PRICE DISCOVERY AND LIQUIDITY IN A FRAGMENTED STOCK MARKET

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Mao Ye

January 2011

PRICE DISCOVERY AND LIQUIDITY IN A FRAGMENTED STOCK MARKET

Mao Ye, Ph. D.

Cornell University 2011

One of the most striking changes in U.S. equity markets has been the proliferation of trading venues. My dissertation studies the impact of market fragmentation on liquidity and price discovery from three different perspectives.

The first section, coauthored with Maureen O'Hara, examines how fragmentation of trading is affecting the quality of trading. We use newly-available trade reporting facilities volumes to measure fragmentation levels in individual stocks, and we use a matched sample to compare execution quality and efficiency of stocks with more and less fragmented trading. We find market fragmentation generally reduces transaction costs, as measured by effective spread and realized spread, and increases execution speeds. Fragmentation does increase short-term volatility, but prices are more efficient in that they are closer to being a random walk.

The second section focuses on a particular type of new trading mechanism, crossing network, in which buy and sell orders are passively matched using the price set by the stock exchange. The results show that the crossing network harms price discovery and the relative lack of revealed information most strongly affects stocks with high uncertainty in their fundamental values. I find that an increase in the uncertainty of the fundamental value of the asset increases the transaction costs in both markets, but stocks with higher fundamental value uncertainty are more likely to have higher market shares in the crossing network. The impact of different allocation rules in the

crossing network on market outcomes is also examined.

The third section tests the theoretical prediction of the second essay. I find that crossing networks have lower effective spread and price impact of trade, but they also have lower execution probability and speed of trade. Non-execution is positive correlated with price impact, decreases in trading volume and increases in volatility. Crossing networks have higher market share for stocks with lower volatility and higher volume. We also find that the underlying assumption in previous literature, that stocks with higher effective spreads have higher reductions in effective spread by trading in crossing networks, is not supported by data.

BIOGRAPHICAL SKETCH

Mao Ye was born in Yangzhou, China. He earned his bachelor's degree in accounting at Southeast University, China; a master's degree in finance from Renmin University, China; and a master's degree in economics at the University of British Columbia. He began his Ph.D. study at Cornell University at 2005 and passed both microeconomics and macroeconomics Ph.D. qualification exams without taking first year courses. During his five years at Cornell, he made two other achievements in addition to his academic ones. In 2006, he was elected as a trustee by Cornell students, making him the first trustee from Mainland China among all Ivy League Institutions. In October 5, 2008, he became father of Cornelia, who has just completed her “study” at Cornell Child Care Center infant room 5.

To My Parents, Xi and Cornelia

ACKNOWLEDGMENTS

The most difficult part to write in my dissertation is the acknowledgement. No matter how hard I recollect, I am still afraid that I may miss someone because of the number of people who has helped me.

My interest in theoretical market microstructure started from the course I took from Professor David Easley. Professor Maureen O’Hara inspired my research interest in empirical market structure through her course. Both courses required a paper, but the papers I turned in are not part of this dissertation, because Professor Easley and Professor O’Hara have spent huge amount of time on helping me to find more interesting and well-defined research questions. I called Professor Gideon Saar when I met some difficulties in writing the term paper for Professor O’Hara’s class, and he gave me detailed instructions even before he knew who I was. Certainly, he gave me even more help after he joined my Ph.D. committee.

Many other faculty members from the Economics Department, Johnson School of Management, Applied Economics and Management and Hotel School at Cornell University either commented on different versions of my dissertation, or helped me with the presentation of the dissertation. They are Talia Bar, Levon Barseghyan, Daniel Benjamin, Yaniv Grinstein, Yongmiao Hong, Ming Huang, Bob Jarrow, Andrew Karolyi, Mark Leary, Peter Liu, Qingzhong Ma, Karel Mertens, Roni Michaely, David Ng, Viktor Tsyrennikov and Xiaoyan Zhang. Keith Hjortshoj taught me how to edit my dissertation through his course in writing and my way of presenting my dissertation was polished through the help from Kimberly Kenyon and Stew Markel at Cornell Center for Teaching Excellence.

I learnt from Professor Danyang Xie 7 years ago that I would learn as much from my fellow students as I would learn from my professors. My fellow students Alyssa Anderson, Haiqiang Chen, David De Angelis, Ram Dubey, Shawn Kong, Yelena

Larkin, Hao Li, Maximilian Mihn, Zhaogang Song, Russell Toth, Liheng Xu, Liyan Yang and Zhuo Zhong either gave me comment on my dissertation, or helped me when I had difficulty in math or programming. Their help certainly supports Professor Xie's hypothesis.

The financial support from NASDAQ Education Foundation Ph.D. dissertation fellowship is highly appreciated. I also would like to thank Eric Maroney, Daniel Wszolek and Professor Jennifer Wissink. They provided me with much better working conditions or support than what I could even dream to get as a Ph.D. student.

I would not know that the theoretical model in the second part of my dissertation could be solved analytically without the help of my wife Xi Yang. As usual, I would like to thank my parents. Their curiosity for research inspired me to pursue an academic career. I also want to thank my parents in law, who went 6,000 miles from China to help us take care of my daughter Cornelia. Without their help, this dissertation would be completed at least one year later.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

IS MARKET FRAGMENTATION HARMING MARKET QUALITY?

One of the more striking changes in U.S. equity markets has been the proliferation of trading venues. While the traditional exchanges continue to execute orders, they now face a host of competitors ranging from electronic platforms such as ECNS (electronic communication networks) and ATS (alternative trading systems), to the trading desks of broker/dealer firms, and even to a variety of new entrants such as futures and options markets. The addition of these new trading venues has created a marketplace in which equity trading can take place in ways and places unimagined but a few years ago. And these changes are not just confined to U.S. markets. European equity trading has seen dramatic growth of electronic platforms such as Chi-X and BATS, and even Canada, where the Toronto Stock Exchange enjoyed a virtual monopoly on trading, has experienced fragmentation with the addition of electronic venues Alpha, Pure and MATCH Now[1].

What is less clear is how this fragmentation of trading is affecting the quality of trading. Certainly, the addition of new trading venues has increased competition, forcing the traditional exchanges to lower trading charges and other fees.[2] The proliferation of venues has also provided a wealth of trading options to the trading community, fostering innovations such as reductions in latency and more sophisticated crossing networks. But there is a deeper concern that fragmentation of trading may also be

---

[1] From its launch in 2007, Chi-X has now captured 19% of EU trading volume market share and in June 2010 it was the second largest trading venue in terms of volume. Alternative trading venues have grown rapidly in Canada following the launch of the consortia-owned Alpha trading system on November 7, 2008. As of March 20010, ATSs have captured 33% of the trading volume in Canada.

[2] See, for example, "NYSE Adjusts Charges in Bid to Draw Traders", Wall Street Journal, Feb. 3, 2009, which discusses the NYSE's strategy of lowering trading fee rebates to attract more high frequency traders.

harming the quality of markets by reducing the liquidity available not only in individual markets but in the aggregate market as well. Such a degradation of market quality could occur, for example, if fragmentation reduced the enforcement of time priority across markets, thereby dis-incentivizing traders from posting limit orders. A related concern is that because many of the new trading platforms are proprietary systems, not all traders can access all trading venues. This raises the specter that markets may not be fragmenting so much as they are fracturing into many disparate pieces.

In this research, we investigate how fragmentation is affecting equity market quality. This question has long interested researchers but empirical investigations have been limited by the difficulty of measuring both the extent of fragmentation and the quality of executions in diverse venues. Our analysis draws on new data sources to provide better metrics for addressing these issues. We calculate the extent of fragmentation in individual stocks by using volumes reported by the newly-established Trade Reporting Facilities (TRFs). Whereas before off-exchange volume was simply aggregated with exchange-executed volume for reporting purposes, now exchanges must report only their on-exchange volumes, with off-exchange volumes handled by TRFs.[3] Because all trades must be reported to the consolidated tape, TRF data provides an accurate measure of the trades being executed in non-exchange venues.[4]

[3] TRFs were mandated by the SEC as a condition for approval of Nasdaq's application for exchange status. The SEC required that as of March 5, 2007, all non-exchanges must report to a trade reporting facility, which in turn would report trades to the consolidated tape.

[4] TRF data does not disaggregate trades into specific execution venues so we cannot determine the specific volume of trading in each of the many non-exchange venues. We can determine the aggregate off-exchange volume per stock, however, giving us comparable, and much improved, metrics for fragmentation. An alternative fragmentation metric is the volume of trade executed away from the listing exchange. Results using the two fragmentation metrics are similar, but for brevity we report only the TRF results.

To address market quality issues, we use SEC Rule 605 data, which is a set of execution metrics reported monthly on a per stock basis by all execution venues.[5] This data was generously provided to us by TAG/Audit, and it allows us to compare execution quality as measured by effective spreads, realized spreads and execution speeds across stocks with more fragmented or more consolidated trading. We also use more standard TAQ microstructure data to investigate quality issues related to price efficiency. Our analysis here examines short-term return volatility and variance ratio tests.

Determining the effects of fragmentation on execution quality is complicated by endogeneity issues. As previously demonstrated (see SEC (2001); Boehmer (2005)), different stocks may have different costs of trading for reasons unrelated to fragmentation. For example, small stocks generally have higher trading costs. If small stock trading is also more likely to fragment, then finding higher trading costs for fragmented stocks may be spurious due to the failure to control for firm size. Additionally, market-related issues (see Bessembinder (2003); Boehmer, Jennings and Wei (2007)) may lead to fragmentation for reasons unrelated to the trading costs of stocks. If particular venues only trade specific stocks, a finding of lower trading costs for fragmented stocks may be spurious due to a failure to control for this selection bias.

Previous research has addressed these endogeneity concerns in a variety of ways, including matched samples, regression analysis and the Heckman correction. We use each of these approaches in our research. We use the Heckman correction to test for

[5] Rule 605 data arises from an SEC requirement that all market centers publicly disclose on a monthly basis execution quality statistics. Not all trade executions must be included, but data must be provided for orders meeting the following criteria: orders must be held; limit price must be less than 10 cents from the quote; order must be straight market or limit order; and the order must be for 10,000 shares or less. Bennett and Wei (2006) also use what was then known as SEC 11Ac1-5 data to address market quality in their study of firms moving from the Nasdaq to the NYSE, as do Goldstein et al (2008) in their interesting study of competition for Nasdaq securities.

selection bias in how stocks fragment across markets, and we use a matched sample approach to compare the execution quality of stocks with more fragmented trading to that of stocks with more consolidated trading. We also use regression analysis to investigate more fully how spreads are affected by fragmentation and other economic variables.

Our analysis yields a number of results. We provide compelling new evidence on the extent and nature of fragmentation in U.S. equity markets. We find that off-exchange venues are executing almost 30% of all equity volume. While fragmentation levels vary widely across stocks, all firms now exhibit fragmented trading, and major markets and TRFS now trade virtually all stocks. These results are in stark contrast with earlier findings that only sub-sets of stocks fragmented and that markets were selective regarding the stocks they chose to trade. Results from the Heckman correction confirm that selection bias is not a factor in explaining the relation of fragmentation and market quality.

Turning to the main focus of our paper, we find fragmented stocks generally have lower transaction costs and faster execution speed. The specific effects of this fragmentation differ across firm sizes, and it differs as well for NYSE-listed and Nasdaq-listed firms. For large firms, fragmentation is associated with faster execution time. For small firms, effective spreads are lower, but there are no significant effects on speed. For NYSE-listed stocks, large, liquid stocks appear to gain the most from fragmentation, whereas for Nasdaq-listed stocks, it is small, illiquid stocks benefiting from fragmentation. Fragmented stocks (particularly on the NYSE) do have higher short-term return volatility, but prices appear to be more efficient in the sense that they are closer to being a random walk. These efficiency effects also exhibit differences with respect to firm size and listing venues. Regression analysis provides confirming evidence that market quality, as measured by effective spreads, is not harmed by market fragmentation.

An immediate application of our results is to the on-going policy debate regarding the desirability of allowing fragmentation to occur in markets. In the United States, fragmentation was an expected outgrowth of Reg NMS, particularly because of the changes required by Rule 611 (the "trade through" rule). Our research provides a first analysis of how market quality as measured by transactions costs and efficiency measures has fared in this new market structure. [6] In Europe and in Canada, fragmentation is more nascent, and our results may be helpful for regulators struggling to decide whether to encourage or discourage more off-exchange trading. In many emerging markets, off-exchange trading is prohibited.[7] Our finding that fragmentation does not appear to have detrimental effects on market quality suggests reconsidering such policies.

We caution, however, that as with prior empirical work, our analysis has limitations. We do not have trade data identified by specific trading locale, limiting our ability to relate how execution quality differences reflect differences in particular trading mechanisms. We also do not observe many factors that could influence routing decisions, such as payment for order flows, the use of indications of interest (IOIs), or smart routers. These data deficiencies limit our ability to address the *ex ante* causes of fragmentation. More recently, concerns have arisen regarding the stability of

[6] Reg NMS, originally proposed in August 2005, entailed a variety of changes to market linkages and structure. Among the most important changes was Rule 611 which essentially imposed a price priority rule across all market centers. By requiring that orders must be sent to the market center with the best price, this rule allowed for greater competition by non-exchange venues. Rule 611 was very contentious, and was only fully implemented for all stocks in October 2007.
[7] China, for example, strictly prohibits all off-exchange trading, as do most Asian markets.

fragmented markets in abnormal market conditions. These conditions do not arise during our sample period, so an analysis of stability issues is beyond the purview of our research.[8] Our analysis is thus best viewed as providing empirical evidence on the *ex post* relation between fragmentation and market quality in normal market settings.

This chapter is organized as follows. The next section sets out theoretical arguments surrounding market consolidation and fragmentation, endogeneity issues and our empirical testing approach. Section 1.2 sets out the data and sample period, and discusses the roles played by trade reporting rules and the newly-established trade reporting facilities. Section 1.3 presents results on the current state of fragmentation, both in the aggregate and conditional on firm and market characteristics. Section 1.4 presents empirical results from the Heckman correction, matched sample investigation, and regression analysis of how fragmentation affects various metrics of market quality. Section 1.5 is a short conclusion.

1.1 Fragmentation versus Consolidation

1.1.1 Theory and Empirical Evidence

Whether trading is best consolidated into a single setting or dispersed across multiple venues has long interested researchers. The arguments underlying this debate generally rely on features of the trading process (specifically, the fixed cost structure of markets and network externalities) on the one hand, and the role of competition on the other. Traditionally, setting up exchanges was extremely costly. Trading involved not only expenses related to the trading platform, but also to ancillary services such as

[8] The causes of aberrant market behavior on May 6, 2010, generally known as the flash crash, remain undetermined.

monitoring and listing functions, and costs of clearing and settlement.[9] With much of this cost fixed, it followed that the larger the scale, the smaller could be the trading cost per share traded, and so the greater the gains from consolidation. Network externalities convey a similar benefit in that the ability to match buyers and sellers is greater the more buyers and sellers there are in a market, and so trading costs also benefit from greater scale. Thus, the notion that "liquidity begets liquidity" favors consolidation, even leading some to view exchanges as natural monopolies. Of course, the downside of a monopoly is that it behaves non-competitively, so one argument for fragmentation is that the increased competition it engenders reduces trading costs.

Much of the early theoretical work looking at fragmentation and consolidation argued in favor of consolidation. Mendleson (1987) was perhaps the first to advance the network argument, while Pagano (1989) argued that equilibrium with trading in two markets was inherently unstable as orders would naturally gravitate to the market with greater liquidity. Chowdry and Nanda (1991) advanced another case for consolidation by arguing that adverse selection costs increase with the number of markets trading the asset. Madhavan (1995) argued that consolidated markets would not fragment if trade disclosure rules were mandatory across markets, but would do so otherwise. In his model with non-disclosure, dealers benefit from fragmentation by being less competitive, and informed traders and large traders also benefit by being able to hide trades. Madhavan stated that "fragmentation increases price volatility and induces other distortions as well."[10]

---

[9] See, for example, Macey and O'Hara (1999) for a discussion of issues relating to exchange and trading system functions.
[10] See Madhavan (1995) pg. 581.

More recent research focused on whether competitive effects might shift the arguments in favor of fragmented markets. Economides (1996) argued that welfare losses connected with monopoly providers are not offset by network externalities, suggesting welfare improvement can obtain under fragmentation. Harris (1993) noted that markets fragment in part because traders differ in the types of trading problems that they confront. Hendershott and Mendelson (2000) demonstrated that fragmentation can reduce inventory risk of individual dealers. Bias (1993) proposed conditions under which fragmentation would have no effect on market quality where quality is measured by the mean of spreads.

Empirically, Battalio (1997) found that spreads narrowed on the NYSE after a third-market broker (Madoff Securities) initiated trading. Boehmer and Boehmer (2003) found a similar positive effect on liquidity when the NYSE began trading ETFs listed on the American Stock Exchange. Fong, Madhavan, and Swan (2001) found positive effects on trading costs for large Australian stocks executed off-exchange. Foucault and Menkveld (2008) looked at competition for Dutch stocks between EuroSETS, the London Stock Exchange trading platform, and NSC, the trading platform of Euronext Amsterdam. They concluded that liquidity as measured by depth increased when trading expanded, supporting the notion that fragmentation may be the better outcome.

Yet, other empirical work reaches a different conclusion. Bennett and Wei (2006) examine stocks voluntarily moving from the more fragmented Nasdaq market to the more consolidated NYSE, and find overall execution costs fell when the stocks began trading on NYSE. A study by the SEC (2001) also found lower effective spreads on NYSE than on Nasdaq for a matched sample of stocks, although other execution quality measures were mixed. Gajewski and Gresse (2007) examine trading in Europe, finding that trading costs are lower in a centralized order book than when orders are split between an order book and competing dealers.[11] Amihud, Lauterbach and Mendelson (2003) provide evidence from warrant exercise on the Tel Aviv Stock Exchange that consolidation is more beneficial. Overall, the empirical evidence to date is mixed as to whether market quality is higher in a fragmented or consolidated market.

1.1.2. Testing for Fragmentation Effects

An immediate challenge to testing for fragmentation effects on market quality are the endogeneity issues noted previously. Endogeniety problems can arise if firm, market, and order characteristics influence market quality measures for reasons unrelated to fragmentation. Regression analysis provides one way to control for such differences and we use variables suggested by Bessembinder (2003), Madhavan (2000), and Stoll (2000) to investigate these effects. Another approach to deal with this problem (see SEC (2001); Boehmer (2003)) is to construct a matched sample of firms differing only with respect to fragmentation levels. In Section 1.4 we discuss in more detail our matched sample analysis.

Potentially more challenging endogeneity problems arise if markets selectively choose which stocks to trade. This was clearly an issue in earlier studies of fragmentation. Bessembinder (2003) found that of the 500 NYSE listed stocks in his sample, other markets centers only traded between 77 and 163 stocks. Boehmer, Jennings and Wei (2007) had 1435 stocks in their sample, but only 258 traded continuously in market centers other than the listing market.

[11] Domowitz et al (2008) add a new dimension to this debate by looking at execution statistics for orders left in a single dark pool as opposed to sent sequentially to many dark pools. They find that resting orders in a single venue enhances execution quality, consistent with an inter-temporal consolidation story.

A standard approach to control for selection bias is the Heckman correction. This approach first uses a probit model to capture the variables affecting the proportion of trade in TRFs (our fragmentation measure). The second stage then regresses effective spreads (a market quality measure) on variables affecting market quality including the fragmentation level (more precisely, the Inverse Mills ratio calculated in the first step). We develop this analysis in more detail in Section 1.4.

Finally, endogeneity problems can also arise at the order level. As markets fragment, orders go to new locales and leave old ones. We can see the execution metrics in both venues, but we cannot know whether the types of orders moving to the new venue are the same as orders remaining on the old venue. We control for this potential bias by comparing execution metrics only for specific order types.

1.2. Measurement Issues, Data, and Sample Selection

1.2.1. Measuring Market Fragmentation

Market fragmentation refers to the extent trades execute in different locales. Traditionally, U.S. listed securities traded only on stock exchanges, or since 1971 on Nasdaq, but this has changed dramatically. New technologies gave rise to trading venues such as electronic communication networks (ECNs) and alternative trading systems (ATS's), and regulatory changes removed barriers that generally favored exchange locales. Of particular importance was the passage of Regulation National Market System (or Reg NMS) in 2005 which changed order routing priorities and imposed caps on access charges that exchanges and other venues could impose.

The result has been an explosion of trading venues, with more than 40 trading platforms available to traders in 2008. Table 1.1 lists these trading venues, which include seven

U.S. registered stock exchanges, 5 ECNs, 20 or more ATS platforms, as wells as a variety of new entrants to equities trading such as the Chicago Board of Options Exchange, the International Securities Exchange (an electronic options market), and the Chicago Mercantile Exchange (a futures market). Add to this the internalization of orders by the more than 100 broker/dealer firms, and the number of venues executing trades becomes larger still.

Table 1.1: Trading Venues for U.S. Equities

This table gives trading venues executing equity trades during the period January-June 2008. ECNS refers to Electronic Communication Networks and ATS refers to Alternative Trading Systems.

EXCHANGES	ECNS	ATS	
NASDAQ	BATS	ITG POSIT	CITIMATCH
NEW YORK STOCK EXCHANGE	DIRECTEDGE	BIDS	CS CROSSFINDER
ARCHIPELAGO	TRADEBOOK	LEVEL	LX
NATIONAL STOCK EXCHANGE	LAVA	LIQUIDNET	MLXN
AMERICAN STOCK EXCHANGE	TRACK	MATCHPOINT	SIGMA X
CHICAGO STOCK EXCHANGE		INSTINET	MORGAN STANLEY POOL
PHILADELPHIA STOCK EXCHANGE		MILLENNIUM	UBS PIN
BOSTON STOCK EXCHANGE		PIPELINE	BNY CONVERGEX
INTERNATIONAL STOCK EXCHANGE		PULSE	FIDELITY CROSS STREAM
CHICAGO BOARD OPTIONS EXCHANGE		ESPEED AQUA	LAVA ATS

Ideally, one would measure fragmentation by simply collecting data on trade executions by venue on a per-stock basis. Unfortunately, such data is not available. To understand why, it is useful to differentiate between execution and reporting venues. In the U.S., all trades of listed equity securities must be reported to the consolidated tape. Until recently, only exchanges could report trades, meaning that any off-exchange venue had to report trades to an exchange, which in turn would report those trades to the tape. Such trades would indicate only the reporting venue's identifier, resulting in the reported trades of Nasdaq, for example, including both trades executed there and trades only reported there. This aggregation limited previous studies of fragmentation as it was not possible to know where trades actually executed. Several studies, including SEC (2001) and Bennett and Wei (2006), simply assumed that Nasdaq was more fragmented than NYSE, and analyzed differences between market executions using venue as a proxy for fragmentation.

In addition to complicating matters for researchers, reporting protocols raised important competitive issues. As exchanges and markets converted to for-profit status, exchange volumes became a competitive metric, with venues vying for listing business based on their claims of market size. The SEC, responding to concerns of bias in these numbers, required that trades only reported on venues be separated from trades actually executed there. Such segregation would be accomplished by the establishment of Trade Reporting Facilities that would report directly to the consolidated tape. As of March 5, 2007, all non-exchange executed trades must report to a TRF.

In our analysis, we use TRF volumes to measure fragmentation on a stock-by-stock basis. Because exchange-reported volume now includes only trades executed on that exchange, TRF data provide an accurate measure of each stock's volume executing in off-exchange venues. These data are not perfect, however, in that we cannot determine specific volumes for non-exchange execution venues (by individual ECN or ATS, for example). [12] Consequently, our TRF number is not a homogenous measure, reflecting as it does fragmentation into what are often very diverse trading platforms.[13]

1.2.2. Measuring Market Quality

Market quality refers to a market's ability to meet its dual goals of liquidity and price discovery. In general, markets with lower transactions costs are viewed as higher quality, as are markets in which prices exhibit greater efficiency. While these concepts are straightforward in theory, actually measuring such effects is problematic. Transactions costs can be measured in a variety of ways, and different traders place different value on different execution features. Market efficiency is even more difficult to measure, with a variety of proxies used in the literature to capture this concept.

We use three measures to capture the transactions cost aspect of market quality: effective spread, realized spread, and execution speed. As discussed later in the paper, Rule 605 data is based on orders, not simply on trade executions.[14] Thus, the effective spread is given by twice the difference of the trade price minus the midpoint of the

---

[12] Due to concerns about the size and significance of off-exchange trading venues such as dark pools, the SEC has proposed adopting a uniform method for reporting equity trading volumes by venue. Such a reporting protocol would provide greater transparency into where volume is actually executing. As of June 2010, however, this proposal has not been adopted, although some venues have begun voluntary reporting.

[13] While all reporting exchanges have established Trade Reporting Facilities, over our sample period only NYSE TRF, Nasdaq TRF, and National Stock Exchange (NSX) TRF were active. In addition, the Alternative Trade Facility (ADF) also operated as a TRF. The ADF was originally created by NASD in response to Nasdaq's conversion to for-profit status. The ADF includes both a reporting and display facility, allowing trading platforms who do not wish to post quotes on Nasdaq an alternative venue in which to display quote and trade information.

[13] Boehmer (2005) provides an excellent discussion of the properties and potential problems with Rule 11Ac1-5 data, which is now known as Rule 605 data.

[14] Boehmer (2005) provides an excellent discussion of the properties and potential problems with Rule 11Ac1-5 data, which is now known as Rule 605 data.

consolidated best bid or offer at the time of order receipt. Effective spread is a standard measure in microstructure, and it captures the overall cost of executing the trade from the point-of-view of a trader submitting a marketable order.[15] Realized spread is twice the difference between the execution price and the midpoint of the consolidated quote five minutes after the trade. Realized spread is sometimes viewed as a proxy for the profits available to market makers in making the trade. Execution speed measures the time from order receipt until execution. For some traders, speed is more important than spread. In general, faster markets are viewed as higher quality.

We measure price efficiency using two standard proxies from the literature: short term volatility and variance ratios. Short-term volatility is the return volatility measured over a 15-minute interval. The SEC views excessive short-term volatility as a negative metric of market quality in that some groups of traders may be disadvantaged by short-term price movements unrelated to long term fundamentals. [16] The variance ratio (see Lo and MacKinlay (1988)) captures the notion that, in an efficient market, prices should approximate a random walk. The variance ratio is defined as the absolute value of the ratio of the variance of 30 minute log returns divided by 2 times the variance of 15 minute log returns minus one. The closer this number is to zero, the more prices behave like a random walk, and so the more efficient is the market.

---

[15] A trader can also submit a non-marketable order, which is typically a limit order to trade at some price not currently at the market. For such a trader, transaction costs would also have to include some measure of non-execution risk.

[16] SEC Concept release No. 34-61358 notes: "short term price volatility may harm individual investors if they are persistently unable to react to changing prices as fast as high frequency traders. As the Commission previously has noted, long term investors may not be in a position to access and take advantage of short term price movements. Excessive short term volatility may indicate that long-term investors, even when they initially pay a narrow spread, are being harmed by short-term price movements that could be many times the amount of the spread. "

Hasbrouck (1993) suggests using a variance decomposition approach to measure price efficiency in markets. This approach uses signed order flow to separate the noise variance component of price movements from the information-based variance component. We do not use this approach because the aggregation of volumes across various market venues means that TRF trades are not homogenous. An added complication is the difficulty of assigning trade direction, an increasingly important problem as more and more trades take place within quoted spreads.

1.2.3. Data and Sample Selection

The time period for our analysis is January 2 – June 30, 2008. The data are drawn from TAQ, CRSP, and SEC Rule 605 data provided to us by TAG Audit. Trading volume and price information are taken from TAQ data. We also use TAQ data to calculate short-term return volatility and variance ratios. We use CRSP data to provide information on market capitalization and price. We use Rule 605 data to provide execution quality measures relating to transactions costs. SEC Rule 605 requires all stock exchanges, dealers, and other market centers executing trades to provide specific data on selected order executions. These data must be provided monthly on a stock by stock basis.[17] The data do not include all executed trades and are limited to specific order types.

[17] These include all orders meeting the following criteria: orders must be held; the limit price must be less than 10 cents from the quote; order must be straight market or limit order; and the order must be for less than 10,000 shares.

Table 1.2: Sample Selection Criteria

The sample is selected from all listed securities in January 2, 2008. We remove all securities that are not included in CRSP at December 31, 2007. Those include warrants, preferred, and units bundled with warrants. We apply CRSP filters to remove non-common stock equities, common stocks of non-U.S. companies, close-end funds, Real Estate Investment Trusts, and Americus Trust components and dual class stock. Volume and quote filters are applied to eliminate infrequently traded stocks and low price stocks.

Criterion	**NASDAQ**	**NYSE**
CRSP Filter (December 31, 2007)		
All securities in Jan 2, 2008	3134	3251
No data in CRSP on December 31, 2007	-104	-762
Non-common stock equities (ADRs, units, certificates and Shares of Beneficial Interest)	-159	-564
Common stocks of non-U.S. companies, close-end funds, Real Estate Investment Trusts and Americus Trust Components, ETFs	- 211	-551
Dural class stock	-123	-145
	2537	1229
Volume and Quote Filter (January 2, 2008-March 31, 2008)		
Missing volume, any day	-507	-17
Price<5	-442	-46
Mean daily volume<1000	0	0
Final Sample	1588	1166

We use data based on marketable limit orders for 9999 shares or less. This data captures the largest category of transactions and seems most representative of general market quality, but it does mean that our analysis does not capture all trading in a stock. Market centers report data separately, so the data must be aggregated to provide an average execution metric for each stock. We used data provided by TAG/Audit to form a volume-weighted average execution measure for each stock. The data exhibit substantial outliers, so following standard practice we winsorize the data to set outliers to the 2.5 and 97.5 percentile levels.

Table 1.2 gives information on our sample selection criteria. We begin with all listed stocks on NYSE and Nasdaq. We follow Boehmer (2005) and apply standard filters to remove non-common equities, dual class shares, REITS, and common stocks of non-US companies. We also exclude stocks with prices below $5.00, with mean daily volume below 1000 shares, and stocks not in the CRSP data base. Our final sample is 2754 stocks, with 1588 firms being Nasdaq-listed and 1166 firms listed on the NYSE. We refer to this as the universe sample.

We use a smaller sub-sample of stocks in testing for market quality differences which we refer to as the select sample. We form this smaller sample by selecting from our universe sample every tenth stock listed on NYSE (112 stocks) and every tenth stock listed on Nasdaq (150 stocks). In the matched-pairs analysis, discussed later, we augment these 262 stocks with an additional 262 stocks chosen to match the selected stocks on attributes of price and market capitalization.

1.3 Market Fragmentation

How fragmented is trading in U.S. equity markets? We address this basic question by first looking at trading volumes across the various executing and reporting venues for the period January – March 2008. During this interval there were 9 exchanges, 3 TRFs, and the ADF reporting trades. Table 1.3 provides data on trading volumes reported by each venue. As is apparent, Nasdaq had the largest volume, followed by New York Stock Exchange. Archipelago, the fourth largest venue, is part of NYSE group, but it is treated as a separate location for regulatory reporting purposes (combining ARCA and NYSE volume results in larger overall volume than on Nasdaq). The data also show that regional exchanges (i.e. National Stock Exchange, American Stock Exchange, Chicago Stock Exchange, and Philadelphia Stock Exchange) execute a very small fraction of trades in the market.[18] Similarly, new non-equity exchange entrants (the Chicago Board of Options Exchange and the International Stock Exchange) did not establish any significant market presence during this time period.

Table 1.3: Consolidated volume by reporting venue

The consolidated volumes of all securities listed in NYSE, NASDAQ, American Stock Exchange (now known as NYSE Alternext U.S.) and NYSE ARCA. Sample period is from January 2, 2008 to March 31, 2008.

Trading Venue	Volume in Millions of Shares	Share of Total Volume in percent
Consolidated Volume	495548	100
NASDAQ	153743	31.025
NYSE	105418	21.273
NASDAQ TRF	88302	17.819
ARCA	82305	16.609
NYSE TRF	31643	6.385
National Stock Exchange TRF	12207	2.463
National Stock Exchange	7701	1.554
International Stock Exchange	5259	1.061
American Stock Exchange	2872	0.58
ADF	2684	0.542
Chicago Stock Exchange	2260	0.456
Chicago Board Options Exchange	717	0.145
Philadelphia Stock Exchange	439	0.089
Boston Stock Exchange	0	0
American Stock Exchange TRF	0	0
Boston Stock Exchange TRF	0	0
International Stock Exchange TRF	0	0
Chicago Stock Exchange TRF	0	0
ARCA TRF	0	0
Chicago Board Options Exchange TRF	0	0
Philadelphia Stock Exchange TRF	0	0

This is not the case for Trade Reporting Facilities, which rank 3rd, 5th, and 6th in overall trade volume, reporting in aggregate approximately 27% of trading volume. The

[18] The Boston Stock Exchange, which was acquired by the Nasdaq, was not active during this time period. Similarly, while most exchanges had set up TRFs, most of these were not active during our sample period. For an interesting discussion of the evolution of regional exchanges see Arnold et al (1999).

overall role of TRFs can be better seen in Figure 1.1, which depicts the share of trading volume for all Nasdaq-listed equities, AMEX-listed equities and NYSE-listed equities. For Nasdaq-listed equities, more than one-third of trading volume is taking place in TRFs. For NYSE and AMEX-listed securities, TRFs play a smaller role, but still report almost 25% of volume in those stocks.[19] By any metric, TRFs report a substantial fraction of total U.S. equity volume. As these trades are actually executing in myriad off-exchange venues, fragmentation is clearly an important feature of US equity markets.

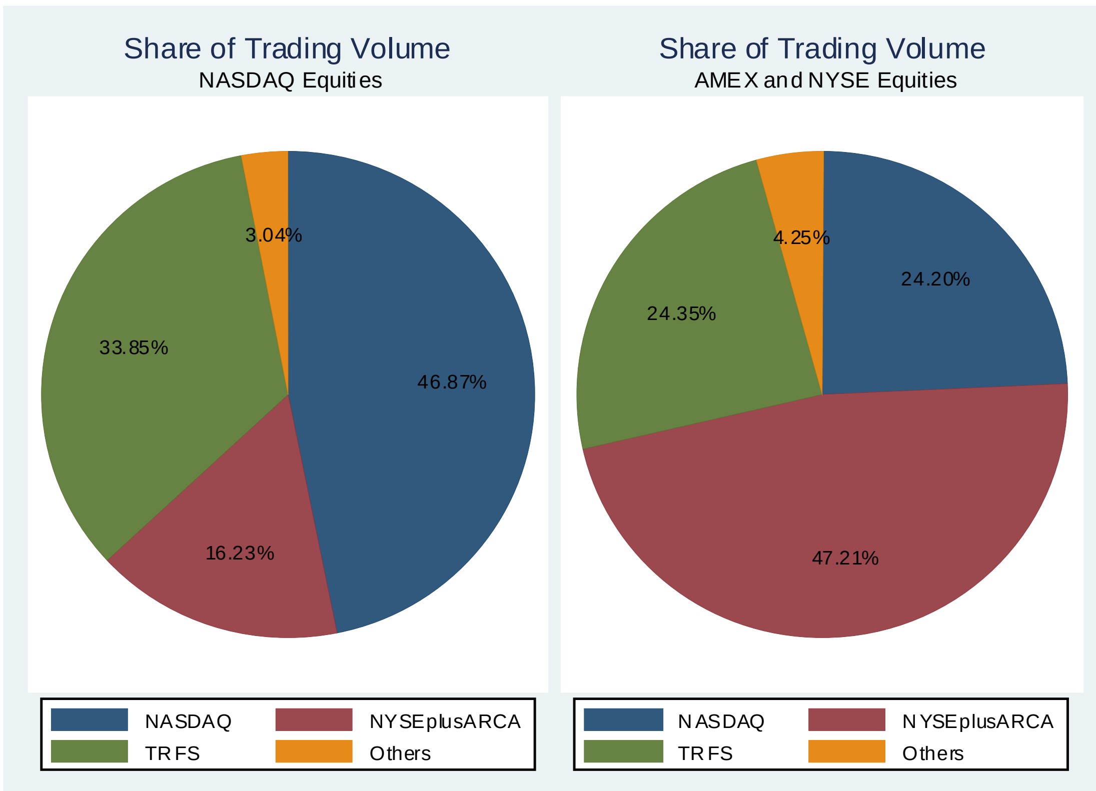

Figure 1.1: This figure gives the percentage share of trading volume for all NASDAQ, AMEX (now know as NYSE Alternext U.S.) and NYSE-listed equities. The sample period is from January 2, 2008 to March 31, 2008. The NASDAQ sample has 3348 equities and the NYSE and AMEX sample has 5414 equities.

[19] This finding that trading in Nasdaq-listed stocks is more fragmented than trading in NYSE-listed stocks is consistent with the intuition of earlier researchers such as SEC (2001) and Bennett and Wei (2006).

How important fragmentation is for individual stocks can be seen from the distribution of volumes across listed securities. As Figure 1.2 (a) shows, individual Nasdaq-listed stock TRF trading ranges from a low of approximately 15% to a high of greater than 75% of volume. For individual NYSE-listed stocks, depicted in Figure 1.2(b), dispersion is smaller, but at the upper range TRFs report almost 40% of volume in some stocks. Equally significant, fragmentation is the reality for all stocks; there are no stocks in our sample with zero TRF volumes.

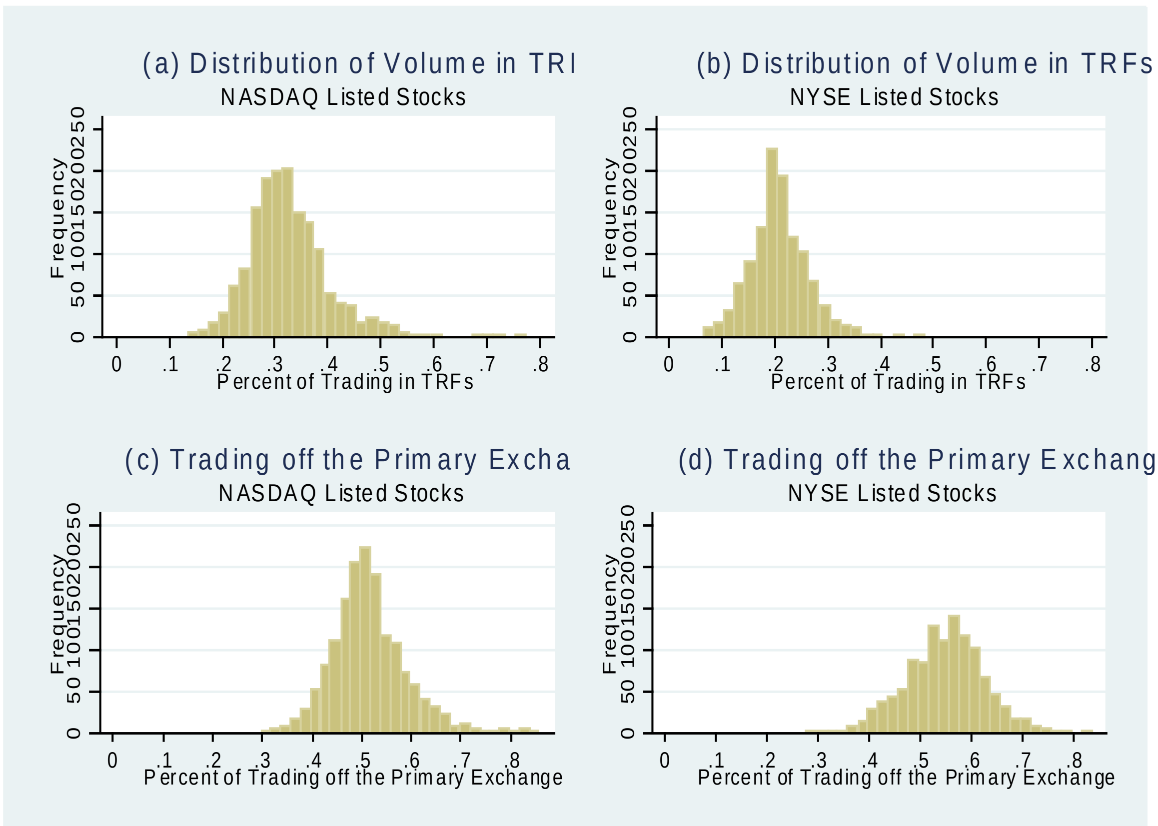

Figure 1.2: Distribution of Volume in Trade Reporting Facilities (TRFs) and Off-Primary Exchange

Figures (a) and (b) demonstrate the distribution of share of volume in TRFs for the 1588 NASDAQ and 1166 NYSE stocks in our filtered sample. The x axis demonstrates the share of volume in TRF, with each bin has a width of 0.02. The y axis counts the number of shares that fall in each bin. Figures (c) and (d) provide the distribution of the share of volume in each stock executing off of the primary listing market. The sample period is from January 2, 2008 to March 31, 2008

Which venues trade particular stocks? Table 1.4 shows the number of stocks with reported trades in each venue. The data show that TRFs trade all 2754 stocks in our universe. This is also the case for Nasdaq. Although the NYSE only trades stocks listed on NYSE, Archipelago trades both NYSE and Nasdaq-listed issues. Thus, analysis at NYSE group level is not subject to a selection bias, nor is it the case for the regional exchanges which collectively also trade all 2754 stocks. The new entrants to equity trading (ISE and CBOE) are not trading every issue, but as noted earlier their market share is negligible. These results illustrate important features of the current competitive landscape for equity trading. The sheer size of TRF volumes testifies to the important competitive challenges that off-exchange trading is posing for established markets. Both NYSE and Nasdaq have been losing market share to TRF venues, and regional exchanges are diminishing in importance as well. For at least some stocks (i.e. those in the right tail of the volume distributions), it appears that TRF trading is now the "market" in terms of trade execution.

But what types of stocks are most likely to trade in TRF venues? We investigate this is more detail in the next section, but we can provide some basic analysis by looking at simple fragmentation patterns by firm size and listing venue. Nasdaq stocks are generally smaller than NYSE-listed firms, so in our universe sample we divide the firms listed on each exchange into large, medium, and small sub-samples based upon firm market capitalization as of January 2, 2008.

Table 1.4: Number of sample stocks traded in each venue
This table gives the number that are traded, or in the case of TRFs reported, in each venue. There are 2754 stocks in our sample. TRF refers to a trade reporting facility, and ADF refers to the Alternative Display Facility. The other trading venues were not active during our sample period.

Venue	Number of Stocks Traded
NASDAQ	2754
National Stock Exchange	2754
Arcapelago	2754
NASDAQ TRF	2754
National Stock Exchange TRF	2754
NYSE TRF	2754
ADF	2751
International Stock Exchange	2674
Chicago Stock Exchange	2502
Chicago Board Options Exchange	1717
New York Stock Exchange	1166
Philadelphia Stock Exchange	690
American Stock Exchange	15

Table 1.5: Fragmentation for large, medium and small NYSE and NASDAQ listed stocks

The total sample has 1166 NYSE-listed stocks and 1588 NASDAQ-listed stocks. Large stocks are the largest one third of stocks in each market, small stocks are the smallest one-third and medium stocks are in-between. The sample period is from January 2, 2008 to March 31, 2008. The asterisks ***, **, and * indicate significance level of one percent, five percent or ten percent.

Panel A – TRF Volumes				
	NYSE Stocks		NASDAQ Stocks	
	Observations	Mean	Observations	Mean
Large	388	0.219	529	0.301
Medium	389	0.205	529	0.314
Small	389	0.204	530	0.368
Panel B – Differences in TRF Volumes				
	Difference	P-Value	Difference	P-Value
Large-Medium	0.014***	0.00	-0.032***	0.00
Large-Small	0.016***	0.00	-0.095***	0.00
Medium-Small	0.002	0.35	-0.063***	0.00

Table 1.5 demonstrates different fragmentation patterns across NYSE-listed and Nasdaq-listed stocks. For Nasdaq stocks, TRF fragmentation is more important, and it affects small stocks more than it does large stocks. Many venues reporting to TRFs are crossing networks or ECNs, and these venues provide traders with opportunities to transact within the spread. Because small stocks tend to have the highest trading costs, these data are consistent with off-exchange locales attracting order flow by providing a more competitive alternative for high trading cost stocks.

For NYSE-listed stocks, TRFs play a smaller but still very significant role. Interestingly, for NYSE-listed stocks, fragmentation is higher for large stocks than it is for small stocks. Large NYSE stocks are the basis for most major stock market indices, and so these stocks are particularly attractive to institutional investors. Crossing networks provide institutions greater ability to trade large orders, while ECNs have typically featured faster execution speeds than the NYSE platform. Greater fragmentation for large NYSE stocks may reflect competition by alternative trading venues for institutional traders.

In summary, we find that U.S equity markets feature substantial fragmentation. There is considerable dispersion in fragmentation across individual stocks and across different listing venues. We now turn to investigating whether there are also differential effects of fragmentation on market quality.

1.4 Fragmentation and Market Quality

If fragmentation affects market quality, then we would expect to find significant differences in market quality metrics between stocks with greater fragmented trading and those with more consolidated trading. In this section, we provide a variety of empirical analyses to investigate this issue. Because these analyses rely on firm-specific order execution data, we analyze the 262 firm select sample composed of every 10th firm listed on Nasdaq and NYSE. This provides a large, random sample of firms to test for market quality differences.

We first use the Heckman correction to investigate whether selection bias across markets affects the relation of fragmentation and market quality as measured by effective spreads. We then use a matched-pairs investigation to control for other firm-specific factors that could affect market quality. In this matched-pairs analysis, we examine a broader range of market quality metrics relating to both transactions costs and market efficiency. Finally, we provide evidence from regression analysis to control for a larger set of factors potentially affecting the relationship of fragmentation and market quality.

1.4.1. Selection Bias and Markets: The Heckman Correction

As noted earlier, a bias can arise if markets or trading venues selectively chose stocks to trade. An econometric specification to control for selection bias is the two-stage estimation procedure commonly referred to as the Heckman correction (see Heckman (1979). This approach has been applied to compare trading costs across various trading venues by Madhavan and Cheng (1997), Bessembinder and Venkataraman (2004), and Conrad, Johnson and Wahal (2003).[20] The first stage of the Heckman correction is to run a Probit model for choice of venue. The Probit estimation then produces a new variable which is included with regressors as controls for selectivity bias in a second

[20] See Bessembinder (2003b) for discussion of this approach and its application to trading cost comparisons.

stage regression of market quality. Because our fragmentation measure is a proportion, we use the Probit model specification for when the dependent variable is continuous between 0 and 1 (see Fleiss, Levin and Paik (2003)). For comparability with previous work, we use effective spreads as the market quality measure.

Suppose that the proportion of trade in TRFs is determined by the following model:

$$TRFpercent_i = \Phi(Z_i\gamma + u_i)$$

where Φ is the standard normal cumulative distribution function (cdf) and Z_i are the economic variables to explain market fragmentation. We estimate the regression:

$$\Phi^{-1}(TRFpercent_i) = Z_i\gamma + u_i$$

Given the estimate $\hat{\gamma}$, we then compute the inverse Mills ratio $\hat{\lambda}_i = \dfrac{\phi(Z_i\hat{\gamma})}{\Phi(Z_i\hat{\gamma})}$, where ϕ is the standard normal pdf and Φ is the standard normal cdf function.

The second stage of the procedure is to run the regression:

$$effective\_spread_i = X_i\beta + \theta\hat{\lambda}_i + \varepsilon_i$$

where X_i are the variables to explain market quality (here captured by effective spread) including the fragmentation level. A simple test of selection bias is given by the t-statistic on $\hat{\lambda}_i$ If the $\hat{\lambda}_i$ is not significant, we can reject the presence of a sample selection problem.

For the choice of X_i and Z_i, we follow Bessembinder (2003b). The explanatory variables Z_i include the logarithm of market cap of the stock on January 2, 2008, the logarithm of average daily trading volume from January, 2 2008 to March 31, 2008, the average order size and the average price impact. [21] We also run the two-stage analysis excluding the price inverse variable. The X_i include the TRF percent, the log of number of trades, the price inverse, the average trade size, and an indicator variable equal to 1 when the listing market is Nasdaq and 0 otherwise.

Table 1.6: Regression Results with Heckman Correction

Panel A presents the estimates from the probit model of the likelihood that an order is executed in TRFs. The dependent variable is the probit transformation of proportion of volume executed in TRFs. logmkt_cap is the log of the market cap in January 2, 2008. logvol is the log of consolidated volume from January 2, 2008 to March 31, 2008. trade_size is the average trade size from January 2, 2008 to March 31, 2008. price_impact is the average price impact from April 1, 2008 to June 30, 2008. Panel B presents the second-stage regression that use inverse Mills ratio obtained from the first stage regression to correct for endogeneity. TRFpercent is the share of consolidated volume executed in TRFs from January 2, 2008 to March 31, 2008. logtradenumber is the total number of trades from January 2, 2008 to March 31, 2008. price_inverse is 1 over the closing price in January 2, 2008. , dummy equals one if the stock is listed on Nasdaq and 0 otherwise. Standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1

Panel A: First-stage Probit Regression		
	(3.1)	(2)
VARIABLES	pro_TRF	pro_TRF
logmkt_cap	-0.0585***	-0.0457**
	-0.0154	-0.0188
logvol	-0.00676	-0.0104
	-0.0144	-0.0148
trade_size	2.417***	2.251***
	-0.231	-0.27
price_inverse		0.578
		-0.485
Constant	-0.656***	-0.724***
	-0.075	-0.0941
Observations	262	262
R-squared	0.413	0.416

[21] Bessembinder (2003) also uses quoted spread in different trading venues as an explanatory variable in his analysis of market competition. We do not include this variable because we have a different focus in our analysis and the quoted spread in our dataset is aggregated quoted spread across all market centers.

Table 1.6: Panel B: Second-stage Heckman Correction		
VARIABLES	effective_spread	effective_spread
mills_ratio	8.622	11.53
	-5.906	-7.488
TRFpercent	-9.263**	-9.262**
	-3.739	-3.735
logtradenumber	-1.852***	-1.906***
	-0.223	-0.244
price_inverse	-44.92***	-39.74***
	-6.903	-9.092
trade_size	33.36***	36.74***
	-9.77	-11.35
dummy	1.842***	1.844***
	-0.56	-0.56
Constant	4.851	0.709
	-8.06	-10.32
Observations	262	262
R-squared	0.435	0.435

The results in Table 1.6 show two important results. First, the lack of significance on the Mills ratio also means that we can reject the hypothesis of a selection bias in the data. Consequently, selection bias at the market level is not the important problem that it was for investigators of earlier fragmentation studies. We caution, however, that our results are at the TRF and market center level. We cannot, and do not, investigate how orders fragment across the individual ATS, ECNs, and broker/dealer desks reporting to the TRF where selection issues may still be present.[22]

[22] A second difficulty is that we do not have individual orders (or even trades) in each stock but rather overall traded volumes. This difference matters because orders are now typically split into pieces and routed to multiple venues. Boehmer, Jennings and Wei (2007) analyze the order routing decision across trading venues. They find that "broker-dealers face competitive pressures to route to low-cost and/or fast execution venues", which is consistent with fragmentation being driven by competitive factors. They note, however, that practices such as payment for order flow, or the use of IOIs may also be explaining order flow, but the unavailability of data precludes analysis of these effects.

Second, we find that after implementing the Heckman correction an increase in TRF trading decreases effective spreads. This is direct evidence that market fragmentation does not appear to harm market quality as captured by effective spreads. The results in Panel B also show that spreads are positively related to trade size and to listing on Nasdaq, and negatively related to the number of trades (a proxy for volume) and to the price level. These latter results are consistent with the findings of previous research.

1.4.2. Matched Pairs Analysis

Another standard approach for investigating market quality differences is a matched pairs analysis. Such an analysis can control for firm-specific factors than can influence market quality measures. Following Davies and Kim (2008), we match firms based on market capitalization, price, and listing exchange. Thus, using our select sample of 150 Nasdaq-listed firms and 112 NYSE-listed firms, we seek a corresponding firm on Nasdaq or NYSE, respectively, that minimizes the matching error given by:

$$D_{ij} = \left| \frac{MCAP_i}{MCAP_j} - 1 \right| + \left| \frac{PRC_i}{PRC_j} - 1 \right|$$

For each pair of stocks, we place the stock with the higher TRF volume into the fragmented group, and the other stock into the consolidated group. By construction, firms in the TRF-fragmented sample have higher TRF volumes, but otherwise are identical to firms in the consolidated sample. We refer to this as the “pairs sample”. We use data from the period January- March 200 to sort the matched pairs into fragmented and consolidated samples, and we use execution data from April –June 2008 to test for statistical differences in the two samples with respect to market quality measures.

1.4.2.1 Execution Quality Results

We first investigate whether fragmentation affects transactions costs which we measure using effective spreads, realized spreads, and execution speeds.[23] Table 1.7 Panel A provides evidence on these trading cost measures across the fragmented and consolidated samples. In the post-Reg NMS world, effective spreads are extremely low, with average spreads in the 3-4 cent range. The data show that effective spreads are lower in the fragmented sample on average by .29 cents, with median spreads lower by .11 cents. These results are statistically significant. As effective spreads measure trading costs from a trader's perspective, this result is consistent with the competitive effects of fragmentation into TRFs being greater than the network externality effects of consolidation. Fragmentation also lowers average execution speed, with significant differences on the order of 7 seconds between the consolidated and fragmented samples. Realized spreads are not significantly different between the two samples.

Panel B reports results segmented by firm size. We divided the 262 pairs of stocks into two groups based on market capitalization. We find that fragmentation tends to benefit large and small stocks, but in different ways. Effective spreads are statistically significantly lower for small stocks but are essentially unchanged for large stocks. Average execution speed falls for large stocks, but it is unaffected for small stocks. These differential effects across firm sizes suggest that different forces may be at work in explaining why trading fragments for different firm types.

[23] The data exhibit substantial outliers, so following standard practice we winsorize the data to set outliers to the 2.5 and 97.5 percentile levels. We report both t-tests based on averages and Wilcoxon signed ranked tests based on medians.

Table 1.7: Execution Quality for Consolidated and Fragmented Samples

Panel A contains the pair-wise difference of execution quality statistics of the112 NYSE pairs and 150 NASDAQ pairs in our sample. Those pairs are matched based on market capitalization and closing price on January 2, 2008. We consider marketable limit order of all sizes executed in all market centers. Effective spread and realized spread are in cents and average speed is in seconds. All variables are calculated using weighted averages based on executed shares across different sizes and market centers in the SEC 605 data. Panel B contains the pair-wise difference of execution quality statistics of large and small stocks based on market cap. Each category has one half of the observation in our 262 pairs of NYSE and NASDAQ stocks. The sample period for execution statistics is from April 2008 to June 2008. The asterisks ***, **, and * indicate significance level of one percent, five percent or ten percent.

Panel A. Overall Pairs Sample

		Consolidate	**Fragment**	**Consolidate -Fragment**	**p-value**
Effective Spread					
T-test	Mean	3.61	3.33	0.29*	0.07
Wilcoxon Signed Rank Test	Median	2.48	2.26	0.11**	0.05
Realized Spread					
T-test	Mean	0.97	1.07	-0.09	0.31
Wlicoxon Signed Rank Test	Median	0.56	0.47	-0.08	0.25
Average Speed					
T-test	Mean	86.58	79.18	7.40*	0.08
Wlicoxon Signed Rank Test	Median	64.11	55.74	3.68*	0.07

Panel B. Large versus Small Stocks

		Large Stocks		**Small Stocks**	
Effective Spread		**Consolidate-Fragment**	**p-value**	**Consolidate-Fragment**	**p-value**
T-test	Mean	0.13	0.33	0.45**	0.05
Wilcoxon Signed Rank Test	Median	0.04	0.36	0.23**	0.03
Realized Spread					
T-test	Mean	0.11	0.34	-0.30	0.11
Wilcoxon Signed Rank Test	Median	0.01	0.43	-0.24	0.13
Average Speed					
T-test	Mean	10.12**	0.03	4.68	0.31
Wilcoxon Signed Rank Test	Median	4.33**	0.03	2.84	0.34

To investigate this further, we examine in Table 1.8 execution costs segmented by firm size for Nasdaq-listed stocks and for NYSE-listed stocks. Segmenting by firm sizes across markets helps us to control for listing standard effects as well for the fact that Nasdaq-listed stocks are smaller in general than NYSE listed stocks. Looking first at Nasdaq results, we find significant differences in both average and median effective spreads for small firms. These differences are consistent with small fragmented firms having lower spreads than their consolidated matched firms. This effect is not statistically significant for large firms. Turning to NYSE results, we find no significant effects on spreads, but average speeds are improved by fragmentation for small firms. Because NYSE firms are larger overall, this result clarifies that execution speeds improvements are accruing not to the largest firms but rather to firms in the lower half of the NYSE size distribution. Overall, our results suggest that fragmentation as measured by TRF volumes generally helps small firms, and does not harm larger firms.

Our implication of these findings is that conflicting results in the literature may be at least partially due to sample selection biases. Bennett and Wei (2006), for example, find that both effective spreads and execution speeds decrease for their sample of firms moving their listing from Nasdaq to NYSE. They attribute these beneficial effects to the consolidation of trading on NYSE relative to Nasdaq, and so conclude that fragmentation is harmful to stocks. But most stocks shifting from Nasdaq to NYSE are the larger stocks in Nasdaq, and as we show here fragmentation has no significant effects on those stocks. A more likely explanation for Bennett and Wei's result are different trading rules or corporate governance requirements between the two venues.[24]

---

[24] One such rule could be the NYSE requirement in place during their sample period that specialists faced restrictions on the size and movement of spreads. Macey, O'Hara, and Pompilio (2009) found that firms delisted from the NYSE had differential effects on trading costs when moving to the Pink Sheets. While the spreads of large firms actually decreased due to the sub-penny pricing allowed on the Pink Sheets, the

Table 1.8 Execution Quality for Large and Small Stocks by Listing Venue

The table contains the pair-wise difference of execution quality statistics of in each market based on market cap. Panel A has 112 pairs of NYSE stocks and Panel has 150 pairs of NASDAQ stocks. The NYSE and NASDAQ samples are divided into large and small stocks based on the market cap on January 2, 2008. We consider marketable limit order of all sizes executed in all market centers. Effective spread and realized spread are in cents and average speed is in seconds. All three variables are calculated using weighted averages based on executed shares across different sizes and market centers in SEC 605 data. The sample period for execution quality is from April 2008 to June 2008. The asterisks ***, **, and * indicate significance level of one percent, five percent or ten percent.

		Large Stocks		Small Stocks	
Panel A: NYSE Stocks					
Effective spread		**Consolidate-Fragment**	**p-value**	**Consolidate-Fragment**	**p-value**
T-test	Mean	0.21	0.11	0.16	0.33
Wilcoxon Signed Rank Test	Median	0.08	0.14	0.08	0.45
Realized Spread					
T-test	Mean	0.12	0.35	-0.02	0.47
Wilcoxon Signed Rank Test	Median	-0.03	0.47	-0.27	0.34
Average Speed					
T-test	Mean	7.14	0.13	7.96	0.23
Wilcoxon Signed Rank Test	Median	1.02	0.30	6.92*	0.08
Panel B: NASDAQ Stocks					
		Large Stocks		**Small Stocks**	
Effective Spread		**Consolidate-Fragment**	**p-value**	**Consolidate-Fragment**	**p-value**
T-test	Mean	-0.05	0.46	0.78**	0.04
Wilcoxon Signed Rank Test	Median	0.01	0.48	0.29**	0.02
Realized Spread					
T-test	Mean	-0.15	0.36	-0.25	0.25
Wilcoxon Signed Rank Test	Median	-0.23	0.28	0.10	0.39
Average Speed					
T-test	Mean	5.77	0.28	8.80	0.25
Wilcoxon Signed Rank Test	Median	5.58	0.24	6.34	0.31

spreads of small and medium-sized firms increased. These authors attribute this worsening to the cross-subsidization of smaller stocks by larger stocks on the NYSE.

1.4.2.2. Market Efficiency Results

Could fragmentation harm other aspects of market quality? To address this issue, we look at differences across the fragmented and consolidated pairs with respect to two standard measures of efficiency, specifically, the short term return volatility and the variance ratio. We divide the trading day into 26 fifteen-minute intervals starting at 9:30 a.m.[25] We calculate return over each interval based on the spread midpoint at the beginning and ending of each interval.[26]

Short-term volatility is defined as the standard deviation of these returns over the three-month period. Greater volatility is viewed as a trading friction, so the lower the volatility the more efficient the market. The variance ratio is the absolute value of one minus the ratio of the variance of 15-minute log returns to one-half of the variance of 30-minute log returns. A ratio of zero is consistent with stocks following a random walk, hence, a smaller number is better in terms of efficiency (see Lo and MacKinlay (1988)).

The results in Table 1.9 reveal interesting divergences in the effects of fragmentation across trading venues. We find weak negative results with respect to volatility: fragmented stocks are more volatile as measured by medians (but not means) for the overall pairs sample. The results for the variance ratio, however, point to the opposite result. The variance ratio is significantly smaller for the fragmented sample, consistent with prices of these stocks behaving more like a random walk.

[25] We also computed the short-term volatility, return autocorrelation and variance ratio for 5 minute intervals and the results are similar.

[26] An interesting problem arises with respect to the treatment of the close-open period. Deleting this period introduces noise into the variance ratio test because the sums of log returns from 3:45 p.m. to 4:00 p.m. and log returns from 9:30 a.m. to 9:45 a.m.(both one period log returns) is not equal to the log return from 3:35 p.m. to 9:45 a.m. (the two period log return). To deal with this heteroscedacticity problem, we included the overnight return, although statistically whether we include the close-to-open interval has a very limited impact.

Table 1.9: Price Efficiency for Consolidated and Fragmented Samples

Panel A contains the pair-wise difference of price efficiency statistics of the combined 112 NYSE pairs and 150 NASDAQ pairs in our sample. Those pairs are matched based on market capitalization and closing price on January 2, 2008. We divide the regular daily trading hour into 26 15-minute intervals and also consider the time between date t close and date t+1's open as an interval. Short term volatility measures the standard deviation of return for the interval. Variance ratio is the absolute value of 1 minus the ratio of variance of one interval log return to one half of the variance of two interval log return. Smaller numbers in both measures mean more efficiency. Panel B contains the pair-wise difference of price efficiency for stocks listed in different markets. The sample period for execution quality is from April 2008 to June 2008. The asterisks ***, **, and * indicate significance level of one percent, five percent or ten percent.

Panel A: Overall Pairs Sample		**Consolidate**	**Fragment**	**Consolidate-Fragment**	**p-value**
Short-term Volatility (in Percent)					
T-test	Mean	0.728	0.749	-0.021	0.11
Wilcoxon Signed Rank Test	Median	0.642	0.716	-0.030**	0.05
Variance Ratio					
T-test	Mean	0.179	0.163	0.017***	0.01
Wilcoxon Signed Rank Test	Median	0.166	0.153	0.014***	0.01

Panel B: Pairs Sample by Listing Venue

		NASDAQ		**NYSE**	
		Consolidate-Fragment	**p-value**	**Consolidate-Fragment**	**p-value**
Short-term Volatility (in Percent)					
T-test	Mean	0.024	0.16	-0.081***	0.00
Wilcoxon Signed Rank Test	Median	0.005	0.25	-0.061***	0.00
Variance Ratio					
T-test	Mean	0.019**	0.02	0.014	0.11
Wilcoxon Signed Rank Test	Median	0.016**	0.02	0.009	0.12

Table 1.10: Price Efficiency for Large and Small Stocks in Each Market

The table contains the pair-wise difference of price efficiency in each market based on market cap. The 112 NYSE pairs and 150 NASDAQ pairs are both divided into large and small stocks based on the market cap on January 2, 2008. Each category has one half of the observations. We divide the regular daily trading hour into 26 15-minute intervals and also consider the time between today's close and tomorrow's open as an interval. Short term volatility measures the standard deviation of return for the interval. Variance ratio is the absolute value of 1 minus the ratio of variance of one interval log return to one half of the variance of two interval log return. Autocorrelation means the absolute value of first order autocorrelation of each interval. Because of our standardization, small numbers in all three measures mean more efficiency. The sample period for execution quality is from April 2008 to June 2008. The asterisks ***, **, and * indicate significance level of one percent, five percent or ten percent.

Panel A: NYSE Stocks

		Large Stocks		Small Stocks	
		Consolidate-Fragment	**p-value**	**Consolidate-Fragment**	**p-value**
Short-term Volatility (in Percent)					
T-test	Mean	-0.052**	0.02	-0.11***	0.00
Wilcoxon Signed Rank Test	Median	-0.037**	0.03	-0.10***	0.00
Variance Ratio					
T-test	Mean	0.008	0.31	0.019*	0.10
Wilcoxon Signed Rank Test	Median	0.016	0.20	0.009	0.17

Panel B: NASDAQ Stocks

		Large Stocks		Small Stocks	
		Consolidate-Fragment	**p-value**	**Consolidate-Fragment**	**P-value**
Short-term Volatility (in Percent)					
T-test	Mean	-0.026	0.19	0.074**	0.02
Wilcoxon Signed Rank Test	Median	-0.031	0.14	0.055**	0.03
Variance Ratio					
T-test	Mean	0.012	0.15	0.026**	0.05
Wilcoxon Signed Rank Test	Median	0.013	0.13	0.032**	0.04

Examining these results by listing-firm market reveals that the positive effects on variance ratios are due to Nasdaq-listed firms; these effects are not significant for the NYSE-listed sample. Similarly, the data show no fragmentation effects on volatility for Nasdaq stocks, but an increase in volatility for NYSE stocks These findings raise the intriguing possibility that fragmentation has enhanced the efficiency of Nasdaq-listed firms, while simply increasing volatility for NYSE-listed firms.

To investigate these effects in more detail, we divide our sample into size groups by listing market. Table 1.10 presents these results. The Nasdaq-listed results clearly indicate that fragmentation is uniformly beneficial for small stock efficiency. Small fragmented stocks have lower volatility than their consolidated counterparts, and they also have lower variance ratios. Large Nasdaq stocks exhibit no statistical differences between fragmented and consolidated firms. Overall, these results suggest that for Nasdaq stocks fragmentation has helped some stocks without harming others.

For the NYSE-listed sample, results are more complex. Consolidated stocks have lower volatility for large and small stocks. However, there is weak statistical evidence from the variance ratio test that prices for small fragmented stocks are closer to being a random walk. Fragmentation thus appears to raise volatility for NYSE-listed stocks but does not appear to harm (and may actually help) other metrics of price efficiency.

1.4.3. Regression analysis

A third empirical approach to investigate fragmentation effects is regression analysis. Regression specifications to study trading costs issues have been used by numerous authors including Bessembinder and Kaufman (1997), Madhavan (2000), Stoll (2000), and Bessembinder (2003). Bessimbinder (2003) argues that regression analysis is particularly appropriate for studying fragmentation issues in that it can control for variations in types of stocks traded in each market or variations in types of orders and market conditions.

We first investigate the relationship between effective spreads and fragmentation. Using the select sample of 262 firms, we ran the following regressions:

$$\begin{aligned} effective\_spread_i &= \alpha + \beta_1 \log trade_i + \beta_2 tradesize_i + \beta_3 price\_inverse_i \\ &+ \beta_4 TRFpercent_i + \beta_5 dummy_i + \varepsilon_i \end{aligned} \tag{1.1}$$

$$\begin{aligned} effective\_spread_i &= \alpha + \beta_1 \log trade_i + \beta_2 tradesize_i + \beta_3 price\_inverse_i \\ &+ \beta_4 sd_i + \beta_5 \log mkt\_cap_i + \beta_6 TRFpercent_i + \beta_7 dummy_i + \varepsilon_i \end{aligned} \tag{1.2}$$

where *logtrade* is the log of the number of trades in stock I from Jan.2 - March 31, 2008, *trade size* is the average trade size for stock i, *price inverse* is 1/price where price is the closing price of stock i on January 2, 2008, *sd* is the standard deviation of the return of stock i from Jan.2 - March 31, 2008, log market cap is the market capitalization of stock i on January 1, 2008, *TRF percent* is the percentage of orders in stock i executing in TRFs, and *dummy* is an indicator variable equal to 1 when the listing market is Nasdaq and 0 otherwise. The variables in these specifications are suggested by Bessembinder (2003), Madhavan (2002) and Stoll (2000).

The results in Table 1.11 show that fragmentation lowers effective spreads. In both specifications, the coefficient on the TRF variable is negative and statistically significant, consistent with our earlier results on the effect of fragmentation on spreads. The regressions also show that spreads are lower for actively traded stocks, higher priced stocks and stocks traded on NYSE, and spreads are higher for stocks with greater volatility volatile, larger trade sizes, and market capitalization.

Table 1.11: Regression Results

This table gives results from regressions where for each stock effective spread is the average effective spread across all market centers from April 1 -June 30, 2008, logtrade is the log of the number of trades from January 2 - March 31, 2008, trade_size is the average trade size from January 2 - March 31, 2008, price_impact is the average price impact equal to the difference between average effective spread and realized spread across all market centers from April 1 - June 30, 2008, price_inverse is 1/closing price in January 2, 2008, sd is the standard deviation of daily stock return from January 2- March 31, 2008 , logmkt cap is the log of market capitalization in January 2, 2008, TRF percent is the percentage of volume reported to the TRFs from January 2- March 31, 2008, dummy equals one if the stock is listed on Nasdaq and 0 otherwise. Standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

VARIABLES	(1) effective_spread	(2) effective_spread
logtrade	-1.592***	-2.014***
	(0.134)	(0.226)
trade_size	20.80***	16.84***
	(4.642)	(4.857)
price_inverse	-50.70***	-46.33***
	(5.669)	(6.646)
sd		40.94***
		(14.60)
logmkt_cap		0.583**
		(0.284)
TRFpercent	-9.772***	-8.710**
	(3.731)	(3.712)
dummy	1.837***	1.678***
	(0.562)	(0.557)
Constant	16.46***	14.54***
	(1.304)	(1.503)
Observations	262	262
R-squared	0.430	0.449

Table 1.12: Regression Using Pairwise Differences (262 Matched Pairs, 524 Observations)

This table shows the result based on matched samples. Δ shows the differences between the Consolidate and Fragmented paired stocks. effective_spread is the average effective spread across all market centers from April 1 -June 30, 2008. logtrade is the log of the number of trades from January 2 - March 31, 2008, trade_size is the average trade size from January 2 - March 31, 2008, price_inverse is 1/closing price in January 2, 2008, price is closing price in January 2, 2008, sd is the standard deviation of daily stock return from January 2- March 31, 2008 , logmktcap is the log of market capitalization in January 2, 2008, Standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

	(3.1)	(2)	(3)	(4)
VARIABLES	Δ effective_spread	Δ effective_spread	Δ effective_spread	Δ effective_spread
Δ logtrade	-1.186***	-1.321***	-1.139***	-1.277***
	(0.177)	(0.191)	(0.176)	(0.190)
Δ trade_size	9.792***	7.254**	10.34***	8.035**
	(3.377)	(3.431)	(3.214)	(3.284)
Δ pinverse	8.284	19.07		
	(58.08)	(57.05)		
Δ price			0.107**	0.0926**
			(0.0420)	(0.0415)
Δ sd		26.66***		25.31***
		(9.743)		(9.669)
Δ logmktcap		6.066**		5.489**
		(2.473)		(2.455)
Constant	0.330	0.337*	0.334*	0.344*
	(0.201)	(0.197)	(0.197)	(0.194)
Observations	262	262	262	262
R-squared	0.181	0.220	0.201	0.234

We also looked at the relationship between fragmentation and effective spreads use pair-differences in our matched-pairs sample. We use similar control variables to those used in regression (1.1) and (1.2), and add additional variables to control for residual matching errors. In the previous literature, the difference of prices enters the regression in two different ways (see Boehmer, (2005); Huang and Stoll (1996)), so we ran the following regressions:

$$\Delta effective\_spread_i = \alpha + \beta_1 \Delta \log trade_i + \beta_2 \Delta trade\_size_i + \beta_3 \Delta price\_inverse_i + \varepsilon_i$$
(1.3)

$$\Delta effective\_spread_i = \alpha + \beta_1 \Delta \log trade_i + \beta_2 \Delta trade\_size_i + \beta_3 \Delta price\_inverse_i + \beta_4 sd_i + \beta_5 \log mkt\_cap_i + \varepsilon_i \quad (1.4)$$

$$\Delta effective\_spread_i = \alpha + \beta_1 \Delta \log trade_i + \beta_2 \Delta trade\_size_i + \beta_3 \Delta price_i + \varepsilon_i \quad (1.5)$$

$$\Delta effective\_spread_i = \alpha + \beta_1 \Delta \log trade_i + \beta_2 \Delta trade\_size_i + \beta_3 \Delta price + \beta_4 sd_i + \beta_5 \log mkt\_cap_i + \varepsilon_i \quad (1.6)$$

The dummy variable does not enter into these four equations because each pair has the same listing market. A positive α in these regressions implies that the consolidated group has higher transaction cost, and Table 1.12 shows that this is case. In general, we find that consolidated stocks' trading cost is about 0.33-0.34 cent higher than the fragmented group, which is similar to the result we find using the matched sample approach.

1.5 Conclusions

Is market fragmentation harming market quality? Our results suggest the answer is generally no. From a transactions cost perspective, fragmentation appears to reduce effective spreads and increase execution speeds. While the magnitude of these effects differs across listing and size regimes, we find that fragmentation is particularly beneficial for small stocks, suggesting that fragmentation has increased competition for traditionally less liquid stocks. Moreover, while short-term volatility appears to have increased particularly for NYSE-listed stocks, overall efficiency seems to be enhanced in that stocks with more fragmented trading exhibit price behavior closer to being a random walk. These results suggest that fragmentation has enhanced the competitive nature of U.S. equity markets without degrading its transactional or informational efficiency.

One might wonder how these ameliorative effects arise given the presumed positive network externality effects that arise from consolidated trading. We believe the answer is that while U.S. equity markets are spatially fragmented, they are, in fact, virtually consolidated. The development of sophisticated order routing combined with the existence of a consolidated tape and the "trade through" rule have resulted in a single virtual market with many points of entry. This allows the positive benefits of greater competition and specialization to prevail without the negative effects that accompany the loss of consolidation.

This result has particular importance for the debates surrounding fragmentation in global markets. In Europe, the development of multi-lateral trading facilities (MTFs) is accelerating the movement of trades away from the established exchanges. However, the lack of a consolidated tape collecting price feeds from all execution venues greatly inhibits the ability to establish market-wide trade-through protection. Without such protection, it is hard to see how a single virtual market can emerge. Similarly, in Canada, fragmentation has begun, but there is not yet regulatory policy regarding access to new venues, nor a trade-through rule to require orders to flow to the most competitive venue. It remains to be seen whether benefits from fragmentation can emerge without such protections.

Our results may have particular importance for developing economies. Emerging economies have traditionally banned off-exchange trading, but the benefits of new trading technologies can be substantial if combined with appropriate regulatory protections. In China, for example, putting in place trade-through protection and unified trade and price reporting protocols could set the stage for substantial improvements in market quality. Conversely, in markets where such protections have not or cannot be implemented, fragmentation is likely to be more detrimental than not, suggesting that off-exchange trading prohibitions may be appropriate.

Finally, the recent “flash crash” has raised concerns that fragmentation may raise stability issues for markets. Our analysis does not include any periods of instability, and as yet this conjecture is unproven. But these concerns underscore the importance of understanding how market structure affects market performance. We believe this is an important issue for future research.

REFERENCES

Amihud, Y., B. Lauterbach, and H. Mendelson, 2003, The Value of Trading Consolidation: Evidence from the Exercise of Warrants, Journal of Financial and Quantitative Analysis, 38 (4), 829-846.

Arnold, T., P. Hersch, J. H. Mulherin, and J. Netter, 1999, Merging Markets, Journal of Finance, 54(3), 1083-1107.

Battalio, R. H., 1997, Third market broker-dealers: Cost Competition or cream skimmers?" Journal of Finance, 52(3.1), 341-352.

Bennett, P. and L. Wei, 2006, Market structure, fragmentation, and market quality, Journal of Financial Markets, 9, 49-78.

Bessembinder, H., 2003, Trade Execution Costs and Market Quality after Decimalization, Journal of Financial and Quantitative Analysis.

Bessembinder, H., 2003b, Selection Biases and Cross-Market Trading Cost Comparison, Working paper, Eccles School of Business, University of Utah.

Bessembinder, H., and H. M. Kaufman, 1997, A cross-exchange comparison of execution costs and information flow for NYSE-listed stocks, Journal of Financial Economics 46, 293-319.

Bessembinder, H. and K. Venkataraman, 2004, Does an Electronic Stock Exchange Need an Upstairs Market?" Journal of Financial Economics.

Biais, B., 1993, Price formation and equilibrium liquidity in fragmented and centralized markets, Journal of Finance 48, 157-185.

Boehmer, E., 2005, Dimensions of execution quality: Recent evidence for U.S. equity markets, Journal of Financial Economics, 533-582.

Boehmer, E., and B. Boehmer, 2003, Trading our neighbor's ETF': Competition or fragmentation?, Journal of Banking and Finance.

Boehmer, E., R. Jennings, and L. Wei, 2007, Public disclosure and private decisions: Equity market execution quality and order routing, Re view of Financial Studies, 20(2), 315-358.

Chowdry, B., and V. Nanda, 1991, Multimarket trading and market liquidity, Review of Financial Studies, 4(3), 483-511.

Conrad, J., K. Johnson, and S. Wahal, 2003, Institutional Trading and Alternative Trading Systems, Journal of Financial Economics.

Davies, R. J. and S. S. Kim, 2008, Using matched samples to test for differences in trade execution costs, Journal of Financial Markets.

Domowitz, I., Finkelshteyn, I., and H. Yegerman, 2009, Cul de sacs and highways: An optical tour of dark pool trading performance, Journal of Trading.

Economides, N., 1996, The economics of networks, International Journal of Industrial Organization, 14, 673-699.

Fleiss,J.L, B. Levin and M.C.Paik, Statistical Methods for Rates and Proportions, 3rd edition, (Wiley:2003)

Fong, K., A. Madhavan, and P. L. Swan, 2001, Why do markets fragment? A panel-data analysis of off-exchange trading, Working Paper, University of Sydney.

Foucault, T. and A. Menkveld, 2008, Competition for order flow and smart order routing systems, Journal of Finance, 63(3.1), 119-158.

Gajewski, J-F, and C. Gresse, 2007, Centralised order books versus hybrid order books: A paired comparison of trading costs, Journal of Banking and Finance, 31(9), 2906-2924.

Goldstein, M., Shkillo, A., Van Ness, B., and R. Van Ness, 2008, Competition in the market for NASDAQ securities, Journal of Financial Markets, 11, 113-143.

Harris, L., 1993, Consolidation, Fragmentation, Segmentation, and Regulation, Financial Markets, Institutions & Instruments, 2 (5), 1-28.

Hasbrouck, J., 1993, Assessing the Quality of a Security Market: A New Approach to Transaction- Cost Measurement, The Review of Financial Studies 6, 191-212.

Heckman, J. 1979, Sample Selection Biases as a Specification Error, Econometrica, 47, 153-162.

Hendershott, T., and H. Mendelson, 2000, Crossing Networks and Dealer Markets: Competition and Performance, Journal of Finance, 55 (5), 2071-2115.

Klock, M., and T. McCormick, 1999, The Impact of Market Maker Competition on Nasdaq Spreads, The Financial Review, 34, 55-75.

Lo, A., and C. MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, Review of Financial Studies, 1(3.1), 41-66.

Macey, J., and M. O'Hara, 1999, The regulation of exchanges and alternative trading systems: A law and economics perspective, Journal of Legal Studies, 28(3.1), 17-54.

Macey, J., O'Hara, M. and D. Pompilio, 2009, Down and out in the stock market: The law and economics of the delisting process, Journal of Law and Economics, 51,683-713.

Madhavan, A., 1995, Consolidation, fragmentation, and the disclosure of trading information, Review of Financial Studies, 8(3), 579-603.

Madhavan, A., 2000, Market microstructure: A survey, Journal of Financial Markets 3, 205-258.

Madhavan, A. and M. Cheng, 1997, In search of liquidity: Block trades in the upstairs and downstairs markets, Review of Financial Studies, 10, 175-203.

Mendelson, H., 1987, Consolidation, fragmentation, and market performance, Journal of Financial and Quantitative Analysis, 22(2), 189-208.

Pagano, M., 1989, Trading volume and asset liquidity, Quarterly Journal of Economics, 104(2), 255-274.

Securities and Exchange Commission, 2001, Report on the Comparison of Order Execution Across Equity Market Structures.

Stoll, H. R., 2000, Friction, Journal of Finance 55, 1479-1514.

CHAPTER 2

A GLIMPSE INTO THE DARK:

PRICE FORMATION, TRANSACTION COST AND MARKET SHARE OF THE CROSSING NETWORK

2.1 Introduction

No trading mechanisms are more controversial than crossing networks, defined by the Securities and Exchange Commission (SEC) as "systems that allow participants to enter unpriced orders to buy and sell securities. Orders are crossed at a specific time at a price derived from another market." (SEC (1998)).[27] According to a report by Tabb group, crossing networks account for 11.5% of average daily volume in the U.S. [28] Recently, this kind of trading platform has received increased public attention, partly because of "the industry's curious choice of the name ‘dark pool’."[29] In June 2009, NYSE Euronext executive vice president Thomas Callahan asked Congress to pressure the SEC to reexamine its regulatory regime for dark pools because they can "harm price discovery and worsen short term volatility."[30] A month later, Nasdaq CEO Robert Greifeld made an even more aggressive claim in a letter to SEC Chairman Mary Shapiro, in which he called for the elimination of dark pools because they are "market structure policies that do not contribute to public price formation and market transparency."[31] On the other hand, proponents of dark pools claim that "undisplayed liquidity adds to execution quality" and "on behalf of all investors, dark liquidity adds to execution."[32] The SEC

---

[27] Examples of crossing networks are Goldman Sachs Sigma X, ITG POSIT, Liquiditynet and Pulse Trading BlockCross. For a list of these crossing networks, refer to Domowitz, Finkelshteyn and Yegerman (2009).

[28] "Study: 'Dark Pools' Account for 4% of European Trades", Wall Street Journal, November 2, 2009.

[29] "Exchanges should unite to end flash orders", by Nasdaq CEO Robert Greifeld, Financial Times, August 6, 2009. Interestingly, even though Robert Greifeld is an opponent of "dark pools", he acknowledges that "dark pool" is a "misnomer and has inevitably gained a negative connotation." In the survey paper by Degryse, Achter and Wuyts (2008) on dark pools, all of the papers cited are actually on crossing networks.

[30] "NYSE Euronext Asks Congress to Press the SEC On Dark Pools", Trader's Magazine, June 10, 2009,

[31] "Dark Pools Fire Back at Call for Ban", Wall Street Journal, July 31, 2009, .

[32] "Dark Pools Fire Back at Call for Ban", Wall Street Journal, July 31, 2009.

now considers dark pools an area that "will have a significant impact on individual investors and the readers of the financial press,"[33] and they are "expected to solicit comments and meet with proponents and opponents of dark pools."[34]

The finance literature provides surprisingly little insight into crossing networks and the following questions are not well addressed. Do crossing networks harm price discovery? Do crossing networks increase stock price volatility? How and to what extent do crossing networks affect liquidity and transaction costs in the public exchange? Should crossing networks provide so-called "fair access" to all traders? What stock properties create a comparative advantage for crossing networks over exchanges or vice versa?

Two obstacles prevent the previous literature from addressing these questions. First, in a model including both an exchange and a crossing network that also allows price discovery, we face two dimensions of uncertainty: price uncertainty and execution uncertainty. Current literature has not found a way to characterize these two uncertainty simultaneously. In classical models of price discovery, execution is guaranteed. Therefore, the volume is equal to the traders' order size, and the only uncertainty of profits comes from price. When not all submitted orders are executed, traders' expected profits are based on the expected volume as opposed to the submitted order size. Therefore, to define a profit function and optimal strategy, we first need to know the conditional expectation of the volume based on the order size, which usually resides within a very complex functional form. When the price is also a random variable, the problem becomes even more complex, because the profit is now a function of two

[33] Speech by SEC Chairman Mary Schapiro, June 18, 2009.
[34] "SEC Plays Keep-Up in High-Tech Race", Wall Street Journal, August, 20 2009.

random variables. Second, if traders with better information can choose where to trade, the uninformed agents must guess informed traders' strategies in both markets, making the learning problem difficult to characterize.

Due to these two obstacles, the previous literature on crossing networks relies on very strong assumptions. For example, Dönges and Heinemann (2006) and Degryse, Achter and Wuyts (2009) eliminate price uncertainty and assume that transaction costs in the exchange are fixed. These two assumptions enable these authors to focus on the complex problem of execution probability. However, the fixed price precludes them from studying price discovery and price volatility. Crossing networks also have no impact on the transaction costs in the exchange, which are assumed to be fixed. In Hendershott and Mendelson (2000), traders with better information cannot choose where and how much they trade. Even under this strong assumption, the Hendershott and Mendelson model can only be analytically solved as the liquidity order flow goes to infinity, that is, when both information asymmetry and execution uncertainty disappear.

This paper contributes to the literature by solving the obstacles encountered when considering two-dimensional uncertainty and two simultaneous markets. An analytical solution is obtained, which in turn provides theoretical predictions consistent with the empirical literature and sheds some light on the current policy debate. The paper also provides a number of predictions to be tested, which are summarized in the conclusion.

This paper extends the frameworks of Grossman and Stiglitz (1980) and Kyle (1985) frameworks to multiple markets: an exchange with guaranteed execution but also with price impact, and a crossing network with no price impact but without guaranteed execution. Two mechanisms are essential for this model. First, the informed trader needs to balance two types of trading costs. In the exchange, his order has guaranteed execution, but each trade shifts the market price in an unfavorable direction. The informed trader's order does not have a direct price impact in the crossing network, but it has a probability impact in that the execution probability decreases as the order size increases. Second, the price impact of the informed trader's trade on the exchange not only affects his profit in the exchange, but it also creates an externality to his profit in the crossing network, where orders are matched at the price set by the exchange. This externality makes the informed trader trade less aggressively in the exchange than the Kyle model predicts.

This reduction in informed trading in the exchange makes the order flow in the exchange less informative than it is in the Kyle model. Therefore, price discovery is reduced, as the order flow reveals less information to the market maker. The relative lack of revealed information most strongly affects stocks with high fundamental value uncertainty, because information on those stocks is more valuable to the informed trader, which creates a higher incentive for him to hide in the crossing network.

However, less informed trading in the exchange decreases the adverse selection problem and increases the liquidity of the exchange as measured by Kyle's λ. This is in contrast to the prediction of previous literature based on cream-skimming. Cream-skimming predicts that the creation of new trading mechanism worsens the liquidity in the primary exchange, because new trading platforms may attract liquidity traders out of the primary exchange while leaving the informed traders in the exchange. Therefore, the adverse selection problem in the primary exchange becomes more serious, and liquidity of the primary exchange is harmed. Empirically, Fong, Madhavan and Swan (2004) do not find that crossing networks increase the adverse selection problem of the primary exchange, and Gresse (2006) finds that crossing networks increase the liquidity of the exchange, which suggests that cream-skimming from the crossing networks must be offset by other mechanisms. My model suggests that the externality of price impact on the crossing network is one such mechanism.

Due to its ability to characterize both price impact and non-execution, this model generates predictions on the relationship between these two transaction costs, which explain several anomalies found in the empirical literature. Ready (2009) finds empirically that stocks with a higher volatility are more likely to be traded in crossing networks, while Dönges and Heinemann (2006) suggest the opposite on theoretical grounds. My predictions differ from those of the Dönges and Heinemann model because their model assumes a fixed transaction cost in the exchange, whereas the transaction costs in both markets are endogenous in this model. My model shows that both price impact and non-execution probability are positively correlated with volatility, but an increase in fundamental value uncertainty creates a comparative advantage for the crossing network, because the informed trader has a higher incentive to hide his trading in the crossing network. Also, I show that crossing networks may have a higher market share for stocks with lower execution probability, which provides an explanation for the empirical anomaly raised by Ready (2009) that crossing networks' volumes are not high in stocks where the likelihood of finding counterparts is expected to be high.

The behavior of the optimizing informed trader leads to a rather surprising prediction on execution probability. I find that an increase in liquidity trading in the crossing network may decrease execution probability, because the resulting increase in informed trading may be greater than the increase in liquidity trading. This prediction is opposite of the predictions of the model with no informed trader (Dönges and Heinemann (2006)) and of the model with exogenous informed traders (Hendershott and Mendelson (2000)). This counterintuitive result is driven by differences in market structure. While the market maker in the exchange can actively adjust quotes to protect himself from the informed trader, a crossing network with a fixed allocation rule is passive. As liquidity trading in the crossing network increases, the informed trader considers the crossing network to be more favorable and moves even more trades to the crossing network than the liquidity traders do.

The paper then shows how the crossing network can change its allocation rules to protect itself from the informed trader. To my knowledge, this is the first paper that studies how market outcomes are affected by different allocation rules in the crossing network. The main discussion in the paper is based on the rule that the informed trader trades first. Then two alternative rules are considered. One is to give the informed trader a lower trading priority, and the other is to exclude the informed trader from the market altogether. Both of these strategies decrease the non-execution probability due to a decrease in the level of adverse selection in the crossing network, but price impact increases as adverse selection in the exchange increases. Price discovery, however, is always enhanced by these two strategies. Interestingly, while the main purpose of the crossing network adopting one of these two strategies is to increase the execution probability, these strategies also minimize the negative impact of crossing networks on price discovery. As a result, the proposed change to enforce fair access in crossing networks will have two undesirable consequences. First, fair access always harms price discovery, because informed traders will hide in the crossing network. Second, fair access will lead to higher adverse selection problems in crossing networks. Liquidity traders on the same side as the informed trader would be crowded out. If a liquidity order is executed, it will be more likely to be an order on the wrong side of the market.

The chapter is organized as follows. Section 2.2 provides institutional details about the crossing networks. Section 2.3 develops the model and solves for the unique linear equilibrium. Section 2.4 analyzes the crossing network's impact on information revelation and price volatility. Section 2.5 analyzes the price impact and the non-execution probability as well as the relationship between them. Section 2.6 considers the competition for order flow between the exchange and the crossing network. Section 2.7 considers the impact of different allocation rules on price discovery and liquidity. Section 2.8 concludes the paper and discusses the directions for future research.

2.2 Institutional Details

In crossing networks, traders anonymously enter unpriced buy and sell orders. The trade is priced by reference to a price derived from some other market. Crossing networks originated in the early 1970s as private phone-based networks among buy-side traders. In the 1980s, crossing networks went electronic with the introduction of Instinet and POSIT. Currently, there are about 40 crossing networks in the U.S. and 60 globally. A partial list of them can be found in Domowitz, Finkelshteyn and Yegerman (2009). As a thorough description of trading procedures of crossing networks would be voluminous, I focus my introduction on three key elements that define crossing networks and distinguish their types.

First, crossing networks all have a benchmark price, which can be bid-ask midpoint, closing price, volume weighted average price, or national best bid and offer price. Here are examples offered by Hasbrouck (2007). For some crossing networks, the price is determined after the quantity match. In ITG's POSIT system, for example, potential buyers and sellers enter quantities to buy or sell, which are not made visible. At the time of the crossing, the system matches buyers and sellers and the execution price is the midpoint of best bid and ask in the listing exchange. To discourage either side from manipulating the price in the listing market to obtain a favorable matching price, the exact time of a cross is random within a time window. For some other crossing networks, price is determined before the quantity match. For example, the Instinet closing cross allows traders to submit orders after the regular market closes. These orders will be matched and executed at closing price. Because price is determined before the quantity match, crossing networks need to be designed to discourage predatory trading.[35] For example, Instinet cancels crosses when there are news announcements and monitors participants, expelling those whose strategies appear to be news driven.

Second, prices of the crossing networks do not have the market-clearing function because they are derived from other markets. If buy and sell orders are not balanced, only the side with fewer orders can be fully executed. Therefore, crossing networks need proprietary matching algorithms to determine the trading priority for the side with the larger quantity. Examples of basic allocation rules include the time priority rule and the pro rata rule; rules in reality may be complex functions of these basic rules and are mostly confidential.[36] As crossing networks are not public exchanges, their customers can be selected and some traders can be excluded. This can be considered as an extreme allocation rule in which some traders always get 0 execution. Crossing networks' preferred customers are "buy-side" firms, particularly those who manage "passive portfolios" such as index funds. Two kinds of traders are often excluded from the

[35] A strategy of predatory trading involve submitting orders in response to news announcements made after the determination of the closing price in the hopes of picking off unwary counterparties.

[36] As the paper will show, allocation rules are the key for crossing networks to minimize the adverse selection problem created by informed traders. Therefore, crossing networks adopt complex allocation rules, keep them confidential, and frequently change them.

crossing network. The first kind is potentially informed traders such as hedge funds, brokers and proprietary traders from sell-side firms; the second kind is traders who submit small orders to extract information contained in the order flow.

Finally, crossing networks differ in their matching frequency. Some only match orders once a day, whereas others may match several times a day or have continuous matching.

There are several advantages to trade in crossing networks. First, there are usually no bid-ask spreads in crossing networks, as buy and sell orders are executed at the same price. Second, trades also do not have price impacts, as their prices are independent of order sizes. Conditional on execution, crossing networks usually have lower transaction costs than does the exchange (Keim and Madhavan (1998), Conrad, Johnson and Wahal (2003), Næs and Ødegaard (2006) and Sofianos and Jeria (2008)). In addition, institutional traders like to use crossing networks because they prevent information leakage. If information associated with an institutional order leaked out, opportunistic front runners could trade in advance of the order in the same direction, thereby driving the price in an unfavorable direction.

The three benefits of trading in crossing networks prompt Conrad, Johnson and Wahal (2003) and Ready (2009) to ask why crossing networks are not more widely used. The answer is that the probability of execution in crossing networks is significantly lower than that in the exchange. Gresse (2006) finds that the execution probability of the crossing network is as low as 2.63% to 4.13%, whereas the order fill rate in the exchange is as high as 90% (Keim and Madhavan (1995) and Perold and Sirri (1993)).[37]

[37] These numbers are for all types of orders. The fill rate for limit orders, especially nonmarketable limit orders, are lower. Hasbrouck and Saar (2009) find that the fill rate for nonmarketable limit order is 6.37%.

If we measure trading costs for both executed orders and nonexecuted orders using the implement shortfall developed by Perold (1988), we can say that crossing networks have lower execution costs but higher opportunity costs.

Non-execution can occur for noninformational reasons. The Hendershott and Mendelson model shows that even if there is no information asymmetry, expected probability of execution cannot be higher than 70% because of random mismatch of geometrically distributed buy and sell order flow. Non-execution can also occur for informational reasons. On one side of the market, there are both liquidity and informed traders, and on the other side there are only liquidity traders. The noninformational and informational causes of non-execution have different implications. In a world without information asymmetry, the expected price change is 0 after each trade. On the other hand, non-execution caused by informational sources has an adverse selection effect. An order on the same side as the informed order may be crowded-out by the informed trader. On the other hand, an executed order is more likely to be on the wrong side of the market. By analyzing an institutional buyer, Næs and Ødegaard (2006) show that stocks that fail to execute in the crossing network have significantly higher cumulative abnormal returns than stocks that successfully execute, an indication of the adverse selection problem in crossing networks.

2.3 Model

2.3.1 Setup of the Model

This model is a variation of the canonical strategic trade model developed by Kyle (1985). I consider a two-period model with two markets: an exchange and a crossing network. A single risky asset is traded by three types of agents: a risk-neutral informed trader, many liquidity traders, and a market maker. The asset has a stochastic liquidation value $\widetilde{v_g}$ with$E(\widetilde{v_g}) = p_0$. The informed trader observes in advance the realization of $\widetilde{v_g}$, denoted v_g, and submits $x_e \in \mathbb{R}$ to the exchange and $x_d \in \mathbb{R}$ to the crossing network to maximize the value of his information.

As in the Kyle model, liquidity traders are passive players and their motives for trade are not explicitly modeled. Dönges and Heinemann (2006), Degryse, Achter and Wuyts (2009), Foster, Gervais and Ramaswamy (2007) and Hendershott and Mendelson (2000) show that heterogeneous liquidity preferences of liquidity traders, which represent their willingness to pay for the immediacy of execution, can lead to non-zero liquidity trading in both markets. These four papers also show that crossing networks generate new liquidity traders who are unwilling to trade in the exchange. In addition to liquidity preference, trade size is also a consideration in the choice of market. Index funds or other institutional traders who manage passive portfolios may make large liquidity trades, which will cause a substantial price impact if they trade in the exchange. Therefore, these traders may opt to trade in the crossing network. Conversely, small traders may prefer the exchange because the price impact of their trade is trivial. It is very hard to incorporate the choice of both informed and liquidity traders when there are different kinds of trading mechanism.[38] The previous literature either assumes that there are no informed traders (Parlour and Seppi (2003)) or that there are only exogenous informed traders (Hendershott and Mendelson (2000)). The assumption of exogenous liquidity traders is certainly closer to the standard assumption in the market microstructure literature.[39]

---

[38] Chowdhry and Nanda (1991) and Baruch, Karolyi and Lemmon (2007) model the choice of both informed and liquidity traders when assets are traded in several markets with the same trading mechanism.

[39] In reality, exchanges and crossing networks coexist, which is a strong indication of liquidity trading in both markets because informed traders cannot trade among themselves. My model reduces to the Kyle model if there are no liquidity traders in the crossing network. No equilibrium exists when all of the liquidity traders are in the crossing network. The reason is as follows. Suppose that the informed trader

The liquidity order flow in the exchange is denoted as $\widetilde{u_e} \in \mathbb{R}$. For the convenience of modeling, the buy order flow and sell order flow are defined separately in the crossing network. Let $\widetilde{u_d^b} \in \mathbb{R}^+$ and $\widetilde{u_d^s} \in \mathbb{R}^+$ be the unsigned aggregate liquidity buy and sell order flows in the crossing network, respectively. I assume that $\widetilde{v_g}$, $\widetilde{u_e}$, $\widetilde{u_d^b}$ and $\widetilde{u_d^s}$ are independently distributed, and their distributions will be specified below.

The timing of events is depicted in Figure 2.1. At time 0, all four random variables are realized. The informed trader observes v_g but does not observe $\widetilde{u_e}$, $\widetilde{u_d^b}$ or $\widetilde{u_d^s}$. His trading strategy $\{X_e, X_d\}$ assigns an order size in the exchange and the crossing network to each v_g. The crossing network only accepts orders before time 1, even though it opens in the second period, because after time 1, the informed trader knows both the realization of $\widetilde{v_g}$ and the price $\tilde{p}$. He could then compare these two values and conduct predatory trading in the crossing network. At time 1, when the exchange opens, the market maker observes the aggregate order flow in the exchange $\tilde{y} = \widetilde{x_e} + \widetilde{u_e}$ but cannot know the individual values of $\widetilde{x_e}$ and $\widetilde{u_e}$. He also does not know $\widetilde{v_g}$, $\widetilde{x_d}$, $\widetilde{u_d^b}$ or $\widetilde{u_d^s}$. The market maker sets the semi-strong efficient price $\tilde{p}$ to clear the imbalance between the buy and sell orders. The market maker's pricing rule is P, which assigns to each outcome of $\tilde{y}$ a price $\tilde{p}$ based on his conjecture of the informed trader's strategy $\{X_e, X_d\}$. At time 2, the crossing network opens. $\tilde{p}$ is used to match the buy and sell orders. The stock liquidates at the end of period 2. As there is no market maker to offer liquidity for the trade imbalance in the crossing network, only the side with less volume gets full execution; the side with more volume, on the other hand, gets partial execution.

---

trades in the exchange; then his information would be fully revealed to the dealer based on Grossman and Stiglitz (1980) and he would earn 0 profit. Therefore, the informed trader would not trade in the exchange. However, if the informed trader does not trade in the exchange, then the exchange shuts down, and there is no price for the crossing network. Alternatively, it can be assumed that the market maker sets a price equal to p_0 when nobody trades. A price of p_0, however, would lead the informed trader to trade in the exchange. Therefore, no equilibrium exists.

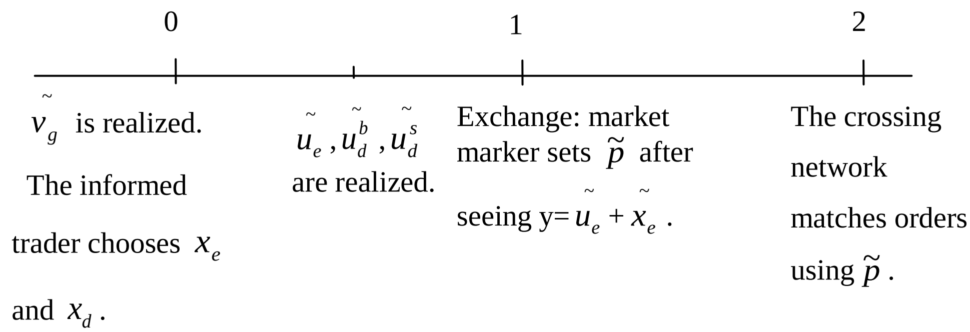

Figure 2.1: Time line of the model

Because $\tilde{p}$ is determined before time 2, the informed order size x_d does not have a direct price impact in the crossing network.[40] This would lead the risk-neutral informed trader to submit an infinite size x_d. To rule out this possibility, I follow Hendershott and Mendelson (2000), Seppi (1997), Parlour and Seppi (2003) and Foucault and Menkveld (2008) and assume that there is an up-front order submission cost of c per share, which applies to both the exchange and the crossing network. Because all orders in the exchange are executed, c can be understood as the commission, whereas the cost in the crossing network will "capture any incremental opportunity or shoe leather costs investors bear when trading from off the exchange." (Seppi (1997)). My model holds for any positive c so the value of c can be set to be arbitrarily small.

Because of the cost c, the informed trader's profit per unit is $(v_g - c) - \tilde{p}$ when he buys and is $\tilde{p} - (v_g + c)$ when he sells. The cost c has an asymmetric effect for the informed trader: it increases the fundamental value to the informed buyer and decreases the fundamental value to the informed seller. Moreover, when $v_g \in [p_0 - c, p_0 + c]$, the

[40] At equilibrium, x_d indirectly impacts the price because the market maker has a correct belief regarding X_d.

potential revenue of trading is lower than the up-front submission cost. Therefore, the informed trader neither buys nor sells. To rule out the possibility of non-trading, the Hendershott and Mendelson model assumes that informed traders always trade. In this paper, I make the following assumption on v_g, whose only purpose is to make all of the proofs in the paper rigorous. Without the assumption, it can be proven that the model holds asymptotically by letting c approach 0.

To define the distribution of $\widetilde{v_g}$, I assume that a new variable $\tilde{v}$, which is the fundamental value of the asset to the informed trader after adjusting for c, follows the normal distribution

$$\tilde{v} \sim N(p_0, {\sigma_v}^2).$$

Then $\widetilde{v_g}$ is defined by the following transformation of $\tilde{v}$:

$$\widetilde{v_g} = \begin{cases} \tilde{v} + c \; when \; \tilde{v} \geq p_0 \\ \tilde{v} - c \; when \; \tilde{v} < p_0 \end{cases} \tag{2.1}$$

Intuitively, this transformation means that the value of the information is normally distributed after deducting the up-front submission cost. In addition, $\Pr(\widetilde{v_g} \in [\,p_0 - c, p_0 + c]) = 0$; thus, the information always leads the informed trader to submit orders. The Kyle model makes assumption about $\tilde{v}$ directly because it does not model commission or other submission cost; by contrast, I need to make assumptions about $\widetilde{v_g}$, because the up-front submission cost is part of the model. The value $\widetilde{v_g}$ will only show up in the intermediate steps of the proofs, whereas the major results of the paper only contain $\tilde{v}$ because c cancels out in the derivation. Finally, $\widetilde{v_g}$ and $\tilde{v}$ are informationally equivalent because they have a one-to-one mapping such that $E(.\,|\widetilde{v_g}) = E(.\,|\tilde{v})$. In addition, it is easy to show that $E(\widetilde{v_g}) = E(\tilde{v}) = p_0$.

As in the Kyle model, I assume that the liquidity order flow in the exchange, $\widetilde{u_e}$, follows a normal distribution with mean 0 and variance σ_e^2, where $\widetilde{u_e} > 0$ represents a net buy order flow and $\widetilde{u_e} < 0$ represents a net sell order flow. The standard deviation σ_e serves as a proxy for the level of liquidity trading in the exchange.[41]

The unsigned liquidity buy, $\widetilde{u_d^b}$, and the unsigned liquidity sell, $\widetilde{u_d^s}$, follow power law distributions. The fact that U.S. trading volume follows a power law distribution has been found by Gopikrishnan, Plerou, Gabaix and Stanley (2000). This result is extended to France and the UK by Gabaix, Gopikrishnan, Plerou and Stanley (2006) and by Plerou and Stanley (2007).

The following equation defines $\widetilde{u_d^s}$. The distribution of $\widetilde{u_d^b}$ can be similarly defined.

The cumulative distribution function (C.D.F.) of $\widetilde{u_d^s}$ is

$$F_{s(}z;k) = P\left(\widetilde{u_d^s} \le z\right) = \begin{cases} 0, & for\ z < 0 \\ 1 - \sqrt{\frac{k}{z+k}}, & for\ z \ge 0 \end{cases} \qquad (2.2)$$

which also implies the following probability distribution function (P.D.F.) of $\widetilde{u_d^s}$

$$f_{s(}z;k) = \begin{cases} 0, & for\ z < 0 \\ \frac{1}{2}k^{\frac{1}{2}}(z+k)^{-\frac{3}{2}} & for\ z \ge 0 \end{cases} \qquad (2.3)$$

The parameter *k*, which is called the scale, is an inherent parameter of the distribution. Figure 2.2 shows that a distribution with a higher *k* stochastically dominates a distribution with a lower k. Therefore, k captures the level of liquidity trading in the

[41] The unsigned order flow, $|\widetilde{u_e}|$, , follows a folded normal distribution with $E|\widetilde{u_e}| = \sqrt{\frac{2}{\pi}}\sigma_e$. Thus, the expected size of uninformed order flow is linear in σ_e.

crossing network. As *k* increases, the level of liquidity trading in the crossing network increases.

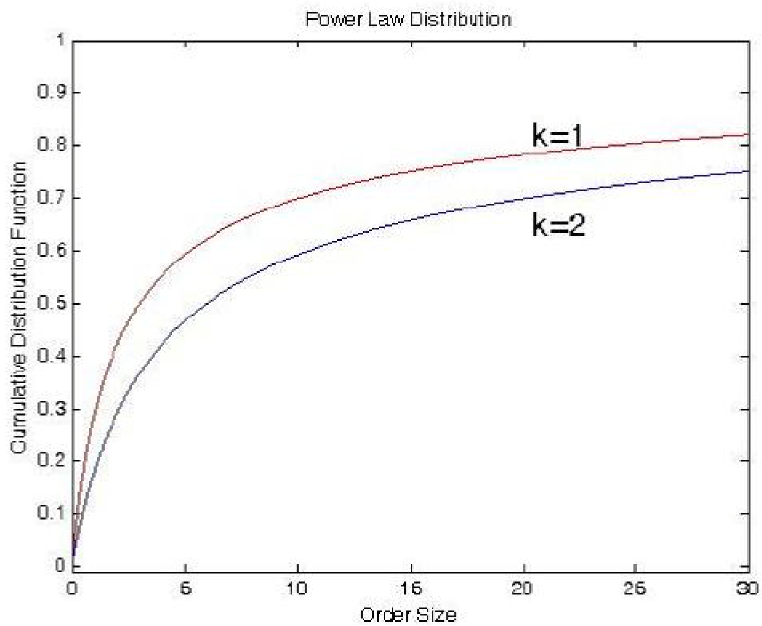

Figure 2.2: The cumulative distribution function of the power law distribution

This figure illustrates the cumulative distribution function (C.D.F.) of the power law distribution for different *k*. The horizontal axis measures the order size and the vertical axis measures the value of the C.D.F. A distribution with larger k stochastically dominates a distribution with smaller *k*.

2.3.2 Allocation Rules

Market microstructure is the study of the process and outcomes of exchanging assets under explicit trading rules (O'Hara (1995)). The key rule affecting the market outcomes in my model is the allocation rule of the crossing network. First, let us consider a rule that defines a scenario in which the crossing network can successfully exclude the informed trader. Then, the exchange in my model reduces to that in the Kyle model. Each trade reveals half of the information, and the information revelation is independent of the fundamental value uncertainty σ_v and the level of noise trading in the exchange σ_e. Therefore, price discovery is not harmed. A crossing network operating under the conditions defined by the informed-excluded rule still has a non-execution problem because of the random mismatch between buyers and sellers. However, it is easy to verify that a change in the level of liquidity trading k will not affect the non-execution probability in this model. Hence, instead of capturing the network externality in which more liquidity traders lead to a higher execution probability (Dönges and Heinemann (2006) and Hendershott and Mendelson (2000)), my model focuses on the non-execution caused by informed trading.

Certainly, it is unrealistic to expect that the crossing network can always exclude the informed trader. Næs and Ødegaard (2006) find evidence of informed trading in the crossing network by examining cumulative abnormal returns of the stocks. In addition, there is a proposed policy change to enforce "fair access" to the crossing network. Suppose this policy change is implemented; then the crossing network cannot exclude any trader. I consider the case where informed trader can trade in the crossing network, which is both realistic and also sheds some light on the effect of the proposed "fair access" policy.

This model can be analytically solved in two cases. In the first case, the informed trader has priority over liquidity traders. This case can be understood as a crossing network with a time priority rule and the informed trader, who has better technology and information, trades faster than the liquidity traders do. In the second case, liquidity buyers and sellers trade first and the informed trader can only trade with the residual of the liquidity order flow. This rule corresponds to the situation in which the informed trader is detected and is placed at the end of the queue.[42] These two rules impose an upper bound and lower bound on the informed trader's impact in the crossing network, where all other rules can be considered combinations of these two extremes. Solving the model under other allocation rules such as a pro rata rule is a formidable task.[43] Fortunately, the qualitative results of the informed-first and the liquidity-first rules are very similar because the only effect of the liquidity-first rule is to decrease the amount of liquidity trading available to the informed trader, whereas the key mechanisms driving the results remain the same. The discussions and proofs in this paper will focus on the informed-first rule, and the liquidity-first rule will be considered in the section on alternative allocation rules.

2.3.3 Equilibrium

In this subsection, I solve the model in two steps. The first step derives the profit function, which is in a complex functional form. Then the model is solved in the second step. The final part of this subsection discusses the comparative statics of the

---

[42] In reality, crossing networks do not like informed traders, and they have all kinds of anti-gaming techniques to minimize the impact of potentially informed order flow.

[43] Pro rata rule means that all the traders get an equal proportion of executed shares for each share he or she submits. If we want to solve the problem under the pro rata rule, we must calculate the expectation of the ratio of two random variables. That is, we must generate the expected volume for the informed trader who wants to submit x_d buy orders in terms of the following function form $\widetilde{x_m} = \begin{cases} x_d * \frac{\widetilde{u_d^s}}{\widetilde{u_d^b}+x_d} & if\ \widetilde{u_d^s} \leq \widetilde{u_d^b} + x_d \\ x_d & \widetilde{u_d^s} > \widetilde{u_d^b} + x_d \end{cases}$. Then, it is very hard to write the functional form of $\mathrm{E}(\widetilde{\mathrm{x_m}}|\mathrm{x_d})$ for the following two reasons. First, the expectation of ratio of two random variables is generally not equal to the ratio of their expectation even if the two variables are independent. Therefore, to calculate $\mathrm{E}(\widetilde{\mathrm{x_m}}|\mathrm{x_d})$ we first need to calculate the joint density of random variables $\widetilde{u_d^s}$ and $\widetilde{u_d^b} + x_d$ or or the expectation of the ratio distribution $\frac{\widetilde{u_d^s}}{\widetilde{u_d^b}+x_d}$. To make things worse, the probability of execution can never be greater than 1. This makes us unable to follow the limited cases in which ratio distribution is well defined. So the best we can do is to express $\mathrm{E}(\widetilde{\mathrm{x_m}}|\mathrm{x_d})$ as integrals. For most distributions of $\widetilde{u_d^s}$ and $\widetilde{u_d^b}$, the integral is difficult or impossible to express in terms of a finite number of elementary functions.

equilibrium, which provides the intuitions for the results on price discovery, volatility, transaction costs and market share.

The definition of equilibrium is as follows:

Definition 2.1*: A rational expectation equilibrium is an informed order submission strategy* X_e *,* X_d *and a market maker's pricing rule P such that the following two conditions hold:*

1. Profit Maximization: For any alternative strategy X_e', X_d' *and any realization v (or* v_g*)*

$$E[\tilde{\pi}(X_e, X_d, P|\tilde{v} = v] \geq E[\tilde{\pi}(X_e', X_d' \ , P|\tilde{v} = v];$$

2. Market Efficiency: The random variable p satisfies

$$\tilde{p}(X_e, X_d, P) = \ E[\tilde{v}|\tilde{y} = \widetilde{x_e} + \widetilde{u_e}]$$

There are two additional comments regarding the market efficiency condition. First, the market efficiency condition implies $\tilde{p} = \ E[\tilde{v}|\tilde{y}]$ instead of $\tilde{p} = \ E[\widetilde{v_g}|\tilde{y}]$ because commission in the exchange is not part of the price. The market maker sets a price equal to $E[\tilde{v}|\tilde{y}]$ and also collects the commission. Second, the market maker needs to conjecture not only the informed strategy in the exchange but also the informed trader's strategy in the crossing network. In the rational expectation equilibrium, the market maker's conjecture on the informer trader's strategies needs to be correct. Therefore, although the market maker can only observe $\tilde{y}$, he can infer $\widetilde{x_e}$ as well as $\tilde{v}$ and $\widetilde{x_d}$ based on his correct belief of X_e and X_d.

I then solve for the equilibrium X_e and X_d as well as the pricing rule P by guessing and verifying. Suppose that $x_e = X_e(v) = \alpha + \beta v$ and $P(y) = \mu + \lambda y$, which are both linear functions. As I will show in the proof, the functional form and coefficient of X_d is uniquely defined by X_e and P.

2.3.3.1 Expected Profit

The risk-neutral informed trader wants to maximize his expected profit. Since informed buying and selling are separable and symmetric in the model, I currently focus on the case in which $v_g \geq p_0 + c$ $(v \geq p_0)$, which represents a scenario in which the informed trader wants to buy. Suppose that the informed trader chooses $\{x_e, x_d\}$. Then he expects that the market maker will set the price at $\tilde{p} = \mu + \lambda(x_e + \widetilde{u_e})$. His expected profit per executed share then becomes $E\left[v_g - \tilde{p}\right] = v_g - \mu - \lambda x_e$ because $\mathrm{E}(\widetilde{u_e}) = 0$. Therefore, his expected profit in the exchange is:

$$E(\widetilde{\pi_e}) = E\left[\left(v_g - \mu - \lambda(x_e + \widetilde{u_e})\right)x_e - cx_e\right] = \left(v_g - \mu - \lambda x_e\right)x_e - cx_e = (v - \mu - \lambda x_e)x_e \qquad (2.4)$$

To determine his expected profit in the crossing network, we first must find the relationship between submitted shares (x_d) and executed shares $\widetilde{x_m}$. $\widetilde{x_m}$ depends on both x_d and the realization of $\widetilde{u_d^s}$. If x_d is larger than the liquidity sell order flow $\widetilde{u_d^s}$, then $\widetilde{u_d^s}$ shares are executed; otherwise, x_d shares are executed. Therefore,

$$\widetilde{x_m}\left(\widetilde{u_d^s}, x_d\right) = \begin{cases} \widetilde{u_d^s} & when \;\; \widetilde{u_d^s} \leq x_d \\ x_d & when \;\; \widetilde{u_d^s} > x_d \end{cases} \qquad (2.5)$$

Thus, the expected number of executed shares conditional on x_d is

$$E(\widetilde{x_m}|x_d) = \int_0^{x_d} z f_s(z)\, dz + x_d \int_{x_d}^{+\infty} f_s(z) dz \qquad (2.6)$$

Lemma 2.1 states the relationship between the numbers of submitted shares and executed shares.

***Lemma 2.1**: $E(\widetilde{x_m}|x_d)$ is increasing in x_d, and if $x_d \neq 0$ the probability of execution $\frac{E(\widetilde{x_m}|x_d)}{x_d}$ decreases with x_d.*

***Proof**: Let* $G(x_d) = E(\widetilde{x_m}|x_d)$. *Then*

$$G'(x_d) = x_d f_s(x_d) - x_d f_s(x_d) + \int_{x_d}^{+\infty} f_s(z)dz = \int_{x_d}^{+\infty} f_s(z)dz > 0$$

$$\left(\frac{G(x_d)}{x_d}\right)' = \frac{G'(x_d)x_d - G(x_d)}{x_d^2} = \frac{x_d\int_{x_d}^{+\infty} f_s(z)dz - \int_0^{x_d} z f_s(z)dz - x_d\int_{x_d}^{+\infty} f_s(z)dz}{x_d^2} = \frac{-\int_0^{x_d} z f_s(z)dz}{x_d^2} < 0$$

The proof does not depend on the functional form of $E(\widetilde{x_m}|x_d)$. The proof holds as long as $f_s(z) > 0$ almost surely when $z \in [0, +\infty)$. Lemma 2.1 captures the probability impact of the order x_d: as the informed trader increases the order size in the crossing network, the expected volume increases but the execution probability decreases. As the matching price is independent of quantity, his total expected revenue increases but his marginal expected revenue decreases.

The crossing network matches orders based on the price in the exchange, $v_g - \mu - \lambda(x_e + \widetilde{u_e})$. Thus, the informed trader's expected profit in the crossing network is

$$E(\widetilde{\pi_d}) = E[\left(v_g - \mu - \lambda(x_e + \widetilde{u_e})\right)\widetilde{x_m}\left(\widetilde{u_d^s}, x_d\right) - cx_d]$$

$$= (v_g - \mu - \lambda x_e)E(\widetilde{x_m}|x_d) - cx_d$$

$$= (v_g - \mu - \lambda x_e)\left(\int_0^{x_d} z f_s(z)dz + x_d\int_{x_d}^{+\infty} f_s(z)dz\right) - cx_d \quad (2.7)$$

The informed trader's optimization problem is to choose $\{x_d, x_e\}$ to maximize his two-period profit. That is,

$$\operatorname*{Max}_{x_d, x_e} E(\tilde{\pi}) = E(\widetilde{\pi_d} + \widetilde{\pi_e}) =$$

$$(v - \mu - \lambda x_e)x_e + (v_g - \mu - \lambda x_e)\left(\int_0^{x_d} z f_s(z)dz + x_d\int_{x_d}^{+\infty} f_s(z)dz\right) - cx_d \quad (2.8)$$

2.3.3.2 Equilibrium Order Submission Strategy and Pricing Rule

Equation (2.8) has two first-order conditions. The first-order condition with respect to x_d is

$$\frac{\partial E(\tilde{\pi})}{\partial x_d} = (\, v_g - \mu - \lambda x_e) \int_{x_d}^{+\infty} z f_s(z) dz - c = 0$$

$$\Rightarrow (\, v_g - \mu - \lambda x_e)(1 - F_s(x_d)) = c$$

$$\Leftrightarrow (\, v_g - \mu - \lambda x_e) Pr(\widetilde{u_d^s} > x_d) = c \qquad (2.9)$$

Equation (2.9) has a very intuitive explanation. The term $(\, v_g - \mu - \lambda x_e)$ represents the informed trader's per unit profit conditional on execution, and $Pr(\widetilde{u_d^s} > x_d)$ is the probability that the x_d^{th} unit is executed. The informed trader chooses x_d such that the marginal profit $(\, v_g - \mu - \lambda x_e) Pr(\widetilde{u_d^s} > x_d) =$ is equal to the order submission cost c. The first-order condition with respect to x_e is

$$\frac{\partial E(\tilde{\pi})}{\partial x_e} = v - \mu - \lambda x_e - \lambda x_e - \lambda \left(\int_0^{x_d} z f_s(z) dz + x_d \int_{x_d}^{+\infty} f_s(z) dz \right) = 0$$

$$\Leftrightarrow \; v - \mu - \lambda x_e = \lambda x_e + \lambda E(\widetilde{x_m} | x_d) \qquad (2.10)$$

Equation (2.10) also has an intuitive explanation. $v - \mu - \lambda x_e$ is exactly the same as the Kyle model. $v - \mu - \lambda x_e$ captures the gain from buying one more share due to the increase in volume. The increase in volume, however, leads the market maker to increase the price by λ causing the informed trader to lose λx_e. Therefore, the informed trader in the Kyle model chooses the optimal x_e based on the trade-off of the volume and the price impact. The crossing network adds another trade-off to the model. The informed trader's order not only has a price impact on the exchange, but also affects his profit in the crossing network. In this sense, his trade in the exchange creates some "externality" for his trade in the crossing network. By trading one more unit in the exchange, the informed trader's expected profit in the crossing network decreases by $\lambda E(\widetilde{x_m}|x_d)$. It is this "externality" that drives the result of the model.

To solve for x_d, the expression of $F_s(x_d)$is plugged into (2.9). The optimal level of x_d, denoted as x_d^*, is

$$(v_g - \mu - \lambda x_e)(\sqrt{\frac{k}{x_d^* + k}}) = c$$

$$\Rightarrow x_d^* = \frac{k}{c^2}(v_g - \mu - \lambda x_e)^2 - k$$

$$\Rightarrow x_d^* = \frac{k}{c^2}(v + c - \mu - \lambda x_e)^2 - k \qquad (2.11)$$

When $v_g \leq p_0 - c$ $(v \leq p_0)$, it is easy to show that the informed strategy is $x_d^* = -\frac{k}{c^2}(\mu + \lambda x_e - v_g)^2 + k = -\frac{k}{c^2}(\mu + \lambda x_e - v + c)^2 + k$. The informed strategies take different forms when $v \leq p_0$ and when $v \geq p_0$, but the expected executed shares, $E(\widetilde{x_m}|x_d^*)$, and the expected profit in the crossing network $E(\widetilde{\pi_d^*})$, , have the same functional form in v. The intuition is that the discontinuity we create for v_g finally cancels out with c. The following lemma summarizes the result.

Lemma 2.2 *The informed trader's order submission strategy is*

$$\begin{cases} x_d^* = \frac{k}{c^2}(v + c - \mu - \lambda x_e)^2 - k & for\ v \geq p_0 \\ x_d^* = -\frac{k}{c^2}(\mu + \lambda x_e - v + c)^2 + k & for\ v < p_0 \end{cases}.$$

which leads to $E(\widetilde{x_m}|x_d^*) = 2\frac{k}{c}(v-\mu-\lambda x_e)$ *and* $E(\pi_d^*) = \frac{k}{c}(v-\mu-\lambda x_e)^2$

Proof. *see the appendix*

Lemma 2.2 states that the informed trader's optimal strategy x_d^* is a quadratic function of v. This quadratic strategy leads to a linear relationship between the expected executed shares $E(\widetilde{x_m}|x_d^*)$ and v. The expected profit in the crossing network is then a quadratic function of v. The linear relationship between $E(\widetilde{x_m}|x_d^*)$and v is the key to obtaining a close-formed solution for the model. This relationship enables my model to merge with the workhorse structure of the rational expectation model, that is, the linear normal framework developed by Grossman and Stiglitz (1980) and Kyle (1985). Therefore, I extend this literature to a market without guaranteed execution.

To solve for the optimal x_e, the expression of $E(\widetilde{x_m}|x_d^*)$ is substituted into (2.10), giving

$$v-\mu-\lambda x_e-\lambda x_e-2\lambda\frac{k}{c}(v-\mu-\lambda x_e)=0 \qquad (2.12)$$

Denote $K=(k/c)$ and the expression for the optimal value of x_e is

$$x_e^* = \frac{1-2\lambda K}{2\lambda-2\lambda^2 K}v - \frac{1-2\lambda K}{2\lambda-2\lambda^2 K}\mu \qquad (2.13)$$

Comparing the coefficient with the conjecture $x_e = \alpha + \beta v$ yields

$$\beta = \frac{1-2\lambda K}{2\lambda-2\lambda^2 K} \qquad (2.14)$$

$$\alpha = -\beta\mu. \qquad (2.15)$$

The market maker sets the clearing price equal to the conditional expectation of v.

$$p = \mu + \lambda y = E\{\tilde{v}|\tilde{x}_e + \tilde{u}_e = y\} = E\{\tilde{v}|\alpha + \beta\tilde{v} + \tilde{u}_e = y\}. \qquad (2.16)$$

The normality of $\tilde{v}$ and $\tilde{u}_e$ makes the conditional expectation linear, and the Projection Theorem yields

$$\lambda = \frac{Cov(\tilde{v},\tilde{y})}{Var(\tilde{y})} = \frac{\beta\sigma_v^2}{\beta^2\sigma_v^2+\sigma_e^2} = \frac{\beta}{\beta^2+\sigma_e^2/\sigma_v^2} = \frac{\beta}{\beta^2+R} \qquad (2.17)$$

$$\mu = p_0 - \lambda(\alpha + \beta p_0). \qquad (2.18)$$

where I define $R = \frac{\sigma_e^2}{\sigma_v^2}$ to simplify the notation. Combining (2.15) and (2.18) yields

$$(\lambda\beta - 1)(p_0 - \mu) = 0. \qquad (2.19)$$

From Equation (2.17), we know $\lambda\beta = (\beta^2)/(\beta^2 + R) < 1$, which implies that

$$\mu = p_0 \text{ and } \alpha = -\beta p_0. \qquad (2.20)$$

λ can then be solved for by substituting (2.14) into (2.17) to give

$$\lambda = \frac{\frac{1-2\lambda K}{2\lambda-2\lambda^2 K}}{(\frac{1-2\lambda K}{2\lambda-2\lambda^2 K})^2+R} \Rightarrow$$

$$R(2\lambda - 2\lambda^2 K)^2 = (1 - 2\lambda K). \qquad (2.21)$$

β is uniquely defined by (2.14) for any λ. In turn, α is uniquely defined by (2.15). Then the key to solve the model is to solving (2.21), which is a depressed quartic equation in

λ.[44] Closed form solutions can be obtained using the Ferrari method. However, analytical forms of the solutions are not presented explicitly because they are overwhelmingly complex in R and K. It is easier to prove the existence and uniqueness of the solution and conduct comparative statics by analyzing (2.14) and (2.21). The analytical forms of the solutions are available upon request from the author.

When $K = 0$, meaning that the crossing network does not exist, the model degenerates into the Kyle model. It is easy to see from (2.14) and (2.21) that there is a unique solution $\beta = \sqrt{R} = \frac{\sigma_e}{\sigma_v}$ and $\lambda = (1/(2\sqrt{R})) = \frac{\sigma_e}{\sigma_v}$. For $K > 0$, the existence and uniqueness of the solution is established by the following two Lemmas.

Lemma 2.3 *Existence of real solutions: for any $K > 0$, there are exactly two real solutions $\lambda_1 \in (0,1/2K)$ and $\lambda_2 \in (-\infty, 0)$ for (2.21). β_1 that corresponds to the solution $\lambda_1 > 0$ is also greater than 0; β_2 that corresponds to the solution $\lambda_2 < 0$ is also smaller than 0. For $K = 0$, there is a unique solution $\beta_1 = \sqrt{R} = \frac{\sigma_e}{\sigma_v}$ and $\lambda_1 = (1/(2\sqrt{R})) = \frac{\sigma_v}{2\sigma_e}$.*

Proof. *see the appendix*

Next, Lemma 2.4 states the uniqueness of the linear equilibrium. The solution $\beta_2 < 0$ corresponds to "bluffing", a scenario in which the informed trader trades in the wrong direction to mislead the price and then benefits from the resulting mispricing by

[44] A depressed quartic equation is a quartic equation with no cubic term. In the 16th century, Italian mathematician Lodovico Ferrari found the formula to express the solution of any depressed quartic equation in terms of its coefficients.

matching orders in the crossing network. "Bluffing" does not constitute an equilibrium in my framework because under the rational expectation framework, the market maker should have correct beliefs of the informed trader's strategies. If the informed trader is bluffing, the market maker should know that the informed trader is bluffing and should set $\lambda_2 < 0$, meaning that the market maker decreases the price with the net buy order flow and raises the price by observing the net sell order flow. However, conditional on $\lambda_2 < 0, \beta_2 < 0$ is not optimal for the informed trader.

Lemma 2.4 *Uniqueness of the solution: for $K > 0$, only the solution with $\lambda_1 > 0, \beta_1 > 0$ constitutes an equilibrium.*

Proof. see the appendix

Lemmas 2.3 and 2.4 and (2.14), (2.20) and (2.21) establish the unique linear equilibrium in this model, which is characterized by the unique solutions $\beta^* = \beta_1$ and $\lambda^* = \lambda_1$:

Theorem 2.1 *There exists a unique linear equilibrium in which: the informed trader trades* $X_e(v) = \beta^*(v - p_0)$, $X_d(v) = \begin{cases} \frac{k}{c^2}(v + c - p_0 - \lambda^* x_e)^2 - k & for\ v \geq p_0 \\ -\frac{k}{c^2}(p_0 + \lambda^* x_e - v + c)^2 + k & for\ v < p_0 \end{cases}$, *and the price function is* $P(y) = p_0 + \lambda^* y$. λ^* *is the unique positive solution of the equation* $R(2\lambda^* - 2\lambda^{*2}K)^2 = (1 - 2\lambda^* K)$. $\beta^* = \frac{1-2\lambda^* K}{2\lambda^* - 2\lambda^{*2}K}$. $R = \frac{\sigma_e^2}{\sigma_v^2} > 0$ *and* $K = \frac{k}{c} \geq 0$ *are the parameters of the model.*

2.3.4 Comparative Statics

The equilibrium of the model is characterized by β^* and λ^*. Lemma 2.5 provides the comparative statics for β^*, $\lambda^*$and x_d^*.These comparative statics provide the intuition to explain the results on price discovery, transaction costs and market share.

Lemma 2.5 *The informed trader trades in both markets unless $v = p_0$. β^* increases in σ_e and decreases in k and σ_v; λ^* decreases in k and σ_e and increases in σ_v; the size of informed trader's order in the crossing network, $|x_d^*|$, increases in k and σ_v and decreases in σ_e.*

Unless the signal is of zero value ($v = p_0$), the informed trader always trades in both markets for the following reason. When the informed trader does not trade in the exchange, the price impact of the trade is zero. Therefore, the informed trader always finds it profitable to trade at least some small amount in the exchange. Similarly, the execution probability in the crossing network approaches 1 when the informed trader only wants to trade an infinitesimal amount. Therefore, the informed trader always trades in both markets when the signal is valuable.

Because the informed trader always trades in the exchange when $v \neq p^0$, the market maker can infer the informed trader's signal through the order flow. However, compared to the case without crossing network ($k = 0$), the informed trader wants to trade less in the exchange when a crossing network exists because the price impact of his trade in the exchange imposes a negative externality on his profit in the crossing network. Therefore, β^* decreases, and the aggregated order flow becomes less informative. Meanwhile, the price impact λ^* decreases due to a decrease in the level of informed trading in the exchange.

An increase in the fundamental value uncertainty increases the informed trader's order size in the crossing network but decreases his order size in the exchange, because an increase in the fundamental value uncertainty increases the value of the information for the informed trader. Therefore, the informed trader has a greater incentive to hide in the crossing network.

2.4 Price Informativeness and Volatility

2.4.1 Price Discovery

One of the most important functions of the securities market is to provide price discovery (O'Hara, 2003). This subsection will show that the crossing network reduces price discovery, which is an intuitive result. Because the price impact of trading in the exchange creates an externality affecting the informed trader's profit in the crossing network, the informed trader chooses to trade less in the exchange. Therefore, the order flow becomes less informative, and price discovery is impeded. Next, I will give a formal proof of this result and also study the determinants of the size of the effect.

The market maker sets the price based on y, which is the signal he receives.

$$y = \widetilde{x_e^*} + \widetilde{u_e} = \beta^*(\tilde{v} - p_0) + \widetilde{u_e} \qquad (2.22)$$

Rearranging terms yields

$$\theta \equiv \frac{y}{\beta^*} + p_0 = \tilde{v} + \frac{\widetilde{u_e}}{\beta^*} \qquad (2.23)$$

where θ is an informationally equivalent transformation of the observed order flow y that has the same mean as the underlying asset. Conditional on v, θ is distributed as

$N(v, \frac{\sigma_e^2}{\beta^{*2}})$.The externality of the price impact on the crossing network results in a decrease of β^* and an increase of $\frac{\sigma_e^2}{\beta^{*2}}$.

Price informativeness is defined in the same way as that in the Kyle model, which is equal to 1 minus the ratio of the posterior variance of $\tilde{v}$ to the prior variance of $\tilde{v}$, denoted

$$e = 1 - \frac{var(\tilde{v}|\tilde{p})}{var(\tilde{v})} \qquad (2.24)$$

Note that $var(\tilde{v}|\tilde{p}) = var(\tilde{v}|\tilde{y}) = var(\tilde{v}|\tilde{\theta})$ because $\tilde{\theta}$, $\tilde{p}$ and $\tilde{y}$ are informationally equivalent.

When the price is perfectly informative, $e = 1$. When the price is pure noise, $e = 0$. Bayes' rule states that the posterior variance $var(\tilde{v}|\tilde{p})$ can be expressed as the following function of the prior variance and the variance of the signal[45]:

$$var(\tilde{v}|\tilde{p}) = var(\tilde{v}|\tilde{\theta}) = \left(\frac{1}{\sigma_v^2} + \frac{\beta^{*2}}{\sigma_e^2}\right)^{-1} = \frac{\sigma_v^2}{1+\beta^{*2}\frac{\sigma_v^2}{\sigma_e^2}} \qquad (2.25)$$

Plugging equation (2.14) and (2.21) into (2.25), I obtain

$$e = 1 - \frac{1}{1 + \frac{\beta^{*2}\sigma_v^2}{\sigma_e^2}} = 1 - \frac{1}{1 + (\frac{1-2\lambda^*K}{2\lambda - 2\lambda^{*2}K})^2 \frac{\sigma_v^2}{\sigma_e^2}}$$

$$= 1 - \frac{1}{1+R(1-2\lambda^*K)\frac{1}{R}} = 1 - \frac{1}{2-2\lambda^*K} \qquad (2.26)$$

[45] For a derivation of the formula, see O'Hara (1995) appendix to Chapter 3.

When $K \equiv k/c = 0$, e reaches its maximum, and its value is 0.5 for any value of R. A non-empty crossing network changes these two results. When $K > 0$, then $e < 0.5$, meaning that the crossing network always makes the price less informative.

Theorem 2.2 states the relationship between price discovery and the four exogenous variables in the paper.

Theorem 2.2 *The crossing network harms price discovery as the price informativeness measure, e, reaches its maximum when* $k = 0$*. If* $k = 0$*, e is independent of* σ_e*,c,and* σ_v*; otherwise, e is uniquely determined by* $\frac{\sigma_e c}{k\sigma_v}$ *and increases in* $\frac{\sigma_e c}{k\sigma_v}$ *.*

Proof. *see the appendix.*

Theorem 2.2 has two interesting implications. First, when $k = 0$, my model degenerates into the Kyle model. Therefore, the information revelation is independent of σ_v and σ_e. The existence of a crossing network changes this prediction. Second, the four exogenous variables in the model (the levels of liquidity trading in the crossing network and the exchange, the fundamental value uncertainty and the up-front submission cost) all affect price discovery, and the degree to which these variables affect price discovery is determined by the ratio $\frac{\sigma_e c}{k\sigma_v}$. If, for example, both σ_e and k double, the level of information revelation does not change. Additionally, an increase in the level of liquidity trading k is equivalent to a similar increase in the value of σ_v in terms of their impact on price discovery, because it is the product of these two variables that determines price discovery.

Theorem 2.2 is proven under the assumptions that the crossing network attracts additional liquidity traders and that the liquidity order flow in the exchange remains the same. Suppose that the crossing network not only creates its own liquidity traders, but also steals some liquidity traders from the exchange. Then σ_e will decrease and k will increase. Price discovery will be further reduced because Theorem 2.2 states that a decrease of liquidity trading in the exchange and an increase in liquidity trading in the crossing network reduce price discovery.

Information revelation is less for stocks with higher values of σ_v, because the information on $\tilde{v}$ is more valuable for stocks with higher fundamental value uncertainty. Therefore, the informed trader has a greater incentive to hide his information in the crossing network, thereby decreasing the informativeness of the price.

Figure 2.3 illustrates the relationship between e and exogenous parameters σ_e, k and σ_v. The graphs of k and σ_v are identical, meaning that an increase in the liquidity trading in the crossing network has the same effect as an increase in the fundamental value uncertainty.

2.4.2 Volatility

The price volatility is measured by $var(\tilde{p})$. Under my framework, e and $var(\tilde{p})$ have the following relationship:

Lemma 2.6 $var(\tilde{v}) = var(\tilde{v}|\tilde{p}) + var(\tilde{p}) and\ var(\tilde{p}) = e\sigma_v^2$

Proof. *see the appendix*

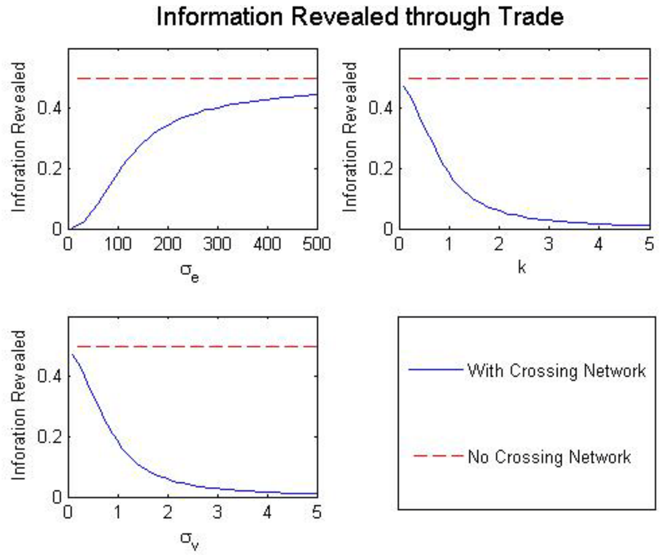

Figure 2.3: This figure demonstrates the information revealed through trade. σ_e is the proxy for liquidity trading in the exchange; σ_v is the fundamental value uncertainty; and k is the proxy for liquidity trading in the crossing network. For all three panels, σ_e*=100*, σ_v *=1*, *k=1* and *c=0.01* unless otherwise specified.

Lemma 2.6 states the relationship between the price volatility and price discovery. The Projection Theorem decomposes the prior variance of $\tilde{v}$ into two parts: the part that can be explained by $\tilde{p}$, $var(\tilde{v}|\tilde{p})$, and the part that can not be explained by $\tilde{p}$, $var(\tilde{p})$. More efficient price discovery means that $var(\tilde{v}|\tilde{p})$ should be smaller, which implies that $var(\tilde{p})$ is higher.

The comparative statics of $var(\tilde{p})$ are summarized in Corollary 2.1 and Figure 2.4. Corollary 2.1 shows that price volatility is positively correlated with fundamental value uncertainty, meaning that the observed price volatility can serve as a proxy for the underlying fundamental value uncertainty.

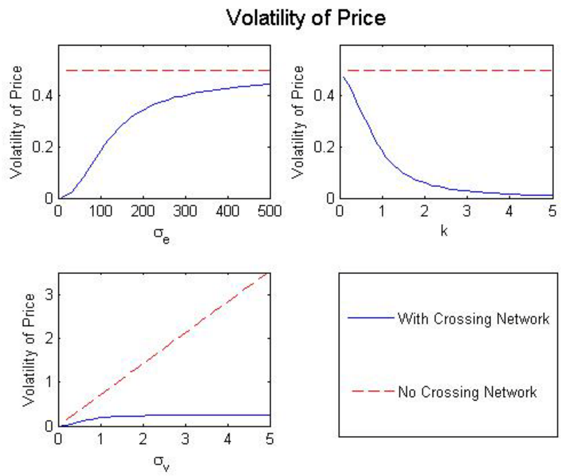

Figure 2.4: This figure demonstrates the price volatility. σ_e is the proxy for liquidity trading in the exchange; σ_v is the fundamental value uncertainty; and k is the proxy for liquidity trading in the crossing network. For all three panels, *σ_e=100, σ_v =1, k=1* and *c=0.01* unless otherwise specified.

However, in a two-period setup, Corollary 2.1 establishes a very surprising relationship between σ_v and $var(\tilde{p})$.When there is a crossing network, no matter how large the fundamental value uncertainty, price volatility can never be higher than $\frac{\sigma_e^2 c^2}{4k^2}$, a constant that is independent of σ_v. The pattern can be seen clearly in the bottom left panel in Figure 2.4. In the benchmark Kyle model, price volatility is linear in the fundamental value uncertainty. The result changes when there is a crossing network. When σ_v is low, an increase in σ_v causes a substantial increase in the price volatility, though the increase is less than a linear increase. The price volatility still increases in $\sigma_v$when σ_v is high, but the change is minuscule and approaches 0 as σ_v increases.

Corollary 2.1 $var(\tilde{p}) = \lambda^{*2}(2 - 2\lambda^* \frac{k}{c})\sigma_e^2$ *. The crossing network decreases price volatility because $var(\tilde{p})$ is largest when $k = 0$. When $k \neq 0$, $var(\tilde{p})$ decreases in $\frac{\sigma_e c}{k}$and increases in σ_v. However, for ∀ σ_v, $var(\tilde{p}) < \frac{\sigma_e^2 c^2}{4k^2}$*

Proof. *see the appendix*

This surprising result, however, has an intuitive explanation under the two-period setup. Higher σ_v values indicate that information regarding the fundamental value becomes more valuable for the informed trader. The informed trader then has a greater incentive to hide his information in the crossing network and trade less in the exchange. Therefore, the information revelation e decreases, which counteracts the increase in σ_v. The overall effect is that $var(\tilde{p})$ cannot be greater than $\frac{\sigma_e^2 c^2}{4k^2}$ because $var(\tilde{p})$=$e\sigma_v^2$. A more intuitive way to understand this is to consider the realization of $\tilde{v}$ as "private news". When there is a deep crossing network, the informed trader will hide his trade in the crossing network. In the short run, the market maker may not even know that some "news" actually took place, nor is the market marker able to tell whether the news was positive or negative. Therefore, he is not able to adjust quotes in the short run, and thus we can see a decrease in the price volatility.

In conclusion, the crossing network reduces both price discovery and price volatility. In the framework of this paper, it is impossible for the crossing network to both harm price discovery and increase price volatility, because lower price discovery directly implies lower price volatility.

2.5 Liquidity and Transaction Costs

This section addresses three questions. I first analyze price impact, the measure I employ to capture the liquidity and transaction cost of the exchange. Next, liquidity and transaction cost of the crossing network as measured by the non-execution probability are analyzed. The final part of this section analyzes the relationship between price impact and non-execution, both of which are endogenous in this model.

2.5.1 Price Impact

For the exchange, λ^* serves as the inverse measure of the liquidity or as a direct measure of price impact. Theorem 2.3 shows the relationship between λ^* and the exogenous variables $\frac{k}{c}$ and $\frac{\sigma_e^2}{\sigma_v^2}$.

Theorem 2.3. The existence of a crossing network puts an upper limit on λ^*in that $\lambda^* < \frac{c}{2k}$ for any $\frac{\sigma_e^2}{\sigma_c^2}$. λ^* is decreasing in $\frac{\sigma_e^2}{\sigma_c^2}$ and $\frac{k}{c}$.

Proof. See the appendix.

In the benchmark Kyle model, the price impact can go to infinity. This possibility is ruled out by a crossing network because of the externality of price impact on the crossing network. This externality makes the informed trader trade less aggressively, decreases the information asymmetry problem in the exchange, and thereby leads the market maker to set a less aggressive price. Figure 2.5 compares λ^* and β^* with the corresponding parameters in the Kyle model. Note that λ^* is always smaller than the Kyle λ. In addition, the Kyle λ converges to infinity as $\frac{\sigma_e}{\sigma_v} \to 0$, whereas in this model $\lambda^* \to \frac{c}{2k} = \frac{1}{2k} = \frac{1}{2K}$ as $\frac{\sigma_e}{\sigma_v} \to 0$.

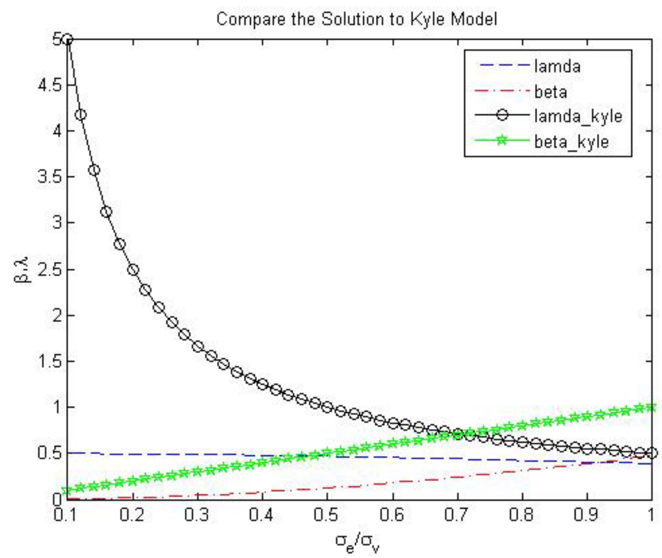

Figure 2.5: This figure compares the equilibrium λ and β with their counterparts in the Kyle model. I choose K≡(k/c)=1 to fit λ and β in the same graph.

Theorem 2.3 states that the crossing network improves the liquidity of the exchange if it can attract new liquidity traders. Dönges and Heinemann (2006), Degryse, Achter and Wuyts (2009) and Hendershott and Mendelson (2000) all show that crossing networks can attract liquidity traders who are unwilling to trade in the exchange. Hendershott and Mendelson (2000) also show that the crossing network may improve the liquidity of the exchange by attracting new liquidity traders. The Hendershott and Mendelson model, however, assumes exogenous informed traders. Therefore, the mechanism driving their result is that a fixed amount of informed order flow is diluted by an increase of liquidity order flow. My result shows that even if the informed trader can choose how much to trade, the crossing network can still benefit the exchange if it can attract new liquidity traders.

Certainly, what I find is only one mechanism, and there are other mechanisms working against the mechanism I find. The claim that crossing networks definitely harm the exchanges is usually based on the implicit assumption of so called "cream-skimming", in which crossing networks draw liquidity traders away from exchanges, leaving mostly informed traders trade in the exchange. However, there are at least two factors working against cream-skimming. First, when the liquidity traders move to the crossing network, the informed trader will follow them. Additionally, new trading mechanisms may attract new liquidity traders who are unwilling to trade in other platforms.

Empirically, cream-skimming has been rejected. Gresse (2006) finds that more crossing trading decreases the transaction costs of the exchange. Fong, Madhavan and Swan (2004) also do not find that crossing networks have an adverse effect on the exchange. Both papers suggest that cream-skimming, if it exists, must be counteracted by some other effects. This paper provides one possible explanation to their findings.[46]

2.5.2 Non-execution

The traditional measures of transaction costs, bid-ask spread and price impact, are irrelevant for the crossing networks. A measure of transaction cost in the crossing network is non-execution. Interestingly, though non-execution in the crossing network and price impact in the exchange are different dimensions of execution costs, they share the same underlying factor - order imbalance. In the crossing network, orders fail to execute because of the imbalance between buy and sell orders. Larger order imbalances lead to higher non-execution probabilities. In classical models of exchanges, the transaction costs is also an implicit function of order imbalance. This imbalance may be caused by non-informational factors, which are the focus of inventory models such as Ho and Stoll (1983). Order imbalance can also be caused by informational factors. On

[46] An alternative explanation is that competition cause the exchange to decrease the trading cost and increase the liquidity.

one side of the market, there are both liquidity and informed traders, and on the other side there are only liquidity traders. In the models of Glosten and Milgrom (1985), Easley and O'Hara (1987), and Kyle (1985), orders would be imbalanced without the intervention of the market maker. In both inventory- and information-based models, order imbalances are positively correlated with transaction costs. The only difference between the crossing network and the exchange is that there is no market maker to clear the order imbalance in the crossing network. Therefore, instead of causing a higher price impact, a higher order imbalance in the crossing network leads to a lower execution probability.

As with order imbalance, non-execution can also be explained by non-informational and informational factors. It is important to separate these two parts is important because they have different implications. In a world without information asymmetry, the expected price change is 0 after each trade. Conversely, non-execution caused by informational sources has an adverse selection effect: the future price is more likely to move in the unfavorable direction of the executed orders because the executed orders are more likely to be on the wrong side of the market.

Previous literature on non-execution has focused on non-informational factors. In Dönges and Heinemann (2006) and Hendershott and Mendelson (2000), non-execution is primarily a function of liquidity externality.[47] In these models, buy and sell order flows follow independent geometric distributions. As the mean of the geometric distribution increases, the non-execution probability decreases because it is easier to find a potential match in a deep market. These two models imply that the major task for the crossing network is to attract more liquidity traders. Crossing networks that cannot

[47] Informed traders in Hendershott and Mendelson (2000) have a passive role.

attract enough liquidity traders will fail due to a low execution probability, whereas crossing networks with many liquidity traders will attract even more liquidity traders.

This paper supplements the Hendershott and Mendelson model and the Dönges and Heinemann model by focusing on the informational causes of non-execution.[48] When non-execution is caused by the informed trader, the execution probability may decrease as the amount of liquidity trading increases. This is because the information asymmetry problem cannot be mitigated simply by an increase in liquidity trading. The present paper shows that an increase in the liquidity order flow leads to a greater increase in the informed order flow, which results in a lower execution probability. Therefore, a crossing network cannot increase its execution probability only by attracting more liquidity traders. It also requires anti-gaming strategies to defend itself from informed traders. More discussions of anti-gaming strategies can be found in section 2.7.

Before continuing, an important distinction must be emphasized. There are two measures of non-execution. The probability of non-execution for the informed trader's order depends on three random variables: $\tilde{v}$, $\widetilde{u_e}$ and the liquidity volume on the opposite side. The informed trader considers this probability in his optimization problem. The non-execution probability of the entire market also involves the liquidity trader on the same side as the informed trader. The proof presented here is based on the non-execution probability for the informed trader. The non-execution probability for the entire market depends on the four random variables v, $\tilde{u}_e$, $\tilde{u}_d^b$ and $\tilde{u}_d^s$ and cannot be solved analytically. However, the simulated results are explained by the theorems presented in this subsection because liquidity traders are passive players in this model.

[48] The noninformational causes of non-execution does not play a role in my model. Suppose that there are no informed traders in this model; then the execution probability is independent of k.

Theorem 2.4 first states the relationship between execution probability and two endogenous variables: price discovery and price impact. This theorem also relates the execution probability to the exogenous variables. The proof is conditional on the realization of the fundamental value v. However, the result of Theorem 2.4 also holds unconditionally.[49]

Theorem 2.4 *The probability of execution for the informed trader conditional on* v, $exe_i = \frac{E(\widetilde{x_m}|x_d^*,v)}{x_d^*}$,*has the following form*

$$exe_i = \frac{2c}{2c+|(v-p_0)|(1-\lambda^*\beta^*)} = \frac{2c}{2c+|(v-p_0)|(1-e)}.$$

The probability of execution, exe_i *, decreases in upfront submission cost c and realization of fundamental value* v, *and increases in* $\frac{\sigma_e}{k\sigma_v}$.

Proof. See the appendix

In my model, price impact, information revelation and non-execution are endogenously determined. Theorem 2.4 establishes the relationship between these values. For example, when price discovery e increases, exe_i decreases, reflecting the intuition that more information revelation in the exchange causes the informed trader to move his trading from the exchange to the crossing network, which increases the order imbalance and leads to lower execution probability. Theorem 2.4 also relates exe_i , to the

[49] It is very straightforward to see that the results of Theorem 2.4 holds unconditionally for c, σ_e and k because they are true for any realization of v. The result for σ_v is less obvious, because by changing σ_v , the distribution of $\tilde{v}$ is also altered, thus making it impossible to compare the result state-by-state. Therefore, I can only compare the result by simulation. The result is not presented here because of space considerations and because it is very similar to that in Figure 2.6.

exogenous variables. An increase in v increases the profit per matched unit, which decreases exe_i, because now the informed trader demands a lower execution probability to break even. The informed trader's information is more valuable for stocks with higher fundamental value uncertainty σ_v. Therefore, he has a greater incentive to hide his information in the crossing network, which leads to a lower execution probability. The execution probability increases with $\sigma_e$because an increase in liquidity trading in the exchange attracts part of the informed order flow from the crossing network to the exchange.

The most surprising result is that the execution probability decreases with the level of liquidity trading in the crossing network, a prediction opposite of that suggested by Dönges and Heinemann (2006) and Hendershott and Mendelson (2000). In these two models, an increase in liquidity trading has two effects. The first effect provides a network externality, and the second effect dilutes a fixed amount of informed order flow. However, when the informed trader optimizes, he will increase his order size more than the increase in liquidity order flow. This paper provides two mathematically rigorous proofs of this result. One is based on comparative statics, which is shown under the proof for Theorem 2.4. The other proof is to decompose the total impact of an increase in liquidity trading into two effects: volume effect and price effect. Details of these two proofs can be found in Appendix A and B. However, the reason why the execution probability decreases with the level of liquidity trading in the crossing network can be understood intuitively. The key to explain this result is the difference in market structure. While the market maker in the exchange can actively adjust quotes to protect himself from the informed trader, a crossing network with a fixed allocation rule is passive. As liquidity trading in the crossing network increases, the informed trader considers the crossing network more favorable and moves even more trades to the crossing network than the liquidity traders do.

The non-execution probability for the entire market depends on three degrees of uncertainty: $\widetilde{x_d^*}$(or $\tilde{v}$), $\widetilde{u_d^b}$ and $\widetilde{u_d^s}$ .To obtain the non-execution probability for the entire market, I first define the matched volume as:

$$\widetilde{Vol}_d = \begin{cases} 2\min\left(\left|\widetilde{x_d^*} + \widetilde{u_d^b}\right|, \left|\widetilde{u_d^s}\right|\right) & for \;\; \widetilde{x_d^*} > 0 \\ 2\min\left(\left|\widetilde{x_d^*} + \widetilde{u_d^s}\right|, \left|\widetilde{u_d^b}\right|\right) & for \;\; \widetilde{x_d^*} \leq 0 \end{cases} \qquad (2.27)$$

When $\widetilde{x_d} > 0$, the informed trader wants to buy. In this case, the buy side has both informed and liquidity orders, while the sell side only has liquidity traders. The number of matched shares is equal to twice the number of shares on the side with fewer submitted shares. The definition is similar when $\widetilde{x_d} \leq 0$. The probabilities of execution for the entire market are then defined as

$$\widetilde{exe}_w = \frac{\widetilde{Vol}_d}{\left|\widetilde{x_d^*}\right| + \left|\widetilde{u_d^b}\right| + \left|\widetilde{u_d^s}\right|} \qquad (2.28)$$

The non-execution probability is simply 1 minus the execution probability.

$$\widetilde{nonexe}_w = 1 - \widetilde{exe}_w \qquad (2.29)$$

Figure 2.6 demonstrates the pattern of non-execution probability for the entire market with respect to the exogenous variables. The non-execution probability for the entire market also decreases in the level of liquidity trading in the exchange and increases in fundamental value uncertainty. Figure 2.6 also shows that non-execution for the entire market is a decreasing function of the liquidity trading in the crossing network.

Therefore, a crossing network cannot increase its execution probability only by attracting more liquidity traders; the crossing network also needs anti-gaming strategies to defend itself from informed traders. A discussion of anti-gaming strategies can be found in Section 2.7.

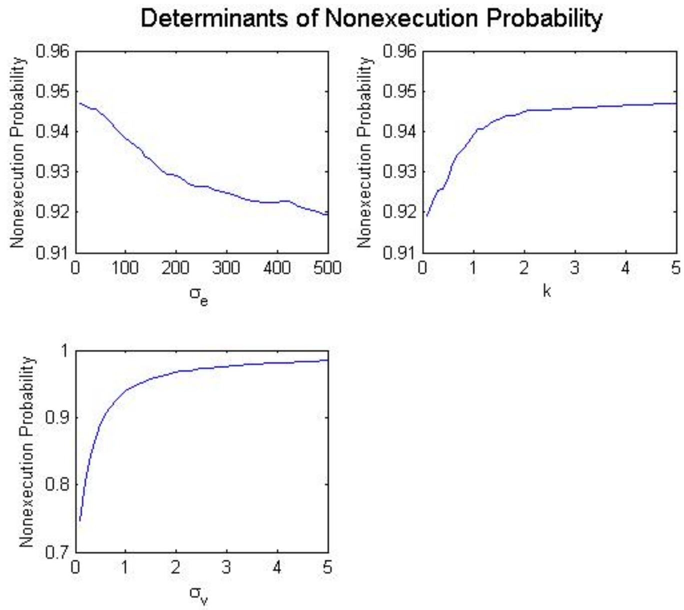

Figure 2.6: This figure demonstrates the simulated non-execution probability. σ_e is the proxy for liquidity trading in the exchange; σ_v is the fundamental value uncertainty; and k is the proxy for liquidity trading in the crossing network. For all three panels, *σ_e=100, σ_v =1, k=1* and *c=0.01* unless otherwise specified.

2.5.3 The Relation between Price Impact and Non-execution

Due to its ability to characterize both price impact and non-execution, this model generates predictions on the relationship between these two transaction costs. Figure 2.7 illustrates the patterns of λ^* and of simulated $E(\widetilde{nonexe_w})$with respect to exogenous variable σ_v,[50] which is consistent with the empirical findings of Næs and Skjeltorp (2003) that stocks that are hard to execute in the crossing network are more volatile than

[50] $E(\widetilde{nonexe_w})$ is the average of one million simulated $E(\widetilde{nonexe_w})$ For each simulation, four realizations of $\tilde{v}$, u_e , u_d^b and u_d^s are independently drawn, $\tilde{x}_d$ is obtained using the informed trader's optimal strategy of (2.11) and $\widetilde{nonexe_w}$ is obtained by Equations (2.27), (2.28) and (2.29)

stocks that are easy to execute in the crossing network[51] and that the non-execution probability is positively correlated with the price impact. My model shows that both price impact and non-execution probability increase as σ_v increases and that price impact and non-execution probability are positively correlated. This positive correlation can be easily understood through the informed trader's effort to balance his trading cost in these two markets.

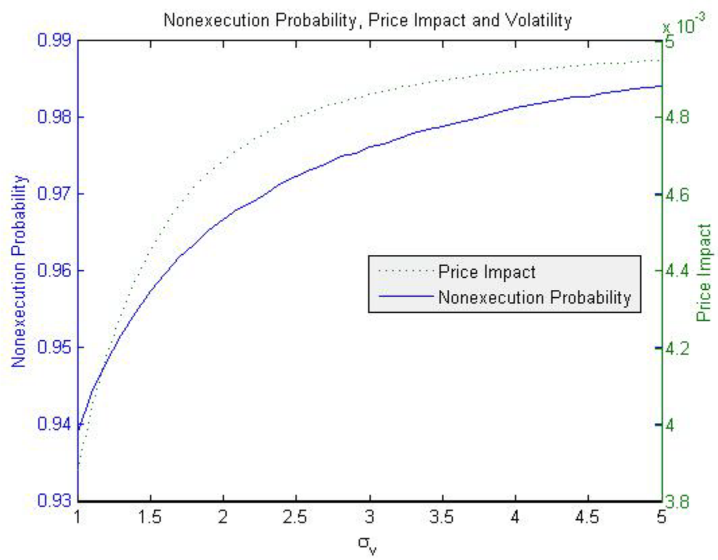

Figure 2.7: This figure illustrates the pattern of price impact and simulated non-execution probability with respect to fundamental value uncertainty σ_v. I set σ_e*=100, k=1* and *c=0.01*

Figure 2.7 also sheds some light on the research on competition between different trading mechanisms. Because price impact and non-execution probability are highly correlated, they may provide limited insight in studying the competition between the exchange and the crossing network. Conditional on execution, stocks with a higher

[51] Certainly, the volatility in their empirical model is observable price volatility, but σ_v in my model is the unobservable fundamental value uncertainty. However, Corollary 2.1 states that the observable price volatility and the unobservable fundamental value uncertainty are positively correlated. Therefore, price volatility can serve as a proxy for σ_v.

price impact may result in larger savings by trading in the crossing network. However, this potential savings may be offset by lower execution probability. Therefore, the relationship between price impact, non-execution and market share is ambiguous. Two verifications of this conjecture are presented in Ready (2009). First, Ready (2009) finds that the bid-ask spread, described by Amihud (2002) as price impact for "standard-size transactions", cannot explain the market share of crossing networks. Additionally, Ready finds that crossing network volumes are not the highest in the highest volume stocks, where the likelihood of finding a counterpart should be the highest. Both findings suggest that transaction costs may be poor explanatory variables for competition between the exchange and the crossing network. I will examine other explanatory variables in the section on market share.

2.6 The Order Splitting Strategy and Market Share

The results on market share are driven by the informed trader's order splitting strategy as well as the liquidity trading in each market. The first part of this section will focus on the order splitting strategy of the informed trader. This order splitting strategies is the key to understanding the simulated results regarding market share in the second subsection.

2.6.1 The Order Splitting Strategy

At equilibrium, the informed trader submits x_e^* to the exchange and x_d^* to the crossing network. However, the market share depends on $E(\widetilde{x_m}|x_d^*)$ because only part of the order x_d^* is executed. Theorem 2.5 states the order splitting strategy.

Theorem 2.5 $\frac{E(\widetilde{x_m}|x_d^*)}{x_e^*} = 1 - \frac{1}{1-2\lambda^* K}$. $\frac{E(\widetilde{x_m}|x_d^*)}{x_e^*}$ *is uniquely determined by* $\frac{\sigma_e c}{k\sigma_v}$ *and decreases with* $\frac{\sigma_e c}{k\sigma_v}$*. For a given* $\frac{\sigma_e c}{k\sigma_v}$, $\frac{E(\widetilde{x_m}|x_d^*)}{x_e^*}$ *is independent of* v.

Proof*. See the appendix*

Theorem 2.5 states that an increase in σ leads the informed trader to submit relatively more shares to the exchange, whereas an increase in k makes him submit relatively more shares to the crossing network. If σ_e and k increase at the same rate, the informed trader maintains the ratio of trades made in these two markets. An increase in c discourages informed trading in the crossing network, while an increase in fundamental value uncertainty encourages trading in the crossing network, because an increase in fundamental value uncertainty provides the informed trader with a higher incentive to hide information in the crossing network. Interestingly, while an increase in σ_v makes the informed trader trade relatively more in the crossing network, an increase in the realization of v would not. This is because an increase in v would cause similar increases in $E(\widetilde{x_m}|x_d^*)$ and x_e^*, as both are linear functions of v.

2.6.2 Market Share

The market share of the crossing network follows the intuition behind the informed trader's order splitting strategy, and the result can be obtained through simulation. The prerequisite to study the market share is to define the volume in each market.

In the exchange, if the informed and liquidity order flow are on the same side of the market (buy or sell), the market maker must trade with both the informed and the liquidity traders. Thus, the volume is the absolute value of the sum of the informed and liquidity order flows. If the informed order flow and the liquidity order flow are on different sides of the market, they can trade with each other and the market maker only needs to offset the imbalance between these two order flows. Thus, the volume is equal to the side with the larger order flow. Therefore, the volume of the exchange is

$$\widetilde{Vol}_e = \begin{cases} |\widetilde{x_e^*} + \tilde{u}_e| & \widetilde{x_e^*} * \tilde{u}_e > 0 \\ \max(|\widetilde{x_e^*}|, |\tilde{u}_e|) & for\ \widetilde{x_e^*} * \tilde{u}_e \leq 0 \end{cases} \quad (2.30)$$

The volume of the crossing network has already been defined by (2.27). Thus, the market share of the crossing network is defined as

$$share_d = E(\frac{\widetilde{Vol}_d}{\widetilde{Vol}_e + \widetilde{Vol}_d}) \quad (2.31)$$

An increase in k or σ_e first causes an exogenous increase in liquidity volume in the crossing network or the exchange, respectively, and the results are enhanced by an increase in the informed volume based on Theorem 2.5. Therefore, it is very straightforward to see that $share_d$ increases with k and decreases with σ_e.

The relationship between $share_d$ and σ_v is shown in Figure 2.8. In the previous section, it was demontrated that non-execution and price impact both increase in σ_v .An increase in the fundamental value uncertainty σ_v , however, gives a comparative advantage to the crossing network, because the informed trader has a high incentive to hide his information for stocks with high σ_v.

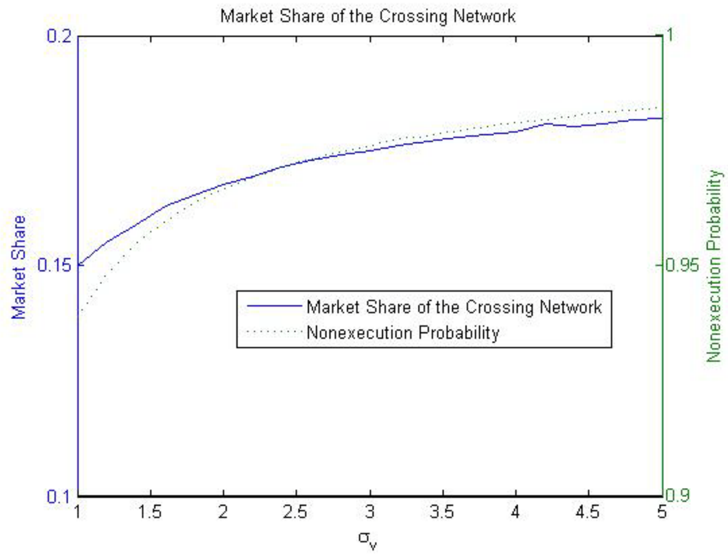

Figure 2.8: This figure illustrates the market share of the crossing network with respect to the exogenous variable σ_v. The non-execution probability is also in this figure. σ_e*=100*, σ_v *=1*, *k=1* and *c=0.01*

Conversely, the Dönges and Heinemann model predicts that the crossing network has a smaller market share for stocks with higher volatility. In their model, the disutility of unexecuted orders is higher for stocks with higher volatility. Therefore, traders prefer to trade higher volatility stocks in the exchange because of the guaranteed execution. The Dönges and Heinemann model, however, assumes a fixed trading cost in the exchange, whereas my model has an endogenous trading cost in the exchange. An increase in σ_v leads to an increase of transaction costs both in the exchange and in the crossing network, but during these increases, the crossing network has a comparative advantage over the exchange. Empirically, Ready (2009) finds that crossing networks have a higher market share for stocks with higher volatility, which supports theoretical prediction made herein.

2.7 Alternative Allocation Rules

The analyses in the previous four sections is based on the informed-first rule. This rule can be understood as the crossing network having a time priority rule, and that the informed trader, who monitors the market more frequently and has better order submission technology, reacts to the market faster than do the liquidity traders. The assumption of an informed-first rule gives an upper limit to the informed trader's maximum impact in the crossing network. As shown in the section on the non-execution probability, the informed-first rule leads to a low execution probability and this problem cannot be mitigated by increasing the number of liquidity traders participating in the crossing network. In reality, crossing networks tend to minimize the impact of informed traders by using a variety of anti-gaming strategies. Of particular importance are strategies that exclude potentially informed traders from the market, or at least give such traders lower priority. One way to do so is to restrict the crossing networks' customers to buy-side traders, especially traders with passive portfolios.[52] Some crossing networks provide "watchdogs" to detect patterns of abuse.[53] Other crossing networks provide credibility rating reports that allow investors to see their counterparties' track records and to opt out of interacting with certain investors.[54]

Let us consider a case in which all liquidity traders manage to trade before the informed trader. Therefore, the liquidity order flow available to the informed trader decreases and can potentially be zero. This provides a lower bound to the informed trader's impact

[52] Chris Heckman, managing director if ITG, said, "More than anything, we pride ourselves on the constituency of POSIT, which is 95% buy-side to buy-side." He also said that "All POSIT participants are effectively passive in nature, searching for natural liquidity." in "Fair game?", The Trade Magazine, April-June 2008.

[53] For example, Liquidnet constantly monitors its system to look for patterns of abuse and to notify members when a trader appears to be gaming. During its seven-year history, the 500-member venue has suspended approximately 100 members. in "Fair game?", The Trade Magazine, April-June 2008.

[54] For example, BIDS trading provides score cards that track the past trading behavior of its users and enable members to filter out counter-parties with suspect behavior. "Big Traders Dive Into Dark Pools" in Business Week, October 3, 2007.

when there is "fair access". Other allocation rules, like the pro rata rule, lie in between the informed-first rule and the liquidity-first rule. Solving the problem for other allocation rules is a formidable task, but at least it is known that the market outcome for other rules should fall between those of the informed-first rule and the liquidity-first rule. A sketch of the proofs of the theorems under the liquidity-first rule can be obtained from the author. These proofs follow exactly from the proofs of all of the theorems under the informed-first rule, except that the liquidity order flow available to the informed trader is smaller.

Two findings emerge from the liquidity-first rule, which are summarized in Figures 2.9 and 2.10. First, the qualitative results for the informed-first rule still hold for the liquidity-first rule, suggesting that the results presented in the previous sections are robust under different allocation rules. Intuitively, the only effect of the liquidity-first rule is to decrease the liquidity trading available to the informed trader. Quantitatively, this is equivalent to a decrease in the value of k. The qualitative results of the model are driven by two mechanisms: the externality of the price impact in the exchange on the crossing network and the choice of the informed trader between price impact and probability impact. Both will hold as long as $k \neq 0$. The left panel of Figure 2.9 shows that the non-execution probability decreases in the level of liquidity trading in the exchange and increases in the fundamental value uncertainty under the liquidity-first rule, both of which are consistent with the prediction of Theorem 2.4. The right panel of Figure 2.9 shows that price impact under liquidity first rule decreases in the level of liquidity trading in the exchange and increases in fundamental value uncertainty, which is also the same as the result established under the informed-first rule. Figure 2.10 shows that price discovery under the liquidity-first rule is less than price discovery without a crossing network. Additionally, price discovery is higher for stocks with higher liquidity trading in the exchange, lower for stocks with higher fundamental value uncertainty, and lower for stocks with higher liquidity trading in the crossing network; these results are the same as those established by Theorem 2.2. Because the impact of the informed trader with respect to other allocation rules lies between the informed-first rule and the liquidity-first rule, it is expected that the qualitative results should be the same for other allocation rules as well.

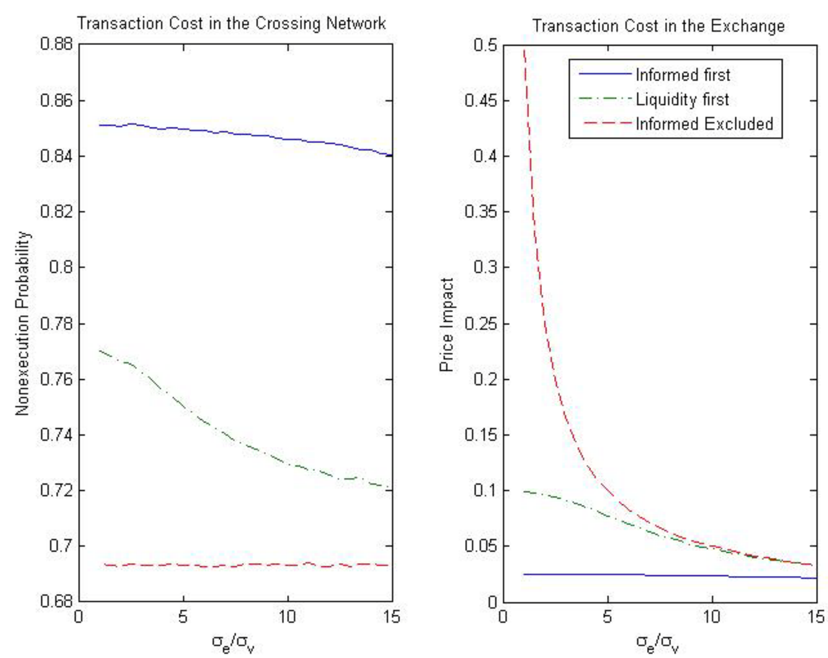

Figure 2.9: This figure illustrates the transaction costs of the crossing network and the exchange under three different allocation rules. The figure shows that non-execution probability is the lowest when the informed trader is excluded from the crossing network. However, the price impact is the highest when the informed trader is excluded from the crossing network. On the contrary, the informed-first rule leads to the lowest price impact in the exchange but the highest non-execution in the crossing network. The liquidity-first rule is in the middle. I choose (k/c)=10 to fit all the three lines in the same figure.

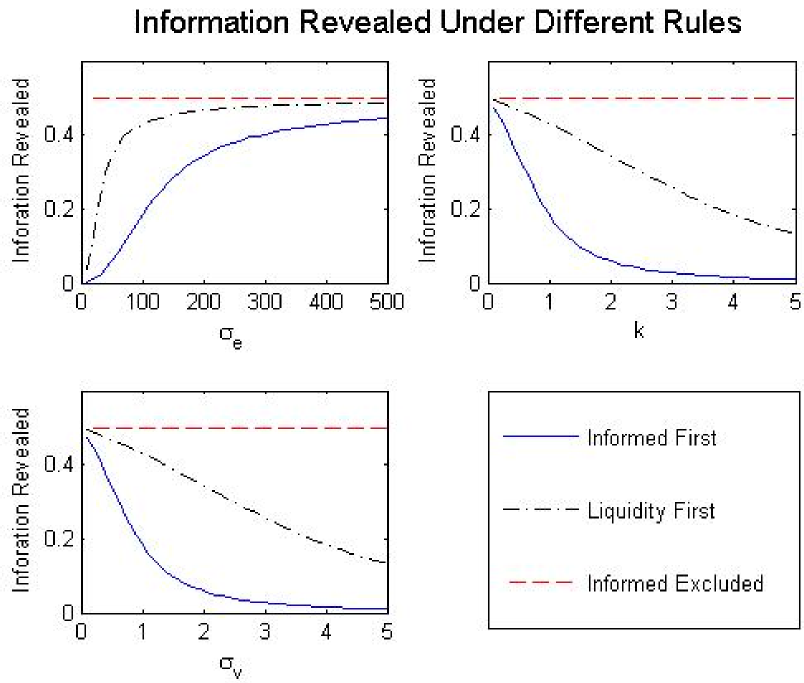

Figure 2.10: This figure demonstrates information revealed through trade under three different allocation rules. σ_e is the proxy for liquidity trading in the exchange; σ_v is the fundamental value uncertainty; and k is the proxy for liquidity trading in the crossing network. For all three panels, $\sigma_e = 100$, $\sigma_v = 1$, $k=1$ and $c=0.01$ unless otherwise specified.

The second finding is that the liquidity-first rule leads to quantitatively different results from the informed-first rule. Figure 2.9 demonstrates that the liquidity-first rule leads to a decrease in the non-execution probability and an increase in the price impact compared to those of the informed-first rule. Under the liquidity-first rule, the liquidity order flow available to the informed trader decreases, which causes the informed trader to decrease his order size in the crossing network and increase his order size in the exchange. Compared to the informed-first rule, the adverse selection problem in the exchange increases and the adverse selection problem in the crossing network decreases. Figure 2.10 demonstrates that price discovery is higher under the liquidity-first rule than under the informed-first rule. Under the liquidity-first rule, the informed trader faces a decrease of liquidity order flow in the crossing network. Therefore, he chooses to trade more aggressively in the exchange, hence revealing more information and increasing price discovery as compared to the informed-first rule.

The liquidity-first rule imposes a lower limit on the informed trader's impact when the crossing network is open to all. The crossing network, however, can be even more aggressive to the informed trader. Because crossing networks are not public exchanges, they can select their customers and exclude others from the market, a practice seen as unfair by their opponents. Currently, the SEC is considering regulation to enforce "fair access" to crossing networks.

Figure 2.9 shows that the non-execution probability under the informed-excluded rule is even lower than it is under the liquidity-first rule; the transaction cost in the crossing network is the lowest under the informed-excluded rule. However, the informed-excluded rule corresponds to the highest price impact in the exchange. As the informed trader cannot trade in the crossing network, the price impact in the exchange does not create any externality for his profit in the crossing network. Therefore, the informed trader will trade more aggressively than he does when he has access to the crossing network. In turn, the adverse selection problem in the exchange increases and so does the price impact. The aggressive trading of the informed trader in the exchange, however, leads to more information revelation compared to the informed-first and liquidity-first rules. Figure 2.10 shows that if the crossing network can exclude the informed trader, price discovery is e=(1/2), which is the same as price discovery without the crossing network. When the informed trader is excluded from the crossing network, this model degenerates to the Kyle model. Therefore, price discovery is equal to (1/2) no matter how much liquidity order flow the crossing network attracts or how large the fundamental value uncertainty is.[55] Now that the informed trader's order in the exchange does not create any externality for his profit in the crossing network, he trades more aggressively in the exchange and reveals more information.

What would happen if the crossing network mistakenly excludes some liquidity traders?[56] Theorem 2.2 predicts that price discovery is enhanced by lower liquidity trading in the crossing network. In conclusion, excluding traders always increases price discovery compared to price discovery under "fair access".

Figures 2.9 and 2.10 also show an interesting pattern: the execution probability is positively correlated with price discovery under different allocation rules. The crossing network has incentive to limit the impact of the informed trader with the intention of increasing its execution probability. These anti-gaming strategies, at the same time, also enhance price discovery. Therefore, the strategies to defend the crossing network's self-interests also minimize the negative impact of the crossing network on price discovery. However, the proposed policy change to enforce "fair access" will have two negative effects. First, price discovery is harmed. Second, fair access will increase the information asymmetry problem in the crossing network.

2.8 Conclusion

This paper studies the impact of the crossing network (dark pool) on the public exchange by extending the Grossman and Stiglitz (1980) and the Kyle (1985)

[55] An exception is the extreme case in which all liquidity traders are in the dark pool, which cannot be sustained as an equilibrium.
[56] If all liquidity traders are excluded, the crossing network collapses because no one will trade with the informed trader. Thus, the model again degenerates to the Kyle model.

frameworks to multiple markets, including one without guaranteed execution. A number of testable hypotheses are generated; some of these hypotheses have been verified, but others are yet to be tested.

This paper also sheds some light on a recent policy debate. For example, passionate opponents of crossing networks argue that excluding some traders from crossing networks is unfair. However, if "fair" is defined as giving different classes of traders (liquidity and informed alike) the same terms of trade, public exchanges also lack this kind of fairness. It is well documented that exchanges provide inferior quotes to potentially informed traders (Barclay, Hendershott and McCormick, (2003)). As crossing networks do not have the ability to adjust quotes, they can only rely on the allocation rules to protect themselves from informed traders. One extreme rule is to exclude informed traders altogether.

It was also shown that the crossing network harms price discovery if the informed trader can trade in the crossing network. Interestingly, although the crossing network harms price discovery, it is in the crossing network's best interest to minimize this negative impact. Informed trading in the crossing network causes lower execution probability and decreases the competitiveness of the crossing network. The crossing network has an incentive to change its allocation rules to limit the impact of the informed trader, which simultaneously limits its negative impact on price discovery.

This paper also provides explanations for several surprising empirical findings. The literature on cream-skimming predicts that crossing networks will increase the transaction costs of the exchange, which is rejected empirically by Fong, Madhavan and Swan (2004) and Gresse (2006). This paper shows one possible mechanism to drive that result: the existence of a crossing network creates an externality for the informed trader's order in the exchange, which makes the informed trader trade less in the exchange and thus decreases the adverse selection problem. Also, Dönges and Heinemann (2006) predict that the crossing network should have a lower market share for stocks with higher volatility, but this very intuitive result is rejected by Ready (2009). The prediction made in this paper differs from that of Dönges and Heinemann, because my model endogenizes the trading cost of the exchange and because volatility in this model has information content.

Several other theoretical predictions remain to be tested. This paper emphasizes non-execution and its association with the transaction costs in the exchange, an area with very limited existing literature. An important area for future research is the exploration of the patterns and determinants of non-execution because most traditional measures of transaction costs are irrelevant for crossing networks. The simplest and most natural question to ask is the following: does non-execution, like transaction costs in the exchange, also contain both non-informational and informational factors? Næs and Skjeltorp (2003) find some evidence of information-based non-execution by studying the abnormal cumulative return, but a more direct way to test this hypothesis is to see whether proxies for liquidity externality, such as trading volume, and proxies for information asymmetry, such as probability of informed trading (PIN) (Easley, Kiefer and O'Hara (1996)), both have explanatory power for non-execution. Additionally, one may ask the following question: because the informed trader tends to "arbitrage" transaction costs in different markets, does this arbitrage finally equalize transaction costs of different markets? This would lead the non-execution probability to follow the same patterns of price impact. Although patterns of price impact are well studied (Chan and Lakonishok (1993,1995 and 1997), Keim and Madhavan (1995,1996,1997 and

1998), Madhavan and Cheng (1997) and Stoll (2000)), the empirical research on non-execution probability has just begun. Exploring this new line of research should prove fruitful. Also, the prediction that the crossing network harms price discovery and decreases price volatility can be tested using an event study or by constructing matched stock samples.

Several theoretical extensions can be made from this paper. First, this model can be extended to multiple periods. In reality, crossing networks are "dark" because they provide pre-trade opaqueness. The executed volume, however, needs to be reported to the consolidated tape. Then the market maker will be able to observe some signal in the crossing network and draw a new inference regarding the fundamental value, which may change the prediction of this model. Second, including multiple informed traders in the framework of Holden and Subrahmanyam (1992) may also generate new predictions. The result may also change when the model includes endogenous signal acquisition instead of an exogenous signal. Third, the liquidity traders are passive in the model. It would be interesting to see what happens if they can react to price impact and non-execution probability. Finally, my model rules out the bluffing equilibrium, in which the informed trader trades in the wrong direction so as to mislead the price and then benefits from the mispricing created by matching orders in the crossing network. However, bluffing may be possible under some other circumstances. An analysis of the conditions under which bluffing exists would make a very interesting study.

APPENDIX 2.A

Proof. for Lemma 2.2

The expected volume for any x_d is $E(\widetilde{x_m}|x_d) = \int_0^{x_d} z f_s(z)\, dz + x_d \int_{x_d}^{+\infty} f_s(z) dz = \int_0^{x_d} z * \frac{1}{2} k^{\frac{1}{2}} (z+k)^{-\frac{3}{2}} dz + x_d \int_{x_d}^{+\infty} \frac{1}{2} k^{\frac{1}{2}} (z+k)^{-\frac{3}{2}} dz = k^{\frac{1}{2}} \left[(z+k)^{\frac{1}{2}} + k(z+k)^{-\frac{1}{2}} \right] |_0^{x_d} - x_d k^{\frac{1}{2}} (z+k)^{-\frac{1}{2}} |_{x_d}^{+\infty} = 2k^{\frac{1}{2}} (x_d + k)^{\frac{1}{2}} - 2k$

Therefore

$$E(\widetilde{x_m}|x_d^*) = 2k^{\frac{1}{2}}(x_d^* + k)^{\frac{1}{2}} - 2k = 2\frac{k}{c}\left(v_g - \mu - \lambda x_e\right) - 2k = 2\frac{k}{c}\left(v_g - \mu - \lambda x_e - c\right) = 2\frac{k}{c}(v - \mu - \lambda x_e)$$

$$E(\pi_c^*) = E(\tilde{p})E(\widetilde{x_m}|x_d^*) - x_d^* * c = \left(v_g - \mu - \lambda x_e\right)\frac{2k}{c}(v - \mu - \lambda x_e) - \left(\frac{k}{c^2}\left(v_g - \mu - \lambda x_e\right)^2 - k\right)c = \left(v_g - \mu - \lambda x_e\right)\frac{2k}{c}(v - \mu - \lambda x_e) - \frac{k}{c}\left(\left(v_g - \mu - \lambda x_e\right)^2 - c^2\right) =$$

$$\frac{k}{c}\left(2v_g - 2\mu - 2\lambda x_e\right)(v - \mu - \lambda x_e) - \frac{k}{c}\left(v_g - \mu - \lambda x_e + c\right)\left(v_g - \mu - \lambda x_e - c\right) =$$

$$\frac{k}{c}\left(2v_g - 2\mu - 2\lambda x_e\right)(v - \mu - \lambda x_e) - \frac{k}{c}\left(v_g - \mu - \lambda x_e + c\right)(v - \mu - \lambda x_e) =$$

$$\frac{k}{c}\left(2v_g - 2\mu - 2\lambda x_e - \left(v_g - \mu - \lambda x_e + c\right)\right)(v - \mu - \lambda x_e) = \frac{k}{c}(v - \mu - \lambda x_e)^2$$

Proof. for Lemma 2.3

Define function $y = f(\lambda) = \frac{(1-2\lambda K)}{(2\lambda - 2\lambda^2 K)^2}$

$f(\lambda)$ is continuous on $(-\infty, 0) \cup \left(0, \frac{1}{K}\right) \cup \left(\frac{1}{K}, +\infty\right)$

$$f'(\lambda) = \left(\frac{-2K\left(2\lambda - 2\lambda^2 K\right)^2 - (1-2\lambda K)*2*\left(2\lambda - 2\lambda^2 K\right)(2-4\lambda K)}{(2\lambda - 2\lambda^2 K)^4}\right)$$

The denominator is equal to 0 when $\lambda = 0$ or $\lambda = \frac{1}{K}$, and greater than 0 otherwise, the numerator $=$ $(2\lambda - 2\lambda^2 K)[-2K(2\lambda - 2\lambda^2 K) - (1-2\lambda K) * 2 * (2-4\lambda K)] =$

$(2\lambda - 2\lambda^2 K)(-12\lambda^2 K^2 + 12\lambda K - 4) = 24(\lambda^2 K - \lambda)\left(\lambda^2 K^2 - \lambda K + \frac{1}{3}\right) =$

$24\lambda(\lambda K - 1)\left[\left(\lambda K - \frac{1}{2}\right)^2 + \frac{1}{12}\right]$

So it is easy to see

$$\begin{cases} f'(\lambda) > 0 \; when \; \lambda < 0 \\ f'(\lambda) < 0 \; when \; 0 < \lambda < \left(\frac{1}{K}\right) \\ f'(\lambda) > 0 \; when \; \lambda > \left(\frac{1}{K}\right) \end{cases}$$

In addition, $\lim_{\lambda \to -\infty} f(\lambda) = 0, \lim_{\lambda \to 0^-} f(\lambda) = +\infty,$

$\lim_{\lambda \to 0^+} f(\lambda) = +\infty,\ \lim_{\lambda \to \left(\frac{1}{K}\right)^-} f(\lambda) = -\infty, \lim_{\lambda \to \left(\frac{1}{K}\right)^+} f(\lambda) = -\infty, \lim_{\lambda \to +\infty} f(\lambda) = 0$

and $f\left(\frac{1}{2K}\right) = 0$

So $y = f(\lambda) = R, (R > 0)$ has two real solutions, $\lambda_1^* \in \left(0, \frac{1}{2K}\right)$ and $\lambda_2^* \in\ (-\infty, 0)$

$\beta = \left(\frac{1-2\lambda K}{2\lambda - 2\lambda^2 K}\right) = R(2\lambda - 2\lambda^2 K) = 2\lambda R(1 - \lambda K)$. when $\lambda \in (-\infty, 0)$, $\beta < 0$, when $\lambda \in \left(0, \frac{1}{2K}\right)$, $\beta > 0$

Proof. for Lemma 2.4

Denote $\pi = E(\widetilde{\pi_d} + \widetilde{\pi_e})$ and the second order derivatives of π are $\left(\frac{\partial^2 \pi}{\partial x_e^2}\right) = -2\lambda$, $\left(\frac{\partial^2 \pi}{\partial x_d^2}\right) = -\left(v_g - \mu - \lambda x_e\right) f_s(x_d^*)$, and $\left(\frac{\partial^2 \pi}{\partial x_d \partial x_e}\right) = -\lambda \int_{x_d}^{+\infty} f_s(z)\, dz$. So the Hessian of π is

$$\begin{vmatrix} -2\lambda & -\lambda \int_{x_d}^{+\infty} f_s(z)\, dz \\ -\lambda \int_{x_d}^{+\infty} f_s(z) dz & -\left(v_g - \mu - \lambda x_e\right) f_s(x_d^*) \end{vmatrix}$$

For $\lambda_2^* \in (-\infty, 0)$, the first-order principle minor is positive. Therefore, the Hessian matrix with λ_2^* can not be negative semidefinite and the necessary condition for profit maximization is violated.

When $\lambda \in \left(0, \frac{1}{2K}\right)$, the first-order principle minor is negative. Now I need to show that the second order principle minor is positive, that is:

$$-2\lambda\,(-(v_g - \mu - \lambda x_e)f_s(x_d^*)) - \left(-\lambda \int_{x_d}^{+\infty} f_s(z)dz\right)^2 > 0$$

Combining (2.3) and (2.11) yields

$$f_s(x_d^*) = \frac{1}{2}k^{\frac{1}{2}}\left(\frac{k}{c^2}\left(v_g - \mu - \lambda x_e\right)^2 - k + k\right)^{-\frac{3}{2}} = \frac{c^3}{2k}(v_g - \mu - \lambda x_e)^{-3}$$

So $-2\lambda\,(-(v_g - \mu - \lambda x_e)f_s(x_d^*)) - (-\lambda \int_{x_d}^{+\infty} f_s(z)dz)^2 = -2\lambda\,(-(v_g - \mu - \lambda x_e)\frac{c^3}{2k}\left(v_g - \mu - \lambda x_e\right)^{-3} - \left(-\lambda\left(\frac{c}{(v_g-\mu-\lambda x_e)}\right)\right)^2 = \lambda\frac{c^3}{k}\left(v_g - \mu - \lambda x_e\right)^{-2} - \lambda^2 c^2\left(v_g - \mu - \lambda x_e\right)^{-2} = \lambda\frac{c^3}{k}\left(v_g - \mu - \lambda x_e\right)^{-2}\left(1 - \left(\frac{k}{c}\right)\lambda\right)$

Because $\lambda \in \left(0, \frac{1}{2K}\right)$ and $\frac{k}{c} \equiv K$, $\left(1 - \frac{k}{c}\lambda\right) > 0$. Also $\lambda > 0$ and $\frac{c^3}{k}\left(v_g - \mu - \lambda x_e\right)^{-2} > 0$. Therefore, the second principle minor for $\lambda_1^* \in \left(0, \frac{1}{2K}\right)$ is greater than 0. So Hessian for λ_1^* is negative definite, which is the sufficient condition for profit maximization.

Proof. for Lemma 2.5

The proof of Lemma 2.5 is divided by three parts. First, I will prove that both β^* and x_d^* are positive unless $v = p_0$. Then I will show the comparative statics of λ^* and β^*. Finally, I will show the comparative statics of x_d^*.

$\beta^* > 0$ follows from Lemma 2.3. For x_d^*, I focus the proof on the case when $v > p_0$, and the case for $v < p_0$is simply symmetric.

When $v > p_0$, $x_d^* = \frac{k}{c^2}(v + c - p_0 - \lambda^* x_e)^2 - k = \frac{k}{c^2}(v - p_0 - \lambda^*\beta^*(v - p_0) + c)^2 - k = \frac{k}{c^2}((v - p_0)(1 - \lambda^*\beta^*) + c)^2 - k$.

From equation (2.17), we know that $\lambda^*\beta^* = \frac{\beta_*^2}{\beta_*^2 + R} < 1$. Therefore, $(v - p_0)(1 - \lambda^*\beta^*) + c > c$ when $v > p_0$. So

$$x_d^* = \frac{k}{c^2}\left((v - p_0)(1 - \lambda^*\beta^*) + c\right)^2 - k > \frac{k}{c^2}(c)^2 - k = 0$$

The comparative statics of λ^* and β^* follow from the implicit function rule. Equations (2.14) and (2.21) defines two implicit functions $\lambda^*(R, K)$ and $\beta^*(R, K)$, where $R = \frac{\sigma_e^2}{\sigma_v^2}$and $K = \frac{k}{c}$. Fix c, the sign of $\frac{\partial\lambda^*}{\partial K}$ is the same as $\frac{\partial\lambda^*}{\partial k}$, and the sign of $\frac{\partial\beta^*}{\partial K}$ is the same as $\frac{\partial\beta^*}{\partial k}$.

Denote

$$\begin{cases} F^1(\lambda^*, \beta^*; R, K) = (2\lambda^* - 2\lambda^{*2}K)\beta^* - (1 - 2\lambda^*K) = 0 \\ F^2(\lambda^*, \beta^*; R, K) = R(2\lambda^* - 2\lambda^{*2}K)^2 - (1 - 2\lambda^*K) = 0 \end{cases}$$

Take total derivatives with respect to R and K I get

$$\begin{bmatrix} (2\lambda^* - 2\lambda^{*2}K) & (2 - 4\lambda^*K)\beta + 2K \\ 0 & 2R(2\lambda^* - 2\lambda^{*2}K)(2 - 4\lambda^*K) + 2K \end{bmatrix} \begin{bmatrix} \frac{\partial\beta^*}{\partial R} & \frac{\partial\beta^*}{\partial K} \\ \frac{\partial\lambda^*}{\partial R} & \frac{\partial\lambda^*}{\partial K} \end{bmatrix} =$$

$$\begin{bmatrix} 0 & 2\lambda^{*2}\beta - 2\lambda^* \\ -(2\lambda^* - 2\lambda^{*2}K)^2 & 2R(2\lambda^* - 2\lambda^{*2}K)2\lambda^{*2} - 2\lambda^* \end{bmatrix}$$

Therefore, $\frac{\partial\lambda^*}{\partial R} = \frac{\left(2\lambda^* - 2\lambda^{*2}K\right)^2}{2R(4\lambda^*K - 2)(2\lambda^* - 2\lambda^{*2}K) - 2K}$.

For $K > 0$ $R > 0$ and $\lambda^* \in \left(0, \frac{1}{2K}\right) \Rightarrow \begin{cases} 2R(4K\lambda^* - 2) < 0 \\ 2\lambda^* - 2\lambda^{*2}K > 0 \\ -2K < 0 \end{cases}$

So $2R(4K\lambda^* - 2)(2\lambda^* - 2\lambda^{*2}K) - 2K < 0 \Rightarrow \frac{\partial\lambda^*}{\partial R} < 0$

$\frac{\partial\beta^*}{\partial R}=\frac{(4\lambda^*K-2)\beta-2K}{(2\lambda^*-2\lambda^{*2}K)}\frac{\partial\lambda^*}{\partial R}$, where $(4\lambda^*K-2)\beta-2K=(4\lambda^*K-2)\frac{1-2\lambda^*K}{2\lambda^*-2\lambda^{*2}K}-2K=\frac{-(1-2\lambda^*K)^2}{2\lambda^*-2\lambda^{*2}K}-2K.$

For $K>0,\lambda^*\in\left(0,\frac{1}{2K}\right),2\lambda^*-2\lambda^{*2}K>0\Rightarrow\frac{-(1-2\lambda^*K)^2}{2\lambda^*-2\lambda^{*2}K}-2K<0\ \Rightarrow\frac{(4\lambda^*K-2)\beta-2K}{(2\lambda^*-2\lambda^{*2}K)}<0$

So $\frac{\partial\beta^*}{\partial R}=\frac{(4\lambda^*K-2)\beta-2K}{(2\lambda^*-2\lambda^{*2}K)}\frac{\partial\lambda^*}{\partial R}>0$

$\frac{\partial\lambda^*}{\partial K}=\frac{2R\left(2\lambda^*-2\lambda^{*2}K\right)(2\lambda^{*2}-2\lambda^*)}{2R(2\lambda^*-2\lambda^{*2}K)(2-4\lambda^*K)+2K}$

from equation (2.21) I know that $R=\frac{(1-2\lambda^*K)}{(2\lambda^*-2\lambda^{*2}K)^2}$

$\Rightarrow\frac{2R\left(2\lambda^*-2\lambda^{*2}K\right)2\lambda^2-2\lambda}{2R(2\lambda^*-2\lambda^{*2}K)(2-4\lambda^*K)+2K}=\frac{-\lambda^{*3}K}{3\lambda^{*2}K^2-3\lambda^*K+1}=-\frac{\lambda^{*3}K}{3\left[\left(\lambda^*K-\frac{1}{2}\right)^2+\left(\frac{1}{12}\right)\right]}<0$ when $K>0$ and $\lambda^*\in(0,\frac{1}{2K})$

$\frac{\partial\beta^*}{\partial K}=\frac{2\lambda^{*2}\beta^*-2\lambda^*+[(4\lambda^*K-2)\beta^*-2K]\left(\frac{\partial\lambda^*}{\partial K}\right)}{(2\lambda^*-2\lambda^{*2}K)}$ where

$2\lambda^{*2}\beta^*-2\lambda^*=2\lambda^{*2}\frac{1-2\lambda^*K}{(2\lambda^*-2\lambda^{*2}K)}-2\lambda^*=\frac{-2\lambda^{*2}}{(2\lambda^*-2\lambda^{*2}K)}$

$[(4\lambda^*K-2)\beta^*-2K]\frac{\partial\lambda^*}{\partial K}=\left[(4\lambda^*K-2)\left(\frac{1-2\lambda^*K}{(2\lambda^*-2\lambda^{*2}K)}\right)-2K\right]\frac{\partial\lambda^*}{\partial K}=\left[(4\lambda^*K-2)\left(\frac{1-2\lambda^*K}{(2\lambda^*-2\lambda^{*2}K)}\right)-2K\right]\frac{\partial\lambda^*}{\partial K}=\frac{-4\lambda^{*2}K^2+4\lambda^*K-2}{2\lambda^*-2\lambda^{*2}K}*\frac{-\lambda^{*3}K}{3\lambda^{*2}K^2-3\lambda^*K+1}=\frac{4\lambda^{*5}K^3-4\lambda^{*4}K^2+2\lambda^{*3}K}{(2\lambda^*-2\lambda^{*2}K)(3\lambda^{*2}K^2-3\lambda^*K+1)}$

$\frac{\partial\beta^*}{\partial K}=\frac{2\lambda^{*2}\beta^*-2\lambda^*+[(4\lambda^*K-2)\beta^*-2K]\left(\frac{\partial\lambda^*}{\partial K}\right)}{(2\lambda^*-2\lambda^{*2}K)}=\frac{-2\lambda^{*2}}{(2\lambda^*-2\lambda^{*2}K)^2}+\frac{4\lambda^{*5}K^3-4\lambda^{*4}K^2+2\lambda^{*3}K}{(2\lambda^*-2\lambda^{*2}K)^2(3\lambda^{*2}K^2-3\lambda^*K+1)}=\frac{4\lambda^{*5}K^3-10\lambda^{*4}K^2+8\lambda^{*3}K-2\lambda^{*2}}{(2\lambda^*-2\lambda^{*2}K)^2(3\lambda^{*2}K^2-3\lambda^*K+1)}=\frac{2\lambda^*K-1}{2(3\lambda^{*2}K^2-3\lambda^*K+1)}<0$ when $\lambda^*\in\left(0,\frac{1}{2K}\right)$

Finally, we need to prove that $\frac{\partial x_d^*}{\partial R}$, we focus on the case when $v > p_0$. The case when $v < p_0$ can be proved symmetrically

$\frac{\partial x_d^*}{\partial R} = \frac{k}{c^2} 2[(1-\lambda^*\beta^*)(v-p_0)+c][(v-p_0)(-\lambda^*\frac{\partial \beta^*}{\partial R} - \beta^*\frac{\partial \lambda^*}{\partial R})]$. As $\lambda^*\beta^* = \frac{\beta_*^2}{\beta^{*2}+R} < 1$ and $v - p_0 > 1$, the sign of $\frac{\partial x_d^*}{\partial R}$ is determined by $\left(-\lambda^*\frac{\partial \beta^*}{\partial R} - \beta^*\frac{\partial \lambda^*}{\partial R}\right)$

We know that $\frac{\partial \beta^*}{\partial R} = \frac{(4\lambda^*K-2)\beta - 2K}{(2\lambda^* - 2\lambda^{*2}K)}\frac{\partial \lambda^*}{\partial R}$. Therefore, $\left(-\lambda^*\left(\frac{\partial \beta^*}{\partial R}\right) - \beta^*\left(\frac{\partial \lambda^*}{\partial R}\right)\right) = \left(-\lambda^*\left(\frac{(4\lambda^*K-2)\beta-2K}{(2\lambda^*-2\lambda^{*2}K)}\right) - \beta^*\right)\frac{\partial \lambda^*}{\partial R} = \frac{2K(1-\lambda^*\beta^*)}{2-2\lambda^*K}\frac{\partial \lambda^*}{\partial R}$

Because $\lambda^*\beta^* = \frac{\beta_*^2}{\beta^{*2}+R} < 1, \lambda^* \in \left(0, \frac{1}{2K}\right)$ and $\frac{\partial \lambda^*}{\partial R} < 0$, $\frac{\partial x_d^*}{\partial R} < 0$

To know the sign of $\frac{\partial x_d^*}{\partial k}$, we only need to know the sign of $\frac{\partial x_d^*}{\partial K}$ when c is fixed

Note that $x_d^* = \frac{k}{c^2}\left((1-\lambda^*\beta^*)(v-p_0)+c\right)^2 - k = \frac{K}{c}(v+c-p_0-\lambda x_e)^2 - Kc$

So $\frac{\partial x_d^*}{\partial K} = \frac{1}{c}\left((1-\lambda^*\beta^*)(v-p^0)+c\right)^2 - c + \left[\frac{K}{c}2\left((1-\lambda^*\beta^*)(v-p^0)+c\right)\left(-\frac{\partial \lambda^*}{\partial K}\beta^* - \lambda^*\frac{\partial \beta^*}{\partial K}\right)\right]$

Note that $\frac{1}{c}((1-\lambda^*\beta^*)(v-p_0)+c)^2 - c = \frac{1}{c}[((1-\lambda^*\beta^*)(v-p_0)+c)^2 - c^2] = \frac{1}{c}[((1-\lambda^*\beta^*)(v-p_0))((1-\lambda^*\beta^*)(v-p_0)+2c)] > 0$

$\frac{K}{c}2((1-\lambda^*\beta^*)(v-p_0)+c)(-\frac{\partial \lambda^*}{\partial K}\beta^* - \lambda^*\frac{\partial \beta^*}{\partial K}) > 0$ because $\left((1-\lambda^*\beta^*)(v-p_0)+c\right) > 0$, $\frac{\partial \lambda^*}{\partial K} < 0$, $\frac{\partial \beta^*}{\partial K} < 0$, $\beta^* > 0$ and $\lambda^* > 0$. Therefore, $\frac{\partial x_d^*}{\partial K} > 0$

Proof. for Lemma 2.6

Because $\tilde{p} = \widetilde{p_0} + \lambda\tilde{y}$, $\tilde{p}$ is informationally equivalent to $\tilde{y}$, implying $var(\tilde{v}|\tilde{p}) = var(\tilde{v}|\tilde{y})$. Because $\tilde{p} = E(\tilde{v}|\tilde{y})$, $var(\tilde{p}) = var(E(\tilde{v}|\tilde{y}))$. $var(\tilde{v}) = var(\tilde{v}|\tilde{y}) +$

$var(E(\tilde{v}|\tilde{y}))$ follows directly from the Projection Theorem, where the variance of $\tilde{v}$ is decomposed into two parts: the part that can be explained by $\tilde{y}$, $var(E(\tilde{v}|\tilde{y}))$, and the part that can not be explained by $\tilde{y}$, $var(\tilde{v}|\tilde{y})$. Therefore, $var(\tilde{v}) = var(\tilde{v}|\tilde{p}) + var(\tilde{p})$ and $var(\tilde{p}) = e\sigma_v^2$ follows (2.24)

Proof. for Theorem 2.2

From (2.26) e is uniquely determined by $2 - 2\lambda^* K$. Denote $2 - 2\lambda^* K \equiv l$, then $\lambda^* = \frac{2-l}{2K}$. So

$$R(2\lambda^* - 2\lambda^{*2}K)^2 = (1 - 2\lambda^* K) \Rightarrow R\lambda^{*2}(2 - 2\lambda^* K)^2 = (1 - 2\lambda^* K)$$

$$\Rightarrow R\left(\frac{2-l}{2K}\right)^2 l^2 = l - 1 \Rightarrow \frac{R}{4K^2}(2-l)^2 l^2 = l - 1 \Rightarrow \frac{1}{4}\frac{\sigma_e^2 c^2}{k^2\sigma_v^2}(2-l)^2 l^2 = l - 1$$

The last step shows that the solution of l only depends on $\frac{R}{K^2} = \left(\frac{\sigma_e c}{k\sigma_v}\right)^2$. There are multiple solution for the equation, but as $2 - 2\lambda^* K \equiv l$ and that λ^* is unique. There should be only one solution satisfies optimization conditions set up the model. Therefore, l is uniquely determined by $\frac{R}{K^2} = \left(\frac{\sigma_e c}{k\sigma_v}\right)^2$

From equation (2.26), $e = \frac{1}{2}$ when $k = 0$.

When $k \neq 0$, we know that l is uniquely determined by $\frac{\sigma_e c}{k\sigma_v}$. Denote $\frac{\sigma_e c}{k\sigma_v} = N$. Then $\frac{1}{4}\frac{\sigma_e^2 c^2}{k^2\sigma_v^2}(2-l)^2 l^2 = l - 1 \Leftrightarrow \frac{1}{4}N^2 = \frac{l-1}{(2-l)^2 l^2}$

Totally differentiate both side of the equation results in $\frac{1}{2}NdN = \frac{(2-l)l(3l^2-6l+4)}{(2-l)^4 l^4}dl \Rightarrow \frac{dl}{dN} = \frac{1(2-l)^3 l^3 N}{2(3l^2-6l+4)}$.

$l \equiv 2 - 2\lambda^* K$, $2 - l = \frac{\lambda^*}{2K}$ and $\lambda^* \in \left(0, \frac{1}{2K}\right) \Rightarrow (2-l)^3 > 0$ and $l^3 > 0$. Also $N > 0$ and $(3l^2 - 6l + 4) = 3(l-1)^2 + 1 > 0$. So $\frac{dl}{dN} > 0$ and l is increasing in N. It is easy to see from (2.26) that e increases in l. So e increases in $N \equiv \frac{\sigma_e c}{k\sigma_v}$.

Proof. for Theorem 2.3

From Lemma 2.3, we know that $\lambda^* \in \left(0, \frac{1}{2K}\right)$ so $\lambda^* < \frac{c}{2k}$.

The relationship between λ^* and exogenous variables is proved in Lemma 2.5.

Proof. for Theorem 2.4

We prove for the case where $v \geq p_0$. From Lemma 2.2 and $\mu = p_0$

$\frac{E(\widetilde{x_m}|x_d^*)}{x_d^*} = \frac{2\frac{k}{c}(v-p_0-\lambda^* x_e^*)}{\frac{k}{c^2}(v+c-p_0-\lambda^* x_e^*)^2 - k} = \frac{2c(v-p_0-\lambda^* x_e^*)}{(v+c-p_0-\lambda^* x_e^*)^2 - c^2} = \frac{2c(v-p_0-\lambda^* x_e^*)}{(v+c-p_0-\lambda^* x_e^* + c)(v+c-p_0-\lambda^* x_e^* - c)} =$

$\frac{2c}{2c+v-p_0-\lambda^* x_e^*}$

For $x_e^* = \beta^*(v-p_0)$, $\frac{E(\widetilde{x_m}|x_d^*)}{x_d^*} = \frac{2c}{2c+(v-p_0)(1-\lambda^*\beta^*)}$

From (2.14), (2.26) and Lemma 2.6. $\lambda^*\beta^* = \lambda^* \frac{1-2\lambda^* K}{2\lambda^* - 2\lambda^{*2} K} = \frac{1-2\lambda^* K}{2-2\lambda^* K} = e = \frac{var(\tilde{p})}{\sigma_v^2}$

It is straightforward to see $v \uparrow \Rightarrow\ 2c + (v-p_0)(1-\lambda^*\beta^*) \uparrow \Rightarrow \frac{2c}{2c+(v-p_0)(1-\lambda^*\beta^*)} \downarrow$.

The comparative statics for $\frac{\sigma_e}{k\sigma_v}$ follows Theorem 2.2. $\frac{\sigma_e}{k\sigma_v} \uparrow \Rightarrow\ e \uparrow \Rightarrow 2c + (v-p_0)(1-e) \downarrow \Rightarrow \frac{2c}{2c+(v-p_0)(1-\lambda^*\beta^*)} \uparrow$.

To find the relationship between c and exe_i, notice that $\frac{1}{exe_i} = \frac{2c+(v-p_0)(1-\lambda^*\beta^*)}{2c} = 1 + \frac{(v-p_0)(1-\lambda^*\beta^*)}{2c} = 1 + \frac{(v-p_0)}{2c(2-2\lambda^* K)} = 1 + \frac{(v-p_0)}{4(c-\lambda^* k)}$

Now we prove that $(c-\lambda^* k)$ increases in c: $\frac{\partial(c-\lambda^* k)}{\partial c} = 1 - k\frac{\partial \lambda^*}{\partial c} = 1 - k\frac{\partial \lambda^*}{\partial K}\frac{\partial K}{\partial c}$

For $K = \frac{k}{c}$, $\frac{\partial K}{\partial c} = -\frac{k}{c^2}$, from proof of Lemma 2.5 we know $\frac{\partial \lambda^*}{\partial K} = \frac{-\lambda^{*3} K}{3\lambda^{*2}K^2 - 3\lambda^* K + 1} \Rightarrow$

$\frac{\partial(c-\lambda^* k)}{\partial c} = 1 - k\frac{-\lambda^{*3}K}{3\lambda^{*2}K^2-3\lambda^*K+1}\left(-\frac{k}{c^2}\right) = 1 - \frac{\lambda^{*3}K\left(\frac{k}{c^2}\right)k}{3\lambda^{*2}K^2-3\lambda^*K+1} = 1 - \frac{\lambda^{*3}K^3}{3\lambda^{*2}K^2-3\lambda^*K+1} =$

$\frac{-\lambda^{*3}K^3+3\lambda^{*2}K^2-3\lambda^*K+1}{3\lambda^{*2}K^2-3\lambda^*K+1} = \frac{(1-\lambda K)^3}{3\left[\left(\lambda^*K-\left(\frac{1}{2}\right)\right)^2+\left(\frac{1}{12}\right)\right]}$

For $\lambda^* \in \left(0, \frac{1}{2K}\right)$, $(1-\lambda K)^3 > 0$ and $\frac{\partial(c-\lambda^* k)}{\partial c} > 0$. So $c - \lambda^* k$ increases in c and $\frac{1}{exe_i} = 1 + \frac{(v-p_0)}{4(c-\lambda^* k)}$ decreases in c, and exe_i is increasing in c.

Proof. for Theorem 2.5

$\frac{E(\widetilde{x_m}|x_d^*)}{x_e^*} = \frac{2\frac{k}{c}(v-p_0-\lambda^* x_e^*)}{x_e^*} = \frac{2K\left(v-p_0-\lambda^*\beta^*(v-p_0)\right)}{\beta^*(v-p_0)} = \frac{2K(1-\lambda^*\beta^*)(v-p^0)}{\beta^*(v-p_0)} = \frac{2K(1-\lambda^*\beta^*)}{\beta^*}$ plug (2.14) into the equation I obtain

$$\frac{E(\widetilde{x_m}|x_d^*)}{x_e^*} = \frac{2K\left(1-\lambda^*\left(\frac{1-2\lambda^*K}{2\lambda^*-2\lambda^{*2}K}\right)\right)}{\left(\frac{1-2\lambda^*K}{2\lambda^*-2\lambda^{*2}K}\right)} = \frac{2K(\lambda^*)}{1-2\lambda^*K} = \frac{1}{1-2\lambda^*K} - 1$$

From the proof of Theorem 2.2 we know that $l \equiv 2 - 2\lambda^* K$ is uniquely determined by $\frac{\sigma_e c}{k\sigma_v}$ and increases in $\frac{\sigma_e c}{k\sigma_v}$. Therefore $l - 1 \equiv 1 - 2\lambda^* K$ is also uniquely determined by $\frac{\sigma_e c}{k\sigma_v}$ and increases in $\frac{\sigma_e c}{k\sigma_v}$. So $\frac{E(\widetilde{x_m}|x_d^*)}{x_e^*} = \frac{1}{1-2\lambda^*K} - 1$ is uniquely determined by $\frac{\sigma_e c}{k\sigma_v}$ and decreases in $\frac{\sigma_e c}{k\sigma_v}$.

Proof. for Corollary 2.1

$var(\tilde{p}) = var(p_0 + \lambda^*\tilde{y}) = \lambda^{*2}var(\tilde{y}) = \lambda^{*2}var(\beta^*\widetilde{x_e} + \widetilde{u_e}) = \lambda^{*2}(\beta^{*2}\sigma_v^2 + \sigma_e^2)$

From (2.14) $\beta^* = \frac{1-2\lambda^* K}{2\lambda^* - 2\lambda^{*2} K}$

$\lambda^{*2}(\beta^{*2}\sigma_v^2 + \sigma_e^2) = \lambda^{*2}(\left(\frac{1-2\lambda^* K}{2\lambda^* - 2\lambda^{*2} K}\right)^2 \sigma_v^2 + \sigma_e^2)$

from (2.21) we know $\left(\frac{1-2\lambda^* K}{2\lambda^* - 2\lambda^{*2} K}\right)^2 = R(1 - 2\lambda^* K)$, and $R = \frac{\sigma_e^2}{\sigma_v^2}$ so

$\lambda^{*2}(\left(\frac{1-2\lambda^* K}{2\lambda^* - 2\lambda^{*2} K}\right)^2 \sigma_v^2 + \sigma_e^2) = \lambda^{*2}(R(1 - 2\lambda^* K)\sigma_v^2 + \sigma_e^2) = \lambda^{*2}((1 - 2\lambda^* K)\sigma_e^2 + \sigma_e^2) = \lambda^{*2}(2 - 2\lambda^* K)\sigma_e^2$

Denote $f(\lambda^*; \sigma_e^2, K) = \lambda^{*2}(2 - 2\lambda^* K)\sigma_e^2$ so $var(\tilde{p}) = f(\lambda^*; \sigma_e^2, K)$

$\frac{\partial f}{\partial \lambda} = (4\lambda^* - 6\lambda^{*2} K)\sigma_e^2 = 4\lambda^* \left(1 - \frac{3}{2}\lambda^* K\right) \sigma_e^2$

Therefore $\frac{\partial f}{\partial \lambda} > 0$ when $\lambda^* \in \left(0, \frac{2}{3K}\right)$

From Lemma 2.3 and Lemma 2.4, $\lambda^* \in \left(0, \frac{1}{2K}\right)$. Therefore,

$f(\lambda^*; \sigma_e^2, K) < f(\frac{1}{2K}; \sigma_e^2, K) = \left(\frac{1}{2K}\right)^2 \left(2 - 2\frac{1}{2K}K\right) \sigma_e^2 = \frac{\sigma_e^2}{4K^2} = \frac{c^2\sigma_e^2}{4k^2}$.

APPENDIX 2.B

This part provides another way to understand why the informed trader increases order size more than the liquidity order flow.

Theorem 2.6 *Suppose* $\{\lambda^*(k_1), x_e^*(k_1), x_d^*(k_1)\}$ *and* $\{\lambda^*(k_2), x_e^*(k_2), x_d^*(k_2)\}$ *are optimal solutions for* $k_1 < k_2$*. An increase from* k_1 *to* k_2 *can be decomposed into the following two effects:*

Volume effect: suppose that the informed trader's strategy in the exchange is fixed as $x_e^*(k_1)$*, then the price impact remains as* $\lambda^*(k_1)$ *and the expected profit per matched unit does not change. Denote* ${x_d}'(k_2)$ *as the optimal choice of the informed trader in the crossing network conditional on* $\lambda^*(k_1)$ *and* $x_e^*(k_1)$ *. Then* $\frac{E\left(\widetilde{x_m}\middle|x_d'(k^2)\right)}{x_d'(k^2)} = \frac{E\left(\widetilde{x_m}\middle|x_d^*(k_1)\right)}{x_d^*(k_1)}$*, meaning that the execution probability does not change.*

Price effect: $v_g - p_0 - \lambda^*(k_1)x_e^*(k_1) < v_g - p_0 - \lambda^*(k_2)x_e^*(k_2)$ *, meaning that the optimal profit per matched share is higher with* k_2*, which implies* $\frac{E\left(\widetilde{x_m}\middle|x_d'(k_2)\right)}{x_d'(k_2)} > \frac{E\left(\widetilde{x_m}\middle|x_d^*(k_2)\right)}{x_d^*(k_2)}$.

Combining the volume effect and price effect results in $\frac{E\left(\widetilde{x_m}\middle|x_d^*(k_1)\right)}{x_d^*(k_1)} > \frac{E\left(\widetilde{x_m}\middle|x_d^*(k_2)\right)}{x_d^*(k_2)}$.

Proof. *Volume effect: Suppose we fix* ${x_e}'(k_2) = x_e^*(k_1)$ *. Because* $x_e^* = \beta^*(v - p_0)$*,* $x_e'(k_2) = x_e^*(k_1)$ *is equivalent to* $\beta'(k_2) = \beta^*(k_1)$*. From (2.17) we know* $\lambda'(k_2) =$

$\lambda^*(k_1)$ when both β^* and R are fixed. So $v_g - p_0 - \lambda'(k^2){x_e}'(k_2) = v_g - p_0 - \lambda^*(k^1)x_e^*(k_1)$.

Denote $v_g - p_0 - \lambda^*(k_1)x_e^*(k_1) = \pi_1$. First-order condition (2.11) and Lemma 2.2 implies that

$x_d^*(k_1) = \frac{k_1}{c^2}\pi_1^2 - k_1$ *and* $E(\widetilde{x_m}|x_d^*(k_1)) = 2\frac{k_1}{c}(\pi_1 - c)$

$x_d'(k_2) = \frac{k_2}{c^2}\pi_1^2 - k_2$ *and* $E(\widetilde{x_m}|x_d'(k_2)) = 2\frac{k_2}{c}(\pi_1 - c)$

Therefore $\left(\frac{E\left(\widetilde{x_m}\middle|x_d'(k_2)\right)}{x_d'(k_2)}\right) = \left(\frac{E\left(\widetilde{x_m}\middle|x_d^*(k_1)\right)}{x_d^*(k_1)}\right) = \frac{2\frac{1}{c}\pi_1}{\frac{1}{c^2}\pi_1^2 - 1} = \frac{2c}{\pi_1 + c}$. So execution probability does not change.

Price effect: Denote $\pi_2 = v_g - p_0 - \lambda^*(k_2)x_e^*(k_2) = v_g - p_0 - \lambda^*(k_2)\beta^*(k_2)(v - p_0)$.

Theorem 2.3 implies that $\lambda^*(k_2) < \lambda^*(k_1)$ and $\beta^*(k_2) < \lambda^*(k_2)$ when $k_2 > k_1$. So $\pi_2 > \pi_1$.

First-order condition (2.11) and Lemma 2.2 implies that

$x_d^*(k_2) = \frac{k_2}{c^2}\pi_2^2 - k_2$ *and* $E(\widetilde{x_m}|x_d^*(k_2)) = 2(\frac{k_2}{c})(\pi_2 - c)$

So $\frac{E\left(\widetilde{x_m}\middle|x_d^*(k_2)\right)}{x_d^*(k_2)} = \frac{2\frac{k_2}{c}(\pi_2 - c)}{\frac{k_2}{c^2}\pi_2^2 - k_2} = \frac{2c}{\pi_2 + c} < \frac{2c}{\pi_1 + c} = \frac{E\left(\widetilde{x_m}\middle|x_d'(k_2)\right)}{x_d'(k_2)}$

Combining volume effect and price effect we get $\frac{E\left(\widetilde{x_m}\middle|x_d^*(k_2)\right)}{x_d^*(k_2)} < \frac{E\left(\widetilde{x_m}\middle|x_d'(k_2)\right)}{x_d'(k_2)} = \frac{E\left(\widetilde{x_m}\middle|x_d^*(k_1)\right)}{x_d^*(k_1)}$*. So an increase of liquidity trading in the crossing network decreases execution probability.*

The volume effect implies that if the profit per matched unit does not change, the informed trader increases his order size at the same ratio as the increase in liquidity trading. Then the probability of execution remains the same. This can be seen from the first-order condition (2.9). The informed trader's optimization problem in the crossing network is to choose the execution probability so that his expected marginal profit, which is the product of profit per matched unit and execution probability, is equal to up-front submission cost c. When profit per matched unit and c are the same, the informed trader will choose the same execution probability, that is, he will increase his order size at the same ratio as the liquidity traders.

What drives the result is the price effect. As the liquidity trading in the crossing network increases, the informed trader finds that the externality of price impact on the crossing network increases. Therefore, he chooses to trade less in the exchange and more in the crossing network. His smaller order size in the exchange decreases the adverse selection problem in the exchange and causes a smaller price impact of trade. Therefore, his profit per matched unit increases. When the profit per matched unit increases, the informed trader requires a lower execution probability to make marginal revenue equal to marginal cost in the crossing network.

REFERENCES

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

Barclay, M. J., T. Hendershott, and D. T. McCormick, 2003, Competition among trading venues: Information and trading on electronic communications networks, Journal of Finance 58, 2637-2665.

Baruch, S., and GA Karolyi, M. Lemmon, 2007, Multimarket trading and liquidity: Theory and evidence, Journal of Finance 62, 2169-2200.

Battalio, R., 1997, Third market broker-market makers: Cost competitors or cream skimmers? An empirical analysis, Journal of Finance 52, 341-352.

Bessembinder, H., and H. M. Kaufman, 1997, A cross-exchange comparison of execution costs and information flow for NYSE-listed stocks, Journal of Financial Economics 46, 293-319.

Chan, L. K. C., and J. Lakonishok, 1993, Institutional trades and intraday stock price behavior, Journal of Financial Economics 33, 173-199.

Chan, L. K. C., and J. Lakonishok, 1995, The behavior of stock prices around institutional trades, The Journal of Finance 50, 1147-1174.

Chan, L. K. C., and J. Lakonishok, 1997, Institutional equity trading costs: NYSE versus Nasdaq, The Journal of Finance 52, 713-735.

Chowdhry, B., and V. Nanda, 1991, Multimarket trading and market liquidity, Review of Financial Studies , 483-511.

Conrad, J., K. M. Johnson, and S. Wahal, 2003, Institutional trading and alternative trading systems, Journal of Financial Economics 70, 99-134.

Degryse, H., M. Van Achter, and G. Wuyts, 2008, Dynamic order submission strategies with competition between a market maker market and a dark pool, Journal of Financial Economics .

Degryse, H., M. Van Achter, Mark and G. Wuyts, Shedding Light on Dark Liquidity Pools(November 18, 2008). TILEC Discussion Paper No. 2008-039.

Domowitz, I., I. FInkelshteyn, and H. Yegerman, Cul de Sacs and Highways: An Optical Tour of Dark Pool Trading Performance, The Journal Of Trading 4, 16-22.

Dönges, J., and F. Heinemann, 2006, dark pool versus market maker Market: Unique Equilibrium in the Allocation of Order Flow .

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

Fong, K., A. Madhavan, and P. Swan, 2004, Upstairs, Downstairs: Does the Upstairs Market Hurt the Downstairs .

Foster, D. P., S. Gervais, K. Ramaswamy, and O. T. Drive, The Benefits of Volume-Conditional Order-Crossing, .

Foucault, T., and A. Menkveld, 2008, rCompetition for order flow and smart order routing sys' temss, Journal of Finance 63, 119-158.

Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley, 2006, Institutional Investors and Stock Market Volatility*, Quarterly Journal of Economics 121, 461-504.

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

Gopikrishnan, P., V. Plerou, X. Gabaix, and H. E. Stanley, 2000, Statistical properties of share volume traded in financial markets, PHYSICAL REVIEW-SERIES E- 62, 4493-4496.

Gresse, C., 2006, The Effect of Crossing-Network Trading on market maker Market's Bid-Ask Spreads, European Financial Management 12, 143-160.

Grossman, S. J., and J. E. Stiglitz, 1980, On the impossibility of informationally efficient markets, The American Economic Review 70, 393-408.

Hendershott, T., and H. Mendelson, 2000, Crossing networks and market maker markets: competition and performance, Journal of Finance , 2071-2115.

Hasbrouck, J., and G. Saar, 2009, Technology and liquidity provision: the blurring of traditional definitions, Journal of financial Markets 12, 143-172.

Hasbrouck, J., 2007, Empirical market microstructure: The institutions, economics and econometrics of securities trading (Oxford University Press, USA).

Ho, T., and H. Stoll, 1981, Optimal market maker pricing under transaction and return uncertainty, Journal of Financial Economics 9, 47-73.

Holden, C. W., and A. Subrahmanyam, 1992, Long-lived private information and imperfect competition, Journal of Finance , 247-270.

Keim, D. B., and A. Madhavan, 1995, Anatomy of the trading process empirical evidence on the behavior of institutional traders, Journal of Financial Economics 37, 371-398.

Keim, D. B., and A. Madhavan, 1996, The upstairs market for large-block transactions: Analysis and measurement of price effects, The Review of Financial Studies 9, 1-36.

Keim, D. B., and A. Madhavan,1997, Transactions costs and investment style: An inter-exchange analysis of institutional equity trades, Journal of Financial Economics 46, 265-292.

Keim, D. B., and A. Madhavan, 1998, The Cost of Institutional Equity Trades, Financial Analysts Journal 54, 50-69

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

Madhavan, A., and M. Cheng, 1997, In search of liquidity: Block trades in the upstairs and and downstairs markets, The Review of Financial Studies 10, 175-203.

Næs, R., and B. A. Ødegaard, 2006, Equity trading by institutional investors: To cross or not to cross? Journal of Financial Markets 9, 79-99.

Næs, R., and J. A. Skjeltorp, 2003, Equity trading by institutional investors: Evidence on order submission strategies, Journal of Banking and Finance 27, 1779-1817.

O'Hara, M., 1997, Market microstructure theory (Wiley).

O'Hara, M., 2003, Presidential address: Liquidity and price discovery, The Journal of Finance 58, 1335-1354.

Parlour, C. A., and D. J. Seppi, 2003, Liquidity-based competition for order flow, Review of Financial Studies 16, 301-343.

Perold, A., and E. Sirri, 1994, The cost of international equity trading, Unpublished working paper, Harvard Business School .

Perold, A. F., and E. C. Schulman, 1988, The free lunch in currency hedging: Implications for investment policy and performance standards, Financial Analysts Journal 44, 45-50.

Plerou, V., and H. E. Stanley, 2007, Tests of scaling and universality of the distributions of trade size and share volume: Evidence from three distinct markets, Physical Review E 76, 46109.

Plexusgroup, 2004. The Changing Role of the Buyside Trader. Commentary 81.

Ready, M.,(2008) Determinants of Volume in Dark Pools, Working paper

Securities and Exchange Commission, 1998, Regulation of Exchanges and Alternative Trading Systems. Release Number 34-40760, 8 December.

Seppi, D., 1997, Liquidity provision with limit orders and a strategic specialist, Review of Financial Studies 10, 103-150.

Sirri, E., 2008, Keynote Speech at the SIFMA 2008 Dark Pools Symposium.

Sofianos, G. and Jeria, D. Quantifying the SIGMA X Crossing Benefit: The Sequel 2008, Journal of Trade, Fall, 84-91

Stoll, H. R., 2000, Friction, The Journal of Finance 55, 1479-1514.

CHAPTER 3

TRANSACTION COSTS AND MARKET SHARE OF CROSSING NETWORKS

3.1. Introduction

A major research topic in market microstructure is the measurement of transaction costs and the examination of their patterns. In his presidential address to the American Financial Association, Stoll (2000) provides seven measures of transaction costs (quoted spread, effective spread, traded spread, covariance of price changes, covariance of quote changes, price impact of trade and opening volatility) and examines their relationships. Yet all seven of these measures seem irrelevant for crossing networks, defined by the Securities and Exchange Commission (SEC) as "systems that allow participants to enter unpriced orders to buy and sell securities. Orders are crossed at a specific time at a price derived from another market."[57] (SEC, 1998). Crossing networks have grown exponentially in the past several years and account for 15.19% of average daily U.S. security trading volume.[58] Recently, this kind of trading platform has received even more public attention, not only because of two requests for comments from the SEC on crossing networks, but also because of “the industry's curious choice of the name ‘dark pool.’”[59] Despite their importance, there is very limited study of transaction costs in crossing networks. The purpose of this paper is to provide an empirical measure of transaction costs in crossing networks from publicly available data, to examine the pattern of transaction costs in crossing networks, and to study the

---

[57] Crossing networks have begun to allow limit orders, however, the fact that crossing networks cross orders using prices derived from other markets has not been changed. Limit orders in the crossing networks will not participate in the cross if the cross price is not as good as the limit price.
58 Rosenblatt Securities Monthly Dark Liquidity Tracker, April 27, 2010, pp2.
[59] "Exchanges should unite to end flash orders," by Nasdaq CEO Robert Greifeld, Financial Times, August 6, 2009.

competition between crossing networks and traditional trading platforms, such as stock exchanges and electronic communication networks.

The traditional measures of transaction costs are less relevant for crossing networks because of their unique trading mechanism. Crossing networks usually use the price set by other markets to match buy and sell orders. The match is conducted anonymously, and crossing networks have proprietary allocation rules to decide the priorities of trading when buy and sell orders are not balanced. Therefore, the price impact of trade is technically 0 for crossing networks because price is determined in other markets and is independent of order size. The effective spread is also 0 if buy and sell orders are matched using quoted-midpoint, which is the business model of many crossing networks. The other five measures of transaction costs in Stoll (2010) are either 0 or do not exist at all for crossing networks. [60]

The major transaction cost of crossing networks is non-execution: only the side with fewer shares gets full execution, while the side with more shares does not get full execution. Theoretical studies on crossing networks (Hendershott and Mendelson (2000), Dönges and Heinemann (2006), Degryse, Van Achter and Wuyts, (2009) and Ye (2010)) all focus on the choice between guaranteed execution with a higher bid-ask spread, or the price impact of trade in exchanges and a lower bid-ask spread or price impact but lower execution probability in crossing networks. However, non-execution probability is a missing piece in most empirical work on crossing networks due to the

[60] We cannot calculate quoted spreads from crossing networks because they do not have their own quotes. For the same reason, we cannot compute covariance of quote changes for crossing networks. The traded spread is also 0 because buy and sell orders are matched using the same price. We cannot compute covariance of price changes for crossing networks because they do not have their own prices. Opening volatility depends on opening price, which is again irrelevant for crossing networks because they do not have their own prices at all.

lack of data. A direct consequence of this omission is two empirical puzzles in the literature. First, Conrad, Johnson and Wahal (2003) and Keim and Madhavan (1998) find that crossing networks consistently have lower transaction costs than stock exchanges. Then, a natural question is why crossing networks are not more widely used (Conrad, Johnson and Wahal (2003) and Ready (2009). The regulator also has the concern that crossing networks will "continue to expand indefinitely." (SEC, 2009) Second, Ready and Ray (2010) find that the market shares of crossing networks are not higher for stocks with higher bid-ask spreads, whose reductions in transaction costs should be higher in crossing networks. Our paper shows that non-execution probability and its cross-sectional variation can address these empirical puzzles.

We contribute to the literature first by constructing a measure of transaction costs of crossing networks, based on publicly available data. Probability of non-execution is derived from SEC 605 data, which in turn allow us to study the pattern and determinants of this dimension of transaction costs and explain the competition between crossing networks and stock exchanges. We find that execution probability in the crossing networks is only 4.11 percent for NYSE stocks and 2.17 percent for NASDAQ stocks, which is significantly lower than the fill rate of the stock exchange. This low fill rate can potentially offset the reduction in effective spread and price impact in crossing networks.

We then extend the literature on cross-sectional variation of transaction costs (Stoll (2000) and Madahvan (2000)) to non-execution in crossing networks. Our empirical findings are consistent with the theoretical prediction of Ye (2010) that non-execution probability in crossing networks should follow a similar cross-sectional pattern as price impacts in stock exchanges. More broadly, non-execution should also follow the cross-sectional pattern of a bid-ask spread because the bid-ask spread can be observed as the "price impact of trading standard size order" (Amihud, 2002). The intuition of the Ye model is that rational traders would move their trades between crossing networks and stock exchanges until they are indifferent between non-execution probability in crossing networks and price impacts in exchanges. We employ three methods for testing this hypothesis. First, we show that non-execution, like bid-ask spread, can also be explained by informational and non-informational causes. Second, we show that non-execution is positively correlated to the price impact of trade, that is, stocks with a higher price impact of trade have a higher non-execution probability. Finally, we show that cross-sectional variation of non-execution can be well explained by the same underlying trading characteristics that explain cross-sectional differences in effective spread and price impact. The close association between effective spread, price impact and non-execution provides an explanation for Ray (2010) and Ready (2009). Ray (2010) finds that crossing networks do not have a higher market share for stocks with higher effective spreads. This can be explained by the positive correlation between the bid-ask spread and non-execution. Conditional on execution, stocks with a higher bid-ask spread should have a higher reduction in transaction costs in crossing networks. However, stocks with higher bid-ask spreads are also stocks with lower fill rates in crossing networks. Therefore, the higher potential saving conditional on trading success is counteracted by the higher failure rate of trade. Ready (2009) questions why crossing network volume is not higher for stocks with the highest volume, where the likelihood of finding counterparties should be highest. Our paper does show that the execution probability is increasing in trading volume, but it is well-known that bid-ask spreads also decrease with trading volume (Stoll (2000 and 2003) and Madahvan (2000)). Stocks with higher volume have lower transaction costs in both stock exchanges and crossing networks. Therefore, there may not be a comparative advantage for crossing networks with higher volume stocks.

The final question we ask is on the competition between exchanges and crossing networks. Particularly, we test the theoretical hypotheses of the competing models of Dönges and Heinemann (2006) and Ye (2010). The Dönges and Heinemann model predicts that market share of the crossing network decreases in the volatility of the stock, while the Ye model predicts the opposite. These two models have different results because they focus on two different aspects of competition. In Dönges and Heinemann, no traders have better information than other traders do. The disutility of missing the trading opportunity is higher for stocks with higher volatility. Therefore, traders move to the stock exchange for guaranteed execution when price volatility increases. The Ye model, however, includes a trader with better information about the true value of the stock. An increase in stock volatility increases the value of the information for the informed trader, giving the informed trader a higher incentive to hide in the crossing network. Certainly, Dönges and Heinemann (2006) and Ye (2010) only focus on one effect of an increase in volatility. In reality, both effects should play a role, and determining which force is stronger is an empirical issue. Using data from 2005-2007, Ready (2009) finds that stocks with higher volatility have a higher market share, implying that the effect found by Ye (2010) is stronger in that sample period. Using the data from January 2010 – March 2010, we find that stocks with higher volatility have lower market shares in crossing networks. This is consistent with the finding of Buti, Rindi and Werner (2010), which covers a more recent period but uses a different dataset. We believe that the discrepancy between Ready (2009) and Buti, Rindi and Werner (2010) and Ye (2010) is due to differences in sample periods. Currently, crossing networks have better anti-gaming strategies to exclude traders with private information or give them less priority to trade. Therefore, while the effect in Ye (2010) dominates the effect in Dönges and Heinemann (2006) from 2005-2007, informed traders play a relatively less important role now due to better anti-gaming strategies.

Due to the limitation of data, there are very few empirical studies on non-execution and market share of crossing networks. To my knowledge, there are only three academic studies on non-execution probability based on proprietary datasets with limited sample coverage. Gresse (2006) finds that aggregate execution probabilities in Posit Europe were 2.63% from July 1, 2000 to December 31, 2000 and 4.13% from January 1, 2001 to June 30, 2001. However, no further analysis has been done on cross-sectional patterns of non-execution probability. Næs and Ødegaard (2006) and Næs and Skjeltorp (2003) examine non-execution probability, based on three days of trading data from one institutional trader (the Government Petroleum Fund in Norway), and the number of orders on one of these three days was “too small to perform reliable statistical tests.” [61] Because of data limitation, the study on competition between crossing networks and stock exchanges (Ready (2009) and Ray (2010)) have to reply on assumptions about transaction costs in crossing networks. Both papers assume that stocks with a higher effective spreads have higher reductions in effective spreads by trading in crossing networks. We show, however, that this assumption is not supported by empirical data for NASDAQ stocks.

This chapter is organized as follows. Section 3.2 provides institutional details of crossing networks. Section 3.3 describes how this paper relates to the existing literature and develops the hypotheses to be tested. Section 3.4 describes our data and sample

[61] Næs and Skjeltorp (2003), pp1789, footnote 16.

selection criteria. Section 3.5 provides preliminary result on the measurements of transaction costs in crossing networks and stock exchanges as well as the competition of these two trading platforms. Section 3.6 examines cross-sectional variation of non-execution probability. Section 3.7 studies the competition between exchanges and crossing networks. Section 3.8 concludes the chapter.

3.2. Institutional Details

Crossing networks originated in the early 1970s as private, phone-based networks among buy-side traders. In the 1980s, with the introduction of Instinet and POSIT, the networks became electronic. Currently, there are about 40 crossing networks, which execute 15.76% of U.S. equity trading volume. The trading mechanism of crossing networks changes very fast, and there are many types of crossing networks. Some modern crossing networks do not even fit exactly into the traditional definition of crossing network in SEC (1998). However, there are three key elements that define crossing networks and distinguish their types.

First, crossing networks all use prices from other markets to match buy and sell orders. These prices, which are often called benchmark prices, can be bid-ask midpoint, closing price, volume-weighted average price, or national best bid and offer price. Some crossing networks may have more than one benchmark price. For example, Goldman Sachs Sigma X has midpoint peg orders, which participate in the crossing at the quoted midpoint. It also has peg-at-bid and peg-at-ask orders. Orders in crossing networks, however, do not participate in the formation of prices but simply free-ride the price discovery in stock exchanges. Trading in crossing networks should have no direct price impact because price is determined before order matching: an increase of buy orders does not increase the price but simply increases the execution probability for sell orders and decreases execution probability for buy orders. However, Ye (2010) shows that trading in crossing networks can impact price indirectly because the agent who sets the price has a rational expectation of other traders' strategies in crossing networks.

Second, because prices in crossing networks are derived from other markets, they do not have the market-clearing function. Crossing networks need proprietary matching algorithms to determine the trading priority for the side with the larger quantity. Examples of basic allocation rules include the time priority rule and the pro rata rule; rules in reality may be complex functions of these basic rules and are mostly confidential. As crossing networks are not public exchanges, their customers can be selected, and some traders can be excluded. This can be considered an extreme allocation rule in which some traders always get 0 execution. Crossing networks' preferred customers are "buy-side" firms, particularly those who manage "passive portfolios" such as index funds. Two kinds of traders are often excluded from the crossing network. The first kind are the potentially informed traders, such as hedge funds, brokers and proprietary traders from sell-side firms; the second kind are traders who submit small orders to extract information contained in the order flow. SEC (2009) proposed a “free access rule,” such that every trader has access to crossing networks. However, whether or not “fair access” will increase market quality is still an open question.

Finally, crossing networks differ in their matching frequency. In the past, most crossing networks only matched orders once or several times a day. Currently, more and more crossing networks conduct continuous matching.

There are several advantages to trade in crossing networks. First, trades also do not have direct price impacts, as their prices are independent of order sizes. Second, buyers (sellers) do not pay the bid-ask spread if their orders are matched at midpoint or the bid (ask) price. Conditional on execution, crossing networks usually have lower transaction costs than does the exchange (Keim and Madhavan (1998), Conrad, Johnson and Wahal (2003), Næs and Ødegaard (2006) and Sofianos and Jeria (2008)). In addition, institutional traders like to use crossing networks because they prevent information leakage. If information associated with an institutional order leaked out, opportunistic front-runners might trade in advance of the order in the same direction, thereby driving the price in an unfavorable direction.

The three benefits of trading in crossing networks prompt Conrad, Johnson and Wahal (2003) and Ready (2009) to ask why crossing networks are not more widely used. The answer is that the probability of execution in crossing networks is significantly lower than that in the exchange. If we measure trading costs for both executed orders and non-executed orders using the implement shortfall developed by Perold (1988), we can say that crossing networks have lower execution costs but higher opportunity costs.

3.3. Related Literature

The theoretical literature on transaction costs in stock exchanges can be classified into two lines (O’Hara, 1995). The first line is inventory models, such as Stoll (1978), Ho and Stoll (1981) and Amihud and Mendelson (1980), in which information is symmetric. More recent literature, such as Kyle (1985), Glosten and Milgrom (1985) and Easley and O'Hara (1987), focus on the transaction costs incurred by information asymmetry. Only recently has theoretical literature on transaction costs in crossing networks been published, but it can also be divided into two lines similar to the literature on transaction costs in stock exchanges.

Even if there is no information asymmetry, non-execution can still arise because of a random mismatch of buyers and sellers. Dönges and Heinemann (2006) and Hendershott and Mendelson (2000)[62] emphasize the relationship between network externality and non-execution probability. All other things being equal, non-execution probability should decrease in order arrival rate. The more shares that arrive to the market, the higher the probability to find a potential match, and the lower the non-execution probability. Non-execution can also be a consequence of information asymmetry: on one side of the market, there are both informed and uninformed traders, and on the other side, there are only uninformed traders. Therefore, an increase of informed trading relative to uninformed trading would increase non-execution probability. The non-execution caused by information asymmetry is the focus of Ye (2010). The Ye model also predicts that non-execution probability should increase when volatility increases.

We first examine whether or not non-execution contains both informational and non-informational causes by regressing non-execution probability on proxies of network externality (number of shares submitted to crossing networks and consolidated trading volume) and proxy of information asymmetry (the price impact of trade). Then, we test the following two hypotheses. Hypothesis 1 follows the prediction of network externality models such as Dönges and Heinemann (2006) and Hendershott and Mendelson (2000).

[62] There are informed traders in the Hendershott and Mendelson model. However, those informed traders cannot choose how much and where to trade. Therefore, the results in the Hendershott and Mendelson model are driven by traders without private information.

Hypothesis 1: non-execution decreases as trading volume increases

Hypothesis 2 is the implication of Ye (2010) model.

Hypothesis 2: non-execution increases as volatility increases.

Then, we want to examine the association between non-execution and measures of transaction costs in the exchange. In Dönges and Heinemann (2006) and Degryse, Van Achter and Wuyts (2009), non-execution is assumed to have 0 correlation with transaction costs in the exchange, whereas Ye (2010) predicts that non-execution should have a positive correlation with the price impact of trade because rational traders who can trade in both exchanges and crossing networks would move their trades until they are indifferent between these two dimensions of transaction costs. Therefore, we have hypothesis 3:

Hypothesis 3: Non-execution probability is positively correlated with price impact.

Finally, we examine the competition between crossing networks and exchanges. There are two papers on this topic. Ready (2009) finds that crossing networks' market share is not higher for the highest volume stocks, where the likelihood of finding counterparty should be the highest. Ready's explanation is that institutional trader's face other constraints besides minimizing transaction costs. Of particular importance is the soft dollar constraint. Ray (2010) finds that the market shares of crossing networks also do not have a monotonic relationship with effective spread. Starting from stocks with the lowest effective spread, the market shares of crossing networks first increase and then decrease with effective spread. Ray explains that it is because people who use crossing networks have concerns of possible gaming for stocks with higher effective spread.

Both of these papers, however, do not have data on transaction costs in crossing networks. Therefore, their analysis relies on assumptions about the transaction costs in crossing networks. For example, Ray (2010) implicitly assumes that the effective spread in crossing networks is 0, and Ready (2009) assumes that potential savings by using crossing networks is a fixed proportion of the spread. Both assumptions imply that effective spread in exchanges and the reduction in effective spread by trading in crossing networks should have correlation coefficient of 1. This hypothesis can certainly be tested using our data. However, we believe it is more informative to test whether the correlation efficient is positive or negative. So we have hypothesis 4 and hypothesis 4'.

Hypothesis 4: The reduction in effective spread by trading in crossing networks and effective spread in exchanges have a correlation coefficient of 1.

Hypothesis 4': The reduction in effective spread by trading in crossing networks and effective spread in exchanges have a positive correlation coefficient

Certainly, hypothesis 4' is weaker than hypothesis 4. If hypothesis 4' is rejected, so is hypothesis 4.

Finally, we want to test the competing hypothesis of Dönges and Heinemann (2006) and Ye (2010) on the market share of crossing networks. In the Ye model, an increase in volatility increases the value of information for informed traders, giving them higher incentives to hide their trading in crossing networks. Therefore, an increase in volatility may increase the market share of crossing networks. The benefit of hiding information in crossing networks does not exist in Dönges and Heinemann model because no traders have better information. The only effect of increased volatility is to increase the disutility of failed trade. Therefore, crossing networks have lower market shares for stocks with higher volatility. Due to the difficulty to model crossing networks, both the Ye model and the Dönges and Heinemann model can only focus on one side of the mechanism: Ye model has informed trader but passive uninformed trader whereas the Dönges and Heinemann model has no informed trader. We believe that both informational and non-informational factors should play a role in determining the market share of crossing networks, and this paper tests which effect plays a more important rule. The hypothesis is stated as follows:

Hypothesis 5: Crossing networks have higher market shares for stocks with higher volatility.

3.4 Data

We apply four datasets in this study. SEC 605 data is used to calculate effective spread, the price impact of trade and non-execution probability. In the United States, each market center that is not registered as a stock exchange must post a link of its SEC 605 report on the Financial Industry Regulatory Authority (FINRA) website. [63] We compare the list of these market centers with the list of crossing networks in Domowitz, Finkelshteyn and Yegerman (2009) to identify our sample of crossing networks. CRSP data is used to identify the sample of stocks as well as the trading characteristics of those stocks.

3.4.1 Measures of Transaction Costs and Market Share

Five measures of transaction costs are generated through SEC 605 data: effective spread, realized spread, price impact, execution speed and non-execution probability.

[63] http://apps.finra.org/datadirectory/1/marketmaker.aspx

Aside from non-execution probability, all other measures are widely studied in the literature, using SEC 605 data[64], (Boehmer, Jennings and Wei (2007), Boehmer (2005), Bennett and Wei (2006), Lipson (2003), Bessembinder (2003), O'Hara and Ye (2010)). Effective spread measures the total price impact of trade (Boehmer, Jennings and Wei (2007)), which can be decomposed into temporary price impact (realized spread) and permanent price impact. The share-weighted average of effective spreads in the SEC 605 report is calculated, for buy orders, as double the amount of difference between the execution price and the midpoint of the consolidated best bid and offer at the time of order receipt and, for sell orders, as double the amount of difference between the midpoint of the consolidated best bid and offer at the time of order receipt and the execution price. The realized spread excludes the effects of the information content of order flow. It is defined, for buy orders, as double the amount of difference between the execution price and the midpoint of the consolidated best bid and offer five minutes after the time of order execution and, for sell orders, as double the amount of difference between the midpoint of the consolidated best bid and offer five minutes after the time of order execution and the execution price. Price impact, the permanent component of effective spread, is defined as twice the change in the quote midpoint from order receipt to five minutes after the trade, or the difference between effective spread and realized spread. Execution speed is defined as the time between order receipt and execution. [65]

Our paper is novel because of the measurement of non-execution of crossing networks obtained from SEC 605 data. Execution probability is defined as the ratio of executed

---

[64] The data is also called Dash 5 data or SEC 11Ac1-5 data in early studies.

[65] Execution speed is not a variable in raw SEC 605 data, though some vendors of SEC 605 data provide execution speed data. We generate execution speed from raw SEC data using the same formula as those vendors, which is defined as the following weighted average:

speed= (shares executed with price improvement*average speed for shares executed with price improvement shares + executed at the quote*average speed for shares executed at the quote + shares executed outside the quote*average speed for shares executed outside the quote)/(shares executed with price improvement + shares executed at the quote + shares executed outside the quote) shares to covered shares in SEC 605 reports. SEC rules require each market center to report any market order or limit order (including immediate-or-cancel orders) received by a market center during regular trading hours at a time when a consolidated best bid and offer is being disseminated, and, if executed, is executed during regular trading hours, but shall exclude any order for which the customer requests special handling for execution.[66] Meanwhile, SEC 605 reports also require market centers to report the cumulative number of shares of covered orders executed at the receiving market center. Therefore, we have a measure of non-execution probability defined as follows.

$$Nonexecution = 1 - \frac{number\ of\ shares\ executed}{number\ of\ shares\ covered} \qquad (3.1)$$

The advantage of this measure is that it is calculated from public available data. As the dataset for crossing networks is very hard, if not impossible, to obtain, a measure based on public data provides an easy proxy for non-execution probability in empirical studies. This proxy allows us to answer some questions that are not addressed in the literature.

We acknowledge, however, the proxy for non-execution, as well as the measures of effective spread, realized spread, price impact and execution speed has their limitations, which impose a constraint on the type of question we can ask. We focus on the cross-sectional comparison of non-execution probability and market share across different stocks in this paper because of the following limitations in our data.

First, our measures for non-execution, effective spread, realized spread, price impact and execution speed only cover certain sizes and types of orders received by each market center. Orders of 10000 shares or more are not in SEC 605 data. More

[66] The special handling orders include, but are not limited to, orders to be executed at a market opening price or a market closing price, orders submitted with stop prices, orders to be executed only at their full size, orders to be executed on a particular type of tick or bid, orders submitted on a "not held" basis, orders for other than regular settlement, and orders to be executed at prices unrelated to the market price of the security at the time of execution.

importantly, market centers now have some discretion on the types of orders they include in their SEC 605 report. The discretion comes from the fact that current order types, especially order types in crossing networks are much more complex than the order types defined in SEC 605 rule in 1998.[67] As some new order types do not follow the standard definition of market or limit orders, market centers can choose whether they include them in the SEC 605 report or not. Therefore, while early works using SEC 605 data focus on the comparison between execution qualities of different market centers, we take a different approach because it is possible now for some market centers to exclude some types of orders to improve their execution quality. Comparing the crossing sectional variation of execution statistics has much less problem: market centers may exclude some order types to improve their execution statistics, but they should manipulate their execution statistics in the some way for all the stocks. We assume that the way orders are excluded would not systematically affect the relative cost for different stocks.

Equation (3.1) also tends to overestimate non-execution probability because it does not account for shares cancelled before execution. An alternative measure is

$$Nonexecution = 1 - \frac{\text{number of shares executed}}{\text{number of shares covered-number of cancelled shares}} \qquad (3.1')$$

However, (3.1') would greatly underestimate non-execution probability because several market centers in our sample treat all non-executed shares as canceled shares.[68] Therefore, there is not even cross-sectional variation in non-execution probability for these market centers because non-execution probability is always 0. Two other reasons

[67] For an introduction for the list of current order types, please see Johnson (2010).

[68] This is mainly because of order types in crossing networks. For example, Liquidnet SEC 605 reports only include their Immediate or Cancel orders, while their time in force orders are exempt from SEC 605 reporting requirements. Therefore, the execution probability is always 1 because orders are either executed immediately or cancelled.

make us choose measure (3.1) instead of measure (3.1'). First, equation (3.1) is the measure of the fill rate recommended by SEC 605 final rule[69]. Second, the aggregated non-execution probability generated by excluding cancelled shares is much more consistent with previous literature (Gresse, 2006).

Fortunately, non-execution is used as a dependent variable in all but one case for this paper. As long as the measurement error is not correlated with the explanatory variables, measurement error in dependent variable does not affect the consistency of the estimation of slope terms. [70] There is only one place where we use non-execution as an independent variable. In this case, measurement (3.1') can potentially serve as an instrument for measurement (3.1). However, measurement (3.1') is such a poor instrument that we cannot reject the hypothesis that they have 0 partial correlations measurement (3.1).[71] This is due to the fact that non-execution measured by (3.1') has very limited variation. Non-executions for many stocks are close to 1 because a number of market centers simply count non-executed shares as canceled shares. Fortunately, measurement error does not change the answer for the question we want to ask. Under classical errors-in-variables assumption, measurement error causes attenuation bias, in that it leads the estimated coefficient closer to 0 than the true value. In addition, the standard error of estimation increases. Therefore, even if non-execution has impact on the dependent variable, we may conclude it does not have an impact due to the measurement error. In section 3.6, we find that non-execution has impact on market share of crossing networks, and we believe the result would be stronger if there is no measurement error.

---

[69] See footnote 51 of SEC 605 final rule.

[70] Because equation (3.1) is an underestimate, we would get an underestimate of the intercept, which is rarely a cause for concern. (Wooldridge, 2006)

[71] Results are not reported for brevity but are available upon request.

We also exclude orders received by a market center but executed elsewhere. Because crossing networks sometimes route their unexecuted orders to other types of trading platforms, including orders routed to other market centers may exaggerate the execution probability for crossing networks. We do conduct the analysis including orders executed in other market centers, and the results are similar.

We also obtain the market share of crossing networks from SEC 605 data, which is defined as

$$\textit{market share} = \frac{\textit{shares executed in crossing networks}}{\textit{shares executed in stock exchanges} + \textit{shares executed in crossing networks}} \quad (3.2)$$

Our market share data may also have some bias because of double counting. Market centers have different protocols for reporting their executed shares. Some market centers may report single-counted volume, where only the number of matched shares is reported. Some other market centers may report double-counted volume, where both buy and sell volume is counted. Once again, focusing on crossing sectional variation will be less a problem. While double-counting in a market center may increase its market share relative to other market centers, we assume that it does not systematically affect the relative market share across different stocks.

Our analysis focuses on all market and marketable limit orders in the SEC 605 data. We apply the filter in Bessembinder (2003) and eliminate an observation if the effective spread is greater than four dollars or less than -0.5 dollars. We also drop a stock if it has an average execution time greater than one trading day either in crossing networks or in stock exchanges. As our study focuses on a cross-sectional comparison, we aggregate the number of shares executed and the number of shares covered in all our sample crossing networks and exchanges in equation (3.1), and we aggregate the number of shares executed in all sample crossing networks and exchanges in equation (3.2). Our choice for sample crossing networks and exchanges is specified in section 3.4.2 and 3.4.3.

3.4.2. Sample Crossing Networks

To choose the sample crossing networks for this paper, we start from the list of crossing networks in Domowitz, Finkelshteyn and Yegerman (2009). [72] Then we compare the list with the master file provided by Financial Industry Regulatory Authority (FINRA), which has names of all the market centers reporting the SEC 605 data through the FINRA website. Because every market center that is not a stock exchange needs to post the links of their SEC 605 data on the FINRA website, theoretically, all the execution data for orders of size 9999 or less in the crossing networks is included in SEC 605 data. However, data from several crossing networks are not available, either because their orders are exempt from SEC 605 reports or because the owners of those crossing networks merge their crossing networks data with execution data from other trading platforms they own. The complete list of crossing networks in Domowitz, Finkelshteyn and Yegerman (2009) and our filters to select the final sample are demonstrated in Table 3.1.

[72] See page 20 of Domowitz, Finkelshteyn and Yegerman (2009) for the list. Though the list is not complete, it includes all current major crossing networks.

Table 3.1: Sample of Crossing Networks

This table shows the list of crossing networks in Domowitz, Finkelshteyn and Yegerman (2009) and our filters to choose the final sample of crossing networks. The sample period is January, 2010 to March 2010.

1. Orders are exempted from Sec 605 report for at least one month		
Pipeline, Pulse Block Cross, Bids and eSpeed Aqua		
2: Exchange owned crossing networks		
International Stock Exchange Midpoint Match, NASDAQ (End of Day Cross; Open; Intraday; Continuous) and NYSE Match Point		
3. Broker-dealer owned crossing networks that do not report independently		
Credit Suisse Cross Finder, Bloomberg BlockHunt, Citadel ExSvs, Citi Markets LIQUIFI, Fidelity CrossStream, Knight Securities Knight Match, Morgan Stanley (Trajectory Cross, MS Pool), Merrill Lynch (MLXN;AXP), State Street Lattice and UBS PIN		
4. Final sample		
Provider	Name	Rank
Goldman Sachs	Sigma X	2
Consortiums	Level	5
Barclays	Barclays ATS	7
ITG	Posit Now	10
Instinet	Data contains multiple crossing networks operated by Instinet	11
Liquidnet	Data contains multiple crossing networks operated by Liquidnet	12
NYFIX	Millennium	13
BNY ConvergEx	ConvergEx Cross	16

In our sample period, all orders from Pipeline, Pulse Block Cross and Bids trading are exempted from SEC 605 reports, and the reports from eSpeed are empty for both January and February 2010. Therefore, we exclude these four crossing networks from our sample. We also need to exclude crossing networks owned by NYSE, NASDAQ and International Stock Exchange from our sample because data of these crossing networks are not reported independently. For the same reason, we need to drop crossing networks owned by Bloomberg, Citadel, Citi Markets, Credit Suisse, Fidelity, Knight Securities, Merrill Lynch, State Street and UBS because their data is mixed with execution data of other trading platforms of the same market center.

Our final sample includes eight crossing networks. Two are independent crossing networks: Level and Liquidnet. Goldman Sachs Sigma X, ITG POSIT Now, Instinet and Barclays ATS are broker-dealer-owned crossing networks but file their independent SEC 605 reports.[73] BNY ConvergEx group files two SEC 605 reports: BNY ConvergEx and ConvergEx's Millennium ATS.[74] I also include these two market centers because both of them are crossing networks. We compare our sample of 8 crossing networks with the report of Rosenblatt Securities on aggregated volume of crossing networks and find that our sample of crossing networks ranks 2, 5, 7, 10, 11, 12, 13 and 16 in all 17 crossing networks they track, which is a pretty representative sample[75].

3.4.3. Sample Exchanges

Sirri (2008) divides trading venues into two categories based on whether or not they display quotes as an integral part of their business models. Roughly speaking, "quoted venues," according to Sirri (2008), include exchanges and ECNs, even if they may offer one or more dark liquidity services through hidden orders or reserved orders. Because

[73] Sigma X is reported as an independent center with a market center code SGMA, while orders executed under other platforms of Goldman Sachs is reported under the market center code GSCO. Barclays ATS has a market center code LATS, while the execution, through other trading platforms of Barclays Capital is reported in market center with a code LEHM. ITG reports its executions through Posit under the market center name Posititnow and market center code TACT. According to my conversation with the compliance office of Instinet, orders reported through the market center INCA are its crossing network execution reports.

[74] BNY ConvergEX files two reports because BNY ConvergEX acquired NYFIX Millennium in 2009. However, the latter still keeps its old market center code NYFX.

[75] Rosenblatt Securities Monthly Dark Liquidity Tracker, April 27, 2010, pp4.

the trading mechanisms of ECNs and exchanges are similar, and some ECNs, such as BATS, actually gained status as exchanges, we use the word "exchange" in this paper instead of "quoted venue," for the sake of brevity.

I collect the SEC 605 reports of all the stock exchanges that filed the reports from January 2010 to March 2010. [76] These include NYSE, NYSE Amex, NYSE ARCA, NASDAQ (including the acquired Boston Stock Exchange and Philadelphia Stock Exchange), National Stock Exchange, Chicago Stock Exchange, International Stock Exchange and Chicago Board of Option Stock Exchange. I also include data from Bats and Direct Edge in our sample. SEC (2010) shows that these market centers represented 73.6% of total U.S. trading volume in September 2009, whereas other quoted venues only executed 1%. Therefore, we believe these trading platforms represent quoted venues as a whole.[77]

3.4.4. Sample Stocks

We use CRSP data to choose our sample of stocks and to measure the characteristics of different firms. To avoid the possibility that contemporaneous quarterly observations produce spurious associations, we apply the CRSP data from October 2009 to December 2009. I start from CRSP data, applying standard filters to remove non-

[76] NASDAQ recently acquired Boston Stock Exchange and Philadelphia Stock Exchange. Boston Stock Exchange still files SEC 605 reports, but there are no covered orders. Philadelphia Stock Exchange no longer files SEC 605 reports.

[77] One potential problem is that now exchanges also have their own crossing networks, which may be included in their SEC 605 reports. We believe that the impact of exchange-owned crossing network has very limited impact on the overall execution quality in SEC 605 data for exchanges. Firstly, the exchange-owned crossing networks only execute a small fraction of the exchanges' volume. Furthermore, much of exchange-owned crossing networks' volume is not included in SEC 605 data because SEC 605 data excludes any order for which the customer requests special handling for execution, including, but not limited to, orders to be executed at a market opening price or a market closing price. SEC 605 data also excludes orders executed after regular trading. As much of the volume of exchange-owned crossing networks is executed through their after-market crossings based on closing prices, this further reduces the problem that exchange-reported data also includes exchanges' executions through their crossing networks.

common equities, dual class shares, REITS, and common stocks of non-US companies. There are 3,610 stocks in our CRSP sample. Following Boehmer (2005), we also drop 304 stocks with average dollar volumes of less than $20,000, 501 stocks with average prices of less than $3 and three stocks with active trading days of two or less. Altogether, we have 1,151 NYSE stocks and 1,651 NASDAQ stocks in our sample.

3.5. Preliminary Results

Bessembinder (2003) and Boehmer, Jennings and Wei (2007) find that market centers competing with the listing exchanges only trade a small subset of stocks traded in the primary market. [78] Among the 2,802 stocks in our sample, there are 10 NYSE and 23 NASDAQ stocks that have no SEC 605 coverage, both in sample crossing networks and sample exchanges. Therefore, we delete these 33 stocks from our sample. Among the remaining 2,769 stocks, there is only one stock (ISRL) with executed volume in exchanges but not in crossing networks. Therefore, our finding is closer to O'Hara and Ye (2010), who find that competing market centers virtually trade all the stocks.[79] Our data also shows that there are no stocks that trade in crossing networks but not exchanges, which is obvious because crossing networks need the exchange to provide the price. For convenience, we delete ISRL from our sample and expect that the result would not change because of the deletion.

Our final sample has 1,141 NYSE stocks and 1,627 NASDAQ stocks. For each stock, we compute its effective spread, realized spread, price impact and executed speed in crossing networks as the weighted average of these variables across all crossing

78 Bessimbinder (2003) found that of the 500 NYSE listed stocks in his sample, other markets centers only traded between 77 and 163 stocks during his 2002 sample period. Boehmer, Jennings and Wei (2007) had 1,435 stocks in their sample, but only 258 traded continuously in market centers other than the listing market.

79 O'Hara and Ye (2010) find that ECN and Alternative Trading System trade all stocks in their sample.

networks. Execution probability in crossing networks is defined as the ratio of all shares executed in crossing networks to all shares covered by crossing networks. Effective spread, realized spread, price impact, execution speed and execution probability across all exchanges are defined in a similar way.

Table 3.2: Execution Quality in Crossing Networks and Exchanges

This table demonstrates the average execution quality measure across our sample crossing networks and exchanges as well as their pairwise difference. The sample period is from January, 2010 to March 2010. The asterisks ***, **, and * indicate significance level of one percent, five percent or ten percent.

Panel A: NYSE Stocks

		Crossing Networks	**Exchanges**	**Crossing Networks-Exchanges**	**p-value**
Effective Spread (in cents)					
T-test	Mean	1.61	2.02	-0.41***	0.0000
Wilcoxon Signed Rank Test	Median	0.97	1.27	-0.24***	0.0000
Effective Spread/Price (in Basis Points)					
T-test	Mean	8.42	10.34	-1.92***	0.0000
Wilcoxon Signed Rank Test	Median	4.64	5.77	-0.94***	0.0000
Price Impact (in cents)					
T-test	Mean	0.80	1.86	-1.07***	0.0000
Wilcoxon Signed Rank Test	Median	0.52	1.19	-0.64***	0.0000
Price Impact/Price (in Basis Points)					
T-Test	Mean	4.66	9.45	-4.79	0.0000
Wilcoxon Signed Rank Test	Median	2.48	5.71	-2.92	
Average Speed (in Seconds)					
T-test	Mean	32.10	13.29	18.81***	0.0000
Wilcoxon Signed Rank Test	Median	21.25	4.39	15.02***	0.0000
Execution Probability					
T-test	Mean	4.11	31.47	-27.36***	0.0000
Wilcoxon Signed Rank Test	Median	3.69	30.52	-26.46***	0.0000

Table 3.2 Panel B: NASDAQ Stocks

		Crossing Networks	Exchanges	Crossing Networks-Exchanges	p-value
Effective Spread (in Cents)					
T-test	Mean	4.47	4.72	-0.25**	0.0377
Wilcoxon Signed Rank Test	Median	1.49	2.25	-0.46***	0.0000
Effective Spread/Price (in Basis Points)					
T-test	Mean	45.01	47.72	-2.71**	0.0114
Wilcoxon Signed Rank Test	Median	13.48	20.18	-3.76***	0.0000
Price Impact					
T-test	Mean	1.55	2.73	-1.17***	0.0000
Wilcoxon Signed Rank Test	Median	0.56	1.51	-0.73***	0.0000
Price Impact/Price (in Basis Points)					
T-test	Mean	16.86	26.77	-9.92***	0.0000
Wilcoxon Signed Rank Test	Median	4.39	13.99	-5.96***	0.0000
Average Speed					
T-test	Mean	51.72	12.93	38.80***	0.0000
Wilcoxon Signed Rank Test	Median	29.77	4.11	22.80***	0.0000
Execution Probability					
T-test	Mean	2.17	26.48	-24.31***	0.0000
Wilcoxon Signed Rank Test	Median	1.65	25.88	-23.40***	0.0000

Table 3.2 shows the summary statistics of effective spread, price impact, execution speed and execution probability in exchanges and crossing networks as well as their

pairwise comparison. [80] We provide a measure both in cents and in basis points (standardized by the average of closing prices). The first observation is that trading in the crossing network has a non-zero effective spread and price impact of trade [81], although, technically, trading in crossing networks should have no price impact, and effective spread should also equal 0 if orders are matched using the quoted midpoint. A non-zero effective spread can easily be explained by order types in crossing networks. At Goldman Sachs Sigma X, for example, traders can enter orders pegged at mid-quote, but they can also enter orders pegged at bid and pegged at ask. If a buyer sends a pegged at ask order, he still needs to pay the spread. However, the price impact of trade in the crossing network should be 0 by definition (See Hasbrouck (2007)) and Gresse (2006)) because the price is determined before the quantity match. The positive price impact of trade has two possible explanations. First, Ye (2010) predicts that trade in the crossing network has an indirect price impact because rational agencies should draw a correct inference on hidden order flows in the crossing network by observing the order flow in the exchange. Therefore, trading in the crossing network moves prices indirectly. Second, the impact of trade is measured based on the price five minutes after the trade. At that point, the trade is reported through the consolidated tape. Though the trader's identity and the executed venues are not reported, the size of the trade still reveals some information and moves the price.

Despite the non-zero effective spread and price impact, crossing networks do have lower effective spreads and price impacts. The average (proportional) effective spread in crossing networks is 0.41 cents (1.92 basis points) lower than that in exchanges for NYSE stocks, and average (proportional) effective spread is 0.25 cents (2.71 basis

[80] I also do the comparison with the 33 stocks that are traded in the stock exchange but not in the crossing networks, and the results are similar.
[81] P value is equal to 0.0000 but not reported.

points) lower in crossing networks than that in exchanges for NASDAQ stocks. Median (proportional) effective spread in the crossing network is 0.24 cents (0.94 basis points) lower for NYSE stocks and 0.46 cents (3.76 basis points) lower for NASDAQ stocks. The reduction in the price impact of trade is economically more significant. Trading in crossing networks reduces the (proportional) price impact by 1.07 cents (4.79 basis points) in terms of the mean, and 0.64 cents (2.92 basis points) in terms of the median for NYSE stocks, and 1.17 cents (9.92 basis points) in terms of mean, and 0.73 cents (5.96 basis points) in terms of the median for NASDAQ stocks. The reason we find a larger reduction in price impact is because of the design of crossing networks. The major function of crossing networks is to reduce the price impact of trade because price is determined before the order match. Traders still need to pay a bid-ask spread if the match price is different from the mid-quote. Therefore, we see a larger effect on the price impact of trade than on effective spread. [82]

However, trading in crossing networks also has downsides. First of all, crossing networks are slower than exchanges. On average, crossing networks take 18.81 more seconds to execute an NYSE order and 38.8 more seconds to execute a NASDAQ order. More importantly, the execution probability of crossing networks is significantly lower than that of the exchange. The average execution probability for crossing networks is only 4.11% for NYSE stocks and 2.17% for NASDAQ stocks, while in exchanges the mean execution probabilities are 31.47% and 26.48%, respectively.[83] Certainly, as we

[82] The other possible explanation is that crossing networks cream-skim stock exchanges by picking the less informed orders to execute. Though cream-skimming provides explanations for the lower price impact of trade for dealers or Electronic Communication Networks competing with the exchange, it is less likely to be an explanation for the price impact of trade in crossing networks because crossing networks do not set their own prices.

[83] There are several reasons why execution probability of exchange is not close to 100%. First, we do not consider cancelled orders in our calculation. Second, the depth of the market is given. Therefore, a large order with a marketable limit price may be only partially filled because of the limited depth. Third, in SEC 605 data, an order is marketable when its price is better than the quote at the time of order receipt. However, price may move between the time of order receipt and order execution. Finally, Hasbrouck and

mentioned in Section 3.3, our measure tends to underestimate the execution probability. However, our execution probability for crossing networks is close to the finding in the previous literature (See Gresse (2006)). The results that crossing networks have lower execution speeds and execution probabilities are both statistically significant when we conduct t-tests for the mean and Wilcoxon signed rank tests for the median.

Figure 3.1 shows the cross-sectional variation of market shares of crossing networks. Panel (a) shows that the market share of crossing networks, in terms of executed shares from market and marketable orders, ranges from 0.4% to 38.3%. The mean market share of crossing networks, in terms of executed market and marketable limit orders, is 8.2%, and the median is 7.3%. However, measuring market share of crossing networks based on executed shares underestimates the true impact of crossing networks (Sirri (2008) and Hendershott and Mendelson (2000)). Panel (b) shows market shares of crossing networks in terms of submitted shares. The mean market share of submitted market and marketable limit orders in crossing networks is 51%, and the median market share is 52%, which implies that the number of shares in market or market limitable orders submitted to the crossing network is larger than their counterparts submitted to the exchanges.

As is mentioned in section 3.4, different market center may select the types of orders that they report. By focusing on market and marketable limit orders, we mitigate this problem. Still, there might be different interpretation of “market and marketable limit orders” across different market centers. As a result, our comparison between crossing networks and stock exchanges need to be explained with caution. Therefore, we will

---

Saar (2009) find the wide use of fleeting orders, which are cancelled immediately if orders are not executed.

focus on crossing stock comparison but not cross market comparison because of the rationale we discussed in section 3.4.

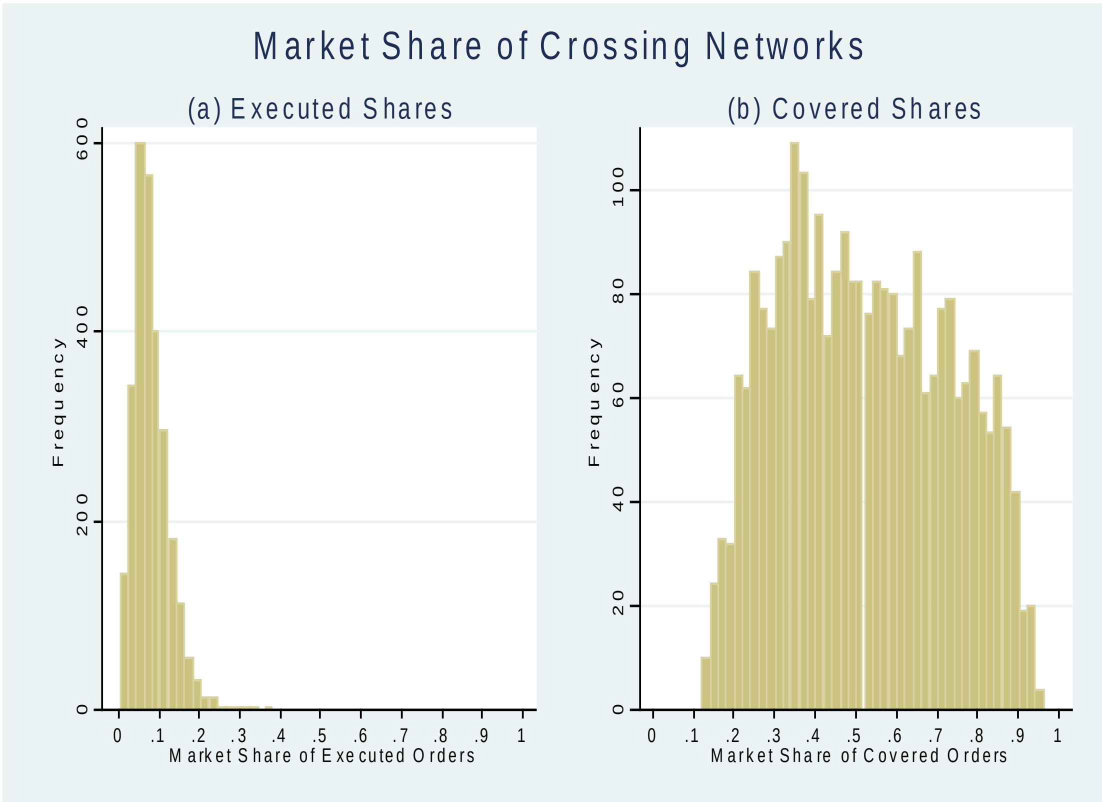

Figure 3.1. Market Share of Crossing Networks
This figure shows cross sectional variation on market share of crossing networks for market and marketable limit orders. Panel (a) shows market share of executed shares and Panel (b) shows market share of covered shares. Our sample has 1141 NYSE stocks and 1627 NASDAQ stocks. The sample period is from January 2010 to March 2010

3.6. Patterns of Non-execution Probability

This section examines the pattern of non-execution. First, we examine whether or not both informational and non-informational factors account for cross-sectional variation of non-execution probability by regressing non-execution on a proxy for network externality and a proxy for information asymmetry. Then, we provide two tests for Ye's (2010) hypothesis that non-execution follows similar patterns to price impact. The first test is to examine the correlation between non-execution and price impact and effective spread. The second test examines whether or not variables that explain cross-sectional variation in effective spread and price impact also explain the cross-sectional variation in non-execution probabilities.

3.6.1. Informational and Non-informational Factors of Non-execution

An important line of research in the market microstructure literature models transaction costs incurred by non-informational and informational causes. Transaction costs incurred by non-informational factors are the focus of inventory models, such as those of Stoll (1978), Ho and Stoll (1981) and Amihud and Mendelson (1980). The role of adverse information costs is emphasized by Copeland and Galai, Kyle (1985), Glosten and Milgram (1985) and Easley and O'Hara (1987). Following these theoretical works, Glosten and Harris (1988), Stoll (1989), Choi, Salandro and Shastri (1988), George, Kaul and Nimalendran (1991), Lin, Sanger and Booth (1995) and Huang and Stoll (1997) investigate the empirical methods to decompose bid-ask spreads into different components.

Crossing networks have quite different mechanisms from markets with intermediates. However, we can still consider that transaction costs are due both to inventory factors and adverse information factors. Non-execution can occur for non-informational reasons. Even if there is no information asymmetry, non-execution exists because of random mismatches between buyers and sellers. The major difference between crossing networks and markets with intermediates is that there is no market-maker to bear the inventory holding cost in crossing networks. Therefore, a larger mismatch leads to a higher non-execution probability instead of a higher bid-ask spread. Hendershott and Mendelson (2000) and Dönges and Heinemann (2006) demonstrate that non-execution caused by non-informational factors decreases as order arrival rates increase. The higher the arrival rate, the easier it is to find a potential match. Therefore, non-execution is inversely correlated with order arrival rate. Hendershott and Mendelson (2000) and Ye (2010), however, show that non-execution can also result from informational factors: while uninformed traders are equally likely to buy or sell, informed traders are always on one side of the market. Again, the major difference between crossing networks and markets with intermediates is that there are no market makers to offset this order imbalance. Therefore, higher information asymmetry results in a higher non-execution probability instead of a higher bid-ask spread or price impact.

It is important to separate these two causes of non-execution because they have different implications. Suppose that non-execution is caused by non-informational factors. Then price is equally likely to go up or down. Therefore, non-execution would not be a cost for risk-neutral traders if they are patient. Surely, if traders are risk-averse or if they prefer immediate execution, non-execution would be a cost. The non-execution caused by informational factors has a very different implication. Uninformed traders are equally likely to be on either side of the market, whereas traders with private information can only be on one side of the market. Non-execution caused by information asymmetry has the following implication. Orders on the correct side of the market are less likely to be executed. If an order is executed, it is more likely that it is on the wrong side of the market. Price tends to move in the opposite direction on executed shares.

Ideally, if we have data for each order submitted to crossing networks, we can decompose the non-execution into informational and non-informational parts, following methods similar to decomposing the bid-ask spread. Because we only have aggregate-level data, we can only conduct the following test. We want to examine whether or not both proxies for network externality and information asymmetry have explanatory powers for cross-sectional variation in non-execution.

As non-execution is a proportion defined as the ratio between executed shares and covered shares, it is bounded between 0 and 1. Therefore, we use standard logit transformation to transform the variable nonexe_c, which is defined as non-execution probability in crossing networks according to measurement 1.[84] A new dependent variable logit_nonexe is defined as

$$logit\_nonexe = ln(\frac{nonexe\_c}{1 - nonexe\_c}) \qquad (3.3)$$

This transformation maps the original variable, which was bounded by 0 and 1, to the real line. (See Fleiss, Levin and Paik (2003)). We also do the regression without logit transformation and the results are similar.

We use the permanent price impact of trade of each stock in exchanges as a proxy for information asymmetry for each stock. The other proxy for information asymmetry is permanent price impact normalized by the stock price. Dummy variable NASDAQ is equal to 1 if the stock is listed in NASDAQ and is equal to 0 if the stock is listed in NYSE. We do have the data for number of shares submitted to the crossing networks, which may serve as a proxy variable for order arrival rate or network externality in crossing networks. However, the number of shares submitted to crossing networks has the following endogenity problem: while the number of shares submitted can affect non-execution probability, the causality can go to the opposite direction. Lower non-

[84] Kennedy (2003) consider "using a linear function form when the dependent variable is a fraction" as one of the common mistakes easily to be made. He recommends logistic transformation for dependent variables.

execution probability increases the attractiveness of crossing networks and thereby the number of shares submitted to crossing networks. To solve the endogenity problem, we use the consolidated volume traded in all exchanges as an instrument for total number of shares submitted to crossing network. Panel A of Table 3.3 first show that consolidated volume is highly correlation with number of shares submitted to crossing networks. There is also good reason to believe that consolidated volume is exogenous in this regression. Non-execution may affect the number of shares routed to crossing networks v.s. stock exchanges and it may even affect total number of shares routed to these two centers because of failed trades. However, it is much less likely that non-execution can affect the consolidated trading volume, which is determined by portfolio management purposes.

Therefore, we run the following 2 regressions using both OLS and 2SLS with the log of consolidated volume in all trading venues as an instrument variable.

$$logit\_nonexe_i = \alpha + \beta_1 \log nse_i + \beta_2 pimpact\_e_i + \beta_3 NASDAQ_i + \varepsilon_i \qquad (3.4)$$

$$logit\_nonexe_i = \alpha + \beta_1 \log nse_i + \beta_2 propimpact\_e_i + \beta_3 NASDAQ_i + \varepsilon_i \qquad (3.5)$$

where *nse* is the number of shares entered into crossing networks, *pimpact_e* is price impact of trade of each stock in stock exchanges, and *propimpact_e* is the price impact of trade of each stock in stock exchanges divided by the price.

Table 3.3: Informational and Noninformational Factors of Non-execution

Panel A examines the correlation of lognse and its instrument (lognse), and Panel B regresses non-execution on proxies for network externality and information asymmetry using both OLS and IV estimation. logit_nonexe is the logit transformation of Non-execution probability in crossing networks. logvol is the log of average trading volume of each stock, pimpact_e is the price impact of trade for each stock in exchanges and propimpact_e is price impact normalized by average closing price for the stock. lognse is the total number of shares entered into the crossing network. Dummy variable NASDAQ is equal to 1 if the stock is listed in NASDAQ and 0 if the stock is listed in NYSE. The sample period is from January, 2010 to March, 2010.

Panel A: Partial Correlation between lognse and its instrument (logvol)

COEFFICIENT	lognse	lognse
pimpact_e	-0.0331***	
	(0.00295)	
NASDAQ	-0.290***	-0.263***
	(0.0244)	(0.0249)
logvol	0.993***	1.010***
	(0.00691)	(0.00748)
propimpact_e		-0.00165***
		(0.000428)
Constant	4.425***	4.152***
	(0.0975)	(0.105)
Observations	2768	2768
R-squared	0.925	0.922

Panel B

	OLS		IV	
	(1)	(2)	(3)	(4)
COEFFICIENT	logit_nonexe	logit_nonexe	logit_nonexe	logit_nonexe
pimpact_e	0.0100***		0.000830	
	(0.00338)		(0.00341)	
lognse	-0.455***	-0.429***	-0.500***	-0.474***
	(0.00733)	(0.00749)	(0.00689)	(0.00808)
NASDAQ	0.0882***	0.0717***		0.00831
	(0.0280)	(0.0274)		(0.0279)
propimpact_e		0.00437***		0.00297***
		(0.000462)		(0.000474)
Constant	11.48***	11.00***	12.31***	11.81***
	(0.135)	(0.137)	(0.119)	(0.148)
Observations	2768	2768	2768	2768
R-squared	0.705	0.713	0.700	0.709

Standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1

The coefficients for *lognse* in these four regressions are all negative and statistically significant, which is a strong indication of network externality effect. We also find after we control for proxy for network externality in crossing networks, we still find that information asymmetry play a role in explaining non-execution probability. Table 3.3 tell us that if two stocks have the same order arrival rate, the stock with higher price impact of trade has higher non-execution probability, which is an indication of informational cause of non-execution. However, the effect seems weaker because the coefficient is only statistically significant for 3 of the 4 specifications.

3.6.2. Correlation between Non-execution, Price Impact and Effective Spread

Ye (2010) predicts that the non-execution probability and (permanent) price impact of trade should have positive correlations, and Panel A in Table 3.4 demonstrates that it is indeed the case. The non-execution probabilities in crossing networks and price impacts of trade in exchanges have a positive correlation of 0.2842, and the correlation is statistically significant. We also do another robustness test: because we know from summary statistics that crossing networks also have a price impact of trade, and exchanges also have non-execution, we examine whether or not the difference between non-execution probability in crossing networks and stock exchanges is positively correlated with the difference between price impact in stock exchanges and crossing networks. We find that the correlation is 0.1742, meaning that stocks that have a higher reduction in price impact by trading in crossing networks also have a higher increase in non-execution probability in crossing networks. Therefore, the potential savings in price impact costs is counteracted by the lower fill rate. In conclusion, we cannot reject Hypothesis 3 that non-execution and price impact are positively correlated.

If we study the correlation between the effective spread, the sum of (permanent) price impact and realized spread (temporary price impact), the relationship becomes weaker. This is not surprising because Ye's prediction is on the relationship between permanent price impact and non-execution. Our results show that when we add the temporary component of price impact, the correlation between transaction costs in exchanges and those in crossing networks becomes weaker. Panel B and C in Table 3.4 demonstrates that stocks with higher effective spreads in exchanges also have higher non-execution probabilities in crossing networks. However, the results for their differences are weaker. For NYSE stocks, stocks with a higher reduction in effective spread by trading in crossing networks also a higher increase in non-execution probability in crossing networks. This result, however, is not true for NASDAQ stocks, where stocks with a higher reduction in effective spread actually have a lower increase in non-execution probability. This negative correlation may be a consequence of the other negative correlation: Table 3.4 demonstrates that stocks with higher effective spreads actually have a lower reduction in effective spread by trading in crossing networks.

The negative correlation between effective spreads in exchanges and the reduction in effective spreads by trading in crossing networks is contrary to the assumption of Ready (2009) and Ray (2010). Ray assumes that trading in crossing networks has a 0 effective spread, and Ready (2010) assumes that the potential cost saving is a fixed proportion of the effective spread. Under these two assumptions, effective spread and the reduction in effective spread by trading in crossing networks should have a correlation coefficient of 1. For NASDAQ stocks, however, Panel C shows that not only these two variables do not have correlation of 1, they do not even have a positive correlation. Therefore, both hypothesis 4 and hypothesis 4' are rejected.

Table 3.4: Correlation between Non-execution, Price Impact and Effective Spread

This table shows the cross-sectional measure among different measure of transaction costs. nonexe_c and nonexe_e are non-execution probability in crossing networks and exchanges, respectively. pimpact_c and pimpact_e are price impact of trade in crossing networks and stock exchanges, which are defined as the average effective spread minus the average realized spread in crossing networks and stock exchanges. espread_c and espread_e are average effective spread in crossing networks and stock exchanges. The sample period is from January, 2010 to March 2010. The asterisks ***, **, and * indicate significance level of one percent, five percent or ten percent.

Panel A: Correlation between Non-execution and Price Impact

	nonexe_c	pimpact_e	nonexe_c-nonexe_e	pimpact_e-pimpact_c
nonexe_c	1.0000			
pimpact_e	0.2842*** (0.0000)	1.0000		
nonexe_c-nonexe_e	0.2136*** (0.0000)	0.2739*** (0.0000)	1.0000	
pimpact_e-pimpact_c	0.1132*** 0.0000	0.6124*** (0.0000)	0.1742*** (0.0000)	1.0000

Panel B: Correlation between Non-execution and Effective Spread: NYSE stocks

	nonexe_c	espread_e	nonexe_c-nonexe_e	espread_e-espread_c
nonexe_c	1.0000			
espread _e	0.2078*** (0.0000)	1.0000		
nonexe_c-nonexe_e	0.3397*** (0.0000)	0.2928*** (0.0000)	1.0000	
espread_e-espread_c	0.1437*** 0.0000	0.8219*** (0.0000)	0.2277*** (0.0000)	1.0000

Panel C: Correlation between Non-execution and Effective Spread: NASDAQ stocks

	nonexe_c	espread_e	nonexe_c-nonexe_e	espread_e-espread_c
nonexe_c	1.0000			
espread _e	0.3142*** (0.0000)	1.0000		
nonexe_c-nonexe_e	0.2772*** (0.0000)	0.3066*** (0.0000)	1.0000	
espread_e-espread_c	-0.0141 (0.5702)	-0.2537*** (0.0000)	-0.0855*** (0.0006)	1.0000

3.6.3. Regression Result on Non-execution

The relationship between characteristics of a stock and its transaction cost in the exchange is one of the strongest and most robust relations in finance. (Stoll, 2000 and 2003). If non-execution follows a similar pattern as price impact, we expect that trading characteristics that can explain cross-sectional variation of price impact can also explain cross-sectional variation of non-execution probability.

Therefore, we run the following regression:

$$logit\_nonexe_i = \alpha + \beta_1 logmktcap_i + \beta_2 logvol_i + \beta_3 logprice_i + \beta_4 sd_i + \beta_6 NASDAQ_i + \varepsilon_i \quad (3.6)$$

where *logit_nonexe* is the logit transformation of non-execution probability in crossing networks. *logmktcap* is average market cap for each stock. *logvol* is the log of average trading volume of each stock, *logprice* is the log of average closing price. *sd* is the standard deviation of daily stock return. Market cap, volume, price and volatility are the control variable in Madhavan (2000), Boehmer (2005) SEC (2001). The latter two papers also find that transaction cost of NASDAQ is higher than that of NYSE after he controls market cap, volume, price and volatility. Therefore, we add NASDAQ as a dummy variable.[85]

Table 3.5 shows that large stocks and stocks with higher volume have lower non-execution probabilities. This result is consistent with Hypothesis 1. We also find support for Hypothesis 2: stocks with higher volatility have higher non-execution probabilities. Interestingly, we demonstrate that stocks with higher prices have lower

[85] Some other regressions on cross-sectional variation of transaction cost added more variables. Stoll (2000), for example, also adds number of trades and the imbalance between buy and sell side to his regression. We do not include the buy and sell imbalance because non-execution should be uniquely determined by buy and sell imbalance. If we know the number of shares demanded and supplied, we know the execution probability. Number of trades suffers from endogenity issues: while number of trades can affect non-execution probability, non-execution probability affects number of trades. A high non-execution probability may increase the number of partial filled orders and increases the number of trades.

non-execution probabilities. While it is easier to explain that large stocks, frequently traded stocks and lower volatility stocks are more likely to have lower non-execution probabilities, the association between price and non-execution probability is less obvious. We believe that it is because stocks with higher prices have lower transaction costs in stock exchanges, and non-execution probability follows a similar pattern to transaction cost in exchanges because rational agents can move their trades between exchanges and crossing networks to balance the trading costs in these two markets. We also find that NASDAQ stocks have higher non-execution probabilities than NYSE stocks. Again, these variables explain of cross-sectional variation in non-execution probability, which is also a strong result.

Table 3.5: Regress Non-execution on Stock Characteristics

This table demonstrates the relationship between stock characteristics and non-execution. The sample period is from January, 2010 to March, 2010. logit_nonexe is the logit transformation of non-execution probability in crossing networks. logmktcap is average market cap for each stock. logvol is the log of average trading volume of each stock, logprice is the log of average closing price. sd is the standard deviation of daily stock return. Dummy variable NASDAQ is equal to 1 if the stock is listed in NASDAQ and 0 if the stock is listed in NYSE.

COEFFICIENT	logit_nonexe
logmktcap	-0.0414**
	(0.0211)
logvol	-0.475***
	(0.0143)
logprice	-0.0765***
	(0.0249)
sd	0.123***
	(0.0264)
NASDAQ	0.0817***
	(0.0270)
Constant	10.66***
	(0.122)
Observations	2768
R-squared	0.736

Standard errors in parentheses
*** p<0.01, **p<0.05, * p<0.1

3.7. Competition between Different Trading Platforms

There are two empirical puzzles in previous literature on the competition between crossing networks and exchanges. The focus of these two puzzles is whether or not traders face other objectives or constraints besides minimizing transaction costs (Ready, 2009). Ready (2009) finds that the market share of crossing networks does not have a monotonic relationship with the volume of the stocks. He questions why crossing networks do not have a higher market share for stocks with higher volume, which have a higher probability of finding a potential match. Ready ascribes this anomaly to soft-dollar arrangements. Ray (2010) finds that the market share of crossing networks is not higher for stocks with higher effective spreads, and Ready (2009) finds that the market share of crossing networks in fact decreases with effective spread. Ray (2010) explains that it is because people who use crossing networks have concerns about possible gaming for stocks with higher effective spreads. Ready (2009) ascribes this pattern to soft-dollar arrangements. In conclusion, both Ready (2009) and Ray (2010) consider incentives other than minimizing transaction costs as explanations for these puzzles.

Ready (2009) and Ray (2010) do not have data to measure the effective spread and non-execution in crossing networks. Therefore, their analysis relies on assumptions about transaction costs in crossing networks: both papers assume that reductions in effective spreads by trading in crossing networks increases linearly with effective spread, and neither paper has an empirical measure of non-execution. After we account for the differences in effective spreads and non-execution in crossing networks, we find that minimizing transaction costs alone is able to explain the cross-sectional variation of market shares of crossing networks.

Similar to non-execution probability, our dependent variable in market share regression, is also a proportion. Therefore, we do the logit transformation to market share. The result without logit transformation is also similar.

$$logit\_share = ln(\frac{share}{1-share}) \qquad (3.7)$$

The market share regression, however, need to be run with caution because of endogenity and measurement error issues. First, while non-execution probability and effective spread in different trading venues can affect order routing decision and market share, market share of crossing networks certainly can also affect non-execution probability and effective spread. To deal with the endogenity issue, we use the market share of crossing networks from April 2010 to June 2010 as the dependent variable, while the non-execution measure and effective spread measure are from January 2010 to March 2010. Because the execution statistics comes with a two month lag, Boehmer, Jennings and Wei (2007) use execution statistics of previous month to explain order routing decisions.

Column (1) in Table 3.6 regresses market shares of crossing networks on effective spreads in exchanges.

$$logit\_share_i = \alpha + \beta_1 espread\_e_{i,-1} + \beta_2 NASDAQ_{i,-1} + \varepsilon_i \qquad (3.8)$$ [86]

Column (2) in Table 3.6 also regresses market shares of crossing networks on effective spreads, but adds more control variable.

[86] Technically, all the lagged independent variable we generate from SEC 605 data are from the first quarter of 2010, and all the lagged independent variables generated from CRSP are from the last quarter of 2009. CRSP data for the first quarter of 2010 is not available through CRSP yearly update when we conduct this study. However, we believe that cross-sectional pattern of market cap, volume, price and listing venue would not change in three months.

$$logit\_share_i = \alpha + \beta_1 espread\_e_{i,-1} + \beta_2 logmktcap_{i,-1} + \beta_3 logvol_{i,-1} + \beta_4 logprice_{i,-1} + \beta_5 sd_{i,-1} + \beta_6 NASDAQ_{i,-1} + \varepsilon_i \quad (3.9)$$

We find a similar pattern to Ready (2009) and Buti, Rindi and Werner (2010): crossing network's market share decreases in effective spreads. However, these two regressions do not consider the fact that crossing networks also have effective spread. Also, non-execution is not in the regression. Therefore, we run the following regression with only difference in effective spread and non-execution as explanatory variable.

$$logit\_share_i = \alpha + \beta_1 d\_espread_{i,-1} + \beta_2 d\_prob_{i,-1} + \beta_3 NASDAQ_{i,-1}, + \varepsilon_i \quad (3.10)$$

The result is summarized in column (3) of Table 3.6. It demonstrates that an increase of effective spread of exchanges, relative to crossing networks, and a decrease of non-execution probability in crossing networks, relative to exchanges, decreases the market share of crossing networks. This simple regression demonstrates that the market shares of crossing networks are consistent with the incentive of cost minimization. Crossing networks attract traders when they offer relative higher reductions in effective spreads and have relatively low non-execution probabilities.

Finally, we add stock characteristics in equation (3.10) and run regression (3.11).

$$logit\_share_i = \alpha + \beta_1 d\_espread_{i,-1} + \beta_2 d\_prob_{i,-1} + \beta_3 logmktcap_{i,-1} + \beta_4 logvol_{i,-1} + \beta_5 logprice_{i,-1} + \beta_6 sd_{i,-1} + \beta_7 NASDAQ_{i,-1} + \varepsilon_i \quad (3.11)$$

Still, we find that a decrease of non-execution probability in the crossing networks increases its market share. We also find that an increase of effective spread of exchanges relative to crossing networks increases market share of crossing networks, though the effective is not statistically significant, which may be because the differences in effective spread can be explained by market cap, volume, volatility and price.

Table 3.6: Market Shares of Crossing Networks

This table demonstrates the relationship between market share of crossing networks and stock characteristics and transaction cost in both crossing networks and stock exchanges. The sample period is. logit_share is the logit transformation of market share of crossing networks from January, 2010 to March, 2010. $espread\_e_{-1}$ is the average effective spread in stock exchanges from January, 2010 to March, 2010. d_prob-$_1$ is equal to nonexeution probability in crossing networks minus Non-execution probability in exchanges from January, 2010 to March, 2010. $d\_espread_{-1}$ is equal to effective spread in exchanges minus effective spread in crossing networks from January, 2010 to March, 2010. $logmktcap_{-1}$ is average market cap for each stock. $logvol_{-1}$ is the log of average trading volume of each stock, $logprice_{-1}$ is the log of average closing price. sd-$_1$ is the standard deviation of daily stock return. Dummy variable $NASDAQ_{-1}$ is equal to 1 if the stock is listed in NASDAQ and 0 if the stock is listed in NYSE. $logmktcap_{-1}$, logvol-$_1$, sd_{-1}, $logprice_{-1}$ and $NASDAQ_{-1}$ are measured use CRSP data from October, 2009 to December, 2009.

	(1)	(2)	(3)	(4)
COEFFICIENT	Logitshare	Logitshare	Logitshare	Logitshare
$espread\_e_{-1}$	-0.0241***	-0.0146***		
	(0.00167)	(0.00191)		
Logmktcap-$_1$		0.0114		-0.0330*
		(0.0186)		(0.0186)
Sd_{-1}		-0.0591**		-0.0651***
		(0.0253)		(0.0246)
$Logvol_{-1}$		0.0759***		0.0936***
		(0.0133)		(0.0125)
$Logprice_{-1}$		0.111***		0.192***
		(0.0223)		(0.0234)
NASDAQ-$_1$	0.403***	0.587***	0.284***	0.495***
	(0.0222)	(0.0238)	(0.0220)	(0.0243)
$d\_espread_{-1}$			0.00753***	0.00410
			(0.00281)	(0.00264)
d_prob_{-1}			-0.0173***	-0.0162***
			(0.00117)	(0.00131)
Constant	-2.696***	-4.242***	-2.273***	-3.676***
	(0.0170)	(0.115)	(0.0362)	(0.127)
Observations	2740	2740	2740	2740
R-squared	0.141	0.230	0.146	0.256
Standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				

We find market shares of crossing networks are higher for stocks with higher volume, and that market shares of crossing networks are higher for high-priced stocks and NASDAQ stocks. In addition, Hypothesis 5 is rejected because we find that stocks with higher volatility have lower market shares in crossing networks, meaning that the effect of the informed trader hiding his trade (Ye, 2010) is not as significant as the effect found by Dönges and Heinemann (2006). This result is consistent with the time series pattern found by Rosenblatt Securities (2009 and 2010): that the aggregated market share of crossing networks decreases in volatility. Buti, Rindi and Werner (2010) also find the same pattern, using a different dataset from this study. On the contrary, Ready (2009) finds that stocks with higher volatility have a higher market share in the crossing network, using a sample from 2005 to 2007. One possible explanation is that now crossing networks have better anti-gaming strategies for excluding informed traders from their market (Ye, 2010). Therefore, informed trading in crossing networks now plays a less important role.

3.8. Conclusion

This paper examines non-execution and market shares of crossing networks. We verify the theoretical prediction of Ye (2010): that non-execution should follow similar patterns as price impact and non-execution increases in volatility of stocks. Non-execution also decreases in trading volume, which supports the network externality argument in Dönges and Heinemann (2006) and Hendershott and Mendelson (2000). We also find that market shares of crossing networks decreases in volatility, suggesting that the effect modeled in Ye (2010) is not as strong as the effect modeled in Dönges and Heinemann (2006). Aside from testing the empirical predictions of theoretical models, we also test whether or not the underlying assumptions in the literature (Ready (2009) and Ray (2010)) are supported by empirical data. We find that the reductions in effective spreads by trading in crossing networks are not positively correlated with effective spread. This contradiction provides alternative explanations for the puzzles found in empirical literature on competition between trading platforms.

There are several possible extensions of the paper. One interesting question to ask is whether or not non-execution follows a similar time-series pattern as price impact, that is, whether or not non-execution is higher in months when the price impact of trade is higher. Competition among different crossing networks also raises interesting questions. Trading in the crossing network certainly has network externalities. Therefore, crossing networks have a natural tendency to consolidate. Crossing networks with the largest numbers of buyers and sellers should have the highest matching probability and then attract traders from other crossing networks, which results in an even higher execution probability. In reality, there are several competing crossing networks, which can be considered "peers." We need explanations for this coexistence conundrum. The most natural explanation is that crossing networks are in the process of consolidation. The way to test that hypothesis is to see whether or not the market share of leaders continues to increase and the market share of followers decreases. The second explanation is specialization. Although there are several peer crossing networks in total trading volume, there are no such relationships at the stock level. There is a leader for each individual stock. Specialization can also be at the order level. Some crossing networks may have a comparative advantage in handling large orders, while some others specialize in small orders. We defer these questions to our future work.

The key for us in addressing questions that are not answered in the literature is the new application of SEC 605 data – this paper is the first one to use data issued by crossing networks to compute an empirical measure of transaction costs in crossing networks. As this measure is from public data, it can be easily applied to other studies. On the other hand, however, the ability for us to address questions in the literature is constrained by the availability and quality of the data. We address some questions but provide limited or no answers to other questions. For example, we find evidence that non-execution has informational and non-informational causes, but we are not able to decompose these two factors because we do not have order-level data. In fact, if we had order-level data, the first thing we would do is polish the empirical measure proposed in this paper. Non-execution is only a rough measure of transaction costs in crossing networks. We do not know the opportunity cost of unfilled orders based on the implement shortfall approach (Perold, 1988), and we cannot compare the difference between short-term alphas for filled and unfilled orders (Jeria and Sofianos, 2008). The SEC (2010) proposes a policy change on the transparency of crossing network data, and we expect new data will provide us with more insights on crossing networks.

REFERENCES

Amihud, Y., and H. Mendelson, 1980, Deallership Market. Market-Making with Inventory, Journal of Financial Economics 8, 31-53.

Bennett, P., and L. Wei, 2006, Market structure, fragmentation, and market quality, Journal of Financial Markets 9, 49-78.

Bessembinder, H., 2003, Selection biases and cross-market trading cost comparisons, Unpublished manuscript, University of Utah .

Boehmer, E., 2005, Dimensions of execution quality: Recent evidence for US equity markets, Journal of Financial Economics 78, 553-582.

Boehmer, E., R. Jennings, and L. Wei, 2007, Public disclosure and private decisions: Equity market execution quality and order routing, Review of Financial Studies 20, 315.

Choi, J. Y., D. Salandro, and K. Shastri, 2009, On the estimation of bid-ask spreads: Theory and evidence, Journal of Financial and Quantitative Analysis 23, 219-230.

Conrad, J., K. M. Johnson, and S. Wahal, 2003, Institutional trading and alternative trading systems, Journal of Financial Economics 70, 99-134.

Domowitz, I., I. FInkelshteyn, and H. Yegerman, 2009 Cul de Sacs and Highways: An Optical Tour of Dark Pool Trading Performance, The Journal Of Trading 4, 16-22.

Dönges, J., and F. Heinemann, 2006, Dark pool versus market maker Market: Unique Equilibrium in the Allocation of Order Flow. Working Paper

Degryse, H., M. Van Achter, and G. Wuyts, 2009, Dynamic order submission strategies with competition between a market maker market and a dark pool, Journal of Financial Economics.

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

Fleiss,J.L, B. Levin and M.C.Paik, Statistical Methods for Rates and Proportions, 3rd edition, (Wiley:2003)

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

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

Gresse, C., 2006, The Effect of Crossing-Network Trading on Dealer Market's Bid-Ask Spreads, European Financial Management 12, 143-160.

Hasbrouck, J., and G. Saar, 2009, Technology and liquidity provision: the blurring of traditional definitions, Journal of financial Markets 12, 143-172.

Hasbrouck, J., 2007, Empirical market microstructure: The institutions, economics and econometrics of securities trading (Oxford University Press, USA).

Ho, T., and H. Stoll, 1981, Optimal market maker pricing under transaction and return uncertainty, Journal of Financial Economics 9, 47-73.

Huang, R. D., 2002, The Quality of ECN and Nasdaq Market Maker Quotes, The Journal of Finance 57, 1285-1319.

Johnson, B. 2010, Algorithmic Trading and DMA: An introduction to direct access trading strategies, 4Myeloma Press, London

Keim, D. B., and A. Madhavan, 1998, The Cost of Institutional Equity Trades, Financial Analysts Journal 54, 50-69

Kennedy, P., 2003, A guide to econometrics (The MIT Press).

Lin, J. C., G. C. Sanger, and G. G. Booth, 1995, Trade size and components of the bid-ask spread, Review of Financial Studies 8, 1153-1183.

Madhavan, A., 2000, Market Microstructure: A Survey, Journal of Financial Markets 3, 205-258.

Næs, R., and J. A. Skjeltorp, 2003, Equity trading by institutional investors: Evidence on order submission strategies, Journal of Banking and Finance 27, 1779-1817.

O'Hara, M., 1997, Market microstructure theory (Wiley).

---. 2003, Presidential address: Liquidity and price discovery, The Journal of Finance 58, 1335-1354.

O'Hara, M., and M. Ye, Is Market Fragmentation Harming Market Quality? Forthcoming, Journal of Financial Economics

Perold, A. F., and E. C. Schulman, 1988, The free lunch in currency hedging: Implications for investment policy and performance standards, Financial Analysts Journal 44, 45-50.

Ray, S. (2010) A Match in the Dark: Understanding Crossing Network Liquidity, Working Paper

Ready, M.,(2009) Determinants of Volume in Dark Pools, Working paper

Rosenblatt Securities (2009 and 2010). Monthly Dark Liquidity Trader.

Securities and Exchange Commission, 1998, Regulation of Exchanges and Alternative Trading Systems. Release Number 34-40760, December 8.

Securities and Exchange Commission, 2009, Regulation of Non-Public Trading Interest. Release No. 34–60997; November 29.

Securities and Exchange Commission, 2010, Concept Release on Equity Market Structure. Release No. 34-61358. 14 January

Sirri, E., 2008, Keynote Speech at the SIFMA 2008 Dark Pools Symposium.

Sofianos, G. and Jeria, D. Quantifying the SIGMA X Crossing Benefit: The Sequel 2008, Journal of Trade, Fall, 84-91

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

Stoll, H. R., 2000, Friction, The Journal of Finance 55, 1479-1514.

Wooldridge, J. M., 2009, Introductory econometrics: A modern approach (South Western Cengage Learning).

Ye, M., 2010, A Glimpse into the Dark - Price Formation, Transaction Cost and Market Share of the Crossing Network. Working Paper