability of informed trading (EKOP)

Panel A The Effect of Regulation Fair Disclosure			
	Model	Data	Findings
Straser(2002)[*1]	HS[*2]	randomly selected 130 among S&P 500 stocks for the period between July 18 2000 and January 31 2001	AS[*3] increases after the adoption of Regulation Fair Disclosure
Chiyachantana, Jiang, Taechapiroontong, and Wood(2004)	GH, GKN, LSB	The pre-FD period: Nov. 1. 1999 to Aug. 15. 2000. The post-FD period: Oct. 23. 2000. to Jul. 31. 2001. 1,125 firms that experience earning announcements	AS in the prior period of earning announcements decreases after Regulation Fair Disclosure becomes effective
Panel B The relation between Ownership and Adverse Selection Component			
	Model	Data	Findings
Heflin and Shaw(2000)	LSB, HS	260 firms in 1988-1989	A positive relation between block-holder ownership and AS
Dennis and Weston(2001)[*4]	HS	Roughly 5,500 firms from 4^{th} quarter of 1997 to 4^{th} quarter of 1998	A positive relation between institutional ownership and AS of HS
Sarin, Shastri, and Shastri(2000)	GH, GKN	786 Firms listed on the American Stock Exchange or New York Stock Exchange in 1985	Mixed results: AS of GH increases with institutional ownership, but AS of GKN decreases with it. When insider trading is conrolled, insider ownership has insignificant relation with AS
Jiang and Kim(2005)	GH, LSB	316 pairs of matched firms: US and Non-US stocks listed on NYSE	A negative relation between AS and institutional ownership for the non-US stocks

Other argue against Regulation FD (see Straser (2002)). Academic research finds varying impacts associted with Regulation FD[1].

Chiyachantana, Jiang, Taechapiroontong, and Wood (2004) focus on how Regulation FD affects trading activity and information asymmetry surrounding earning announcements. Employing the Glosten and Harris (1988) model, the George, et al. (1991) model and the Lin, et al. (1995) model, Chiyachantana et al. (2004) finds that the adverse selection components of the period prior to earning announcements decrease, compared to those in before Regulation FD is adopted. Chiyachantana et al. (2004) also finds that the adverse selection components in the days immediately after earning announcements decrease after Regulation FD is adopted.

On the other hand, Straser (2002) investigates whether Regulation FD induces companies to commit to higher or lower levels of voluntary disclosures by focusing on the changes in information asymmetry. Straser uses two information asymmetry proxies such as the probability of informed trading from Easley et al. (1996)[2] and the adverse selection component of the Huang and Stoll (1997) model. Straser finds a significant increase in the adverse selection component after the adoption of Regulation FD.

Panel B of Table 4.2 provides another examole of diverse results based on spread decomposition models. Heflin and Shaw (2000) and Dennis and Weston (2001) find a positive relationship between blockholding and the adverse selection component. Jiang and Kim (2005) and Sarin, Shastri, and Shastri (2000), however, do not find a positive relationship.

Heflin and Shaw (2000) examine the relation between block ownership and market liquidity. Heflin and Shaw find that firms who have greater blockholders[3] have larger adverse selection components when they employ the Lin, et al. (1995) model and the Huang and Stoll (1997) model. In support of Heflin and Shaw (2000), Dennis and Weston (2001) find that the adverse selection component of the Huang and Stoll (1997) model is positively correlated with the percentage of insider ownership and institutional ownership.

[1] For the view of policy makers, the impact of Regulation FD is important since policy makers want to protect the general investing public.

[2] I used this model in Chapter Six.

[3] Blockholders are entities holding at least 5% of a firm's outstanding common shares.

Jiang and Kim (2005) provides contrasting findings to Heflin and Shaw(2000) and Dennis and Weston(2000). Jiang and Kim (2005) find a negative relation between the adverse selection cost and institutional ownership when they investigate non-US stocks cross-listed on the New York Stock Exchange. When Jiang and Kim (2005) employ the Glosten and Harris (1988) model and the Lin, et al. (1995) model, they find that lower institutional ownership can explain a higher adverse selection component. Meanwhile, Sarin, Shastri, and Shastri (2000)[1] find a mixed relation between adverse selection costs and institutional ownership. When they estimate the adverse selection component using the Glosten and Harris (1988) model they find a negative relation between the adverse selection component and institutional ownership. However when they employ the George, et al. (1991) model, they find a positive relation. Therefore extensive research employs one or more of spread decomposition models and is confronted with contradicting findings.

Table 4.2 suggests that spread decomposition models do not provide invariant empirical results. Without knowing the effectiveness of spread decomposition models, financial researchers can have only a limited understanding about the topics they examine.

4.2.2 The Papers Evaluating Spread Decomposition Models

Researchers recognise an empirical concern shown in Section 4.2.1 that is associated with measuring the adverse selection component, and they attempt to test the effectiveness of spread decomposition models. For instance, Neal and Wheatley (1998), Van Ness, Van Ness, and Warr (2001), and De Winne and Majois (2004) examine the effectiveness of the adverse selection costs based on different spread decomposition models.

[1]When Sarin, Shastri, and Shastri(2000) investigate the relation between stock liquidity and ownership structure, they find that the adverse selection component increases with insider ownership given that they do not control the positive relation between the frequency of insider trading and the level of insider ownership while they find insignificant relation between the adverse selection component and insider ownership after they control for insider trading.

Neal and Wheatley (1998) test two models: Glosten and Harris(1988) and George, et al (1991). Neal and Wheatley employ the intuition that closed-end funds may have less information asymmetry about Net Asset Value (NAV) since the *Wall Street Journal* reports NAV every Monday. Thus, less information asymmetry about the liquidation value of closed-end funds will lead to lower adverse selection costs in the bid-ask spread. Neal and Wheatley compare the estimated adverse selection components of closed-end funds with those of control stocks.

Neal and Wheatley find that the estimated adverse selection components for closed-end funds are unexpectedly large and significant, but the adverse selection components for control stocks are much larger than those of closed-end funds. Estimates of the adverse selection component from the Glosten and Harris (1988) model are found to be on average 19%, while estimates of the adverse selection component from the George, et al. (1991) model is on average 52%. For control stocks, the Glosten and Harris (1988) model provides the adverse selection components that are 34% at the spread, while the George, et al. (1991) model suggests 65%.

Van Ness, Van Ness, and Warr (2001) evaluate five spread decomposition models[1] by examining the relation between estimated adverse selection components and other information asymmetry measures commonly used in corporate finance. Van Ness, et al (2001) find no relation between the adverse selection components and information asymmetry proxies such as analyst forecast errors, market-to-book, R&D expenses, and intangible asset. They suggest one of two explanations: (1) the adverse selection components of spread decomposition models are poorly measured or (2) information asymmetry proxies in corporate finance are poorly measured.

De Winne and Majois (2004) evaluate the effectiveness of spread decomposition models within an order-driven market like Euronext Brussels. Their sample is composed of 19 stocks belonging to the Belgian Bel20 index. De Winne and Majois first examine whether the estimated spread components are plausible and consistent with each other. Secondly, similar to Van Ness, et al (2001), they analyze whether the estimated components measure information

[1] Van Ness et al (2001) evaluate the same five models as I evaluate here.

asymmetry. They find that the adverse selection components of spread decomposition models except for the George, et al. (1991) model are highly correlated with each other. They argue that Huang and Stoll (1997) two-way decomposition model, Lin, et al. (1995), and Madhavan, et al. (1997) are the best models to employ in an order-driven market because there are not inventory components in the order-driven market. Meanwhile, when Relative hidden depth, Trade imbalance, Limit order proportion, Depth imbalance, and Volatility are used for proxies of information asymmetry and are compared with the estimated adverse selection components, De Winne and Majois find that only limited order proportion and Volatility have expected positive relation with estimated adverse selection components. De Winne and Majois also conclude like Van Ness, et al. (2001) that the proxies for information asymmetry may be poorly selected or spread decomposition models may measure poorly adverse selection components.

4.2.3 Summary

While an extensive literature employs spread decomposition models, few researches have evaluated the effectiveness of spread decomposition models. If there was a way of evaluating spread decomposition models, then academic researchers may better understand the information asymmetry surrounding trading process and policy makers may evaluate better the impact of new regulation on market participants.

Chapter 5

Empirical Evaluation of Spread Decomposition Models

5.1 Introduction

This chapter has an objective that is to provide an empirical hypothesis and a set of results that evaluate spread decomposition models in terms of the adverse selection cost. Additionally, I explore the magnitude of the adverse selection cost for ETFs. Chapter Four discusses the extensive employment of spread decomposition in the finance literature and also shows that little attention has been devoted to the effectiveness of these models. Part of the reason is that an informed trader's trading activities are, in practice, unknown and therefore one may not know the exact loss of the market-makers to informed trading, i.e. the adverse selection cost of the spread. Therefore, it is difficult to estimate the adverse selection cost without knowing how much informed trading takes places with the market-maker during the trading process. However, I postulate in this chapter that combining the theoretical contribution of Subrahmanyam (1991) and Gorton and Pennacchi (1993) with the empirical characteristics of ETFs provides a theoretically-supported empirical hypothesis to evaluate the spread decomposition models[1].

[1]Note that spread decomposition models provide measures for other components such as inventory holding and order processing components, which also are important topics in the empirical research. However, this chapter mainly concentrates on the adverse

5.2 Theoretical Backgrounds for an Empirical Hypothesis

This section reviews the contribution made by Subrahmanyam (1991), in which the basket security[1] has lower informed trading than average individual constituent security. In addition to Subrahmanyam, Gorton and Pennacchi (1993) provide the same intuition when they assume risk-averse traders.

5.2.1 Subrahmanyam (1991)

Using the framework of Kyle (1985)[2], in which a market-maker sets prices using a linear pricing rule, Subrahmanyam (1991) shows that the pricing parameter decreases with the number of informed traders. This decrease in the pricing parameter leads to collectively smaller losses of the liquidity traders because the lower pricing parameter gives liquidity traders lower variability of a security. Then Subrahmanyam demonstrates that liquidity traders have their lowest trading loss when they all trade in the basket security and this situation happens when the pricing parameter[3] of the basket security is less than the weighted sum of the pricing parameter for individual constituent securities[4].

The condition that the equilibrium pricing parameter of the basket security is less than the weighted sum of the equilibrium pricing parameter of securities is satisfied as the impact of informed trading based on a constituent security offsets the impacts of informed trading based on other constituent securities in the basket

selection component of spread decomposition models and evaluates adverse selection component effectiveness across spread decomposition models.

[1]Subrahmanyam (1991) defines a basket security as a security whose liquidation payoff is equal to a weighted average of the payoffs of the individual component securities. Thus the basket security represents the share of a portfolio that is composed of different assets.

[2]In Kyle (1985) model, there is a single informed trader, the market-maker and liquidity traders. After a single informed trader places market orders (the market-maker observes net order flow), the market-maker sets a price which is the expected value of the security. In this frame work Kyle shows that there is a rational expectation equilibrium and market prices will eventually incorporate all available information.

[3]The parameter is an equilibrium value when all discretionary liquidity traders concentrate on the basket security

[4]The weighted sum of individual parameters is an equilibrium value when liquidity traders transacts individual securities

security and the impact of security-specific informed trading is diversified away in the end. With the diversified effect of security-specific information on the basket security, liquidity traders face smaller trading losses due to informed trading when they trade in the basket security.

Put another way, Subrahmanyam demonstrates that there are three possible Nash equilibriums[1]: the first equilibrium is that all liquidity traders transact the basket security; the second equilibrium is that some liquidity traders transact only the basket security and the others trade individual securities; and the third equilibrium is that all liquidity traders focus on the individual securities.

Among these Nash equilibriums, the second is dominated by the other two because of higher transaction costs. Since the trading loss of liquidity traders is considered as their transaction cost, the total loss of liquidity traders to informed trading is greatest in the second Nash equilibrium, where some liquidity traders prefer trading the basket security to individual securities and others prefer the opposite. The reason for the higher trading loss of liquidity traders is that trading individual securities leads to higher transaction costs when individual securities have higher variability than the basket security. When they have lower variability than the basket security, trading the basket security leads to higher transaction costs for liquidity traders. Thus, in either case liquidity traders embraces higher transaction cost.

The third Nash equilibrium can be dominant when the component-security-specific information of the basket security has greater variability than the weighted security-specific information of individual securities. Since greater variability leads to greater transaction costs, trading the basket security means higher transaction cost than trading individual securities. However, as the number of individual securities in the basket security becomes large, it is highly likely that the security-specific information in the basket security has less impact, causing the variability of the basket security due to security-specific information to become less than the weighted variability of individual securities. The third

[1]The Nash equilibrium is a kind of optimal collective strategy in a game when two or more players are involved. In the Nash equilibrium, no player has anything to gain by changing only his or her own strategy. If each player has chosen a strategy and no player can benefit by changing that strategy while the other players keep theirs unchanged, then the current set of strategy choices and the corresponding payoffs constitute a Nash equilibrium.

Nash equilibrium occurs only when the variability of the basket is greater than the weighted variability of individual securities.

As a result, the first Nash equilibrium is dominant since the variability of the basket security is highly likely to be less than the weighted average variability of individual securities. Liquidity traders in the first Nash equilibrium minimise their trading costs by trading in the basket security as the number of individual constituent securities increases.

Subrahmanyam (1991) extends his model to the case where informed traders have common factor information[1] such as industry-specific private information. For example, an informed trader can have private information on the oil industry, even though he/she does not have such information about a specific oil company. Subrahmanyam shows that the existence of such factor-informed traders also leads to similar predictions that can be derived from security-specific information: i.e. liquidity traders face smaller trading loss to factor-informed traders when liquidity traders trade in the factor-based basket security than in individual securities. In addition to this prediction, it is easily confirmed that the basket security generates lower adverse selection costs than the factor-based basket security. One reason is that factor-specific information will be diversified away in the basket security[2], but not in the factor-based basket security. Meanwhile, security-specific information will be completely diversified away in both the basket security and the factor-based basket security.

5.2.1.1 The Basic Model with the Basket Security

Securities(Individual Security and the Basket Security): N securities are simultaneously traded over a single period. One period model is assumed, thus trading occurs at time 0 and is liquidated at time 1. When $\bar{S}_i$ is the value of security i at time 0 and S_i is the liquidation value at time 1, the value of security

[1] In Subrahmanyam (1991), factor information is related to industry information, and not overall economic information like the inflation rate. He assumes that factor-informed traders know private information on a specific industry.

[2] In this context, the basket security is the one that reflects the overall security market, and not a single industry.

i depends on two main factors: systematic factor and security-specific factor. i.e. the value of the security i at time 1 is expressed by

$$S_i = \bar{S}_i + \beta_i\gamma + \varepsilon_i, \tag{5.1}$$

where i=1,..., N, ε_i and γ have mutually independent normal distribution with zero mean. $\beta_i\gamma$ and ε_i are considered as the systematic and security-specific components of the security value, respectively. Thus γ is some systematic risk common to all securities and β_i represents the sensitivity of the value of security i to the common risk factor γ.

The basket security is defined by the security whose liquidation payoff is identical to a weighted average of the payoffs of its underlying constituent securities.

Then the value of the basket security S_m at time 1 is expressed by

$$S_m \equiv \sum_{i=1}^{N} w_i\bar{S}_i + \sum_{i=1}^{N} w_i\beta_i\gamma + \sum_{i=1}^{N} w_i\varepsilon_i, \tag{5.2}$$

where w_i is the weight of security i in the basket.

Types of traders (Informed Traders and Liquidity Traders): There are k_i numbers of informed traders in the security i who know signals about ε_i and γ at time 0. The informed traders can take their positions between individual security i and basket security.

Two types of liquidity traders exist: discretionary liquidity traders and non-discretionary liquidity traders. Discretionary liquidity traders are defined as traders who want to trade all securities and do not have any preference to individual securities and basket securities. Non-discretionary liquidity traders buy/sell a fixed order size in a particular security which is either the basket security or constituent securities. Liquidity traders (discretionary and non-discretionary) submit a random quantity z_i, which is normally distributed with mean zero and is independent of γ and ε_i.

Informed traders and liquidity traders submit their orders to the market-maker and do not know the market clearing price when they do so.

Equilibrium: Because of competition, the profit of the market-maker is expected to be zero and to absorb the net trade in the security ϖ_i. The price set by the market-maker is

$$P_i = E(S_i|\varpi_i) \tag{5.3}$$

Like Kyle(1984, 1985), the market-maker sets prices by a linear pricing rule:

$$P_i = \bar{S}_i + \lambda_i \varpi_i, \tag{5.4}$$

where the quantity λ_i represents the pricing parameter for the security i, thus $1/\lambda_i$ is a measure of the liquidity or the depth of the market for security i.

Lemma 1: Each informed trader j submits an optimal order for security i x_{ji} where

$$x_{ji} = \frac{\varepsilon_i}{(k_i+1)\lambda_i} \tag{5.5}$$

The equilibrium value of λ_i is given by

$$\lambda_i = \frac{1}{k_i+1}\sqrt{\frac{k_i Var(\varepsilon_i)}{Var(z_i)}} \tag{5.6}$$

(Proof)

The objective function of informed trader j is

$$E(x_{ji}(S_i - P_i)|\varepsilon_i) \tag{5.7}$$

When the informed trader j considers the order of all other informed traders as $\beta\varepsilon_i$, he maximises

$$\begin{aligned} E(x_{ji}(S_i - P_i)|\varepsilon_i) &= E(x_{ji}\{\bar{S}_i + \beta_i\gamma + \varepsilon_i - (\bar{S}_i + \lambda_i\varpi_i)\}|\varepsilon_i) \\ &= E(x_{ji}\{\beta_i\gamma + \varepsilon_i - \lambda_i\varpi_i\}|\varepsilon_i) \\ &= E(x_{ji}\varepsilon_i - x_{ji}\lambda_i(x_{ji} + (k_i-1)\beta\varepsilon_i + z_i)|\varepsilon_i) \end{aligned} \tag{5.8}$$

Maximising the objective function of informed trader j with respect to x_{ji} produces

$$x_{ji} = \left(\frac{1}{2\lambda_i} - \frac{(k_i - 1)\beta}{2}\right)\varepsilon_i \tag{5.9}$$

Setting x_{ji} equal to $\beta\varepsilon_i$ yields $\beta = \frac{1}{(k_i+1)\lambda_i}$.

Meanwhile, normality and the zero profit condition of the market-maker make λ_i be the regression coefficient in the forecast of $\varepsilon_i + \gamma$ on the order flow ϖ_i. When informed traders know only security-specific information,

$$\lambda_i = \frac{Cov(\varepsilon_i, \varpi_i)}{Var(\varpi_i)} \tag{5.10}$$

where

$$Cov(\varepsilon_i, \varpi_i) = Cov(\varepsilon_i, k\beta\varepsilon_i + z_i) = k\beta Var(\varepsilon_i) \tag{5.11}$$

$$\begin{aligned} Var(\varpi_i) &= Var(x_{ji} + (k-1)\beta\varepsilon_i + z_i) \\ &= Var(\beta\varepsilon_i + (k-1)\beta\varepsilon_i + z_i) \\ &= k^2\beta^2 Var(\varepsilon_i) + Var(z_i) \end{aligned} \tag{5.12}$$

λ_i results to a quadratic equation in λ_i and the positive root of the solution yields lemma 1. (QEV)

Lemma 1 suggests the equilibrium of individual security i, and the optimal λ_i is decreasing in the number of informed traders. Meanwhile Lemma 2 provides the equilibrium of the basket security.

Lemma 2: Each trader possessing information about security i submits an order x_{im} in the basket security, where

$$x_{im} = \frac{w_i\varepsilon_i}{\lambda_m(k_i + 1)} \tag{5.13}$$

The equilibrium pricing parameter λ_m in the market for the basket security is expressed by

$$\lambda_m = \sqrt{\left(\sum_{i=1}^{N} w_i^2 Var(\varepsilon_i)\frac{k_i}{(k_i+1)^2}\right) \Big/ Var(z_m)} \tag{5.14}$$

(Proof) the proof procedure is similar to that in Lemma 1.

In the view of a discretionary liquidity trader q, he/she knows his/her demand for security i which is $w_i l_q$. Thus, if the discretionary liquidity trader trades with informed traders, then the maximum loss of the discretionary liquidity trader is either $\sum_{i=1}^{N} w_i^2 \lambda_i (l_q)^2$ or $\lambda_m (l_q)^2$. The former maximum loss is realised when the discretionary liquidity traders trade in the individual securities: the latter maximum loss when trade in the basket security. Since all pricing parameters decreases monotonically in the total variance s of liquidity trading in the respective markets, the losses of individual discretionary liquidity traders are lowest when they all trade in either the individual securities or the basket security.

Proposition 1 shows that the losses of individual discretionary liquidity traders are minimised when they all trade in the basket security. This happened under the condition that λ_m with all the discretionary liquidity trades concentrated in the basket is less than $\sum_{i=1}^{N} w_i^2 \lambda_i$ with all such trades concentrated in the securities.

Proposition 1: A necessary and sufficient condition for the losses of each discretionary liquidity trader to be minimised by concentration of trade in the basket is that

$$\sum_{i=1}^{N} w_i \sqrt{\frac{var(\varepsilon_i)/K}{var(l)+var(r)}} > \sqrt{\frac{\sum_{i=1}^{N} w_i^2 \left[var(\varepsilon_i)/K_i\right]}{var(l)+var(m)}} \tag{5.15}$$

Under this condition, there always exists a Nash equilibrium in which discretionary liquidity traders all trade in the basket.

(Proof)

Based on lemmas 1 and 2, when all discretionary liquidity traders trade in the securities, the pricing parameter in security i λ_i^* is computed by

$$\lambda_i^* = \frac{\lambda_i}{w_i} = \sqrt{\frac{Var(\varepsilon_i)}{w_i^2[Var(l)+Var(r)]K_i}} \tag{5.16}$$

where K_i represents $\frac{(k_i+1)^2}{k_i}$.

When all liquidity traders purchase/sell in the basket security, however, the pricing parameter in the basket is expressed by

$$\lambda_m^* = \sqrt{\sum_{i=1}^{N} \frac{Var(w_i\varepsilon_i)}{K_i} \Big/ (Var(l)+Var(m))} = \lambda_m \tag{5.17}$$

if and only if the condition of proposition 1 holds, then the condition $\lambda_m^* = \lambda_i < \sum_{i=1}^{N} w_i\lambda_i = \sum_{i=1}^{N} w_i^2\lambda_i^*$ holds.

For the detailed proof of proposition 1, see Subrahmanyam (1991). (QEV)

The proposition 1 suggests that discretionary liquidity traders are concentrated on the basket security when $\sum\limits_{i=1}^{N} w_i\sqrt{\frac{var(\varepsilon_i)/K}{var(l)+var(r)}}$ is greater than $\sqrt{\frac{\sum_{i=1}^{N} w_i^2[var(\varepsilon_i)/K_i]}{var(l)+var(m)}}$. In other words, when the variability of private information about the basket security $\sqrt{\sum_{i=1}^{N} w_i^2\,[var(\varepsilon_i)]}$ is less than the weighted sum of the variability of private information about individual securities $\sum\limits_{i=1}^{N} w_i\sqrt{var(\varepsilon_i)}$, discretionary liquidity traders are concentrated on trading the basket security to minimize their trading loss. Thus, the diversification of private information in the basket security leads to the basket security having lower adverse selection costs.

5.2.1.2 Systematic-factor Informed Traders:

Let us consider the case of systematic-factor informed traders. Systematic-factor informed traders are aware of the systematic component of security value. i.e. security-i-specific informed traders observe $\varepsilon_i + \mu_i$ and systematic-factor informed traders observe $\gamma + \nu$. Meanwhile $var(\mu_i) = \theta$ and $var(\nu) = \kappa$ and *i=1, ...,*

N. Additionally, $\mu_i, \ldots, \mu_N$ and ν are mutually independent of each other and independent of $\varepsilon_i, \ldots, \varepsilon_N$ and γ.

Then Lemma 3 provides the equilibrium for a given level of liquidity trading in each market. (For the brevity, I do not show the proof of Lemma 3 and proposition 2. However, the way to prove Lemma 3 and proposition 2 is similar to Lemma 1 and 2 and proposition 1.)

Lemma 3 With systematic-factor information trading, the equilibrium pricing parameters in the basket security and in security i for given liquidity trade variances $var(z_m)$ and $var(z_i)$ in these markets are, respectively, given by

$$\lambda_m = \sqrt{\left[\left(\sum_{i=1}^{N} \frac{w_i^2 var^2(\varepsilon_i)}{K_i[var(\varepsilon_i)+\theta_i]}\right) + \frac{(\sum_{i=1}^{N} w_i\beta_i)^2 var^2(\gamma)}{G[var(\gamma)+\kappa]}\right] \Big/ var(z_m)} \tag{5.18}$$

, and

$$\lambda_i = \sqrt{\left(\frac{var^2(\varepsilon_i)}{K_i[var(\varepsilon_i)+\theta_i]}\right) + \frac{\beta_i^2 var^2(\gamma)}{G[var(\gamma)+\kappa]} \Big/ var(z_i)} \tag{5.19}$$

K_i is same as in the proposition 1 while $G \equiv (g+1)^2/g$.

The lemma 3 considers that the dealer faces the additional adverse selection component due to the systematic-factor informed traders.

Proposition 2 With systematic-factor information trading, a necessary and sufficient condition for the transaction costs of the discretionary liquidity traders to be minimised by concentration of trade in the basket is that

$$\sum_{i=1}^{N} w_i \sqrt{\left(\frac{T_{ei}}{K_i} + \frac{\beta_i^2 T_g}{G}\right) \Big/ [var(l) + var(r)]}$$

$$>$$

$$\sqrt{\left[\sum_{i=1}^{N} \frac{w_i^2 T_{ei}}{K_i} + \left(\sum_{i=1}^{N} w_i\beta_i\right)^2 \frac{T_g}{G}\right] \Big/ [var(l) + var(m)]} \tag{5.20}$$

where $T_{ei} \equiv \frac{var^2(\varepsilon_i)}{var(\varepsilon_i)+\theta_i}$ and $T_g \equiv \frac{var^2(\gamma)}{var(\gamma)+\kappa}$. Under this condition, there always exists a Nash equilibrium in which discretionary liquidity traders all trade in the basket.

Proposition 2 is explained in two ways: when the systematic-factor sensitivities (β_i) are the same sign and systematic-factor informed traders trade based on systematic-factor information, securities are perfectly correlated to each other and then systematic-factor trading is symmetric in the basket security and in the individual securities. When the systematic-factor sensitivities (β_i) differ in sign across securities, the diversification effect of systematic-factor information arises in the basket security. When proposition two holds, the diversification effect is strengthened. This is true unless $var(m) < var(r)$. $var(m)$ is the variance of the total nondiscretionary liquidity trade in the basket security and $var(r)$ is the variance of the total demands of the nondiscretionary traders in security i.

When proposition two holds, the discretionary liquidity traders trade in the basket security and then λ_m is less than $\sum_{i=1}^{N} w_i^2 \lambda_i$. In this case, the uninformed trading of the basket security has less loss to informed traders than the aggregate loss of corresponding uninformed trades in the underlying securities.

5.2.2 Gorton and Pennacchi (1993)

Gorton and Pennacchi (1993) extend the work of Subrahmanyam (1991) by changing the risk preference of liquidity traders. They show that the presence of the basket security allows liquidity traders to shun the expected transaction losses incurred when they transact with security-specific informed traders. As the basket security reflects less impact of security-specific information than individual securities, liquidity traders concentrate on the basket security to reduce expected losses. This leads to lower adverse selection costs associated with the basket security than those concerned with the individual security.

The main difference between Subrahmanyam (1991) and Gorton and Pennacchi (1993) are the assumptions about the risk-averseness of liquidity traders. Subrahmanyam assumes risk-neutral liquidity traders, while Gorton and Pennacchi assume risk-averse liquidity traders.

5.3 Exchange-Traded Funds

I contend here that Exchange-Traded Funds (ETF) successfully capture the characteristics of a basket security as suggested by Subrahmanyam (1991) and Gorton and Pennacchi (1993). Each ETF share reflects the collective performance of a portfolio whose constituents can be anything from gold to stock. Additionally, the portfolio is usually a particular market index and the ETF aims to achieve the same return as that index.

There are two ways to own an ETF; one is to buy the ETF on the stock exchange, the other is to create the ETF by depositing the constituents of an ETF into a depository institution. Any individual investors can buy (sell) ETFs on the exchange, but those who create (redeem) ETFs are usually institutional investors including market-makers. Gastineau (2001) provides a good introduction to ETFs. Kostovetsky (2003) compares index mutual funds with Exchange-Traded Funds, and a further explanation of ETFs can be found on the websites of the American Stock Exchange (AMEX) and the National Association of Stock Dealers Automated Quotation (NASDAQ).

In this chapter, I deal with spread decomposition models so the focus is on ETFs whose constituents are common stocks.

5.3.1 Characteristics of ETFs

5.3.1.1 Transparency

ETFs have higher levels of transparency regarding their constituent securities, compared to mutual funds[1]. The US Securities and Exchange Commission explains that an ETF's investment objective is to achieve the same return as a particular market index. One way to achieve this goal is to invest in all constituent securities of the market index with the same weight as the market index. Another method is to choose representative securities in the constituents of the market index and to invest in them. Whichever investment method is adopted, ETF

[1] A mutual fund is a company that pools money from many investors and invests the money in stocks, bonds, short-term money-market instruments, other securities or assets, or some combination of these investments

investors can clearly identify which constituent securities compose the market index and how is the weight of each security in the market index.

Though ETFs usually follow the overall market index representing either whole stock market movement or whole bond market movement, there are some occasions when an ETF concentrates on a specific industry. This is possible because some market indices focus on the market movement of a specific industry. For example, the SPDR Consumer Staple (XLP) ETF is composed of only stocks that cover food and drug retailing, beverages, food products, tobacco, house hold products, and personal products. These types of ETFs are called Sector ETFs and focus on one industry, in this case, the consumer staple industry.

Even though an ETF is composed of some stocks that represent either the broad market or a specific industry, investors purchasing the ETF receive the prospectus of the ETF or the product description. With this information, ETF investors can easily know which stock comprises the ETF they invest in its weight within the ETF. With this information the ETF investors can compute the Net Asset Value[1] (NAV) of the ETF and its intraday/daily return. Furthermore, in the case where an ETF investor knows which market index is tracked by the ETF, he or she can compute the intraday/daily return of the ETF based on the return of the market index.

This higher transparency of an ETF leads to the indifference between the intrinsic value (NAV) of the ETF and its trading price. The higher transparency causes the information asymmetry between informed traders and uninformed traders[2] to be smallest because the movement of the tracked market index reveals the movement of intrinsic value which is based on the private information of informed traders.

5.3.1.2 Flexibility of Transaction

ETF shares are traded continuously on exchanges during the trading hours. This intraday trading characteristic is similar to the Closed-end index funds[3].

[1] The Net Asset Value of an ETF = the total assets of the ETF – the total liabilities of the ETF. Per Share NAV = the Net Asset Value / the number of shares outstanding

[2] Market makers and liquidity traders

[3] A Closed-end Fund is one of three types of Investment Company. The two other types of Investment Company are Mutual Funds (Open-end) and Unit Investment Trusts. Closed-end

Due to the ability to undertake intra-day transactions, the ETF can reflect all information throughout the day when its corresponding index reflects any new information due to its constituent security.

ETFs can be sold short. As suggested by Diamond and Verrecchia (1987), the short selling might encourage the transmission of information to prices. While ETF shares are allowed to sell short, common stocks and closed-end funds can not be sold short. Thus, ETFs have a relative advantage in terms of incorporating information to prices.

Therefore, whatever the trading reason is, traders are able to buy/sell ETFs when they need. i.e. when informed traders have new private information, they can trade ETFs. Meanwhile when liquidity traders need liquidity they can sell ETFs, or when they need to invest in ETFs they can buy them.

5.3.1.3 Creation and Redemption

The creation and redemption processes associated with ETF are important features in that they contribute towards diminishing the discount[1] or premium of the ETF through arbitrage trading. The creation of ETF shares means that authorised participants[2] including market-makers and institutional investors can issue ETFs shares by depositing the appropriate number of shares of the constituent stocks. The redemption of ETF shares means that authorised participants redeem the proper number of constituent stocks by depositing the ETF shares.

Whenever the ETF trading price is greater than its intrinsic value (NAV) (i.e. ETFs have premium), authorised participants make arbitrage profit by the creation of ETFs. On the contrary, whenever the trading price of ETFs

funds are traded on the Exchange and the price is determined by the market demand and supply. Index funds have an objective that is to achieve the same return as a particular market index.

[1]Discount (premium) occurs when the Net Asset Value (NAV) of the fund is different from the transaction price of the fund. The discount puzzle of closed-end funds is originated from the phenomenon that closed-end funds have sold for less than the market value of their holdings. And discount (premium) continues to exist under the possibility of arbitrage profit trading.

[2]Authorised participants are institutional investors who are permitted to purchase Creation Units directly from, or to redeem Creation Units directly with, the fund. To be an Authorised Participant, an entity must be a participant in the Depository Trust Company and must enter into an agreement with the fund's Distributor.

is less than the intrinsic value (NAV), (i.e. ETFs are discounted) authorised participants make an arbitrage profit by the redemption of ETFs. Thus, the creation and redemption process of ETFs, with their intra-day transaction, mitigates the magnitude of the discount or premium. With intra-day and short-selling transacting, ETFs have insignificant premium/discount. This prediction is confirmed by Elton, Gruber, Comer, and Li (2002) when they use SPDR[1]. In case of SPDR, the premium or discount disappears in a day.

5.3.1.4 Basket Security

By definition, an ETF share is a security representing a bundle of its constituent securities. The Net Asset Value (NAV) of ETF shares is determined by the weighted average of constituent security values. Since the prospectus of an ETF addresses the composition of the ETF and the prospectus is publicly accessible, any investors can compute the NAV of the ETF in principle when they know the value of the component assets. However, even though an ETF investor does not know the constituent securities of the ETF, he or she can still compute the return of the ETF because the ETF usually track one market index and the return of the market index is publicly announced.

It is possible that the introduction of a basket security squeezes the liquidity of constituent stocks, as is discussed in Subrahmanyam (1991). This is reasonable when limited liquidity traders exist. However, in practice, the introduction of a basket security generates more liquidity traders. The empirical findings of Boehmer and Boehmer (2003) and Tse and Erenburg (2003) suggest that when ETFs trading on the American Stock Exchange (AMEX) are newly introduced on the New York Stock Exchange (NYSE), the market liquidity rises rather than falls.

5.3.1.5 Summary

The high levels of transparency associated with ETFs, flexibility of transactions, and the diversification of security-specific information allow ETFs to be used to evaluate spread decomposition models. An ability to transact on an intraday

[1]SPDR is an ETF that follows the S&P 500 index.

basis along with the ability to short sell and the ability to create (redeem) ETFs allows for smaller discount or premium[1]. The underlying security-specific information is expected to be diversified away in the ETFs. All these features suggest that any liquidity traders transacting ETFs do not need to worry about the trading loss arising from the constituent security-specific informed traders. Those liquidity traders only care about industry or gross economic information, which is considered as public.

5.4 Empirical Design

5.4.1 Empirical Hypothesis

Combining the theoretical results of Subrahmanyam (1991) and Gorton and Pennacchi (1993) with the characteristics of ETFs provides the empirical hypothesis to evaluate the performance of spread decomposition models.

Hypothesis: ETFs have lower adverse selection costs than their matched control stocks. Furthermore I distinguish broad-market ETFs and industry-wide ETFs. Broad-market ETFs follow the movement of the overall stocks market and industry-wide ETFs track the movement of stocks that belong to a specific industry. If a spread decomposition model correctly measures the adverse selection cost, then broad-market ETFs and industry-wide ETFs also show significant difference from control securities. The relationship is expressed by

$$\begin{bmatrix} \text{Adverse Selection} \\ \text{cost of} \\ \text{ETFs} \end{bmatrix} \leq \begin{bmatrix} \text{Adverse Selection} \\ \text{cost of} \\ \text{matched Control Stocks} \end{bmatrix}$$

5.4.2 Diversification of ETFs

One important assumption for the empirical investigation in section 5.4.1 is that ETFs are basket securities. It is possible that an ETF behave like equity

[1]Small discount or premium means that the net asset value of an ETF is the same as the trading price. It is hard for informed traders to gain from arbitrage trading.

since the ETF is strongly related to one constituent stock. In other words, one constituent stock is heavily weighted in the ETF[1] and the private information of that constituent stock causes a move in the ETF. In this case the ETF is not fully diversified and is less likely to co-vary with the movement of the overall stock market which is considered as most diversified. Thus, I check the correlation of the stock market and ETFs.

I simply look into how ETF returns are correlated with the returns of the stock market indices. In other words, if a stock market index represents a fully-diversified market portfolio, then the correlation between ETFs and stock indices are close to one under the assumption that they are also fully- diversified basket securities. The S&P 500 stock index is used as a stock market index.

I compute the Pearson correlation coefficients between the returns of ETFs and the return of the S&P 500 stock index. The average Pearson correlation coefficient of broad-market ETFs is 0.8975 and of industry-wide ETFs is 0.7096. The higher Pearson correlation coefficient of broad-market ETFs implies that broad-market ETFs are more diversified than industry-wide ETFs. When I consider that some ETFs do not track, but highly correlate with, the S&P 500 stock index, I can contend that the constituent stock in my ETF sample does not strongly affect the movement of the ETF. For detailed statistics, see Appendix 5.A.

5.5 Data

US stock market data are employed[2]. The New York Stock Exchange (NYSE) provides the Trade and Quote (TAQ) consolidated trade database, which contains intra-day trades and quotes for all NYSE-listed and non-listed securities. TAQ data set includes information about quotes and trades such as stock symbol, bid (ask) size, bid (ask) price, quoted and transaction times, quoted and transacted exchanges, volume of trade, sale condition and the correction of recorded trade.

From the TAQ data set, I exclude the following files:

[1]This case is the third Nash equilibrium of Subrahmanyam (1991). i.e. the variability of the basket security is heavily dependent on the variability of one constituent stock.

[2]The same data are employed in Chapter Six and Seven.

(1) quotes if either the ask or bid price is less than or equal to zero;

(2) quotes if either the ask size or the size is less than or equal to zero;

(3) quotes if the bid-ask spread is greater than \$5 or less than zero;

(4) trades before the open or after the close of the stock market ;

(5) trades if the price or volume is less than or equal to zero;

(6) trade price, p_t, if $|(p_t - p_{t-1})/p_{t-1}| > 0.5$;

(7) ask quote, a_t, if $|(a_t - a_{t-1})/a_{t-1}| > 0.5$;

(8) bid quote, b_t, if $|(b_t - b_{t-1})/b_{t-1}| > 0.5$.

These screening procedures are conducted since these files represent typing errors. Huang and Stoll (1996) and Chung, Chuwonganant, and McCormick (2004) used a similar exclusion procedure.

5.5.1 Selection of Sample ETFs

I consider only ETFs whose constituents are US common stocks. While over two hundred ETFs were trading on the US stock markets in December 2006, some of these have as their constituents the securities of foreign countries. For instance, iShares MSCI – United Kingdom ETF is composed of UK stocks, but is traded on the US stock exchange. This example suggests that US investors cannot trade fully the constituent stocks when they have security-specific information about constituent stocks of international ETFs. Due to different trading hours and different trading protocol, US investors may have difficulty in trading foreign stocks. This trading difficulty leads to the impossibility that the private information of foreign constituent stock is immediately reflected on the international ETF. Therefore, this study concentrates on the ETFs that contain US stocks as constituents. I also exclude, for similar reasons, ETFs whose constituents are bonds or commodities like Gold.

After I collect US stock based ETFs, I distinguish broad-market ETFs from industry-wide ETFs. The discerning criterion follows the classification used by the NASDAQ and AMEX. Both exchange websites provide information about which ETFs are a broad-market or single-industry. The lists of broad-market ETFs are shown in Table 5.1 and industry-wide ETFs in Table 5.2. The sample period covers October 2005 to December 2005. The three months sample period is employed to estimate the components of spread decomposition models. The last columns of the Table 5.1 and Table 5.2 show the distance measure to choose matched control stock that will be explained in the next section.

5.5.2 Selection of Control Stocks

5.5.2.1 The Justification of Matching Method

If I compare the weighted average adverse selection cost of constituent stocks with the adverse selection cost of the ETF, this approach will directly test the empirical hypothesis that ETFs have lower adverse selection cost than control stocks. However, this approach is empirically problematic, because not all ETF prospectuses reveal all the constituent stocks of the ETFs. In some ETFs like SPY it is clearly known to what the constituent stocks are, because SPY tracks the S&P 500 stock index. However, other ETFs like IWV does not show all their constituent stocks since IWV tracks the Russell 3000 index that comprises 3000 stocks. Note that even though an ETF prospectus does not show all the constituent stocks, this does not cause a problem with the transparency of ETFs, since all ETFs clearly mention what stock market index they track in their prospectus and the return of an ETF is perfectly correlated with the return of the stock market index associated with the ETF.

Even if ETFs have a higher transparency and all the constituent stocks, a computation problem may exist. For example, when I compute the adverse selection cost of the constituent stocks of IWV, I need to compute the adverse selection cost of 3000 stocks.

These empirical problems prevent a direct comparison of ETFs and their constituent stocks.

One solution is to match each ETF to a control security. Clarke and Shastri (2001) employ a control matching method with three variables[1]. They compare NYSE ordinary securities and closed-end funds. Neal and Wheatley (1998) employ trading volume to match closed-end funds and ordinary securities. They examine two spread decomposition models such as Glosten and Harris (1988) and George, Kaul, and Nimalenddran (1991).

[1]Price, trading volume, and market capitalisation are used in their matching process.

Table 5.1: Broad market ETFs and Matched Control Stocks

Inception is the date an ETF became active for the first time. Product name is the name by which an ETF is known in the marketplace. Control is the abbreviated name of the control stock used thereafter. Company name is the name of the control stock. Volume represents trading volume. Dist is a computed measure of a chosen control stock that minimizes $\left\{\frac{(X_E - X_C)}{(X_E + X_C)/2}\right\}^2$, where X r epresents *Vol*, *P* , and σ. Subscript C and E represent Control securities and ETFs, respectively. Vol_C (Vol_E) is the average daily volume of a control stock (ETF), P_C (P_E) is the average daily price of the control stock (ETF), σ_C (σ_E) is the standard deviation of daily return of control stock (ETF).

Symbol	Inception	Product Name	Tracked Index	Volume	Control	Company Name	Volume	Dist
DIA	20-Jan-98	DIAMONDS	Dow Jones Industrial Average	6956065	MO	ALTRIA GROUP INC	6943895	0.40600
DSG	25-Sep-00	streetTRACKS DJ Wilshire Small Cap Growth ETF	Dow Jones Wilshire Small Cap Growth Index	2025	UTL	UNITIL CORP	2306	0.94454
DSV	25-Sep-00	streetTRACKS DJ Wilshire Small Cap Value ETF	Dow Jones Wilshire Small Cap Value Index	3048	Y	ALLEGHANY CORP DE	5215	0.82484
ELV	25-Sep-00	streetTRACKS DJ Wilshire Large Cap Value ETF	Dow Jones Wilshire Large Cap Value Index	7009	STU	STUDENT LOAN CORP	9403	0.45560
IJH	22-May-00	iShares S&P MidCap 400 Index Fund	S&P MidCap 400 Index	169623	HB	HILLENBRAND INDS INC	171178	0.13214
IJJ	24-Jul-00	iShares S&P MidCap 400 Value Index Fund	S&P MidCap 400/BARRA Value Index	201975	CYN	CITY NATIONAL CORP	201515	0.06283
IJK	24-Jul-00	iShares S&P MidCap 400 Growth Index Fund	S&P MidCap 400/BARRA Growth Index	97709	CBSH	COMMERCE BANCSHARES INC	110748	0.14406
IJR	22-May-00	iShares S&P SmallCap 600 Index Fund	S&P SmallCap 600 Index	1289262	NOC	NORTHROP GRUMMAN CORP	1371482	0.02893
IJS	24-Jul-00	iShares S&P SmallCap 600 Value Index Fund	S&P SmallCap 600/BARRA Value Index	176246	MRBK	MERCANTILE BANKSHARES CORP	172236	0.03032
IJT	24-Jul-00	iShares S&P SmallCap 600 Growth Index Fund	S&P SmallCap 600/BARRA Growth Index	94821	ESE	ESCO TECHNOLOGIES INC	77898	0.40958
IVE	22-May-00	iShares S&P 500 Value Index Fund	S&P 500/BARRA Value Index	247568	MCY	MERCURY GENERAL CORP NEW	199573	0.29923
IVV	15-May-00	iShares S&P 500 Index	S&P 500 Index	768209	CB	CHUBB CORP	813104	0.24485
IVW	22-May-00	iShares S&P 500 Growth Index Fund	S&P 500/BARRA Growth Index	239596	UB	UNIONBANCAL CORP	277196	0.13872
IWB	15-May-00	iShares Russell 1000	Russell 1000 Index	477685	TMK	TORCHMARK CORP	501623	0.09213
IWD	22-May-00	iShares Russell 1000 Value	Russell 1000 Value Index	675448	BCR	BARD C R INC	690060	0.23370
IWF	22-May-00	iShares Russell 1000 Growth	Russell 1000 Growth Index	734865	LH	LABORATORY CORP AMERICA HLDGS	687262	0.16967
IWM	22-May-00	iShares Russell 2000	Russell 2000 Index	23289239	GE	GENERAL ELECTRIC CO	20435200	0.45624
IWN	24-Jul-00	iShares Russell 2000 Value	Russell 2000 Value Index	864953	CMA	COMERICA INC	879142	0.03500
IWO	24-Jul-00	iShares Russell 2000 Growth	Russell 2000 Growth Index	893392	ABC	AMERISOURCEBERGEN CORP	896554	0.01034
IWP	17-Jul-01	iShares Russell Midcap Growth Index Fund	Russell MidCap Growth Index	132775	IDXX	IDEXX LABORATORIES INC	140180	0.20170
IWR	17-Jul-01	iShares Russell Midcap Index Fund	Russell MidCap Index	113631	WPS	WPS RESOURCES CORP HOLDING CO	117257	0.22242
IWV	22-May-00	iShares Russell 3000	Russell 3000 Index	115660	EEP	ENBRIDGE ENERGY PARTNERS LP	71156	0.33076
JKH	02-Jul-04	iShares Morningstar Mid Growth Index	Morningstar Mid Growth Index	10479	COKE	COCA COLA BOTTLING CO CONS	8920	0.76761
JKJ	02-Jul-04	iShares Morningstar Small Core Index	Morningstar Small Core Index	8153	SFSW	STATE FINANCIAL SERVICES CORP	8661	0.37526
JKK	07-Jul-04	iShares Morningstar Small Growth Index	Morningstar Small Growth Index	4482	BF	BROWN FORMAN CORP	4604	0.04657
JKL	06-Jul-04	iShares Morningstar Small Value Index	Morningstar Small Value Index	7284	BARI	BANCORP RHODE ISLAND INC	7146	0.41629
MDY	04-May-95	MidCap SPDRs	Standard & Poor's MidCap 400 Index	2050306	LEH	LEHMAN BROTHERS HOLDINGS INC	1871851	0.17939
PEY	09-Dec-04	POWERSHARES E T F TRUST	Dividend Achievers 50 Index	161148	BFIN	BANKFINACIAL CORP	161813	0.02420
PWC	01-May-03	PowerShares Dynamic Market Portfolio	Dynamic Market Intellidex Index	76239	KMR	KINDER MORGAN MANAGEMENT LLC	74134	0.08513
PWO	01-May-03	PowerShares Dynamic OTC Portfolio	Dynamic OTC Intellidex Index	30390	CFFN	CAPITOL FEDERAL FINANCIAL	33019	0.17123
QQQQ	10-Mar-99	Nasdaq-100 Index Tracking Stock	Nasdaq-100 Index	80030105	INTC	INTEL CORP	49778952	0.75761
VB	26-Jan-04	Vanguard Small-Cap VIPERs	MSCI U.S. Small Cap 1750 Index	19803	TRH	TRANSATLANTIC HOLDINGS INC	21171	0.01875
VBK	26-Jan-04	Vanguard Small-Cap Growth VIPERs	MSCI U.S. Small Cap Growth 1750 Index	14712	YANB	YARDVILLE NATIONAL BANCORP	11115	0.38712
VBR	26-Jan-04	Vanguard Small-Cap Value VIPERs	MSCI U.S. Small Cap Value 1750 Index	18067	WSFS	WSFS FINANCIAL CORP	21606	0.17911
VO	26-Jan-04	Vanguard Mid-Cap VIPERs	MSCI U.S. Mid Cap 450 Index	30070	GBL	GAMCO INVESTORS INC	33953	0.49701
VTI	31-May-01	Vanguard Total Stock Market VIPERs	MSCI U.S. Broad Market Index	130410	PDX	PEDIATRIX MEDICAL GROUP	161601	0.59788
VTV	26-Jan-04	Vanguard Value VIPERs	MSCI U.S. Prime Market Value Index	25357	ERIE	ERIE INDEMNITY CO	25329	0.04370
VV	27-Jan-04	Vanguard Large-Cap VIPERs	MSCI U.S. Prime Market 750 Index	40312	IBOC	INTERNATIONAL BANCSHARES CORP	39429	0.40106
VXF	27-Dec-01	Vanguard Extended Market VIPERs	Wilshire 4500 Completion Index	16689	BANF	BANCFIRST CORP	17106	0.48578

Table 5.2: Industry-wide ETFs and Matched Control Stocks

Inception is the date an ETF became active for the first time. Product name is the name by which an ETF is known in the marketplace. Control is the abbreviated name of the control stock used thereafter. Company name is the name of the control stock. Volume represents trading volume. Dist is a computed measure of a chosen control stock that minimizes $\left\{\frac{(X_E-X_C)}{(X_E+X_C)/2}\right\}^2$, where X represents Vol, P , and σ. Subscript C and E represent Control securities and ETFs, respectively. Vol_C (Vol_E) is the average daily volume of a control stock (ETF), P_C (P_E) is the average daily price of the control stock (ETF), σ_C (σ_E) is the standard deviation of daily return of control stock (ETF).

Symbol	Inception	Product Name	Sector/ tracked industry	Volume	Control	Company Name	Volume	Dist
BDH	04-Apr-00	BROADBAND HOLDRS TRUST	BroadBand	173621	ACTL	ACTEL CORP	176076	0.33099
BHH	24-Feb-00	B 2 B INTERNET HOLDRS TRUST	B2B Internet	44428	QVDX	QUOVADX INC	55826	0.13258
HHH	24-Sep-99	INTERNET HOLDRS TRUST	Internet	407012	CERN	CERNER CORP	390190	0.11846
IAH	25-Feb-00	INTERNET ARCHITECTURE HOLDRS TR	Internet architecture	81576	PQE	PROQUEST CO	109364	0.14775
IBB	05-Feb-01	iShares Nasdaq Biotechnology	Health	977192	IVGN	INVITROGEN CORP	914890	0.01781
IDU	12-Jun-00	iShares Dow Jones U.S. Utilities Sector Index Fund	Utilities	69692	UIL	U I L HOLDINGS CORP	52253	0.60363
IGE	22-Oct-01	iShares Goldman Sachs Natural Resources	Natural Res	79937	PVA	PENN VIRGINIA CORP	72575	0.51036
IGM	13-Mar-01	iShares Goldman Sachs Technology	Technology	47970	LIFE	LIFELINE SYSTEMS INC	41324	0.56037
IGN	10-Jul-01	iShares Goldman Sachs Networking	Technology	102742	MRCY	MERCURY COMPUTER SYSTEMS	117866	0.32216
IGV	10-Jul-01	iShares Goldman Sachs Software	Technology	174032	IMN	IMATION CORP	212600	0.13439
IGW	10-Jul-01	iShares Goldman Sachs Semiconductor	Technology	246871	ATK	ALLIANT TECHSYSTEMS INC	261171	0.07306
IIH	25-Feb-00	INTERNET INFRASTRUCTURE HLDRS TR	Internet Infrastructure	59964	AMSWA	AMERICAN SOFTWARE INC	54876	0.21440
IYC	12-Jun-00	iShares Dow Jones U.S. Consumer Services Sector Index Fund	the Dow Jones U.S. Consumer Services Index	43092	BL	BLAIR CORP	47593	0.67407
IYE	12-Jun-00	iShares Dow Jones U.S. Energy Sector Index Fund	Natural Res	118121	MIDD	MIDDLEBY CORP	90222	0.46611
IYG	12-Jun-00	iShares Dow Jones U.S. Financial Services Index Fund	Financial	29085	BOKF	B O K FINANCIAL CORP	38275	0.82270
IYH	12-Jun-00	iShares Dow Jones U.S. Healthcare Sector Index Fund	Health	151492	UTR	UNITRIN INC	168439	0.33408
IYJ	12-Jun-00	iShares Dow Jones U.S. Industrial Sector Index Fund	the Dow Jones U.S. Industrials Index	23892	VTRU	VERTRUE INC	24972	0.87923
IYK	12-Jun-00	iShares Dow Jones U.S. Consumer Goods Sector Index Fund	the Dow Jones U.S. Goods Index	42301	ROG	ROGERS CORP	66490	0.76408
IYM	12-Jun-00	iShares Dow Jones U.S. Basic Materials Sector Index Fund	the Dow Jones U.S. Basic Materials Index	75651	CW	CURTISS WRIGHT CORP	82823	0.44893
IYT	06-Oct-03	iShares Dow Jones Transportation Average Index Fund	the Dow Jones Transportation Average Index	136176	FLA	FLORIDA EAST COAST IND INC	99534	0.44449
IYZ	22-May-00	iShares Dow Jones U.S. Telecommunications Sector Index Fund	Communications	239857	WFSL	WASHINGTON FEDERAL INC	251932	0.18530
OIH	06-Feb-01	OIL SERVICE HOLDRS TRUST	Oil Services	5491767	IBM	INTERNATIONAL BUSINESS MACHS COR	5905220	0.37137
RKH	26-Jun-00	REGIONAL BANK HOLDRS TRUST	Regional Bank	655056	ZION	ZIONS BANCORP	5927711	0.69530
RTH	02-May-01	RETAIL HOLDRS TRUST	Retail	3672783	TGT	TARGET CORP	3765931	0.41729
SMH	26-May-00	SEMICONDUCTOR HOLDRS TRUST	Semi-conductor	18421435	DELL	DELL INC	16852963	0.03253
UTH	26-Jun-00	UTILITIES HOLDRS TRUST	Utilities	379027	EQT	EQUITABLE RESOURCES INC	350667	0.42407
VAW	26-Jan-04	Vanguard Materials VIPERs	Natural Res	11656	SCL	STEPAN CO	11654	0.82357
VCR	26-Jan-04	Vanguard Consumer Discretionary VIPERs	the MSCI U.S. Investable Market Consumer Discretionary Index	7929	FSCI	FISHER COMMUNICATIONS INC	9067	0.81483
VDE	29-Sep-04	Vanguard Energy VIPERs	Natural Res	39581	TNC	TENNANT CO	36556	0.37205
VFH	26-Jan-04	Vanguard Financials VIPERs	Financial	7275	COBH	PENNSYLVANIA COMMERCE BANCORP IN	5380	0.73764
VPU	26-Jan-04	Vanguard Utilities VIPERs	Utilities	17187	CHG	C H ENERGY GROUP INC	32182	0.54525
WMH	01-Nov-00	WIRELESS HOLDRS TRUST	Wireless	9893	FSCI	FISHER COMMUNICATIONS INC	9067	0.92708
XLB	22-Dec-98	Select Sector SPDR	Materials	2496942	MAS	MASCO CORP	22654060	0.14562
XLE	22-Dec-98	SECTOR SPDR TRUST	Energy	14027384	XOM	EXXON MOBIL CORP	18119117	0.11357
XLK	22-Dec-98	Select Sector SPDR	Technology	1050464	CVC	CABLEVISION SYSTEMS CORP	12408000	0.18775
XLU	22-Dec-98	Select Sector SPDR	Utilities	1934689	SO	SOUTHERN CO	19062030	0.00633
XLV	22-Dec-98	Select Sector SPDR	Health care	831734	PFG	PRINCIPAL FINANCIAL GROUP INC	9298460	0.25221
XLY	22-Dec-98	Select Sector SPDR	Consumer Discretionary	1036512	CTL	CENTURYTEL INC	11189010	0.01587

5.5.2.2 How to Construct Matched Control Stocks

As a second best procedure, I use control stocks that are comparable to ETFs. To control the inventory holding and order processing costs, I choose three factors suggested by Madhavan (2000). Madhavan provides that the variability in the bid-ask spread is explained by volume, risk, and price.

A chosen control stock minimises the following summation value:

$$\left(\frac{Vol_C - Vol_E}{(Vol_C + Vol_E)/2}\right)^2 + \left(\frac{P_C - P_E}{(P_C + P_E)/2}\right)^2 + \left(\frac{\sigma_C - \sigma_E}{(\sigma_C + \sigma_E)/2}\right)^2 \qquad (5.21)$$

Vol_C (Vol_E) is the average daily volume of a control stock (ETF), P_C (P_E) is the average daily price of the control stock (ETF), σ_C (σ_E) is the daily return volatility of a control stock (ETF).

I used a three-month test period for the selection of control securities, which is from 1^{st} July to 30^{th} September 2005. From this test period, control securities are selected to have less than one summation value. Additionally, trading volume is considered because higher trading volume of ETFs and lower volume of control securities can mislead the analysis. I confirmed with the paired-t test that ETFs do not differ from matched control securities as a result trading volume.

Furthermore, industry classification also is considered when control securities for industry-wide ETFs are selected. This is attained by choosing control securities that belong to the same industry as component companies of industry-wide ETFs. For industry-wide ETFs the control stock must also be drawn from the same industry as the ETF. For this we use the two digits SIC code.

Meanwhile three factors can be regarded as controlling for the order-processing cost and the inventory holding cost. The increase in volume helps to diminish the inventory holding risk of the market-maker since the market-maker easily adjusts his or her inventory stocks. The volatility also relates with the inventory holding risk when prices move in the adverse direction to the expectation of the market-maker. Prices control for the order processing cost. Higher prices have less order processing cost.

This matching method is extensively adopted in the literature. Stoll (2000) checks the variability of cross-sectional spread and finds that four factors (price, standard deviation of return, trading volume, and market capitalisation) explain the cross-sectional relations of spread well. Hatch and Johnson (2002) also use the similar matching method when they investigate the effect of specialist firm acquisitions on market quality. Likewise, Brockman and Chung (2001) also employ similar matching method.

5.6 Empirical Results

5.6.1 Sample Statistics

Table 5.3 and 5.4 illustrate sample statistics including average price, average trade size, average quoted spread, average effective spread, and so on. Table 5.3 gives the sample statistics of broad-market ETFs and matched control stocks. Table 5.4 reports the sample statistics of industry-wide ETFs and matched control stocks.

The average dollar-quoted spread of broad-market ETFs is 0.13 dollars while that of matched control stocks is 0.41. The average dollar-quoted spreads of industry-wide ETFs and matched control stock also have similar values, 0.14 and 0.27 respectively. Thus the dollar-quoted spreads of ETFs (Broad-market and Industry-wide) appear to be lower than those of matched control stocks.

The dollar-effective spread measure also shows a similar pattern to the dollar-quoted spread. The average dollar-effective spread of broad-market (industry-wide) ETFs is 0.04 and the average dollar-effective spread of their matched control stocks is 0.07.

The low quoted and effective spreads are consistent with recent studies by Elton, Gruber, Comer, and Li (2002) and Boehmer and Boehmer (2003) who acknowledge the high levels of liquidity-associated with ETFs.

5.6.2 Estimation Results

The hypothesis test results depending on the spread decomposition model are shown in the order of Glosten and Harris (1988), Madhavan, Richardson, and

Table 5.3: Sample Statistics of Broad-market ETFs and Matched Control Stocks

This table presents the mean spread values for broad market ETFs and control stocks. Trade price is the mean sample traded price, Trade size is the mean sample trade size .The quoted s is the difference between the quoted ask and the quoted bid price. The effective s is the difference between the midpoint price and the subsequent transaction price. Standard deviation is the standard deviation of the spread estimates. Percent refers to percentage values of the spread.

ETF	Trade price	Trade size	Quoted sprd	Percent Quoted sprd	Effective sprd	Percent Effective sprd	Control	Trade Price	Trade size	Quoted sprd	Percent Quoted sprd	Effective sprd	Percent Effective sprd
DIA	105.78	1560	0.03	0.03	0.01	0.01	MO	74.07	740	0.17	0.23	0.05	0.06
DSG	79.24	377	0.32	0.40	0.09	0.11	UTL	25.61	266	0.45	1.76	0.15	0.60
DSV	61.38	575	0.25	0.41	0.07	0.12	Y	297.21	200	2.20	0.74	0.73	0.25
ELV	68.97	504	0.22	0.32	0.06	0.09	STU	221.97	189	1.99	0.89	0.70	0.31
IJH	71.58	652	0.09	0.13	0.03	0.04	HB	47.07	391	0.29	0.63	0.06	0.13
IJJ	68.41	756	0.11	0.16	0.04	0.05	CYN	70.80	329	0.28	0.39	0.08	0.11
IJK	73.36	1020	0.12	0.17	0.05	0.06	CBSH	52.66	218	0.10	0.19	0.04	0.07
IJR	56.80	1354	0.11	0.19	0.03	0.05	NOC	55.41	561	0.27	0.49	0.05	0.09
IJS	62.91	634	0.12	0.18	0.04	0.06	MRBK	56.83	217	0.09	0.15	0.03	0.06
IJT	113.67	695	0.21	0.18	0.07	0.07	ESE	43.04	327	0.34	0.79	0.09	0.21
IVE	63.61	659	0.08	0.13	0.03	0.04	MCY	58.85	256	0.45	0.77	0.09	0.16
IVV	122.56	1261	0.08	0.07	0.02	0.02	CB	92.66	322	0.27	0.29	0.05	0.05
IVW	58.45	663	0.10	0.16	0.03	0.06	UB	67.98	436	0.17	0.25	0.05	0.07
IWB	66.84	935	0.10	0.15	0.03	0.04	TMK	53.30	525	0.30	0.57	0.07	0.13
IWD	68.34	1280	0.11	0.16	0.03	0.04	BCR	65.68	415	0.77	1.17	0.12	0.19
IWF	50.61	1611	0.09	0.18	0.02	0.05	LH	50.36	504	0.20	0.40	0.04	0.07
IWM	65.64	3142	0.02	0.04	0.01	0.02	GE	34.74	982	0.04	0.12	0.01	0.02
IWN	65.37	1490	0.11	0.17	0.03	0.05	CMA	57.55	338	0.30	0.53	0.06	0.10
IWO	68.27	1336	0.10	0.15	0.03	0.04	ABC	76.36	361	0.41	0.54	0.07	0.10
IWP	91.64	976	0.13	0.14	0.04	0.05	IDXX	68.84	215	0.14	0.20	0.05	0.07
IWR	86.18	722	0.12	0.14	0.04	0.04	WPS	54.91	348	0.32	0.58	0.06	0.11
IWV	71.07	1011	0.10	0.14	0.04	0.05	EEP	47.34	355	0.44	0.92	0.07	0.16
JKH	75.91	638	0.12	0.16	0.04	0.06	COKE	44.91	186	0.56	1.26	0.20	0.45
JKJ	71.64	769	0.20	0.28	0.06	0.08	SFSW	37.53	267	0.12	0.32	0.04	0.10
JKK	66.91	658	0.13	0.20	0.04	0.07	BF	73.58	385	0.29	0.40	0.07	0.10
JKL	70.33	813	0.24	0.35	0.05	0.08	BARI	34.99	332	0.45	1.28	0.17	0.48
MDY	130.62	1422	0.07	0.05	0.02	0.01	LEH	120.51	394	0.49	0.41	0.06	0.05
PEY	14.93	927	0.07	0.48	0.03	0.18	BFIN	14.11	332	0.15	0.99	0.06	0.39
PWC	44.98	598	0.15	0.32	0.05	0.12	KMR	47.33	351	0.15	0.31	0.05	0.10
PWO	48.76	550	0.22	0.45	0.08	0.16	CFFN	33.83	220	0.13	0.38	0.05	0.14
QQQQ	40.05	4362	0.01	0.03	0.01	0.02	INTC	24.94	3309	0.02	0.06	0.01	0.03
VB	58.98	699	0.13	0.23	0.05	0.08	TRH	61.93	204	0.93	1.51	0.10	0.17
VBK	57.57	552	0.21	0.35	0.06	0.11	YANB	35.12	196	0.34	0.96	0.11	0.32
VBR	61.05	398	0.17	0.28	0.06	0.09	WSFS	61.52	220	0.36	0.59	0.13	0.21
VO	63.13	572	0.13	0.20	0.05	0.07	GBL	44.84	187	0.26	0.57	0.09	0.19
VTI	121.04	595	0.19	0.16	0.05	0.04	PDX	80.18	285	0.36	0.45	0.10	0.12
VTV	56.26	748	0.11	0.20	0.04	0.07	ERIE	52.65	272	0.15	0.28	0.05	0.09
VV	54.84	710	0.10	0.19	0.04	0.07	IBOC	29.71	297	0.12	0.40	0.04	0.13
VXF	88.98	630	0.29	0.33	0.09	0.10	BANF	81.25	151	1.23	1.61	0.50	0.66
Mean	70.94	996	0.13	0.21	0.04	0.07	Mean	65.44	412	0.41	0.63	0.12	0.18
StdDev	23.04	742	0.07	0.11	0.02	0.04	StdDev	50.77	502	0.46	0.43	0.16	0.15
Median	66.91	722	0.12	0.18	0.04	0.06	Median	54.91	327	0.29	0.53	0.06	0.12

Table 5.4: Sample Statistics of Industry-wide ETFs and Matched Control Stocks

This table presents the mean spread values for the industry-wide ETFs and the control stocks. Trade price is the mean sample trade price, the trade size is the mean sample trade size. The quoted sprd is the difference between the quoted ask and the quoted bid price. The effective sprd is the difference between the midpoint price and the subsequent transaction price. Standard deviation is the standard deviation of the spread estimates. Percent refers to percentage values of the spread.

ETF	Trade price	Trade size	Quoted sprd	Percent Quoted sprd	Effective sprd	Percent Effective sprd	Control	Trade Price	Trade size	Quoted sprd	Percent Quoted sprd	Effective sprd	Percent Effective sprd
BDH	18.67	1472	0.08	0.44	0.03	0.13	ACTL	13.99	288	0.05	0.34	0.02	0.11
BHH	2.51	1644	0.09	3.69	0.02	0.69	QVDX	2.68	682	0.07	2.56	0.02	0.91
HHH	64.77	867	0.15	0.24	0.03	0.04	CERN	89.36	269	0.27	0.30	0.12	0.13
IAH	35.41	820	0.11	0.30	0.03	0.08	PQE	31.00	358	0.18	0.57	0.06	0.21
IBB	75.43	1211	0.17	0.23	0.06	0.08	IVGN	67.34	339	0.09	0.13	0.03	0.04
IDU	76.47	585	0.13	0.17	0.04	0.06	UIL	49.20	209	0.40	0.82	0.08	0.16
IGE	85.74	445	0.37	0.43	0.12	0.14	PVA	56.28	267	0.85	1.51	0.12	0.21
IGM	46.81	645	0.16	0.35	0.06	0.12	LIFE	35.10	240	0.19	0.55	0.06	0.17
IGN	30.92	704	0.09	0.29	0.03	0.09	MRCY	20.40	345	0.06	0.29	0.02	0.10
IGV	40.57	1105	0.09	0.23	0.03	0.08	IMN	42.99	254	0.28	0.65	0.07	0.17
IGW	59.31	1035	0.12	0.21	0.04	0.07	ATK	73.27	377	0.30	0.41	0.08	0.11
IIH	3.81	1212	0.12	3.13	0.02	0.63	AMSWA	5.96	455	0.20	3.43	0.06	1.10
IYC	58.77	1211	0.15	0.26	0.05	0.08	BL	38.79	187	0.62	1.60	0.10	0.27
IYE	83.64	565	0.15	0.18	0.05	0.06	MIDD	77.29	206	0.29	0.38	0.10	0.14
IYG	110.72	446	0.18	0.17	0.06	0.05	BOKF	45.95	203	0.17	0.37	0.06	0.13
IYH	61.43	664	0.13	0.21	0.04	0.07	UTR	46.02	226	0.37	0.80	0.08	0.17
IYJ	56.50	644	0.13	0.23	0.04	0.08	VTRU	36.35	178	0.26	0.72	0.10	0.27
IYK	52.95	539	0.18	0.34	0.06	0.11	ROG	37.87	215	0.44	1.18	0.09	0.24
IYM	48.80	591	0.12	0.26	0.04	0.09	CW	58.64	288	0.30	0.51	0.07	0.12
IYT	72.23	1251	0.13	0.18	0.04	0.06	FLA	42.53	244	0.22	0.53	0.08	0.18
IYZ	23.19	1277	0.05	0.20	0.01	0.06	WFSL	23.15	289	0.04	0.16	0.01	0.06
OIH	120.78	920	0.17	0.14	0.05	0.04	IBM	83.95	491	0.26	0.31	0.03	0.04
RKH	137.18	1385	0.20	0.15	0.05	0.04	ZION	73.40	272	0.08	0.10	0.02	0.03
RTH	95.27	2007	0.28	0.29	0.09	0.09	TGT	54.39	514	0.17	0.31	0.03	0.06
SMH	36.02	1791	0.04	0.11	0.01	0.03	DELL	31.19	1960	0.03	0.09	0.01	0.03
UTH	113.63	979	0.18	0.16	0.05	0.05	EQT	37.59	434	0.25	0.67	0.05	0.15
VAW	57.64	404	0.14	0.25	0.05	0.09	SCL	25.40	637	0.29	1.15	0.09	0.37
VCR	52.07	1034	0.14	0.26	0.05	0.10	FSCI	46.16	174	0.49	1.05	0.17	0.37
VDE	71.76	348	0.31	0.43	0.09	0.12	TNC	45.19	377	0.38	0.85	0.08	0.18
VFH	55.28	405	0.24	0.42	0.10	0.17	COBH	33.30	239	0.82	2.48	0.27	0.83
VPU	65.45	362	0.20	0.31	0.06	0.09	CHG	46.45	225	0.54	1.17	0.17	0.37
WMH	58.99	350	0.31	0.53	0.06	0.10	FSCI	46.16	174	0.49	1.05	0.17	0.37
XLB	28.56	4464	0.03	0.10	0.01	0.03	MAS	29.25	681	0.17	0.57	0.04	0.15
XLE	49.45	2295	0.04	0.07	0.01	0.03	XOM	57.89	766	0.08	0.14	0.02	0.03
XLK	21.01	1379	0.02	0.11	0.01	0.04	CVC	25.24	1179	0.14	0.54	0.03	0.11
XLU	31.41	1958	0.04	0.12	0.01	0.04	SO	34.61	602	0.10	0.29	0.02	0.06
XLV	31.01	1350	0.04	0.11	0.01	0.03	PFG	48.76	422	0.24	0.50	0.03	0.07
XLY	32.34	2091	0.05	0.16	0.02	0.05	CTL	32.98	375	0.14	0.44	0.03	0.10
Mean	57.01	1117	0.14	0.41	0.04	0.11	Mean	43.32	411	0.27	0.78	0.07	0.22
StdDev	30.98	772	0.08	0.73	0.03	0.14	StdDev	19.88	330	0.20	0.72	0.05	0.24
Median	55.89	1007	0.13	0.23	0.04	0.08	Median	42.76	288	0.25	0.55	0.06	0.15

Roomans (1997), George, Kaul, and Nimalenddran (1991), Lin, Sanger, and Booth (1995), and Huang and Stoll (1997).

5.6.2.1 Glosten and Harris (1988)

In the Glosten and Harris (1988) model, the estimated function is $\Delta P_t = c_0\Delta Q_t + c_1\Delta Q_t V_t + z_0 Q_t + z_1 Q_t V_t + U_t$. From the estimated results, the implied spread is computed by $2(c_0+c_1V_t)+2(z_0+z_1V_t)$ and the percent adverse selection cost is computed by $2(z_0+z_1V_t)/\{2(c_0+c_1V_t)+2(z_0+z_1V_t)\}$. (For the details of the Glosten and Harris model, see section 4.1.1). In Table 5.5, Panels A and B provide the results using the Glosten and Harris (1988) model for the broad-market ETFs and the industry-wide ETFs, respectively. The coefficients c_o and c_1 measure the order processing component of the spread. The c_0 coefficient measures the fixed component and c_1 measures the variable component as a function of trading volume.

The estimates of c_0 for both ETFs and control securities are uniformly positive and significant. This indicates that the order processing costs unrelated to firm size tend to be positive for both groups of securities. Comparisons of the mean values of c_0 for the broad-market and industry-wide ETFs show that these costs are similar across the two groups and are not dissimilar to the average costs associated with the control securities. When the c_1 estimates are negative it indicates that order processing costs decline with trade size. Although both ETFs and the control securities display a tendency for the c_1 coefficients to be positive, few of these coefficients are significant for the ETFs. This shows that for both ETFs and control securities, order processing costs do not tend to fall with trade size.

The coefficients z_0 and z_1 measure the adverse selection component of the spread. The z_0 coefficient measures the part of the adverse selection component that is independent of trade size and the z_1 measures the element linearly related to trade size. In most cases the z_0 estimates associated with the broad-market and the industry-wide ETFs are small, negative and insignificant. However, the estimates of z_0 for the control securities are much more likely to be significant

Table 5.5: Model Estimatses Using Glosten and Harris (1988)

The model estimated is $\Delta P_t = c_0\ \Delta Q_t + c_1\ \Delta(Q_t V_t) + z_0 Q_t + z_1 Q_t V_t + U_t$, where P_t is the security price at time t, Q_t is the transaction indicator variable (+1 for a buyer initiated transaction -1 for a seller initiated trade). V_t is the trade size. c_0 is the inventory cost unassociated with trade size, c_1 is the inventory cost related to trade size. z_o is the adverse selection cost unrelated to trade size while z_1 reflects adverse selection costs related to trade size, the c_1 and z_1 coefficients are multiplied by 1000000. Implied Spread is $[2(c_0 + c_1 V)+2(z_0 + z_1 V)]$. Implied is the magnitude of the spread implied by the model estimates, percentage ASC is the estimated percentage of the spread due to adverse selection costs. Dollar adverse selection cost is percent adverse selection cost times average quoted spread.
An ***indicates that a coefficient is significant at a 1% level and ** at a 5% level and * at a 10% level.

Panel (a) Broad-Market ETFs and their control securities

ETFs	C0	C1	Z0	Z1	Mean Volume	Implied Spread	Percent ASC	Controls	C0	C1	Z0	Z1	Mean Volume	Implied Spread	Percentage ASC
DIA	0.0054***	0.1030***6	0.0016***	0.043***	1560	0.0145	0.2350	MO	0.0071***	0.042***	0.0013***	0.0009	740	0.0168	0.1554
DSG	0.0474***	-2.5666	0.0089	-0.0412	377	0.1107	0.1601	UTL	0.0258***	-3.4163	0.0137***	26.872**	266	0.0914	0.4557
DSV	0.0450***	-0.6326	-0.0081	0.2500	575	0.0734	-0.2168	Y	0.0496	135.10	0.3963	-637.69	200	0.8541	0.6274
ELV	0.0616***	-1.7935	-0.0231**	7.8182	504	0.0840	-0.4321	STU	0.4653***	-394.65	-0.0490	220.63	189	0.7685	-0.0207
IJH	0.0142***	0.302***	0.0002	-0.1825*	652	0.0291	0.0078	HB	0.0056***	-0.815	0.0024*	0.8195	391	0.0159	0.3391
IJJ	0.0242***	0.0161	-0.0014***	0.2438*	756	0.0459	-0.0544	CYN	0.0084***	-1.936**	0.0045***	3.203***	329	0.0265	0.4167
IJK	0.0308***	-0.0062	-0.0002	0.2675*	1020	0.0619	0.0036	CBSH	0.0082***	0.553***	0.0046***	0.594***	218	0.0260	0.3602
IJR	0.0098***	0.0301	0.0005**	0.0388	1354	0.0206	0.0499	NOC	0.0042***	0.496***	0.0030***	-0.38***	561	0.0146	0.3880
IJS	0.0250***	0.0534	-0.0024***	0.1045	634	0.0453	-0.1038	MRBK	0.0118***	1.122***	0.0052***	-0.071	217	0.0345	0.3029
IJT	0.0441***	-0.4885	-0.0063***	0.4847	695	0.0756	-0.1590	ESE	0.0136***	-0.6757	0.0033**	-0.0412	327	0.0333	0.1950
IVE	0.0156***	0.333***	0.0006*	-0.0707	659	0.0327	0.0320	MCY	0.0064***	-0.4372	0.0048*	1.207	256	0.0228	0.4465
IVV	0.0058***	-0.0178	0.0033***	0.157***	1261	0.0186	0.3761	CB	0.0070***	0.593***	0.0044***	0.0064	322	0.0232	0.3777
IVW	0.0205***	0.0510	0.0001	0.0046	663	0.0413	0.0060	UB	0.0080***	0.586**	0.0057***	-0.6592	436	0.0273	0.3961
IWB	0.0170***	0.0676	-0.0005	-0.1064*	935	0.0329	-0.0389	TMK	0.0047***	0.1576	0.0019**	-0.376**	525	0.0129	0.2604
IWD	0.0127***	0.201***	0.0000	-0.0238	1280	0.0259	0.0016	BCR	0.0040***	0.745***	0.0039***	-0.95***	415	0.0155	0.4455
IWF	0.0110***	0.0072	-0.0008***	0.0067	1611	0.0205	-0.0755	LH	0.0065***	0.362***	0.0023***	-0.0658	504	0.0178	0.2532
IWM	0.0040***	0.064***	0.0012***	0.019***	3142	0.0110	0.2421	GE	0.0043***	0.013***	0.0002***	0.011***	982	0.0092	0.0573
IWN	0.0066***	0.0186*	0.0025***	-0.0034	1490	0.0183	0.2753	CMA	0.0049***	0.788***	0.0028***	-0.2267	338	0.0158	0.3468
IWO	0.0057***	0.175***	0.0026***	0.0534	1336	0.0173	0.3028	ABC	0.0036***	1.067***	0.0045***	-0.165	361	0.0169	0.5299
IWP	0.0311***	0.056	0.0006	-0.1060	976	0.0633	0.0169	IDXX	0.0137***	3.027***	0.0080***	-1.2763*	215	0.0441	0.3503
IWR	0.0285***	0.1615	0.0002	-0.0134	722	0.0577	0.0081	WPS	0.0111***	0.39	0.0048***	-0.3616	348	0.0319	0.2922
IWV	0.0222***	0.1563	-0.0005	0.0137	1011	0.0439	-0.0212	EEP	0.0146***	1.9161	0.0065***	-1.4508	355	0.0426	0.2849
JKH	0.0370***	-1.4917	0.0064	0.7526	638	0.0859	0.1601	COKE	0.0626***	0.5908	0.0164***	-9.8877	186	0.1546	0.1889
JKJ	0.0385***	-8.5779	-0.0082	-4.2818	769	0.0408	-0.5641	SFSW	0.0407***	-36.571	-0.0045	11.4517	267	0.0585	-0.0503
JKK	0.0295***	3.4322	-0.0059	-0.3820	658	0.0510	-0.2424	BF	0.0128***	0.3496	0.0042***	0.4953	385	0.0350	0.2730
JKL	0.0193*	-0.6496	-0.0004	0.9747	813	0.0378	-0.0061	BARI	0.0071***	1.5869	0.0083***	-1.7448	332	0.0306	0.5145
MDY	0.0064***	0.236***	0.0041***	-0.0276	1422	0.0217	0.3795	LEH	0.0083***	0.74***	0.0063***	-0.42***	394	0.0294	0.4175
PEY	0.0098***	0.0507	-0.0004*	0.0597	927	0.0190	-0.0348	BFIN	0.0100***	0.617**	0.0023***	0.3128	332	0.0253	0.1941
PWC	0.0220***	-0.2628	-0.0003	0.2435	598	0.0434	-0.0064	KMR	0.0155***	-1.5212	0.0029*	2.8083	351	0.0378	0.2071
PWO	0.0247***	-0.6814*	0.0004	0.8211	550	0.0506	0.0333	CFFN	0.0101***	0.9453	0.0039***	-0.0409	220	0.0284	0.2723
QQQQ	0.0033***	0.007***	0.0003***	0.013***	4362	0.0073	0.0942	INTC	0.0043***	-0.0007*	0.0005***	0.017***	3309	0.0097	0.1066
VB	0.0283***	1.2173	0.0033	-0.2304	699	0.0646	0.0980	TRH	0.0371***	-2.8386	0.0024	10.5293	204	0.0820	0.1093
VBK	0.0350***	2.248**	-0.0009	-2.4315*	552	0.0680	-0.0654	YANB	0.0229***	4.1124	0.0131***	-6.1452	196	0.0712	0.3345
VBR	0.0235***	-0.7621	0.0019	-0.0688	398	0.0502	0.0754	WSFS	0.0477***	1.2557	0.0207***	-0.1376	220	0.1371	0.2996
VO	0.0225***	-0.0794	-0.0037***	0.5681	572	0.0382	-0.1755	GBL	0.0072	95.23***	0.0114	-58.068	187	0.0510	0.0203
VTI	0.0285***	-0.1856	-0.0039***	0.2856	595	0.0494	-0.1507	PDX	0.0174***	0.9422	0.0130***	-2.2052	285	0.0600	0.4081
VTV	0.0194***	-0.1644	-0.0012	-0.2183	748	0.0358	-0.0781	ERIE	0.0100***	0.699***	0.0055***	-0.0198	272	0.0315	0.3529
VV	0.0184***	-0.6346	-0.0008	1.1789	710	0.0360	0.0001	IBOC	0.0107***	0.691**	0.0047***	0.4196	297	0.0315	0.3131
VXF	0.0329***	0.0042	-0.0039	-0.4550	630	0.0575	-0.1461	BANF	0.0266***	9.1099	0.0154***	9.3352	151	0.0895	0.3772
Ave.	0.0228	-0.2564	-0.0009	0.1477	996	0.0436	-0.0003	Ave.	0.0266	-4.5900	0.0144	-11.119	412	0.0759	0.3003
StdDev	0.0136	1.6577	0.0050	1.5352	742	0.0232	0.1913	StdDev	0.0735	69.54	0.0636	109.43	502	0.1574	0.1605
Med	0.0222	0.0161	-0.0002	0.0137	722	0.0413	0.0016	Med	0.0100	0.5908	0.0045	-0.0412	327	0.0314	0.3097
Sign	< 0.0001	0.3368	1	0.3368		< 0.0001	0.5224	Sign	< 0.0001	0.0034	< 0.0001	0.5224		< 0.0001	< 0.0001
Signed Rank	0.0015	0.0646	< 0.0001	0.4499		0.4254	< 0.0001								

Panel (b) Industry-wide ETFs and their control securities

ETFs	C0	C1	Z0	Z1	Mean Volume	Implied Spread	Percent ASC	Controls	C0	C1	Z0	Z1	Mean Volume	Implied Spread	Percentage ASC
BDH	0.0081***	-0.0475	0.0007**	0.0358	1472	0.0175	0.0777	ACTL	0.0052***	0.324***	0.0020***	0.0137	288	0.0144	0.2734
BHH	0.0049***	0.0177	-0.0002	-0.0599	1644	0.0093	-0.0648	QVDX	0.0050***	-0.0349	0.0017***	0.01711	682	0.0133	0.2544
HHH	0.0076***	0.322***	0.0018***	-0.1846**	867	0.0189	0.1726	CERN	0.0203***	0.721***	0.0081***	0.4645*	269	0.0574	0.2884
IAH	0.0064***	-0.092	0.0005	0.0932	820	0.0138	0.0839	PQE	0.0087***	-0.0605	0.0017	-1.0776	358	0.0199	0.1260
IBB	0.0112***	0.303***	0.0012***	0.0289	1211	0.0254	0.0914	IVGN	0.0076***	0.774***	0.0041***	0.161***	339	0.0239	0.3484
IDU	0.0221***	0.3719	-0.0003	-0.7441*	585	0.0432	-0.0350	UIL	0.0057*	8.4205	0.0138***	-11.3625	209	0.0378	0.6052
IGE	0.0597***	-0.4704	-0.0020*	-1.1152	445	0.1140	-0.0432	PVA	0.0090***	-1.1589	0.0070**	0.8167	267	0.0318	0.4587
IGM	0.0275***	-0.3034	-0.0008	0.1909	645	0.0532	-0.0267	LIFE	0.0199***	-0.0955	0.0073***	-0.2297	240	0.0542	0.2660
IGN	0.0153***	-0.12	-0.0008**	0.1988	704	0.0291	-0.0401	MRCY	0.0066***	0.2369***	0.0023***	0.126**	345	0.0181	0.2620
IGV	0.0132***	0.148***	0.0010*	-0.1031	1105	0.0285	0.0613	IMN	0.0088***	2.4455**	0.0073***	-3.147**	254	0.0321	0.4076
IGW	0.0168***	0.0946*	0.0011**	0.0322	1035	0.0360	0.0635	ATK	0.0130***	0.4851	0.0094***	0.7866	377	0.0457	0.4241
IIH	0.0042***	0.0464	0.0000	0.0063	1212	0.0085	0.0070	AMSWA	0.0088***	0.0388	0.0024***	-0.3717	455	0.0220	0.1990
IYC	0.0290***	0.0331	-0.0032**	-0.0179	1211	0.0515	-0.1257	BL	0.0140***	-1.5332	0.0071***	10.359***	187	0.0454	0.3949
IYE	0.0157***	1.013***	0.0028***	-0.1694	565	0.0379	0.1400	MIDD	0.0313***	-3.36***	0.0138***	2.834**	206	0.0899	0.3191
IYG	0.0247***	1.5275	0.0117***	-1.9979	446	0.0724	0.2984	BOKF	0.0114***	1.88***	0.0059***	-2.311***	203	0.0344	0.3160
IYH	0.0255***	0.0509	-0.0008*	-0.2769	664	0.0489	-0.0418	UTR	0.0039***	2.1465	0.0073***	-4.93**	226	0.0212	0.5839
IYJ	0.0255***	0.7892	-0.0009	0.0968	644	0.0503	-0.0344	VTRU	0.0201***	1.1058	0.0113***	-3.6185	178	0.0620	0.3426
IYK	0.0316***	-0.1519	-0.0016*	-0.1630	539	0.0598	-0.0554	ROG	0.0088***	-0.5088	0.0142***	3.2548	215	0.0472	0.6315
IYM	0.0280***	-0.501***	-0.0016**	-0.0299	591	0.0520	-0.0636	CW	0.0122***	5.178*	0.0071*	-9.303**	288	0.0363	0.2448
IYT	0.0172***	0.0151	-0.0009	0.0497	1251	0.0326	-0.0532	FLA	0.0189***	-0.2942	0.0050	7.4279	244	0.0515	0.2666
IYZ	0.0085***	-0.0183	0.0002	0.1272	1277	0.0176	0.0416	WFSL	0.0054***	0.294***	0.0012***	0.232**	289	0.0136	0.1911
OIH	0.0117***	0.449***	0.0037***	0.2310***	920	0.0319	0.2478	IBM	0.0058***	0.345***	0.0019***	0.138***	491	0.0159	0.2560
RKH	0.0132***	0.526***	0.0036***	-0.2121***	1385	0.0343	0.1922	ZION	0.0079***	1.224***	0.0050***	0.0305	272	0.0264	0.3785
RTH	0.0067***	0.176***	0.0026***	0.0256*	2007	0.0195	0.2757	TGT	0.0069***	0.136***	0.0024***	0.0123	514	0.0188	0.2614
SMH	0.0029***	0.058***	0.0006***	0.0378***	1791	0.0072	0.1872	DELL	0.0041***	0.017***	0.0007***	0.03***	1960	0.0099	0.1571
UTH	0.0113***	0.672***	0.0030***	-0.0586	979	0.0298	0.2005	EQT	0.0074***	-0.0164	0.0042***	-0.2393	434	0.0230	0.3558
VAW	0.0284***	4.7962	-0.0072*	-1.3716	404	0.0452	-0.3446	SCL	0.0434***	-4.251	0.0127	8.4379	637	0.1174	0.3072
VCR	0.0336***	-0.8617	-0.0015	-0.0757	1034	0.0622	-0.0515	FSCI	0.0461***	-6.4991	0.0146***	23.4804*	174	0.1275	0.2944
VDE	0.0316***	0.702	-0.0035**	-0.04545	348	0.0563	-0.1315	TNC	0.0150***	2.9422*	0.0054	-2.9631	377	0.0408	0.2106
VFH	0.0202***	1.1335	0.0000	0.3485	405	0.0418	0.0094	COBH	0.0669***	-14.8013	0.0165*	7.7471	239	0.1621	0.2164
VPU	0.0313***	4.0424	-0.0055*	4.0007	362	0.0574	-0.1414	CHG	0.0320***	-15.2052	0.0056	5.2492	225	0.0706	0.1893
WMH	0.0247***	4.2912	0.0040	-4.6582	350	0.0570	0.0812	FSCI	0.0461***	-6.4991	0.0146***	23.48*	174	0.1275	0.2944
XLB	0.0035***	0.022***	0.0009***	-0.0026	4464	0.0089	0.1796	MAS	0.0038***	0.074***	0.0014***	0.0072	681	0.0105	0.2771
XLE	0.0048***	0.089***	0.0012***	0.0519***	2295	0.0126	0.2130	XOM	0.0050***	0.026***	0.0015***	0.017***	766	0.0131	0.2287
XLK	0.0034***	0.037***	0.0007***	0.0095*	1379	0.0083	0.1736	CVC	0.0042***	0.045***	0.0019***	-0.025*	1179	0.0122	0.3038
XLU	0.0046***	0.084***	0.0008***	0.0008	1958	0.0112	0.1524	SO	0.0049***	0.055***	0.0018***	-0.0007	602	0.0133	0.2652
XLV	0.0052***	0.0204*	0.0002	-0.0017	1350	0.0107	0.0291	PFG	0.0048***	-0.0097	0.0015***	-0.1838	422	0.0124	0.2283
XLY	0.0052***	0.042***	0.0000	0.0256***	2091	0.0108	0.0169	CTL	0.0037***	0.0802	0.0021***	-0.1484	375	0.0115	0.3489
Ave.	0.0170	0.5082	0.0003	-0.1606	1117	0.0349	0.0457	Ave.	0.0145	-0.6665	0.0061	1.4529	412	0.0417	0.3098
StdDev	0.0123	1.2348	0.0030	1.1044	772	0.0230	0.1347	StdDev	0.0147	4.2755	0.0047	6.6627	331	0.0375	0.1139
Med	0.0142	0.0717	0.0002	-0.0005	1007	0.0323	0.0351	Med	0.0087	0.0502	0.0052	0.0171	289	0.0291	0.2804
Sign	< 0.0001	0.0017	0.2559	1		< 0.0001	0.2559	Sign	< 0.0001	0.2559	< 0.0001	0.2559		< 0.0001	< 0.0001
Signed Rank	0.1134	0.9717	< 0.0001	0.6746		0.4892	< 0.0001								
Ave.	0.0199	0.1209	-0.0003	-0.0045	1056	0.0393	0.0225	Ave.	0.0207	-2.6537	0.0103	-4.9151	412	0.0612	0.3036
StdDev	0.0132	1.5048	0.0041	1.3403	754	0.0234	0.1665	StdDev	0.0533	49.2999	0.0453	77.7814	424	0.1283	0.1331
Med	0.0193	0.0464	0.0000	0.0067	813	0.0378	0.0078	Med	0.0088	0.3241	0.0047	0.0064	322	0.0315	0.2944
Sign	< 0.0001	0.0028	0.4944	0.4944		< 0.0001	0.1711	Sign	< 0.0001	0.0028	< 0.0001	0.8199		0.0133	< 0.0001
Signed Rank	0.0008	0.1469	< 0.0001	0.7879		0.8861	< 0.0001								

and positive. The mean adverse selection cost is insignificant for both the broad-market and the industry-wide ETFs while for both sets of control securities the mean estimates of z_o are positive and significant. This suggests that this adverse selection component is not important for the ETFs but is important for the control securities.

Easley and O'Hara (1987) developed a model in which informed traders choose to trade larger amounts at any given price. This model predicts that the z_1 coefficients should be positive to ensure that the cost of engaging in larger trades reflects the likelihood of higher adverse selection costs. The results for both the ETFs and the control securities indicate no clear relationship between trade size and adverse selection costs. Many of the z_1 coefficients are negative and few are statistically significant (any sign) for either the ETFs or the control securities. For the broad-market ETFs, six of the z_1 coefficients are positive and significant and only five are for the control securities. A similar picture is apparent for the industry-wide ETFs as only five z_1 coefficients are significantly positive and eleven for the control securities.

The high number of negative z_0 and z_1 coefficients associated with the ETFs cause overall estimates of the adverse selection component as a percentage of the spread to be negative for many of the ETFs(eighteen adverse selection costs are negative for broad-market ETFs and fifteen for the industry-wide ETFs)[1]. In contrast, percentage adverse selection costs for the control securities are almost always positive. In general the results using this model suggest an absence of adverse selection costs associated with ETFs but positive adverse selection costs associated with the control securities.

The average percent adverse selection costs of control stocks is similar to those of Neal and Wheatley (1998). This slight difference is due to the difference between the Glosten and Harris (1988) Model in my analysis and the model in Neal and Wheatley (1998), in which dummy variables are employed to consider higher opening and closing bid-ask spread. The average/median adverse selection cost of broad-market ETF (-0.0003/0.0016) is smaller than the average/median

[1]These results may reflect strategic behaviour that causes informed traders to trade smaller qualities to mask their identity. Empirical evidence for example by Barclay and Warner (1993) suggests that medium sized trades are more likely to be initiated by informed traders.

adverse selection cost of matched control stocks (0.3003/0.3097). The same pattern appears in the industry-wide ETFs.

The two bottom rows of each panel shows the non-parametric test p-value (Wilcoxon sign test and Wilcoxon signed rank sum test), since parametric methods (T-tests and Paired T-tests) basically provide the same results as non-parametric results. Non-parametric methods take advantage of not assuming the population distribution of tested value. Furthermore, non-parametric tests used in this thesis are based on percent adverse selection costs.

The Wilcoxon sign tests of both ETFs are insignificant in the case of percent adverse selection costs, whilst control securities show positive percent adverse selection costs significantly different from zero. The Wilcoxon signed rank sum tests demonstrate that both ETFs have clearly different percent adverse selection costs from control securities.

The five bottom rows of the industry-wide ETFs panel (Panel (b) of Table 5.5) represent the test statistics of all ETFs and control securities. Their results are consistent with the results of broad-market ETFs and industry-wide ETFs. While control securities have on average 60 % adverse selection cost of the spread, ETFs have around 20 % adverse selection cost of the spread. This difference is supported by the signed rank sum test.

In summary, the Glosten and Harris (1988) model accepts the empirical hypothesis that ETFs have lower adverse selection cost than matched control stocks. This empirical hypothesis is true in all cases of full ETFs, broad-market ETFs, and industry-wide ETFs. Note that the estimated implied spreads of ETFs and control securities do not report any difference. This result may reflect the fact that the matched control securities are properly selected in the light of the Glosten and Harris (1988) model.

Table 5.6: Model Estimatses Using Madhavan, Richardson, and Roomans (1997)

The ϕ is the dollar inventory cost, and θ is the dollar adverse selection cost. The estimated implied spread is calculated as $S=2(\theta+\phi)$. The percentage adverse selection component of the spread is $2\theta/2(\theta+\phi)$. Dollar adverse selection cost is computed by the average quoted spread times estimated percent adverse selection costs. . Dollar adverse selection cost is percent adverse selection cost times average quoted spread. An ***indicates that a coefficient is significant at a 1% level and ** at a 5% level and * at a 10% level.

Panel (a) Broad-Market ETFs and their control securities

ETFs	Inventory (ϕ)	Adverse Selection (θ)	Implied Spread $2(\phi+\theta)$	Percent Adverse Selection $2\theta/2(\theta+\phi)$	Controls	Inventory (ϕ)	Adverse Selection (θ)	Implied Spread $2(\phi+\theta)$	Percent Adverse Selection $2\theta/2(\theta+\phi)$
DIA	0.0052***	0.0091***	0.0286	0.6345	MO	0.0047***	0.0169***	0.0432	0.7823
DSG	0.0068	0.0109	0.0354	0.6154	UTL	0.0054	0.0195***	0.0497	0.7836
DSV	0.0133	0.0083	0.0433	0.3857	Y	0.0274	0.0303	0.1153	0.5251
ELV	0.0223*	0.0018	0.0481	0.0749	STU	0.0264	-0.0223	0.0081	-5.4898
IJH	0.0166***	0.0035***	0.0402	0.1721	HB	0.0126***	0.0137***	0.0528	0.5209
IJJ	0.0190***	0.0025***	0.0431	0.1169	CYN	0.0016	0.0152***	0.0337	0.9034
IJK	0.0203***	0.0021	0.0447	0.0936	CBSH	0.0028***	0.0128***	0.0313	0.8193
IJR	0.0184***	0.0025***	0.0417	0.1199	NOC	0.0049***	0.0112***	0.0323	0.6948
IJS	0.0212***	0.0006	0.0436	0.0295	MRBK	0.0049***	0.0160***	0.0417	0.7656
IJT	0.0287***	-0.0050**	0.0472	-0.2130	ESE	0.0040	0.0104***	0.0289	0.7225
IVE	0.0246***	-0.0022***	0.0448	-0.0992	MCY	0.0041	0.0107**	0.0296	0.7250
IVV	0.0207***	0.0007*	0.0428	0.0338	CB	0.0039***	0.0119***	0.0316	0.7546
IVW	0.0205***	0.0008	0.0426	0.0382	UB	0.0085***	0.0054***	0.0278	0.3854
IWB	0.0197***	0.0011	0.0417	0.0536	TMK	0.0135***	-0.0044***	0.0182	-0.4878
IWD	0.0195***	0.0009**	0.0408	0.0459	BCR	-0.0050***	0.0168***	0.0237	1.4174
IWF	0.0187***	0.0009***	0.0392	0.0476	LH	0.0051***	0.0087***	0.0276	0.6333
IWM	0.0155***	0.0010**	0.0331	0.0615	GE	0.0133***	0.0149***	0.0565	0.5285
IWN	0.0141***	0.0022***	0.0326	0.1362	CMA	0.0069***	0.0074***	0.0286	0.5190
IWO	0.0118***	0.0038***	0.0311	0.2415	ABC	0.0011***	0.0079***	0.0180	0.8782
IWP	0.0126***	0.0034*	0.0320	0.2135	IDXX	0.0063***	0.0159***	0.0444	0.7171
IWR	0.0133***	0.0034***	0.0334	0.2014	WPS	0.0126***	0.0156***	0.0564	0.5529
IWV	0.0146***	0.0032**	0.0355	0.1785	EEP	0.0019	0.0140***	0.0318	0.8817
JKH	0.0148	0.0090	0.0477	0.3781	COKE	0.0037	0.0334***	0.0742	0.9005
JKJ	0.0218	0.0029	0.0496	0.1187	SFSW	0.0137**	0.0045	0.0364	0.2473
JKK	0.0324	-0.0045	0.0559	-0.1600	BF	-0.0031	0.0175***	0.0289	1.2123
JKL	0.0293*	0.0015	0.0615	0.0473	BARI	-0.0096**	0.0229***	0.0265	1.7228
MDY	0.0212***	0.0052***	0.0527	0.1973	LEH	0.0053***	0.0094***	0.0292	0.6406
PEY	0.0168***	0.0080***	0.0496	0.3229	BFIN	-0.0007	0.0133***	0.0251	1.0555
PWC	0.0167***	0.0072***	0.0479	0.3026	KMR	0.0060***	0.0076***	0.0271	0.5582
PWO	0.0172***	0.0058***	0.0462	0.2534	CFFN	0.0023***	0.0162***	0.0370	0.8766
QQQQ	0.0169***	0.0049***	0.0437	0.2253	INTC	0.0069***	0.0069***	0.0276	0.5000
VB	0.0128**	0.0097	0.0449	0.4299	TRH	0.0126	0.0018	0.0288	0.1276
VBK	0.0140***	0.0088	0.0456	0.3852	YANB	0.0213***	0.0274***	0.0973	0.5625
VBR	0.0134***	0.0096**	0.0460	0.4175	WSFS	0.0101**	0.0284***	0.0771	0.7373
VO	0.0217***	0.0003	0.0441	0.0148	GBL	0.0125	0.0178	0.0606	0.5876
VTI	0.0199***	0.0045**	0.0488	0.1835	PDX	0.0059	0.0134*	0.0387	0.6926
VTV	0.0201***	0.0033	0.0468	0.1411	ERIE	0.0028***	0.0119***	0.0294	0.8069
VV	0.0192***	0.0047	0.0478	0.1965	IBOC	0.0098***	0.0108***	0.0413	0.5234
VXF	0.0190***	0.0064	0.0509	0.2531	BANF	0.0008	0.0192***	0.0399	0.9615
Ave.	0.0181	0.0037	0.0435	0.1767	Ave.	0.0069	0.0131	0.0399	0.5448
StdDev	0.0054	0.0038	0.0070	0.1813	StdDev	0.0074	0.0095	0.0213	1.0505
Med	0.0187	0.0033	0.0441	0.1721	Med	0.0053	0.0134	0.0318	0.7171
Sign	< 0.0001	< 0.0001	< 0.0001	< 0.0001	Sign	< 0.0001	< 0.0001	< 0.0001	< 0.0001
SignedRank	< 0.0001	< 0.0001	0.1450	< 0.0001					

Panel (b) Industry-wide ETFs and their control securities

ETFs	Inventory (ϕ)	Adverse Selection (θ)	Implied Spread $2(\phi+\theta)$	Percent Adverse Selection $2\theta/2(\theta+\phi)$	Controls	Inventory (ϕ)	Adverse Selection (θ)	Implied Spread $2(\phi+\theta)$	Percent Adverse Selection $2\theta/2(\theta+\phi)$
BDH	0.0049***	0.0158***	0.0413	0.7653	ACTL	0.0017***	0.0069***	0.0173	0.8033
BHH	0.0051***	0.0120***	0.0342	0.7006	QVDX	0.0051***	0.0091***	0.0283	0.6433
HHH	0.0184***	0.0032***	0.0431	0.1475	CERN	0.0014***	0.0187***	0.0401	0.9325
IAH	0.0157***	0.0036*	0.0386	0.1854	PQE	0.0066***	0.0083***	0.0298	0.5576
IBB	0.0144***	0.0044***	0.0375	0.2340	IVGN	0.0054***	0.0086***	0.0278	0.6150
IDU	0.0154***	0.0048***	0.0402	0.2362	UIL	0.0051	0.0107	0.0317	0.6750
IGE	0.0237***	0.0043**	0.0560	0.1532	PVA	0.0053	0.0132**	0.0368	0.7144
IGM	0.0244***	0.0026	0.0540	0.0950	LIFE	0.0047***	0.0127***	0.0349	0.7279
IGN	0.0233***	0.0016**	0.0499	0.0646	MRCY	0.0047***	0.0131***	0.0355	0.7365
IGV	0.0197***	0.0029***	0.0451	0.1285	IMN	0.0060***	0.0139***	0.0397	0.6997
IGW	0.0183***	0.0044***	0.0454	0.1953	ATK	0.0024	0.0108***	0.0265	0.8184
IIH	0.0167***	0.0037***	0.0408	0.1828	AMSWA	0.0022***	0.0077***	0.0199	0.7790
IYC	0.0182***	0.0010	0.0385	0.0538	BL	0.0019	0.0140***	0.0318	0.8777
IYE	0.0165***	0.0050***	0.0430	0.2314	MIDD	0.0010	0.0241***	0.0501	0.9609
IYG	0.0145*	0.0078	0.0447	0.3502	BOKF	0.0024**	0.0141***	0.0331	0.8537
IYH	0.0183***	0.0044***	0.0454	0.1944	UTR	0.0178***	0.0184***	0.0724	0.5071
IYJ	0.0239***	-0.0001	0.0477	-0.0028	VTRU	0.0139***	0.0186***	0.0649	0.5729
IYK	0.0262***	-0.0017	0.0491	-0.0692	ROG	0.0024	0.0172**	0.0392	0.8777
IYM	0.0280***	-0.0016	0.0527	-0.0623	CW	0.0039	0.0162**	0.0402	0.8041
IYT	0.0245***	0.0006	0.0501	0.0228	FLA	0.0046	0.0124**	0.0341	0.7291
IYZ	0.0203***	0.0010***	0.0425	0.0452	WFSL	0.0138***	0.0185***	0.0645	0.5720
OIH	0.0167***	0.0088***	0.0511	0.3442	IBM	0.0118***	0.0100***	0.0434	0.4590
RKH	0.0155***	0.0071***	0.0451	0.3144	ZION	0.0163***	0.0216***	0.0756	0.5701
RTH	0.0131***	0.0082***	0.0427	0.3851	TGT	0.0162***	-0.0074***	0.0177	-0.8395
SMH	0.0124***	0.0081***	0.0411	0.3952	DELL	0.0021***	0.0139***	0.0319	0.8706
UTH	0.0097***	0.0101***	0.0396	0.5081	EQT	0.0031***	0.0111***	0.0285	0.7803
VAW	0.0138**	0.0063	0.0402	0.3127	SCL	0.0130	0.0079	0.0419	0.3788
VCR	0.0156**	0.0083	0.0478	0.3489	FSCI	0.0131*	0.0220***	0.0703	0.6269
VDE	0.0190***	0.0087***	0.0553	0.3135	TNC	0.0142*	-0.0038	0.0207	-0.3645
VFH	0.0189***	0.0076**	0.0530	0.2852	COBH	0.0087	0.0107	0.0387	0.5528
VPU	0.0178***	0.0064	0.0485	0.2637	CHG	0.0083	0.0089	0.0344	0.5196
WMH	0.0195***	0.0062	0.0512	0.2405	FSCI	0.0131*	0.0220***	0.0703	0.6269
XLB	0.0141***	0.0046***	0.0374	0.2465	MAS	0.0053***	0.0089***	0.0285	0.6267
XLE	0.0138***	0.0047***	0.0369	0.2523	XOM	0.0113***	0.0187***	0.0599	0.6234
XLK	0.0126***	0.0043***	0.0337	0.2536	CVC	0.0075***	0.0135***	0.0419	0.6446
XLU	0.0103***	0.0042***	0.0290	0.2891	SO	0.0097***	0.0036***	0.0267	0.2712
XLV	0.0094***	0.0036***	0.0260	0.2784	PFG	0.0064***	0.0101***	0.0330	0.6141
XLY	0.0086***	0.0032***	0.0236	0.2701	CTL	0.0074***	0.0184***	0.0516	0.7144
Ave.	0.0166	0.0050	0.0432	0.2409	Ave.	0.0074	0.0126	0.0398	0.6088
StdDev	0.0055	0.0036	0.0077	0.1740	StdDev	0.0049	0.0065	0.0158	0.3299
Med	0.0166	0.0044	0.0431	0.2435	Med	0.0057	0.0129	0.0352	0.6439
Sign	< 0.0001	< 0.0001	< 0.0001	< 0.0001	Sign	< 0.0001	< 0.0001	< 0.0001	< 0.0001
SignedRank	< 0.0001	< 0.0001	0.1833	< 0.0001					
Ave.	0.0173	0.0043	0.0433	0.2083		0.0071	0.0128	0.0399	0.5764
StdDev	0.0054	0.0037	0.0073	0.1795		0.0063	0.0081	0.0187	0.7784
Med	0.0172	0.0042	0.0437	0.1973		0.0054	0.0132	0.0341	0.6926
Sign	< 0.0001	< 0.0001	< 0.0001	< 0.0001		< 0.0001	< 0.0001	< 0.0001	< 0.0001
SignedRank	< 0.0001	< 0.0001	0.0528	< 0.0001					

5.6.2.2 Madhavan, Richardson, and Roomans (1997)

The Madhavan, et al. (1997) model employs the generalised method of moments and has as population moments;

$$E\begin{pmatrix} x_t x_{t-1} - x_t^2\rho \\ |x_t| - (1-\lambda) \\ m_t - \alpha \\ (m_t - \alpha)x_t \\ (m_t - \alpha)x_{t-1} \end{pmatrix} = 0, \tag{5.22}$$

where $m_t = P_t - P_{t-1} - (\phi+\theta)x_t + (\phi+\rho\theta)x_{t-1}$. Based on population moments conditions, the Madhavan, et al. (1997) model identifies four parameters vectors (θ, ϕ, λ, ρ) and a constant drift α. The estimated implied spread is calculated by $2(\theta+\phi)$ and the percentage adverse selection cost is computed by $\%\theta = \frac{2\theta}{2(\theta+\phi)}$. In the Madhavan, et al. (1997) model, estimates of ϕ indicate dollar inventory costs while θ provides estimates of the dollar adverse selection costs. (For the details of the Madhavan et al (1997) model, see section 4.1.4)

Table 5.6 contains the results of the Madhavan, et al. (1997) model for the broad-market and industry-wide ETFs respectively. Estimates of θ for the ETFs are uniformly positive and almost always significant. For the controls, the overwhelming majority of these coefficients are also positive and significant. These results indicate that average dollar inventory costs are higher for the ETFs than for the control securities since the mean value of θ is 0.0037 for broad-market and 0.005 for industry-wide ETFs as opposed to 0.0131 and 0.0126 for their respective controls. Values of θ are uniformly positive for the control securities but are negative for the ETFs. The higher values of θ associated with the control securities cause the mean percentage adverse selection component of the spread to be noticeably higher in the control securities when compared to either the broad-market or the industry-wide ETFs. Comparisons of the mean percentage adverse selection costs of broad-market and industry-wide ETFs suggest that the industry-wide ETFs have slightly higher adverse selection costs than the broad-market ETFs. When the dollar adverse selection costs is computed by percent adverse selection cost times average quoted spread, the average dollar adverse selection of broad-market ETFs is lower than that of the control securities.

Each panel shows Wilcoxon sign test statistics and signed rank sum test. The bottom five rows of industry-wide ETFs (panel (b) of Table 5.6) show summary statistics of all ETFs and controls. The Wilcoxon sign test demonstrates that all estimated adverse selection and inventory holding costs are significantly positive in both cases of broad-market and industry-wide ETFs. The Wilcoxon signed rank sum test p-values suggest that ETFs have smaller percent adverse selection costs than control securities. The same interpretation appears in the dollar adverse selection costs. The Madhavan, et al. (1997) model accepts the empirical hypothesis.

Note that the estimated implied spreads of ETFs and control securities do not report any difference when broad-market ETFs and industry-wide ETFs are considered independently. This result may reflect the fact that the matched control securities are properly selected in view of the Madhavan, et al. (1997) model.

5.6.2.3 George, Kaul, and Nimalendran (1991)

The George, et al. (1991) model is estimated from $RD_{it} = \alpha_i + \frac{\beta_i}{2}$ ($Q_{it} * s_{qit}$ - $Q_{it-1} * s_{qit-1}$) + V_t and this function is a modified George, et al. (1991) model and is used in the Neal and Wheatley (1998). The estimated β_i represents the proportional order processing cost including inventory holding cost. Thus, the percentage adverse selection cost is calculated by $(1 - \beta_i)$. (For the details of the George et al (1991) model, see section 4.1.2)

For both ETFs and the control securities, the order processing (β_i) and adverse selection $(1 - \beta_i)$ costs are always positive and lead to substantive component costs. Mean percentage order processing and adverse selection costs are positive and significant for both broad-market and industry-wide ETFs. For both broad-market and industry-wide ETFs, the adverse selection costs account for the highest proportion of the spread representing about two-thirds of the costs for broad-market ETFs and about 58% of the spread for industry-wide ETFs. For all securities the intercept (α_i) is insignificant.

The average percent adverse selection costs of broad-market ETFs and industry-wide ETFs are 0.6562 and 0.5876, respectively. Meanwhile the percent

Table 5.7: Model Estimatses Using George, Kaul, and Nimalenran (1991)

The model estimated is $2RD_t = \pi_0 + \pi_1 \Delta(S_{qt}Q_t) + V_t$, where RD_t are transaction returns less mid-point returns. The coefficient, π_1, represents the percentage order processing cost. The percentage adverse selection costs are calculated as 1- π_1. Dollar adverse selection cost is percent adverse selection cost times average quoted spread. An ***indicates that a coefficient is significant at a 1% level and ** at a 5% level and * at a 10% level.

Panel (a) Broad-Market ETFs and their control securities

ETFs	Intercept (π_0)	Percentage Order Processing (π_1)	Percentage Adverse Selection (1- π_1)	Controls	Intercept (π_0)	Percentage Order Processing (π_1)	Percentage Adverse Selection (1- π_1)
DIA	0.0000	0.4488***	0.5512***	MO	0.0001	0.6988***	0.3012***
DSG	-0.0214	0.2164***	0.7836***	UTL	-0.0040	0.5103***	0.4897***
DSV	-0.0004	0.1716***	0.8284***	Y	-0.0087	0.3535***	0.6465***
ELV	0.0147	0.0812***	0.9188***	STU	-0.1051	0.5446***	0.4554***
IJH	-0.0002	0.5063***	0.4937***	HB	-0.0012	0.2774***	0.7226***
IJJ	0.0000	0.2004***	0.7996***	CYN	-0.0004	0.3719***	0.6281***
IJK	0.0006	0.5227***	0.4773***	CBSH	0.0002	0.9496***	0.0504***
IJR	0.0004	0.5006***	0.4994***	NOC	-0.0002	0.5367***	0.4633***
IJS	-0.0008	0.5812***	0.4188***	MRBK	0.0001	0.5087***	0.4913***
IJT	-0.0002	0.2999***	0.7001***	ESE	-0.0030	0.5031***	0.4969***
IVE	-0.0002	0.2087***	0.7913***	MCY	0.0012	0.3541***	0.6459***
IVV	0.0000	0.1882***	0.8118***	CB	-0.0019	0.2886***	0.7114***
IVW	-0.0002	0.3012***	0.6988***	UB	0.0002	0.4121***	0.5879***
IWB	-0.0002	0.1571***	0.8429***	TMK	0.0006	0.3427***	0.6573***
IWD	-0.0001	0.1496***	0.8504***	BCR	-0.0003	0.2539***	0.7461***
IWF	-0.0002	0.1983***	0.8017***	LH	0.0001	0.3084***	0.6916***
IWM	0.0000	0.3943***	0.6057***	GE	0.0000	0.2183***	0.7817***
IWN	0.0006	0.6368***	0.3632***	CMA	0.0000	0.3012***	0.6988***
IWO	0.0002	0.6281***	0.3719***	ABC	-0.0001	0.3131***	0.6869***
IWP	0.0006	0.2027***	0.7973***	IDXX	0.0000	0.7593***	0.2407***
IWR	-0.0003	0.1530***	0.8470***	WPS	0.0001	0.3100***	0.6900***
IWV	-0.0002	0.6193***	0.3807***	EEP	-0.0002	0.2956***	0.7044***
JKH	0.0005	0.0335***	0.9665***	COKE	-0.0016	0.7122***	0.2878***
JKJ	0.0040	0.0501***	0.9499***	SFSW	0.0042	0.1664***	0.8336***
JKK	0.0013	0.5733***	0.4267***	BF	0.0001	0.3336***	0.6664***
JKL	0.0006	0.0819***	0.9181***	BARI	-0.0018	0.6062***	0.3938***
MDY	0.0000	0.3170***	0.6830***	LEH	0.0011	0.2154***	0.7846***
PEY	0.0003	0.6110***	0.3890***	BFIN	0.0000	0.9715***	0.0285***
PWC	0.0005	0.4124***	0.5876***	KMR	0.0002	0.3925***	0.6075***
PWO	0.0010	0.6553***	0.3447***	CFFN	0.0000	0.9199***	0.0801***
QQQQ	0.0000	0.5782***	0.4218***	INTC	0.0000	0.0415***	0.9585***
VB	0.0003	0.6124***	0.3876***	TRH	-0.0058	0.1543***	0.8457***
VBK	-0.0020	0.2973***	0.7027***	YANB	-0.0011	0.6626***	0.3374***
VBR	0.0018	0.2005***	0.7995***	WSFS	-0.0011	0.5585***	0.4415***
VO	-0.0004	0.3289***	0.6711***	GBL	0.0022	0.3930***	0.6070***
VTI	-0.0002	0.1013***	0.8987***	PDX	0.0036	0.3204***	0.6796***
VTV	-0.0026	0.5112***	0.4888***	ERIE	0.0000	0.5709***	0.4291***
VV	-0.0004	0.2726***	0.7274***	IBOC	-0.0002	0.5970***	0.4030***
VXF	-0.0055	0.4031***	0.5969***	BANF	-0.0011	0.9606***	0.0394***
Ave.	-0.0209	0.3438	0.6562	Ave.	-0.3182	0.4612	0.5388
StdDev	0.4406	0.1954	0.1954	StdDev	1.6886	0.2342	0.2342
Med	-0.0007	0.3012	0.6988	Med	-0.0021	0.3925	0.6075
Sign	0.7493	< 0.0001	< 0.0001	Sign	0.5224	< 0.0001	< 0.0001
SignedRank	0.4499		0.0137				

Panel (b) Industry-wide ETFs and their control securities							
ETFs	Intercept (π_0)	Percentage Order Processing (π_1)	Percentage Adverse Selection (1- π_1)	Controls	Intercept (π_0)	Percentage Order Processing (π_1)	Percentage Adverse Selection (1- π_1)
BDH	0.0059	0.4373***	0.5627***	ACTL	0.0000	0.5890***	0.4110***
BHH	0.0000	0.1992***	0.8008***	QVDX	-0.0001	0.5544***	0.4456***
HHH	0.0000	0.4283***	0.5717***	CERN	-0.0001	0.9057***	0.0943***
IAH	0.0013	0.5037***	0.4963***	PQE	0.0006	0.5135***	0.4865***
IBB	0.0000	0.7049***	0.2951***	IVGN	0.0000	0.0617***	0.9383***
IDU	-0.0005	0.2287***	0.7713***	UIL	-0.0021	0.2720***	0.7280***
IGE	-0.0012	0.3154***	0.6846***	PVA	0.0011	0.2477***	0.7523***
IGM	0.0000	0.5055***	0.4945***	LIFE	0.0009	0.7235***	0.2765***
IGN	0.0003	0.3686***	0.6314***	MRCY	0.0000	0.3602***	0.6398***
IGV	-0.0018	0.3930***	0.6070***	IMN	0.0000	0.3230***	0.6770***
IGW	0.0005	0.4568***	0.5432***	ATK	-0.0014	0.3919***	0.6081***
IIH	-0.0001	0.2481***	0.7519***	AMSWA	-0.0002	0.5820***	0.4180***
IYC	0.0006	0.3570***	0.6430***	BL	0.0004	0.3102***	0.6898***
IYE	0.0006	0.6217***	0.3783***	MIDD	-0.0001	0.8800***	0.1200***
IYG	0.0013	0.4679***	0.5321***	BOKF	0.0001	0.7860***	0.2140***
IYH	0.0004	0.3416***	0.6584***	UTR	-0.0013	0.4045***	0.5955***
IYJ	-0.0015	0.2107***	0.7893***	VTRU	-0.0016	0.8158***	0.1842***
IYK	-0.0007	0.2530***	0.7470***	ROG	0.0004	0.2720***	0.7280***
IYM	-0.0013	0.5466***	0.4534***	CW	-0.0005	0.2368***	0.7632***
IYT	0.0008	0.5170***	0.4830***	FLA	0.0010	0.4024***	0.5976***
IYZ	-0.0001	0.4415***	0.5585***	WFSL	0.0000	0.5691***	0.4309***
OIH	0.0000	0.3669***	0.6331***	IBM	0.0000	0.1329***	0.8671***
RKH	0.0001	0.3145***	0.6855***	ZION	0.0000	0.1708***	0.8292***
RTH	0.0002	0.5239***	0.4761***	TGT	-0.0001	0.4355***	0.5645***
SMH	0.0000	0.2777***	0.7223***	DELL	0.0000	0.2844***	0.7156***
UTH	0.0000	0.8614***	0.1386***	EQT	0.0004	0.3792***	0.6208***
VAW	0.0026	0.2172***	0.7828***	SCL	0.0125	0.3084***	0.6916***
VCR	0.0029	0.5821***	0.4179***	FSCI	-0.0031	0.6581***	0.3419***
VDE	0.0028	0.5638***	0.4362***	TNC	-0.0093	0.3962***	0.6038***
VFH	-0.0067	0.8154***	0.1846***	COBH	0.0005	0.5885***	0.4115***
VPU	0.0000	0.1627***	0.8373***	CHG	-0.0095	0.6638***	0.3362***
WMH	-0.0012	0.3504***	0.6496***	FSCI	-0.0031	0.6581***	0.3419***
XLB	-0.0001	0.6845***	0.3155***	MAS	0.0002	0.4445***	0.5555***
XLE	0.0000	0.2913***	0.7087***	XOM	0.0000	0.3072***	0.6928***
XLK	0.0001	0.3470***	0.6530***	CVC	-0.0002	0.5245***	0.4755***
XLU	0.0001	0.4824***	0.5176***	SO	-0.0001	0.3034***	0.6966***
XLV	0.0001	0.2174***	0.7826***	PFG	-0.0001	0.1399***	0.8601***
XLY	-0.0001	0.7446***	0.2554***	CTL	-0.0001	0.4059***	0.5941***
Ave.	0.0135	0.4124	0.5876	Ave.	-0.0389	0.4429	0.5571
StdDev	0.1784	0.1751	0.1751	StdDev	0.3128	0.2107	0.2107
Med	0.0010	0.3826	0.6174	Med	-0.0007	0.4022	0.5978
Sign	0.6271	< 0.0001	< 0.0001	Sign	0.6271	< 0.0001	< 0.0001
SignedRank	0.2034		0.7815				
Ave.	0.0000	0.3864	0.6136	Ave.	-0.0018	0.4544	0.5456
StdDev	0.0034	0.1907	0.1907	StdDev	0.0122	0.2222	0.2222
Med	0.0000	0.3669	0.6331	Med	0.0000	0.3962	0.6038
Sign	1	< 0.0001	< 0.0001	Sign	0.0847	< 0.0001	< 0.0001
SignedRank	0.1557		0.0395				

adverse selection costs of broad-matched control stocks and industry-matched control stocks are 0.5388 and 0.5571, respectively. Full ETFs have higher percent adverse selection costs than control securities.

In respect of percent adverse selection costs, broad-market ETFs are greater than control securities and industry-wide ETFs are insignificantly greater than control securities. Hence, the George, et al. (1991) model rejects the empirical hypothesis in view of percent adverse selection costs.

Table 5.8: Model Estimates Using Lin, Sanger, and Booth (1995)

The fractional adverse selection cost is computed from the model: $M_{t+1} - M_t = \lambda z_t + \varepsilon_{t+1}$. λ represents the fractional adverse selection cost of the spread, M_t represents the spread midpoint and z_t represents the signed effective half-spread. The error is ε_{t+1}. . Dollar adverse selection cost is percent adverse selection cost times average quoted spread. An ***indicates that a coefficient is significant at a 1% level and ** at a 5% level and * at a 10% level.

Broad ETFs	λ	Control Stocks	λ	Industry ETFs	λ	Control Stocks	λ
DIA	0.7066***	MO	0.4737***	BDH	0.5229***	ACTL	0.5619***
DSG	0.7676***	UTL	0.4797***	BHH	0.7034***	QVDX	0.4128***
DSV	0.4087***	Y	0.7306***	HHH	0.8524***	CERN	0.4913***
ELV	0.2834***	STU	0.7753***	IAH	0.8245***	PQE	0.5483***
IJH	0.8030***	HB	0.4806***	IBB	0.6575***	IVGN	0.4002***
IJJ	0.5248***	CYN	0.7177***	IDU	0.8288***	UIL	0.8110***
IJK	0.7360***	CBSH	0.4660***	IGE	0.5081***	PVA	0.6284***
IJR	0.8415***	NOC	0.7101***	IGM	0.7689***	LIFE	0.6966***
IJS	0.7991***	MRBK	0.4358***	IGN	0.6989***	MRCY	0.2857***
IJT	0.5978***	ESE	0.6887***	IGV	0.5095***	IMN	0.7596***
IVE	0.7025***	MCY	0.6803***	IGW	0.8021***	ATK	0.7448***
IVV	0.8819***	CB	0.5108***	IIH	0.7274***	AMSWA	0.4380***
IVW	0.6205***	UB	0.7143***	IYC	0.9320***	BL	0.6574***
IWB	0.5924***	TMK	0.7659***	IYE	0.6754***	MIDD	0.4530***
IWD	0.5327***	BCR	0.7237***	IYG	0.9311***	BOKF	0.6349***
IWF	0.6323***	LH	0.7162***	IYH	0.7614***	UTR	0.8740***
IWM	0.6109***	GE	0.6529***	IYJ	0.7426***	VTRU	0.6408***
IWN	0.6167***	CMA	0.8377***	IYK	0.6109***	ROG	0.6288***
IWO	0.6190***	ABC	0.8256***	IYM	0.7129***	CW	0.8495***
IWP	0.5924***	IDXX	0.5566***	IYT	0.7522***	FLA	0.6929***
IWR	0.6162***	WPS	0.7877***	IYZ	0.6209***	WFSL	0.4849***
IWV	0.8631***	EEP	0.7112***	OIH	0.5284***	IBM	0.6838***
JKH	0.0989	COKE	0.6174***	RKH	0.8371***	ZION	0.4366***
JKJ	0.3686***	SFSW	-0.1750**	RTH	0.7617***	TGT	0.6489***
JKK	0.1361	BF	0.8155***	SMH	0.5869***	DELL	0.7031***
JKL	0.7754***	BARI	0.5454***	UTH	0.8625***	EQT	0.6769***
MDY	0.8160***	LEH	0.4607***	VAW	0.6628***	SCL	0.6953***
PEY	0.8410***	BFIN	0.4515***	VCR	1.0746***	FSCI	0.5259***
PWC	0.7527***	KMR	0.8729***	VDE	0.9004***	TNC	0.5860***
PWO	0.9703***	CFFN	0.4615***	VFH	0.9140***	COBH	0.4298***
QQQQ	0.5731***	INTC	0.4128***	VPU	0.5867***	CHG	0.6043***
VB	0.3626***	TRH	0.7644***	WMH	0.7319***	FSCI	0.5259***
VBK	0.6478***	YANB	0.6775***	XLB	0.7707***	MAS	0.7322***
VBR	0.7293***	WSFS	0.4173***	XLE	0.5618***	XOM	0.6299***
VO	0.7575***	GBL	0.8445***	XLK	0.6617***	CVC	0.6156***
VTI	0.5591***	PDX	0.6912***	XLU	0.5327***	SO	0.7212***
VTV	0.8470***	ERIE	0.5436***	XLV	0.5097***	PFG	0.7802***
VV	0.8516***	IBOC	0.7388***	XLY	0.6425***	CTL	0.7294***
VXF	0.8785***	BANF	0.5798***				
Ave.	0.6491	Ave.	0.6195	Ave.	0.7176	Ave.	0.6163
StdDev	0.2008	StdDev	0.1904	StdDev	0.1397	StdDev	0.1355
Med	0.6478	Med	0.6803	Med	0.7201	Med	0.6324
Sign	< 0.0001	Sign	< 0.0001	Sign	< 0.0001	Sign	< 0.0001
SignedRank	0.4254			SignedRank	0.0094		
Ave.	0.6829		0.6179				
StdDev	0.1756		0.1645				
Med	0.7034		0.6489				
Sign	< 0.0001		< 0.0001				
SignedRank	0.0221						

5.6.2.4 Lin, Sanger, and Booth (1995)

In the Lin, et al (1995) model, the percentage adverse selection cost is directly estimated from the function $Q_{t+1} - Q_t = \lambda z_t + \varepsilon_{t+1}$. The λ is the estimated percentage adverse selection cost. (For the details of the Lin et al (1991) model, see section 4.1.3)

Table 5.8 shows summary statistics of ETFs and matched control stocks. The average percent adverse selection costs of ETFs are 0.6491 and 0.7176, in the order of broad-market ETFs and industry-wide ETFs. Meanwhile, their matched control stocks have 0.6195 and 0.6163 percent adverse selection cost, respectively. The Wilcoxon signed rank sum test of percent adverse selection cost suggests that broad-market ETFs and industry-wide ETFs have significantly smaller percent adverse selection costs than their matched control securities.

5.6.2.5 Huang and Stoll (1997)

The Huang and Stoll (1997) model is based on the generalised method of moments. The population moments are

$$E(Q_{t-1}|Q_{t-2}) = (1 - 2\pi)Q_{t-2} \tag{5.23}$$

$$\Delta M_t = (\alpha + \beta)\frac{S_{t-1}}{2}Q_{t-1} - \alpha(1 - 2\pi)\frac{S_{t-2}}{2}Q_{t-2} + \varepsilon_t \tag{5.24}$$

From these population moments, I estimate parameter vector (α, β, π). The parameter α represents the percent adverse selection cost. (For the details of the Huang and Stoll (1997) model, see section 4.1.5)

Table 5.9 reports the summary statistics of ETFs and matched control stocks. Broad-market ETFs (industry-wide ETFs) have, on average, 0.4151 (0.4018) percent adverse selection cost. Interestingly, matched control securities show negative adverse selection costs, -0.0143 and -0.0149 respectively. These negative adverse selection costs are common problem in estimating the Huang and Stoll (1997) model. Even Huang and Stoll (1997) show negative adverse selection costs when they estimate adverse selection costs in their paper.

Table 5.9: Model Estimates Using Huang and Stoll (1997)

Panel (a) Broad-Market ETFs and their control securities

The estimated moment conditions are $E(Q_{t-1}|Q_{t-2}) = (1-2\pi)Q_{t-2}$ and $\Delta M_t = 0.5(\alpha+\beta)S_{t-1}Q_{t-1} - 0.5\alpha(1-2\pi)Q_{t-2} + \varepsilon_t$. α is the proportion of the half spread attributable to the adverse selection costs and β is the inventory cost. M_t is the midpoint and Q_t the transaction type . At time t order processing costs are $(1-\alpha-\beta)$. An ***indicates that a coefficient is significant at a 1% level and ** at a 5% level and * at a 10% level. Dollar adverse selection cost is percent adverse selection cost times average quoted spread.

ETFs	Adverse Selection (α)	Inventory (β)	Controls	Adverse Selection (α)	Inventory (β)
DIA	0.3920***	0.1978***	MO	-0.0211***	0.1354***
DSG	0.4105***	0.1954***	UTL	-0.0192	0.1232***
DSV	0.4209***	0.1825**	Y	-0.0336***	0.1470***
ELV	0.4054***	0.1933**	STU	-0.0031	0.0891***
IJH	0.4537***	0.0894***	HB	-0.0142	0.1466***
IJJ	0.4578***	0.0826***	CYN	-0.0003	0.1482***
IJK	0.4704***	0.0771***	CBSH	-0.0029	0.1597***
IJR	0.4535***	0.0863***	NOC	-0.0280***	0.1450***
IJS	0.4634***	0.0709***	MRBK	-0.0276***	0.1475***
IJT	0.4712***	0.0662***	ESE	-0.0044	0.1438***
IVE	0.4535***	0.0790***	MCY	-0.0083	0.1134***
IVV	0.4311***	0.1087***	CB	0.0160***	0.1248***
IVW	0.4389***	0.0893***	UB	-0.0381***	0.1566***
IWB	0.4400***	0.0740***	TMK	-0.0186***	0.1171***
IWD	0.4466***	0.0649***	BCR	0.0073**	0.1467***
IWF	0.4463***	0.0513***	LH	-0.0066	0.1161***
IWM	0.3931***	0.1406***	GE	-0.0178***	0.1549***
IWN	0.3824***	0.1497***	CMA	-0.0018	0.1496***
IWO	0.3762***	0.1553***	ABC	-0.0747***	0.2674***
IWP	0.3843***	0.1184***	IDXX	-0.0286***	0.1719***
IWR	0.3997***	0.0766***	WPS	-0.0294***	0.1405***
IWV	0.4022***	0.0676***	EEP	0.0050	0.1332***
JKH	0.3431***	0.2157*	COKE	0.0047	0.1413***
JKJ	0.2960***	0.3076**	SFSW	0.0035	0.0734
JKK	0.4080***	0.1015	BF	0.0225***	0.1164***
JKL	0.3944***	0.1141**	BARI	-0.0066	0.1744***
MDY	0.3592***	0.1846***	LEH	-0.0105***	0.1243***
PEY	0.3594***	0.1652**	BFIN	0.0317**	0.0996***
PWC	0.3880***	0.1299***	KMR	-0.0169	0.1387***
PWO	0.3748***	0.1396***	CFFN	-0.0104	0.1705***
QQQQ	0.3746***	0.1329***	INTC	-0.0254***	0.1604***
VB	0.4268***	0.0513	TRH	-0.0070	0.0955***
VBK	0.4296***	0.0492	YANB	-0.0564***	0.1915***
VBR	0.4330***	0.0870**	WSFS	-0.0376***	0.1563***
VO	0.4375***	0.0492*	GBL	-0.0368	0.1991**
VTI	0.4543***	0.0420***	PDX	-0.0257**	0.1392***
VTV	0.4356***	0.0753*	ERIE	-0.0025	0.1421***
VV	0.4328***	0.0890**	IBOC	-0.0094	0.1357***
VXF	0.4478***	0.0670	BANF	-0.0263	0.2042***
Average	0.4151	0.1133	Average	-0.0143	0.1446
StdDev	0.0390	0.0583	StdDev	0.0203	0.0343
Median	0.4268	0.0893	Median	-0.0105	0.1438
Sign	< 0.0001	< 0.0001	Sign	< 0.0001	< 0.0001
Signed Rank	< 0.0001	0.0049			

Panel (b) Industry-wide ETFs and their control securities					
ETFs	Adverse Selection (α)	Inventory (β)	Controls	Adverse Selection (α)	Inventory (β)
BDH	0.3808***	0.1949***	ACTL	-0.0567***	0.2427***
BHH	0.4258***	0.1388***	QVDX	-0.0113	0.1089***
HHH	0.4299***	0.1421***	CERN	-0.0022	0.1569***
IAH	0.4256***	0.1361***	PQE	-0.0246*	0.1354***
IBB	0.4326***	0.1342***	IVGN	-0.0275***	0.1634***
IDU	0.4324***	0.1255***	UIL	-0.0333***	0.1496***
IGE	0.4336***	0.1070***	PVA	-0.0074	0.1013***
IGM	0.4313***	0.1176***	LIFE	-0.0139***	0.1296***
IGN	0.4363***	0.1106***	MRCY	-0.0272***	0.1469***
IGV	0.4188***	0.1435***	IMN	-0.0296***	0.1726***
IGW	0.4153***	0.1412***	ATK	-0.0200*	0.1917***
IIH	0.4405***	0.1123***	AMSWA	-0.0178**	0.1867***
IYC	0.3953***	0.0870**	BL	0.0429***	0.0747***
IYE	0.3751***	0.1269***	MIDD	-0.0295***	0.1545***
IYG	0.3604***	0.1655***	BOKF	0.0175**	0.1238***
IYH	0.3687***	0.1268***	UTR	-0.0262***	0.1367***
IYJ	0.3660***	0.1349***	VTRU	-0.0368***	0.1571***
IYK	0.3901***	0.1252***	ROG	-0.0011	0.0899***
IYM	0.3990***	0.1063***	CW	0.0080	0.1293***
IYT	0.4011***	0.0890***	FLA	-0.0224	0.1742***
IYZ	0.3983***	0.0948***	WFSL	-0.0395***	0.1615***
OIH	0.3574***	0.1683***	IBM	0.0000	0.1178***
RKH	0.3777***	0.1268***	ZION	-0.0641***	0.2067***
RTH	0.4011***	0.1187***	TGT	-0.0217***	0.1244***
SMH	0.3948***	0.1178***	DELL	-0.0045	0.1514***
UTH	0.3768***	0.1132***	EQT	0.0034	0.1343***
VAW	0.4094***	0.0641	SCL	-0.0212	0.1284*
VCR	0.4468***	0.0467	FSCI	-0.0225	0.1742***
VDE	0.4233***	0.0778***	TNC	-0.0165	0.1131***
VFH	0.4064***	0.1026	COBH	0.0197*	0.1125***
VPU	0.4443***	0.0636*	CHG	0.0030	0.1438***
WMH	0.4465***	0.0540**	FSCI	-0.0225	0.1742***
XLB	0.3881***	0.1524***	MAS	-0.0162***	0.1314***
XLE	0.3757***	0.1748***	XOM	-0.0324***	0.1435***
XLK	0.3707***	0.1801***	CVC	0.0106*	0.1276***
XLU	0.3627***	0.1867***	SO	-0.0165***	0.1131***
XLV	0.3627***	0.1708***	PFG	-0.0153***	0.1188***
XLY	0.3668***	0.1558***	CTL	0.0101**	0.1310***
Average	0.4018	0.1246	Average	-0.0149	0.1430
StdDev	0.0285	0.0365	StdDev	0.0208	0.0328
Median	0.4001	0.1261	Median	-0.0172	0.1360
Sign	< 0.0001	< 0.0001	Sign	0.0017	< 0.0001
Signed Rank	< 0.0001	0.0429			
Average	0.4085	0.1189	Average	-0.0146	0.1438
StdDev	0.0347	0.0488	StdDev	0.0204	0.0333
Median	0.4080	0.1176	Median	-0.0165	0.1421
Sign	< 0.0001	< 0.0001	Sign		
Signed Rank	< 0.0001	0.0001			

The negative adverse selection costs are estimated commonly. Hatch and Johnson (2002) also find many economically unreasonable adverse selection costs when they use the Huang and Stoll (1997) model. Thus, they employ the Madhavan, et al. (1997) model to estimate adverse selection costs. They conjecture that the negative adverse selection costs appear as trading frequency and persistence in trading type increase.

Meanwhile, the Huang and Stoll (1997) model rejects the empirical hypothesis in terms of percent adverse selection costs. The Wilcoxon signed rank sum tests demonstrate that there are clear difference between ETFs and their matched control securities.

5.6.2.6 Discussion

The Empirical Hypothesis So far only the Glosten and Harris (1988) model and the Madhavan, et al. (1997) model provide estimates of the adverse selection component associated with ETFs that are significantly lower than for the control stocks.

Using the Glosten and Harris (1988) model average adverse selection costs across all ETFs are numerically positive but statistically insignificant. However, mean costs associated with control stocks are in the region of 30%, and are therefore comparable to those identified by Glosten and Harris (1988) and Neal and Wheatley (1998). The Wilcoxon signed rank test indicates that the costs associated with ETFs are lower than those of the control stocks. The results obtained using the Madhavan, et al. (1997) model suggest that both ETFs and the control securities have higher estimated costs than the Glosten and Harris (1988) model. Using the Madhavan, et al. (1997) model adverse selection costs across all ETFs are 0.2083 but are 0.5764 for the control stocks. Although both costs are significantly positive, those associated with ETFs are significantly lower than for the control securities.

The other three spread decomposition models indicate much higher adverse selection costs associated with the ETFs and suggest that ETFs do not have lower costs than the control securities. The George, et al. (1991) model estimates mean adverse selection costs to be 0.6136 for the ETFs in comparison to 0.5464 for the control stocks (the estimates for the control securities are comparable to Neal and Wheatley (1998)'s estimates for individual securities). For the George, et al. (1991) model, both full ETFs and the control securities have positive adverse selection costs which are statistically different from each other while the Huang and Stoll (1997) model shows that ETF's have significantly higher adverse selection costs than control firms. The Lin, et al (1995) model suggests that mean adverse selection costs associated with the ETFs are statistically higher when compared to the control firms.

When we examine the broad-market ETFs and the industry-wide ETFs separately, this pattern is repeated for both groups. The Glosten and Harris (1988) and the Madhavan, et al. (1997) models suggest that both broad-market and industry-wide ETFs have lower mean costs than the control stocks. The George, et al. (1991) model suggests that broad-market ETFs have higher adverse selection costs than the control firms, while the Lin, et al. (1995) model shows that industry-wide ETFs have higher adverse selection costs than the control securities. The Huang and Stoll (1997) model shows clearly that ETFs have higher adverse selection costs than control securities.

5.7 Robustness Test

5.7.1 Exchange Test

In the US there are several stock exchanges such as New York Stock Exchange (NYSE), American Stock Exchange (AMEX), NASDAQ, and regional exchanges (Pacific, Boston, Philadelphia, and CBOE). Even though I did not consider where each trade occurs, some studies[1] suggest that different exchanges have different transaction cost due to different trading procedure. For instance, the NYSE and AMEX adopt a specialist system whereas the NASDAQ employs dealer systems. In the NYSE and AMEX, one specialist is assigned to one share, while in the NASDAQ several dealers compete for one share. Due to this different trading system, Huang and Stoll (1996) find the NASDAQ to have higher transaction

[1] This is discussed in Chapter Seven

costs than the NYSE. Thus, I analyse whether this different trading exchange would affect the adverse selection of each model.

The exchange trading mechanism has been shown to exert an important influence over liquidity and to also influence the magnitude of the information asymmetry between traders. Huang and Stoll (1996), in their cross-market analysis of liquidity, showed that bid-ask spreads for stocks with the same characteristics were higher on the NASDAQ dealer market than on the NYSE auction market. Madhavan (1992) showed how the information aggregation process of a call auction can reduce market failures caused by information asymmetry. Biais, Hillion, and Spatt (1999) and Cao, Ghysels, and Hatheway (2000) show that the pre-opening period prior to the execution of a call auction is important for price discovery as traders learn from indicative prices disseminated by the exchange. Moreover, Schnitzlein (1996) experimental study indicates that informed traders earn smaller profits in call markets and are less able to exploit uninformed traders, suggesting that lower information asymmetries are associated with this trading system.

The ETFs in my sample trade on a range of markets and for most of them trading takes place on more than one market. Since the trading mechanism of an exchange can influence the information asymmetry and, therefore, the adverse selection costs present, it is necessary to demonstrate that my results are not driven by exchange-related effects, and hence, these are controlled by the following cross-section model that regresses the adverse selection estimates against a set of control variables.

$$ASC_i = \alpha + \delta_1 ASCDUM + \beta Vol_i + \delta_2 AMEX + \delta_3 NASDAQ + \delta_4 Pacific + e_i \quad (5.25)$$

where ASC is the estimated adverse selection cost of security i (both ETFs and control securities). ASCDUM has a value of unity if the adverse selection cost is for a control stock but has a zero otherwise. The coefficient δ_1 captures the extent to which the adverse selection cost of the control securities is above or below that of the ETFs. The value of δ_1 will be positive if ETFs have lower adverse selection costs. Vol is the average daily trading volume of each security and will capture the effect that volume has on the adverse selection estimates.

In this model I also control for the exchange. I examine each trade associated with the ETF's and the control securities and identify the location of each trade. I then identify the exchange on which most of the trading is undertaken. If AMEX is the main exchange for trading I signify the security with unity but a zero otherwise. If the NASDAQ is the main exchange for trading I signify the security with unity but a zero otherwise. If the Pacific exchange is the main exchange for trading I signify this as unity for that security but a zero otherwise. Although both types of securities also trade on the NYSE, CBOE, Philadelphia and Boston exchanges, none of these is a main market where the majority of trading takes place. A security that mainly trades on a market outside those listed is classified as other and is reflected in α. The coefficients δ_i therefore captures the impact that the dominant exchange has on the adverse selection costs of the two types of securities. The results from the estimation of this model are contained in the panel (a) of Table 5.10.

It is also possible that the market that is not dominant could influence the estimated adverse selection costs. To capture this possibility I undertake the following regression which captures the proportion of trading taking place on each of the following exchanges, NYSE, AMEX, NASDAQ, Boston, Pacific, and Philadelphia.

$$ASC_i = \begin{array}{c} \alpha + \delta ASCDUM + \beta Vol_i + \gamma_1 NYSE + \gamma_2 AMEX \\ +\gamma_3 NASDAQ + \gamma_4 Boston + \gamma_5 Pacific + \gamma_6 Phil + e_i \end{array} \tag{5.26}$$

where the coefficients γ_1 to γ_6 capture the impact that the intensity of trading on the different exchanges has on the adverse selection costs of the ETF's and control securities. The α is the constant. The e_i is the cross-sectional error. I undertake the regression five times using the adverse selection cost estimates of the five models. The results from the estimation of this model are contained in the panel (b) of Table 5.10.

In the regression for the Madhavan, et al. (1997) model, an ETF 'ELV' and its matched control security 'STU' is removed because STU has unreasonably higher percent adverse selection costs, -548.98{

Table 5.10: Controlling for the Stock Exchange

Panel (a) Dummy variable Model

In this table we report estimates from the following model.

$ASC_i = \alpha + \delta_1 ASCDUM + \beta Vol_i + \delta_2 AMEX + \delta_3 NASDAQ + \partial_4 Pacific + e_i$

where, ASC is the estimated adverse selection costs, ASCDUM has a value of unity if the security is a control but has a zero otherwise, Vol is the average daily trading volume of security i, AMEX, NASDAQ and Pacific have a value of 1 if security i mainly trades on this market but has a value of zero otherwise. The cross-sectional error is e_i and α, β and δ_j are coefficients.
GH stands for Glosten and Harris (1988), GKN for George, et al. (1991), LSB for Lin, et al. (1995), MRR for Madhavan, et al. (1997), and HS for Huang and Stoll (1997)
The value in parenthesis is t-value, and *** is for 1% significance level, ** for 5% significance level, and * for 10% significance level.

	GH	MRR	GKN	LSB	HS
Intercept α	-0.3212 (-1.66)	0.7898** (2.04)	-0.0614 (-0.22)	0.3218 (1.36)	0.4052*** (10.66)
ASCDUM δ_1	**0.3697***** **(8.71)**	**0.4318***** **(5.15)**	**-0.0588** **(-0.95)**	**-0.0043** **(-0.08)**	**-0.4211***** **(-50.62)**
Volume β	0.0316 (1.67)	-0.0591 (-1.57)	0.0675** (2.45)	0.0340 (1.48)	-0.0023 (-0.63)
AMEX δ_2	-0.0107 (-0.16)	-0.0149 (-0.11)	0.0440 (0.45)	0.0539 (0.66)	0.0351*** (2.67)
NASDAQ δ_3	-0.0333 (-0.50)	-0.0348 (-0.26)	0.0598 (0.61)	-0.0090 (-0.11)	0.0229* (1.75)
PACIFIC δ_4	0.1675** (2.34)	0.0532 (0.38)	-0.0792 (-0.76)	-0.0108 (-0.12)	0.0126 (0.90)
Adjusted R-squared	0.5531	0.3953	0.0537	0.0271	0.9837

Panel (b) model with percent variables

In this table we report estimates from the following model.

$ASC_{ik} = \alpha + \delta ASCDUM + \beta Vol_i + \gamma_1 NYSE + \gamma_2 AMEX + \gamma_3 NASDAQ + \gamma_4 Boston + \gamma_5 Pacific + \gamma_6 Phil + e_i$,

where the variables NYSE, AMEX, NASDAQ, Boston, Pacific, and Philadelphia are variables that capture the proportion of trading undertaken on these exchanges. The α, β and δ's are coefficients Each exchange variable represents the fraction of trading in each exchange. For example, NYSE variable represents the fraction of trading occurred in the NYSE. ASCDUM variable is dummy variable where is 0 in case of ETF and 1 in case of control stock. Volume variable is average log volume during the sample period. The value in parenthesis is t-value, and *** is for 1% significance level, ** for 5% significance level, and * for 10% significance level.

	GH	MRR	GKN	LSB	HS
Intercept α	-0.3957* (-1.91)	0.5033 (1.22)	0.1587 (0.54)	0.3935 (1.62)	0.3718*** (9.03)
ASCDUM δ	**0.3712***** **(8.38)**	**0.5259***** **(6.06)**	**-0.1512**** **(-2.41)**	**-0.0223** **(-0.43)**	**-0.4004***** **(-45.48)**
Volume β	0.0402** (2.10)	-0.0111 (-0.29)	0.0432 (0.1124)	0.0257 (1.15)	-0.0014 (-0.37)
NYSE γ_1	0.1585 (1.15)	-0.4799* (-1.78)	0.8574*** (4.39)	0.7917*** (4.89)	0.0065 (0.24)
AMEX γ_2	-0.0221 (-0.20)	-0.1394 (0.63)	0.0809 (0.51)	0.2160 (1.64)	0.0879*** (3.94)
NASDAQ γ_3	-0.2433* (-1.77)	-0.2624 (-0.97)	01785 (0.92)	-0.1317 (-0.82)	0.0331 (1.22)
Boston γ_4	-1.1037 (-1.25)	-0.5229 (-0.30)	1.9343 (1.55)	0.7866 (0.76)	-0.0218 (-0.12)
Pacific γ_5	0.3711** (2.19)	-0.2495 (-0.75)	-0.1963 (-0.82)	-0.0577 (-0.29)	0.0375 (1.11)
Philadelphia γ_6	20.0180 (0.62)	-199.9802 (-3.18)	-33.1528 (-0.73)	8.3488 (0.22)	-1.3868 (-0.22)
Adjusted R-squared	0.5865	0.4497	0.1764	0.1713	0.9845

Panel (a) of Table 5.10 shows that after controlling for the dominant market, consistent results occur. Using the Glosten-Harris estimates of the adverse selection costs for the regression Table 5.10 shows that the Pacific exchange appears to have a positive effect on adverse selection costs but the coefficient associated with δ_1 shows that even after controlling for these effects, the ETF's have significantly lower adverse selection costs than their controls. For the

Madhavan, et al. (1997) model none of the exchange dummies is significant but δ_1 is insignificantly positive suggesting that ETF's have lower adverse selection costs. For the other three models ETFs are not shown to have lower adverse selection costs. For the Huang and Stoll (1997) model the δ_1 is found to be significant confirming that using this model ETFs have higher adverse selection costs than their controls. The George, et al. (1991) and Lin, et al. (1995) models show that the ETF and control security adverse selection costs are not different from each other.

Panel (b) of Table 5.10 shows that the proportion of trading on an exchange has a stronger impact on the adverse selection estimates. For the Glosten and Harris (1988) model the γ's associated with the NASDAQ and the Pacific are significant. The NASDAQ has a negative effect on the adverse selection estimates while the Pacific increases adverse selection costs. However, the controls are still shown to have higher adverse selection costs. With the Madhavan, et al. (1997) model the ETFs have lower adverse selection costs. For the George, et al. (1991) and the Huang and Stoll (1997) model ETFs continue to have higher adverse selection costs. The Lin, et al. (1995) model continues to provide insignificantly different adverse selection costs for the controls and the ETFs.

The percent adverse selection costs test results show that even after controlling for the exchange the Glosten-Harris model and the Madhavan, et al. (1997) models show that ETFs have lower adverse selection costs than their controls.

5.7.2 Regression Tests

This section provides another robustness result of the empirical hypothesis test. This study employs a control matching method to test the empirical hypothesis, but there is a possible concern with the matching procedure, that being selection bias. i.e. matching ETFs with wrongly selected control securities.

However, three variables (trading volume, trading price, and volatility) used in the matching procedure are commonly acknowledged as affecting adverse selection costs. Thus, the percent adverse selection costs of five spread decomposition models are regressed on these three variables and a control dummy variable is also included in the regression model, which is represented by

$$ASC = \alpha + \beta_1 \text{Con\_Dum} + \beta_2 \text{Vol} + \beta_3 \text{Ret\_Vol} + \beta_4 \text{Price} + \varepsilon \tag{5.27}$$

ASC represents the percent adverse selection cost of spread decomposition models. Con_Dum variable is one in the case of control security and zero in the case of ETF.

Table 5.11 reports the regression results based on each model. The Glosten and Harris (1988) model and the Madhavan, et al. (1997) model report significantly positive coefficients of the Con_Dum variable. Meanwhile the George, et al. (1991) model and the Lin, et al. (1995) model have insignificant β_1. Additionally the Huang and Stoll (1997) model reports significantly negative coefficient β_1.

The other variables such as Volume, Return volatility and Price report that most coefficients are insignificant, implying that these variables do not discern ETFs and control securities.

Table 5.11: Regression Results Using Control Variables

	GH	MRR	GKN	LSB	HS
Full ETF					
	Parameter	Parameter	Parameter	Parameter	Parameter
Interecept α	-0.0640	0.1563	0.6316***	0.7156***	0.4170***
Con_Dum β_1	0.2774***	0.4351***	-0.0273	-0.0340	-0.4208***
Volume β_2	0.0548	-0.0757	0.2194	-0.1799	-0.0497*
Return Volatility β_3	4.3240	6.1120	-9.2831*	-6.2517	-0.7168
Price β_4	0.0007**	0.0000	0.0008*	0.0002	0.0000
Broad-market ETFs					
	Parameter	Parameter	Parameter	Parameter	Parameter
Interecept α	-0.0405	0.0959	0.7030***	0.6969***	0.4316***
Con_Dum β_1	0.3183***	0.4972***	-0.0441	0.0167	-0.4236***
Volume β_2	0.0628	-0.1541	0.1521	-0.1276	-0.0469
Return Volatility β_3	-2.4732	16.0644	-20.4445	-10.5832	-2.0118
Price β_4	0.0007	-0.0005	0.0012**	0.0003	0.0000
Industry-wide ETFs					
	Parameter	Parameter	Parameter	Parameter	Parameter
Interecept α	-0.0634	0.1157	0.6515***	0.7538***	0.4082***
Con_Dum β_1	0.2654***	0.3173***	0.0129	-0.0449	-0.4190***
Volume β_2	0.2187	0.1913	1.1256*	-0.2755	-0.0852
Return Volatility β_3	3.1743	12.0932	-9.5867	-10.5737**	0.1360
Price β_4	0.0013**	0.0001	-0.0001	0.0012**	-0.0001

5.8 Concluding Remarks

The objective of this chapter is to evaluate the performance of spread decomposition models with respect to the adverse selection costs. Evaluating spread decomposition models requires setting up the empirical hypothesis that Exchange-Traded Funds (ETFs) have lower adverse selection costs than control stocks. Furthermore, I examine how much adverse selection costs spread decomposition models provide depending on broad-market ETFs, industry-wide ETFs and control securities. Specifically, I investigate whether broad-market ETFs have zero adverse selection costs, industry-wide ETFs have close-to-zero adverse selection costs, and matched control securities provide non-zero adverse selection costs according to spread decomposition models.

The hypothesis that ETFs have lower adverse selection costs than matched control securities is supported by Subrahmanyam (1991) and Gorton and Pennacchi (1993), in which they suggest the information asymmetry in the basket securities are diversified away. Specifically both papers show that the information asymmetry in the basket security is lower than the weighted average information asymmetry of the constituent securities. This happens because the information asymmetry of constituent securities is crossed out each other. Additionally, the features of ETFs support that ETFs have lower information asymmetry. The creation / redemption, transparency etc make ETFs a proper tool to evaluate spread decomposition models.

Five commonly-used spread decomposition models are tested: Glosten and Harris (1988), Madhavan et al. (1997), George et al. (1991), Lin et al. (1995), and Huang and Stoll (1997). Based on the percent adverse selection costs, the Glosten and Harris (1988) model and the Madhavan, et al. (1997) model satisfy the hypotheses, implying that these two models provide the lower adverse selection costs of ETFs than control securities. The other three spread decomposition models, however, show the higher adverse selection costs of ETFs compared to control securities. These differences between ETFs and control securities are confirmed by non-parametric comparison.

I check the hypothesis test result with two regression models. The first regression model employs exchange variables to consider the impact of exchange on the adverse selection cost. The second model is based on the control procedure to check whether or not the control procedure affects the adverse selection cost. These two regression results do not change the result of the hypothesis test in that the Glosten and Harris (1988) model and the Madhavan, et al. (1997) model provide lower adverse selection costs than control securities.

Subrahmanyam (1991) and Gorton and Pennacchi (1993) and the characteristics of ETFs imply that broad-market ETFs have zero adverse selection costs, industry-wide ETFs close-to-zero adverse selection costs, control stocks non-zero adverse selection costs. This is examined as an empirical question.

A Appendix 5.A

Table 5.A.1: The Correlation between ETFs and the S&P 500 Index

Symbol	Correlation Coefficient	Symbol	Correlation Coefficient
DIA	0.9316	BDH	0.6571
DSG	0.8281	BHH	0.3855
DSV	0.8030	HHH	0.5453
ELV	0.8921	IAH	0.6890
IJH	0.8896	IBB	0.6890
IJJ	0.8985	IDU	0.7183
IJK	0.8803	IGE	0.5586
IJR	0.8832	IGM	0.8158
IJS	0.8807	IGN	0.6915
IJT	0.8752	IGV	0.7144
IVE	0.9614	IGW	0.6835
IVV	0.9761	IIH	0.5061
IVW	0.9631	IYC	0.8804
IWB	0.9773	IYE	0.5669
IWD	0.9425	IYG	0.8449
IWF	0.9416	IYH	0.7564
IWM	0.8836	IYJ	0.9073
IWN	0.8896	IYK	0.8686
IWO	0.8731	IYM	0.7812
IWP	0.9072	IYT	0.7609
IWR	0.9067	IYZ	0.7032
IWV	0.9760	OIH	0.4673
JKH	0.8389	RKH	0.8105
JKJ	0.8766	RTH	0.7777
JKK	0.8275	SMH	0.6390
JKL	0.8789	UTH	0.6895
MDY	0.8858	VAW	0.7623
PEY	0.8811	VCR	0.7916
PWC	0.9203	VDE	0.5479
PWO	0.8462	VFH	0.8094
QQQQ	0.8644	VPU	0.7245
VB	0.8745	WMH	0.7380
VBK	0.8429	XLB	0.7828
VBR	0.8710	XLE	0.5517
VO	0.9018	XLK	0.8285
VTI	0.9774	XLU	0.6989
VTV	0.9530	XLV	0.7499
VV	0.9580	XLY	0.8833
VXF	0.8812		
Average	0.8985	Average	0.7099
Std. Dev.	0.0458	Std. Dev.	0.1226

Chapter 6

Using The PIN to Measure Information Asymmetry

6.1 Introduction

This chapter incorporates an alternative measure of information asymmetry which is called the probability of informed trading (PIN). Easley, Kiefer, O'Hara, and Paperman (1996) developed a measure for the probability of trading caused by privately-informed traders. This measure is complementary to the adverse selection cost of bid-ask spread. The potential existence of informed traders causes the adverse selection cost in the bid-ask spread (see Glosten and Milgrom (1985) etc in Chapter Three). Therefore, the adverse selection cost can be measured in dollar terms since the adverse selection cost is a component of bid-ask spread[1]. Easley, et al. (1996) estimate the existence of informed traders as the probability of informed trading. That is, the probability of informed trading measures the chance that a trade is initiated by informed traders. The probability of informed trading is computed using the actual number of buy (sell) trades.

As a complementary measure to the adverse selection cost, this probability of informed trading is employed in various empirical studies. Section 6.2 displays some examples that use the probability of informed trading as a methodological tool.

[1]See Chapter Five. The models include George, Kaul , and Nimalenran (1991), Lin, Sanger, and Booth (1995), Huang and Stoll (1997), and Madhavan, Richardson, and Roomans (1997).

This chapter investigates whether ETFs have a lower probability of informed trading than control stocks. Since the probability of informed trading is parallel to the adverse selection costs in terms of information asymmetry in the trading process, ETFs should have a lower probability of informed trading than control stocks. This is the hypothesis I seek to examine in this chapter.

6.2 Empirical Use of the PIN

Easley, et al. (1996) developed the PIN measure using the Bayesian approach[1]. Since then many papers have employed the PIN measure to address a variety of issues, such as informed trading (Heidle and Huang (2002); Fuller (2003)), equity carve-outs (Fu (2002)), post-earning announcement drift (Vega (2006)), and accrual profits[2] (Liu and Qi (2006)).

Heidle and Huang (2002) examine whether the PIN reflects possible differences in the level of information asymmetry that exist in auction and dealer markets, and whether it is reflected in the bid-ask spread. They found that the PIN increases when shares move from being listed on the NASDAQ to either being listed on the NYSE or the AMEX. It was found that the PIN decreases with the move from the AMEX to the NASDAQ and the PIN is unchanged when firms change their listing from the AMEX to the NYSE. They also find that the change in the PIN is related to the change in the bid-ask spread when shares are relocated from one market to another.

Fuller (2003) also employs the PIN to test whether the magnitude of a price reaction to a dividend signal is related to the amount of informed trading. Fuller finds that there is an inverse relationship between the market reaction to a dividend signal and the amount of informed trading.

[1] In their model, a market-maker updates his belief by observing the order flow. At the beginning of trading, the market-maker does not know whether or not an information event occurs and whether that information event is either good or bad. The market-maker changes his belief by observing coming order type and frequency. This process is based on Bayesian inference.

[2] Accrual profits come from accrual basis accounting. In accrual basis accounting, profits are recorded and recognised when revenues incur and are matched with related expenses in a process known as matching or expense matching.

Fu (2002) investigates how an equity carve-out[1] affects the information asymmetry of investors in the parent company. By using the PIN as a measure of information asymmetry, Fu finds that after an equity carve-out, the average PIN of the parent company is significantly diminished. Additionally, when the carved-out division is in a different industry from the parent company, the decrease in the PIN is more pronounced.

Vega (2006) examines how public/private information affects the stock price drift after an earnings announcement[2]. The PIN, public news surprises, and the number of days a particular firm is mentioned in the news, are used to differentiate between the type of information (public or private) and the arrival rate of traders (informed/uninformed). Vega finds that as the days of the news coverage increase, stocks display significant post-earnings announcement drift. When the PIN is high, the stock experiences a low or insignificant drift. In addition to this, the PIN is related to public news surprises displaying insignificant post-earnings announcement drift. This is interpreted as showing that post-earnings announcement drifts are small when more information about the true value of an asset is released, either privately or publicly, and when investors agree and trade on this information.

Liu and Qi (2006) use the PIN to determine whether informed investors can exploit possible arbitrage opportunities experienced by a high-accrual companies through having lower abnormal returns than lower accrual companies. They find that the returns to the accruals-based strategy are larger when stocks with higher PIN are traded. The results are robust even when they consider transaction costs affecting accruals-based strategies.

These papers make a small example of how to use the PIN measure in the financial research. All these papers use the PIN as an information asymmetry measure, while no paper questions whether the PIN properly capture information

[1]Equity carve-out means initial public sale of a minority position (20% or less, typically) in the stock of a subsidiary or division by a larger corporation. The parent company will often spin off the remaining shares to its existing stockholders at a future date.

[2]In the financial economic literature, stock price is assumed to reflect all available information. However, empirical finding suggests that stock price reacts slowly to an information event. This phenomenon is called stock price drift.

asymmetry. This chapter will provide an evidence of this question. The concerns with the PIN are discussed in later section (Section 6.4).

6.3 The Basic PIN Model

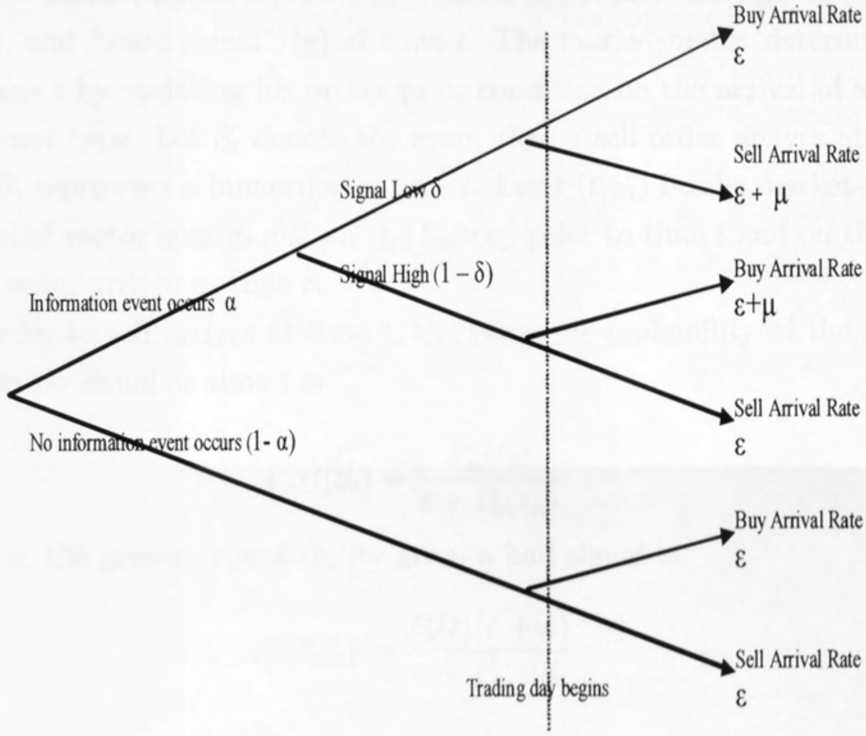
6.3.1 Setup

In the model suggested by Easley, et al. (1996), Nature determines whether an information event will occur and which type of information event it will be. Information events are independently distributed and the probability of the occurrence of such an event is α. i.e. the probability that the information event occurs is α, and the probability that the information event does not occur is $1-\alpha$. When an information event occurs, the probability that the information is good is $1-\delta$, and the probability that the information is a bad signal for the securities is δ. The occurrence and type of information event is determined before the first trade takes place. Thus, there are three types of information events (no news, a positive, or negative information event).

$(V_i)_{i=1}^{I}$ is a random variable representing the value of the asset at the end of trading days $i = 1, \ldots, I$. The value of the asset conditional on a good signal on day i is $\overline{V}_i$; similarly it is $\underline{V}_i$ conditional on a bad signal on day i. The value of the asset if no signal arises on day i is denoted as V_i^*. It is assumed that $\underline{V}_i < V_i^* < \overline{V}_i$. On any day, informed and uninformed traders arrive on the exchange at different rates[1], but each arrival rate of informed (uninformed) traders follows an independent Poisson process. The arrival rates of informed traders are equal to μ. Similarly, uninformed buyers and sellers follow an arrival rate ε. All informed traders are risk-neutral and competitive.

The tree diagram in Figure 6.1 illustrates the probability structure of the trading process. Nature determines the occurrence of an information event at the first node of the tree with probability α, and then nature determines the type of information event with probability δ. These decisions occur before trading starts. For instance, on a good signal day, buy orders arrive with the composition of $\varepsilon+\mu$ rates, and sell orders arrive at the rate of ε.

[1] In other words, traders send their orders to the exchange.



α is the probability of an information event, δ is the probability of a low signal, μ is the rate of informed trade arrival, and ε is the rate of uninformed buy and sell trade arrivals. Nodes to the left of the dotted line occur once per day before the trading.

Figure 6.1: The Trading Procedure in the Model

6.3.2 Model

The market-maker employs Bayes rules to update his/her beliefs about the nature of an information event throughout the trading day because he/she does not know the occurrence nor type of the information event. The market-maker uses the arrival of trades and the rate of trading to do this. Let $P(t) = (P_n(t),\ P_b(t),\ P_g(t))$ be the market-maker's prior belief about the events "no signal" (n), "bad signal" (b), and "good signal" (g) at time t. The market-maker determines the quote at time t by updating his or her prior condition on the arrival of an order of the relevant type. Let S_t denote the event that a sell order arrives at time t; similarly, B_t represents a buy order at time t. Let $P(t|S_t)$ be the market-maker's updated belief vector conditional on the history prior to time t and on the event that a sell order arrives at time t.

If an order to sell arrives at time t, the posterior probability of the market-maker given no signal at time t is

$$P_n(t|S_t) = \frac{P_n(t)\varepsilon}{\varepsilon + P_b(t)\mu} \tag{6.1}$$

Similarly, the posterior probability given a bad signal is

$$P_b(t|S_t) = \frac{P_b(t)(\varepsilon + \mu)}{\varepsilon + P_b(t)\mu} \tag{6.2}$$

and the posterior probability given a good signal is

$$P_g(t|S_t) = \frac{P_g(t)\varepsilon}{\varepsilon + P_b(t)\mu} \tag{6.3}$$

The corresponding probabilities in the case of a buy order arrival are

$$\begin{aligned} P_n(t|B_t) &= \frac{P_n(t)\varepsilon}{\varepsilon + P_g(t)\mu} \\ P_b(t|B_t) &= \frac{P_b(t)\varepsilon}{\varepsilon + P_g(t)\mu} \\ P_g(t|B_t) &= \frac{P_g(t)(\varepsilon + \mu)}{\varepsilon + P_g(t)\mu} \end{aligned} \quad . \tag{6.4}$$

The ask quote at time t on day i is

$$a(t) = \frac{P_g(t)(\varepsilon + \mu)\overline{V_i} + P_b(t)\varepsilon \underline{V_i} + P_n(t)\varepsilon V_i^*}{\varepsilon + P_g(t)\mu} \quad (6.5)$$

and the bid quote at time t on day i is

$$b(t) = \frac{P_b(t)(\varepsilon + \mu)\underline{V_i} + P_g(t)\varepsilon \overline{V_i} + P_n(t)\varepsilon V_i^*}{\varepsilon + P_b(t)\mu}. \quad (6.6)$$

Note that the ask quote at time t is the weighted average of the asset value on day i: similarly, the bid quote at time t is the weighted average value. The weights for the ask are the posterior probabilities of the market-maker when he or she receives a buy order, and the weights for the bid are the posterior probabilities of the market-maker when a sell order is received.

Then the bid-ask spread is determined by

$$\Sigma(t) = \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}\left(\bar{V_i} - E[V_i|t]\right) + \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}\left(E[V_i|t] - V_i\right), \quad (6.7)$$

where $E[V_i|t] = P_n(t)V_i^* + P_b(t)V_i + P_g(t)\bar{V_i}$ is the expected value of the asset conditional on the history of trade prior to time t.

The bid-ask spread at time t is composed of two terms; the first term is equal to the probability of an informed buy times the expected loss to an informed buyer, and the second term is symmetric for sells. The bid-ask spread in the Easley, et al. (1996) model arises from asymmetric information, that is, the spread does not reflect other components such as inventory holding and order processing.

The Easley, et al. (1996) model provides the probability of informed trading PIN(t), exactly

$$PIN(t) = \frac{\mu(1 - P_n(t))}{\mu(1 - P_n(t)) + 2\varepsilon}. \quad (6.8)$$

The initial belief of the market-maker on the probability of informed trading can be interpreted as the unconditional share of informed trading.

$$PIN(t) = \frac{\mu \cdot \alpha}{\mu \cdot \alpha + 2 \cdot \varepsilon} \quad (6.9)$$

The Implication of PIN(t) in equation (6,9): This quantity represents the fractional value of informed trading to total trading. i.e. the denominator $\mu\alpha$ represents the orders from informed traders since μ represents the order arrival rate from informed traders. Additionally 2ε represents the orders from uninformed traders since ε represents the order arrival rate from uninformed traders. Finally, the PIN measure represents the ratio of informed orders to total orders.

Since the PIN measure is the ratio between informed orders and uninformed orders, the PINs of two independent samples can be different even if two independent samples have the same arrival rate of informed traders. Thus, the higher arrival rate of informed orders does not lead to a higher PIN when two independent samples are compared. Moreover, the lower arrival rate of uninformed orders does not lead to a lower PIN. The manifestation of a higher or lower PIN depends on the relative arrival rate of informed orders and uninformed orders.

Meanwhile these arrival rates μ and ε are estimated per day in the model. Since the model assumes that the arrival rate follows a Poisson distribution, the inter-arrival time[1] is on average 1/arrival rate. For example, if the average rate of arrival μ is 10 per day, then the average inter-arrival time is 1/5 day. Additionally the inter-arrival time follows exponential distribution.

6.3.3 The Characteristics of the PIN

Chung, Li, and McInish (2005) find that there is a positive and significant relationship between the PIN and both the price impact of trades and serial correlation in trade direction. In their study, they argue that positive serial correlation in trade direction, or quote revision, is attributed to either the inventory control of a dealer[2] or the trading behaviour of informed traders[3]. They find that the time interval between trades has a significant impact on the serial correlation of trade types when stocks have a higher PIN.

[1] The inter-arrival time means the time between arrivals

[2] Hasbrouck (1991), Chan and Lakonishok (1993), and Chan and Lakonishok (1995)

[3] Covrig and Ng (2004) and Kelly and Steigerwald (2004)

Note that the PIN and the adverse selection cost measure information asymmetry in the trading procedure. While the estimated PIN is the probability that a trade comes from informed traders, the estimated adverse selection cost is an average cost borne by a market-maker to informed traders. The adverse selection cost of the bid-ask spread is the trading loss of the market-maker to informed traders, while the market-maker is likely to trade with uninformed traders. In this case the market-maker earns a trading profit by the amount of the adverse selection cost. Thus, these unexpected trading profits will be counted in the estimation of the adverse selection. i.e. the estimated adverse selection cost will be an average value of real adverse selection losses and unexpected adverse selection profits due to trading with uninformed traders. In contrast to the adverse selection cost, the PIN is a probability measure. When the market-maker holds an incorrect probability, he/she can correct that probability without changing the posted bid/ask price. From these characteristics, the PIN can be updated after each transaction while the adverse selection cost of the spread cannot. Thus, to provide the information content of a trade, the PIN can be computed from each transaction while the adverse selection cost can be computed when total trades are known.

6.4 Empirical Model to Estimate the PIN

6.4.1 Log Likelihood Function of the PIN

Estimating the PIN requires the estimation of the parameters θ=(α, δ, ε, μ) in equation (6,8). The Easley et al. (1996) model employs the likelihood function to estimate the parameter vector (α, δ, ε, μ). It considers three scenarios: a bad-event day, a good-event day, and a no-event day. In addition, buys and sells follow one of three Poisson processes on a given day. The likelihood of observing any sequence of orders on a bad-event day is given by

$$e^{-\varepsilon T}\frac{(\varepsilon T)^B}{B!}e^{-(\mu+\varepsilon)T}\frac{[(\mu+\varepsilon)T]^S}{S!}, \tag{6.10}$$

where B is the number of buys and S is the number of sells and T is total time.

On a no-event day, the likelihood function is

$$e^{-\varepsilon T}\frac{(\varepsilon T)^B}{B!}e^{-\varepsilon T}\frac{(\varepsilon T)^S}{S!} \tag{6.11}$$

On a good-event day the likelihood function is

$$e^{-(\mu+\varepsilon)T}\frac{[(\mu+\varepsilon)T]^B}{B!}e^{-\varepsilon T}\frac{(\varepsilon T)^S}{S!} \tag{6.12}$$

Equations (6.10), (6.11), and (6.12) suggest that the number of buys B and the number of sells S are sufficient statistics to estimate the order arrival rates of buys and sells.

When the type of day is unknown, the likelihood function can be found by the weighted average of Equations (6.10), (6.11), and (6.12). And the weight is the probability that each type of day occurs, i.e. the probability of a no-event day, a bad-event day, and a good event day, is (1-α), $\alpha\delta$, and α(1-δ), respectively. Finally, the likelihood function is

$$\begin{aligned} L((B,S)|\theta) = {} & (1-\alpha) * e^{-\varepsilon T}\frac{(\varepsilon T)^B}{B!}e^{-\varepsilon T}\frac{(\varepsilon T)^S}{S!} \\ & + \alpha\delta * e^{-\varepsilon T}\frac{(\varepsilon T)^B}{B!}e^{-(\mu+\varepsilon)T}\frac{[(\mu+\varepsilon)T]^S}{S!} \\ & + \alpha(1-\delta) * e^{-(\mu+\varepsilon)T}\frac{[(\mu+\varepsilon)T]^B}{B!}e^{-\varepsilon T}\frac{(\varepsilon T)^S}{S!} \end{aligned} \tag{6.13}$$

In case of multiple days, the likelihood function is the product of the daily likelihoods, since days are assumed to be independent.

$$L(M|\theta) = \prod_{i=1}^{I} L(\theta|B_i, S_i) \tag{6.14}$$

where $M = (B_i, S_i)_{i=1}^{I}$ over I days and $\theta = (\alpha,\, \delta,\, \varepsilon,\, \mu)$.

Finally the likelihood function (6.14) is maximised to estimate the parameter vector $\theta = (\alpha,\, \delta,\, \varepsilon,\, \mu)$. However, the likelihood function (6.14) makes a computation problem. Specifically, factorials like B! and S! in the likelihood function (6.14) increase the computation time required to find parameter vectors. In addition, when there is no closed form solution to the maximisation of the likelihood function (6.14), it takes much computation time.

Therefore, Easley, Hvidkjaer, and O'Hara (2004) suggest a log likelihood function to avoid the computation problem. In their paper, they distinguish buy and sell orders of uninformed traders. In other words, buy and sell orders from uninformed traders are assumed to arrive at the market according to an independent Poisson process with daily arrival rates ε_b for buy orders and ε_s for sell orders. After dropping a constant term and re-arranging, the likelihood function (6.14) is transformed into the log-likelihood function[1]:

$$\begin{aligned} &L((B_i, S_i)_{i=1}^{I}|\theta) = \\ &\sum_{i=1}^{I} \left[-\varepsilon_b - \varepsilon_s + M_i(\ln x_b + \ln x_s) + B_i \ln(\mu + \varepsilon_b) + S_i \ln(\mu + \varepsilon_s)\right] \\ &+ \sum_{i=1}^{I} \ln\left[\alpha(1-\delta)e^{-\mu} x_s^{S_i - M_i} x_b^{-M_i} + \alpha\delta e^{-\mu} x_b^{B_i - M_i} x_s^{-M_i} + (1-\alpha) x_s^{S_i - M_i} x_b^{B_i - M_i}\right] \end{aligned} \tag{6.15}$$

where $M_i = (\min(B_i, S_i) + \max(B_i, S_i))/2$, $x_s = \varepsilon_s/(\mu + \varepsilon_s)$, and $x_b = \varepsilon_b/(\mu + \varepsilon_b)$. By using the log-likelihood function (6.15), the computation problem can be reduced.

I employ the log-likelihood function (6.15) to estimate the parameter vector θ. Furthermore, I run the SAS NLP procedure 400 times to maximise the function (6.15). To maximise the log likelihood function (6.15), the program chooses randomly initial values of parameter vector θ. Random initial parameters and 400 simulations could allow me not only to avoid local maximum values, but also to provide the estimated PIN close to the intrinsic PIN. After 300 simulations, there are no noticeable differences between the average values of estimated parameters. (For the method to solve the maximisation problem, refer to Yan and Zhang (2006)).

[1]When the arrival rate of uninformed orders ε is discerned into buy order arrival rate ε_b and sell order arrival rate ε_s, the likelihood function is expressed by
$$\begin{aligned} L((B,S)|\theta) = &(1-\alpha) * e^{-\varepsilon_b T} \frac{(\varepsilon_b T)^B}{B!} e^{-\varepsilon_s T} \frac{(\varepsilon_s T)^S}{S!} \\ &+ \alpha\delta * e^{-\varepsilon_b T} \frac{(\varepsilon_b T)^B}{B!} e^{-(\mu+\varepsilon_s)T} \frac{[(\mu+\varepsilon_s)T]^S}{S!} \\ &+ \alpha(1-\delta) * e^{-(\mu+\varepsilon_b)T} \frac{[(\mu+\varepsilon_b)T]^B}{B!} e^{-\varepsilon_s T} \frac{(\varepsilon_s T)^S}{S!} \end{aligned}$$

6.4.2 Other Issues with the PIN

6.4.2.1 Extreme Values of Parameter α and δ

Other difficulties exist in estimating the PIN because a Maximum Likelihood Estimation method is used to compute it. i.e. extreme parameter values are estimated in the computation, but those extreme parameter values are unlikely to happen in the real market. For example, parameters α and δ as probability measures can have a value between zero and one. If, however, parameters α and δ take on the value of zero or one, then these cases can be considered as corner solutions in the estimation. The reason is that when α takes on a value of zero or one it implies that no information event occurs, or that an information event always occurs. In addition, if δ is one the signal is always good. If δ is zero the signal is always bad. These situations are less likely to happen in real trading on a market. Thus, when the estimated value of parameters α and δ is either zero or one, these estimated values are extreme values and should be removed in the computation of the PIN.

These extreme values of parameters α and δ lead to a biased PIN measure and an incorrectly estimated PIN that can lead to the wrong interpretation of empirical findings in many studies. Thus many researchers develop their own method to avoid this concern. A bootstrapping method is used in Vega (2006), while Brown, Hillegeist, and Lo (2004) filter out the questionable probabilities associated with informed trading[1] to check the robustness of their empirical findings. Recently, Yan and Zhang (2006) have suggested that biased probabilities associated with informed trading arise from a maximum likelihood estimation process. When parameters α or δ are equal to zero or one the estimated parameters α or δ are considered as boundary solutions. They show that the boundary solutions for parameters generate a biased probability of informed trading.

Following Yan and Zhang (2006), I also removed extreme parameters (α and δ which are either greater than 0.99 or smaller than 0.01). Two broad-market ETFs and one industry-wide ETF are deleted from the analysis because these ETFs

[1](1) if $50\mu > \varepsilon$ or $50\varepsilon > \mu$, where $\varepsilon = \varepsilon_b + \varepsilon_s$; (2) if α <0.02 or α >0.98; (3) if δ <0.02 or $\delta > 0.98$; and (4) if $\min(\varepsilon, \mu) < 1$.

have extreme parameters. Thus, the full sample in this chapter is composed of 37 broad-market ETFs and 37 industry-wide ETFs with paired control stocks.

6.5 Full Sample ETFs

First, I provide the estimation results of the PIN based on the full ETF sample. Secondly I describe the results for broad-market ETFs in section 6.6 and the results for industry-wide ETFs in section 6.7.

6.5.1 Sample Statistics of the Parameters of Full ETFs

Table 6.1 provides the summary statistics of each parameter. The adjusted ε_b represents the daily arrival rate of uninformed buy orders that are divided by the number of trades a day. This statistic considers the difference between the number of trades in ETFs and control securities[1]. The adjusted arrival rate of uninformed ETF buy orders is 0.3982 per trade (median rate 0.4113), while the adjusted uninformed buyer of control securities arrives at the average rate of 0.2747 (median rate 0.3233) per trade. This difference is supported by both the paired T-test and Wilcoxon signed rank sum test at the 1% level of significance.

ε_b represents the daily arrival rate of uninformed buy orders. Uninformed buyers of ETFs arrive at the average rate of 82.54 (median rate 57.55) while the uninformed buyer of control securities arrives at the average rate of 45.30 (median rate 28.62). The average arrival rates mean that the uninformed traders of ETFs send buy orders 82.54 times per day and the uninformed traders of control securities send buy orders 45.30 times per day. The median arrival rate can be interpreted in a similar way. The arrival rates of uninformed buy orders are significantly different between ETFs and control securities. This difference is supported by both the paired T-test and Wilcoxon signed rank sum test at the 1% level of significance.

[1]Usually ETFs are considered to have higher trading volume than control securities. This implies that the higher trading volume may lead to the higher arrival rate of informed (uninformed) orders. Thus using the adjusted ε_b and ε_s can relfect the difference in trading volume between ETFs and control securities

The adjusted ε_s is similar to the adjusted ε_b in that the adjusted ε_s represents the daily arrival rate of uninformed sell orders divided by the number of trades a day. The adjusted arrival rate of uninformed ETF sell orders is 0.2824 per trade (median rate 0.3022), while the adjusted uninformed seller of control securities arrives at the average rate of 0.2747 (median rate 0.3285) per trade. Both the paired T-test and Wilcoxon signed rank sum test suggest that there is no difference between ETFs and control securities.

Meanwhile, uninformed sell orders are similar to uninformed buy orders. The average (median) arrival rate of uninformed sell orders shows at a 10% (5%) level that there is a difference between ETFs and the control securities. The median arrival rate of uninformed sell orders is significantly different between ETFs (42.52) and controls (26.04). ETFs have more uniformed sell orders than controls. Therefore, it seems that more uninformed buy/sell orders arrive at the ETF than controls, implying that ETFs are assets preferred by uninformed traders, compared to individual securities.

The adjusted 'μ' represents the arrival rate of informed traders divided by the number of trades a day. The adjusted arrival rates of informed traders are on average 0.6407 (on median 0.4356) a trade in the case of ETFs. The average (median) rates of control stocks are 0.7863 (0.5431) a trade. The interpretation of adjusted informed order arrival is similar to that of the adjusted uninformed order arrival rate. i.e. the informed traders of ETFs send an order on average 0.6407 times a trade and the informed traders of control securities send orders on average 0.7863 times a trade. The adjusted informed order arrival rates suggest that there is no difference between ETFs and control securities. This indifference is supported by the paired t-test and the Wilcoxon signed-rank sum test.

'μ' represents the arrival rate of informed traders. The average (median) arrival rates of informed traders are 64 (55.91) times per day in the case of ETFs and the average (median) rates of control stocks are 47.86 (34.44) times per day. The interpretation of average informed order arrival is similar to that of the average uninformed order arrival rate. i.e. the informed traders of ETFs send an order on average 64 times per day and the informed traders of control securities send orders on average 47.86 times per day. The informed traders of ETFs seem to send on average (median) more orders than the informed traders of control stocks, which is confirmed by the paired t-test and the Wilcoxon signed-rank sum test. There appears to be a difference between ETFs and control stocks in terms of informed order arrival rate.

'α' and 'δ' represent the probability of an information event occurring and the probability of the type of news being of a low signal, respectively. These parameters are assumed to be the decision of nature. The average value 0.4 of 'α' suggests that information events arise on average on 40% of the sample trading days and that the remaining 60% of days do not have any information event. The average value of 'δ' in an ETF sample is 0.40 and this implies that whenever information events occur, the probability that the event is bad (terminal value is low) is on average 0.40. In case of control stocks, the average 'α' is 0.35 and the average 'δ' is 0.47. Based on these values, ETFs are less likely to fall whenever an information event happens, than controls stocks. In case of control stocks the probability of either price increases or price decreases is almost half whenever an information event occurs.

Meanwhile, the estimated values of 'α' and 'δ' suggest that there are differences between ETFs and control stocks. These differences are supported by the paired t-test and Wilcoxon signed-rank sum test.

6.5.2 PIN Statistics of Full ETFs

Table 6.2 shows the PIN statistics using the full sample of ETFs and control stocks. The PIN of the ETF and control stock is computed as $\alpha\mu/(\alpha\mu+\varepsilon_b+\varepsilon_s)$, which is slightly different from Equation (6.9). The reason is that I distinguish between ε_b and ε_s in the PIN estimation procedure (Equation (6.15)), while Equation (6.9) does not distinguish between an uninformed buy order ε_b and an uninformed sell order ε_s.

The paired difference column of Table 6.2 provides mixed results on the hypothesis that the PIN of ETFs is smaller than the PIN of control stocks. The paired difference is computed as the PIN of the ETF minus the PIN of its matched control securities. The positive paired difference means that the ETF has a greater PIN than its matched control stock. When the PIN difference between the ETF and control stock is equally either positive or negative, the average

Table 6.1: The Parameters of Full Sample ETFs

ε_b and ε_s represent the daily arrival rate of uninformed buy orders and of uninformed sell orders. 'μ' represents the arrival rate of informed traders. 'α' and 'δ 'represent the probability of an information event occurring and the probability of the type of news occurring. Adjusted ε_b represents the the daily arrival rate of uninformed buy orders divided by the number of traded a day. Adjusted ε_s represents the the daily arrival rate of uninformed sell orders divided by the number of traded a day. Adjusted 'μ' represents the arrival rate of informed traders divided by the number of trades a day.
The ETF column represents statistics from ETFs and the Control column represents statistics from control stocks. The paired difference column is computed as the parameter of ETF minus paired control stock. 74 paired differences are computed and these differences are used for statistics in the paired difference column. If average (median) paired difference is not close to zero, then the parameter is different between ETFs and control stocks. Negative difference means that the ETF has a smaller PIN than its paired control stock. The value in paired difference represents the average difference and the value in parenthesis is paired difference test statistic. The paired t-test and Wilcoxon signed-rank sum test are used to test whether paired differences are close to zero.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level.
Range represents the difference between Max and Min values.

Parameters	Statistics	ETF	Control	Paired Difference (ETF minus control)
ε_b	Average	82.54	45.30	37.23*** (3.09)
	Median	57.55	28.62	15.52*** (603.5)
	Std. Dev.	84.76	60.94	103.48
	Range	442.20	376.33	817.88
Adjusted ε_b	Average	0.3982	0.2747	0.1234*** (7.15)
	Median	0.4113	0.3233	0.1120*** (1049.5)
ε_s	Average	67.24	44.77	22..46* (1.89)
	Median	42.52	26.04	4.39** (393.5)
	Std. Dev.	80.85	62.90	101.79
	Range	481.50	422.97	903.69
Adjusted ε_s	Average	0.2824	0.2702	0.0121 (0.6421)
	Median	0.3022	0.3285	-0.0092 (47.5)
'μ'	Average	64.00	47.86	16.13** (2.11)
	Median	55.91	34.44	-0.001* (307.5)
	Std. Dev.	48.83	43.68	65.60
	Range	189.10	188.39	328.14
Adjusted 'μ'	Average	0.6407	0.7863	-0.1456 (-0.7832)
	Median	0.4356	0.5431	0.0000 (-137.5)
'α'	Average	0.40	0.35	0.054* (1.83)
	Median	0.42	0.34	0.051** (454.5)
	Std. Dev.	0.15	0.20	0.25
	Range	0.61	0.93	1.44
'δ'	Average	0.40	0.47	-0.066* (-1.96)
	Median	0.42	0.49	-0.76** (-398.5)
	Std. Dev.	0.19	0.19	0.29
	Range	0.76	0.84	1.30

(median) PIN difference is close to zero. The positive average PIN difference, however, implies that more than half of ETFs have a greater PIN than controls.

The average (median) paired difference is 0.0172 (0.0287). These results suggest that ETFs have on average (median) greater PINs than control stocks do. The paired t-test statistic is insignificant, though the Wilcoxon signed-rank sum test statistic is significant at the 5% level. Thus, the PIN differences between ETFs and control stocks show mixed results. In terms of average PIN value, ETFs are not significantly different from the PIN of control securities. The median value suggests that ETFs have a greater PIN than control securities.

Meanwhile, the PIN standard deviation for ETFs is 0.0752 and the PIN standard deviation of control stocks is 0.1082. The range is computed as the difference between the maximum and minimum PIN. The PIN range of control stocks is 0.6290, and is almost twice the PIN range of ETFs. The PIN standard deviation and the PIN range of ETFs are more tightly concentrated in lower PINs than control stocks.

Figure 6.2 clearly confirms this condensed pattern. The bottom row of Table 6.2, No Obs. >0, represents the number of observations that are greater than zero. The No. Obs. >0 of Paired Difference reports the number of observation that the PIN of ETFs is greater than that of its matched control securities. This shows that 48 out of the 74 ETFs display a greater PIN compared to control securities.

Table 6.1 suggests that the uninformed traders of ETFs send more buy/sell orders than the uninformed traders of control securities, while the informed traders of ETFs also send the same amounts of orders as the informed traders of control securities. Despite the difference shown in Table 6.1, Table 6.2 suggests that ETFs have a greater PIN than control stocks.

Table 6.2: The PIN of Full Sample ETFs and Control Stocks

The PIN is computed by $\alpha\mu$ / $(\alpha\mu + \varepsilon\_b + \varepsilon\_s)$ and each parameter is an estimated parameter. Paired Difference is calculated by the PIN of an ETF minus the PIN of its paired control. Paired T-test and Signed Rank Sum Test provide test statistics of paired difference average and median, respectively And the test statistics are shown in parenthesis below average and median value of paired difference column. Both tests are based on a two-tailed test, and the null hypothesis is that the average (median) equals zero. Negative difference means that the ETF has a smaller PIN than its paired control stock. The value in paired difference represents the average difference and the value in parenthesis is paired difference test statistic.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level. No. Obs. >0 in the bottom row represents the number of observation that is greater than zero and thus No. Obs. >0 of paired Difference represents the number of observations in case that the PINs of ETFs are greater than those of their matched controls. Total number of observation is 74.

Statistics	ETF	Control	Paired Difference (ETF minus control)
Average	0.1661	0.1489	0.0172 (1.35)
Median	0.1711	0.1346	0.0287** (432.5)
Std. Dev.	0.0752	0.1082	0.1097
Range	0.3296	0.6290	0.6769
No. Obs.>0	74	74	48

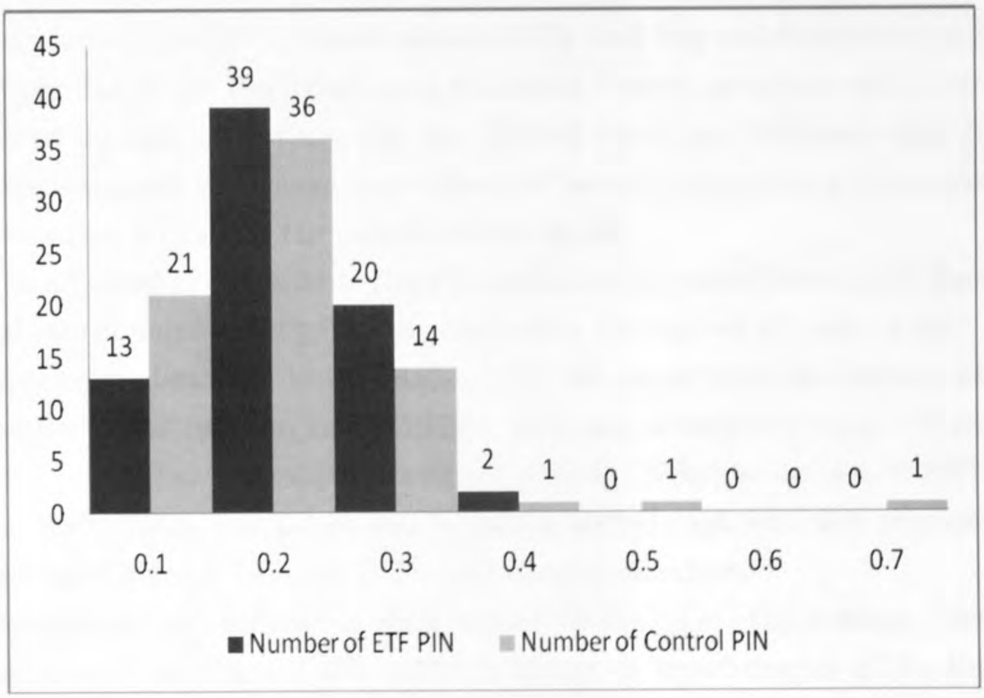


Figure 6.2: The Histogram of Full Sample ETFs

6.6 Broad-market ETFs

6.6.1 Sample Statistics of the Parameters of Broad-Market ETFs

In this section, I discuss broad-market ETFs and paired control stock samples. Table 6.3 shows summary statistics of estimated parameters. The adjusted ε_b is the daily arrival rate of uninformed broad-market ETF (control securities) buy orders that are divided by the number of trades a day. This statistic considers the difference between the number of broad-market ETFs trades and control securities. The adjusted arrival rate of uninformed broad-market ETF buy orders is 0.4423 per trade (median rate 0.4956), while the adjusted uninformed buyer of control securities arrives at the average rate of 0.2952 (median rate 0.3292) per trade. This difference is supported by both the paired T-test and Wilcoxon signed rank sum test at the 1% level of significance.

The average (median) arrival rate of uninformed buy orders ε_b is 88.93 (66.76) in the case of broad-market ETFs, but for control stocks it is 45.24 (23.83). Thus, the uninformed traders of broad-market ETFs send buy orders on average 88.93 times per day, while the uninformed traders of control securities send their buy orders on average 45.24 times per day. Paired t-test and Wilcoxon signed-rank sum tests suggest that there is a difference between uninformed buy orders of broad-market ETFs and the paired control stocks.

The adjusted ε_s is similar to the adjusted ε_b in that the adjusted ε_s is the daily arrival rate of uninformed sell orders divided by the number of trades a day[1]. The uninformed traders send broad-market ETF sell orders with the average rate of 0.2564 per trade (median rate 0.2830), while the uninformed sellers of control securities send their orders at the average rate of 0.2704 (median rate 0.3267) per trade. Both the paired T-test and Wilcoxon signed rank sum test suggest that there is no difference between ETFs and control securities.

Uninformed sell orders ε_s show mixed results. i.e. the average (median) arrival rate of uninformed sell orders is higher in broad-market ETFs than in

[1]Usually ETFs are considered to have higher trading volume than control securities, implying that the number of trading occurrence. This difference may generate the difference in the arrival rate, thus using the adjusted ε_b and ε_s

control stocks. While paired t-test reports no difference between broad-market ETFs and matched control securities, the Wilcoxon signed-rank sum test shows that the arrival rate of uninformed arrival rates are different between broad-market ETFs and control stocks.

The adjusted 'μ' is the arrival rate of informed traders divided by the number of trades a day. The adjusted arrival rates of informed traders are on average 0.8236 (on median 0.3882) a trade in the case of ETFs. The average (median) rates of control stocks are 0.7918 (0.5212) a trade. The informed traders of ETFs send an order on average 0.8236 times a trade and the informed traders of control securities send orders on average 0.7863 times a trade. The paired t-test and the Wilcoxon signed-rank sum test suggest that there is no difference between ETFs and control securities in terms of the adjusted informed order arrival rates.

The informed order arrival rates 'μ' also show mixed results. The paired t-test suggests a difference between broad-market ETFs and control securities, but the Wilcoxon signed rank sum test does not. The average paired difference is 17.18 and the median paired difference is 0.0001.

The probability of an information event occurring 'α' does not display any difference between broad-market ETFs and control securities. This is confirmed by both the paired t-test and Wilcoxon signed rank sum test. In the case of 'δ', broad-market ETFs are different from their matched control stocks. i.e. whenever an information event occurs, the probability that the information event means a bad signal is 0.3869 in the case of broad-market ETFs and 0.4865 for matched control stocks.

6.6.2 PIN Statistics of Broad-Market ETFs

Table 6.4 presents the PIN statistics obtained from broad-market ETFs. The same interpretation of the full ETF sample applies for the broad-market ETFs. The paired t-test and Wilcoxon singed-rank sum test shows insignificant statistics, implying no difference between broad-market ETFs and matched control securities. The paired difference column shows the average (median) PIN difference between ETFs and control stocks as -0.0007 (0.006) respectively. The negative average of paired difference suggests that broad-market ETFs have

Table 6.3: The Parameters of Broad-market ETFs

ε_b and ε_s represent the daily arrival rate of uninformed buy orders and of uninformed sell orders. 'μ' represents the arrival rate of informed traders. 'α' and 'δ 'represent the probability of an information event occurring and the probability of the type of news occurring. Adjusted ε_b represents the the daily arrival rate of uninformed buy orders divided by the number of traded a day. Adjusted ε_s represents the the daily arrival rate of uninformed sell orders divided by the number of traded a day. Adjusted 'μ' represents the arrival rate of informed traders divided by the number of trades a day.
The ETF column represents statistics from ETFs and the Control column represents statistics from control stocks. The paired difference column is computed as the parameter of ETF minus paired control stock. 37 paired differences are computed and these differences are used for statistics of paired difference column. If average (median) paired difference is not close to zero, then the parameter is different between ETFs and control stocks. Negative difference means that the ETF has a smaller PIN than its paired control stock. The value in paired difference represents the average difference and the value in parenthesis is the paired difference test statistic. The paired t-test and Wilcoxon signed-rank sum test are used to test that paired differences are close to zero.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level.
Range represents the difference between Max and Min values.

Parameters	Statistics	ETF	Control	Paired Difference (ETF minus control)
ε_b	Average	88.93	45.24	43.69** (2.17)
	Median	66.76	23.83	32.14*** (168.5)
	Std. Dev.	92.03	68.50	122.02
	Range	442.03	376.26	817.88
Adjusted ε_b	Average	0.4423	0.2952	0.1470*** (6.46)
	Median	0.4956	0.3292	0.1302*** (311.5)
ε_s	Average	62.64	45.10	17.54 (0.88)
	Median	40.98	26.50	4.82* (109.5)
	Std. Dev.	86.42	73.35	120.10
	Range	481.10	422.97	903.69
Adjusted ε_s	Average	0.2564	0.2704	-0.0139 (-0.58)
	Median	0.2830	0.3267	-0.0127 (-54.5)
'μ'	Average	62.37	45.19	17.18* (1.83)
	Median	55.45	34.98	0.0001 (93.5)
	Std. Dev.	47.12	35.65	57.09
	Range	189.10	119.59	239.24
Adjusted 'μ'	Average	0.8236	0.7918	0.0318 (0.09)
	Median	0.3882	0.5212	0.0151 (-9.5)
'α'	Average	0.3900	0.3693	0.206 (0.406)
	Median	0.4193	0.3153	0.035 (71.5)
	Std. Dev.	0.1759	0.2208	0.309
	Range	0.5964	0.9359	1.444
'δ'	Average	0.3869	0.4865	-0.0995** (-2.17)
	Median	0.4046	0.5181	-0.0794** (-148.5)
	Std. Dev.	0.1969	0.1806	0.277
	Range	0.7301	0.7799	1.292

a smaller PIN than paired control stock, while the positive median suggests the opposite. Both tests do not support the argument that a difference between broad-market ETFs and matched controls exists.

Meanwhile the dispersion of broad-market ETFs is smaller than that of control stocks, which appear in the full sample ETFs and control stocks. In the case of broad-market ETFs, the standard deviation is 0.0709 and the range is 0.3270. These standard deviation and range values are smaller than those of control stocks, which are 0.1205 and 0.6282 respectively. These results are shown in Figure 6.3. Additionally, the number of observations which is greater than zero in the Paired Difference column is 19, implying that 19 among 37 ETFs have a greater PIN than their matched controls.

Table 6.4: The PIN of Broad-market ETFs and Control Stocks

The PIN is computed by $\alpha\mu$ / $(\alpha\mu + \varepsilon\_b + \varepsilon\_s)$ and each parameter is an estimated parameter. Paired Difference is calculated by the PIN of an ETF minus the PIN of its paired control. Paired T-test and Signed Rank Sum Test provide test statistics of paired difference average and median, respectively And the test statistics are shown in parenthesis below average and median value of paired difference column. Both tests are based on a two-tailed test, and the null hypothesis is that the average (median) equals zero. Negative difference means that the ETF has a smaller PIN than its paired control stock. The value in paired difference represents the average difference and the value in parenthesis is the paired difference test statistic.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level. No. Obs. >0 in the bottom row represents the number of observation that is greater than zero and thus No. Obs. >0 of paired Difference represents the number of observations in case that the PINs of ETFs is greater than those of their matched control. Total number of observation is 37.

Statistics	ETF	Control	Paired Difference (ETF minus control)
Average	0.1603	0.1673	-0.007 (-0.348)
Median	0.1567	0.1388	0.006 (28.5)
Std. Dev.	0.0709	0.1205	0.1224
Range	0.3270	0.6282	0.6238
No. Obs. >0	37	37	19

6.7 Industry-wide ETFs

6.7.1 Sample Statistics of the Parameters of Industry-wide ETFs

Table 6.5 presents the summary statistics of estimators using industry-wide ETFs. The order arrival rates of industry-wide ETFs are similar to those of full sample

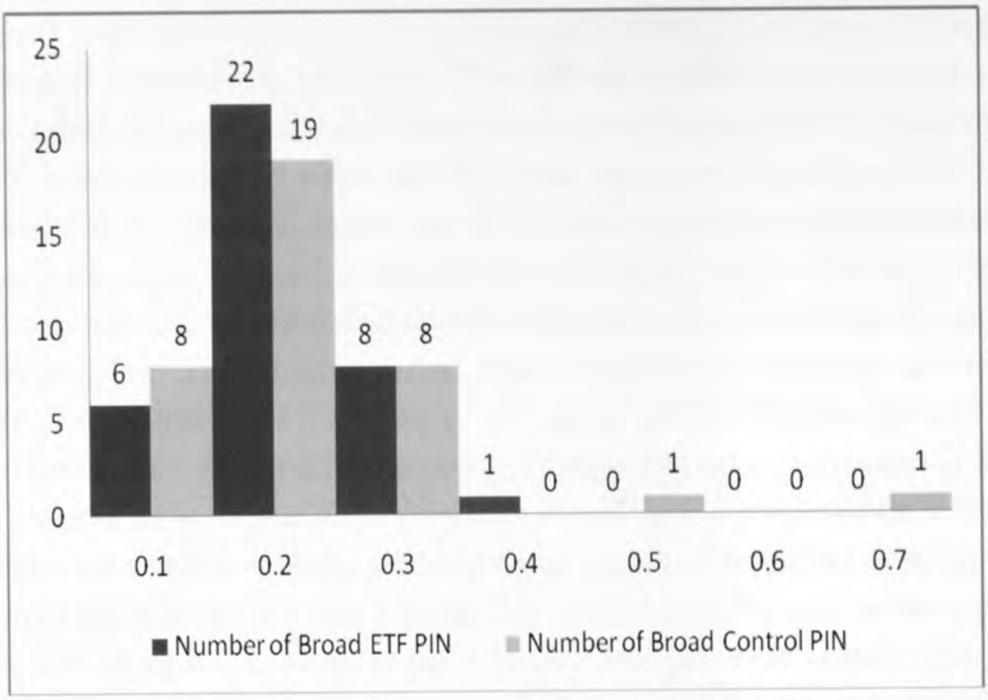


Figure 6.3: The Histogram of Broad-market ETFs

ETFs and the broad-market sample. When industry-wide ETFs are compared to control securities, the adjusted ε_b still shows a difference. i.e. the average (median) adjusted ε_b for industry-wide ETFs are 0.3541 per trade (median rate 0.3866). The average (median) adjusted ε_b for matched control securities are 0.2542 (0.3230) per trade. Moreover the paired T-test and Wilcoxon signed rank sum test confirm the difference at the 1% level of significance.

The uninformed traders send broad-market ETF sell orders with the average rate of 0.3083 per trade (median rate 0.3405), while the uninformed sellers of control securities send their orders at the average rate of 0.2700 (median rate 0.3428) per trade. Both the paired T-test and Wilcoxon signed rank sum test suggest that there is no difference between ETFs and control securities.

The daily arrival rate of uninformed buy and sell orders, ε_b and ε_s respectively, are higher in industry-wide ETFs than in control stocks. For industry-wide ETFs, 76.15 (51.03) is the average (median) arrival rate of uninformed buyer orders and 71.84 (44.39) is the average (median) arrival rate of uninformed seller orders. Meanwhile, control stocks have a 45.36 (29.08) average (median) uninformed buy order arrival rate and a 44.44 (24.46) average (median) uninformed sell order arrival rate. The different arrival rates of uninformed buyer orders are statistically significant between industry-wide ETFs and control stocks in that the paired t-test and Wilcoxon signed rank sum test results are significant at the 5% level. In the case of uninformed sell orders, the paired t-test is significant at the 5% level, although Wilcoxon signed rank test is insignificant.

The adjusted 'μ' is the arrival rate of informed traders divided by the number of trades a day. The adjusted arrival rates of informed traders are on average 0.4577 (on median 0.4721) a trade in the case of ETFs. The average (median) rates of control stocks are 0.7808 (0.5894) a trade. The informed traders of ETFs send an order on average 0.4577 times a trade and the informed traders of control securities send orders on average 0.7808 times a trade. The paired t-test and the Wilcoxon signed-rank sum test suggest that there is no difference between ETFs and control securities in terms of the adjusted informed order arrival rates.

'μ' suggests that the arrival rate of informed orders for industry-wide ETFs and control stocks is similar. This is supported by both the paired t-test and Wilcoxon signed-rank sum test. The median arrival rate of informed orders for ETFs is 65.62, while control stocks have a median arrival rate of 50.53 informed orders.

With regard to 'α', industry-wide ETFs have a significantly different probability of an information event when compared to their matched control securities. The paired t-test statistic is 2.97 and the Wilcoxon signed-rank sum test statistic is 163.5. Since 'α' represents the probability of an information event occurring, the occurrence of an information event is different between industry-wide ETFs and control stocks in terms of median value.

With regard to parameter 'δ', the paired t-test and Wilcoxon signed-rank sum test say that there is no difference between industry-wide ETFs and control stocks.

6.7.2 PIN Statistics of Industry-wide ETFs

Table 6.6 shows the PIN statistics for the industry-wide ETFs. Here I have the same interpretation as those in the full ETF sample and broad-market ETFs.

Table 6.5: The Parameters of Industry-wide ETFs

ε_b and ε_s represent the daily arrival rate of uninformed buy orders and of uninformed sell orders. 'μ' represents the arrival rate of informed traders. 'α' and 'δ 'represent the probability of an information event occurring and the probability of the type of news occurring.
Adjusted ε_b represents the the daily arrival rate of uninformed buy orders divided by the number of traded a day. Adjusted ε_s represents the the daily arrival rate of uninformed sell orders divided by the number of traded a day. Adjusted 'μ' represents the arrival rate of informed traders divided by the number of trades a day.
The ETF column represents statistics from ETFs and the Control column represents statistics from control stocks. The paired difference column is computed as the parameter of ETF minus paired control stock. 37 paired differences are computed and these differences are used for the statistics in the paired difference column. If the average (median) paired difference is not close to zero, then the parameter is different between ETFs and control stocks. Negative difference means that the ETF has a smaller PIN than its paired control stock. The value in paired difference represents the average difference and the value in parenthesis is the paired difference test statistic. The paired t-test and Wilcoxon signed-rank sum test are used to test that paired differences are close to zero.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level.
Range represents the difference between Max and Min values.

Parameters	Statistics	ETF	Control	Paired Difference (ETF minus Control)
ε_b	Average	76.15	45.36	30.78** (2.28)
	Median	51.03	29.08	14.95** (129.5)
	Std. Dev.	77.55	53.28	82.10
	Range	289.76	220.41	356.07
Adjusted ε_b	Average	0.3541	0.2542	0.0999*** (3.89)
	Median	0.3866	0.3230	0.0872*** (215.5)
ε_s	Average	71.84	44.44	27.39** (2.06)
	Median	44.39	24.46	3.97 (94.5)
	Std. Dev.	75.78	51.41	80.84
	Range	236.67	222.74	294.02
Adjusted ε_s	Average	0.3083	0.2700	0.0382 (1.32)
	Median	0.3405	0.3428	-0.0001 (67.5)
'μ'	Average	65.62	50.53	15.09 (1.24)
	Median	57.27	32.17	-0.002 (62.5)
	Std. Dev.	51.08	52.82	73.92
	Range	185.58	188.39	328.14
Adjusted 'μ'	Average	0.4577	0.7808	-0.3230* (-2.02)
	Median	0.4721	0.5894	-0.0049 (-62.5)
'α'	Average	0.4272	0.3399	0.0873*** (2.97)
	Median	0.4299	0.3601	0.0676** (163.5)
	Std. Dev.	0.1348	0.1850	0.178
	Range	0.5382	0.6375	0.651
'δ'	Average	0.4317	0.4649	-0.0332 (-0.66)
	Median	0.4451	0.4672	-0.0733 (-51.5)
	Std. Dev.	0.2006	0.2100	0.3021
	Range	0.7220	0.7613	1.2903

The paired difference column displays the positive average and median values, implying that industry-wide ETFs have a higher PIN than control stocks do. The difference between the industry-wide ETFs and paired control stock is 0.0414 and the median is 0.0405. The paired t-test statistic is significant at the 1% level, as also is the Wilcoxon signed-rank sum test. This suggests that industry-wide ETFs have a higher PINs than their paired control stocks when subjected to the parametric and non-parametric tests. No. Obs. > 0 cell in the paired difference column is 29, suggesting that 78% of industry-wide ETFs have a higher PIN than matched controls.

The histogram in Figure 6.4 is virtually identical to Figure 6.2 and Figure 6.3. Figure 6.4 shows that industry-wide ETFs still have a more concentrated PIN than control stocks. The maximum PIN of the ETFs and matched control securities is located between 0.3 and 0.4, respectively.

Table 6.6: The PIN of Industry-wide ETFs and Control Stocks

The PIN is computed by $\alpha\mu$ / $(\alpha\mu + \varepsilon\_b + \varepsilon\_s)$ and each parameter is an estimated parameter. Paired Difference is calculated by the PIN of an ETF minus the PIN of its paired ETF. Paired T-test and Signed Rank Sum Test provide test statistics of paired difference average and median, respectively And the test statistics are shown in parenthesis below average and median value of paried difference column. Both tests are based on a two-tailed test, and the null hypothesis is that the average (median) equals zero. Negative difference means that the ETF has a smaller PIN than its paired control stock. The value in paired difference represents the average difference and the value in parenthesis is the paired difference test statistic.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level. No. Obs. >0 in the bottom row represents the number of observation that is greater than zero and thus No. Obs. >0 of paired Difference represents the number of observations in the case that the PIN of ETFs is greater than that of its matched controls. Total number of observation is 37.

Statistics	ETF	Control	Paired Difference (ETF minus control)
Average	0.1720	0.1305	0.0414*** (2.78)
Median	0.1757	0.1284	0.0405*** (191.5)
Std. Dev.	0.0798	0.0924	0.0907
Range	0.3199	0.3819	0.4503
No. Obs. >0	37	37	29

6.8 Comparison of ETFs and Control Stocks

The overall results illustrates that there is no difference in the PIN between broad-market ETFs and matched controls, and that industry-wide ETFs have a higher PIN than matched controls. Full ETFs and their matched controls show

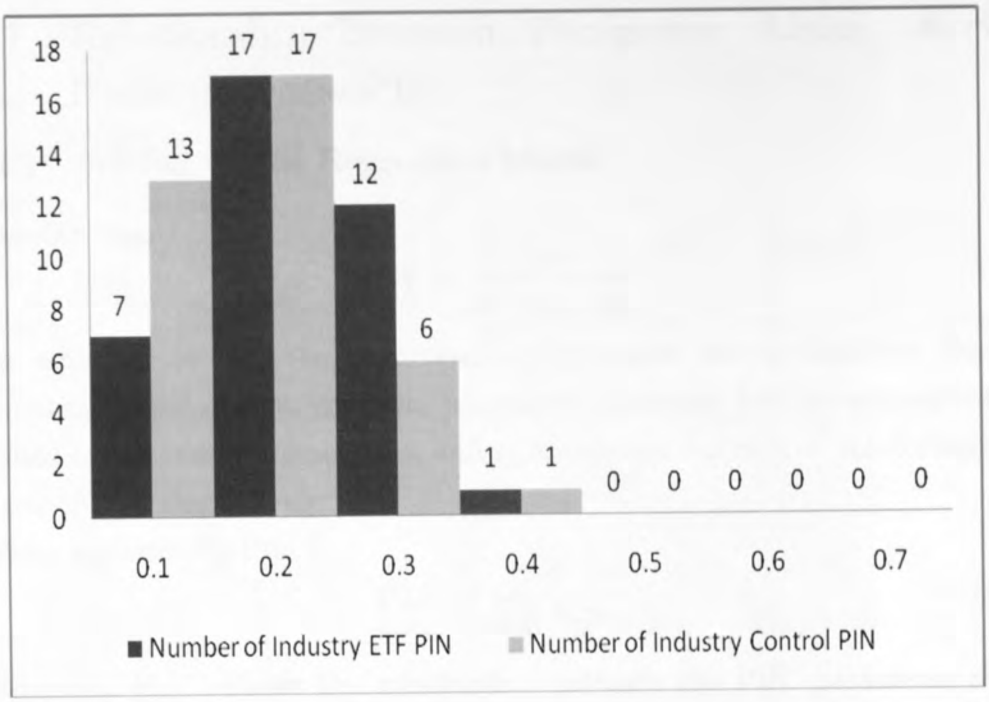


Figure 6.4: The Histogram of Industry-wide ETFs

inconsistent results between the paired t-test and Wilcoxon signed rank sum test. We may think at least that ETFs have the same level of PIN as control securities. This intuition is clearly contradictory to the hypothesis that ETFs should have lower information asymmetry than control securities.

Thus, I consider two different approaches to this puzzling result. The first is to use the regression method (section 6.8.1) to examine the relation between parameter alpha (arrival ratio) and the PIN. The second is to use NASDAQ exchange data (section 6.8.2) to compare ETFs and control securities. The main trading locations for ETFs are the American Stock Exchange (AMEX) and NASDAQ while control securities are mainly traded on the NASDAQ. Constraining NASDAQ quotes and trades limit information asymmetry to one exchange.

6.8.1 Relationship Between Parameter Alpha, Arrival Ratio, and the PIN

6.8.1.1 Setting up the Regression Model

Remember that

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_b + \varepsilon_s} \tag{6.16}$$

In equation (6.16), the parameter α represents the probability that an information event occurs, and the parameter μ stands for the arrival rate of informed orders, and parameters ε_b and ε_s are the arrival rate of uninformed buy and sell orders, respectively.

From equation (6.16)

$$PIN = \frac{\alpha}{\alpha + \frac{\varepsilon_b + \varepsilon_s}{\mu}} \tag{6.17}$$

Equation (6.17) shows the relationship between the PIN, parameter α and the arrival ratio computed as$(\varepsilon_b + \varepsilon_s)/\mu$. The arrival ratio represents the ratio of uninformed orders to informed orders, and is negatively related to the PIN while parameter α is positively related with the PIN. Furthermore, the arrival ratio suggests that as the informed arrival rate increases, the PIN also increases. This relationship suggests that the PIN is an increasing function of the informed order arrival rate while the PIN is a decreasing function of the uninformed order arrival rate. All these relationships are clearly required in Equation (6.17) and lead to the following regression model:

$$PIN = b_0 + b_1 * \alpha + b_2 * AR + e \tag{6.18}$$

where b_0 represents the intercept term, and b_1 and b_2 are coefficients, and e is the error term[1]. α is the estimated parameter of the probability that an information event occurs and AR is $\frac{\varepsilon_b + \varepsilon_s}{\mu}$.

[1] I did not employ dummy regression to see the difference between ETFs and controls when a dummy variable represents zero in the case of ETFs and one in the case of controls. This dummy regression, however, is not proper method for a pair matched sample, since the PIN of an ETF is possibly related with the independent variables of control that are not paired with the ETF.

Table 6.7: Relationship between the PIN and Alpha and Arrival Ratio

The Dependent variable is the PIN of each sample. As an independent variable, parameter α represents the probability of an information event occurring. Another independent variable is the arrival ratio, which is the ratio of uninformed orders arrival rate ($\varepsilon_b + \varepsilon_s$) to informed orders arrival rate (μ). Control represents dummy variable which is one in the case of control securities and zero in the case of ETFs.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level.

Panel (a)						
	Full ETFs		Broad-market		Industry-wide	
	ETFs	Control Stocks	ETFs	Control Stocks	ETFs	Control Stocks
Intercept (b_0)	0.1503*** (6.81)	0.1763*** (10.46)	0.1335*** (5.34)	0.1713*** (6.59)	0.1755*** (4.42)	0.1782*** (7.79)
Parameter α (b_1)	0.1081** (2.06)	-0.0117 (-0.24)	0.1188* (2.03)	0.0265 (0.34)	0.0769 (0.80)	-0.416 (-0.65)
Arrival ratio (b_2)	-0.0048*** (-5.92)	-0.0045*** (-4.24)	-0.0051*** (-3.70)	-0.0046** (-2.49)	-0.0047*** (-4.39)	-0.0043*** (-3.14)
Adj R-squared	0.3253	0.2853	0.2969	0.1608	0.3430	0.3289

6.8.1.2 Regression Results

Table 6.7 shows the estimation results of (6.18). The relationship between the PIN and α is positive. The coefficients (b_1) from full ETFs and broad-market ETFs are significant (0.1081 and 0.1188, respectively). On the contrary, the coefficient of α from industry-wide ETFs is 0.0769 and insignificant. Meanwhile, the coefficients associated with the α from the control stock are insignificant (-0.0117 for full control stocks, 0.0265 for broad-market control stocks, and -0.416 for industry-wide control stocks). This implies that the occurrence of an information event affects the PIN of full ETFs and broad-market ETFs, although the information event has no impact on the PIN in the other cases, including industry-wide ETFs and control securities.

The arrival ratio coefficient in the third row of Table 6.7 implies a negative relationship with the PIN. In other words, the PIN is negatively associated with the arrival ratio, and thus it increases with the arrival rate of an informed order. Meanwhile, as the arrival rate of uninformed orders increases the PIN decreases. A closer look at Table 6.7 demonstrates that ETFs have a greater absolute coefficient value compared to control stocks, implying that ETFs increase by a greater unit of PIN as one unit of an informed order arrives.

For instance, the coefficient of the arrival ratio in the full ETF sample is -0.0048, meaning that when the arrival ratio increases by one unit, the PIN decreases by 0.0048 units. In other words, a unit increase of uninformed order arrival (one unit decrease of informed order arrival) falls by the 0.0048 unit of PIN. Meanwhile, the coefficient of the arrival ratio for its paired control stock is -0.0045, suggesting that a unit increase in the arrival ratio reduces the 0.0045 unit of the PIN. The same unit increase in the arrival ratio decreases by a smaller unit in ETFs than in control stocks. This means that the informed order arrival of ETFs has a greater impact on the PIN than the informed order arrival of control stocks. Therefore, the coefficients of the arrival ratio provide evidence that ETFs have a greater PIN than control stocks.

The same pattern of full ETFs appears in the case of broad-market ETFs and industry-wide ETFs. The arrival coefficient of broad-market ETFs (-0.0051) is smaller than that of matched control stocks (-0.0046). Additionally, industry-wide ETFs have a smaller arrival ratio coefficient (-0.0047) than matched control securities (-0.0043). All these results suggest that when an uninformed order arrives the PIN in relation to the ETFs decrease more than that of the control security. In the case of an informed order, the PIN of the ETFs increases more than the PIN of control securities.

Finally I compare broad-market ETFs with industry-wide ETFs. While the coefficient of the arrival ratio in broad-market ETFs is -0.0051, that of industry-wide ETFs is -0.0047. The same unit increase of the arrival ratio affects broad-market ETFs and industry-wide ETFs differently. In the broad-market ETFs as one informed order arrives the PIN increases by 0.0051 units. In the industry-wide ETFs one informed order arrival increases the PIN by 0.0047 units.

6.8.2 The NASDAQ Exchange

This section compares the PIN of ETFs and control securities using quotes and trades from only one exchange, the NASDAQ. The previous section suggests the unexpected evidence that ETFs have a higher PIN than control securities. It is possible that this unexpected result comes from the fact that I used National Best Bid and Offer prices (NBBO). While NBBO provides the smallest bid-ask spread at the trading time and market-makers are required to match their quote with NBBO, a disadvantage is that NBBO comprises bid/ask quotes from many exchanges. Thus, this section limits bid/ask quotes on the NASDAQ. The reason why the NASDAQ is chosen is that both ETFs and control securities have enough trading data on the NASDAQ to analyse.

When I re-construct data into NASDAQ trading data, five pairs are lost[1]. The ETFs of lost pairs are IWP and JKJ in broad-market ETFs and IYE, IYG, and VFH in industry-wide ETFs. Thus the final sample for this section is 72 ETFs and theirs matched control stocks. The detailed statistics of parameters are shown in Appendix 6.A.

Table 6.8: The PIN of ETFs and Control Securities on the NASDAQ

The PIN is computed by $\alpha\mu$ / $(\alpha\mu + \varepsilon\_b + \varepsilon\_s)$ and each parameter is an estimated parameter. Paired Difference is calculated by the PIN of an ETF minus the PIN of its paired control. Paired T-test and Signed Rank Sum Test provide test statistics of paired difference average and median, respectively And the test statistics are shown in parenthesis below average and median value of paired difference column. Both tests are based on a two-tailed test, and the null hypothesis is that the average (median) equals zero. Negative difference means that the ETF has a smaller PIN than its paired control stock. The value in paired difference represents the average difference and the value in parenthesis is paired difference test statistic.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level. No. Obs. >0 in the bottom row represents the number of observation that are greater than zero and thus No. Obs. >0 of paired Difference represents the number of observations where the PIN of controls is greater than that of its matched ETF. Total number of observation is 36.

Statistics	NASDAQ ETF	NASDAQ Control	Paired Difference (ETF minus control)
Full			
Average	0.2394	0.2804	-0.0410** (-2.26)
Median	0.2326	0.2876	-0.0101 (-256)
Std. Dev.	0.1054	0.1564	0.1540
Range	0.5645	0.6141	0.6996
No. Obs. >0	72	72	33
Broad-Market			
Average	0.2453	0.2827	-0.0373 (-1.65)
Median	0.2295	0.2878	-0.0073 (-84.5)
Std. Dev.	0.1164	0.1329	0.1375
Range	0.5645	0.5934	0.6484
No. Obs. >0	37	37	16
Industry-wide			
Average	0.2331	0.2781	-0.0450 (-1.54)
Median	0.2339	0.2637	-0.0128 (-47)
Std. Dev.	0.0936	0.1800	0.1718
Range	0.3642	0.6124	0.6648
No. Obs. >0	35	35	17

[1]This is due to the removal of extreme value of α and δ.

6.8.2.1 PIN Statistics of NASDAQ ETFs and Control Securities

Table 6.8 shows the PIN statistics for ETFs and control securities. The overall results suggest that there is no difference in the PIN of ETFs and control securities. These results are contradictory to the hypothesis in this chapter that ETFs have a lower probability of informed trading (PIN) than control stocks. The finding in NASDAQ ETFs/controls, however, is similar to the findings in this chapter.

I have the same interpretation as those in the full ETF sample and broad-market ETFs and industry-wide ETFs. The paired difference column displays the negative average and median values, implying that industry-wide ETFs have a smaller PIN than control stocks do. Only the average paired difference between full ETFs and matched control securities show significance at the 5% level. These results suggest that the PIN measures do not show difference between ETFs and their paired control stocks on the NASDAQ exchange.

The histogram in Figure 6.5 shows a virtually identical pattern in Figure 6.2, Figure 6.3 and Figure 6.4. The maximum PIN of ETFs is between 0.6 and 0.7 but the maximum PIN of control stocks is between 0.5 and 0.6. In addition to Figure 6.5, No. Obs. > 0 cell in the paired difference column is 33 for the full ETFs sample, suggesting that 45.83% of the ETFs have a greater PIN than controls when the sample is used. Table 6.8 and Figure 6.5 suggest that even though only NASDAQ data is considered, the PINs of ETFs and matched controls is not significantly different.

6.9 Concluding Remarks

The Probability of Informed Trading (PIN) can be considered as an alternative measure to the adverse selection cost of spread decomposition model. It is, therefore, anticipated that the probability of informed trading is smaller in ETFs than in control stocks.

The overall analysis based on the PIN measure suggests that ETFs have the same level of PINs as control securities (some have a larger PIN). The full ETF sample shows mixed results, in which the paired t-test does not demonstrate a

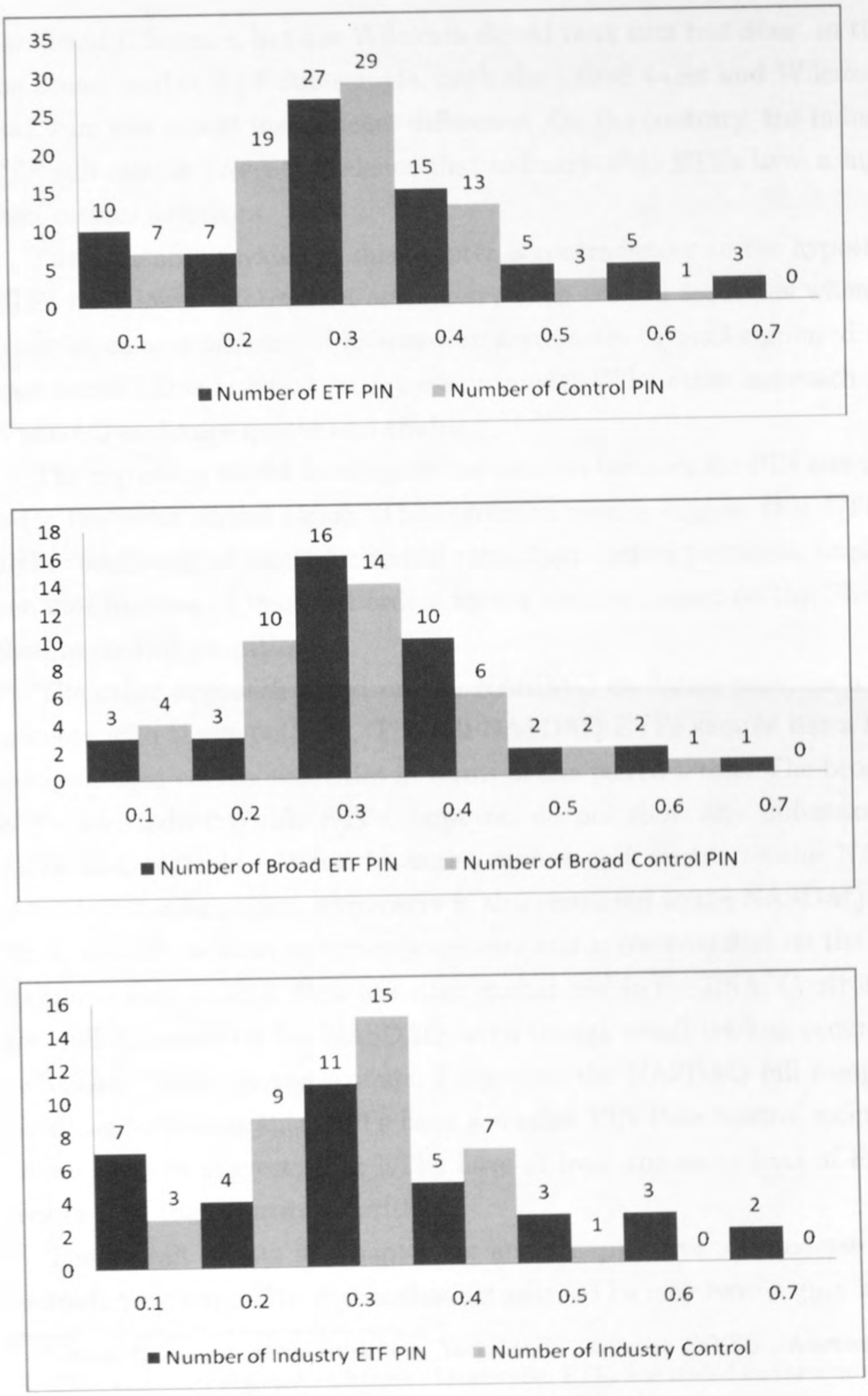


Figure 6.5: The Histogram of ETFs on the NASDAQ

significant difference, but the Wilcoxon signed rank sum test does. In the case of the broad-market ETF sub-sample, both the paired t-test and Wilcoxon signed rank sum test report insignificant difference. On the contrary, the industry-wide ETF sub-sample presents evidence that industry-wide ETFs have a higher PIN than control securities.

The evidence provided in this chapter is contradictory to the hypothesis that ETFs have lower information asymmetry than control securities when the PIN is employed as a measure of information asymmetry. Thus I employed two other approaches. One is based on regression model. The other approach uses only NASDAQ exchange quotes and trades.

The regression model investigates the relation between the PIN and parameter α (or the order arrival ratio). The regression results suggest that ETFs have a higher coefficient of the order arrival ratio than control securities, implying that one unit increase of informed orders have a greater impact on the PIN in ETFs than in control securities.

The other approach based on the NASDAQ exchange provides a case that matches with the hypothesis. The full NASDAQ ETFs sample has a lower PIN than matched control securities in terms of the paired t-test. The broad-market ETFs and industry-wide ETFs, however, do not show any difference between ETFs and controls. When I restrict trades and quotes to the NASDAQ, I assume that information asymmetry is also restricted to the NASDAQ exchange. Thus, any information asymmetry appears and is resolved first on the NASDAQ exchange, even though there are other exchanges[1] in the USA. Control securities are traded mainly on the NASDAQ, even though small trading occurs on other exchanges. With these restraints, I interpret the NASDAQ full sample case as insufficient evidence that ETFs have a smaller PIN than control securities. The overall analysis suggests that ETFs have at least the same level of PIN as or a greater PIN than control securities.

The overall results in Chapter six and Chapter five are understood in two contradictory ways. The first is that actually ETFs may have higher information

[1]Other exchanges include the New York Stock Exchange (NYSE), American Exchange (AMEX), and other regional exchanges. Meanwhile, ETFs are traded on these stock exchanges and the AMEX or NASDAQ are among the main trading places.

asymmetry than control securities. This contention is contradictory to the implications of Subrahmanyam (1991) and Gorton and Pennacchi (1993), because ETFs should have lower information asymmetry if ETFs are basket securities.

The second is that PIN also may not perform very well to measure information asymmetry. This contention may follow the case of three spread decomposition models in chapter five. George et al (1991), Lin et al (1995), and Huang and Stoll (1997) have higher adverse selection costs of ETFs, compared to common stocks.

Note that even though PIN and three spread decomposition models show higher information asymmetry in ETFs than control stocks, This does not mean the rejection of these models. The only thing chapter five and six show is that there is an empirical caveat if these models measure correctly information asymmetry. That empirical caveat is that ETFs should have lower adverse selection costs than control stocks. When any models do not capture this empirical feature, academic researchers need to be careful to employ these models. The reason is that these models already show that they do not capture what they should capture empirically.

The reason that spread decomposition models are discussed first is that spread decomposition models have several variant. For instance, George et al (1991) are originally based on covariance between trades, but Madhavan et al (1997) employ trading direction to capture adverse selection cost. While the PIN also has few variants, the variant PIN models also are based on the same Bayesian approach as the PIN. This makes me to evaluate first spread decomposition models.

A Appendix 6.A

Table 6.A.1: The Parameters of Full ETFs/Controls on the NASDAQ

ε_b and ε_s represent the daily arrival rate of uninformed buy orders and of uninformed sell orders. 'μ' represents the arrival rate of informed traders. 'α' and 'δ 'represent the probability of an information event occurring and the probability of the type of news occurring.
ETF column represents statistics from ETFs and Control column represents statistics from control stocks. Paired difference column is computed as the parameter of ETF minus paired control stock. 72 paired differences are computed and these differences are used for the statistics in the paired difference column. If average (median) paired difference is not close to zero, then the parameter is different between ETFs and control stocks. Negative difference means that the ETF has a smaller PIN than its paired control stock. The value in paired difference represents the average difference and the value in parenthesis is the paired difference test statistic. Paired t-test and Wilcoxon signed-rank sum test are used to test whether paired differences are close to zero.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level.
Range represents the difference between Max and Min values.

Parameters	Statistics	ETF	Control	Paired Difference (ETF minus control)
ε_b	Average	27.64	47.80	-20.16** (-2.24)
	Median	13.56	11.56	-4.16*** (-463)
	Std. Dev.	34.14	88.61	76.34
	Range	157.50	612.53	573.10
ε_s	Average	23.64	41.84	-18.19** (-2.47)
	Median	11.12	13.65	-5.61*** (-545)
	Std. Dev.	33.41	72.36	62.35
	Range	157.17	489.58	461.82
'μ'	Average	27.26	46.39	-19.13*** (-3.54)
	Median	20.46	30.60	-11.06*** (-583)
	Std. Dev.	23.57	47.54	45.81
	Range	110.06	226.96	204.37
'α'	Average	0.42	0.47	-0.0460* (-1.68)
	Median	0.41	0.47	-0.0586** (-351)
	Std. Dev.	0.17	0.16	0.2316
	Range	0.78	0.67	1.2861
'δ'	Average	0.40	0.52	-0.1153*** (-43.93)
	Median	0.44	0.51	-0.1267*** (-648)
	Std. Dev.	0.19	0.15	0.2485
	Range	0.78	0.69	1.2634

Table 6.A.2: The Parameters of Broad-market ETFs/Control on the NASDAQ

ε_b and ε_s represent the daily arrival rate of uninformed buy orders and of uninformed sell orders. 'μ' represents the arrival rate of informed traders. 'α' and 'δ 'represent the probability of an information event occurring and the probability of the type of news occurring.
ETF column represents statistics from ETFs and Control column represents statistics from control stocks. Paired difference column is computed as the parameter of ETF minus paired control stock. 72 paired differences are computed and these differences are used for statistics of paired difference column. If average (median) paired difference is not close to zero, then the parameter is different between ETFs and control stocks. Negative difference means that the ETF has a smaller PIN than its paired control stock. The value in paired difference represents the average difference and the value in parenthesis is the paired difference test statistic. The paired t-test and Wilcoxon signed-rank sum test are used to test whether paired differences are close to zero.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level.
Range represents the difference between Max and Min values.

Parameters	Statistics	Broad-market ETF	Broad-market Control	Paired Difference (ETF minus control)
ε_b	Average	27.19	33.62	-6.42 (-0.82)
	Median	14.00	10.88	-3.97 (-97.5)
	Std. Dev.	33.11	45.46	47.18
	Range	118.75	158.95	242.81
ε_s	Average	20.97	29.14	-8.17 (-1.10)
	Median	9.39	12.23	-3.69** (-158.5)
	Std. Dev.	32.27	40.80	44.85
	Range	129.72	163.14	276.73
'μ'	Average	25.66	46.84	-21.18*** (-3.40)
	Median	19.64	31.92	-13.75*** (-218.5)
	Std. Dev.	23.50	42.83	37.88
	Range	110.06	170.77	197.30
'α'	Average	0.45	0.43	0.0222 (0.5407)
	Median	0.44	0.45	-0.0084 (31.5)
	Std. Dev.	0.19	0.16	0.2497
	Range	0.78	0.56	1.2861
'δ'	Average	0.35	0.54	-0.1878*** (-4.55)
	Median	0.35	0.53	-0.1798*** (-258.5)
	Std. Dev.	0.17	0.16	0.2508
	Range	0.69	0.69	1.2634

Table 6.A.3: The Parameters of Industry-wide ETFs/Control on the NASDAQ

ε_b and ε_s represent the daily arrival rate of uninformed buy orders and of uninformed sell orders. 'μ' represents the arrival rate of informed traders. 'α' and 'δ 'represent the probability of an information event occurring and the probability of the type of news occurring.
ETF column represents statistics from ETFs and Control column represents statistics from control stocks. Paired difference column is computed as the parameter of ETF minus paired control stock. 72 paired differences are computed and these differences are used for statistics of paired difference column. If average (median) paired difference is not close to zero, then the parameter is different between ETFs and control stocks. Negative difference means that the ETF has a smaller PIN than its paired control stock. The value in paired difference represents the average difference and the value in parenthesis is the paired difference test statistic. Paired t-test and Wilcoxon signed-rank sum test are used to test whether paired differences are close to zero.
* represents 10% significant level, ** represents 5% significant level, and *** represents 1% significant level.
Range represents the difference between Max and Min values.

Parameters	Statistics	Industry-wide ETF	Industry-wide Control	Paired Difference (ETF minus control)
ε_b	Average	28.11	62.80	-34.69** (-2.11)
	Median	12.16	19.57	-8.28** (-135)
	Std. Dev.	35.66	117.29	96.90
	Range	157.03	612.46	554.01
ε_s	Average	26.46	55.25	-28.79** (-2.24)
	Median	11.68	14.43	-6.23** (-128)
	Std. Dev.	34.82	93.86	75.91
	Range	157.17	489.41	419.34
'μ'	Average	28.95	45.92	-16.96* (-1.87)
	Median	21.28	21.43	-1.29 (-85)
	Std. Dev.	23.87	52.70	53.43
	Range	85.61	226.93	204.37
'α'	Average	0.39	0.51	-0.1182*** (-3.70)
	Median	0.39	0.52	-0.1116*** (-199)
	Std. Dev.	0.12	0.16	0.188
	Range	0.58	0.62	0.793
'δ'	Average	0.45	0.49	-0.0387 (-1.01)
	Median	0.47	0.47	-0.030 (-55)
	Std. Dev.	0.20	0.13	0.2249
	Range	0.78	0.54	0.8887

Chapter 7

Empirical Investigation of Intra-day Behaviour

7.1 Introduction

One important topic in market microstructure is the intra-day behaviour of trading patterns such as bid-ask spread, return volatility, and trading volume. Extensive research provides evidence that intra-day patterns of trading volume, volatility, and bid-ask spread follow a U-shape. However, these U-shaped patterns vary with market structure and the characteristics of the traded asset. For example, McInish and Wood (1992) find U-shaped spreads on the NYSE, while securities on NASDAQ examined by Chan, Christie, and Schultz (1995) show that bid-ask spreads are high in the morning and become flat then significantly decrease at the market close. Jain and Joh (1988) find U-shaped trading volume, while the London Stock Exchange shows a doubly humped intra-day trading volume (Abhyankar, Ghosh, and Limmack (1997)).

ETFs per se have unique characteristics compared to equity (for details see Chapter 5). For example, ETFs embrace less information asymmetry than ordinary stocks. This informational feature of ETFs is supported by Subrahmanyam (1991), who shows that basket securities such as ETFs have less information asymmetry than the weighted average information asymmetry of constituents stocks. Meanwhile ETFs may face less inventory risk than ordinary stocks. The reason is that any market-maker for an ETF can create (redeem) an

ETF when she/he has insufficient(excessive) inventory, compared to the normal level of inventory. i.e. the market-maker for an ETF does not need to trade the ETF on an exchange to adjust his/her. This creation /redemption feature reduces the inventory risk faced by the ETF market-maker[1].

This chapter investigates whether the unique characteristics of ETFs lead to different intra-day behaviour relative to individual stocks.

7.2 Review of Theoretical Intra-day Behaviour

7.2.1 Information Asymmetry

7.2.1.1 Admati and Pfleiderer (1988)

Admati and Pfleiderer (1988) develop a model in which informed traders and discretionary/non-discretionary liquidity traders concentrate trades. They try to answer three questions: (1) why trading is concentrated in a certain time period within the trading day, (2) why returns are volatile in some periods rather than others, and (3) why the periods of higher trading volume tend to be the periods of higher return volatility.

To answer these questions, Admati and Pfleiderer assume that some liquidity traders behave strategically; i.e. they assume that discretionary liquidity traders can choose their trading time strategically. Non-discretionary liquidity traders do not have any discretion in this respect.

Then Admati and Pfleiderer show that liquidity traders trade aggressively in a specific period when more informed traders are involved in trading. The reason is that when informed traders compete with each other, liquidity traders face smaller trading losses with informed traders. Especially when informed traders have the same information[2], liquidity traders have much smaller trading losses

[1]Since ordinary stocks do not have creation/redemption features, the market-maker for ordinary stocks should trade on exchange to adjust their inventory. This leads that when the inventories of market makers are over-balanced (under-balanced) market-makers post more frequently ask (bid) price.

[2]When informed traders have same information, they tends to trade in a direction. This direction gives liquidity traders some hint about the intrinsic value of the asset and liquidity traders can prevent their loss to informed traders.

to more informed traders. If equilibrium exists, liquidity/informed trading is concentrated since liquidity traders still have smaller trading losses when they trade together. Informed traders also have greater trading losses when liquidity traders trade more. Finally, a specific pattern in volume and price behaviour appears.

Simple Model

Model Set-up A single asset is traded over a span of T time periods. The value of the asset in period T is exogenously decided by

$$\tilde{F} = \bar{F} + \sum_{t=1}^{T} \tilde{\delta}_t \tag{7.1}$$

where $\tilde{\delta}_t$, t= *1, 2, ..., T,* are independently distributed random variables, each having a mean of zero. The payoff $\tilde{F}$ represents the liquidation value of the asset. The innovation $\tilde{\delta}_t$ becomes public knowledge in each period t.

All informed and liquidity traders are risk-neutral. Additionally, n_t informed traders are endowed with private information. A privately informed trader observes $\tilde{\delta}_{t+1} + \tilde{\varepsilon}_t$, where $var(\tilde{\varepsilon}_t) = \phi_t$. However, all traders including liquidity traders will know something about the piece of public information one period after privately informed traders observe it.

Liquidity traders are composed of m number of discretionary traders and unlimited number of non-discretionary traders. Discretionary liquidity traders do not need to satisfy their liquidity demands immediately. Thus, discretionary liquidity traders have demands for shares in some period T' and need to satisfy their demands before period T'', where *T'< T"< T*. A discretionary liquidity trader is allowed to trade once between T' and T". $\tilde{Y}^j$ is the total demand of the j^{th} discretionary liquidity trader.

A market-maker decides the trade price for the asset in each period. She/he is also risk-neutral and earns zero expected profits due to competition.

$\tilde{x}_t^i$ is the i^{th} informed trader's order in period t and $\tilde{y}_t^j$ is the order of the j^{th} discretionary liquidity trader in period t. $\tilde{z}_t$ is the total demand for shares by the non-discretionary liquidity traders in period t. In period t, the market-maker must purchase $\tilde{w}_t = \sum_{i=1}^{n} \tilde{x}_t^i + \sum_{j=1}^{m} \tilde{y}_t^j + \tilde{z}_t$ shares in period t. The price in period t set by the market-maker is based on the history of public information $\tilde{\Delta}_t$ and the history of order flows $\tilde{\Omega}_t$. $\tilde{\Delta}_t$= ($\tilde{\delta}_1$, $\tilde{\delta}_2$, ..., $\tilde{\delta}_t$) represent the history of public information until period t and $\tilde{\Omega}_t$=($\tilde{\omega}_1$, $\tilde{\omega}_2$, ..., $\tilde{\omega}_t$) represent the history of order flows until period t. Then the zero expected profit condition requires

$$\tilde{P}_t = E(\tilde{F}|\tilde{\Delta}_t, \tilde{\Omega}_t) \tag{7.2}$$

Equation (7.2) implies that the market-maker employs a linear pricing function based on $\tilde{\Delta}_t$ and $\tilde{\Omega}_t$. Finally, the random variables ($\tilde{Y}^1$, $\tilde{Y}^2$,..., $\tilde{Y}^m$, $\tilde{z}_1$, $\tilde{z}_2$, ..., $\tilde{z}_{T-1}$, $\tilde{\delta}_1$, $\tilde{\delta}_2$, .., $\tilde{\delta}_T$, $\tilde{\varepsilon}_1$, $\tilde{\varepsilon}_2$, ...,$\tilde{\varepsilon}_{T-1}$) are mutually independent and follow multivariate normal distribution with each variable having a mean of zero.

Equilibrium The past trades of informed traders are independent of future public information ($\tilde{\delta}_{t+1}$, $\tilde{\delta}_{t+2}$, .., $\tilde{\delta}_T$) and the liquidity trading in any period is independent of the liquidity trading in any other period. Thus, the market-maker sets up the price in period t that is equal to the expectation of $\tilde{F}$ conditional on all public information observed in period t plus an adjustment that reflects the information contained in the current order flow $\tilde{\omega}_t$. This relation is shown in the equation (7.3), where λ_t is the reciprocal of the Kyle (1985) market-depth parameter.

$$\begin{aligned} \tilde{P}_t(\tilde{\Delta}_t, \tilde{\Omega}_t) &= E(\tilde{F}|\tilde{\Delta}_t) + \lambda_t \tilde{w}_t \\ &= \bar{F} + \textstyle\sum_{\tau=1}^{t} \tilde{\delta}_\tau + \lambda_t \tilde{w}_t \end{aligned} \tag{7.3}$$

Admati and Pfleiderer show that if the market-maker uses a linear pricing strategy, then in equilibrium each informed trader i submits a market order of $\tilde{x}_t^i = \beta_t^i(\tilde{\delta}_{t+1} + \tilde{\varepsilon}_t)$ at time t, where

$$\beta_t^i = \sqrt{\frac{\Psi_t}{n_t(var(\tilde{\delta}_{t+1}) + \phi_t)}} \tag{7.4}$$

The equilibrium value of λ_t is given by

$$\lambda_t = \frac{var(\tilde{\delta}_{t+1})}{n_t + 1}\sqrt{\frac{n_t}{\Psi_t(var(\tilde{\delta}_{t+1}) + \phi_t)}} \quad (7.5)$$

In equations (7.4) and (7.5), Ψ_t represents the total variance of the liquidity trading in period t: $\Psi_t \equiv var(\sum_{j=1}^{m} \tilde{y}_t^j + \tilde{z}_t)$.

In equilibrium j^{th} discretionary liquidity trader attempts to minimise his/her expected transaction costs subject to meeting his/her liquidity demand $\tilde{Y}^j$. Equation (7.6) measures the transaction cost at time $t \in [T', T'']$ as the difference between the paid cost of the j^{th} discretionary liquidity trader for the security, and the expected value of the security.

$$E\left(\left(P_t(\tilde{\Delta}_t, \tilde{\Omega}_t) - \tilde{F}\right)\tilde{Y}^j|\tilde{\Delta}_t, \tilde{\Omega}_{t-1}, \tilde{Y}^j\right) \quad (7.6)$$

Equation (7.6) can be simplified to $\lambda_t(\tilde{Y}^j)^2$ and suggests that the transaction cost is minimised when λ_t is smallest in the period $t^* \in [T', T'']$.

Implication With these assumptions, Admati and Pfleiderer outline three implications. The first is that equilibrium always exists in which all discretionary liquidity trading occurs in the same period, so called concentrated trading patterns. Additionally, only these concentrated trading equilibria are robust, because the concentrated trading of discretionary liquidity traders minimises the trading cost borne by discretionary liquidity traders even if there exists an equilibrium in which discretionary liquidity traders do not trade in the same period.

The second implication provides three relationships: $V_{t^*}^L > V_t^L$ and $V_{t^*}^I > V_t^I$and $V_{t^*}^M > V_t^M$ for $t \neq t^*$ in an equilibrium in which all discretionary liquidity trading occurs in period t^*. Moreover,

$$V_t^I \equiv \sqrt{var\left(\sum_{i=1}^{n} \tilde{x}_t^i\right)} = \sqrt{var(n_t\beta_t(\tilde{\delta}_{t+1} + \tilde{\varepsilon}_t))} = \sqrt{n_t\Psi_t} \quad (7.7)$$

$$V_t^L \equiv \sum_{j=1}^{m} \sqrt{var(\tilde{y}_t^j)} + \sqrt{var(\tilde{z}_t)} \quad (7.8)$$

$$V_t^M \equiv \sqrt{var(\tilde{w}_t)} \tag{7.9}$$

$$V_t \equiv V_t^I + V_t^L + V_t^M \tag{7.10}$$

The second implication means that volume when discretionary liquidity trading occurs is higher than volume when discretionary liquidity trading does not occur. This is because discretionary liquidity trading promotes an increase in the volume of informed trading indirectly and the volume of non-discretionary liquidity trading directly.

The third implication states that when the number of informed traders is constant over time, the extent to which prices reveal private information is constant and the variance of the price change is constant. When the measure Q_t represents the extent to which prices reveal private information i.e. $Q_t \equiv var(\tilde{\delta}_{t+1}|\tilde{P}_t)$ and the measure R_t represents the variance of the price change i.e. $R_t \equiv var(\tilde{P}_t - \tilde{P}_{t-1})$, two equations are established under the assumption that $n_t = n$ for every t. n is the number of informed traders. The first is that $Q_{t*} = Q_t$ for every t, and the second is that $R_{t*} = R_t = 1$ for every t. These equations suggest that the variance of price changes R_t does not change depending on the number of informed traders. The overall rate at which information comes to the market Q_t is stable under the same number of informed traders. Moreover, R_t and Q_t are independent of the variance of liquidity trading in period t.

The Extension of the Simple Model Admati and Pfleiderer (1988) show the existence of concentrated trading in their simple model. In the concentrated trading period, both informed traders and liquidity traders arrive more frequently. Admati and Pfleiderer extend their simple model to the case of endogenous information acquisition, the case when informed traders observe different signals, the case when discretionary liquidity traders can trade several times within the periods in the interval $[T', T'']$. All these cases provide similar results to the simple model case. That is, the concentrated trading pattern exists and informed traders are higher in the concentrated trading period when liquidity traders are higher.

Admati and Pfleiderer find the same results as the simple model when different timing constraints for liquidity traders, risk-averse liquidity traders, and correlated demands of liquidity traders, are considered.

7.2.1.2 Foster and Viswanathan (1993)

While Admati and Pfleiderer (1988) focus on the concentrated trading of liquidity traders, Foster and Viswanathan (1993) concern themselves with the release of public information, and particularly, with its type. Although public information does not affect expected trading volume and price variability in the Admati and Pfleiderer (1988) model, Foster and Viswanathan (1993) show that the variance of prices and expected trading volume depend on public information released at the start of trading. If public information is different from the expectation of investors, then price volatility and trading volume will increase. With multiple-periods and with long-lived private information, the profits of informed traders vanish due to the competition among informed traders.

The Basic Model The model consists of I risk-neutral informed traders, a market-maker, and many liquidity traders in one-period. v represents the terminal value of an asset and is known to only informed traders. s represents public information and correlates with the terminal value of the asset value and is observed by all traders. The i^{th} informed trade submits an order $x_i(I, s, v)$, which means the function of informed traders, public information, and the terminal value of the asset. Liquidity traders submit in total l quantity of orders that are uncorrelated with v and s. Thus, the net order flow is $y = \sum_{i=1}^{I} x_i(I, s, v) + l$. The market-maker is risk-neutral and sets up prices to make zero expected profits, given the public information and the net order flow.

The terminal value of the asset, the public information, and the liquidity trading follow jointly a multivariate elliptically-contoured class of distribution (ECC):

$$\begin{pmatrix} s \\ v \\ l \end{pmatrix} \sim ECC\left(\begin{pmatrix} s_0 \\ p_0 \\ 0 \end{pmatrix}, \Sigma, \psi(\cdot) \right) \tag{7.11}$$

where $\psi(\cdot)$ is a function that determines the shape of the density.

In the class of elliptically contoured distribution (ECC), the density function has ellipse contour lines. ECC includes the normal distribution, multivariate t distribution, a mixture of normal, and multivariate double exponential distribution. One characteristic of ECC is linear conditional expectations. For example, given a random vector $a = (a_1, a_2)'$, let $b = \alpha' a$ and $c = \gamma' a$. Then $E[b|c] = \lambda c$ for all α and γ if and only if a is a member of the elliptically-contoured class.

Σ in distribution (7,11) represents the unconditional variance-covariance matrix.

$$\Sigma = \begin{pmatrix} \sigma_s^2 & \sigma_{sv} & 0 \\ \sigma_{sv} & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_l^2 \end{pmatrix} \tag{7.12}$$

The function $\psi(\cdot)$ in distribution (7.11) can be any continuous function satisfying the restriction of a characteristic function. The restriction is that liquidity trading is semi-independent of $s - s_0$ and $v - p_0$. The semi-independence is defined by two random variables a and b that are semi-independent if $E[a|b]$= 0 and $E[b|a]$= 0.

Then, consider the conditional distribution of $v - p_0$ and l given the realised public information s. This conditional distribution is an elliptically-contoured distribution.

$$\begin{pmatrix} v - p_0|s \\ l|s \end{pmatrix} \sim ECC \left(\begin{array}{l} \begin{pmatrix} (\sigma_{sv}/\sigma_s^2)(s - s_0) \\ 0 \end{pmatrix}, \\ g\left(\frac{(s-s_0)^2}{\sigma_s^2}\right) \begin{pmatrix} \sigma_v^2 - (\sigma_{sv}^2/\sigma_s^2) & 0 \\ 0 & \sigma_l^2 \end{pmatrix}, \psi(\cdot) \end{array} \right) \tag{7.13}$$

where function $g(\cdot)$ is equal to unity for all values of the public information and depends on function $\psi(\cdot)$.

The market-maker sets the price $p(y, s)$ given the net order flow and the realised public information. Specifically, the market-maker decides the price $p(y, s)$ to satisfy $p(y, s) = E[v|y, s]$. Meanwhile the informed trader submits the order of $x_i(I, s, v)$ to maximise $E[(v - p(y, s))x_i(I, s, v)|v, s]$.

Equilibrium and Implications Then Foster and Viswanathan show that the unique linear Nash equilibrium is symmetric and is of the form

$$\begin{aligned} x_i(I,s,v) &= \beta(v - E[v|s]) \\ p(y,s) &= E[v|s] + \lambda y \end{aligned} \tag{7.14}$$

where $\lambda = \frac{\sqrt{I}}{I+1}\sqrt{\frac{\sigma_v^2 - (\sigma_{sv}^2/\sigma_s^2)}{\sigma_l^2}}$ and $\beta = \frac{1}{\sqrt{I}}\sqrt{\frac{\sigma_l^2}{\sigma_v^2 - (\sigma_{sv}^2/\sigma_s^2)}}$. It is clear that λ and β are independent of the public information and of the particular elliptically-contoured distribution used. However, the trading volume and the informativeness of prices depend on the public information s and the order flow y. The Nash equilibrium is dependent on the unconditional variance-covariance matrix, not on the specific elliptically-contoured distribution used.

The second implication is that $Var[v - p(y,s)|s,y]$ is conditionally heteroskedastic in cases of elliptically-contoured distributions, but is constant in the case of normal distributions. In the case of non-normal contoured distributions, therefore, the conditional variance $Var[v - p(y,s)|s,y]$ differs according to the variation between public information and the unconditional expected values of the public information $|s - s_0|$ and with the absolute net order flow $|y|$.

All expected trading volume from informed traders, liquidity traders, and the market-maker, depends on the public information, and finally total expected trading volume relies on the public information in the case of non-normal elliptically-contoured distributions. Meanwhile, the proportion of trading volumes from informed traders, liquidity traders, and the market-maker, are constant in the compound normal distribution. Therefore, the public signal affects trading volume and is the source of conditional heteroskedasticity. This implies that as public information is implied on the price, trading volume will decrease and higher trading volume may reflect diverse information of informed traders.

Finally, Foster and Viswanathan show that under the assumption that $g(\cdot)$ is monotonic in $|s - s_0|$, the conditional variances of the terminal value and the trading volume given the public information are monotonic in $|s - s_0|$. This implies that a larger difference between the realised and expected public information leads to higher volatility and volume. Thus, higher information shocks induce higher volatility and volume.

The Extension of the Basic Model In the simple model, Foster and Viswanathan show that public information affects the expected trading volume of all traders and the market-maker, and that as public information is realised quite differently from expected, volatility and volume are unexpectedly higher. Meanwhile, all these results are based on the exogenously determined number of informed traders.

In extending their model, Foster and Viswanathan consider that traders can become informed by paying the cost c. Then they show that the conditional variance of the terminal value and the volume brings increases in the size of unexpected public information, and these increases are discontinuous with the number of informed traders.

Foster and Viswanathan consider the case of long-lived information. i.e. before the first trading period, informed traders know the terminal value of the asset that will occur after multiple trading periods. Informed traders can choose one trading period, though the market-maker observes the order of informed traders and adjusts the market price. Thus, current public information includes past prices and volume. Meanwhile, Foster and Viswanathan consider two cases of liquidity demand: the first is that liquidity demands are constant in each subsequent trading period, and the second is that liquidity demands decline in the subsequent trading period.

When the distribution is elliptically contoured, except for normal distribution, trading volume and the conditional variance of the difference between terminal value and market price at t is conditional on the price and the order flow history until $t-1$. This leads to auto-correlated trading volume. However, when the distribution is normal, trading volume is positively auto-correlated. Especially, the first auto-covariance declines with the number of lags used, i.e. $Cov(V_t, V_{t-1}) > Cov(V_{t-1}, V_{t-2}) > 0$ and $Cov(V_t, V_{t-1}) > Cov(V_t, V_{t-2}) > 0$.

Finally, Foster and Viswanathan show that as the trading period is increased, informed traders' total expected profits and order size go to zero. This happens when the number of informed traders is greater than, or equal to, two. However, when liquidity-trading variance in each period is stable and the number of trading periods is increased, informed traders earn finite profits under the competition among informed traders.

7.2.2 Inventory Risk

To the best of my knowledge, no theoretical approach based on inventory risk is made on the issue of intra-day behaviour of bid-ask spread etc. Inventory risk models[1] assume that when market-makers have imbalanced inventory, they adjust these immediately. Amihud and Mendelson (1980), however, provide some brief intuition on the intraday behaviour of bid-ask spreads. Three papers in this section discuss the inventory risk of market-makers empirically. Their findings suggest that market-makers do not always adjust their imbalanced inventory immediately.

Amihud and Mendelson (1980) discuss the inventory risk of market-makers and the bid-ask spread (for details, see Chapter Three). They show that when market-makers are in their preferred inventory level they post smaller bid-ask spreads and place greater bid-ask spreads when their inventory increases.

As trading occurs throughout the day, market-makers accumulate their inventory by the end of the day. Since this higher inventory imbalance is not the preferred level of market-makers, they post greater bid-ask spreads at the market closing. Similarly, Amihud and Medelson address the relationship between the trading volume and inventory level, showing that as market-makers inventory increases, so too does trading volume. Therefore, the behaviour of trading volume during the day can be explained by the inventory risk of market-makers. The intuition that the inventory of market-makers affects bid-ask spreads is supported by the findings of Ho and Macris (1984), in which an AMEX option specialist's trading book is employed. They find that the inventory level of specialist influences the observed transaction returns.

Moreover, empirical findings suggest that market-makers do not adjust their imbalanced inventory immediately. Madhavan and Smidt (1991) set up a linear estimation model to incorporate trading volume and information asymmetry, and

[1] See Chapter Three Stoll (1978a), Amihud and Mendelson (1980), Ho and Stoll (1981),

using transaction data from an NYSE specialist, they find a weak inventory effect. They, however, find a strong information asymmetry effect that leads to higher adverse selection costs. Consistent with Madhavan and Smidt (1991), Hasbrouck and Sofianos (1993) find from NYSE securities that the inventory auto-correlations of some securities are positive and persistent over long lags, when they examine the time horizon of inventory mean reversion. The long lag auto-correlation of inventory suggests that adjusting an imbalanced inventory takes a long time.

Madhavan and Smidt (1993) develop an inter-temporal model of market-maker trades and quotes and find evidence that adjusting the inventory of market-makers is a slow mean reverting process when the specialists' desired inventory is assumed to be constant. When there are unobserved shock to the inventory of the specialists, specialists attempt to revert to the desired inventory level. Previous inventory theory[1] assumes that the change of inventory level happens quickly, implying a fast mean reverting process. However, Madhavan and Smidt (1993) find that the process to revert the desired inventory level takes several days. To reduce the imbalanced inventory by 50% takes on average 7.3 trading days. Their finding suggests that correcting an imbalanced inventory takes moderate periods to adjust.

Madhavan and Sofianos (1998) explain the weak inventory effect by selective timing and the trading direction of market-makers. They collect 200 NYSE securities from the NYSE's Specialist Equity Trade Summary data to examine the role of market-makers. They find that NYSE specialists selectively participate in trading. This finding is different from the assumption of inventory-risk theory models, in that market-makers always take the other side of an external order from informed/uninformed traders, thereby suggesting that market-makers selectively time the magnitude and direction of trades to control inventory. They find that the participation rate of market-makers depends on several factors; the inventory of a market-maker, minimum tick size, signed spread, trade price movement, and trade size. While minimum tick size positively affects the participation rate, other factors have a negative impact on the participation rate.

[1]See section 3.3 in Chapter Three (Stoll (1978b), Amihud and Mendelson (1980), Ho and Stoll (1981), O'Hare and Oldfield (1986))

To sum up, when inventory models[1] assume that market-makers adjust their inventory level immediately, empirical findings suggest that market-makers time their participation in trading and adjust their trading direction to reflect their inventory level and the adjustment of an inventory level takes several days. Additionally no theoretical models based on inventory risk explain the intra-day behaviour of spreads/ return volatility/ trading volume.

7.2.3 Exchange Structure

Madhavan (1992) compares an order-driven market and a quote-driven market. The bid-ask spread, market depth, price efficiency, and price variance are used to measure two different mechanisms.

7.2.3.1 The Quote-Driven Market.

When information asymmetry is less than an upper bound[2], equilibrium exists at time t_i (i=1,2, ..., T). The upper bound depends on the coefficient of risk aversion and the variance of the initial endowment of the risky asset; hence, as these two variables increase the upper bound also goes up.

Madhavan shows that if equilibrium exists, (a) transaction prices follow a martingale and prices are semi-strong form efficient. (b) the predictive errors are positively correlated. (c) the effective spread is an increasing function of order size and the quoted bid-ask spread decreases over the day. Especially the reduction of spreads is due to the decrease of information asymmetry between dealers and traders.

7.2.3.2 The Order-Driven Market

The order-driven market takes two forms: a continuous auction and a periodic auction. In comparing this to the quote-driven market, I discuss the continuous auction. (Madhavan also shows the periodic auction.)

[1] See Chapter three. Garman (1976), Ho and Stoll (1981), and Amihud and Mendelson (1980).

[2] The upper bound is in the form of a function. The function value is changeable depending on variables.

In the continuous auction market, equilibrium exists when information asymmetry is below an upper bound. In addition to the same variables as the quote-driven market, the upper bound of order-driven market depends on the number of dealers.

Thus, the finite number of dealers is a necessary condition for the existence of equilibrium. On equilibrium in the order-driven market, continuous auction prices do not follow a martingale and are not semi-strong form efficient. The unconditional variability of prices in the continuous auction system is higher than in the continuous dealer system. Past trade prices of the continuous order-driven and the quote-driven markets are equally informative. As the number of dealers increases infinitely the equilibrium price-quantity pair for continuous auction converges to the equilibrium price-quantity pair of the quote-driven mechanism.

If dealers can enter freely in market-making, the quoted bid-ask spread increases in the precision[1] of the trader's signal and the precision of risky asset endowments. The quoted spread decreases with the coefficient of risk aversion and the precision of the market-maker's signal. Meanwhile, market depth behaves in the opposite way to the quoted bid-ask spread. Price variability is negatively related to the precision of the trader's signal, the precision of the market-maker's signal, and the precision of risky asset endowments. Price variability increases as the coefficient of risk aversion rises. Finally, a security with high price volatility has narrower bid-ask spreads than a security whose volatility is low.

Therefore, Madhavan suggests that different market structures generate different features on prices, spread, and variability. Overall, a quote-driven market system is more price efficient than a continuous auction system.

7.2.4 Market Closure

There is an argument that periodic market closure generates several intraday patterns. Hong and Wang (2000) argue that periodic market closure affects investors' trading behaviour, leading to different intra-day patterns. They assume that when the market is open two types of trading occur: hedging trading and speculative trading. Hedging trading arises when investors want to rebalance

[1]Precision is the inverse of variance.

their portfolio assets and speculative trading happens because some investors have private information on the future stock payoffs.

These two types of trading have different impacts when the market is closed. Hedge traders face a higher risk of holding stocks over market closure because they cannot rebalance their portfolio. When the market is closed there is no market price, which is a source of private information. Thus, market closure generates time variation in information asymmetry among traders.

When hedging trading dominates trading activity, both the mean and volatility of stock returns decrease over time during trading periods. When time-varying information asymmetry dominates trading activity, the mean and volatility of stock returns increase over trading periods. The mean return and return volatility are U-shaped when hedging trading dominates at the market open, and time-varying information asymmetry dominates at the market close. The information accumulation during the market closure generates higher trading volume at the market open and the intention of hedge traders to reduce their hedge positions at the market closing generates higher trading volume at the close. This explains the U-shaped trading volume during a trading day.

French and Roll (1986) find that during market closure private information is accumulated. When the market is open private information generates higher returns volatility. They also find that pricing errors make stock returns more volatile during trading hours than during market closure. However, they find that public information makes no difference in volatility between market opening and market closure.

7.3 Review of Empirical Intra-day Behaviour

An extensive literature provides empirical evidence about intra-day patterns. The studies focus on two questions: what kinds of patterns are prominent in the variables representing intra-day behaviour, and whether different markets show diverse intra-day patterns. For example, McInish and Wood (1992) and McInish and Van Ness (2002) address the first type of question. Werner and Kleidon (1996) concern themselves with the second.

7.3.1 Empirical Intra-day Patterns on the NYSE

McInish and Wood (1992) investigated the intra-day behaviour of a time-weighted bid-ask spread, discovering several patterns. The time-weighted bid-ask spreads show U-shaped intraday patterns after some control variables are included in the regression[1]. In addition to the U-shaped spreads, the relationship between time-weighted bid-ask spreads and control variables are as follows: spreads are inversely related to trading activity and the level of competition, but positively related to the level of risk and the amount of information coming to the market[2]. McInish and Van Ness (2002) test intra-day behaviour using different periods and the same methodology. They find similar U-shaped spreads to McInish and Woods (1992).

Jain and Joh (1988) investigate the intra-day distribution of hourly common stock index returns and trading volume, using data consisting of hourly NYSE common stock trading volume and returns. They find that trading volume is highest in the first hour of each day and then decreases monotonically. Toward the end of a trading day, trading volume increases again. Thus, they find that trading volume is U-shaped throughout a day. Furthermore, the pattern of hourly return is similar to that of trading volume. The hourly returns are usually higher during the first and last trading hours than in other trading hours. The U-shaped pattern of trading volume in a day is confirmed by Gerety and Mulherin (1992), who find that trading volume during the closing hour is positively related to expected overnight volatility, while the opening hour is positively related to both expected and unexpected volatilities of the previous night.

Wood, McInish, and Ord (1985) empirically investigate the intra-day behaviour of trade returns, and the standard deviation of returns. They find that trade returns and the standard deviation of returns are high at the opening and the end of the NYSE. Lockwood and Linn (1990) investigate the intra-day behaviour of hourly return variance, using the Dow Jones Industrial Average

[1]In the regression, time-weighted bid-ask spreads are dependent variable. Seven control variables and 12 30-minutes time interval variables are independent variables.

[2]For the trading activity, McInish and Wood (1992) employ the number of trades and the average trade size. The standard deviation of time-weighted spread is used as a risk measure. Normalised trade size is considered as a measure of information. Competition between exchanges is reflected by a regional exchange variable.

(DJIA) returns between 1964 and 1989. They find that the hourly variance is highest during the opening hour and then decreases until early afternoon, after which the hourly variance starts to increase. i.e. hourly variance follows a U-shaped pattern during the trading day.

Lee, Mucklow, and Ready (1993) investigate the intra-day behaviour of spreads, depths, and volume. They find that after firms announce earnings, bid-ask spreads are wider, but when volume is controlled for, the wide spreads do not appear. Specifically, spreads increase in the half-hour containing the earnings announcement, and then remain wider for up to one day, while quoted depths return to non-announcement levels after three hours. Wide spreads and low depths are insignificant when volume increase is considered. Another finding is that higher volume leads to low depths, and lower depth leads to wide spreads.

7.3.2 Empirical Intra-day Patterns on Non-NYSE Exchanges

The examination of the NYSE (section 7.3.1) suggests that many intra-day variables including trading volume and return volatility, display a U-shaped pattern. A naturally-arising question is, whether the U-shaped intra-day pattern is due to either the characteristics of the NYSE, or private information, as Admati and Pfleiderer (1988) consider[1]. Thus, other empirical research explores the intra-day behaviour in other exchanges to see whether the same U-shaped intra-day pattern appears in exchanges that have different institutional characteristics to the NYSE.

Chan, Christie, and Schultz (1995) investigate the intra-day behaviour of securities on the NASDAQ, finding that the bid-ask spread for NASDAQ securities is relatively stable throughout the day after the peak of early morning trading. Additionally, the spread narrows significantly during the closing hours. They argue that the NASDAQ bid-ask spread near market closing is ascribed to the inventory control of the market-makers (dealers on the NASDAQ). Contrary

[1]Note that the explanation of Foster and Viswanathan (1993) is an L-shaped intra-day pattern because long-lived private information causes trading to happen more in early morning. After then, private information is pounded in the trading price and trading volume is stable.

to the pattern of bid-ask spread, they find that return volatility and share volume follow a U-shaped intra-day pattern. They argue that if return volatility and share volume reflect the differential information between informed traders and uninformed traders, the intra-day pattern of bid-ask spread on the NASDAQ is not related to the private information held by informed traders. The selective participation of dealers on the NASDAQ at market closing is consistent with the findings of Madhavan and Sofianos (1998), in which NYSE specialists also selectively participate in trading.

Consistent with Chan, Christie, and Schultz (1995), Chung and Zhao (2003) investigate the intra-day pattern of NASDAQ securities. They raise the question of whether any different intra-day variation occurs on the NYSE and NASDAQ after the Limit Order Display Rule[1] is adopted. Firstly they examine the intra-day variation of the bid-ask spreads of NYSE/NSADAQ securities. They confirm that the intra-day bid-ask spread of the NYSE follows a U-shape and that NASDAQ securities have the same intra-day variation as Chan, Christie, and Schultz (1995). This intra-day variation of spreads is evident before US SEC adopted the Limit Order Display Rule on NASDAQ in 1997. Since the Limit Order Display Rule became effective, the intra-day pattern of NYSE and NASDAQ securities has been similar. i.e. NASDAQ standardised spreads[2] are widest at the market open, narrow sharply during the mid-day, and become stable until the market close. Meanwhile NYSE standardised spreads are wide at the open, narrow during the day, and increase slightly during the last thirty minutes trading. NASDAQ raw spreads, however, remain in a similar intra-day variation as before the rule, showing the greater spreads in the morning and the lowest spreads at the closing.

Chan, Chung, and Johnson (1995) investigate the intra-day pattern that appears on the Chicago Board Options Exchange (CBOE) puts and calls, and on their NYSE-traded underlying stocks. They find that underlying stocks display

[1]The Limit Order Display Rule is part of the Order-handling Rule on NASDAQ that is enacted by U.S. Securities and Exchange Commission (US SEC). The rule requires that limit orders be displayed in NASDAQ best bid and ask when they are better than quotes posted by market-makers.

[2]Standardised spread= (the quoted spread - the mean of the quoted spread) / the standard deviation of the quoted spread.

U-shaped intra-day bid-ask spreads, while call and put options show a different intra-day pattern. The bid-ask spreads of options are high at the market open and then they narrow sharply to level off at the market close. However, the volatility and trading volume from underlying stocks on the NYSE is indifferent from that of options on the CBOE. i.e. the volatility and trading volume have U-shaped intra-day patterns. The U-shaped intra-day pattern of volatility is consistent with the finding of Sheikh and Ronn (1994), who examined the intra-day behaviour of returns on the CBOE

Ekman (1992) investigates the intra-day pattern of the S&P 500 index futures market to see whether the pattern is similar to that of the NYSE in terms of returns, volatility, and the number of traders. He finds that returns and the number of trades show a rough U-shaped intra-day pattern, that is to say, at the market opening, returns and the number of trades are highest, after which they diminish until mid-day and then increase again. The difference from the NYSE is that the number of trades and returns are slightly decreased at the closing of the market. However, Ekman finds that volatility has a U-shaped intra-day pattern.

Abhyankar, Ghosh, and Limmack (1997) focus on the London Stock Exchange (LSE) to investigate the intra-day pattern of variables. They examine whether 835 equities listed on the LSE show the same intra-day bid-ask spread, trading volume, and return volatility patterns as those on the NYSE. They find that the bid-ask spread and return volatility of equities on the LSE show a U-shaped intra-day pattern. Meanwhile trading volume has a double-humped intra-day pattern. i.e. trading volume is high at 9:30 am and at 4:00 pm. Contrary to Abhyankar, et al. (1997), Cai, Hudson, and Keasey (2004) find a similar intra-day bid-ask spread pattern to that of the NASDAQ when they investigate equities on the London Stock Exchange during the period from March to May 2001. i.e. Cai et al. (2004) find that the bid-ask spread is high at the opening and then decreases to stable status for the rest of the day. Meanwhile, Cai et al. (2004) find that trading volume and volatility have the same intra-day pattern as Abhyankar, et al. (1997). i.e. intra-day volume shows a two-humped pattern and volatility has a U-shaped pattern.

Werner and Kleidon (1996) compare intra-day patterns between British stocks cross-listed on both the UK and US stock markets. Consistent with Abhyankar, et al. (1997) and Cai, et al. (2004)[1], Werner and Kleidon find that trading volume on the LSE has a two-humped intra-day pattern and the volatilities of quote price and trade price display a rough U-shaped intra-day pattern. Additionally, the bid-ask spreads on the LSE are high in the morning and continue to decrease until the end of trading, thus making the bid-ask spread at the closing, the lowest. However, the spreads for British stocks cross-listed on the NYSE contrast with the U-shaped spreads for ordinary stocks on the NYSE. British stocks cross-listed on the NYSE have a greater spread at the opening and then decline over the course of the trading day.

Ahn and Cheng (1999) investigate bid-ask spread patterns and depth on the Hong Kong Stock Exchange. This Exchange employs a limit order-driven market[2]. Although the Hong Kong Stock Exchange uses a different trading structure from that of the NYSE, its intra-day pattern is similar to that of the NYSE. i.e. the bid-ask spread and trading volume of the Hong Kong Stock Exchange has a U-shaped pattern, while depth has a reverse U-shaped intra-day pattern.

Andersen, Bollerslev, and Cai (2000) use the Nikki 225 index return to examine whether intra-day volatility exhibits a U-shaped pattern. The Nikki 225 index is a price-weighted index and consists of 225 firms in the first section of the Tokyo Stock Exchange. The selection of a constituent company is subject to certain industry-balance considerations. Nikki 225 index returns display a doubly U-shaped pattern that occurs because the Tokyo Stock Exchange closes for lunch from 11:00 am to 12:30 pm, thereby creating two sessions, a morning and afternoon one. These two different sessions lead to two distinct U-shaped patterns of intra-day volatility.

7.3.3 Market Structure and Stock Returns

Amihud and Mendelson (1987) examine the effects of trading mechanism, employing close-to-close returns and open-to-open returns to compare the

[1] They discern SETS securities and non-SETS securities. They find a reverse J-shaped bid-ask spread pattern on SETS securities and a declining bid-ask spread pattern on non-SETS securities. But both securities show a two humped pattern of trading volume.

[2] NYSE (specialists) and NASDAQ (dealers) use a quote-driven market

periodic call auction market and continuous quotes-driven market. Open-to-open returns come from periodic call auctions and close-to-close returns represent the continuous quote-driven market. This is because the NYSE adopts a periodic call auction mechanism to decide the opening trading price and a continuous quote-driven mechanism to decide the closing trading price. They find that open-to-open returns have higher volatility than close-to-close returns, and that open-to-open returns have a stronger negative dependence on past returns than close-to-close returns. When Fama's market efficiency hypothesis[1] holds (there is no serial correlation of returns), open-to-open returns are highly likely to deviate from the market efficiency hypothesis.

Stoll and Whaley (1990) confirm the finding of Amihud and Mendelson (1986) in that the ratio of variance of open-to-open returns to close-to-close returns is greater than one. Additionally, Stoll and Whaley find that the daytime returns are likely to reverse the previous overnight returns, implying the temporary price deviation at the opening. Meanwhile, the closing price is less likely to deviate to the temporary price since the overnight return is less likely to reverse the previous daytime returns. The temporary deviation of prices at the opening suggests that the implied cost of immediacy at the open, which is provided by market-makers, is significantly higher than at the close.

7.4 Methodology for Intra-day Analysis

The methodology employed by McInish and Van Ness (2002) and McInish and Wood (1992), is used to investigate the intra-day pattern of ETFs. McInish and Wood (1992) and McInish and Van Ness (2002) employ linear estimation that regresses the time-weighted bid-ask spread on a variety of independent variables (control variables and interval dummy variables). They divide the trading period of a day into 13 intervals, each being composed of 30 minutes. The advantage of using this methodology is that whilst controlling the relationship between the independent variables and time-weighted bid-ask spread, the intra-day variation

[1]Market efficiency hypothesis means that stock prices full reflect all available information at time.

can be examined by looking at the interval coefficients. In addition, the Generalized Method of Moment (GMM) is employed to estimate each coefficient.

7.4.1 Dependent Variable

7.4.1.1 Time-weighted Bid-ask Spread

The time-weighted bid-ask spread is based on the proportional quote bid-ask spread, which is (Bid-Ask)/Mid-quote price. The mid-quote price is the middle price between the bid and the ask prices. In other words

$$BAS = \frac{Ask - Bid}{(Ask + Bid)/2} \tag{7.15}$$

Meanwhile, the time-weight is computed as follows:

$$wt_j = \frac{t_{j+1} - t_j}{T' - T} \tag{7.16}$$

In an interval (T, T'), N quotations occur at time t_j, j=1,2, ... , N and the proportional bid-ask spreads BAS_j at time t_j are matched with N quotations. Meanwhile BAS_0 and BAS_{N+1} are the quotation when $t_0 = T$ and $t_{N+1} = T'$, respectively.

In the equation (7,16), the numerator represents the time difference between the current quote time and the following quote time. i.e. the numerator measures how long the current quote lasts. The denominator represents the time for each interval. The first period is less than 30 minutes because there are no standing quotes for the first period. i.e. the first trading interval is from the first quote time to 10:00 am. The first quote for each ETF appears at a different time in each day. Thus, the first trading interval varies with ETFs and the trading day. Except for the first trading interval, the other 12 trading intervals hold for 30 minutes. i.e. the second trading interval is from 10:00 am to 10:30 am, and the last (13th) interval is from 15:30 pm to 16:00 pm.

Finally, the time-weighted average BAS is computed by

$$\sum_{i=0}^{N} \frac{(t_{i+1} - t_i)}{(T' - T)} BAS_i \tag{7.17}$$

In the equation (7.17), the first interval starts with t_1 because BAS_0 does not exist when the market is open. The other 12 intervals start with t_0.

7.4.1.2 Return Volatility

Return volatility is based on the standard deviation of return, and the return is computed from the transaction / mid-quote prices. The mid-quote price is the middle price between the bid and the ask prices.

$$\text{Return}_{t+1} = \frac{\text{Trade Price}_{t+1} - \text{Trade Price}_t}{(\text{Ask}_t + \text{Bid}_t)/2} \tag{7.18}$$

Return volatility is the standard deviation of this return.

$$\text{Return Volatility}_t = \text{Standard deviation}(\text{Return}_t) \tag{7.19}$$

7.4.1.3 Trading Volume

Trading volume is the summation of each trading volume during each 30 minute interval.

$$\text{Trading Volume}_i = \sum_{t=0}^{t=N} TV_{t,i} \tag{7.20}$$

where subscript i represents each 30 minute interval and subscript t represents transaction t during the interval i. TV is a trading volume at time t.

7.4.2 The Model of Intra-day Analysis

The time-weighted BAS, Return Volatility, and Trading volume are dependent variables. Trading activity variables, Risk variables, Information variables and regional variables, are used as control variables. Model (7.21) is about time-weighted bid-ask spreads and includes all control variables discussed in Appendix 7.D.

Model (7.22) is about return volatility and employs some control variables from Model (7.21). As the number of trades increases, return volatility decreases, since higher trading frequency leads to smaller change of return volatility. As trade size increases, return is less likely volatile. Since regional variable represents the frequency of trading on other exchanges, the higher occurrence of trading on other exchanges leads to the lower trading occurrence of AMEX/NASDAQ, implying the lower return volatility of AMEX/NASDAQ. The Nsize variable is positively related with return volatility. Since Nsize represents the arrival of public information, as Nsize increases return volatility rises (Appendix 7.D).

Trading volume regression (7.23) includes three control variables: Size, Regional, and Price. As trade size increases, trading volume rises. The Regional variable has a positive relation with trading volume. As more trades occur on other exchanges, the AMEX and NASDAQ also have more trades (Appendix 7.D).

Meanwhile return volatility regression and trading volume regression results do not depend on which control variable is included in the regression model. (Appendix 7.C.) In the return volatility regression, the intra-day variation of return volatility always shows the same pattern without regarding to control variables. In the case of trading volume regression, the inclusion of the trades variable decides the intra-day variation of trading volume, because the trades variable is an alternative measure to trading volume. Other control variables do not have an impact on the intra-day variation of trading volume[1].

Finally, Hansen (1982)'s Generalised Method of Moments (GMM) is used to estimate each coefficient. The basic model is formulated as:

$$\begin{aligned} BAS_{k,i} = b_0 &+ b_1 Trades_{k,i} + b_2 Size_{k,i} + b_3 Risk1_{k,i} + b_4 Risk2_{k,i} \\ &+ b_5 Nsize_{k,i} + b_6 Regional_{k,i} + b_7 Price_{k,i} \\ &+ \sum_{i=1}^{9} \delta_i D_i + \sum_{i=11}^{13} \delta_i D_i \\ &+ e_{k,i} \end{aligned} \tag{7.21}$$

[1]The regression results depending on different control variables are available on Appendix 7-C.

$$\begin{aligned} Return\ \ Volatility_{k,i} = b_0 + b_1 Trades_{k,i} + b_2 Size_{k,i} \\ +b_5 Nsize_{k,i} + b_6 Regional_{k,i} + b_7 Price_{k,i} \\ +\sum_{i=1}^{9} \delta_i D_i + \sum_{i=11}^{13} \delta_i D_i \\ +e_{k,i} \end{aligned} \tag{7.22}$$

$$\begin{aligned} Trading\ \ Volume_{k,i} = b_0 + b_2 Size_{k,i} + b_6 Regional_{k,i} + b_7 Price_{k,i} \\ +\sum_{i=1}^{9} \delta_i D_i + \sum_{i=11}^{13} \delta_i D_i \\ +e_{k,i} \end{aligned} \tag{7.23}$$

The *Price* variable is in the models (7.21), (7.22) and (7.23) because the level of price affects the size of the bid-ask spread, return volatility, and trading volume. Stoll (1978b) shows that when a stock's price is high, the bid-ask spread of the stock is smaller. (For details, see Chapter Three.) Additionally, the higher price is related to lower return volatility and lower trade volume. Interval dummy variables represent the time interval. i.e. an interval dummy variable of period 1 represents the time between the first quoting time and 10:00 am, and the variable for period 2 is the time between 10:00 am and 10:30 am. The final (13^{th}) interval dummy variable is the time between 3:30 pm and 4:00 pm. In addition, the intercept b_0 represents the interval dummy variable of period 10, which is from 2:00 pm to 2:30 pm. Since the interval variables are dummy variables, each has the value one when trading time falls into the period of each interval dummy variable. Otherwise, the interval variable is zero. Meanwhile $e_{k,i}$ represents the error term.

7.4.3 Control Variables

Appendix 7.D. describes the control variables in the regression models and what control variables are employed.

7.5 The Summary Statistics of ETFs based on Exchange

This section provides the summary statistics of three dependent variables on the AMEX and NASDAQ. I do not consider quotes and trades from the NYSE because these represent a very small fraction of the full sample. Meanwhile, the trading protocol on the AMEX is similar to that on the NYSE, and dissimilar to that on the NASDAQ. The NYSE and AMEX employ a specialist system, while the NASDAQ uses a multiple dealer system. i.e. on the NYSE and AMEX, there is a specialist who manages a stock to provide liquidity for that stock. However, the NASDAQ has multiple dealers who can supply a stock with liquidity. Therefore, using quotes and trades from only the AMEX can explain how the different market structure between the AMEX/NYSE and NASDAQ affects the intra-day behaviour of the bid-ask spread, if at all.

Table 7.1 shows the summary statistics of the bid-ask spread depending on the AMEX and NASDAQ. These show a decreasing pattern. i.e. the time-weighted bid-ask spread is highest at the market opening, and then starts to decrease until the end of the day, when it reaches its lowest point. In particular, the first time interval (9:30 to 10:00) has a higher time-weighted bid-ask spread than daily spread

Wilk's lambda test and the Kruskal-Wallis test examine the equality of interval means. The assumption of the Wilk's lambda is that the mean time-weighted bid-ask spreads of each interval are equal. The hypothesis being tested by the Kruskal-Wallis test is that median time-weighted bid-ask spreads of each interval are equal. Both test statistics suggest the rejection of the hypotheses. This implies that not all intervals have same time-weighted bid-ask spreads.

Paired T-test and Kruskal-Wallis test in the bottom row test whether AMEX and NASDAQ have same time-weighted bid-ask spreads in each time interval. The test results suggest that there are differences between AMEX and NASDAQ, even though the overall variation pattern of NASDAQ spreads is similar to the 1999 NASDAQ raw spreads of Chung and Zhao (2003). The result that AMEX spreads show an L-shaped pattern is inconsistent with the U-shaped spreads that the NYSE/AMEX presume to generate in prior studies of individual securities.

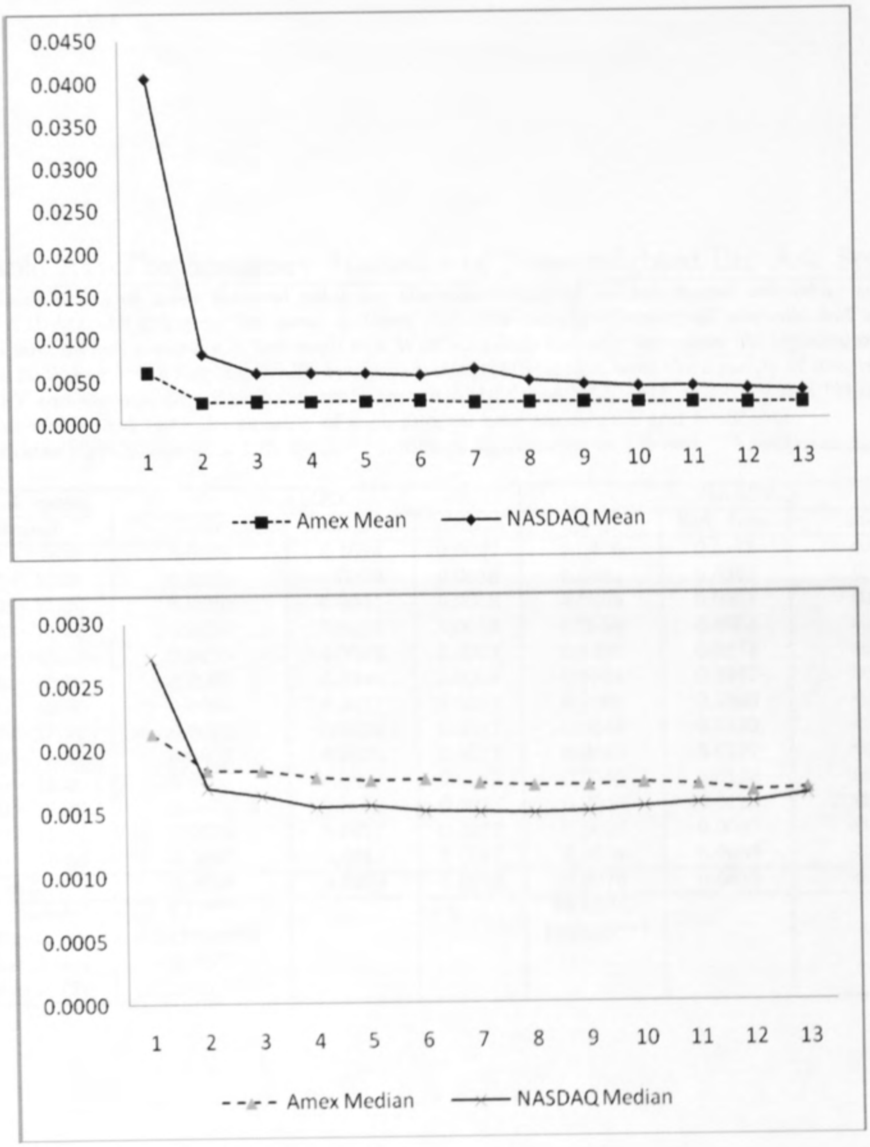


Figure 7.1: Time-weighted Bid-ask Spread

Table 7.1: The Summary Statistics of Time-weighted Bid-Ask Spread

This table provides mean interval value for the time-weighted bid-ask spread according to AMEX and NASDAQ. All refers to the mean of these variables calculated using all intervals and all ETFs. Wilk's lambda test statistics is the result of a Wilk's lambda test that examines the equality of interval values. K-W test is the Kruskal-Wallis non-parametric statistic that tests the equality of interval values. Paired-T test the equality of each interval between AMEX and NASDAQ, and K-W test (2) represents Kruskal-Wallis that tests the equality of each interval between AMEX and NASDAQ.
'*' indicates significance at a 10% level, ** indicates significance at 5% and *** indicates significance at 1%.

Bid-ask spread	AMEX			NASDAQ		
Interval	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
9:30 – 10:00	0.0060	0.1024	0.0021	0.0406	0.2119	0.0027
10:00 – 10:30	0.0025	0.0029	0.0018	0.0081	0.0369	0.0017
10:30 – 11:00	0.0026	0.0041	0.0018	0.0058	0.0237	0.0016
11:00 – 11:30	0.0025	0.0037	0.0018	0.0058	0.0253	0.0015
11:30 – 12:00	0.0025	0.0035	0.0018	0.0057	0.0274	0.0016
12:00 – 12:30	0.0025	0.0040	0.0018	0.0054	0.0227	0.0015
12:30 – 13:00	0.0023	0.0027	0.0017	0.0062	0.0390	0.0015
13:00 – 13:30	0.0022	0.0026	0.0017	0.0048	0.0193	0.0015
13:30 – 14:00	0.0022	0.0024	0.0017	0.0043	0.0127	0.0015
14:00 – 14:30	0.0022	0.0022	0.0017	0.0040	0.0114	0.0015
14:30 – 15:00	0.0021	0.0021	0.0017	0.0040	0.0123	0.0016
15:00 – 15:30	0.0020	0.0017	0.0017	0.0036	0.0097	0.0016
15:30 – 16:00	0.0020	0.0019	0.0017	0.0034	0.0083	0.0016
All	0.0026	0.0273	0.0018	0.0076	0.0613	0.0016
Wilk's Lambda	5.17***			88.59***		
K-W test (1)	517.95***			1053.63***		
Paired-T test	-14.73***					
K-W test (2)	4912240.82***					

Table 7.2 shows the summary statistics of return volatility depending on the AMEX and NASDAQ. These show a crudely U-shaped pattern. i.e. return volatility is highest at the market opening, and then starts to decrease in the middle of the day to reach its lowest point, after which it increases at the market closing. NASDAQ and AMEX show a similar pattern.

The highest return volatility is 0.0011 for the AMEX mean. In the case of the NASDAQ, the highest return volatility is 0.0013. In terms of median, the AMEX and NASDAQ have the highest return volatility, 0.0009. The time interval up to 11:00 has higher return volatility, compared to the daily mean and median.

Wilk's lambda test and Kruskal-Wallis test examine whether interval means are equal. The hypothesis for Wilk's lambda is that the mean return volatilities of each interval are equal. The hypothesis for Kruskal-Wallis test is that median return volatilities of each interval are equal. Both test statistics suggest the rejection of hypothesis. This implies that not all intervals have same return volatilities.

Paired T-test and Kruskal-Wallis test (2) in the bottom row test whether AMEX and NASDAQ have the same return volatilities in each time interval. The test results suggest that there are differences between AMEX and NASDAQ.

Table 7.3 shows the summary statistics of trading volume depending on the AMEX and NASDAQ. The average trading volume of the NASDAQ/AMEX reports a U-shaped pattern. In addition, the median trading volume of the NASDAQ/AMEX shows a U-shaped pattern. The AMEX mean trading volume shows less curvature than the NASDAQ mean trading volume, while the AMEX median trading volume and NASDAQ median trading volume display a similar curvature.

The AMEX has a lower intra-day mean trading volume and a higher intra-day median trading volume than the NASDAQ. These facts suggest that the AMEX trading volume has less extreme values than the NASDAQ trading volume.

Wilk's lambda test and Kruskal-Wallis test examine the equality of interval means. The hypothesis for Wilk's lambda is that the mean trading volumes of each interval are equal. The hypothesis for Kruskal-Wallis test is that median trading volumes of each interval are equal. Both test statistics suggest the

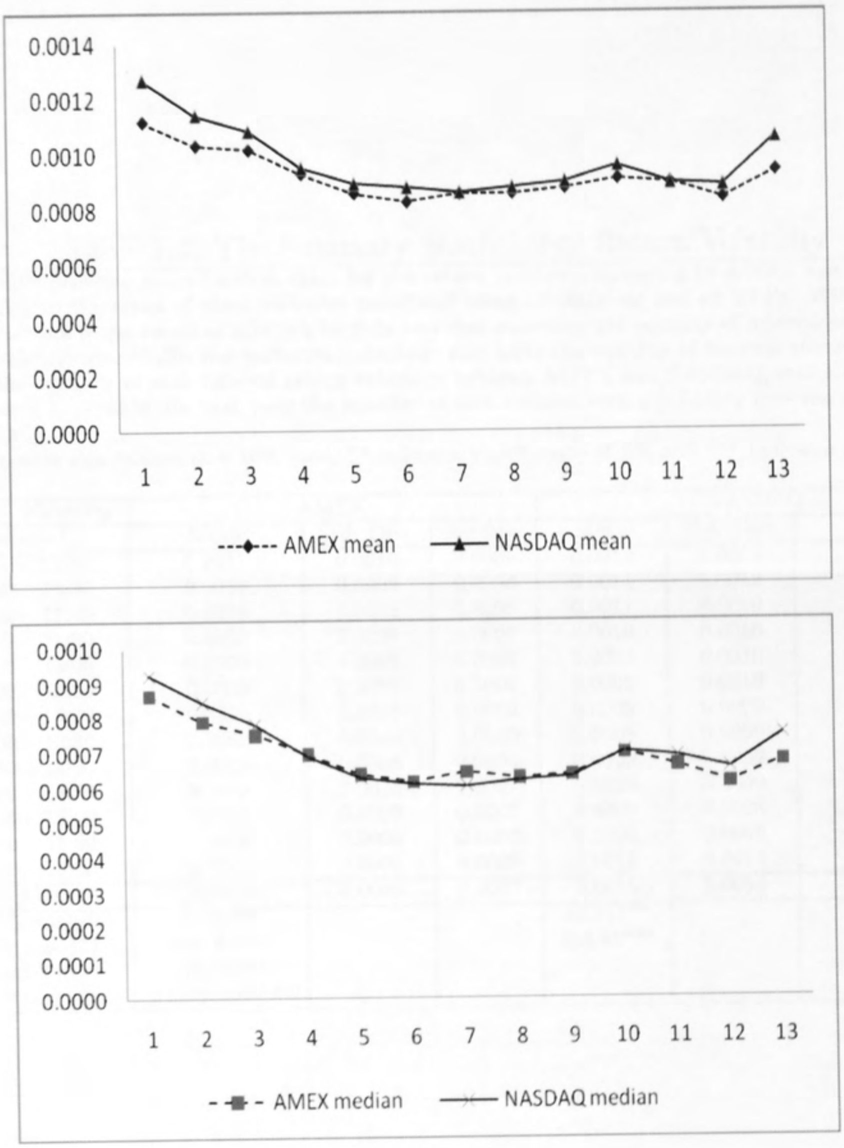


Figure 7.2: Return Volatility

Table 7.2: The Summary Statistics of Return Volatility

This table provides mean interval value for the return volatility according to AMEX and NASDAQ. All refers to the mean of these variables calculated using all intervals and all ETFs. Wilk's lambda test statistics is the result of a Wilk's lambda test that examines the equality of interval values. K-W test is the Kruskal-Wallis non-parametric statistic that tests the equality of interval values. Paired-T tests the equality of each interval return volatility between AMEX and NASDAQ, and K-W test (2) represents Kruskal-Wallis that tests the equality of each interval return volatility between AMEX and NASDAQ.
'*' indicates significance at a 10% level, ** indicates significance at 5% and *** indicates significance at 1%.

Return Volatility	AMEX			NASDAQ		
Interval	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
9:30 – 10:00	0.0011	0.0010	0.0009	0.0013	0.0013	0.0009
10:00 – 10:30	0.0010	0.0009	0.0008	0.0011	0.0012	0.0008
10:30 – 11:00	0.0010	0.0010	0.0008	0.0011	0.0010	0.0008
11:00 – 11:30	0.0009	0.0009	0.0007	0.0010	0.0010	0.0007
11:30 – 12:00	0.0009	0.0008	0.0006	0.0009	0.0010	0.0006
12:00 – 12:30	0.0008	0.0008	0.0006	0.0009	0.0010	0.0006
12:30 – 13:00	0.0009	0.0008	0.0006	0.0009	0.0009	0.0006
13:00 – 13:30	0.0009	0.0008	0.0006	0.0009	0.0009	0.0006
13:30 – 14:00	0.0009	0.0008	0.0006	0.0009	0.0009	0.0006
14:00 – 14:30	0.0009	0.0008	0.0007	0.0010	0.0009	0.0007
14:30 – 15:00	0.0009	0.0009	0.0007	0.0009	0.0008	0.0007
15:00 – 15:30	0.0008	0.0008	0.0006	0.0009	0.0009	0.0007
15:30 – 16:00	0.0009	0.0009	0.0007	0.0011	0.0012	0.0008
All	0.0009	0.0009	0.0007	0.0010	0.0010	0.0007
Wilk's lambda	31.73***			42.71***		
K-W test (1)	689.91***			633.85***		
Paired-T test	-20.75***					
K-W test (2)	3217870.66***					

rejection of hypothesis. This implies that not all intervals have same trading volumes.

Paired T-test and Kruskal-Wallis test in the bottom row suggest that there are differences between AMEX and NASDAQ. This difference is clear in the Figure 7.3.

Table 7.3: The Summary Statistics of Trading Volume

This table provides mean interval value for the trading volume according to AMEX and NASDAQ. All refers to the mean of these variables calculated using all intervals and all ETFs. Wilk's lambda test statistics is the result of a Wilk's lambda test that examines the equality of interval values. K-W test is the the Kruskal-Wallis non-parametric statistic that tests the equality of interval values. Paired-T tests the equality of each interval trading volume between AMEX and NASDAQ, and K-W test (2) represents Kruskal-Wallis that tests the equality of each interval trading volume between AMEX and NASDAQ.
'*' indicates significance at a 10% level, ** indicates significance at 5% and *** indicates significance at 1%.

Trading Volume	AMEX			NASDAQ		
Interval	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
9:30 – 10:00	24592	77879	3200	163254	588151	2800
10:00 – 10:30	19543	63632	2900	117201	443292	2300
10:30 – 11:00	17211	56165	2300	93738	377175	1900
11:00 – 11:30	12891	39605	2100	71858	292052	1700
11:30 – 12:00	11657	34626	1800	60772	283384	1500
12:00 – 12:30	11170	37208	1600	55907	244555	1400
12:30 – 13:00	9445	31718	1500	54375	250402	1200
13:00 – 13:30	10617	67823	1700	49648	205850	1300
13:30 – 14:00	9811	35540	1700	51226	228877	1200
14:00 – 14:30	12821	43100	2000	73069	392884	1500
14:30 – 15:00	12533	39971	1900	67113	293908	1500
15:00 – 15:30	15372	47349	2600	70785	319697	1700
15:30 – 16:00	22167	69182	2900	95651	404472	2100
All	14640	51874	2100	78639	348499	1600
Wilk's lambda	33.93***			29.08***		
K-W test(1)	679.19***			519.91***		
Paired-T test	-38.50***					
K-W test(2)	490271.86***					

Control Variables Detailed statistics and figures for the control variables are shown in Appendix 7.B. Appendix Table 7.B.1 reports Trade variables for the AMEX and NASDAQ in terms of mean and median. Appendix Table 7.B.2 provides the statistics of the trade size variable and Appendix Table 7.B.3 reports the statistics of the normalised trade size variable. Appendix Table 7.B.4 and Appendix Table 7.B.5 report the summary statistics of Risk1 and Risk2 variables, respectively. Appendix Table 7.B.6 reports the summary statistics of the Regional variable. Appendix Table 7.B.7 reports the intra-day variation of trade price.

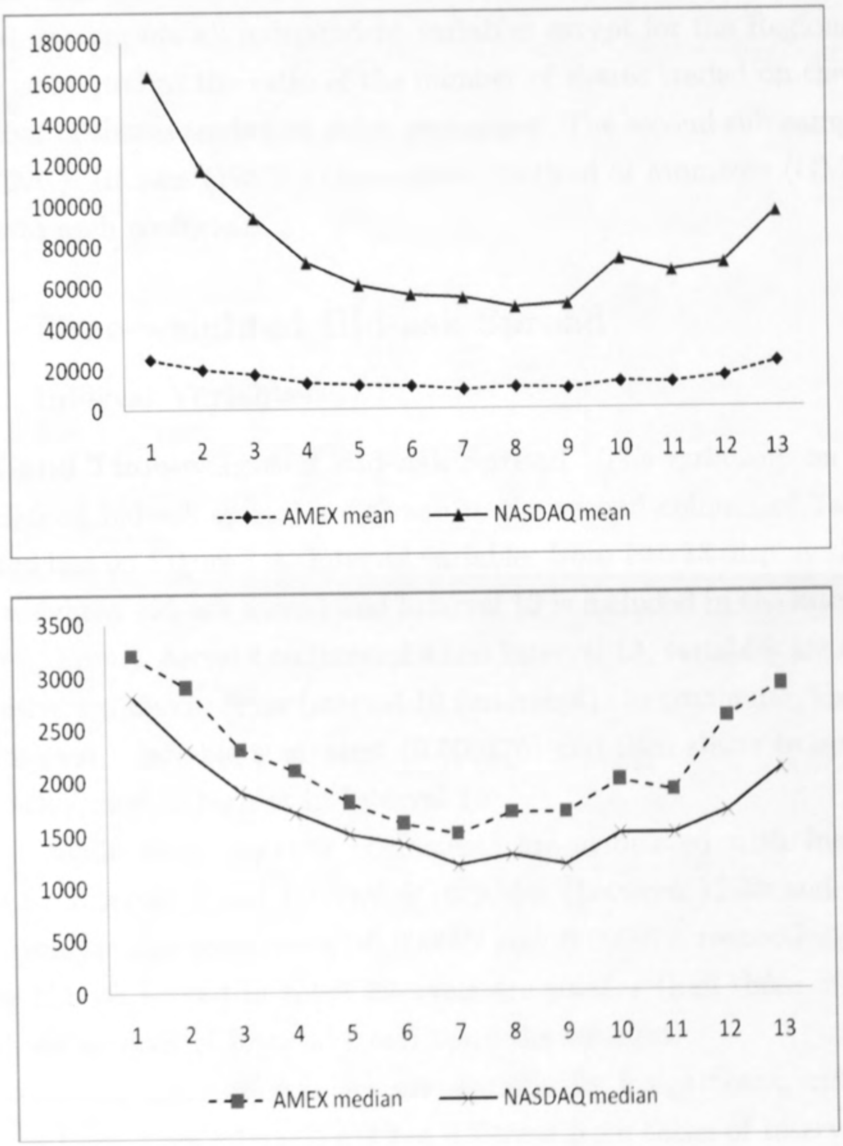


Figure 7.3: Trading Volume

7.6 Linear Regression Results of ETFs

I examined two sub-samples segregated by the stock exchange which the ETFs are traded on. The first uses only the AMEX. Trades and quotes from the AMEX are employed to compute all independent variables except for the Regional variable, which is computed as the ratio of the number of shares traded on the AMEX to the number of shares traded on other exchanges. The second sub-sample is based on NASDAQ. Hansen (1982)'s Generalised Method of Moments (GMM) is used to estimate each coefficient.

7.6.1 Time-weighted Bid-ask Spread

7.6.1.1 Interval Variables

AMEX and Time-weighted Bid-ask Spread The variations on the AMEX time-weighted bid-ask spread are shown in the second column of Table 7.4 and the dotted line on Figure 7.4. Interval variables from 1 to 13 display the variation of time-weighted bid-ask spread and Interval 10 is included in the intercept in the regression. From Interval 1 to Interval 4 and Interval 13, variables are significantly and positively different from Interval 10 (intercept). In particular, the coefficient of the Interval 1 variable is greatest (0.000276) and then starts to decrease, thus the bid-ask spread is highest in Interval 1.

Other significantly negative coefficients are associated with Interval 7 and Interval 8. Interval 7 and Interval 8 variables (between 12:30 and 13:30) have negative and smaller coefficients (-0.000092 and -0.000079, respectively), implying that the bid-ask spread in these intervals are smaller than those of interval 10. The bid-ask spreads of Interval 7 and 8 are the smallest.

All the other interval variables are statistically insignificant, indicating that spreads at these time intervals are not different from those of Interval 10. Thus, the bid-ask spreads are greater at the opening of the market, i.e. from Interval 1 to Interval 4. Overall intra-day variation of the time-weighted bid-ask spread has a weak U-shaped pattern.

The coefficient pattern associated with the Interval variables on the AMEX is similar to those discovered by McInish and Wood (1992) and McInish and Van

Ness (2002), who discover a U-shaped pattern for coefficients. According to the theories of Admati and Pfleiderer (1988) and Foster and Viswanathan (1993), the higher bid-ask spreads at AMEX opening and closing are due to the higher information asymmetry. Hong and Wang (2000) suggest that the higher bid-ask spreads at the AMEX opening are due to information asymmetry, but the higher bid-ask spreads at the AMEX closing are due to market closure. Since data set does not distinguish which trades come from informed trader or liquidity traders, the higher closing bid-ask spreads in AMEX cannot be confirmed due to informed trading (information asymmetry) or to liquidity trading (market closure). This will be a topic of future empirical research when we have more detailed trade and quote data set.

NASDAQ and Time-weighted Bid-ask Spread The third column of Table 7.4 and the solid line on Figure 7.4 report the regression results for the NASDAQ. Interval 1 and Interval 2 variables have statistically significant and positive coefficients (0.001738 and 0.00642, respectively) and the Interval 6 variable is significantly negative and the smallest (-0.000310). No other Interval variables are statistically significant, implying that coefficients are zero. NASDAQ bid-ask spreads have an overall L-shaped pattern.

Interpretation of NASDAQ bid-ask spreads is similar to that of AMEX bid-ask spreads. i.e. the higher bid-ask spreads at NASDAQ opening is due to information asymmetry (Admati and Pfleiderer (1988), Foster and Viswanathan (1993), and Hong and Wang (2000)). The levelled bid-ask spreads at NASDAQ closing suggest that no information asymmetry appears at NASDAQ closing. This suggests that even though ETFs are traded on AMEX and NASDAQ at the same time, ETFs on NASDAQ and those on AMEX are different at the market closing. Thus, this difference may be due to the difference between NASDAQ dealer system and AMEX specialist system.

In addition, an F-test shows whether AMEX and NASDAQ interval variables are the same coefficients. The test statistic suggests that AMEX interval coefficients are different from NASDAQ interval coefficients. The figure 7.4 suggests that NASDAQ spreads are more changeable than AMEX spreads. i.e.

the higher bid-ask spreads in the morning and lower bid-ask spread in the afternoon are shown in NASDAQ compared to the mid-day spread.

Table 7.4: Time-weighted Bid-ask Spread Regression based on Exchanges

*indicates statistical significance at the 0.10 level, ** at the 0.05 level, and *** at the 0.01 level. Dependent variable is Time-weighted bid-ask spread. Trades is the number of trades in each interval. Size is the log of average trade size in each interval. Risk1 is the time-weighted standard deviation of the bid-ask spread. Risk2 is a normalised value that is computed as (the standard deviation of time-weighted bid-ask spread minus the mean of standard deviation)/ the standard deviation of standard deviation. Nsize also is a normalised value that is computed as log ((trade size of interval t - average trade size of total interval)/the standard deviation of trade size of total interval). Regional is the ratio of trades on the NASDAQ (AMEX) to trades on other exchanges in case of NASDAQ (AMEX) regression. Price is the log of average trade price in each interval. Interval variables are dummy variables and represent 30-minute intervals. Interval 1 represents the time interval between 9:30 and 10:00, Interval 2 the time interval between 10:00 and 10:30, Interval 13 between 15:30 and 16:00. Each Interval variable is one when trading time falls into its own representing interval, otherwise zero. Hansen's Generalized Method of Moments are used to estimate coefficients. F-test tests the equality of AMEX interval and NASDAQ interval.

	AMEX	NASDAQ
	Coefficient	Coefficient
Intercept (b0)	0.009700***	0.013607***
Trades (b1)	-0.000003***	-0.000008***
Size (b2)	-0.000266***	-0.000100
Risk1 (b3)	0.488980***	0.551027***
Risk2 (b4)	0.000292***	0.001366***
Nsize (b5)	0.001200***	0.001001*
Regional (b6)	-0.000051***	-0.000270***
Price (b7)	-0.001500***	-0.002280***
Interval 1	0.000276***	0.001738***
Interval 2	0.000145***	0.000642**
Interval 3	0.000067**	0.000075
Interval 4	0.000060*	-0.000170
Interval 5	-0.000014	-0.000250
Interval 6	-0.000048	-0.000310*
Interval 7	-0.000092***	-0.000200
Interval 8	-0.000079***	-0.000050
Interval 9	-0.000043	-0.000240
Interval 11	0.000019	0.000050
Interval 12	-0.000009	-0.000040
Interval 13	0.000074**	0.000013
R^2	0.7413	0.7657
F-test	91.12***	

7.6.1.2 Controling Independent Variables

Appendix 7.D. discusses the control variables in the bid-ask regression model.

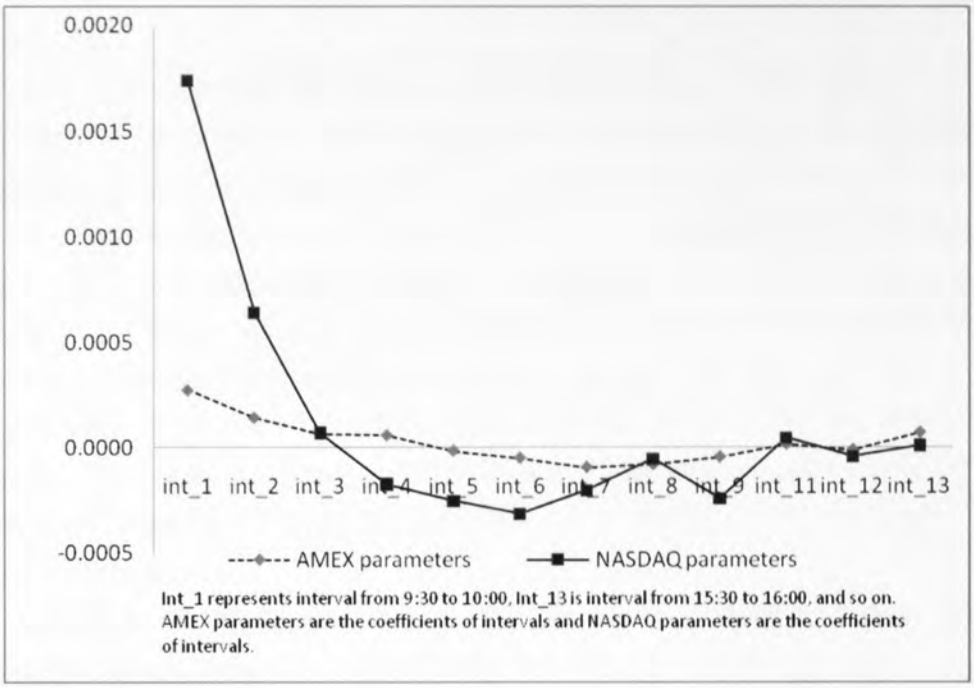


Figure 7.4: The Trend line of Time-weighted Bid-ask Spread Interval Variables

7.6.2 Return Volatility

7.6.2.1 Interval Variables

AMEX and Return Volatility Interval variables from 1 to 13 highlight variations in return volatility. The second column of Table 7.5 shows variations for the AMEX. Only Interval 4 and 11 variables are statistically insignificant. All the other interval variables are statistically significant. Variables from Interval 1 to Interval 3, and Interval 13, are statistically and positively different from zero, implying that return volatility is significantly higher for these intervals than Interval 10. In particular, the coefficient of the Interval 1 variable is greatest (0.000249) and then starts to decrease, thus return volatility is highest in Interval 1.

Significantly negative coefficients appear between Interval 5 and Interval 9 variables. Interval 12 also has a significantly negative coefficient. These significant negative coefficients mean that these intervals have lower return volatility than Interval 10. At mid-day, i.e. Interval 6, the coefficient is negative and smallest

(-0.0008).

Thus, return volatilities are greater at the opening of the market, i.e. from Interval 1 to Interval 3. Overall the intra-day variation of the AMEX return volatility has a rough U-shaped pattern, which is similar to the results of Wood, et al. (1985) and Lockwood and Linn (1990). Admati and Pfleiderer (1988) show that when the number of informed traders is constant over time, the extent to which prices reveal private information is constant and the variance of the price change is constant. The higher return volatility at market open and close means that there are greater numbers of informed traders, implying the higher information asymmetry. Foster and Viswanathan (1991) too provide a similar intuition that when there is higher information asymmetry, higher return volatility appears.

According to Hong and Wang (2000), periodic market closure generates different trading behaviour of different traders. i.e. the higher return volatility at market open is due to the hedging trading, and the higher return volatility at market close is due to time-varying information asymmetry. However, these trading behaviours cannot be confirmed because I cannot discern these two trading in my data set. The U-shaped intraday return volatility appears as Hong and Wang (2000).

NASDAQ and Return Volatility The third column of Table 7.5 reports the regression results for the NASDAQ. Interval 1 to 3, and Interval 13 variables have statistically significant and positive coefficients. Intervals 5 to 12 variables have significantly negative coefficients. The Interval variables' patterns on the NASDAQ seem similar to those on the AMEX in the second column. Return volatility shows a roughly U-shaped pattern. The higher NASDAQ return volatility is interpreted in the same way as the higher AMEX return volatility.

F-test result in the Table 7.5 shows the difference between AMEX return volatility and NASDAQ return volatility. F-test tests whether the coefficients of interval are different between AMEX and NASDAQ. The test statistic is 53.02 and significant at 1% level, meaning that AMEX intraday return volatility is different from NASDAQ intraday return volatility.

The intraday patterns of NASDAQ return volatility are similar to those of NASDAQ, even though F-test suggests the different level of average return volatility between AMEX and NASDAQ. This suggests that higher return volatility are due to higher information asymmetry (Admati and Pfleiderer (1988) and Foster and Viswanathan (1991)).

The understanding of U-shaped AMEX return volatility by Hong and Wang (2000) applies to U-shaped NASDAQ return volatility. Different traders are dominant during different trading hours. Hedge trading may be dominant at NASDAQ open and informed trading due to informed trading may be common at NASDAQ close. Because of the limitation of data set, I cannot discern between hedge trading and informed trading during the trading hours. However, the periodic closure of NASDAQ generates the U-shaped intraday return volatility.

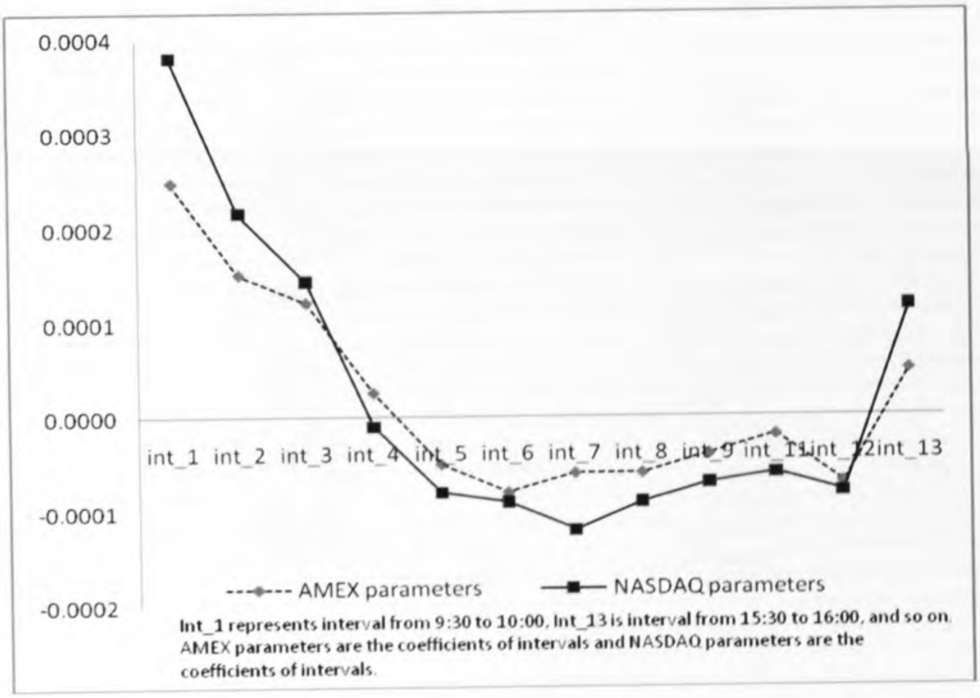


Figure 7.5: The Trend line of Retrun Volatilities Interval Variables

7.6.2.2 Controling Independent Variables

Appendix 7.D. discusses the control variables in the return volatility regression model.

Table 7.5: Return Volatility Regression based on Exchange

*indicates statistical significance at the 0.10 level, ** at the 0.05 level, and *** at the 0.01 level. Return volatility is dependent variable. Trades is the number of trades in each interval. Size is the log of average trade size in each interval. Nsize also is a normalised value that is computed as log ((trade size of interval t - average trade size of total interval)/the standard deviation of trade size of total interval). Regional is the ratio of trades on the NASDAQ (AMEX) to trades on other exchanges in case of NASDAQ (AMEX) regression. Price is the log of average trade price in each interval. Interval variables are dummy variables and represent 30-minute intervals. Interval 1 represents the time interval between 9:30 and 10:00, Interval 2 the time interval between 10:00 and 10:30, Interval 13 between 15:30 and 16:00. Each Interval variable is one when trading time falls into its own representing interval, otherwise zero. Hansen's Generalized Method of Moments are used to estimate coefficients. F-test tests the equality of AMEX interval and NASDAQ interval.

	AMEX	NASDAQ
	Coefficient	Coefficient
Intercept	0.003524***	0.003405***
Trades	-0.000004***	-0.000002***
Size	-0.000140***	-0.000140***
Nsize	0.000642***	0.000659***
Regional	-0.000040***	-0.000070***
Price	-0.000390***	-0.000320***
Interval 1	0.000249***	0.000380***
Interval 2	0.000152***	0.000217***
Interval 3	0.000122***	0.000144***
Interval 4	0.000026	-0.000010
Interval 5	-0.000050***	-0.000080***
Interval 6	-0.000080***	-0.000090***
Interval 7	-0.000060***	-0.000120***
Interval 8	-0.000060***	-0.000090***
Interval 9	-0.000040**	-0.000070***
Interval 11	-0.000020	-0.000060***
Interval 12	-0.000070***	-0.000080***
Interval 13	0.000049**	0.000117***
R^2	0.1394	0.1276
F-test	53.02***	

7.6.3 Trading Volume

7.6.3.1 Interval Variables

AMEX and Trading Volume AMEX trading volume regression is presented in the second column of Table 7.6. The AMEX regression provides a traditional U-shaped intra-day variation of trading volume. At the market open and close, i.e. Intervals 1, 2, 12, and 13, variables have significantly positive coefficients. Interval 7 and 9 variables, which represent the mid-day trading period, have significantly negative coefficients. The U-shaped trading volume is consistent with the findings of Jain and Joh (1988) and Gerety and Mulherin (1992).

The understanding of trading volume is same as that of return volatility. The higher trading volume is due to the higher information asymmetry. Thus, the higher information asymmetry at AMEX open and close leads to the higher trading volume of AMEX in the early morning and late afternoon.

According to Hong and Wang (2000), periodic market closure generates U-shaped intraday trading volume of AMEX. The information accumulation during the market closure generates higher trading volume at the AMEX open and the intention of hedge traders to reduce their hedge positions at the market closing generates higher trading volume at the AMEX close.

NASDAQ and Trading Volume The third column of Table 7.6 reports the regression results for the NASDAQ. The overall intra-day variation of NASDAQ trading volume is similar to the variation of AMEX trading volume. NASDAQ trading volume also shows U-shaped intra-day variation. While Interval 1, Interval 12, and Interval 13 variables have statistically significant positive coefficients, Interval 6, 8, and 9 variables are significantly negative. All the other interval variables are insignificant. These coefficients create a U-shaped pattern.

While AMEX and NASDAQ have U-shaped trading volume, F-test suggests that these trading volume are different from each other. The significance is at 1% level. The NASDAQ trading volume is more changeable than AMEX trading volume.

Even though F-test suggests the different level of trading volume between AMEX and NASDAQ, the intraday patterns of AMEX/NASDAQ trading volumes are u-shaped. This suggests the interpretation of trading volume. Thus the higher trading volumes at NASDAQ open and clsose imply the higher information asymmetry.

Hong and Wang (2000)'s market closures explain the U-shaped intraday trading volume of NASDAQ.

Table 7.6: Trading Volume Regression based on Exchanges

*indicates statistical significance at the 0.10 level, ** at the 0.05 level, and *** at the 0.01 level. Dependent variable is trading volume. Size is the log of average trade size in each interval. Regional is the ratio of trades on the NASDAQ (AMEX) to trades on other exchanges in case of NASDAQ (AMEX) regression. Price is the log of average trade price in each interval. Interval variables are dummy variables and represent 30-minute intervals. Interval 1 represents the time interval between 9:30 and 10:00, Interval 2 the time interval between 10:00 and 10:30, Interval 13 between 15:30 and 16:00. Each Interval variable is one when trading time falls into its own representing interval, otherwise zero. Hansen's Generalized Method of Moments are used to estimate coefficients. F-test tests the equality of AMEX interval and NASDAQ interval.

	AMEX	NASDAQ
	parameters	parameters
Intercept(b0)	-161868***	-591451***
Size (b2)	20708***	96120***
Regional(b6)	4913***	3266***
Price(b7)	10264***	14084***
Interval 1	7949***	50950***
Interval 2	3430***	19298**
Interval 3	2596**	5683
Interval 4	-930	-7216
Interval 5	-1002	-13495*
Interval 6	-1060	-13179*
Interval 7	-2304***	-10312
Interval 8	-1768	-20710**
Interval 9	-1851**	-15102**
Interval 11	680	-5051
Interval 12	3064***	-4120
Interval 13	8394***	13339
R^2	0.1810	0.1334
F-test	98.70***	

7.6.3.2 Controling Independent Variables

Appendix 7.D. discusses the control variables in the trading volume regression model.

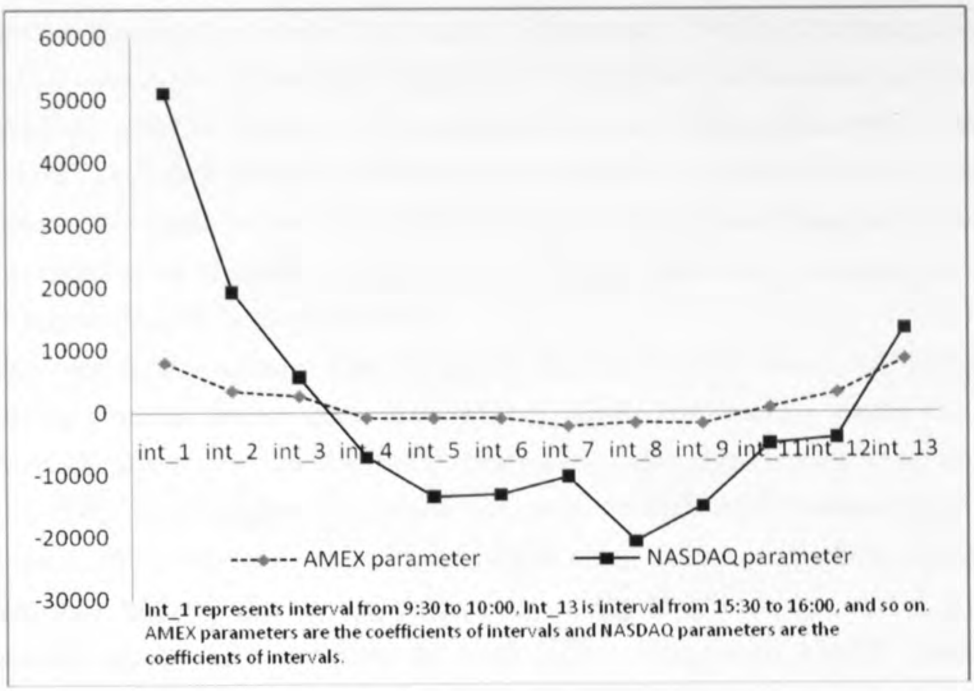


Figure 7.6: The Trend line of Trading Volume Interval Variables

7.6.4 Discussion of Intra-day Pattern

The higher trading volume at market open and close suggests higher information asymmetry and the lower participation of dealers. The higher return volatility at market open and close suggests the higher information asymmetry and the higher inventory-holding risk of dealers. These intraday patterns are not different between AMEX and NASDAQ. Meanwhile, the higher bid-ask spreads appear at market open and the smallest bid-ask spreads appear at noon[1]. These patterns do not discern stock exchanges (AMEX and NASDAQ). Then NASDAQ closing bid-ask spreads are insignificantly higher than the level of afternoon bid-ask spreads[2]. AMEX closing bid-ask spreads are significantly higher than AMEX afternoon bid-ask spreads.

Based on these findings, I face one unexplained intraday pattern, which is NASDAQ closing bid-ask spreads. If higher trading volume and return volatility are due to higher information asymmetry at market closing, then bid-ask spreads

[1]Interval 6 to 8, which is between 12:00 and 13:30.

[2]Interval 10, which is between 14:00 and 14:30.

at market closing too should be higher. However, NASDAQ closing bid-ask spreads do not seem to be higher, implying insignificant information asymmetry.

Another notable thing to be mentioned is the highly changeable bid-ask spreads of NASDAQ, which is shown in the Figure IV. i.e. the changes of intraday coefficients are more severe at NASDAQ than AMEX. These changes depend on the characteristics of stock exchanges (AMEX and NASDAQ), because the same ETFs are traded on both exchanges.

One way to understand this change is that at market open, NASDAQ has higher information asymmetry than AMEX, and after trading starts the high information asymmetry of NASDAQ diminishes more significantly than that of AMEX. This may suggest that since competitive NASDAQ dealers know only their part of ETFs order flow and do not know other dealers order flow, NASDAQ dealers face higher information asymmetry. Meanwhile monopolistic AMEX specialists manage all order flows for their ETFs. This leads AMEX specialists to have lower information asymmetry than NASDAQ dealers.

7.7 Concluding Remarks

In this chapter, I investigate the intra-day variation of ETFs bid-ask spreads, return volatility and trading volume. Specifically, I examine whether these variables are different depending on exchanges and if there is a difference, I conclude that this arises as a result of the variation in trading system used on the exchange. I divide the quotes in my data set into AMEX quotes and NASDAQ quotes. I use quotes from only one exchange to investigate the intra-day variation of bid-ask spreads. The AMEX employs a specialist system whilst the NASDAQ uses a dealer system. I also examined the intra-day variation of return volatility and trading volume.

The time-weighted bid-ask spread regression suggests that AMEX bid-ask spreads show the appearance of the highest bid-ask spread at the market opening, the lowest bid-ask spread at the mid-day, and elevated bid-ask spread at the close. NASDAQ bid-ask spreads are higher in the morning, lowest in the mid-day, slightly elevated at the market closing. These patterns are confirmed by the change of control variables (Appendix 7-C). Two notable things occur: NASDAQ spreads are more changeable than AMEX spreads; and NASDAQ spreads are slightly higher at the market closing.

The return volatility regression shows a U-shaped pattern irrespective of the exchange. The trading volume regression also shows U-shaped intra-day variation. i.e. return volatility and trading volume are higher in the morning and at their lowest at the mid-day. Finally, at the market close, return volatility and trading volume increase.

The leveled NASDAQ closing bid-ask spreads are not easily understood. The reason is that NASDAQ return volatility and trading volume suggest the higher information asymmetry at NASDAQ close. However, this higher information asymmetry does not generate the higher bid-ask spreads at the NASDAQ closing. In addition, the same ETFs on AMEX clearly have higher closing bid-ask spreads, which imply the higher information asymmetry of AMEX.

A Appendix 7.A

Appendix 7-A shows, the estimation results of this study, and the regression results of McInish and Van Ness (2002). The estimation results are similar to each other. When the estimation method is applied to the data from this study, it is easy to find the difference between this data set and the results of McInish and Wood (1992) and McInish and Van Ness (2002) Since I use a similar regression methodology to that of McInish and Van Ness (2002), I compare my estimation results to McInish and Van Ness (2002) with the same sample. McInish and Van Ness (2002) employ 30 stocks that are the constituents of the Dow Johns Industrial Average (DJIA). The sample period lasts from October 1997 to September 1998. They obtain trades and quotes on the New York Stock Exchange (NYSE) to analyse the intra-day behaviour of bid-ask spreads.

The regression equations are the same as equation (7.21). i.e.

$$\begin{aligned} BAS_{i,t} = b_0 + b_1 Trades_{i,t} + b_2 Size_{i,t} + b_3 Risk1 + b_4 Risk2 \\ + b_5 Nsize_{i,t} + b_6 Regional + b_7 Price_{i,t} \\ +12\, Interval\ Dummy\ Variables(1-9, 11-13) \\ + e_{i,t} \end{aligned}$$

The definitions of variables are the same as in McInish and Van Ness. For detailed explanation, see section 7.4.

Appendix Table 7.A.1 reports the results from my estimation method and the results from McInish and Van Ness (2002). Even though specific values of coefficients are slightly different, the overall results are the same. The Trade, Size, Regional, and Price variables have a negative relationship with the time-weighted bid-ask spread. The risk variables (Risk1 and Risk2) and Nsize variables have a positive relationship with the time-weighted bid-ask spread. All these relationships are supported by theoretical literature.

To investigate the bid-ask spreads of ETFs and matched control stocks, I employ the same method as in this appendix. This strategy brings the advantage that it is easy to compare the results of McInish and Van Ness (2002) with my estimation results.

Table 7.A.1: The Comparison of Estimation Results

Trade variable is the number of trades in each interval. *Size* is the log of average trade size in each interval. *Risk1* is the time-weighted standard deviation of the bid-ask spread. *Risk2* is a normalised value that is computed as (the standard deviation of time-weighted bid-ask spread minus the mean of standard deviation)/ the standard deviation of standard deviation. *Nsize* also is a normalised value that is computed as log ((trade size of interval *t* – average trade size of total interval)/the standard deviation of trade size of total interval). *Regional* is the ratio of trades on the NYSE to trades on regional exchanges. *Price* is the log of average trade price in each interval. Interval variables are dummy variables and represent 30- minute intervals. Interval 1 represents the time interval between 9:30 and 10:00, Interval 2 the time interval between 10:00 and 10:30, Interval 13 between 15:30 and 16:00. Each Interval variable is one when trading time falls into its own representing interval, otherwise zero. *indicates statistical significance at the 0.10 level, ** at the 0.05 level, and *** at the 0.01 level.

	McInish and Van Ness (2002)		Estimation Results	
	Coefficients	T-statistics	Coefficients	T-statistics
Intercept	0.00504	121.28**	0.00448	86.74***
Trade	-0.00000	-1.35	-0.00000	-10.88***
Size	-0.00005	-13.78**	-0.00003	-5.87***
Risk1	1.23950	130.31**	1.11994	246.82***
Risk2	0.00032	262.41**	0.00007	38.39***
Nsize	0.00003	17.32**	0.00007	1.37
Regional	-0.00007	-33.88**	-0.00004	-4.29***
Price	-0.00092	-192.07**	-0.00079	-136.97***
Interval 1	0.000128	21.05**	0.000123	15.73***
Interval 2	0.000015	2.74**	0.000054	6.9***
Interval 3	0.000002	0.27	0.000046	5.84***
Interval 4	0.000005	0.88	0.000039	5.00***
Interval 5	0.000002	0.44	0.000019	2.41**
Interval 6	0.000012	2.18**	0.000004	0.56
Interval 7	0.000002	0.32	-0.000023	-2.85***
Interval 8	-0.000005	-0.83	-0.000015	-1.90*
Interval 9	-0.000004	-0.74	-0.000020	-2.47**
Interval 11	0.000002	0.44	0.000025	3.21***
Interval 12	0.000005	0.87	0.000050	6.21***
Interval 13	0.000007	2.20**	0.000055	6.84***
R^2	0.8007		0.7820	

B Appendix 7.B

Appendix Table 7.B.1 reports the average value of the number of trades in 30-minute time intervals, and standard deviation and median values. ETFs and control stocks show a U-shaped intra-day pattern of mean and median values. The number of trade variables represents the activity of traders. Since the number of trades does not distinguish informed traders and liquidity traders, the average and median statistics suggest that when the market is either open or closed, informed and uninformed traders are actively transacting.

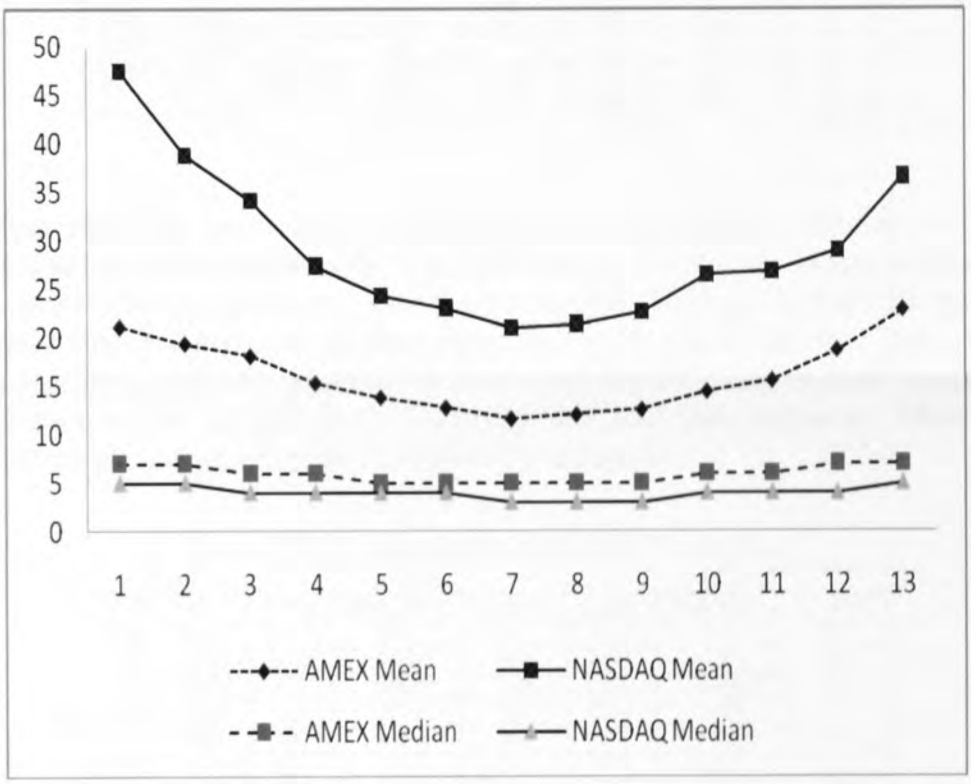


Figure 7.B.1: The Number of Trades

Appendix Table 7.B.2 and Appendix Table 7.B.3 show the average, standard deviation, and median of the trade size and normalised trade size variables. The greatest trade size appears in the first trading interval. In the middle of the day, trade size is minimised. Although trade size is increasing from mid-day to the end of the day, the increase in trade size is not significant since the trade size in the final trading interval is less than the average trade size of all intervals. Trade size is considered as the activity of traders. Thus, in the sample of ETFs and control stocks, traders' activities are higher in the early morning, lower at mid-day, and modest at the end of the day.

Table 7.B.1: Sample Statistics of No. of Trades

Trade	AMEX			NASDAQ		
Interval	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
9:30 – 10:00	21	42	7	47	141	5
10:00 – 10:30	19	40	7	39	120	5
10:30 – 11:00	18	38	6	34	104	4
11:00 – 11:30	15	32	6	27	82	4
11:30 – 12:00	14	28	5	24	72	4
12:00 – 12:30	13	26	5	23	69	4
12:30 – 13:00	12	24	5	21	63	3
13:00 – 13:30	12	25	5	21	66	3
13:30 – 14:00	13	26	5	23	71	3
14:00 – 14:30	14	30	6	26	86	4
14:30 – 15:00	16	32	6	27	86	4
15:00 – 15:30	19	36	7	29	90	4
15:30 – 16:00	23	48	7	37	115	5
All	16	34	6	29	93	4

Meanwhile, the normalised trade size variable also follows a similar intra-day pattern as the trade size variable. The difference is the slightly higher normalised trade size at the end of the day since the normalised trade size at the 13th interval is greater than the average/ median normalised trade size of the day. Thus, when the normalised trade size is viewed as the arrival of public information, the public information arrival is higher in the early morning, and then decreases. After mid-day, the public information arrival modestly increases.

Table 7.B.2: Sample Statistics of Trade Size

Size	AMEX			NASDAQ		
Interval	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
9:30 – 10:00	1134	3659	500	2886	8845	633
10:00 – 10:30	1026	3409	450	2456	9776	545
10:30 – 11:00	921	2891	425	2127	6805	500
11:00 – 11:30	839	2178	400	1835	5659	445
11:30 – 12:00	889	3281	400	1845	7320	443
12:00 – 12:30	874	2930	361	1568	4998	417
12:30 – 13:00	877	7439	364	1528	4785	400
13:00 – 13:30	1364	30670	367	1803	9318	411
13:30 – 14:00	755	2194	360	1689	8426	400
14:00 – 14:30	838	2263	400	1730	10704	413
14:30 – 15:00	786	3249	370	2150	19972	404
15:00 – 15:30	791	1739	392	1623	6093	450
15:30 – 16:00	972	4131	420	1538	3818	475
All	926	9007	400	1898	9107	450

Appendix Table 7.B.4 and Appendix Table 7.B.5 report the sample statistics of Risk1 and Risk2 variables. The Risk1 variable measures the risk of each time interval, while Risk2 measures the risk that can be different from the average risk of the day. Median Risk1 variables of ETFs and control stocks show U-shaped

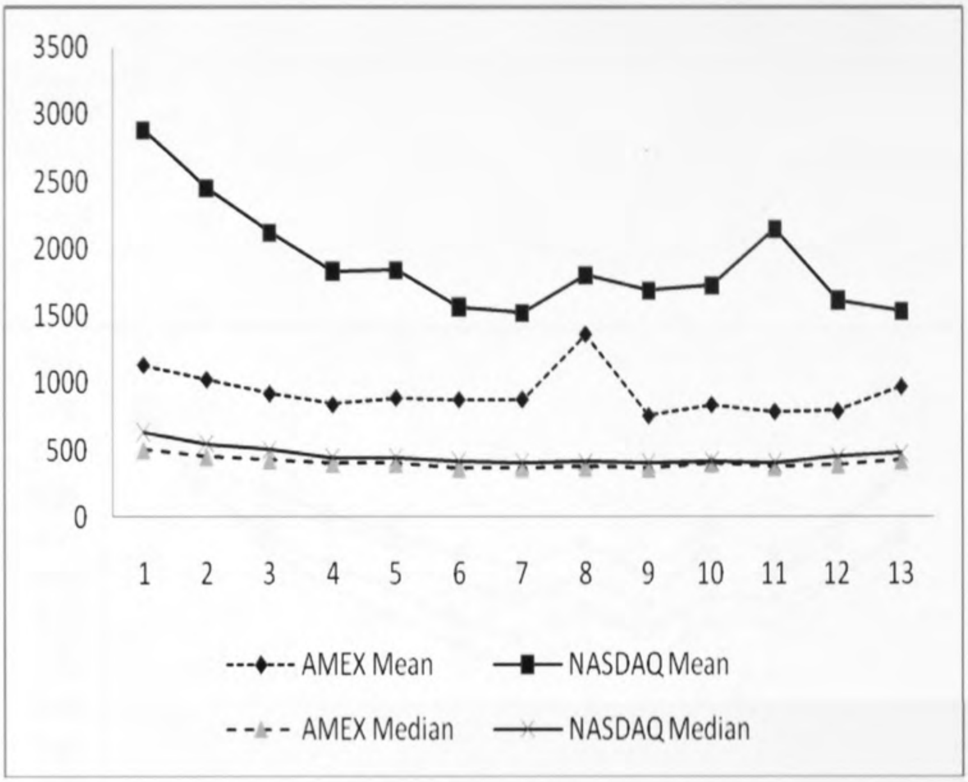


Figure 7.B.2: Trade Size

Table 7.B.3: Sample Statistics of Nsize

Nsize	AMEX			NASDAQ		
Interval	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
9:30 – 10:00	0.9794	0.1145	0.9823	0.9764	0.1282	0.9807
10:00 – 10:30	0.9672	0.1111	0.9641	0.9619	0.1245	0.9626
10:30 – 11:00	0.9612	0.1125	0.9569	0.9514	0.1228	0.9476
11:00 – 11:30	0.9559	0.1111	0.9535	0.9432	0.1227	0.9386
11:30 – 12:00	0.9517	0.1118	0.9490	0.9388	0.1226	0.9318
12:00 – 12:30	0.9464	0.1152	0.9391	0.9306	0.1231	0.9251
12:30 – 13:00	0.9437	0.1155	0.9357	0.9269	0.1252	0.9194
13:00 – 13:30	0.9493	0.1199	0.9409	0.9326	0.1266	0.9264
13:30 – 14:00	0.9416	0.1132	0.9341	0.9280	0.1229	0.9200
14:00 – 14:30	0.9528	0.1125	0.9482	0.9369	0.1228	0.9310
14:30 – 15:00	0.9462	0.1079	0.9413	0.9355	0.1193	0.9306
15:00 – 15:30	0.9541	0.1059	0.9500	0.9410	0.1215	0.9401
15:30 – 16:00	0.9656	0.1068	0.9677	0.9509	0.1178	0.9516
All	0.9550	0.1125	0.9508	0.9426	0.1237	0.9388

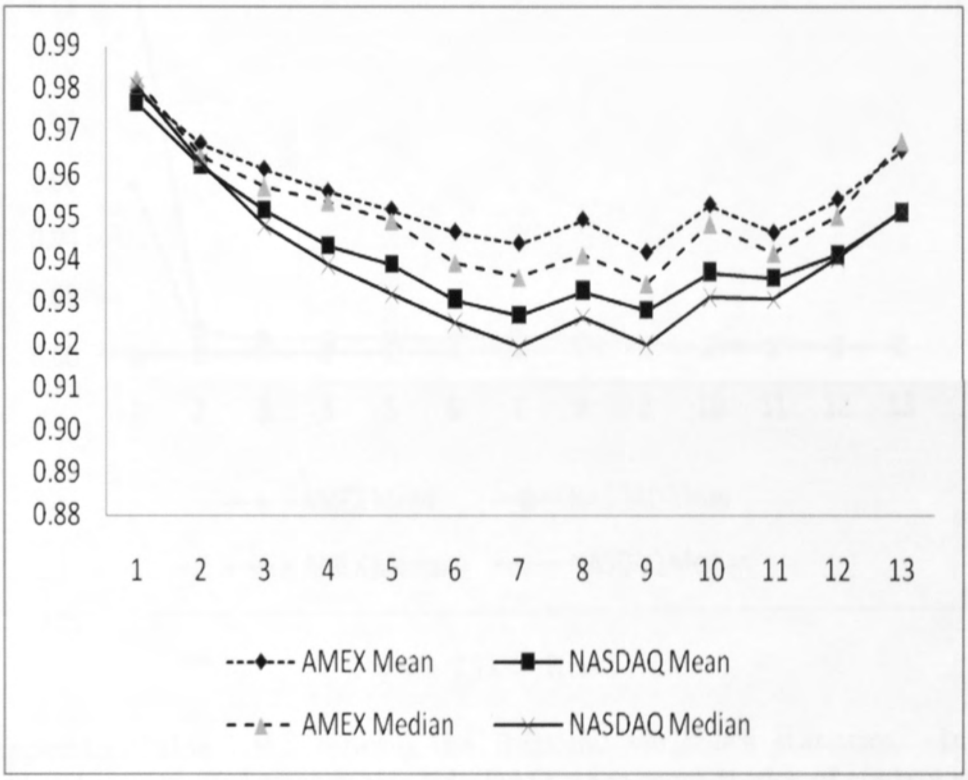


Figure 7.B.3: Nsize

intra-day variation, while mean Risk1 variables are roughly U-shaped. Risk2 variables suggest that the risk in the first trading interval morning is above the average risk of the day, but the risk in the final trading interval is below the average of the day.

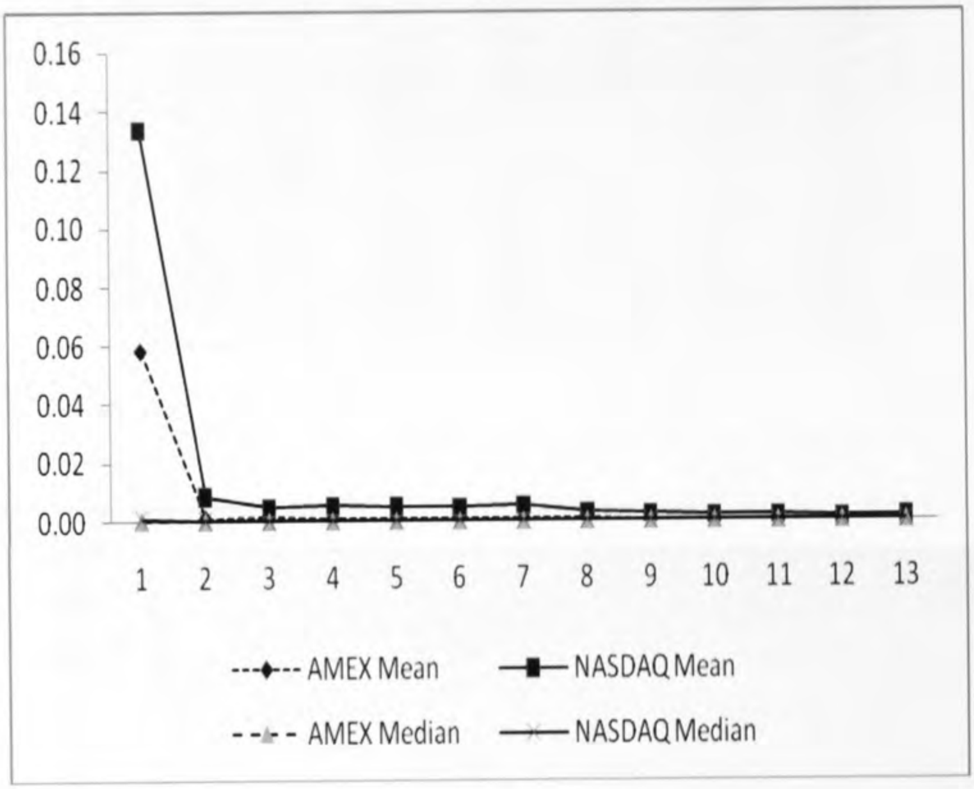


Figure 7.B.4: Risk1

Appendix Table 7.B.6 reports the Regional variable's statistics. In this analysis, the main exchange is the NASDAQ since most trades of control stocks happen there. Meanwhile, trades of ETFs occur on the AMEX, NASDAQ and other exchanges. The Regional variable represents competition from other exchanges. ETFs show a U-shaped intra-day pattern, implying that competition is fierce at opening and closing times. Controls stocks show that competition increases as the market is close to its end. The higher competition of control stocks at the closing of market is possibly due to the inventory control of dealers. As the market is close to its end, dealers aggressively post bid/ask prices to resolve their imbalance due to trades over the day.

Appendix Table 7.B.7 reports sample statistics of trade price. No unique patterns appear in ETFs and control stocks. The Price variable is included in the regression because trade price is negatively associated with the bid-ask spread.

Table 7.B.4: Sample Statistics of Risk1

Risk1	AMEX			NASDAQ		
Interval	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
9:30 – 10:00	0.0585	2.5546	0.0005	0.1334	1.2126	0.0009
10:00 – 10:30	0.0011	0.0048	0.0005	0.0087	0.0819	0.0004
10:30 – 11:00	0.0015	0.0080	0.0005	0.0047	0.0388	0.0003
11:00 – 11:30	0.0013	0.0061	0.0004	0.0054	0.0407	0.0002
11:30 – 12:00	0.0013	0.0062	0.0004	0.0051	0.0511	0.0002
12:00 – 12:30	0.0014	0.0071	0.0004	0.0047	0.0362	0.0002
12:30 – 13:00	0.0011	0.0040	0.0004	0.0055	0.0723	0.0002
13:00 – 13:30	0.0009	0.0032	0.0004	0.0031	0.0224	0.0002
13:30 – 14:00	0.0008	0.0023	0.0004	0.0024	0.0119	0.0002
14:00 – 14:30	0.0007	0.0019	0.0004	0.0020	0.0091	0.0002
14:30 – 15:00	0.0006	0.0012	0.0004	0.0018	0.0078	0.0002
15:00 – 15:30	0.0006	0.0009	0.0004	0.0014	0.0051	0.0002
15:30 – 16:00	0.0005	0.0010	0.0004	0.0016	0.0055	0.0003
All	0.0050	0.6766	0.0004	0.0131	0.3273	0.0002

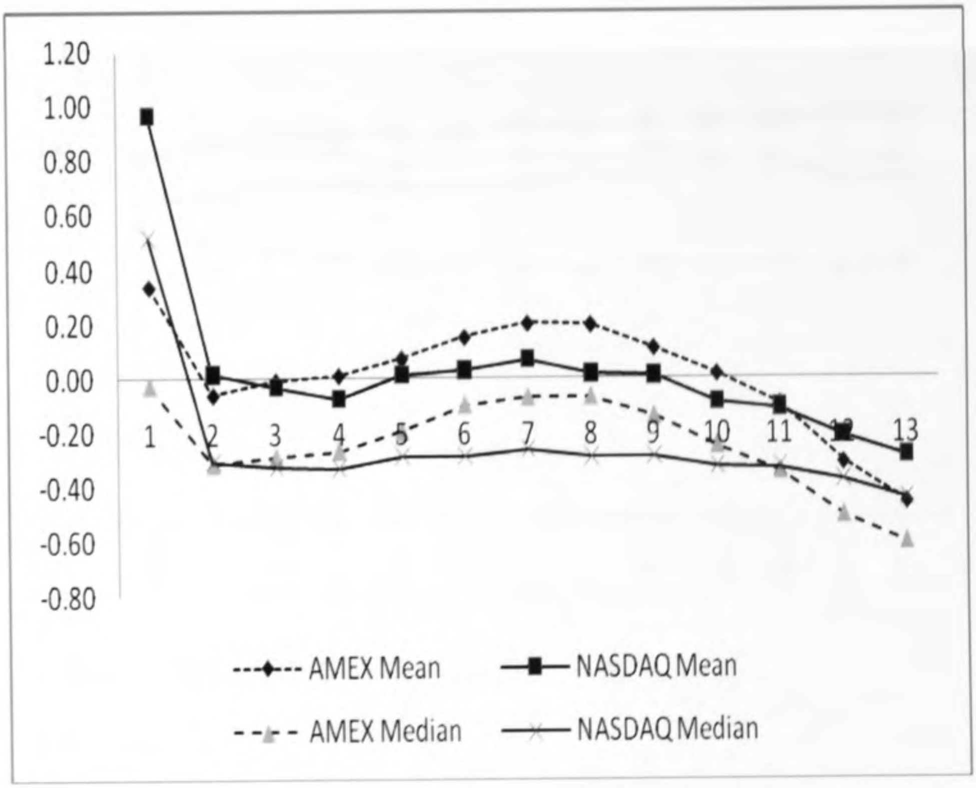


Figure 7.B.5: Risk2

Table 7.B.5: Sample Statistics of Risk2

Risk2	AMEX			NASDAQ		
Interval	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
9:30 – 10:00	0.3379	1.1920	-0.0288	0.9678	1.5596	0.5165
10:00 – 10:30	-0.0601	0.9079	-0.3170	0.0141	0.9563	-0.3081
10:30 – 11:00	-0.0079	0.9284	-0.2908	-0.0322	0.8958	-0.3250
11:00 – 11:30	0.0102	0.9212	-0.2680	-0.0755	0.8659	-0.3341
11:30 – 12:00	0.0716	0.9395	-0.1984	0.0116	0.8723	-0.2890
12:00 – 12:30	0.1499	0.9749	-0.1020	0.0274	0.8876	-0.2902
12:30 – 13:00	0.1998	0.9964	-0.0740	0.0688	0.9063	-0.2630
13:00 – 13:30	0.1961	0.9823	-0.0702	0.0153	0.8851	-0.2908
13:30 – 14:00	0.1103	0.9571	-0.1418	0.0088	0.8782	-0.2902
14:00 – 14:30	0.0150	0.9014	-0.2502	-0.0904	0.8088	-0.3246
14:30 – 15:00	-0.0956	0.8639	-0.3445	-0.1113	0.8038	-0.3288
15:00 – 15:30	-0.3112	0.7769	-0.5069	-0.2134	0.7359	-0.3767
15:30 – 16:00	-0.4575	0.7312	-0.6055	-0.2880	0.7526	-0.4437
All	0.0000	0.9499	-0.2773	0.0000	0.9457	-0.3132

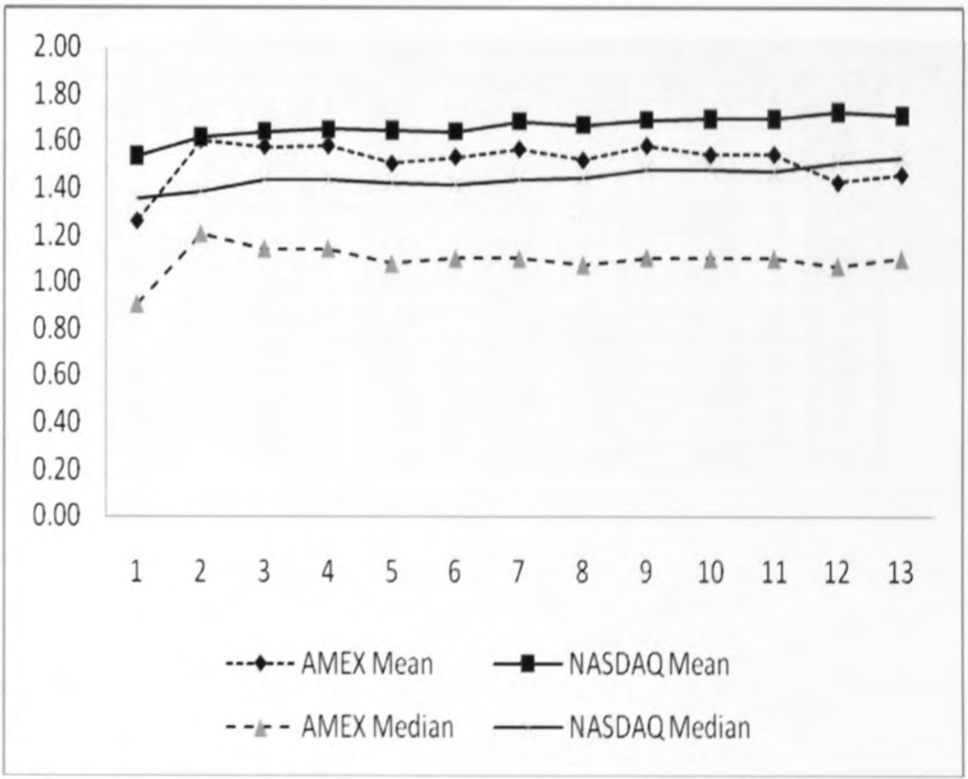


Figure 7.B.6: Regional

Table 7.B.6: Sample Statistics of Regional

Regional	AMEX			NASDAQ		
Interval	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
9:30 – 10:00	1.2631	1.1198	0.9039	1.5360	0.9269	1.3596
10:00 – 10:30	1.6013	1.4042	1.2040	1.6157	1.1129	1.3863
10:30 – 11:00	1.5752	1.3850	1.1400	1.6410	1.1044	1.4380
11:00 – 11:30	1.5790	1.3826	1.1408	1.6489	1.1082	1.4351
11:30 – 12:00	1.5049	1.3397	1.0769	1.6435	1.1290	1.4220
12:00 – 12:30	1.5324	1.3585	1.0986	1.6400	1.1499	1.4153
12:30 – 13:00	1.5653	1.3933	1.0986	1.6834	1.1983	1.4395
13:00 – 13:30	1.5195	1.3657	1.0730	1.6680	1.1651	1.4469
13:30 – 14:00	1.5800	1.4201	1.1040	1.6915	1.1579	1.4824
14:00 – 14:30	1.5429	1.3945	1.0986	1.6963	1.1281	1.4841
14:30 – 15:00	1.5436	1.3967	1.0986	1.6940	1.1441	1.4717
15:00 – 15:30	1.4257	1.2553	1.0678	1.7244	1.1464	1.5122
15:30 – 16:00	1.4591	1.2732	1.0986	1.7101	1.0823	1.5313
All	1.5155	1.3504	1.0986	1.6623	1.1225	1.4469

Meanwhile the average trade price of ETFs is \$66.36 and the average trade price of control stocks is \$36.63.

Table 7.B.7: Sample Statistics of Trade Price

Price	AMEX			NASDAQ		
Interval	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
9:30 – 10:00	68.16	35.57	63.56	69.53	36.42	65.66
10:00 – 10:30	67.21	34.25	62.38	68.36	35.59	64.96
10:30 – 11:00	67.10	34.32	62.31	68.25	35.43	64.91
11:00 – 11:30	67.18	34.39	62.41	68.22	35.04	64.65
11:30 – 12:00	67.42	34.23	62.43	68.14	35.00	64.67
12:00 – 12:30	67.59	34.39	62.89	68.46	35.27	64.86
12:30 – 13:00	67.63	34.31	62.81	68.13	34.96	64.61
13:00 – 13:30	67.77	34.20	62.85	68.38	34.79	64.69
13:30 – 14:00	67.29	34.13	62.81	68.22	34.59	64.86
14:00 – 14:30	67.14	33.99	62.30	68.20	34.79	64.61
14:30 – 15:00	66.96	33.96	62.01	68.25	34.55	64.74
15:00 – 15:30	66.88	33.74	62.00	67.82	34.39	64.53
15:30 – 16:00	66.14	33.17	61.28	68.10	34.10	64.57
All	67.24	34.18	62.41	68.30	34.98	64.82

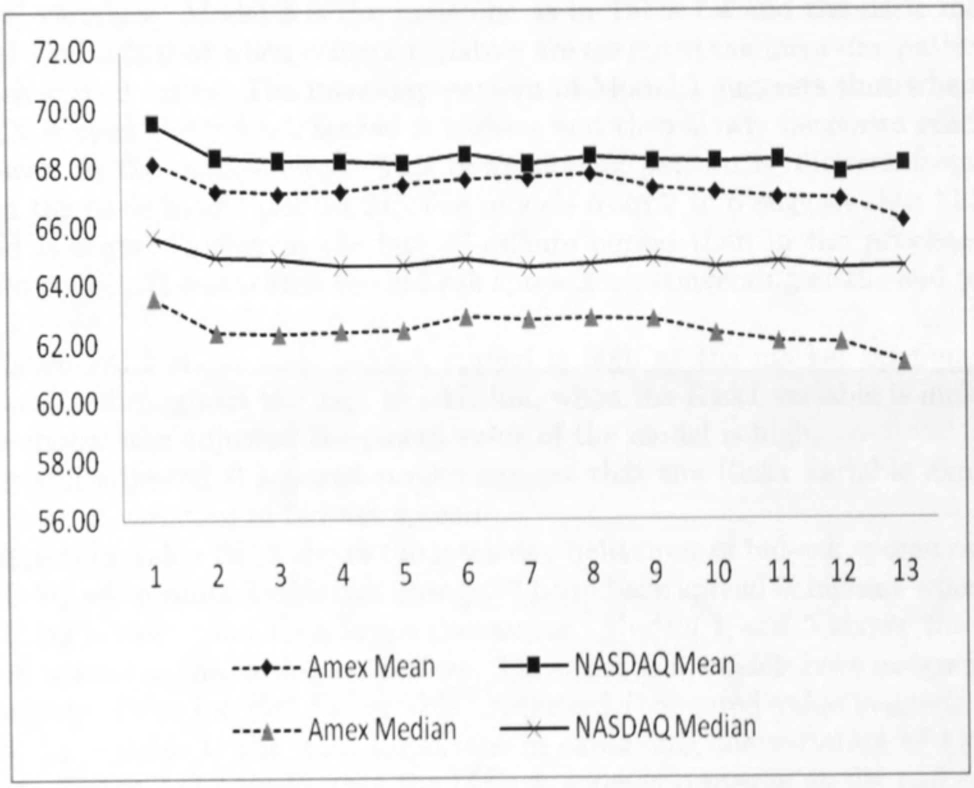


Figure 7.B.7: Trade Price

C Appendix 7.C

C.1 The Effects of Control Variables

C.1.1 The Relation of Bid-ask Spread and Control Variables

Section 7.6.1 suggests that the AMEX has a different pattern of time-weighted bid-ask spread from the pattern of previous research. Thus, this appendix examines whether this pattern is constant with the change of control variables.

Appendix Table 7.C.1 shows 6 regression models, each of which has different control variables. Model 2 is the same one as in Table 7.4 and the basic model. Model 1 reports that when control variables are excluded the intra-day pattern of bid-ask spread exists. The intra-day pattern of Model 1 suggests that when the AMEX is open the bid-ask spread is highest and then slowly decreases reaching its lowest at the market close. This down-sloping pattern is different from the one in the basic model (model 2). The models from 2 to 6 suggest that bid-ask spread is slightly higher in the last 30-minute period than in the previous 30-minute period. It seems that the bid-ask spreads are converging at the end of the day.

Figure 7.C.1 shows that bid-ask spread is high at the market opening and then stable throughout the day. In addition, when the Risk1 variable is included in the model, the adjusted R-squared value of the model is high.

Overall adjusted R-squared results suggest that the Risk1 variable explains large part of variation in bid-ask spread.

Appendix Table 7.C.2 shows the intra-day behaviour of bid-ask spread on the NASDAQ when control variables change. The bid-ask spread is highest when the NASDAQ is open, and then keeps decreasing. Models 1 and 3 shows that the bid-ask spread at the closing is smallest. The other four models have insignificant coefficients of the Interval 13 variable. Adjusted R-squared value suggests that the Risk1 variable is the most important in explaining the variation of bid-ask spread. Figure 7.C.2 shows that the bid-ask spreads converge at the end of the day, whereas in the morning the bid-ask spread is higher.

C.1.2 The Relation of Return Volatility and Control Variables

This section provides the intra-day pattern of return volatility, showing that return volatility is highest in the early morning, and then decreases before rising again at the market close. This pattern is evident in Table 7.C.3 and Table 7.C.4, including figures. This U-shaped pattern does not depend on the control variables. In terms of adjusted R-squared values, Trade and Price variables highly affect the variation of return volatility.

Table 7.C.1: Intra-day Pattern of Bid-ask Spread on the AMEX

*indicates statistical significance at the 0.10 level, ** at the 0.05 level, and *** at the 0.01 level. The dependent variable is Time-weighted bid-ask spread. Trades is the number of trades in each interval. Size is the log of average trade size in each interval. Risk1 is the time-weighted standard deviation of the bid-ask spread. Risk2 is a normalised value that is computed as the standard deviation of time-weighted bid-ask spread minus the mean of standard deviation/ the standard deviation of standard deviation. Nsize also is a normalised value that is computed as log (trade size of interval t – average trade size of total interval)/the standard deviation of trade size of total interval. Regional is the ratio of trades on the AMEX to trades on other exchanges in the case of AMEX regression. Price is the log of average trade price in each interval. Interval variables are dummy variables and represent 30-minute intervals. Interval 1 represents the time interval between 9:30 and 10:00, Interval 2 the time interval between 10:00 and 10:30, Interval 13 between 15:30 and 16:00. Each Interval variable is one when trading time falls into its own representing interval, otherwise zero. Hansen's Generalized Method of Moments is used to estimate coefficients.

	Model1	Model2	Model3	Model4	Model5	Model6
Intercept	0.00219***	0.009703***	0.012477***	0.013091***	0.009041*	0.001802***
Trade		-2.89E-6***	-3.58E-6***	-9.63E-7***	-8.28E-6*	
Size		-0.00027***	-0.0004***	-0.00043***	0.000048	
Risk1		0.488975***	0.14415***			0.51211***
Risk2		0.000292***		0.000627***		0.000291***
Nsize		0.001197***	0.001879***	0.002156***	-0.00314	
Regional		-0.00005***	-9.74E-6	-0.00006***	-0.00002	
Price		-0.0015***	-0.00193***	-0.002***	-0.00177***	
Interval 1	0.003788**	0.000276***	0.000775***	0.000252***	0.002754*	0.00014***
Interval 2	0.000271***	0.000145***	0.000263***	0.00035***	0.000354***	0.000093**
Interval 3	0.000378***	0.000067**	0.000276***	0.000372***	0.000393***	0.000025
Interval 4	0.000296***	0.00006*	0.000193***	0.00025***	0.000255***	0.000039
Interval 5	0.000294***	-0.00001	0.000163***	0.000195***	0.00022***	-0.00002
Interval 6	0.00031***	-0.00005	0.00016***	0.00015***	0.000199***	-0.00005
Interval 7	0.000116**	-0.00009***	0.000061	-2.76E-6	0.000068	-0.00011***
Interval 8	0.000047	-0.00008***	0.000014	-0.00008*	3.209E-6	-0.00008**
Interval 9	0.000048	-0.00004	-5.71E-6	-0.00005	-0.00004	-0.00002
Interval 11	-0.00007	0.000019	-0.00006	-0.00001	-0.00008**	0.000031
Interval 12	-0.0002***	-8.89E-6	-0.00016***	1.857E-7	-0.00016***	-6.46E-6
Interval 13	-0.00017***	0.000074**	-0.00012***	0.000116***	-0.00007	0.000066*
Adj R-Sq	0.0010	0.7413	0.9895	0.2625	0.0036	0.6682

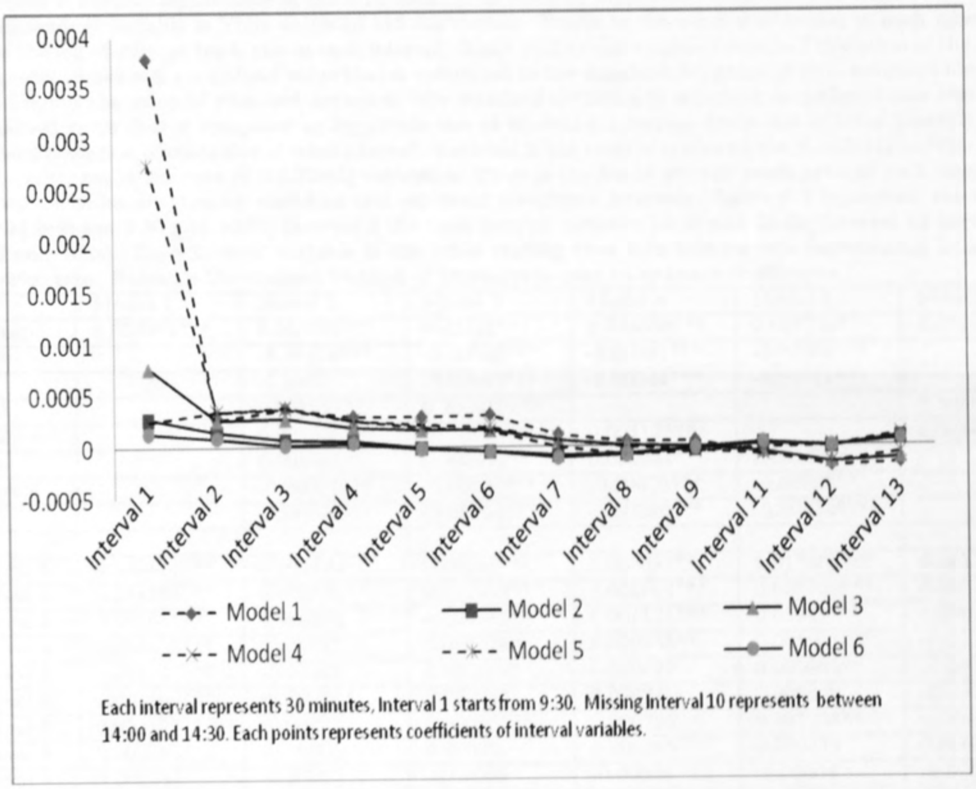


Figure 7.C.1: Intra-day Pattern of Bid-ask Spread on the AMEX

Table 7.C.2: Intra-day Pattern of Bid-ask Spread on the NASDAQ

*indicates statistical significance at the 0.10 level, ** at the 0.05 level, and *** at the 0.01 level. The dependent variable is Time-weighted bid-ask spread. Trades is the number of trades in each interval. Size is the log of average trade size in each interval. Risk1 is the time-weighted standard deviation of the bid-ask spread. Risk2 is a normalised value that is computed as the standard deviation of time-weighted bid-ask spread minus the mean of standard deviation/ the standard deviation of standard deviation. Nsize also is a normalised value that is computed as log (trade size of interval t – average trade size of total interval)/the standard deviation of trade size of total interval. Regional is the ratio of trades on the NASDAQ to trades on other exchanges in the case of NASDAQ regression. Price is the log of average trade price in each interval. Interval variables are dummy variables and represent 30-minute intervals. Interval 1 represents the time interval between 9:30 and 10:00, Interval 2 the time interval between 10:00 and 10:30, Interval 13 between 15:30 and 16:00. Each Interval variable is one when trading time falls into its own representing interval, otherwise zero. Hansen's Generalized Method of Moments is used to estimate coefficients.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	0.003983***	0.013607***	0.02252***	0.024568***	0.027729***	0.002988***
Trade		-8.39E-6***	-0.00001***	-0.00001***	-0.00002***	
Size		-0.0001	-0.00049***	-0.00044***	-0.00077***	
Risk1		0.551027***	0.200215***			0.545636***
Risk2		0.001366***		0.00412***		0.001469***
Nsize		0.001001*	0.002272**	0.002264**	0.001772	
Regional		-0.00027***	-0.00048***	-0.00076***	-0.0009***	
Price		-0.00228***	-0.00359***	-0.00386***	-0.00414***	
Interval 1	0.036574***	0.001738***	0.00948***	0.003581***	0.017058***	0.001365***
Interval 2	0.004138***	0.000642**	0.002252***	0.002604***	0.003135***	0.000597**
Interval 3	0.00183***	0.000075	0.000989***	0.001131***	0.00147***	0.000191
Interval 4	0.001801***	-0.00017	0.000574*	0.000955**	0.001001***	-0.00007
Interval 5	0.001747***	-0.00025	0.000569*	0.000622	0.000965**	-0.00005
Interval 6	0.001465***	-0.00031*	0.000365	0.00021	0.000639	-0.00016
Interval 7	0.002177***	-0.0002	0.000862**	0.000795	0.001338**	0.000073
Interval 8	0.000798**	-0.00005	0.00032	0.000045	0.000445	0.000063
Interval 9	0.000294	-0.00024	-0.00005	-0.00038	-0.00004	-0.00007
Interval 11	-0.00002	0.00005	-0.00006	-0.00005	-0.00012	0.000137
Interval 12	-0.00038	-0.00004	-0.00037*	-0.00006	-0.00047*	0.000028
Interval 13	-0.0006***	0.000013	-0.00035*	0.00035	-0.00037	-0.00007
Adj R-Sq	0.0223	0.7657	0.6851	0.0650	0.0218	0.7710

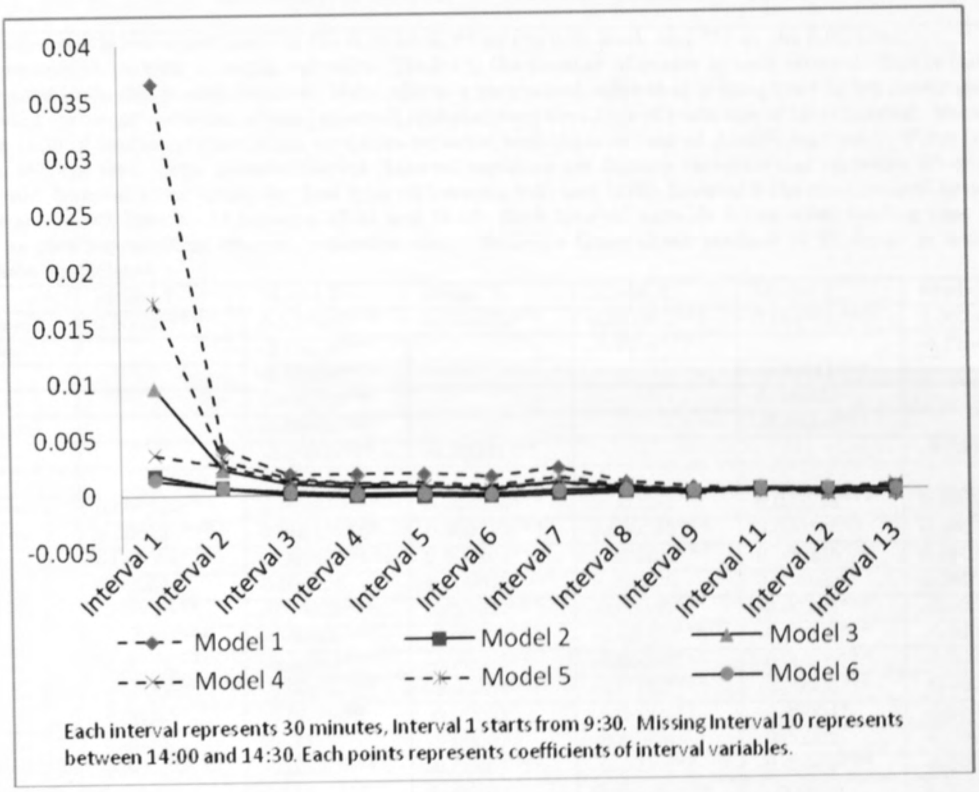


Figure 7.C.2: Intra-day Pattern of Bid-ask Spread on the NASDAQ

Table 7.C.3: Intra-day Pattern of Return Volatility on the AMEX

*indicates statistical significance at the 0.10 level, ** at the 0.05 level, and *** at the 0.01 level.
The dependent variable is return volatility. Trades is the number of trades in each interval. Size is the log of average trade size in each interval. Nsize also is a normalised value that is computed as log (trade size of interval t - average trade size of total interval)/the standard deviation of trade size of total interval. Regional is the ratio of trades on the AMEX to trades on other exchanges in case of AMEX regression. Price is the log of average trade price in each interval. Interval variables are dummy variables and represent 30-minute intervals. Interval 1 represents the time interval between 9:30 and 10:00, Interval 2 the time interval between 10:00 and 10:30, Interval 13 between 15:30 and 16:00. Each Interval variable is one when trading time falls into its own representing interval, otherwise zero. Hansen's Generalized Method of Moments is used to estimate coefficients.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	0.00091***	0.003524***	0.002573***	0.001011***	0.001708***	0.002438***
Trade		-3.65E-6***		-5.9E-6***		-4.8E-6***
Size		-0.00014***			-0.00011***	
Nsize		0.000642***			0.00006	
Regional		-0.00004***			-0.00009***	
Price		-0.00039***	-0.00041***			-0.00035***
Interval 1	0.000212***	0.000249***	0.000216***	0.000253***	0.000214***	0.00025***
Interval 2	0.000124***	0.000152***	0.000125***	0.000155***	0.00014***	0.000149***
Interval 3	0.000108***	0.000122***	0.000104***	0.000132***	0.000115***	0.000124***
Interval 4	0.000021	0.000026	0.000023	0.000026	0.000023	0.000027
Interval 5	-0.00005**	-0.00005***	-0.00005**	-0.00005***	-0.00005***	-0.00005***
Interval 6	-0.00008***	-0.00008***	-0.00008***	-0.00009***	-0.00008***	-0.00009***
Interval 7	-0.00005***	-0.00006***	-0.00005**	-0.00007***	-0.00006***	-0.00006***
Interval 8	-0.00005**	-0.00006***	-0.00004**	-0.00007***	-0.00006***	-0.00006***
Interval 9	-0.00003	-0.00004**	-0.00003	-0.00004**	-0.00004*	-0.00004*
Interval 11	-0.00001	-0.00002	-0.00002	-8.84E-6	-0.00002	-0.00001
Interval 12	-0.00007***	-0.00007***	-0.00007***	-0.00004**	-0.00009***	-0.00005***
Interval 13	0.000027	0.000049**	0.000021	0.000082***	0.000019	0.000067***
Adj R-Sq	0.0088	0.1394	0.0881	0.0685	0.0344	0.1263

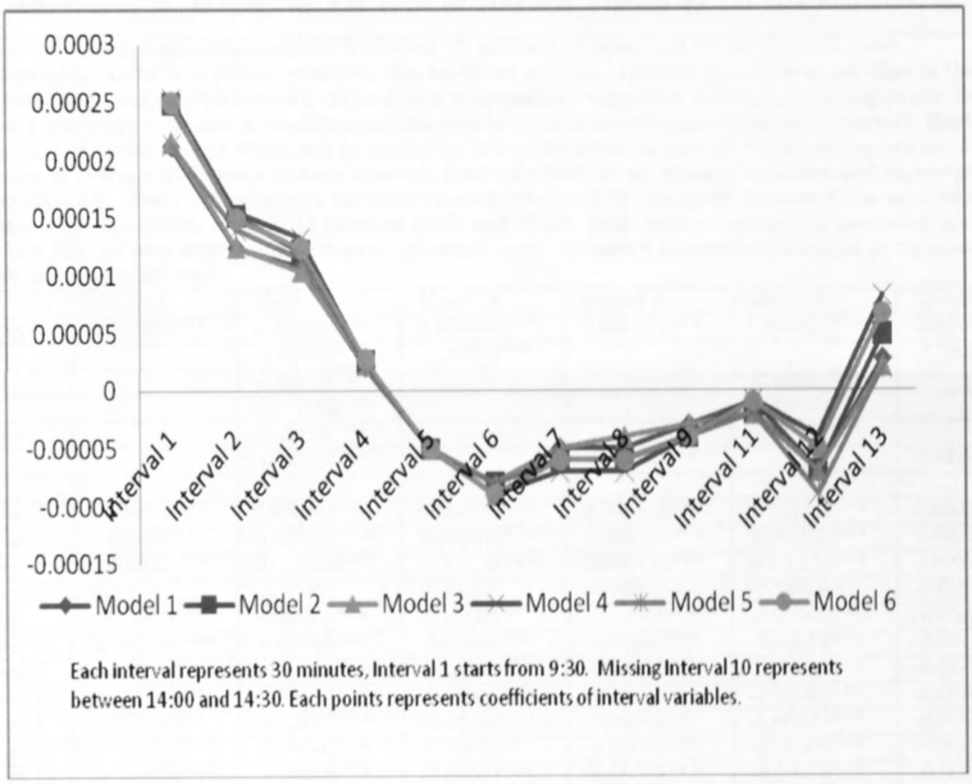


Figure 7.C.3: Intra-day Pattern of Return Volatility on the AMEX

Table 7.C.4: Intra-day Pattern of Return Volatility on the NASDAQ

*indicates statistical significance at the 0.10 level, ** at the 0.05 level, and *** at the 0.01 level.
The dependent variable is return volatility. Trades is the number of trades in each interval. Size is the log of average trade size in each interval. Nsize also is a normalised value that is computed as log (trade size of interval t – average trade size of total interval/the standard deviation of trade size of total interval). Regional is the ratio of trades on the NASDAQ to trades on other exchanges in case of NASDAQ regression. Price is the log of average trade price in each interval. Interval variables are dummy variables and represent 30-minute intervals. Interval 1 represents the time interval between 9:30 and 10:00, Interval 2 the time interval between 10:00 and 10:30, Interval 13 between 15:30 and 16:00. Each Interval variable is one when trading time falls into its own representing interval, otherwise zero. Hansen's Generalized Method of Moments is used to estimate coefficients.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	0.000958***	0.003405***	0.001038***	0.002271***	0.002457***	0.002324***
Trade		-1.82E-6***	-2.29E-6***			-2.26E-6***
Size		-0.00014***			-0.00021***	
Nsize		0.000659***			0.000931***	
Regional		-0.00007***			-0.00007***	
Price		-0.00032***		-0.00032***		-0.00031***
Interval 1	0.000312***	0.00038***	0.000367***	0.000314***	0.000353***	0.000368***
Interval 2	0.000185***	0.000217***	0.000216***	0.000183***	0.000203***	0.000214***
Interval 3	0.000127***	0.000144***	0.000146***	0.000126***	0.000137***	0.000144***
Interval 4	-5.1E-6	-0.00001	-5.81E-6	-6.92E-6	-7.19E-6	-7.59E-6
Interval 5	-0.000060**	-0.00008***	-0.00007***	-0.00007***	-0.00007***	-0.00007***
Interval 6	-0.000080***	-0.00009***	-0.00009***	-0.00008***	-0.00008***	-0.00009***
Interval 7	-0.000100***	-0.00012***	-0.00011***	-0.0001***	-0.00011***	-0.00011***
Interval 8	-0.000080***	-0.00009***	-0.00009***	-0.00008***	-0.00008***	-0.00009***
Interval 9	-0.000060**	-0.00007***	-0.00007***	-0.00006**	-0.00007***	-0.00007***
Interval 11	-0.000070***	-0.00006***	-0.00006***	-0.00006***	-0.00006***	-0.00006***
Interval 12	-0.000070***	-0.00008***	-0.00007***	-0.00008***	-0.00008***	-0.00007***
Interval 13	0.000096***	0.000117***	0.000121***	0.000096***	0.0001***	0.00012***
Adj R-Sq	0.0141	0.1272	0.0701	0.0537	0.0556	0.1082

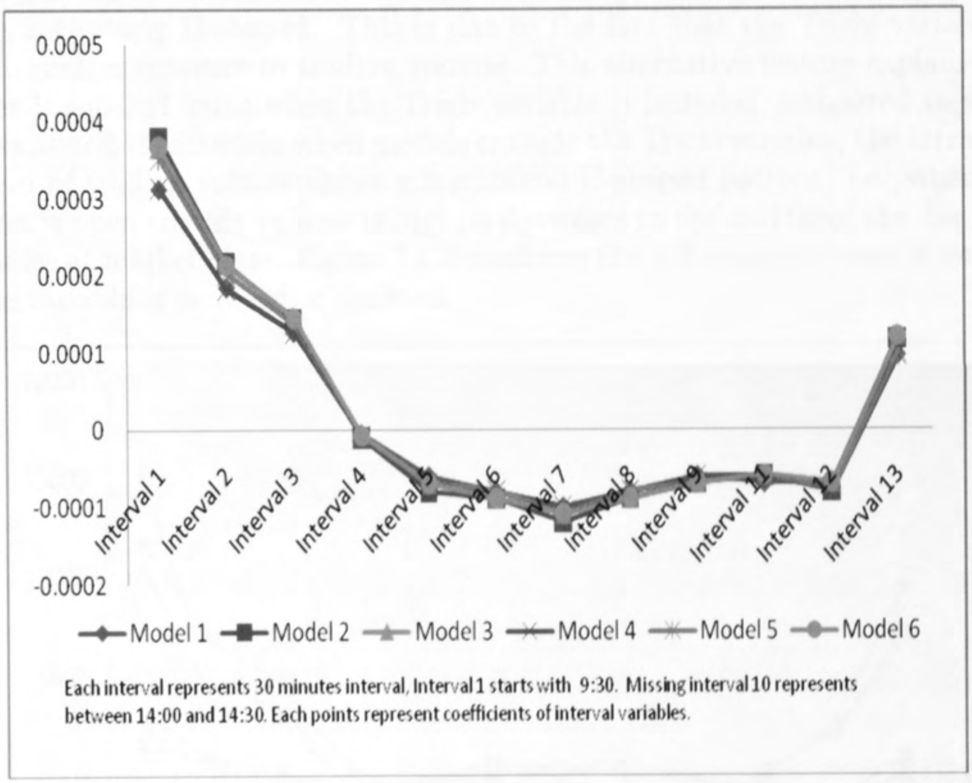
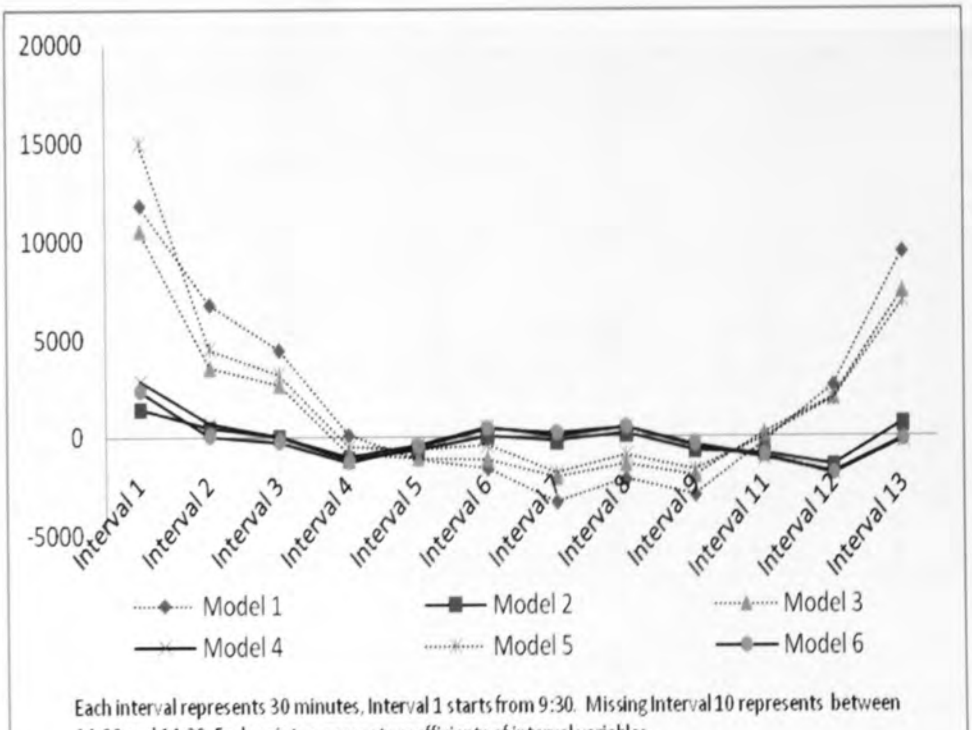


Figure 7.C.4: Intra-day Pattern of Return Volatility on the NASDAQ

C.1.3 The Relation of Trading Volume and Control Variables

This section provides the intra-day pattern of trade volume. Section 7.6.3 suggests that there is a difference in the intra-day pattern of trading volume between the AMEX and NASDAQ. On the AMEX trading volume at the closing is slightly increasing, while trading volume on the NASDAQ is decreasing. This section provides the evidence that the intra-day pattern of trading volume in section 7.6.3 depends on whether the Trade variable is employed in the control variables.

Table 7.C.5 shows that when the Trade variable is included in the model (Model 2, Model 4 and Model 6), the intra-day pattern shows trading volume to be not clearly U-shaped. This is due to the fact that the Trade variable is an alternative measure to trading volume. This alternative feature explains the higher R-squared value when the Trade variable is included, compared to when it is excluded. Meanwhile when models exclude the Trade variable, the intra-day pattern of trading volume shows a traditional U-shaped pattern. i.e. when the market is open trading volume is high, it decreases in the middle of the day, and increases at market close. Figure 7.C.5 confirms the difference between when the Trade variable is included or omitted.



Each interval represents 30 minutes, Interval 1 starts from 9:30. Missing Interval 10 represents between 14:00 and 14:30. Each points represents coefficients of interval variables.

Figure 7.C.5: Intra-day Pattern of Trading Volume on the AMEX

Table 7.C.5: Intra-day Pattern of Trading Volume on the AMEX

*indicates statistical significance at the 0.10 level, ** at the 0.05 level, and *** at the 0.01 level.
The dependent variable is trading volume. Trades is the number of trades in each interval. Size is the log of average trade size in each interval. Risk1 is the time-weighted standard deviation of the bid-ask spread. Risk2 is a normalised value that is computed as the standard deviation of time-weighted bid-ask spread minus the mean of standard deviation/ the standard deviation of standard deviation. Nsize also is a normalised value that is computed as log (trade size of interval t − average trade size of total interval)/the standard deviation of trade size of total interval. Regional is the ratio of trades on the NASDAQ (AMEX) to trades on other exchanges in the case of NASDAQ (AMEX) regression. Price is the log of average trade price in each interval. Interval variables are dummy variables and represent 30-minute intervals. Interval 1 represents the time interval between 9:30 and 10:00, Interval 2 the time interval between 10:00 and 10:30, Interval 13 between 15:30 and 16:00. Each Interval variable is one when trading time falls into its own representing interval, otherwise zero. Hansen's Generalized Method of Moments is used to estimate coefficients.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	12820***	-103568***	-233702***	12441***	-17875***	-65170***
Trade		**1025*****		**1108*****		**1030*****
Size		15872***	30120***			10432***
Risk1		56182***	315030***	-69177***	101483***	
Risk2		1347***	-3570***	1411***	-4258***	
Nsize		-59061***	-106396***	45843***	100477***	
Regional		-1451***	3044***	-83	6595***	
Price		1068***	13090***	-3158***	6437***	
Interval 1	11771***	1415	10524***	2824***	14925***	2352**
Interval 2	6722***	432	3517***	642	4460***	74
Interval 3	4390***	-39	2631	4	3165***	-231
Interval 4	70	-1195**	-902	-1050*	-557	-1284**
Interval 5	-1163	-713	-1117	-457	-661	-511
Interval 6	-1650*	-38	-1177	422	-423	401
Interval 7	-3375***	-273	-2117***	-4	-1874**	188
Interval 8	-2203*	101	-1422	457	-946	504
Interval 9	-3009***	-758	-2007**	-522	-1732*	-386
Interval 11	-287	-929*	139	-1117*	-67	-988*
Interval 12	2551**	-1522**	1966***	-1867***	1841*	-1902***
Interval 13	9346***	624	7375***	-182	6846***	-204
Adj R-Sq	0.0078	0.5725	0.2071	0.5394	0.0786	0.5654

Table 7.C.6 also shows that the pattern of trading volume on the NASDAQ as similar to the patterns on the AMEX. When the Trade variable is included, trading volume on the NASDAQ keeps decreasing from the market open to the market close. When the Trade variable is excluded, however, the intra-day pattern of trading volume on the NASDAQ has a traditional U-shaped pattern. In addition, the exclusion of the Trade variable leads to a lower adjusted R-squared value, implying that the Trade variable and trading volume measure the same thing. Figure 7.C.6 confirms the U-shaped pattern.

Table 7.C.6: Intra-day Pattern of Trading Volume on the NASDAQ

*indicates statistical significance at the 0.10 level, ** at the 0.05 level, and *** at the 0.01 level. The dependent variable is trading volume. Trades is the number of trades in each interval. Size is the log of average trade size in each interval. Risk1 is the time-weighted standard deviation of the bid-ask spread. Risk2 is a normalised value that is computed as(the standard deviation of time-weighted bid-ask spread minus the mean of standard deviation/ the standard deviation of standard deviation. Nsize also is a normalised value that is computed as log (trade size of interval t – average trade size of total interval)/the standard deviation of trade size of total interval. Regional is the ratio of trades on the NASDAQ to trades on other exchanges in the case of NASDAQ regression. Price is the log of average trade price in each interval. Interval variables are dummy variables and represent 30-minute intervals. Interval 1 represents the time interval between 9:30 and 10:00, Interval 2 the time interval between 10:00 and 10:30, Interval 13 between 15:30 and 16:00. Each I nterval variable is one when trading time falls into its own representing interval, otherwise zero. Hansen's Generalized Method of Moments is used to estimate coefficients.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	73069***	-129218***	-784273***	15579***	89286***	-138728***
Trade		**3180*** **		**3257*** **		**3185*** **
Size		20876***	124464***			20456***
Risk1		28007***	-79265**	15847**	-178990**	
Risk2		-2607**	-18458***	-2459**	-20031***	
Nsize		-23234***	-507109***	105533***	285771***	
Regional		-5419***	-14420***	-6041***	-20095***	
Price		-1252	15516***	-2513**	9706***	
Interval 1	90184***	21395***	95288***	25559***	135473***	18289***
Interval 2	44131***	429	26480***	1467	37740***	-73
Interval 3	20669**	-6241	9355	-5490	16969*	-6447
Interval 4	-1211	-5192	-5666	-4876	-3603	-5195
Interval 5	-12297	-4902	-11606	-4721	-11470	-5034
Interval 6	-17161**	-4922	-14449*	-4536	-13382*	-4795
Interval 7	-18693**	1417	-10735	1614	-11380	1266
Interval 8	-23421***	-6200	-20781***	-5766	-20215***	-6298
Interval 9	-21843***	-7740*	-16219**	-7707*	-17373**	-7783*
Interval 11	-5956	-6437	-5876	-6434	-5764	-6318
Interval 12	-2284	-10766**	-4737	-10894**	-4621	-10367**
Interval 13	22582**	-12755***	15070*	-13458***	14832	-12124**
Adj R-Sq	0.0073	0.7597	0.1545	0.7568	0.0364	0.7595

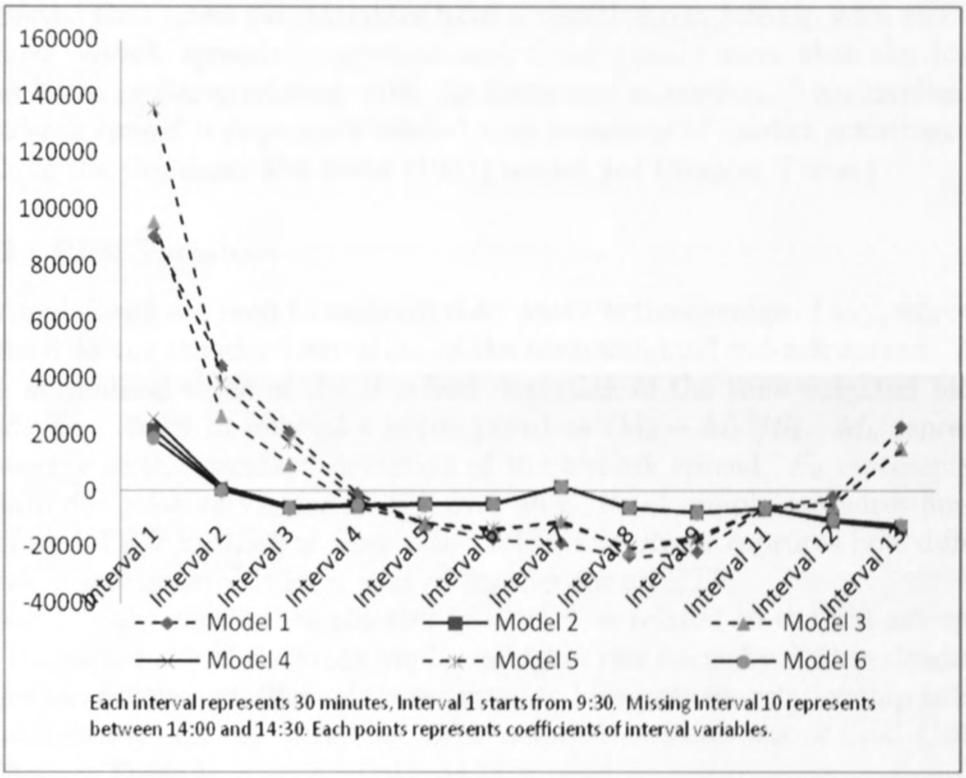


Figure 7.C.6: Intra-day Pattern of Trading Volume on the NADSAQ

D Appendix 7.D

D.1 Control Variables

D.1.1 Trading Activity Variables

Trades and *Size* variables represent the trading activity of an ETF in each interval. $Trade_{k,i}$ is the number of transactions for ETF k in interval i. $Size_{k,i}$ is the natural log of the average number of shares per trade for ETF k in interval i. Following McInish and Van Ness (2002) and McInish and Wood (1992), it is expected that these two variables have a negative relationship with the time-weighted bid-ask spreads. Copeland and Galai (1983) show that the bid-ask spread has a negative relation with the frequency of trading. This implies that the bid-ask spread is negatively related with measures of market activities. (For details of the Copeland and Galai (1983) model, see Chapter Three.)

D.1.2 Risk Variables

Risk1 and *Risk2* are used to measure risk. *Risk1* is the average of $V_{k,i}$, where $V_{k,i}$ is defined as the standard deviation of the time-weighted bid-ask spread. *Risk2* is the normalised value of the standard deviation of the time-weighted bid-ask spread. i.e. *Risk2* in interval i is computed as $(V_{ki} - M_k)/S_k$. M_k represents the average of the standard deviation of the bid-ask spread. S_k represents the standard deviation of $V_{k,i}$ for ETF k over all i. *Risk1* simply measures how the risk of an ETF k is different from other ETFs and *Risk2* captures how different the risk of one interval is from that of another for an ETF.

Stoll (1978b) shows that the risk of a stock is related to the bid-ask spread since the higher risk of the stock implies a higher risk borne by dealers, leading to a larger bid-ask spread. Thus, It is expected to be a positive relationship between bid-ask spreads and risk variables. (For detailed explanations of Stoll (1978b), see Chapter Three.)

D.1.3 Public Information and Regional Variables

Since the *Nsize* variable measures relatively abnormal trade size, *Nsize* represents the arrival of public information to the market. The *Nsize* variable is computed as the normalised value of the *Size* variable. Thus, $Nsize_{i,t}$ is calculated by $(Size_{i,t} - X_i)/D_i$, where X_i is the log of average $Size_i$ and D_i is the log of the standard deviation $Size_i$.

It is anticipated that the bid-ask spread is positively related to the *Nsize* variable. Glosten and Milgrom (1985) show that market-makers will increase their quoted bid-ask spread when it is highly probable that informed traders send orders to market-makers. Easley and O'Hara (1987)show that trade size reflects the information held by informed traders.

It is anticipated that a negative relation exists between the bid-ask spread and the level of competition. When market-makers compete against each other in quoting bid/ask prices, the bid-ask spread should decrease. The decrease in bid-ask spread is true when many exchanges compete against each other. Thus, the *Regional* variable is employed as a measure for the level of competition. The *Regional* variable is computed as the ratio of the number of shares of ETF i traded on all other exchanges, to the number of shares traded on one exchange. For instance, the *Regional* variable for AMEX regression model is the ratio of the number of shares traded on all other exchanges to the number of shares traded on AMEX. For NASDAQ regression models, the denominator is the number of shares traded on NASDAQ. (For the regional exchanges, see the robustness test of Chapter Five.)

D.1.4 Controling Independent Variables of Bid-ask Spread

The second column of Table 7.4 shows the regression results based on AMEX quotes and trades. The regression model is (7.21). The third column of Table 7.4 shows the regression results based on the NASDAQ. Control variables, except for interval variables, show the same pattern as those in the regression results for the AMEX.

Negative Trades and Sizes coefficients are interpreted to mean that when trading activity and trade size increase, the time-weighted bid-ask spread decreases. Regional and Price variables also have negative coefficients, implying that higher competition for posting bid/ask prices leads to a smaller bid-ask spread and a higher trading price causes the bid-ask spread to decrease.

Contrary to the negative coefficients of Trades and Size variables, the risk variables (Risk1 and Risk2) and Nsize variable have positive coefficients. As Risk1 and Risk2 variables rise, the time-weighted bid-ask spread increases. When the Nsize variable increases, so too does the time-weighted bid-ask spread.

D.1.5 Controlling Independent Variables of Return Volatility

The second column of Table 7.5 shows the regression results when return volatility is the dependent variable. The regression model is (7.22). Negative Trades and Sizes coefficients are interpreted as showing that return volatility decreases when trading activity and trade size increase. Regional and Price variables also have negative coefficients, implying that higher competition for posting bid/ask prices leads to a smaller return volatility and a higher trading price causes the return volatility to decrease. Contrary to the negative coefficients of Trades and Size variables, the Nsize variable has positive coefficients. As the Nsize variable rises, so too does return volatility.

The third column of Table 7.5 shows the regression results based on the NASDAQ. Control variables, except for interval variables, show the same pattern as those in the regression results for the AMEX.

D.1.6 Controlling Independent Variables of Trading Volume

The second column of Table 7.6 shows the regression results when trading volume is the dependent variable. The regression model is (7.23). The third column of Table 7.6 shows the regression results based on the NASDAQ. Positive Size coefficients are interpreted to mean that when trading activity increases trading volume rises. The Regional variable has negative coefficients, implying that higher competition for posting bid/ask prices leads to a smaller trading volume on the AMEX/NASDAQ. The Risk1 variable has a positive relationship with trading volume. While the NASDAQ Risk2 variable has a negative relationship with trading volume, the Risk2 variable on the AMEX has a positive coefficient.

Chapter 8

Conclusion

8.1 Conclusion

In this thesis, I have employed Exchange-Traded Funds (ETFs) to examine three questions. The first was to evaluate spread decomposition models. The second was to employ the probability of informed trading as a measure of information asymmetry to Exchange-Traded Funds and control securities. The third was to investigate the intraday pattern of trading patterns for Exchange-Traded Funds. To provide background information to my empirical work I reviewed the literature on bid-ask spreads in Chapter Two to Four. Chapter Five to Seven provide the empirical contributions of my thesis.

In Chapter Two, I reviewed three types of bid-ask spread: the quoted spread, the effective spread and the realised spread. In Chapter Three, I showed why the bid-ask spread is important for my thesis which examines information asymmetry in ETFs. I show that three main reasons exist for the existence of a spread: the order processing cost, the inventory holding cost and the adverse selection cost.

Chapter Four discussed how these components of the bid-ask spread are estimated. I discussed the five most popular spread decomposition models. There were introduced by Glosten and Harris (1988), George, Kaul, and Nimalendran (1991), Lin, Sanger, and Booth (1995), Madhavan, Richardson, and Roomans (1997) and Huang and Stoll (1997).

Chapter Four also discussed how often these spread decomposition models are used in financial research and reviewed the issues they have been used to exercise.

The review showed that spread decomposition models are extensively used because the adverse selection cost is a type of information asymmetry. Moreover the diverse spread decomposition models provide very different estimates of the adverse selection costs. This has generated a concern among financial researchers about which model should be employed. Although some work has been done on model evaluation, this area of work has achieved very little to do. For example, Neal and Wheatley (1998) evaluated the Glosten and Harris (1988) model and the George et al. (1991) model. Van Ness, Van Ness, and War (2001) evaluated the five spread decomposition models addressed in Chapter Four. Neal and Wheatley (1998) and Van Ness et al. (2001) do confirm that the estimated adverse selection costs varied according to spread decomposition models. Moreover, none of these papers have identified a preferred spread decomposition model.

For the empirical work in this thesis, Exchange-Traded Funds (ETFs) were employed. ETFs have some unique features when compared to ordinary stocks. ETFs are highly transparent seucirites and can be easily created/redeemed. Moreover, Subrahmanyam (1991) and Gorton and Pennacchi (1993) show that for a basket security such as an ETF a lower information asymmetry exists than for the weighted average information asymmetry of constituent securities. The lower information asymmetry of the basket security leads to lower adverse selection costs within the bid-ask spread.

Chapter five results suggest that when financial researchers employ one of spread decomposition models they need to be careful which model to be used. Even though the rejection of the empirical hypothesis does not mean the rejection of that spread decomposition model, the rejected model[1] does not capture what it should capture empirically. Thus, employing the rejected spread decomposition model for empirical research possibly misleads researchers to improper conclusion or findings. The reason is that, since the rejected model already once does not capture what it should capture, the rejected model may capture other empirical findings also improperly.

In Chapter Five, the features of ETFs and the work of Subrahmanyam (1991) provide a rationale for evaluating spread decomposition models, leading

[1]The rejected model means the model that does not accept the empirical hypothesis in Chapter five.

to the expectation that ETFs have lower comparison of adverse selection costs derived from ETFs and control securities. This is then applied to five spread decomposition models. The models of Glosten and Harris (1988) and Madhavan et al. (1997) showed that ETFs have lower adverse selection costs than control securities. In contrast, the models of George et al. (1991), Lin et al. (1995) and Huang and Stoll (1997) showed that ETFs have higher adverse selection costs than control securities. This results is used to evaluate the spread decomposition models.

In my thesis I distinguished broad-market ETFs and industry-wide ETFs, because broad-market ETFs should contain higher diversification in information asymmetry than industry-wide ETFs. The Glosten and Harris (1988) model and the Madhavan et al. (1997) model provided a lower adverse selection cost in broad-market ETFs and industry-wide ETFs than control securities. In the George et al. (1991) model, broad-market ETFs had significantly higher adverse selection costs than control securities, while the adverse selection costs of industry-wide ETFs did not show any difference to those of control securities. In the Lin et al. (1995) model, adverse selection cost did not show any difference between broad-market ETFs and matched control securities, while industry-wide ETFs displayed significantly higher adverse selection costs than control securities. Meanwhile, the Huang and Stoll (1997) model showed that higher adverse selection costs appeared in both broad-market ETFs and industry-wide ETFs, compared to control securities. I conclude therefore that only the Glosten and Harris (1988) and Madhavan et al (1997) models provide theoretically consistent results.

In Chapter Six, I used another measure of information asymmetry, that is, the probability of informed trading (PIN). While the adverse selection cost is a measure based on bid-ask spread, the probability of informed trading represents the probability that each transaction is generated by informed traders. When I compared the PIN of full ETFs and that of control securities, I found mixed results: full ETFs have insignificantly higher PINs than control securities in terms of average value, while full ETFs have significantly higher PINs than control securities in terms of median value. In the case of broad-market ETFs and matched control securities, the same level of average PINs and median

PINs appeared. Meanwhile, industry-wide ETFs had higher PINs than control securities, in terms of average and median values. All these results were confirmed by the regression model that I used to test the PIN difference between ETFs and control securities.

Chapter six results imply two contradictory things: if the empirical hypothesis is reasonable, then the higher PIN of ETFs is not acceptable and PIN measure may have some empirical faults. The second thing is that if PIN measure is to capture information asymmetry correctly then the empirical hypothesis is improperly used in this chapter. The empirical hypothesis is supported by theory papers (Subrahmanyam (1991) and Gorton and Pennacchi (1993)) and the characteristics of ETFs. PIN is developed by Easley et al (1996) and widely employed in empirical research. Therefore, chapter six cannot provide which one is correct and further future research is required on this topic.

The final empirical issue I investigated is contained in Chapter Seven. This investigates the intraday pattern of the bid-ask spread, return volatility, and trading volume for ETFs. Previous empirical research showed that while common stocks on the NYSE/AMEX have a U-shaped intraday pattern of bid-ask spreads, common stocks on NASDAQ display an L-shaped pattern. Previous research had also found a U-shaped pattern in intraday return volatility and trading volume. I found that the U-shaped intraday patterns did appear in the return volatility and trading volume of ETFs. Although ETFs had a U-shaped intraday bid-ask spread on AMEX, ETFs on NASDAQ displayed a different intraday pattern. ETFs on NASDAQ had higher bid-ask spreads when the market opened, the lowest bid-ask spread was at midday, bid-ask spreads at the close were slightly higher at the close of the market.

The intraday pattern of ETFs on NASDAQ is puzzling because higher return volatility and trading volume at the end of the day suggest higher information asymmetry. The bid-ask spread at the end of the day, however, is slightly elevated relative to midday spreads, implying that the bid-ask spreads of ETFs on NASDAQ are slightly higher after midday. This suggests that the higher information asymmetry at the end of the day does not affect the bid-ask spread of ETFs.

As I concluded the empirical results of my thesis, I added a weakness of my thesis. Chapter five and six results are based on control sampling method. Even though I followed the most commonly-accepted sampling method, there are still some problems. For instance, financial literature has not discussed what the current control sampling method is controlling for. There is a presumption that control sampling method control for inventory holding cost and order processing cost. No empirical research is provided so far. Thus, if control sampling method is not proper then chapter five and six need to change sampling methodology.

The best way to choose control stocks against ETFs is to compose the constituent stocks of ETFs as control stocks. By doing this, the problems inhabited in control sampling can be removed. When the constituent stocks of ETFs are used, the hypothesis is also fully supported by Subrahmanyam (1991).

References

Abhyankar, A. D., E. Levin Ghosh, and R.J. Limmack, 1997, Bid-ask spreads, trading volume and volatility: Intra-day evidence from the london stock exchange, Journal of Business Finance and Accounting 24, 343–361. 170, 188

Ahn, Hee-Joon, Jun Cai, Yasushi Hamao, and Richard Y.K. Ho, 2002, The components of the bid-ask spread in a limit-order market: evidence from the tokyo stock exchange, Journal of Empirical Finance 9, 399–430. 74

Ahn, Hee-Joon, and Yan-Leung Cheng, 1999, The intraday patterns of the spread and depth in a market without market makers: The stock exchange of hong kong, Pacific-Basin Finance Journal 7, 539–556. 189

Alexander, Gordon J., Evren Ors, Mark A. Peterson, and Paul J. Seguin, 2004, Margin regulation and market quality: a microstructure analysis, Journal of Corporate Finance 10, 549–574. 74

Amihud, Yakov, and Haim Mendelson, 1980, Dealership market:market making with inventory, Journal of Financial Economics 8, 31–53. 20, 21, 22, 180

Amihud, Yakov, and Heim Mendelson, 1987, Trading mechanisms and stock returns: An empirical investigation, Journal of Finance 42, 533–553. 189

Andersen, Torben G., Tim Bollerslev, and Jun Cai, 2000, Intraday and interday volatility in the japanese stock market, Journal of International Financial Markets, Institutions and Money 10, 107–130. 189

Bagehot, Walter, 1971, The only game in town, Financial Analysts Journal 27. 36

Barclay, Michael J., and Terrence Hendershott, 2004, Liquidity externalities and adverse selection:evidence from trading after hours, Journal of Finance 59, 681–710. 8, 9, 15

Barclay, Michael J., and Jerold B. Warner, 1993, Stealth trading and volatility: which trades move prices?, Journal of Financial Economics 34, 281–305. 42, 62

Battalio, Robert, Brian Hatch, and Robert Jennings, 2004, Toward a national market system for u.s. exchange-listed equity options, Journal of Finance pp. 933–962. 8, 9

Biais, Bruno, Pierre Hillion, and Chester Spatt, 1999, Price discovery and learning during the preopening period in the paris bourse, Journal of Political Economy 107, 1218–1248. 124

Blume, Marshall E., and Michael A. Goldstein, 1992, Differences in execution prices among the nyse, the regionals, and the nasd, Rodney L. White Center for Financial Research Paper 4. 9

Boehmer, Beatrice, and Ekkehart Boehmer, 2003, Trading your neighbor's etfs: Competition or fragmentation, Journal of Banking & Finance 27, 1667–1703. 96, 104

Brockman, Paul, and Dennis Y. Chung, 2001, Managerial timing and corporate liquidity: Evidence from actual share repurchases, Journal of Financial Economics 61, 417–448. 74, 75, 104

Brown, Stephen, Stephen A. Hillegeist, and Kin Lo, 2004, Conference calls and information asymmetry, Journal of Accounting and Economics 37, 343–366. 143

Cai, Charlie X., Robert Hudson, and Kevin Keasey, 2004, Intra day bid-ask spreads, trading volume and volatility: Recent empirical evidence from the london stock exchange, Journal of Business Finance and Accounting 31, 647–676. 188

Cao, Charles, Eric Ghysels, and Frank Hatheway, 2000, Price discovery without trading: Evidence from the nasdaq preopening, Journal of Finance 55, 1339–1365. 124

Chakravarty, Sugato, Bonnie F. Van Ness, and Robert A. Van Ness, 2005, The effect of decimalization on trade size and adverse selection costs, Journal of Business Finance & Accounting 32, 1063–1081. 74

Chan, K.C., William G. Christie, and Paul H. Schultz, 1995, Market structure and the intraday pattern of bid-ask spreads for nasdaq securities, Journal of Business 68, 35–60. 170, 186, 187

Chan, Kalok, Y. Peter Chung, and Herb Johnson, 1995, The intraday behavior of bid-ask spreads for nyse stocks and cboe options, Journal of Financial and Quantitative Analysis 30, 329=346. 187

Chan, L, and J Lakonishok, 1993, Institutional trades and intra-day stock price behaviour, Journal of Financial Economics 33, 173–199. 139

Chan, L, and J. Lakonishok, 1995, The behaviour of stock prices around institutional trades, Journal of Finance 50, 1147–1174. 139

Chiyachantana, Chiraphol N., Christine X. Jiang, Nareerat Taechapiroontong, and Robert A. Wood, 2004, The impact of regulation fair disclosure on information asymmetry and trading: An intraday analysis, The Financial Review 39, 549–577. 76, 78

Choi, J.Y., Dan Salandro, and Kuldeep Shastri, 1988, On the estimation of bid-ask spreads: theory and evidence, Journal of Financial and Quantitative Analysis 23. 13

Chou, Robin K., and Huimin Chung, 2006, Decimalization, trading costs, and information transmission between etfs and index futures, Journal of Futures Markets 26, 131–151. 74

Chung, Kee H., Chairat Chuwonganant, and D. Timothy McCormick, 2004, Order preferencing and market quality on nasdaq before and after decimalization, Journal of Financial Economics 71, 581–612. 99

Chung, Kee H., Mingsheng Li, and Thomas H. McInish, 2005, Information-based trading, price impact of trades, and trade autocorrelation, Journal of Banking & Finance 29, 1645–1669. 139

Chung, Kee H., and Xin Zhao, 2003, Intraday variation in the bid-ask spread: Evidence after the market reform, Journal of Financial Research 26, 191–206. 187, 195

Clarke, Jonathan, and Kuldeep Shastri, 2001, Adverse selection costs and closed-end funds, Working Paper University of Pittsburgh. 100

Clarke, Jonathan E., C. Edward Fee, and Shawn Thomas, 2004, Corporate diversification and asymmetric information: evidence from stock market trading characteristics, Journal of Corporate Finance 10, 105–129. 74

Conrad, Jennifer, and Gautam Kaul, 1988, Time-variation in expected returns, Journal of Business 61, 409–425. 59

Copeland, Thomas E., and Dan Galai, 1983, Information effects on the bid-ask spread, Journal of Finance 38, 1457–1469. 36, 37, 38, 43, 239

Covrig, Vicentiu, and Lilian Ng, 2004, Volume autocorrelation, information, and investor trading, Journal of Banking & Finance 28, 2155–2174. 139

De Winne, Rudy, and Christophe Majois, 2004, A comparison of alternative spread decomposition models on euronext brussels, Brussels Economic Review 46, 91–135. 79, 80

Demsetz, H, 1968, The cost of transaction, Quarterly Journal of Economics 82, 33–53. 17, 19

Dennis, Patrick J., and James P. Weston, 2001, Who's informed? an analysis of stock ownership and informed trading, Working Paper, University of Virginia. 78

Diamond, D.W., and R.E. Verrecchia, 1987, Contraints on short-selling and asset price adjustment to private information, Journal of Financial Economics 18, 277–311. 95

Domowitz, Ian, Jack Glen, and Ananth Madhavan, 1998, International cross-listing and order flow migaration: Evidence from an emerging market, Journal of Finance 53, 2001–2027. 74

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2004, Factoring information into returns, Working paper Cornell University. 142

Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman, 1996, Liquidity, information, and infrequently traded stocks, Journal of Finance 51, 1405–1436. 132

Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, Journal of Financial Economics 19, 69–90. 43, 45, 46, 62, 240

Ekman, Peter D., 1992, Intraday patterns in the s&p 500 index futures market, Journal of Futures Markets 12, 365–381. 188

Elton, Edwin J., Martin J. Gruber, George Comer, and Kai Li, 2002, Spiders: Where are the bugs?, Journal of Business 75, 453–472. 96, 104

Flannery, Mark J., Simon H. Kwan, and M. Nimalendran, 2004, Market evidence on the opaqueness of banking firms' assets, Journal of Financial Economics 71, 419–460. 73

French, Kenneth R., and Richard Roll, 1986, Stock return variances: The arrival of information and the reaction of traders, Journal of Financial Economics 17, 5–26. 184

Fu, Huijing, 2002, Information asymmetry and the market reaction to equity carve-outs, Working Paper Texas Christian University. 133

Fuller, Kathleen P., 2003, The impact of informed trading on dividend signaling: a theoretical and empirical examination, Journal of Corporate Finance 9, 385–407. 133

Garbade, K., and Z. Lieber, 1977, On the independence of transactions on the new york stock exchange, Journal of Banking and Finance pp. 151–172. 12

Garman, M, 1976, Market microstructure, Journal of Financial Economics 5, 257–275. 19, 20, 21, 22, 30

Gastineau, Gary L., 2001, Exchange-traded funds: An introduction, The Journal of Portfolio Management 27, 88–96. 93

George, T.J., G Kaul, and M. Nimalenddran, 1991, Estimation of the bid-ask spread and its components: A new approach, Review of Financial Studies 4, 623–656. 56, 59, 66, 100, 107

Gerety, Mason S, and J Harold Mulherin, 1992, Trading halts and market activity: An analysis of volume at the open and the close, Journal of Finance 47, 1765–1784. 185, 210

Gibson, Scott, Rajdeep Singh, and Vijay Yerramilli, 2003, The effect of decimalization on the components of the bid-ask spread, Journal of Financial Intermediation 12, 121–148. 74, 75

Glascock, John L., William T. Hughes, and Sanjay B. Varshney, 1998, Analysis of reit ipos using a market microstructure approach: Anomalous behaviour or asset structure, Journal of Real Estate Finance and Economics 16, 243–256. 74

Glosten, L.R, and Larry E. Harris, 1988, Estimating the components of the bid-ask spread, Journal of Financial Economics 21, 123–142. 56, 57, 58, 59, 60, 66, 73, 75, 78, 79, 80, 100, 104, 107, 110, 111, 122, 123, 126, 127, 128, 129, 130

Glosten, L., and P. Milgrom, 1985, Bid ask and transaction prices in a specialist market with heterogeneously informed traders, Journal of Financial Economics 13, 71–100. 38, 41, 42, 43, 239

Gorton, Gary b., and George G. Pennacchi, 1993, Security baskets and index-linked securities, Journal of Business 66, 1–27. 82, 83, 92

Greene, Jason, and Scott Smart, 1999, Liquidity provision and noise trading: Evidence from the "investment dartboard" column, Journal of Finance 54, 1885–1899. 73, 74

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, Econometrica 50, 1029–1054. 193

Hasbrouck, Joel, 1991, Measuring the information content of stock trades, Journal of Finance 46, 179–207. 139

———, and George Sofianos, 1993, The trades of market makers: An empirical analysis of nyse specialists, Journal of Finance 48, 1565–1593. 181

Hatch, Brian C., and Shane A. Johnson, 2002, The impact of specialist firm acquisitions on market quality, Journal of Financial Economics 66, 139–167. 72, 73, 74, 122

Heflin, Frank, and Kenneth W. Shaw, 2000, Blockholder ownership and market liquidity, Journal of Financial and Quantitative Analysis 35, 621–633. 78

Hegde, Shantaram P., and John B. McDermott, 2003, The liquidity effects of revisions to the s&p 500 index: an empirical analysis, Journal of Financial Markets 6, 413–459. 73, 74

Heidle, Hans G., and Roger D. Huang, 2002, Information-based trading in dealer and auction markets: An analysis of exchange listings, Journal of Financial and Quantitative Analysis 37, 391–424. 133

Hendershott, Terrence, and Charles M. Jones, 2005, Island goes dark: Transparency, fragmentation, and regulation, Review of Financial Studies 18, 743–793. 15

Ho, Thomas, and Hans Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, Journal of Financial Economics 9, 47–73. 26, 27, 29, 30, 33

Ho, Thomas S. Y., and Richard G. Macris, 1984, Dealer bid-ask quotes and transaction prices: An empirical study of some amex options, Journal of Finance 39, 23–45. 180

Hong, Harrison, and Jiang Wang, 2000, Trading and returns under periodic market closure, Journal of Finance 55, 297–354. 183

Huang, Richard D., and Hans R. Stoll, 1996, Delaer versus auction markets: a paired comparison of execution costs on nasdaq and the nyse, Journal of Financial Economics 41, 313–357. 8, 9, 14, 99, 123, 124

———, 1997, The components of the bid-ask spread: A general approach, Review of Financial Studies 10, 995–1034. 56, 69, 70, 71, 72, 107

Jain, Prem C., and Gun-Ho Joh, 1988, The dependence between hourly prices and trading volume, Journal of Financial and Quantitative Analysis 23, 269–283. 170, 185, 210

Jiang, Christine X., and Jang-Chul Kim, 2005, Trading costs of non-us stocks on the new york stock exchange: The effect of institutional ownership, analyst following, and market regulation, Journal of Financial Reseach 28, 439–459. 78

Kelly, David l., and Douglas G. Steigerwald, 2004, Private information and high-frequency stochastic volatility, Studies in Nonlinear Dynamics & Econometrics 8. 139

Kim, In Joon, Kwangsoo Ko, and Seok Kyun Noh, 2002, Time-varying bid-ask components of nikkei 225 index futures on simex, Pacific-Basin Finance Journal 10, 183–200. 74

Kofman, Paul, and James T Moser, 1997, Spreads, information flows and transparency across trading systems., Applied Financial Economics 7, 281–294. 74

Kostovetsky, Leonard, 2003, Index mutual funds and exchange-traded funds, Journal of Portfolio Management 29, 80–92. 93

Kumar, Raman, Atulya Sarin, and Kuldeep Shastri, 1998, The impact of options trading on the market quality of the underlying security: an emprical analysis, Journal of Finance 53, 717–732. 74, 75

Kyle, Albert S., 1985, Continuous auctions and insider trading, Econometrica 53, 1315–1336. 83, 173

Laux, Paul A., 1995, Dealer market structure, outside competition, and the bid-ask spread, Journal of Economic Dynamics and Control 19, 683–710. 22

Lee, Charles M.C., Belinda Mucklow, and Mark J. Ready, 1993, Spreads, depths, and the impact of earnings information: An intraday analysis, Review of Financial Studies 6, 345–374. 46, 186

Lin, Ji-Chai, Gary C. Sanger, and G. Geoffrey Booth, 1995, Trade size and components of the bid-ask spread, Review of Financial Studies 8, 1153–1183. 56, 66, 107

Liu, Qiao, and Rong Qi, 2006, Do we accept accrual profits at our peril?, Financial Analysts Journal 62. 134

Lockwood, Larry J., and Scott C. Linn, 1990, An examination of stock market return volatility during overnight and intraday periods 1964-1989, Journal of Finance 45, 591–601. 185, 207

Madhavan, Ananth, 1992, Trading mechanisms in securities markets, Journal of Finance 47, 607–641. 124, 182

Madhavan, Anand, M. Richardson, and M. Roomans, 1997, Why do security prices changes? a transaction-level analysis of nyse stocks, Review of Financial Studies 10, 1035–1064. 56, 65, 104

Madhavan, Ananth, and Seymour Smidt, 1991, A bayesian model of intraday specialist pricing, Journal of Financial Economics 30, 99–134. 180

———, 1993, An analysis of changes in specialist inventories and quatations, Journal of Finance 48, 1595–1628. 181

Madhavan, Ananth, and George Sofianos, 1998, An empirical analysis of nyse specialist trading, Journal of Financial Economics 48, 189–210. 181, 187

Madhavan, Smidt A., 2000, Market microstructure: a survey, Journal of Financial Markets 3, 205–258. 103

McDonald, Cynthia G, Terry D Nixon, and V Carlos Jr Slawson, 2000, The changing asymmetric information component of reit spreads: A study of anticipated announcements, Journal of Real Estate Finance & Economics 20, 195–210. 74

McInish, Thomas H., and Bonnie F. Van Ness, 2002, An intraday examination of the components of the bid-ask spread, The Financial Review 37, 507–524. 184

———, and Robert A. Van Ness, 2001, Market changes and spread components, implications for international markets, Journal of International Financial Markets, Institutions and Money 11, 31–42. 74

McInish, Thomas H., and Robert A. Wood, 1992, An analysis of intraday patterns in bid/ask spreads for nyse stocks, Journal of Finance 47, 753–764. 170, 184, 185

Neal, Robert, and Simon M. Wheatley, 1998, Adverse selection and bid-ask spreads:evidence from closed-end funds, Journal of Financial Markets 1, 121–149. 63, 79, 100

O'Hare, Maureen, and G. Oldfield, 1986, The microeconomics of market making, Journal of Financial and Quantitative Analysis 21, 361–376. 30, 32, 33, 34, 35

Petersen, Mitchell A., and David Fialkowski, 1994, Posted versus effective spreads: Good prices or bad quotes?, Journal of Financial Economics 335, 269–292. 9

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spreads in an efficient market, Journal of Finance 39, 1127–1139. 8, 10, 13

Sarin, Atulya, Karen A. Shastri, and Kuldeep Shastri, 2000, Ownership structure and stock market liquidity, Working Paper University of Pittsburgh. 78, 79

Schnitzlein, Charles R., 1996, Call and continuous trading mechanisms under asymmetric information: An experimental investigation, Journal of Finance 51, 613–636. 124

Sheikh, Aamir M., and Ehud I. Ronn, 1994, A chanracterization of the daily and intraday behavior of returns on options, Journal of Finance 49, 557–579. 188

Stoll, Hans, 2000, Friction, Journal of Finance 55, 1479–1514. 104

Stoll, Hans R., 1978a, The pricing of security dealer services: An empirical study of nasdaq, Journal of Finance 33, 1153–1172. 19

Stoll, Hans R, 1978b, The supply of dealer services in securities markets, Journal of Finance 33, 1133–1151. 22, 24, 25, 26, 29, 194, 239

Stoll, Hans R., and Robert E. Whaley, 1990, Stock market structure and volatility, Review of Financial Studies 3, 37–71. 190

Straser, Vesna, 2002, Regulation fair disclosure and information asymmetry, Working Paper, University of Notre Dame - Department of Economics. 78

Subrahmanyam, Avanidhar, 1991, A theory of trading in stock index futures, Review of Financial Studies 4, 17–51. 82, 83, 85

Tse, Yiuman, and Grigori Erenburg, 2003, Competition for order flow, market quality, and price discovery in the nasdaq 100 index tracking stock, Journal of Financial Research 26, 301–318. 96

Van Ness, Bonnie F., Robert A. Van Ness, and Richard S. Warr, 2001, How well do adverse selection components measure adverse selection?, Financial Management pp. 5–30. 79, 80

Vega, Clara, 2006, Stock price reaction to public and private information, Journal of Financial Economics 82, 103–133. 134, 143

Venkatesh, P. C., and R. Chiang, 1986, Information asymmetry and the dealers bid-ask spread: A case study of earnings and dividend announcements, Journal of Finance 41, 1089–1102. 42

Wei, Pei-Hwang, 1992, Intraday variations in trading activity, price variability, and the bid-ask spread, Journal of Financial Reseach 15, 265–276. 74, 75

Werner, Ingrid M., and Allan W. Kleidon, 1996, U.k. and u.s. trading of british cross-listed stocks: An intraday analysis of market integration, Review of Financial Studies 9, 619–664. 188

Weston, James P., 2000, Competition on the nasdaq and the impact of recent market reforms, Journal of Finance 55, 2565–2598. 74, 75

Wood, Robert A., Thomas H. McInish, and J. Keith Ord, 1985, An investigation of transactions data for nyse stocks, Journal of Finance 40, 723–739. 185

Yan, Yuxing, and Shaojun Zhang, 2006, An improved estimation method and empirical properties of the probability of informed trading, Working paper Nanyang Technological University. 142, 143