NYU STERN
NEW YORK UNIVERSITY · LEONARD N. STERN SCHOOL OF BUSINESS
SALOMON CENTER FOR THE STUDY OF FINANCIAL INSTITUTIONS

# Preliminary Program for the Stern Microstructure Meeting, Friday, June 1, 2012

Supporting funding is provided by NASDAQ OMX through a grant to the Salomon Center at Stern.

| | |
|---|---|
| Organizer | Joel Hasbrouck, Stern School |
| Program Committee | Tarun Chordia, Goizueta School, Emory University<br>Amit Goyal, HEC Lausanne<br>Charles Jones, Columbia University<br>Bruce Lehmann, UCSD<br>Avanidhar Subrahmanyam, UCLA |

The Stern Microstructure Conference is open to everyone with an interest in market microstructure research. The sessions will be held at the Management Education Center, 44 W. 4th St., NYC (near the southeast corner of Washington Square Park). For more complete directions see http://www.stern.nyu.edu/AboutStern/VisitStern/index.htm. There will be a registration desk in the lobby.

Registration Instructions: E-mail salomon@stern.nyu.edu with "SMC2012" in the subject line. Please indicate if you will be joining us for the dinner the night before. Other inquiries: jhasbrou@stern.nyu.edu.

Please note: Hard copies of the papers will not be available at the conference. A single pdf containing the schedule and all papers is available here.

| | |
|---|---|
| Thursday, May 31 | |
| 6:30 pm | Dinner (open to all registered conference attendees) |
| Friday, June 1 | |
| 8:30 am - 9:00 | Continental Breakfast |
| | Chair for morning sessions: Bruce Lehmann, UCSD |
| 9:00 - 10:00 | Click or Call? Auction versus Search in the Over-the-Counter Market<br>Terrence Hendershott, University of California Berkeley<br>Ananth Madhavan, BlackRock<br><br>Discussant: Larry Harris, USC |
| 10:00 - 11:00 | Liquidity: What you see is what you get?<br>Vincent van Kervel, Tilburg University<br><br>Discussant: Emiliano Pagnotta, Stern School, NYU |
| 11:00 - 11:15 | Break |

| | |
|---|---|
| 11:15 - 12:15 | CDS Auctions<br>Mikhail Chernov, London School of Economics<br>Alexander Gorbenko, London Business School<br>Igor Makarov, London Business School<br>Discussant: Raghu Sundaram, Stern School, NYU |
| 12:15-1:15 | Lunch<br>Frank Hatheway, NASDAQ OMX |
| | Chair for afternoon sessions: Avanidhar Subrahmanyam, UCLS |
| 1:15-2:15 | The Effect of Algorithmic Trading on Liquidity in the Options Market<br>Suchismita (Suchi) Mishra, Florida International University<br>Robert Daigler, Florida International University<br>Richard Holowczack, Baruch College<br><br>Discussant: Liuren Wu, Zicklin School, Baruch College |
| 2:15-3:15 | News Trading and Speed<br>Thierry Foucault, HEC Paris<br>Johan Hombert, HEC Paris<br>Ioanid Rosu, HEC Paris<br><br>Discussant: Pete Kyle, Robert H. Smith School of Business, University of Maryland |
| 3:15-3:30 | Break |
| 3:30-4:30 | How (Un)Informed Is Trading?<br>(formerly titled: Round-the-Clock News, Trading, and Stock Returns)<br>Eric Kelley, University of Arizona<br>Paul Tetlock, Columbia University<br><br>Discussant: Clara Vega, Board of Governors, Federal Reserve System |
| 4:30 | Adjourn |

# Click or Call? Auction versus Search in the Over-the-Counter Market

Terrence Hendershott*

Ananth Madhavan†

March 19, 2012

* Haas School of Business, University of California, Berkeley

† BlackRock Inc.

The views expressed here are those of the authors alone and not necessarily those of BlackRock, its officers, or directors. This note is intended to stimulate further research and is not a recommendation to trade particular securities or of any investment strategy. Information on iShares ETFs is provided strictly for illustrative purposes and should not be deemed an offer to sell or a solicitation of an offer to buy shares of any funds that are described in this presentation. We thank Hank Bessimbinder, Francis Longstaff, Kumar Venkataraman, Alex Sedgwick and Mark Coppejans for their helpful suggestions and MarketAxess® for some of the data used herein. Of course, any errors are entirely our own.

# Click or Call? Auction versus Search in the Over-the-Counter Market


## Abstract

We analyze fixed income executions in an electronic auction venue and in the over-the-counter dealer search market to toward understand the value of an electronic auction process, endogenous venue selection, and the determinants of bond trading costs. Analysis of an extensive sample of corporate bond execution data from January 2010 through April 2011 provides evidence that: (1) Traders select between dealer and auction markets based on a tradeoff between leakage about trading intentions and search costs; (2) Costs of trading are significant relative to a last trade benchmark. (3) Costs are directly related to measures of dealer competition and decline with size; (4) Dealer competition can be predicted as a function of bond attributes, trade size, past activity, and seasonalities; and (5) Competition is a strong determinant of trading cost and can explain our results in terms of trade size. The results demonstrate the value of intelligent sourcing of liquidity and suggest that electronic auctions will play an increasingly important role in the future evolution of over-the-counter markets.

# 1. Introduction

Trading in over-the-counter (OTC) markets typically involves a bilateral transaction with a dealer over the phone with limited pre- or post-trade transparency. Price negotiation with a single dealer enables a trader to control information leakage regarding their positions and intentions, but concedes some temporary bargaining power to the dealer. The advent of new trading technologies, however, allows even relatively illiquid assets to be traded in an electronic auction mechanism where a trader can simultaneously contact multiple potential counterparties. As electronic trading volumes increase across all asset classes, a market structure transition sometimes referred to as "voice to electronic," many questions remain open. How do new, auction-type mechanisms coexist with existing OTC markets, and are there natural limits to their growth? Are they preferred for certain kinds of transactions? What is the impact of the competition between voice and electronic trading on liquidity and trading costs? This paper examines these questions using unique data for all 1.8 million transactions in US investment grade corporate bonds from January 2010-April 2011.

Fixed income markets are a natural candidate for an analysis of the potential and limits for electronic auction trading. Despite being the largest asset class, most fixed income securities are typically not traded on an open exchange.[1] Further, the proliferation and complexity of fixed income instruments further complicates execution and limits the opportunity for standardized exchange trading. Indeed, relative to other asset classes such as equities, price discovery and liquidity sourcing are far more challenging. Hence, the limits of electronic trading are in the sharpest relief in these markets.

We examine traders' choice of electronic auctions versus traditional dealer trading mechanisms using data for all trades in investment grade US corporate bonds which also identify and detail the transactions executed in an electronic auction market. This feature of the data permits an analysis of mechanism choice and costs under the two structures. We show that traders are most likely to use electronic auctions when their search costs are likely high and leakage is less important. We find strong evidence that fixed income costs decrease in size and vary with proxies for risk and inventory carrying costs. We highlight the value of competition among dealers,

[1] Exceptions include fixed-income futures contracts, interdealer electronic brokerage systems, and transactions on certain electronic markets such as MarketAxess and TradeWeb, typically in liquid government issues.

showing that increased auction participation sharply reduces costs. To the extent that technology will continue to reduce search costs, these results support the view that even traditional bastions of over-the-counter trading will face strong competition from auction type markets.

The paper contributes to the growing literature on the evolution of market structure and fixed income markets in particular. Our results are consistent with Biais and Green (2007) who note that there was an active market in corporate and municipal bonds on the NYSE prior to the 1940s. They argue that the decline of exchange trading was driven by the growing importance of institutional investors, who prefer to trade in over the counter markets. That dynamic is shifting again with the advent of new, electronic trading technologies that allow traders to easily engage in multilateral trading. More practically, the results confirm the value to traders and investors from sourcing liquidity widely and using tools to optimally select venue. We show that these gains are significant and can materially influence realized investment returns.

Recent regulatory requirements have improved trade reporting, leading to a growing literature providing valuable insight into the magnitude and determinants of fixed income trading costs in OTC markets dominated by dealers. Edwards, Harris and Piwowar (2007) and Goldstein, Hotchkiss and Sirri (2007) document large transactions costs in corporate bonds, and contrary to most microstructure theories based on asymmetric information or inventory control, the costs of trade are higher for small trades and subsequently decline. Harris and Piwowar (2006) find, for example, that municipal bond trades are significantly more expensive than equivalent sized equity trades, which is surprising given that bonds are lower risk securities. One explanation may be the lack of pre-trade transparency that confers rents to dealers who possess bargaining power in bilateral trading situations. Green, Hollifield, and Schürhoff (2007) develop and estimate a structural model of bargaining between dealers and customers, and conclude that dealers exercise substantial market power.

Bessembinder, Maxwell, and Venkataraman (2007) argue that improvements in post-trade transparency associated with the implementation of the TRACE system provides market participants with better indications of true market value, allowing for a reduction in costs for those bonds included in the TRACE system. While these papers all suggest that the relatively large transactions costs facing bondholders are due to the OTC structure of the bond market, they do not provide insights into the costs in an alternative, electronic market with lower search costs. Nor do they help us understand whether the OTC market in bonds and other asset classes will evolve over time towards a more standardized, exchange traded form.

A number of empirical papers examine trading mechanism choice in financial markets both in terms of electronic versus non-electronic and searching for liquidity. Bessembinder and Kaufman (1997) examine trading costs across different stocks exchanges. Conrad, Johnson, and Wahal (2003) and Barclay, Hendershott, and McCormick (2003) analyze stock trading on electronics markets and traditional exchanges. Much of the equity market research focuses on both information flow across markets and trading costs. Barclay, Hendershott, and Kotz (2006) study the venue choice for U.S. Treasury securities between fully electronic limit order book and human voice broker intermediation. Consistent with our results that more liquid bonds trade more electronically, they find that trading moves from the electronic systems to the voice mechanisms when securities go off the run and liquidity falls. Bessembinder and Venkataraman (2004) examine large equity trades upstairs in a dealer search market versus immediate execution in the electronic limit order book. They along with others find that the electronic venue is used for easier trades.

By contrast, the auction "request for quote" mechanism analyzed here is quite distinct from the more familiar electronic communications networks (ECNs) that characterize many stock and derivative exchanges. In less liquid instruments, mechanisms that require liquidity providers to post continuous and firm quotations face challenges from adverse selection and possible tacit collusion. The auction mechanism described here mitigates these challenges and offers traders an alternative method of sourcing liquidity over the traditional dealer model. For institutional traders with many securities in their trading list, it is difficult to simultaneously call many counterparties in a short interval of time. By contrast, the request for quote auction mechanism allows traders to reach multiple counterparties simultaneously and avoid prolonged negotiations by setting a time limit for the auction. Since the trader can reveal how many other dealers are being queried, this can be a credible way induce dealers to improve pricing. Indeed, the results show that electronic auction markets based on a competitive sealed bid process are a viable and important source of liquidity even in inactively traded instruments.

Our results augment our knowledge of trading costs in fixed income markets and our understanding of the value of the competition inherent in an auction setting. The evolution of bilateral, sequential trading into an auction type framework offers a path from an over-the-counter market to centralized, continuous trading.

The paper proceeds as follows: Section 2 provides an overview of the relevant institutional detail and summarizes our data sources and procedures; Section 3 examines overall trading costs in the corporate bond market as a whole; Section 4 analyzes endogenous mechanism choice and provides a framework to analyze the data; Section 5 contains our empirical results on bidding and trading behavior including estimates in the auction mechanism; and Section 6 concludes and offers some recommendations for public policy.

## 2. Institutions and Data

The data comprise all investment-grade corporate bond trades in Financial Industry Regulatory Authority's (FINRA) Trade Reporting and Compliance Engine (TRACE) from the beginning of 2010 through April 2011, a total of 1.8 million transactions. We augment the TRACE data with specific details on all trades directed to an electronic auction venue, MarketAxess, in the period. MarketAxess was formed in April 2000 to provide institutional investors an electronic trading platform with access to multi-dealer competitive pricing in U.S. high-grade corporate bonds, Eurobonds, emerging markets, high yield, credit default swaps (CDS) and U.S. agency securities. Other electronic venues – most importantly TradeWeb – are also used by fixed income traders, primarily for transactions in treasury bonds where costs are typically very low.

Access to MarketAxess typically involves the fixed costs of integration into an institutions order management system. Once integration is complete traders can simply click to send their orders to a number of dealers. Institutional clients select the number of dealers queried for a request for quote in a bond for a given size trade, providing considerable efficiencies in terms of time relative to the alternative of a sequence of bilateral negotiations. Typically dealers are aware of the number of other dealers queried and the identity of the institutional client. A time period is specified for the quotes to be submitted, also encouraging competitive behavior. At the end of the "bidding" period all dealer quotes are revealed to the client. If satisfied with the responses the client selects the best quote and the trade is executed. Essentially it operates as a sealed bid auction. MarketAxess trades with client prices are reported in TRACE. MarketAxess charges dealers a fee between 0.1 and 0.5 basis points for investment grade bonds.

The TRACE data are comprehensive and indicate the size of the trade and whether each trade is between two dealers or buyer- or seller-initiated with a dealer. This latter feature is relatively new starting in late 2008. Of special interest, the MarketAxess data are unique in several respects in that they identify the number of dealers queried and those that respond. We focus only on bond trades in standard, e.g., non-callable, U.S. investment grade assets, excluding trades in agencies, treasuries, TIPS and mortgages.

Summary statistics are contained in Table 1. While all MarketAxess trades are included in TRACE, for ease of exposition we refer to the non-MarketAxess trades simply as TRACE trades. Interdealer trades in TRACE are also excluded. MarketAxess is designed to facilitate client to dealer transactions, so interdealer trades do not occur on MarketAxess.

[ Insert Table 1 Here]

The total TRACE sample (excluding MarketAxess transactions) is approximately 1.6 million transactions in 4,129 different bonds. By comparison 191,150 transactions were in MarketAxess in 1,579 bonds. The trade size categories in Table 1 are based on dollar value traded and are chosen in accordance with standard industry conventions. The majority of these trades are micro lots (77%) defined as below $100,000 in value. Although we will examine this point in more detail later, some size differences are already apparent in Table 1 between the auction mechanism and the over the counter market. There is a much higher concentration of odd lot ($100,000-$1 Million) sized trades on MarketAxess and it also appears that there are fewer large transactions. Overall, the share of the auction market in overall bond trading is 6.5% in micro (1-100K), 26.1% in odd lot (100K-1M), 21.9% in round lot (1M-5M) and 3.6% of the maximum reported size (5M+) trades.[2] MarketAxess' small market share in micro sized trades could result from smaller retail oriented traders not having access to the platform. This could also explain why the average trade size is larger on MarketAxess for odd and round lot trades. MarketAxess' lower market share in the largest trades likely arises from differences in the trading mechanisms.

Trading occurs in MarketAxess in 1,579 distinct bonds. The average bond characteristics of trades on and off MarketAxess differ noticeably. Younger bonds and bonds with larger issuance size are likely to be more liquid; see Harris and Piwowar (2006) and Edwards, Harris and Piwowar (2007) for evidence of this for the municipal and corporate bonds markets. The average bond age is smaller and the bond issuance size is larger for MarketAxess trades suggesting that the electronic platform may be more effective for easier trades. Table 1 also includes a standard risk measure of the bond's yield spread over treasury times duration. Our proxy for duration is years to maturity. As with issue size and age MarketAxess trades are more prevalent in less risky bonds which are expected to be more liquid. It should be noted that there are likely

[2] Trades greater than five million bonds are reported as five million for investment grade corporate bonds. .

some differences in the clienteles across systems, although most large institutions will likely use both venues. In general, because of the set up costs and sophistication required, we would expect a narrower pool of users for MarketAxess. We address questions of endogenous venue choice later in the paper.

## 3. Market-wide Transactions Costs

A number of approaches have been used to calculate trading costs in sparsely traded fixed income markets. Unlike equity markets where continuous bid and ask quotations are available, corporate bond markets only report transactions. The simplest approach is to compare buy and sell prices of the same bond around the same time to create an imputed spread. As the TRACE data identify whether a transaction is buyer- or seller-initiated, it is possible to compute the imputed spreads straightforwardly. Hong and Warga (2000) follow this approach to estimate what Harris and Piwowar (2006) refer to as a benchmark methodology by subtracting the average price for all sell transactions from the average buy price for each bond each day when there is both a buy and a sell. Imputed spreads, although simple, have some deficiencies. Given the infrequency of trading for many bonds, the same bond, same day criterion limits the amount of usable data. Further, although we know transaction side, trades may be initiated based on market moves (either contrarian or momentum) confounding the estimates. Harris and Piwowar (2006) use a regression approach to utilize more data. We adapt the Harris and Piwowar approach to allow us to more closely follow the transaction cost literature in the equity markets and to more easily include cross-sectional and time series covariates in our cost estimation. In this section we use both the imputed (benchmark) and regressions methodologies.

### 3.1 Imputed Half-Spreads

Table 2 calculates different between the average price for all sell transactions and the average buy price for each bond each day when there is both a buy and a sell. This cost is in basis point for MarketAxess and TRACE for each of the trade size category in Table 1. We divide the different between buy and sell prices by two to measure the one way transaction cost often referred to as half-spreads.

Panel A shows the standard decline in trading costs with trade size. Panel A also reveals substantial costs differences between MarketAxess and TRACE. For odd lot trades MarketAxess averages 8.54 basis points while for TRACE trades the cost is 32.43 basis points, substantially higher. The costs for TRACE fall to 9.12 and 6.39 in the round and maximum trade size categories, respectively. MarketAxess costs fall with trade size also, but more slowly. These TRACE costs are similar in magnitude to previous corporate bonds cost estimates.

[ Insert Table 2 Here]

While MarketAxess costs are lower than TRACE, Table 1 shows that the characteristics of bonds traded via MarketAxess and TRACE differ with bonds likely to be more liquid, e.g., bonds with larger issue sizes, trading more on MarketAxess. An initial attempt to control for differences in the types of bonds is in Panel B of Table 2 where half-spreads are calculated only for bonds traded via both MarketAxess and TRACE. The MarketAxess costs in Panel B are very similar to those in Panel A because virtually all the bonds traded on MarketAxess are also traded on TRACE. The TRACE costs generally are lower in Panel B, but the declines are modest and costs remain higher than in MarketAxess.

Limiting the analysis to only bonds traded on both MarketAxess and TRACE controls for cross-sectional differences in bond characteristics. However, it is still possible that bonds are more likely to trade on MarketAxess on days where liquidity in those bonds in higher. Panel C of Table 2 controls for this by only examining on days where costs for a bond can be calculated for both MarketAxess and TRACE. The costs and cost differences are not greatly affected by further narrowing the sample, but the number of observations falls substantially, especially in the larger trade sizes. A natural approach to controlling for bond characteristics and market conditions is in a multivariate regression framework.

## 3.2 Regression Cost Estimates

To control for bonds characteristics, market conditions, and trade characteristics it is useful to construct a cost of transacting for each trade. Typically, trading costs are defined the difference in return between the actual investment return and the return to a notional or paper portfolio. Costs include commissions, market impact, and opportunity costs. One important unobservable opportunity cost for OTC markets is the time traders spend searching for a counterparty.

We compute percentage transaction costs in basis points relative to a variety of benchmarks but focus here on the last trade in that bond in the interdealer market as most representative. Then the cost or implementation shortfall) is defined as:

$$Cost = ln\left(\frac{TradePrice}{BenchmarkPrice}\right) \times Trade\ Sign \tag{1}$$

Trade sign is the standard buy/sell indicator which is +1 if the client is buying and -1 if the client is selling. It is important to note that cost in (1) is a fraction of trade value, not yield. The trading convention for high-grade corporate bonds is typically for negotiations to take place in terms of yield relative to the yield on a benchmark treasury of similar duration rather than dollar value. However, from the investment perspective, cost or shortfall should be expressed relative to the value of the trade. We compute transactions costs throughout in basis points of value by multiplying equation (1) by 10,000.

Harris and Piwowar (2006), Bessembinder, Maxwell, and Venkataraman (2006), and Edwards, Harris and Piwowar (2007) calculate trading costs with regressions of the change in price between transaction on change in the trade sign. Consistent with our approach interdealer trades are given a sign of zero. Our approach uniquely assigns a cost to each transaction. This allows for more straightforward inclusion of transaction specific covariates of interest, in particular, later when analyzing MarketAxess trades the inclusion of the number of dealers queried and number responding. The disadvantage of calculating a cost for each trade is the information in trade signs and price changes is less fully exploited. In addition, the interdealer price may include costs dealers charge each other for trading. Our baseline regression transaction costs estimates are similar to the costs in Table 2 which do not use any interdealer trades. In unreported results we obtain similar trading costs estimates using the Harris and Piwowar (2006) regression approach that utilizes both prior customer and dealer trade prices as a benchmark.

For cost computations, we favor using the last interdealer price as the benchmark price as this is a good proxy for the mid-quote. Using the last price introduces a bias from bid-ask bounce. Another popular benchmark, the Volume Weighted Average Price (VWAP) suffers from problems of relevance when there are few trades in the day.

### 3.3 Determinants of Costs

Table 3 analyzes trading costs for TRACE and MarketAxess while controlling for market conditions, bond characteristics, and trade size. Before presenting the results we discuss our control variables in detail.

*Bond characteristics:*

- Credit quality – dummy variables for A and B rated bonds.
- Maturity – natural logarithm of the time until the bond matures
- Age – natural logarithm of the time since the bond was issued
- Issue size – natural logarithm of the bond's issue size bond

- Industry sector – dummy variables for whether the bond issuer is in the financial, industrial, or utility sectors.

*Market conditions:*

- Risk term: DTS – the yield spread over treasury times years to maturity.
- Drift term: – change in treasury yield relative to benchmark trade × buy-sell indicator × years to maturity; the controls changes in price due to treasury rate shifts.
- Calendar Time Controls – beginning and ending of week dummy, 1 for Friday and Monday, 0 otherwise; month end dummy, 1 for last trading day of the month, 0 otherwise.

*Trade characteristics:*

- Trade size: Micro, Odd, Round, and Max dummy variables for trades of less than 100,000, 100,000 to 999,999, 1M to less than 5M, and 5M and above, respectively.
- MA – dummy equal to 1 if the trade is on MarketAxess and 0 otherwise.
- MA Micro, MA Odd, MA Round, MA Max – MA dummy variable interacted with trade size dummy variables.

[ Insert Table 3 Here]

Table 3 presents three transaction cost regressions. Standard errors control for contemporaneous correlation across bonds on the same day and time series correlation within bond using the clustering approach of Petersen (2009) and Thompson (2011). The first regression's independent variables are the trade size dummy variables, the MarketAxess interacted trade size dummy variables, the calendar time dummy variables, the rating and industry dummy variables, and the treasury drift variable. All independent variables are demeaned so the trade size dummy variables can be added together to calculate average trading costs.

The constant of 78.38 represents the cost of a micro-sized TRACE trade and is close to the 70.58 cost for a micro trade in Table 2. That is a substantial trade cost relative to small trades in other asset classes. The -32.05 coefficient on the odd-lot dummy variable shows the cost for a TRACE odd-lot trade is 32.05 basis points lower than for a TRACE micro trade; making the TRACE odd-lot cost 46.33 basis points, somewhat higher than the 32.43 basis points in Table 2. Round lot TRACE costs are 20.75 basis points ($= 78.38 - 57.63$). Maximum sized TRACE trades are 19.19 basis points. The cost estimates for TRACE odd, round, and maximum trade sizes being higher in Table 3 as compared to Table 2 could arise from buys and sells occurring in the same day in bonds that are more liquid or on days that are more liquid. So, costs decline with trade size but in a non-linear manner. Beyond a round lot, there appears to be little difference in cost as a function of size.

The MarketAxess interacted trade size dummy variables capture the difference in trading costs between MarketAxess and TRACE for that trade size. For odd-lot, round-lot, and maximum trade sizes costs on MarketAxess are 27.27, 4.50, and 2.80 basis points lower than TRACE, respectively. The differences are comparable to the differences in the simple trading costs measures in Table 2.

The second regression specification in Table 3 adds the bond characteristics for time to maturity, age since issuance, and issuance size as independent variables. The coefficients on these are consistent with Edwards, Harris and Piwowar's (2007) findings that older, longer maturety, small issue bonds have higher transactions costs. Adding these controls increase the costs difference between MarketAxess and TRACE for micro and round trade sizes. For the maximum trade size the relative cost difference between MarketAxess and TRACE becomes positive, 1.81 basis points, but is not statistically significant. For larger sizes client traders have an incentive to invest more time in negotiating better prices over the phone with dealers. As noted above, such costs are not observable in prices. In contrast the time required for an auction in MarketAxess is independent of trade size. Thus, the 5M+ result is consistent with a search model where for large enough sizes the unobserved search effort for TRACE trades increase to offset the gains from lower search costs on MarketAxess. Similarly, this could lead to MarketAxess having smaller trades in the 5M+ category, complicating the cost comparison for the largest trade size.

The third regression specification in Table 3 adds the risk variable as an independent variable. Higher yield bonds with longer duration represent greater inventory risk for dealers. Surprisingly, after controlling for other bond and market characteristics, the risk variable is not statistically significant. The calendar time, industry sector, and rating dummy variables are often not statistically significant so these variables are omitted in some subsequent specifications.

## 4. Endogenous Venue Choice

To move beyond our largely exploratory empirical investigation thus far it is useful to build a more explicit framework to analyze endogenous venue choice. We focus on the tradeoff between the lower search costs a trader enjoys by using an electronic auction mechanism, where a trader can simultaneously request multiple quotes, and the anonymity benefits from bilateral negotiation in the over the counter market. Let $C_0$ denote trader's the expected cost from transacting bilaterally in the OTC market and let $C_m$ denote the expected cost if the trader were instead to select the auction mechanism. The trader selects the auction market if and only if the costs of transacting there are lower than the OTC alternative, i.e., $C_0 > C_m$.

Consider a potential buy order of size $x > 0$. We model the expected price in bilateral OTC trading as the expected value of the asset $v$ plus a premium $C_0 = v + d(x)$, where $d$ is the expected dealer markup which could depend on trade size $x$ and is trader specific, depending on relative bargaining power.[3] From an econometric viewpoint, the relative bargaining position of a trader is unobserved and will create selectivity bias.

Now consider the cost of trading in the auction mechanism. The trader can conduct an auction by simultaneously selecting $M$ dealers to contact, up to a maximum of $\overline{M}$. The choice of $M$ is itself endogenous and will depend also on trade size: Contacting more dealers implies a higher likelihood of responses and hence lower costs but involves additional leakage of information. We assume the number of dealers to query, conditional upon selecting the auction mechanism, is optimally selected and let $M(x)$ denote this function.[4] Let $N$ be the (random) number of dealers responding given that $M(x)$ dealers are queried. We model $E[N] = \lambda(z)$ where $\lambda$ is the hazard rate that depends on a vector of bond and trade characteristics, $z$. In our later empirical analysis, we will estimate this function assuming a Poisson model for responses. We assume the auction fails (i.e., that no dealers respond) with probability $q = \Pr[N = 0] = e^{-\lambda}$. In this case, the trader is forced to go to the bilateral dealer mechanism and incurs additional costs $s(x) \geq 0$ relative to the OTC market. The additional cost has two potential components: First, given $M$ dealers were contacted, there is potential additional leakage of information which is likely positively related to size because larger trades may be more likely to be associated with private information. Even if the trade were not information motivated, knowledge of flows may lead to front running and hence there could be a leakage cost just associated with the fact that a large buyer is in the market. So that in the event there are zero responses, the total purchase cost is $C_0 + s(x)$, where $C_0$ is the expected cost in the bilateral market. Implicit in this framework is the time dimension; the multilateral auction may take longer to run even though it involves a single click than a bilateral negotiation on the phone. Thus, $s$ captures the additional

[3] For simplicity we assume that if the trader engages in bilateral trading, execution will be obtained with certainty albeit at a cost, and hence there is no leakage cost. It is straightforward to extend the model to allow for a positive probability that the negotiations lead to no trade followed by subsequent searches in the future. See Duffie, Garleanu, and Pedersen (2005) for a fully developed sequential search and trading OTC model.

[4] Levin and Smith (1994) examine entry incentives in auctions with stochastic numbers of entrants. They show that in a common value auction (in our setting the common value is the common inventory component across dealers) a seller wants to limit the number of bidders even in the absence of leakage costs.

leakage (over any in the OTC market) result. The term , $s$ is hence trader or trade specific and is not observed by the econometrician.

If $N > 0$, with probability $1 - q$, the auction is viable. Rational dealer behavior suggests that in a sealed bid auction, each dealer will charge their reservation price which is the expect value ($v$) plus a dealer specific inventory based premium which depends on size. A positive premium can arise as a compensation for unwanted risk or if the dealer ascribes a positive probability to being in a monopolistic position should no other dealers respond. Alternatively, a dealer who has an opposite side inventory position may aggressively bid to reduce risk in which case the premium can be negative. We expect there to be cross-sectional dispersion in initial inventory across dealers. The expected total auction cost, $C_m$, for trade size $x$ combines the cost of a failed auction with the costs of a successful auction:

$$C_m = q(C_0 + s(x)) + (1 - q)(v + p(x)) \tag{2}$$

where $p(x)$ is the expected premium. This yields the choice model; a trader chooses the OTC market ($OTC = 1$) if and only if:

$$qs(x) + (1 - q)\big(p(x) - d(x)\big) > 0 \tag{3}$$

For a given size $x$, the higher the costs of leakage $s$, the greater likelihood of searching the dealers over the auction, and vice versa with the dealer markup $d$. Similarly, lower dealer response rates λ (e.g., on less liquid issues) implies a higher $q$ and hence less likelihood of selecting the auction.

The model also shows that costs can vary with size in a nonlinear way. Unlike asymmetric information models, realized cost $C(x)$ will reflect the optimal choice of venue plus the tradeoff between the cost elements:

$$C(x) = min[C_0(x|OTC = 1),\ C_m(x|OTC = 0)]\,. \tag{4}$$

This equation forms the basis for the endogenous choice model we estimate below. For small sizes, leakage $s$ is likely to be minimal and the auction mechanism dominates. This is also the case if dealers are competitive in bidding so that $p(x) = 0$. Beyond a point, as trade size rises, we would expect higher costs from leakage and hence more likelihood of using a dealer mechanism. If the dealer markup $d(x)$ reflects economies of scale or bargaining power, we may observe that realized cost declines with $x$.

## 4.1 Regression Cost Estimates Controlling for Selection Bias

The multivariate regressions in Table 3 control for observable differences in bond characteristics and market conditions. However, there may be unobservable characteristics of trades which affect both the costs of the trade and whether the trade is executed on MarketAxess. The standard econometric approach is an endogenous Heckman switching model.[5] This is a two stage model. In stage 1, a trader chooses venue 1 (auction) over venue 2 (dealer phone search) if he believes that the expected cost of venue 1 are lower than those of venue 2. We model expected costs as

$$c_k = z'\delta_k + \eta_k . \tag{5}$$

Where $z$ is a vector of explanatory terms (including possibly non-linear functions of size) and the error term $\eta$ captures the unobserved costs of search and slippage, as detailed in the model. The auction venue is chosen if

$$c_1 \le c_2 \quad \text{or} \quad z'(\delta_2 - \delta_1) + (\eta_2 - \eta_1) \ge 0 \quad \text{or} \quad z'\delta + \eta \ge 0 , \tag{6}$$

which forms the basis of the Probit equation in Table 4.

In stage 2, we model the true cost equation (if there was no selection bias) as

$$y_k = x'\beta_k + \varepsilon_k . \tag{7}$$

The residual captures unobserved cost factors such as dealer inventory effects. Assuming joint normality:

$$E[y_1 \mid x, z, c_1 \le c_2] = x'\beta_1 + \rho_1 \sigma_{\varepsilon_1} \frac{\varphi(z'\delta)}{\Phi(z'\delta)} \quad = x'\beta_1 + \rho_1 \sigma_{\varepsilon_1} mr(z'\delta) \tag{8}$$

$$E[y_2 \mid x, z, c_1 \ge c_2] = x'\beta_2 - \rho_2 \sigma_{\varepsilon_2} \frac{\varphi(z'\delta)}{1 - \Phi(z'\delta)} = x'\beta_2 - \rho_2 \sigma_{\varepsilon_2} \overline{mr}(z'\delta) \tag{9}$$

where $\rho_k$ is the correlation of $\varepsilon_k$ and $\eta_k$. Essentially to run both regressions requires two Mill's ratio variables (with appropriate dummies) where the denominators are slightly different (they sum to one).

We estimate the probit model for venue selection in Table 4 with the same three specifications in Table 3. Consistent with Table 1 odd- and round-lot trades and bonds with larger issue sizes are more likely to be on MarketAxess. Unlike in Table 3 some of the bond characteristic and calendar time dummy variables are significant. A-rated bonds are more likely to trade on

[5] Madhavan and Cheng (1997) and Bessembinder and Venkataraman (2004) use the procedure to estimate costs for block trades while controlling for selection.

MarketAxess. Trades on Monday and at the end of the month are more likely to be on MarketAxess.

[ Insert Table 4 Here]

Table 5 uses the third probit specification to estimate the second-stage cost model for MarketAxess and TRACE. As before all continuous independent variables are demeaned. The selectivity adjustment (Mill's ratio) terms are *Inv Mill MA* and *Inv Mill TRACE*, respectively. The difference in independent variable coefficients (MarketAxess minus TRACE) shows that MA's relative costs decrease in issue size and increase in age and duration times spread, consistent with easier trades being done on MA. The inverse mills ratio has a negative coefficient, is consistent with our model of selection and where the auction market is chosen for orders with less likelihood of leakage and higher search costs.

[ Insert Table 5 Here]

The results in Table 5 are consistent with MarketAxess being cheaper for average trades. The differences in coefficient estimate across the columns illustrate how much the independent variable must change for the cost differential between MarketAxess and TRACE to change sign. For example, leaving aside the selectivity adjustment (Mill's ratio) terms an average maximum size trade on MarketAxess costs 12.93 basis points ($= 19.14 - 6.21$) as compared to 17.23 basis points ($= 78.87 - 61.64$) on TRACE. The difference in the issue size coefficients is $-6.117 + 2 = -4.117$. The ratio of the costs difference to coefficient difference is $1.04 = -4.3/-4.117$. Because issue size is in natural logarithm units this translates into issue size differing by a factor of 2.83 ($= e^{1.04}$). Put another way, a maximum sized trade in an otherwise average bond would need to have an issue size less than one-third the average issue size, e.g., \$600M as compared to \$1.8B, to be cheaper on TRACE than MarketAxess.

Also noteworthy are the cost differentials for odd lot trades. Very clearly, the auction mechanism is preferred for this size over the OTC alternative. Given the relative size of the coefficients, we would need the slippage term $s$ or the probability of non-trading $q$ to be implausibly large to explain this selection. It is likely that this result reflects some of the differences in the client composition across the venues referred to earlier. Specifically, the results may reflect the fact smaller and less active traders who could benefit from trading their odd lots in an auction framework are unwilling to bear the associated set up costs, closing out this option.

Finally, we have not considered the 30% or so of the time that a MarketAxess auction does not result in a trade. However, we also do not observe phone searches that do not result in trades. In general, to fully capture expected trading costs across venues one needs to run an experiment where orders are randomly sent to different mechanisms. This is a shortcoming of all studies of observed transaction costs.

## 5. Trading and Bidding Behavior in Electronic Auctions

Theory suggests that bidder's behavior is crucial for auction performance (for example, Bulow and Klemperer (2009)). We next turn to the detailed data we have on dealer's bidding behavior in MarketAxess. Figure 1 illustrates the frequency distribution of the number of dealer responses in the auction market, conditional upon at least one response. The modal response is 3. With that number of dealers, the typical auction should approach a competitive outcome. In 7.3 percent of auctions no dealers respond.

[ Insert Figure 1 Here]

Table 6 provides information by trade size on the per auction average number of dealers queried, the percentage of dealers responding, and the percentage of auctions with zero responses. The number of dealer's queried decreases with trade size with 27.70 dealers queried for micro trades and 23.98 queried for maximum trades. This is consistent with a concern by traders about leakage of their intentions increasing in trade size. Despite this the number of dealers responding increases in trade size. Queries decreasing in trade size while dealer responses increasing indicates the dealers are more likely to respond for large trades. This would arise if dealers' costs of participation is fixed, making the participation cost per bond falling in trade size. The percentage of dealers responding increases substantially from 16.8% for micro trades to 29.7% for large trades. An alternative explanation is that large trades are simply done in bonds where dealers are more likely to respond.

[ Insert Table 6 Here]

To control for bond characteristics requires a more formal model of the number of dealers $N$ responding to $M$ queries, i.e., $E[N|M] = \lambda(z)$, We model the number of dealers who respond to a trader's queries using a count data model (as $N = 0,1,2\ldots$) which also naturally allows for zero outcomes or auction failure. It is important to allow for cross-sectional heterogeneity which we expect given that dealer bargaining power and perceptions of leakage will vary across traders. We model this for trade *i* as the outcome of a Poisson distribution with conditional mean:

$$\ln(\lambda_i) = z_i'\beta + u_i \tag{10}$$

where the error term $u_i$ captures individual (unobserved) variation in dealer responses. When $u_i$ has a gamma distribution $\Gamma(1,\theta)$, this yields a negative binomial model. Unlike the Poisson model, we do not restrict the mean and variance of the sample data to be equal. Over dispersion (where the variance exceeds the mean) is quite common with count data and hence the negative binomial is preferred. The distribution of the number of dealer responses $N_i$ in auction *i* conditioned on $z_i$ is

$$P(N_i = n_i \mid z_i) = \frac{\Gamma(n_i+\theta)}{\Gamma(\theta)\Gamma(n_i+1)}\left(\frac{\theta}{\theta+\lambda_i}\right)^{\theta}\left(\frac{\lambda_i}{\theta+\lambda_i}\right)^{n_i} \qquad n_i = 0,1,2,\cdots \tag{11}$$

Table 7 reports estimates of three versions of the dealer response model.[6] Trade size and issue size are positive predictors.

[ Insert Table 7 Here]

The model can help traders better predict auction interest to utilize this mechanism more efficiently and reduce costs. Using equation (1), the probability of the auction failing with no responses at all is $P(N_i = 0 \mid z_i) = \left(\frac{\theta}{\theta+\lambda_i}\right)^{\theta}$, which is the analog to the Poisson probability with no heterogeneity in responses. There is a marked jump in the probability of auction failure as trade size changes from a round to an odd lot. Depending on the model, the difference in the odd and round lot coefficients is approximately 0.2 implying that, all else equal, the probability of auction failure for an odd lot transaction is 1.22 times that of a comparable round lot. Risk is a negative predictor along with age and the end of week and end of month dummy variables. There are fewer dealers in higher yield bonds, which makes sense given dealer risk aversion. It is also interesting to note that the coefficient on the log number of dealers is less than one, which corresponds to the probability of dealers responding declining in the number of dealers queried.

Auction theory predicts that the number of dealer should be closely linked to the auction outcome. Figure 2 shows the costs in basis points as a function of the number of responding

[6] In order to utilize the results from our response model in a trading cost regression model 3 is estimated only for auctions resulting in trades. Approximately two percent of trades cannot be matched to auctions causing the number of observations in model 3 of Table 7 to be slightly less than the number of MarketAxess trades in Table 1.

dealers. Competition lowers costs as is clear from the figure. When only a few dealers respond, costs are high, 20-35 basis points. Observe that mean realized costs go to zero and actually become slightly negative (-3 to -4 basis points) when the number of responses is large, in this case above 10. While estimation error could lead to negative realized costs, recall from the model that this result is consistent with a negative premium $p(x)$, i.e., that some dealers are willing to price aggressively to liquidate unwanted inventory. It is worth noting that this does not mean that the winning dealer loses on the round trip in expectation as they may be able to charge a premium to enter into the position initially.

[ Insert Figure 2 Here]

Table 8 presents cost estimates for the auction market. Model 2 shows that costs declining in the number of dealers responding in Figure 2 is robust to the inclusion of bond and trade characteristics. Essentially, this can be thought of as a regression with a log of frequency response as an explanatory variable. We use the natural logarithm of the bids because Figure 2 shows a clear nonlinearity in costs as a function of bids. Model decomposes the number of bidders into the expected and unexpected number of bidders using specification 3 of the negative binomial dealer response model in Table 7. The logarithm of expected and unexpected responses is taken after subtracting the overall minimum and adding one; this ensures that the minimum of both the logarithm of expected and unexpected responses is zero. The coefficient on the unexpected number of bidders is significantly more negative than the coefficient on the expected number showing that unexpectedly fewer bidders responding is particularly costly.

It is interesting that conditioning on the number of bids in model 2 reverses the ordering of costs for odd, round, and maximum trade sizes with the costs of trading no longer monotonically decreasing in trade size. This difference demonstrates that dealer bidding behavior drives the decline costs in trade size. This contrasts with the standard argument that costs decreasing in trade size is due to market power by intermediaries where the bargaining power the trader in increasing in trade size.[7]

[ Insert Table 8 Here]

---

[7] Bernhardt, Dvoracek, Hughson, and Werner (2005) provide a variant on this argument by modeling the repeated interaction between customers and dealers. They find that dealers offer better prices to more regular customers, and, in turn, these customers optimally choose to submit larger order.

Note that Model (1) and (2) in Table 7 include auctions where there was no trade. Model (3) is only for trades. The number of observations in Model 3 is slightly smaller than in Table 1 because some of the trades could not be matched to auctions, 191,150 in Tables 1 and 5 versus 187,834 in Tables 7 and 8.

It would be interesting to extend this analysis to capture the time dimension of trading (i.e., the duration between order initiation and execution), which as we noted earlier is a factor in determining the expected slippage and hence the trader's strategy. While this statistic is not in the TRACE data, in theory we could look at this for the MA data. It would, for example, also be interesting to examine to link the duration statistic to leakage and also examine dealer behavior in time, i.e., whether some dealers always wait until the end of the allotted auction period or respond right away. These more detailed questions are topics for further analysis.

## 6. Conclusions

As electronic trading expands more rapidly across regions and asset classes, it is natural to ask whether there are limits to the potential scope of auction markets. The fixed income markets are of particular interest given their size and the complexities of the instruments traded. Trading in this asset class still remains very much over-the-counter although electronic auctions – as in equities and derivatives markets – are gaining traction.

Using an extensive sample of corporate bond transactions from January 2010 to April 2011 we show that: (1) There is clear evidence that traders rationally select between dealer and auction markets based on a tradeoff between leakage about trading intentions and search costs; (2) Costs of trading are significant relative to a last interdealer trade benchmark. The analysis also confirms that trading costs decreases as a function of size and provides an explanation for this result; (3) Costs are directly related to measures of dealer competition and trade size; (4) Dealer competition can be predicted as a function of bond attributes, trade size, past activity, and seasonalities; and (5) Competition is a strong determinant of trading cost. There are significant gains to sourcing multiple bids for fixed income transactions. Competition is greater for larger trade sizes, providing an explanation for why costs might decline with size.

From a public policy perspective, the results here shed light on trading costs in fixed income markets and increase our understanding of the value of the competition inherent in an auction setting. It is important to understand that the evolution of bilateral, sequential trading into an auction type framework offers a path from an over-the-counter market to a centralized, continuous trading. The mechanism analyzed here is quite distinct from the more familiar electronic communications networks (ECNs) that many stock and derivative exchanges have evolved into over the past two decades. The auction "request for quote" market made possible by technological advances is a very different system altogether and may offer a way to mitigate the complexities of markets where liquidity providers post their quotes such as the original Nasdaq Automated Quotation System. In less liquid instruments, such markets pose problems of potential collusion among dealers who observe each other's actions or adverse selection from offering free options in the form of continuous and firm quotes. Our results indicate that electronic auction markets based on "sealed bids" are a viable and important source of liquidity even in inactively traded instruments.

## References

Barclay, M., T. Hendershott, and K. Kotz, 2006, "Automation versus intermediation: Evidence from Treasuries going off the run," *Journal of Finance* 61, 2395-2414.

Barclay, M., T. Hendershott, and T. McCormick, 2003, "Automation versus intermediation: Evidence from Treasuries going off the run," *Journal of Finance* 58, 2637-2666.

Bernhardt, D., V. Dvoracek, E. Hughson, and I. Werner, 2005, "Why do larger orders receive discounts on the London Stock Exchange?" *Review of Financial Studies* 18, 1343-1368.

Bessembinder, H., and H. 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., W. Maxwell, and K. Venkataraman, 2006, "Market Transparency, Liquidity Externalities, and Institutional Trading Costs in Corporate Bonds," *Journal of Financial Economics* 82, 251-288.

Bessembinder, H., and K. Venkataraman, 2004, "Does an Electronic Stock Exchange need an Upstairs Market?" *Journal of Financial Economics* 73, 3-36.

Bulow, J., and P. Klemperer, 2009, "Why do Sellers (Usually) Prefer Auctions?" *American Economic Review* 99, 1544-1575.

Biais, B. and R. Green, 2007, "The Microstructure of the Bond Market in the 20th Century," Working paper, Carnegie Mellon University.

Chakravarty, S., and A. Sarkar, 2003, ''Trading costs in three U.S. bond markets,'' *Journal of Fixed Income* 13, 39-48.

Conrad, J., K. Johnson, and S. Wahal, 2003, "Institutional Trading and Alternative Trading System," *Journal of Financial Economics* 70, 99-134.

Duffie, D., N. Garleanu, and L. Pedersen, 2005, "Over-the-Counter Markets," *Econometrica* 73, 1815-1847.

Edwards, A., L. Harris, and M. Piwowar. 2007, "Corporate Bond Market Transparency and Transactions Costs," *Journal of Finance* 62, 1421–51.

Goldstein, M., Hotchkiss, E., Sirri, E., 2007, "Transparency and Liquidity: A Controlled Experiment on Corporate Bonds," *Review of Financial Studies* 20, 235-273.

Green, R., B. Hollifield, and N. Schürhoff, 2007, "Financial Intermediation and the Costs of Trading in an Opaque Market," *Review of Financial Studies* 20, 275-314.

Green, R., B. Hollifield, and N. Schürhoff, 2007, "Dealer Intermediation and Price Behavior in the Aftermarket for New Bond Issues," *Journal of Financial Economics* 86, 643-682.

Harris, L., and M. Piwowar, 2006, ‘‘Municipal Bond Liquidity,’’ *Journal of Finance*, 61, 1330-1366.

Hong, G., and A. Warga, 2000, ‘‘An Empirical Study of Bond Market Transactions,’’ *Financial Analysts Journal*, 56, 32-46.

Levin, D., and J. Smith, 1994, “Equilibrium in Auctions with Entry,” *American Economic Review*, 84, 585-599.

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

Petersen, M., 2009, “Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches,” *Review of Financial Studies* 22, 435-480.

Schultz, P., 2001, ‘‘Corporate Bond Trading Costs: A Peek Behind the Curtain,’’ *Journal of Finance*, 56, 677-698.

Thompson, S., 2011, “Simple formulas for standard errors that cluster by both firm and time” *Journal of Financial Economics* 99, 1-10.

**Table 1**
**Descriptive Statistics**

The table presents descriptive statistics based on a sample of all US investment-grade non-callable corporate bond trades in Financial Industry Regulatory Authority's (FINRA) Trade Reporting and Compliance Engine (TRACE) from January 2010 through April 2011, excluding all interdealer trades and trades directed to MarketAxess, an electronic auction venue. Four trade size categories by dollar size are represented based on market conventions, up to a maximum of \$5M and above. Bond characteristics such as Age and Maturity are measured in years. *DTS* (*duration*×*Spread*) is a risk measure defined the bond's duration multiplied by its yield spread over treasuries.

| | **MarketAxess** | **TRACE** |
|---|---|---|
| Number of Trades | 191,150 | 1,578,024 |
| Micro (1-100K) | 44.6% | 77.1% |
| Odd (100K-1M) | 42.6% | 14.6% |
| Round (1M-5M) | 11.8% | 5.1% |
| Max (5M+) | 1.0% | 3.2% |
| Mean Trade Size (\$000) | | |
| Micro (1-100K) | 29 | 21 |
| Odd (100K-1M) | 320 | 247 |
| Round (1M-5M) | 1,794 | 1,896 |
| Number of Distinct Bonds | 1,579 | 4,129 |
| Issue Size (\$ Billion) | 1.91 | 1.56 |
| Age | 3.20 | 3.96 |
| Maturity | 7.52 | 7.82 |
| DTS (*Duration*×*Spread*) | 10.89 | 14.12 |

**Table 2**
**Benchmark Corporate Bond Trading Costs**

The table presents estimates of one-way trading costs (half the bid-offer spread) for a sample of all US investment-grade non-callable corporate bond trades in Financial Industry Regulatory Authority's (FINRA) Trade Reporting and Compliance Engine (TRACE) from January 2010 through April 2011. Half-spread is defined as the difference between the average price for all sell transactions and the average buy price for each bond-day when there is both a buy and a sell, divided by two and expressed in basis points.

| | **MarketAxess** | | **TRACE** | |
|---|---|---|---|---|
| | **Half-Spread** | **Number of Bond-Days** | **Half-Spread** | **Number of Bond-Days** |
| **Panel A:** All bond-days with both a buy and sell in same size category | | | | |
| Micro (1-100K) | 14.35 | 12,029 | 70.58 | 130,529 |
| Odd (100K-1M) | 8.54 | 9,032 | 32.43 | 33,596 |
| Round (1M-5M) | 5.79 | 1,113 | 9.12 | 13,225 |
| Max (5M+) | 2.76 | 16 | 6.39 | 8,874 |
| **Panel B**: Only bonds with both MarketAxess and TRACE trades in same trade size category | | | | |
| Micro (1-100K) | 14.34 | 12,023 | 58.74 | 73,118 |
| Odd (100K-1M) | 8.49 | 8,995 | 32.98 | 24,294 |
| Round (1M-5M) | 5.72 | 1,108 | 8.90 | 6,754 |
| Max (5M+) | 2.58 | 16 | 4.95 | 471 |
| **Panel C**: Only bond-days with both MarketAxess and TRACE trades in same trade size category | | | | |
| Micro (1-100K) | 14.05 | 9,135 | 48.60 | 9,135 |
| Odd (100K-1M) | 6.09 | 2,164 | 30.43 | 2,164 |
| Round (1M-5M) | 6.22 | 202 | 7.43 | 202 |
| Max (5M+) | 0.36 | 9 | 5.05 | 9 |

**Table 3**
**Regressions on Corporate Bond Trading Costs**

The table presents three regression models for implementation shortfall. Standard errors are in parentheses and control for contemporaneous correlation across bonds and time series correlation within a bond. Independent variables include treasury drift and dummy variables for trade size, trade size interacted with MarketAxess (MA), calendar time, rating, and industry. DTS (*Duration*×*Spread*) is duration multiplied by yield spread. All independent variables are demeaned.

| | **(1)** | **(2)** | **(3)** |
|---|---|---|---|
| Odd | -32.05*** | -27.32*** | -27.44*** |
| | (1.523) | (1.494) | (1.487) |
| Round | -57.63*** | -52.06*** | -52.28*** |
| | (1.765) | (1.832) | (1.807) |
| Max | -59.19*** | -55.04*** | -55.23*** |
| | (1.598) | (1.928) | (1.912) |
| MA Micro | -55.01*** | -45.84*** | -45.71*** |
| | (1.582) | (1.394) | (1.404) |
| MA Odd | -27.27*** | -27.75*** | -27.61*** |
| | (1.308) | (1.514) | (1.529) |
| MA Round | -4.503*** | -6.829*** | -6.626*** |
| | (0.887) | (1.047) | (1.051) |
| MA Max | -2.795** | 1.812 | 1.943 |
| | (1.150) | (1.335) | (1.343) |
| A-Rated | -6.925** | -3.049 | -2.031 |
| | (3.293) | (2.937) | (2.828) |
| B-Rated | 12.79 | 11.60 | 11.49 |
| | (11.47) | (11.02) | (11.01) |
| Industrial | -7.875 | 0.404 | -0.361 |
| | (8.477) | (6.183) | (6.129) |
| Financial | -3.101 | 13.89** | 12.61** |
| | (7.880) | (5.678) | (5.651) |
| Utility | 21.82** | 2.910 | 2.010 |
| | (10.34) | (7.191) | (7.208) |
| Monday | 0.611 | -0.111 | -0.0881 |
| | (0.612) | (0.692) | (0.682) |
| Friday | -1.243** | -1.033 | -1.034 |
| | (0.627) | (0.735) | (0.722) |
| Month-End | 0.566 | -0.0693 | -0.0864 |
| | (0.956) | (1.110) | (1.094) |

**Table 3 (Continued)**

| | (1) | (2) | (3) |
|---|---|---|---|
| Drift | 0.347*** | 0.358*** | 0.359*** |
| | (0.0160) | (0.0157) | (0.0158) |
| ln(Maturity) | | 36.52*** | 34.21*** |
| | | (1.875) | (2.629) |
| ln(Age) | | 4.867*** | 4.476*** |
| | | (1.156) | (1.195) |
| ln(Issue Size) | | -5.194*** | -5.117*** |
| | | (0.366) | (0.356) |
| DTS | | | 0.117 |
| | | | (0.0997) |
| Constant | 78.32*** | 58.80*** | 59.15*** |
| | (7.838) | (6.109) | (6.086) |
| Observations | 1,769,174 | 1,769,174 | 1,769,174 |
| R-squared | 0.055 | 0.100 | 0.100 |

*** p<0.01, ** p<0.05, * p<0.1

**Table 4**
**Endogenous Selection of Trading Mechanism: Stage I Choice Model**

Three probit models for the binary choice between over-the-counter versus electronic trading. Trades in MarketAxess are denoted by 1; zero otherwise. Independent variables include dummy variables for trade size, calendar time, rating, and industry. DTS (Duration×Spread) is duration multiplied by yield spread. All independent variables are demeaned. Standard errors are in parentheses.

| | **(1)** | **(2)** | **(3)** |
|---|---|---|---|
| Odd | 0.897*** | 0.863*** | 0.876*** |
| | (0.0252) | (0.0246) | (0.0230) |
| Round | 0.742*** | 0.684*** | 0.702*** |
| | (0.0247) | (0.0264) | (0.0257) |
| Max | -0.286*** | -0.373*** | -0.357*** |
| | (0.0276) | (0.0315) | (0.0310) |
| A-Rated | 0.224*** | 0.117* | 0.0312 |
| | (0.0649) | (0.0618) | (0.0596) |
| B-Rated | -0.809*** | -0.738*** | -0.715*** |
| | (0.189) | (0.183) | (0.179) |
| Industrial | 0.0523 | -0.0592 | -0.0208 |
| | (0.0839) | (0.0873) | (0.0890) |
| Financial | -0.153** | -0.264*** | -0.153** |
| | (0.0658) | (0.0662) | (0.0697) |
| Utility | -0.837*** | -0.551*** | -0.456*** |
| | (0.0951) | (0.0945) | (0.102) |
| Monday | 0.0318** | 0.0416*** | 0.0404*** |
| | (0.0136) | (0.0148) | (0.0141) |
| Friday | -0.00360 | -0.00653 | -0.00588 |
| | (0.0138) | (0.0150) | (0.0139) |
| Month-End | 0.194*** | 0.201*** | 0.202*** |
| | (0.0184) | (0.0207) | (0.0184) |
| ln(Maturity) | | -0.120*** | 0.0897* |
| | | (0.0274) | (0.0499) |
| ln(Age) | | -0.0302 | -0.00289 |
| | | (0.0254) | (0.0252) |
| ln(Issue Size) | | 0.215*** | 0.212*** |
| | | (0.0194) | (0.0186) |
| DTS | | | -0.0127*** |
| | | | (0.00311) |
| Constant | -1.581*** | -1.437*** | -1.480*** |
| | (0.0813) | (0.0807) | (0.0798) |
| Observations | 1,769,174 | 1,769,174 | 1,769,174 |

*** p<0.01, ** p<0.05, * p<0.1

**Table 5**
**Endogenous Selection of Trading Mechanism: Stage II Cost Model**

Table 5 uses the probit specification of Table 4, model (3), to estimate the second-stage cost model for MarketAxess and TRACE. All continuous independent variables are demeaned. The selectivity adjustment (Inverse Mill's ratio) terms are *Inv Mill MA* and *Inv Mill TRACE*, respectively. Standard errors are in parentheses.

| | **MarketAxess** | **TRACE** |
|---|---|---|
| Odd | -5.543*** | -5.307 |
| | (1.794) | (4.000) |
| Round | -8.187*** | -35.13*** |
| | (1.412) | (2.832) |
| Max | -6.208*** | -61.64*** |
| | (1.622) | (2.294) |
| Drift | 0.423*** | 0.344*** |
| | (0.0169) | (0.0177) |
| ln(Maturity) | 6.896*** | 35.04*** |
| | (0.838) | (2.718) |
| ln(Age) | 3.921*** | 3.263*** |
| | (0.456) | (1.101) |
| ln(Issue Size) | -6.117*** | -2.000*** |
| | (0.500) | (0.495) |
| DTS | 0.177*** | 0.0469 |
| | (0.0483) | (0.0875) |
| Inv Mills MA | -0.222 | |
| | (2.753) | |
| Inv Mills TRACE | | 80.00*** |
| | | (12.14) |
| Constant | 19.14*** | 78.87*** |
| | (5.088) | (2.102) |
| Observations | 191,150 | 1,578,024 |
| R-squared | 0.069 | 0.085 |

*** p<0.01, ** p<0.05, * p<0.1

**Table 6**
**Descriptive Statistics for Number of Dealers in an Auction**

The table presents descriptive statistics from January 2010 through April 2011 for the number of dealers participating in 191,150 electronic auctions. Four trade size categories by dollar size are represented, based on market conventions, up to a maximum of $5M and above. Figures reported are sample means.

| Trade Size | Number of Dealers Queried | Percentage Responding | Percentage with No Response |
|---|---|---|---|
| Micro (1-100K) | 27.70 | 16.8% | 5.9% |
| Odd (100K-1M) | 26.60 | 22.3% | 7.7% |
| Round (1M-5M) | 25.66 | 26.6% | 9.8% |
| Max (5M+) | 23.98 | 29.7% | 15.1% |

**Table 7**
**Negative Binomial Model for Number of Dealers Responding in Auction**

The table presents three regression models for the number of dealers responding for a sample of electronic auctions from January 2010-April 2011. Independent variables include treasury drift and dummy variables for trade size, trade size interacted with MarketAxess (MA), calendar time, rating, and industry. DTS (*Duration*×*Spread*) is duration multiplied by yield spread. All independent variables are demeaned. Robust standard errors in parentheses

| | **(1)** | **(2)** | **(3)** |
|---|---|---|---|
| Ln(Queries) | 0.643*** | 0.591*** | 0.534*** |
| | (0.00684) | (0.00543) | (0.00568) |
| Odd | 0.255*** | 0.263*** | 0.291*** |
| | (0.0117) | (0.0103) | (0.00958) |
| Round | 0.456*** | 0.462*** | 0.511*** |
| | (0.0164) | (0.0127) | (0.0110) |
| Max | 0.490*** | 0.527*** | 0.674*** |
| | (0.0215) | (0.0176) | (0.0138) |
| Monday | 0.0617*** | 0.0599*** | 0.0589*** |
| | (0.00311) | (0.00275) | (0.00260) |
| Friday | -0.0588*** | -0.0555*** | -0.0496*** |
| | (0.00362) | (0.00310) | (0.00289) |
| Month-End | 0.00716 | -0.0186*** | -0.0185*** |
| | (0.00721) | (0.00523) | (0.00494) |
| ln(Maturity) | | -0.0203*** | 0.117*** |
| | | (0.00226) | (0.0183) |
| ln(Age) | | -0.0817*** | -0.0482*** |
| | | (0.00326) | (0.00509) |
| ln(Issue Size) | | 0.318*** | 0.255*** |
| | | (0.0106) | (0.0104) |
| DTS | | | -0.00435*** |
| | | | (0.000869) |
| Constant | -0.683*** | -0.595*** | -0.130*** |
| | (0.0251) | (0.0190) | (0.0308) |
| Observations | 261,306 | 261,306 | 187,834 |

*** p<0.01, ** p<0.05, * p<0.1

**Table 8**
**Corporate Bond Trading Costs in an Auction Mechanism**

The table presents three regression models for trading cost (implementation shortfall) for a sample of electronic auctions from January 2010-April 2011. Independent variables include treasury drift and dummy variables for trade size, bond characteristics, and dealer responses (total, expected, and unexpected) to queries based on the estimates in Table 7. DTS (*Duration*×*Spread*) is duration multiplied by yield spread. All independent variables are demeaned. Robust standard errors in parentheses

| | (1) | (2) | (3) |
|---|---|---|---|
| Odd | -5.373*** | -1.309*** | -2.631*** |
| | (0.478) | (0.420) | (0.458) |
| Round | -7.792*** | -0.454 | -3.157*** |
| | (0.594) | (0.537) | (0.604) |
| Max | -6.277*** | 2.933*** | -1.222 |
| | (1.199) | (1.137) | (1.196) |
| Drift | 0.422*** | 0.424*** | 0.424*** |
| | (0.0169) | (0.0165) | (0.0165) |
| ln(Maturity) | 6.843*** | 9.075*** | 8.395*** |
| | (0.827) | (0.754) | (0.776) |
| ln(Age) | 2.192*** | 1.451*** | 1.677*** |
| | (0.205) | (0.208) | (0.202) |
| ln(Issue Size) | -6.263*** | -1.225*** | -3.288*** |
| | (0.420) | (0.390) | (0.420) |
| DTS | 0.186*** | 0.120*** | 0.144*** |
| | (0.0449) | (0.0382) | (0.0396) |
| ln(Resp) | | -17.07*** | |
| | | (0.640) | |
| ln(Exp_Resp) | | | -14.09*** |
| | | | (0.803) |
| ln(Unexp Resp) | | | -36.14*** |
| | | | (1.529) |
| Constant | 27.46*** | 29.16*** | 140.3*** |
| | (1.365) | (1.223) | (4.709) |
| Observations | 187,834 | 187,834 | 187,834 |
| R-squared | 0.069 | 0.101 | 0.097 |

*** p<0.01, ** p<0.05, * p<0.

**Figure 1: Frequency Distribution of Number of Dealer Responses**

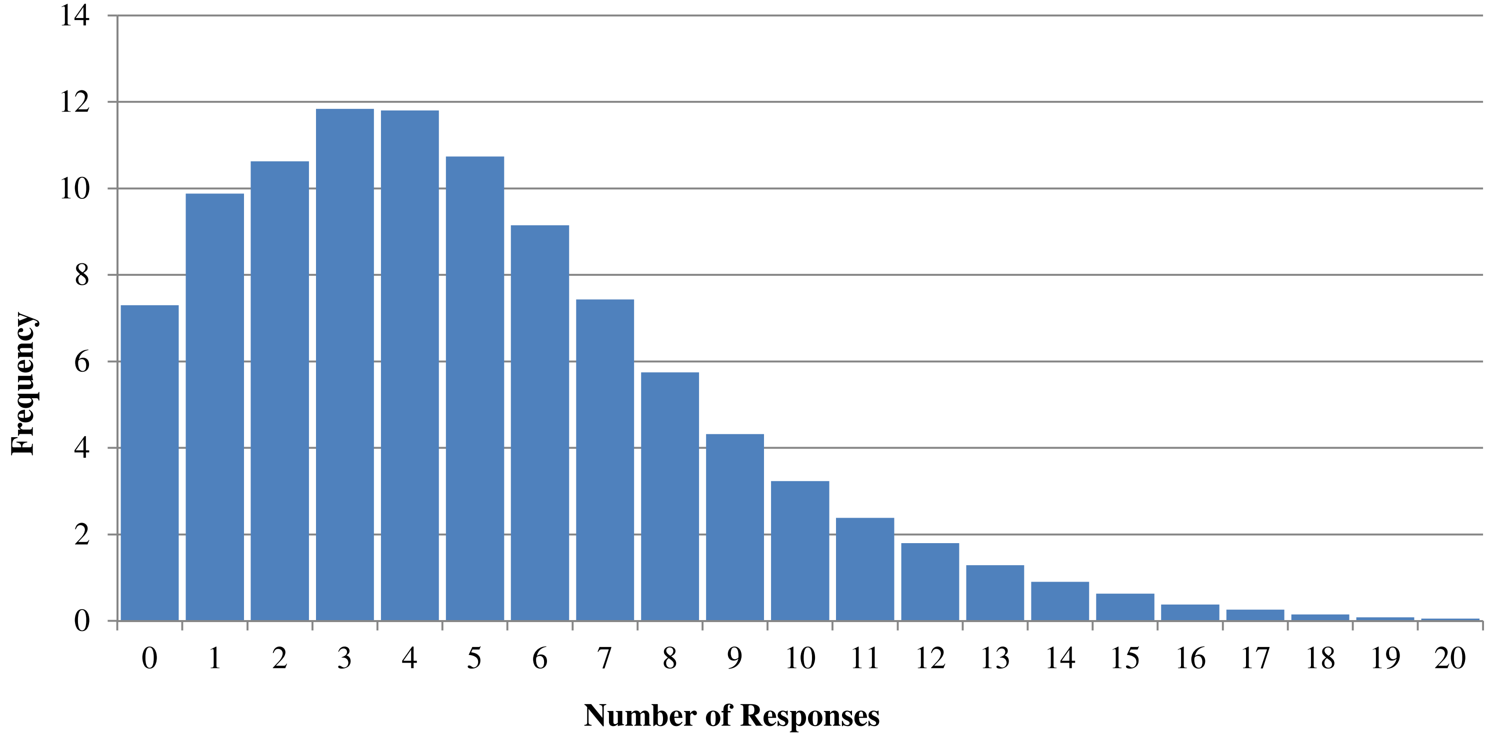
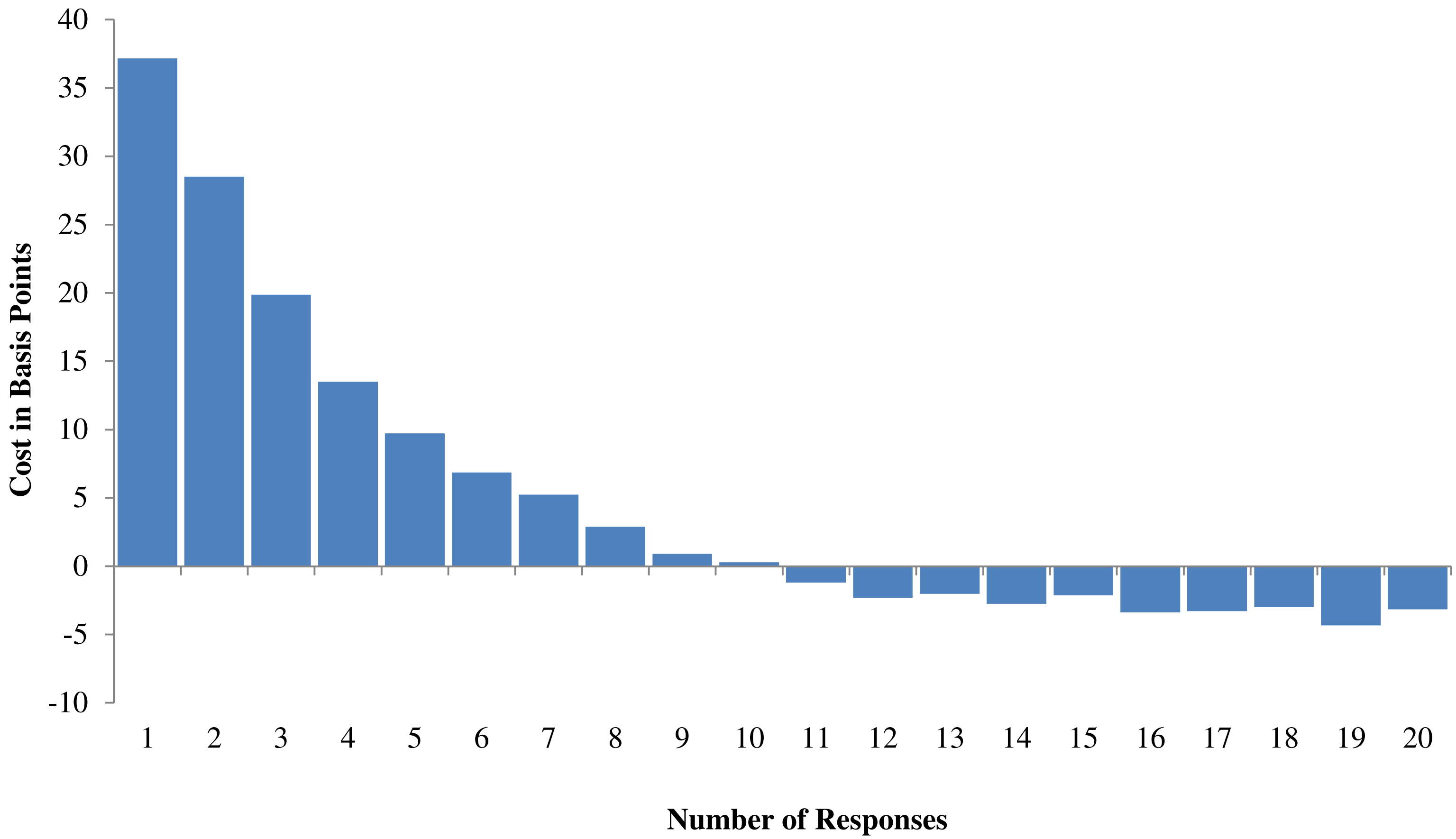


**Figure 2: Transaction Costs in Basis Points by Number of Dealer Responses**



# Liquidity: What you see is what you get?

Vincent van Kervel*

April 2012

**Abstract**

Competition between electronic limit order books improves the overall liquidity of equity markets in most studies. However, my model shows that liquidity offered on the limit order books combined may strongly overestimate the actual liquidity available to investors. The excess is caused by high-frequency traders operating as market makers, who may duplicate their limit order schedules on several venues to increase their execution probabilities. Then, after a trade on one venue they will quickly cancel outstanding limit orders on others. The magnitude of the cancellations depends on the fraction of investors that may access several venues simultaneously, i.e., who use Smart Order Routing Technology (SORT). The reason is that market makers incur higher adverse selection costs when the investor trades at a competing venue first. Consequently, a higher fraction of SORT investors reduces the incentives of market makers to place duplicate limit orders. The empirical results confirm the main prediction of the model, as trades on the most active venues are followed by cancellations of limit orders on competing venues of more than 53% of the trade size.

**JEL Codes:** G10; G14; G15;

**Keywords:** Market microstructure, Fragmentation, Market makers, High Frequency Trading, Smart Order Routing

---

*Department of Finance, Tilburg University, TILEC, CentER. Email: v.l.vankervel@uvt.nl., Phone: +31 13 466 4056. I am indebted to my thesis supervisors Hans Degryse and Frank de Jong. I thank Larry Glosten for invaluable comments and the invitation to stay at Columbia University, whose hospitality is gratefully acknowledged. In addition, I thank Jean-Edouard Colliard, Peter Hoffmann, Albert Menkveld and seminar participants at the University of Amsterdam and the VU University of Amsterdam for suggestions. Comments and suggestions are appreciated. All errors or mine.

# 1 Introduction

Two important trends have drastically changed equity markets in recent years. First, technological innovations have lead to high-frequency trading, a trading strategy whereby a computer algorithm analyzes market data and trades at extremely high speed. Second, competition between trading venues has caused a dispersion of trading volumes and liquidity across venues, i.e., the market has become fragmented. These two changes in the structure of equity markets might strongly affect the optimal behavior of investors. Indeed, as high-frequency traders operate on several trading venues simultaneously, the order flow and liquidity of these venues become strongly interrelated.[1] This paper argues that the interrelation of liquidity across trading venues causes substantial overestimation of liquidity aggregated over these trading venues.

I underpin this hypothesis in a model of competition between two centralized limit order books. The model predicts that high-frequency traders, who supply liquidity by acting as traditional market makers, have an incentive to duplicate their limit orders on both venues. They will do so because this strategy increases their execution probabilities and therefore expected profits. Then, after execution on one venue, they will quickly cancel their limit orders on competing venues. Effectively, the depth aggregated over all venues overstates true liquidity, as a single trade reduces liquidity on many venues simultaneously. The empirical results strongly support this hypothesis: a trade reduces liquidity at competing venues by 53% of the trade size within a second.

My first contribution to the literature is a theoretical model. The model is based on the framework of a pure limit order market with adverse selection of Sandås (2001), extended to a two-venue setting. High-frequency market makers supply liquidity with limit orders, whereas potentially informed traders demand liquidity with market orders. In this setting, only a fraction of the traders has the technological infrastructure to submit market orders to both venues simultaneously, i.e., smart order routing technology (SORT). Effectively, I introduce market segmentation because non-SORT traders cannot access the liquidity of both venues. In equilibrium, a limit order faces higher adverse selection costs when executed against a SORT trader. The reason is that conditional upon execution, there is a probability that she traded on the competing venue already and therefore the combined

[1] The high level of interaction between markets became apparent during the flash crash, i.e., between the E-mini S&P 500 futures and the individual stocks (SEC-CFTC "Findings regarding the market events of May 6 2010", 2010).

trades are larger and more informed on average. Consequently, a lower fraction of SORT traders reduces adverse selection costs and increases equilibrium liquidity supply. However, the additional liquidity follows from market segmentation only, and will be cancelled after a trade on the competing venue; hence the term duplicate liquidity.

The model offers the following predictions. First, due to the information content of a trade on one venue, limit orders on the same side of the order books of competing venues will be cancelled. Cancellations would not occur in a single venue setting, which implies that I must test whether the observed cancellations are significantly greater than zero. Second, after a trade new limit orders will be placed on the opposite side of the limit order books of all venues, consistent with asymmetric information models. Third, duplicate limit orders will only be placed on venues where non-SORT investors operate (mostly the traditional market), because market makers face lower adverse selection costs when trading with non-SORT investors. When all investors use SORT, competition between market makers only improves liquidity, as documented by Glosten (1998) and Foucault and Menkveld (2008). Fourth, the model explains why entrant venues are typically very liquid but execute relatively little volume. Although market makers offer substantial liquidity there, only a small fraction of investors use SORT to tap this liquidity.[2] Fifth, the decision of one investor to adopt smart order routing technology increases her accessible liquidity, but also imposes a negative externality on all other investors (the SORT and non-SORT traders as well as market makers).

My second contribution to the literature is an empirical investigation. The dataset contains the entire limit order books of all relevant trading venues with publicly displayed data, for a sample of FTSE 100 stocks in November 2009. These stocks are traded in a fairly fragmented environment, as the traditional market (the London Stock Exchange) executes 66% of lit trading volume, leaving 34% to four competing venues (Chi-X, Bats Europe, Turquoise and Nasdaq OMX). I test the models main predictions by investigating the short-term correlations between the supply and demand of liquidity across trading venues. Specifically, I regress changes in liquidity supply at the bid or ask side of one venue on contemporeneous and lagged liquidity demand of all five venues, i.e., buy and sell trading volumes. Lagged trading volumes of up to ten seconds away measure investors' responses to trades over time. I sample one observation per 100 milliseconds in order to analyze high-frequency trading strategies. To the best of my knowledge, this paper is the

[2] In the model, transaction costs and trading speeds are assumed equal across trading venues, and therefore do not explain differences in market shares.

first to analyze the impact of high-frequency trading strategies on the liquidity supply across trading venues.

The empirical results strongly support the models' main prediction that trades are followed by limit order cancellations on competing venues. That is, within 100 milliseconds, transactions on the three most active trading venues are followed by cancellations on the *same side* of competing limit order books of 38 to 85% of the transaction size. As a result, liquidity aggregated over all venues overstates liquidity available to investors, since e.g. a 100 share trade reduces liquidity by more than 138 shares. Further, these cancellations increase to 53 to 149% of the transaction size after one second (depending on the trading venue). This finding is particularly relevant to algorithms designed to split up large trades over time, as the liquidity impact of each individual small trade is indeed larger than previously thought. The fact that liquidity shocks immediately spill over to other venues is not captured by static liquidity measures, such as the quoted depth.

The analysis also confirms several other predictions of the model. First, the magnitude of cancellations is stronger on the primary market after trades on entrant venues (46 to 52%), than vice versa (14 to 30%). Second, trades reveal information on the true asset value, as a trade on the ask side is followed by new limit orders on the bid side of 30 to 70% of the tradesize.

The main policy implication of the model is that fair markets require investors to be able to split up trades *simultaneously* across several venues. When a trader leaves a millisecond delay between the split, the market effectively becomes segmented. That is, after high-frequency traders observe the first part of the trade, they will quickly cancel their duplicate limit orders on competing venue before the second part arrives. This high-frequency trading strategy is known as latency arbitrage, and relates to the trend of increasingly faster trading mechanisms.[3]

Most related to this work is literature on competition between electronic limit order books. Pagano (1989) predicts that all trading activity should divert to the trading system with the lowest transaction costs, and only unstable equilibria may exist when two venues have identical cost structures. In contrast, Glosten (1998) shows that two electronic limit order markets can coexist when tick sizes are discrete and time priority rules apply. Since time priority is absent *across* venues, competition between liquidity suppliers increases, which in turn raises aggregate liquidity. This point is further developed in Foucault and

[3]See for example "Trading at the speed of light", Sept 12, 2011 on www.ft.com.

Menkveld (2008), who coin this channel the "queue-jumping" effect. Competition between exchanges also arises through differences in the tick size, where the venue with the smallest tick size becomes most liquid (Biais, Bisière, and Spatt (2010) and Buti, Rindi, Wen, and Werner (2011)). My model confirms all these findings, and adds the presence of duplicate limit orders when some investors cannot access all trading venues.

Empirical research on competition between exchanges typically focusses on its impact on aggregate liquidity and welfare. O'Hara and Ye (2011) find that competition between exchanges reduces the effective cost of trading. Jovanovic and Menkveld (2011) find that competition between an entrant (Chi-X) and incumbant market (Euronext) has an ambiguous effect on total welfare. Degryse, de Jong, and van Kervel (2011) show that aggregate liquidity increases by competition between venues with publicly displayed limit order books, but worsens by competition of opaque markets. Instead, the current paper studies the effect of competing exchanges on the liquidity supply of each of the individual exchanges.

Rather than assuming a fraction of non-SORT traders, market segmentation may also arise when investors have different trading speeds. Biais, Foucault, and Moinas (2011) show that high-frequency trading facilitates the search for trading opportunities, but increases adverse selection costs for slow traders. As a result, the equilibrium level of investment in high-frequency technology exceeds the welfare maximizing level. Hoffmann (2011) shows that high-frequency traders have lower adverse selection costs on average, which in turn causes slow traders to post less aggressive limit orders. The result is an ambiguous effect on welfare, which depends on the stocks fundamental volatility. High-frequency traders might extract rents from liquidity motivated traders when they operate as intermediaries (Cartea and Penalva, 2011). McInish and Upson (2011) provide empirical evidence that high-frequency traders pick off slower traders in the US, due to regulation that effectively causes slow investors to trade against stale quotes. Hasbrouck and Saar (2011) argue that trading speeds affect the competition between liquidity suppliers in a single trading venue. This paper shows that different trading speeds might in fact cause duplicate limit orders in fragmented markets.

Finally, this paper relates to recent research on high-frequency traders who act as market makers (e.g., Jovanovic and Menkveld (2011), Menkveld (2011) and Guilbaud and Pham (2011)). Such market makers gain the bid-ask spread by offering liquidity at both sides of the limit order book, while simultaneously managing advere selection costs (e.g., Glosten and Milgrom (1985), Glosten and Harris (1988)) and inventory risk (e.g., Ho and

Stoll (1981)). This paper extends these works by analyzing market making when trading is fragmented across electronic limit order books.

The remainder of the paper is structured as follows. Section 2 describes the duplicate limit order hypothesis and the model. Section 3 presents the empirical work, after which I conclude.

# 2 The model

In this section I first describe the duplicate limit order hypothesis. Then, I place it into a model that quantifies this duplicate limit order effect. In essence, the model is a combination of Glosten (1998), Sandås (2001) and Foucault and Menkveld (2008). I contribute to Foucault and Menkveld (2008) by allowing for adverse selection. Rather than analyzing the single exchange setting in Sandås (2001), I focus on two competing centralized limit order books. Compared to Glosten (1998), I introduce market segmentation by constraining some traders to have access to one trading venue only, and show how this causes overestimation of consolidated liquidity. Glosten (1998) and Sandås (2001) are special cases of my model.

## 2.1 Duplicate limit order hypothesis

In a fragmented trading environment time priority is not enforced between trading venues, whereas price priority is enforced only when the trader has access to both venues. Price priority implies that limit orders with a better price are executed before those with a worse price, while time priority entails that limit orders placed first are executed first.[4]

Because of the absence of time priority across venues, liquidity suppliers can improve their execution probabilities by placing similar limit orders on several venues simultaneously. After execution on one venue, they will cancel remaining limit orders on the other venues.[5] Therefore, I predict that a trade is immediately followed by cancellations of limit orders on the same side of competing limit order books. A tradeoff arises however, as there is a probability that both limit orders will be executed simultaneously, causing the liquidity supplier to trade too much.

[4] In the US, price priority across markets is enforced by law, Reg NMS. Time priority however, crucial for this hypothesis, is not enforced.

[5] Note that this strategy does not work in a single exchange setting due to time priority.

Who might pursue this duplicate limit order strategy? I argue that high-frequency traders who operate as market makers can strongly benefit from the increased execution probability. That is, placing duplicate limit orders increases their trading rate and expected profits. At the same time, using state of the art technology allows them to monitor several venues simultaneously and cancel limit orders quickly after observing trades and other news. Being able to cancel quickly reduces the expected cost of simultaneous execution and adverse selection.[6] In contrast, this strategy is likely not very attractive to "regular traders." For some traders, the technology required for continuous monitoring might be too expensive. Other traders might use algorithms to optimally split up large quantities over time, in which case they will not cancel limit orders since each child order is part of a large parent order.

## 2.2 Model setup

This section quantifies the cost of executing duplicate limit orders in an adverse selection framework. I show that duplicate limit orders face higher adverse selection costs because conditional upon executing both limit orders, the incoming trade is larger and therefore more informed on average.

Consider two venues, A and B, and two types of investors, market makers and traders. The risk neutral market makers supply liquidity by placing limit orders on one or both venues. They are profit maximizing and use high-frequency trading technology to quickly access all venues. The risk averse traders demand liquidity by placing market orders. Traders have private information or liquidity motives, and therefore want to trade quickly.[7]

The asset has a fundamental value $X_t$, which incorporates all information available up until period $t$. The next periods fundamental value is given by

$$X_{t+1} = X_t + \mu + \varepsilon_{t+1}, \tag{1}$$

where $\mu$ is a trend and $\varepsilon_{t+1}$ a random innovation. Traders have some private information on $\varepsilon_{t+1}$.

[6] In this context, trades on two venues occur simultaneously when the liquidity supplier is not fast enough to adjust his outstanding limit orders after the first trade. Effectively, his quotes are stale when the second trade comes in.

[7] Since traders are risk averse and may have liquidity motives, the market does not break down like in a Kyle (1985) framework.

Trading occurs sequentially over periods indexed by $t$. Each period $t$ consists of three stages. First, the market makers arrive consecutively and place limit orders on one or both venues. They do so until no market maker finds it optimal to place additional limit orders. Then the trader arrives, placing market orders with total size $x$ that consume liquidity on one or both venues. Finally, the market makers update their expected fundamental value of the asset conditional on the incoming market order using some price impact function. Now, the game starts over and is repeated for every trade. Because of high-frequency trading technology, the last two stages might last only a few milliseconds.

Market makers place their limit orders on a discrete pricing grid, $\{p_1, p_2, ...p_k\}$ for the ask side. The current expected value (midpoint) is $p_0$, the best ask price $p_1$ and the tick size $\Delta = p_i - p_{i-1} > 0$. I focus on the ask side only, prices larger than $p_0$, as the bid side is analogous. Denote the number of shares offered on venue $j \in \{A, B\}$ at each price level by $\{Q_{j1}, Q_{j2}, \ldots, Q_{jk}\}$. Price and time priority exists within each venue, but not between venues.

## 2.3 The trader

The trader is randomly drawn from a population of traders, which consists of three types. The first type only goes to venue A (fraction $\alpha$), the second type only to venue B (fraction $\beta$), and the third type uses smart order routing technology (SORT) to access both venues simultaneously (fraction $\gamma = 1 - \alpha - \beta$). Simultaneous is defined here as sending trades to both venues so fast that the market makers are unable to update their limit order schedules inbetween the trades. When both venues offer the same best price, the SORT investors are indifferent as to where to send their trades to. In this case, they simply use a tie-breaking rule, which posits that with probability $\pi$ they first buy shares on venue A, and with probability $(1 - \pi)$ they first buy shares on venue B. The parameters $\alpha, \beta$ and $\pi$ are constant, as I focus on a high-frequency environment.

Four reasons motivate why some investors are not able to trade on both venues simultaneously. First, human traders with access to both venues might trade too slowly, creating a delay of several milliseconds when they split up a trade across two venues. In this case, high-frequency market makers have sufficient time to update their limit order schedules inbetween trades. As a result, human traders effectively have access to one exchange only.[8] This argument seems realistic, as McInish and Upson (2011) show that slow traders are adversely selected because they often trade against stale quotes. Second, smart order routing technology might be too expensive for some traders, as it requires fixed costs for technological infrastructure, software, programmers and access to data feeds etc. Third, fixed costs of sending your trade to two venues could make it more economical to trade through one price but save on the transaction costs (such that fixed clearing and settlement costs are paid only once). Fourth, investors may deliberately decide to split up large trades over time, for example to benefit from the resiliency of liquidity. In this case, they might not need the liquidity offered on additional venues.

The trader is a buyer or seller with equal probability, and has a reservation price $p_m > p_1$ at which she is not willing to buy.[9] The expected order size is similar for all types, with mean $\phi$ and exponential density function

$$f(x) = \frac{1}{\phi}\exp(-\frac{x}{\phi}) \qquad if\ x > 0\ (market\ buy). \tag{2}$$

The cumulative distribution function is $F(\cdot)$. Assuming an exogenous order size simplifies the analysis, although in reality the size depends on several factors, such as the current state of the limit order book, the trader's expected fundamental value and her current holdings. However, I mainly focus on the behavior of the market makers, described next.

## 2.4 The market makers

Since traders are informed about the innovations in the assets true value, market makers update their expectation of the fundamental value based on the size of the incoming trade $x$. The market makers observe trades on both exchanges. They use a price impact function $h(x)$ which is non-decreasing, since buy trades typically contain positive information with respect to the true value (similarly, sells contain negative information). Then, larger orders

[8] In fact, when market makers can update inbetween the two trades, the slow trader will never prefer to split up his trade across two venues. The reason is that the market makers update their limit order schedules symmetrically across two venues. Therefore, if it is optimal to submit the first part of the trade to venue A, then it is also optimal to submit the second part there.

[9] This small assumption prevents the trade from walking up the limit order book too much in case of a thin order book, but does not affect the outcome of the model.

cause more adverse selection costs and greater price impacts.

$$E(X_{t+1}|x) = E(X_t) + \mu + h(x). \tag{3}$$

As in Sandås (2001), the price impact function is linear with coefficient $\lambda$,

$$h(x) = \lambda x. \tag{4}$$

For simplicity, I assume market makers face a fixed limit order *execution cost* $c$, while the *submission cost* is zero.[10] Given this setup, we can calculate the expected profit of a limit order placed on any location $q$ in the queue of limit orders on each venue. The profit depends on the expected value of the asset conditional upon execution of the limit order. For a limit order on venue A, this value is $E(X|x > q)$ when the trader immediately goes to A (denoted $E_q(X)$ for brevity), and $E(X|x > q + Q_{B1})$ when the trader first buys all shares $Q_{B1}$ on venue B and then goes to A. Denote the probability that the incoming order $x$ is larger than $q$ as $\overline{F}_q = 1 - F(q)$, then the profit of a limit order on price level 1 of venue A is

$$\Pi_{A,q} = (\alpha + \gamma\pi)\overline{F}_q(p_1 - c - E_q(X)) + \gamma(1 - \pi)\overline{F}_{q+Q_{B1}}(p_1 - c - E_{q+Q_{B1}}(X)). \tag{5}$$

In the first term, the limit order executes against traders going to venue A only ($\alpha$) and against SORT investors who choose to trade on venue A first ($\gamma\pi$). Then, the expected profit is simply the price minus the fixed cost $c$ and the expected value of the asset conditional on $x > q$. The second term represents SORT traders who first buy all the shares offered at price $p_1$ on venue B, and then buys shares on venue A ($\gamma(1 - \pi)$). Indeed, market makers only realize profits when their limit orders are executed. Therefore, conditional upon execution, the second term of the expected profit is lower since the incoming trade is larger and more informed, i.e., $E(X_{t+1}|x > q + Q_{B1}) > E(X_{t+1}|x > q)$.

Not surprisingly, we observe that the expected profit of limit orders on venue A depends on the number of shares offered on venue B. Therefore, to obtain the equilibrium liquidity supply we need to solve for the profit equations of limit orders on both venues

[10] Placing limit orders is costless on the venues we analyze empirically.

simultaneously. The profitabilty of limit orders on price level 1 of venue B is

$$\Pi_{B,q} = (\beta + \gamma(1-\pi))\overline{F}_q(p_1 - c - E_q(X)) + \gamma\pi\overline{F}_{Q_{A1}+q}(p_1 - c - E_{Q_{A1}+q}(X)). \quad (6)$$

## 2.5 Equilibrium

The model is in equilibrium when no market maker can profitably place an additional limit order on any price level (as in Glosten (1994), Proposition 2). Therefore, the expected profit of the *marginal limit order*, the single share offered at the end of the queue of limit orders, must equal zero for all price levels on each venue. Following Sandås (2001), I substitute $q = Q_{A1}$ (the marginal limit order) and integrate the profit equation over the distribution of the incoming order x, using equations (2), (3) and (4)

$$\Pi_{A1} = \int_{Q_{A1}}^{\infty} (\alpha + \gamma\pi)(p_1 - c - X_t - \lambda x)\frac{1}{\phi}\exp(-\frac{x}{\phi})\mathrm{d}x + \int_{Q_{A1}+Q_{B1}}^{\infty} \gamma(1-\pi)(p_1 - c - X_t - \lambda x)\frac{1}{\phi}\exp(-\frac{x}{\phi})\mathrm{d}x = 0.$$

The first integral goes to infinity, because the marginal limit order is executed for any trade larger or equal to $Q_{A1}$. For demonstrational purposes previous equations contain $\gamma$, which I next substitute with $(1 - \alpha - \beta)$ to calculate the solutions. Solving the integral gives

$$\Pi_{A1} = (\alpha + \pi(1-\alpha-\beta))(p_1 - c - X_t - \lambda(\phi + Q_{A1}))\exp(-\frac{Q_{A1}}{\phi}) + (1-\pi)(1-\alpha-\beta)(p_1 - c - X_t - \lambda(\phi + Q_{A1} + Q_{B1}))\exp(-\frac{Q_{A1}+Q_{B1}}{\phi}) = 0.$$

The zero expected profit condition implies that the first line of the equation is positive while the second line is negative. In equilibrium, the market makers expect to lose to traders that go to venue B first, and profit from traders that go to venue A first. Similarly, for venue B we have

$$\Pi_{B1} = \int\limits_{Q_{B1}}^{\infty} (\beta + \gamma(1-\pi))(p_1 - c - X_t - \lambda x)\frac{1}{\phi}\exp(-\frac{x}{\phi})\mathrm{d}x + \int\limits_{Q_{A1}+Q_{B1}}^{\infty} \gamma\pi(p_1 - c - X_t - \lambda x)\frac{1}{\phi}\exp(-\frac{x}{\phi})\mathrm{d}x = 0,$$

which I solve to obtain

$$\Pi_{B1} = \pi(\beta + (1-\pi)(1-\alpha-\beta))(p_1 - c - X_t - \lambda(\phi + Q_{B1}))\exp(-\frac{Q_{B1}}{\phi}) + \pi(1-\alpha-\beta)(p_1 - c - X_t - \lambda(\phi + Q_{A1} + Q_{B1}))\exp(-\frac{Q_{A1}+Q_{B1}}{\phi}) = 0.$$

The two equations with two unknowns can be solved implicitly[11]

$$\begin{aligned} Q_{A1} &= \frac{p_1 - c - X_t - \lambda\phi}{\lambda} - \frac{\gamma(1-\pi)Q_{B1}}{\gamma(1-\pi) + (\alpha+\gamma\pi)\exp(\frac{Q_{B1}}{\phi})}, \\ Q_{B1} &= \frac{p_1 - c - X_t - \lambda\phi}{\lambda} - \frac{\gamma\pi Q_{A1}}{\gamma\pi + (1-(\alpha+\gamma\pi))\exp(\frac{Q_{A1}}{\phi})}. \end{aligned} \tag{7}$$

The zero expected profit condition holds for prices deeper in the order book too. Now, the expected profit consists of three terms, as a limit order on price level 2 on venue A might get executed by traders of type $\alpha$ (who first buy $Q_{A1}$ and then $Q_{A2}$), by type $\gamma\pi$ (who first buy $Q_{A1}, Q_{B1}$ and then $Q_{A2}$), or by type $\gamma(1-\pi)$ (who first buy $Q_{B1}, Q_{A1}, Q_{B2}$ and then $Q_{A2}$). For brevity, denote $Z(x) = \alpha(p_2 - c - X_t - \lambda x)\frac{1}{\phi}\exp(-\frac{x}{\phi})$, then

$$\Pi_{A2} = \int\limits_{Q_{A1}+Q_{B1}}^{\infty} Z(x)\mathrm{d}x + \int\limits_{Q_{A1}+Q_{B1}+Q_{A2}}^{\infty} Z(x)\mathrm{d}x + \int\limits_{Q_{B1}+Q_{A1}+Q_{B2}+Q_{A2}}^{\infty} Z(x)\mathrm{d}x. \tag{8}$$

I solve the system for prices deeper in the order book in similar fashion to equation (7), which gives an implicit solution of the form $Q_{A2} = f(Q_{A1}, Q_{B1}, Q_{B2}; \mathit{parameters})$, and similarly for other price levels.

[11]This solution is unique. In Equation (7), $\partial Q_{A1}/\partial Q_{B1} < 0$ and $\partial Q_{B1}/\partial Q_{A1} < 0$, since the first term does not depend on $Q_{A1}$ or $Q_{B1}$, while we subtract a second term which consists of non-negative parameters only.

## 2.6 Testable implications

In this section I discuss two static predictions of the model, which hold at each point in time, and two dynamic predictions, which hold before and after a trade.

The static predictions follow from the solutions for $Q_{A1}$ and $Q_{B1}$ in equation (7). The first terms are identical, and equal the optimal quantity offered in a single venue setting (the solution obtained by Sandås (2001)). However, we subtract a non-negative second term, implying that the offered quantities on the individual venues are weakly lower in a fragmented market. The second term reflects the adverse selection costs incurred by the market makers when a SORT trader buys at the competing venue first. Based on the solution, I derive the following static predictions (the proofs are in section A.1 of the Appendix).

**Proposition 1** *When the liquidity at both venues has reached equilibrium:*

1. *An increasing fraction of SORT traders $\gamma$ strictly reduces consolidated depth.*
2. *For $\gamma > 0$ and SORT traders do not always go to venue A first or always to venue B first, i.e., $0 < \pi < 1$, the consolidated liquidity is strictly higher than the liquidity offered in a single exchange setting.*

A larger fraction of SORT traders implies less segmented markets, as many traders have access to the liquidity of both venues simultaneously. Effectively, market makers face higher adverse selection costs when trading with a SORT trader, because there is a probability that she traded on the competing venue already. Therefore, a higher fraction of SORT traders increases expected adverse selection costs and reduces equilbrium liquidity supply. The second part of the proposition relates to the benefit of competition between exchanges as documented by Glosten (1998) and Foucault and Menkveld (2008), which stems from market makers' ability to jump the queue of limit orders on one venue by placing limit orders on the other. Such "queue jumping" erodes the profitability of existing limit orders on the first venue and effectively increases competition between liquidity suppliers.

To demonstrate how the equilibrium liquidity depends on the individual parameters, I take the first derivative from the solution of $Q_{A1}$ with respect to $\alpha, \beta, \gamma$ and $\pi$. The results can be summarized as follows, and hold symmetrically for $Q_{B1}$ (see the Appendix).

**Proposition 2** *Other things equal, the number of shares offered at price level 1 of venue A, $Q_{A1}$:*

1. *Increases by $\pi$, the probability that SORT traders go to venue A first; by $\alpha$, the fraction of investors that only have access to venue A; and by $\beta$, those with access only to venue B when holding $\alpha$ constant.*
2. *Decreases by $Q_{B1}$; by $\gamma$, the fraction of SORT investors; and by $\beta$ when holding $\gamma$ constant.*

As expected, the liquidity offered on venue A strictly increases in $\alpha$, and decreases in $\gamma$. These effects are solely driven by adverse selection costs, which reduce when more investor go to venue A first ($\alpha$ and $\pi$) and increase when more investors go to venue B first and then to venue A ($\gamma(1-\pi)$).[12] Since $\gamma = 1 - \alpha - \beta$, liquidity on A increases with $\beta$ when holding $\alpha$ constant (such that the fraction of smart order routers decreases), and decreases with $\beta$ when holding $\gamma$ constant (such that the fraction of $\alpha$ decreases). An increase in $Q_{B1}$ implies larger adverse selection costs when a SORT trader purchases $Q_{B1}$ first and then $Q_{A1}$.

Notice the following extreme case. When the fraction of SORT investors $\gamma = 0$, the second term of equation (7) becomes zero and the quantity offered on each venue is identical to that of a single exchange setting. However, while consolidated liquidity is twice that of the single exchange setting, investors are not better off since they can trade on one venue only. Effectively, duplicates of limit orders are placed on both venues, such that consolidated liquidity overstates liquidity available to investors by a factor of two.

The market segmentation that arises when $\gamma < 1$ leads to important dynamic changes in market makers' liquidity supply, i.e., before and after a trade. This is the main prediction of the model.

**Proposition 3** *Define the "consolidated liquidity impact" of a trade as the difference in equilibrium consolidated liquidity before and after a one unit trade.*

1. *When $\gamma = 1$ or in the single exchange setting, the consolidated liquidity impact equals one.*

[12] The signs of the derivatives with respect to $\pi$ and $Q_{B1}$ hold on the relevant domain of $Q_{A1}$, i.e., for $Q_{A1}$ smaller or equal to the equilibrium solution.

2. *The consolidated liquidity impact decreases by $\gamma$. For $\gamma < 1$, the consolidated liquidity impact is strictly larger than one, such that the impact of a trade on liquidity is larger than the trade size.*

In the first case, a one unit trade reduces consolidated liquidity with one unit because the price impact of the trade equals the slope of the limit order book. However, a decrease in $\gamma$ reduces the market makers expected adverse selection costs, who in turn increase their liquidity supply. But given that the information content of a trade is held fixed, it must be that the increased liquidity is cancelled after the trade.[13] Effectively, the private information of a trade is incorporated on the competing venue via cancellations of limit orders. The numerical example of the next subsection explains this in more detail.

Market segmentation does not arise when $\gamma = 1$ or in the single exchange setting, so market makers have no incentive to place duplicate limit orders. The channel of duplicate limit orders occurs in addition to the effect that for $\gamma > 0$, consolidated liquidity increases because of competition between liquidity suppliers (the second part of Proposition (1)). In what follows, I define duplicate liquidity as the liquidity impact minus one, i.e., the impact of a trade on consolidated liquidity due to additional cancellations of limit orders.

Next we analyze the impact of a trade on the liquidity supply at the other venue.

**Proposition 4** *Define the "cross-venue liquidity impact" as the difference in equilibrium liquidity at the competing exchange before and after a one unit trade. The cross-venue liquidity impact as always negative, and increases in magnitude by the liquidity of the competing exchange. Further, a similar effect takes place on the bid side after a trade on the ask side, and vice versa.*

The cancellations are simply a consequence of the private information revealed by the trade. The price impact changes the market makers' expected fundamental value, and accordingly they cancel and resubmit their entire limit order schedules around this value. If market makers offer more liquidity on a venue, the cancellations following trades on the competing venue will be larger too. Proposition (2) describes the impact of the models' parameters on $Q_{A1}$, which directly translate into the cross-venue liquidity impact. Also, since a trade on the ask side increases the expected fundamental value, the liquidity supply

[13] The proposition that $\gamma$ reduces consolidated liquidity is proved in the Appendix, which makes this proof redundant.

on the bid side of both venues will increase accordingly. This holds symmetrically for trades on the bid side.

To summarize this section, a *reduction* in the fraction of SORT investors lowers liquidity via less competition between liquidity suppliers, but this negative effect is more than offset by the increase in duplicate limit orders. However, the duplicate limit orders will be cancelled after a trade, as shown in the numerical example of the next section.

## 2.7 Numerical example

In this section I substitute the models parameters with realistic values and analyze the equilibrium. In particular, we are interested in the impact of the fraction of traders that might go to one venue only ($\alpha$ and $\beta$) on outstanding liquidity of both venues.

I choose the following parameter values. The average trade size is 1 unit ($\phi = 1$), the best ask price is £10.00 and the tick size is 0.5 cent, which is the relevant case for the sample stocks with a price of £10.00. The fixed order execution cost $c$ is 0.1 cent (one fifth of the tick size). The price impact of a one unit trade $\lambda = 20$ basis points, and the tie-breaking rule $\pi = 0.5$. The models' results come out clearest when I set the fundamental value $X_t$ just above £9.99, such that the depth in the order book is constant at each price level for the case that all investors use SORT ($\gamma = 1$).[14]

The numerical outcomes for the four best price levels are shown in Table (1), where $\alpha$ and $\beta$ vary. The first row shows the single exchange setting, $\alpha = 1$, which is the Sandås (2001) solution. This is the benchmark case, and shows that 1.53 units are offered on the best price level, and 2.5 units on all subsequent price levels. In this case, the quantities offered beyond the best price are constant because of the tradeoff between improved prices and higher adverse selection costs, which equals the tick size divided by the price impact.

The second column shows the situation where all investors use smart order routing technology ($\gamma = 1, \alpha = \beta = 0$). Compared to the benchmark case, consolidated liquidity is 63% higher on price $p_1$ (2.5 versus 1.53), and identical on all subsequent levels. This corresponds to part 2 of Proposition (1), and shows that liquidity summed over the first price level and beyond indeed dominates the benchmark case. Consistent with competition

[14] Specifically, I set the fundamental value $X_t = 9.993943$. Small changes in $X_t$ relative to the fixed pricing grid cause changes in offered liquidity at the best price, which interact with liquidity on the competing exchange and in turn with liquidity deeper in the order book.

between market makers, the SORT traders reduce the expected profits to market makers with 33% (from 2.24 to 1.50 basis points per trade, in the bottom part of the Table).

From columns three to six the fraction of SORT traders gradually decreases, which increases consolidated liquidity (part 1 of Proposition (1)). When we move towards the full duplicate limit orders case ($\gamma = 0$, column (6)), consolidated liquidity is twice that of the single exchange setting. In each case, the Table reports the market shares and the per trade expected profits to markets makers of both venues.

The main prediction of the model is that a higher fraction of SORT traders reduces the amount of duplicate limit orders (Proposition (3)). To illustrate this point, we analyze the impact of a 2.5 share trade on the liquidity of both limit order books. The price impact is $2.5\lambda = 1$ tick, meaning that all prices shift up exactly one level after market makers have revised their limit orders. When all investors use SORT (column (2)), the trade consumes the 1.25 units of $Q_{A1}$ and 1.25 of $Q_{B1}$, and then the limit order books are immediately in equilibrium (market makers will not need to revise limit orders on the ask side). In effect, there are no duplicate limit orders and the consolidated liquidity impact of a trade is exactly one. In contrast, when no investors use SORT (column (6)), the trade will consume half of the offered liquidity on both venues (1.25 of $Q_{A1}$ and $Q_{B1}$), but then the remaining shares will be cancelled because of the price impact of these trades. In effect, after the market makers revision, the 2.5 share trades reduce liquidity on both venues with 5 shares, meaning that 100% of the order size is cancelled.

The second prediction of the model is that the cross-venue liquidity impact of a trade depends on the liquidity of the competing exchange (Proposition (4)). We confirm this in column (7), where 50% of the investors have access to venue A only and 50% use SORT. Thus, venue A is very liquid compared to B. As above, when we trade 1.25 on venue A and B simultaneously, the entire liquidity schedules shift up one price level. Therefore, cancellations on $Q_{A1}$ are 0.17 (from 1.42-1.25 to 0), and $Q_{A2}$ are 1.12 (from 2.37 to 1.42), whereas $Q_{B1}$ are 0 and $Q_{B2}$ is -0.10, a replenishment (from 1.15 to 1.25). In general, the replenishment implies that after a trade on one exchange, limit orders may get cancelled on the competing venue and replaced on the current venue to restore equilibrium.

Indeed, column (7) shows the realistic setting in Europe, where a large fraction of investors is able to trade only on the traditional venue A (50%), and the remaining investors use SORT to access a new entrant venue B. The model correctly describes the following stylized facts. Despite that only SORT investors trade on venue B, liquidity is still fairly high here: 88% relative to venue A at $p_1$, which reduces to 29% when we sum liquidity on levels $p_1$ to $p_4$. This is consistent with the stocks analyzed empirically. In addition, the liquidity at $p_1$ of venue A is about 93% of that in the single venue setting in column (1), which closely matches the finding of Degryse et al. (2011) that competition between exchanges reduces liquidity of the traditional market with about 10%. Lastly, while the entrant offers a substantial fraction of total liquidity, it has a relatively low market share of trading volume (22%, highly consistent with the empirical results).

# 3 Empirical results

This section first presents a brief overview of the sample stocks trading environment, followed by a data description and an explanation of the liquidity measures, the DepthAsk(X) and DepthBid(X). Then I test the duplicate limit order hypothesis and discuss the results.

## 3.1 Market structure FTSE100 stocks

The FTSE100 stocks are primarily listed on the London Stock Exchange (LSE), which is the fourth largest stock exchange in the world. In the sample period, November 2009, the LSE executes approximately 61% of trading volume (excluding dark pool and OTC volumes).[15] These stocks are traded on the trading system SETS, an electronic limit order market organised by the LSE which integrates market makers liquidity provision. Note that the market makers in the model are regular investors, who use high-frequency technology to operate like traditional market makers. Continuous trading occurs between 08:00 and 16:30, local time.

Once stocks are listed on the LSE, alternative trading venues may decide to organize trading in them as well.[16] Four important entrants have emerged which also employ publicly displayed limit order books: Chi-X, Bats, Turquoise and Nasdaq OMX Europe. These venues are regulated as Multi-lateral Trading Facilities (MTFs), the European equivalent to ECNs. While these entrants in effect have the same market model as the LSE, they differ with respect to trading technology (speed in particular), fixed and variable trading fees, and some of the types of orders that may be placed (e.g., pegging a limit order price

[15] As reported by Fidessa, see http://fragmentation.fidessa.com.

[16] This feature makes the current study inherently different from literature on cross-listings, where firms may choose to list on several exchanges to improve access to global capital.

to the midpoint, such that it always equals the midpoint $+n$ ticks). Investors can demand ("take") liquidity by issuing a market order or supply ("make") liquidity by issuing limit orders at any moment in time. All markets allow for visible, partially hidden (iceberg) and fully hidden limit orders. The hidden portion of iceberg orders becomes visible after (partial) execution of the visible part. Accordingly, limit orders have a price, transparency, time priority within a trading venue, but not between trading venues.

Chi-X started trading in April 2007 and is the most succesful entrant in terms of market share with 24% of trading volume in November 2009. Turquoise and Nasdaq OMX started trading FTSE 100 firms as of September 2008, and Bats two months later. Their market shares are substantially lower, with 5.5%, 1.8% and 7.6%, respectively. In May 2010 Nasdaq OMX closed down, as they did not meet their targeted market shares.[17] As of July 2009, the five trading venues use identical tick sizes, which depend on the stock price. All the new competitors employ a maker - taker pricing schedule, where executed limit orders receive a rebate of 0.18 to 0.20 basis points, while market orders are charged 0.28 to 0.30 basis points of traded value. These make-take fees are relatively small compared to a ticksize of 5 basis points (0.5 pennies) for a £10.00 stock.

The trading venues with publicly displayed limit order books execute approximately 60% of total volume, while the remaining 40% is executed on dark pools, Broker-Dealer Crossing Networks, internalized and Over-The-Counter.

## 3.2 Data

The current analysis is based on a subsample of ten FTSE100 stocks, each randomly selected from one market cap decile of the 100 constituents (i.e., a size stratified sample).[18] The sample period consists of 10 trading days (November 2 - 13, 2009), and high-frequency data are taken from the Thomson Reuters Tick History database. For each stock, the data contain separate limit order books for the five trading venues.

For each transaction, I observe the price, traded quantity and execution time to the millisecond,[19] while for each limit order placement, modification or cancellation, the dataset

[17] See "Nasdaq OMX to close pan-European equity MTF", www.thetradenews.com.

[18] I choose ten stocks during ten trading days as computational limitations prevents me from using the full sample of stocks or more trading days.

[19] If a single market order is executed against several outstanding limit orders, separate messages are generated for each limit order.

reports the timestamp and the ten best prevailing bid and offer prices and their associated quantities.[20]

While the time stamp is per millisecond, I take snapshots of the limit order books at the end of every $100^{th}$ millisecond, resulting in approximately 30 million observations. Higher frequencies are not useful when comparing multiple trading venues, as it may lead to inaccuracies because of latency issues, i.e. millisecond reporting delays. Per snapshot, I observe the available liquidity and that periods trading volumes on the buy and sell sides of every venue. The advantage of taking snapshots is that the data become evenly spaced, such that every observation receives the same weight in the regressions. Accordingly, lagged variables in the regressions become easily interpretable.

I do not directly observe hidden and iceberg limit orders. However, from trades I can construct the executed hidden quantity, based on the state of the order book directly before and after the trade and the traded quantity. As such, I do observe hidden liquidity that gets 'hit' by a market order. These data are identical to those offered by several information vendors, meaning I use the information set available to the market.

Panel A of Table 2 presents summary statistics for the sample stocks. As I select stocks from within each size decile, there is a large variation in market cap: the mean is £21 billion with a £37 billion standard deviation. Accordingly, also trading volume (in shares and pounds) and realized volatility vary substantially. A large part of this variation stems from Itv PLC, the smallest stock in the sample. In contrast, the market shares of the five trading venues are fairly stable between firms, and highly representative for the entire FTSE100 index.

Panel B of Table 2 presents summary statistics on the average number of limit orders and transactions per minute. While the LSE's market share is largest by far, the number of transactions lie much closer together (i.e., Chi-X trades are smaller on average). On top of that, the number of limit orders on Chi-X greatly exceeds the LSEs, on average 218 versus 160 per minute.

Worth mentioning is the ratio of limit orders to trades, which is 31:1 for the LSE, 51:1 for Chi-X and increases as a venues market share goes down to 123:1 for Nasdaq. This is mostly due to high-frequency traders placing many limit orders, and shows that new entrants are highly active despite a small overall market share. Chi-X and Bats are the

[20] This feature of the data makes it difficult to follow a limit order over time, because its location in the order book can changes and may even fall outside the observable range of ten price levels.

most succesfull new competing venues in terms of market share, number of transactions and limit order book activity.

Finally, in Table 3 reports summary statistics on trading volume per minute, denominated in GBPs. For each venue I show the total buy and sell volume (left panel), and the volume executed against hidden limit orders (right panel). Overall, the buy and sell sides are fairly symmetric.

## 3.3 The DepthAsk(X) and DepthBid(X) liquidity measures

This subsection explains the DepthAsk(X) and DepthBid(X) measures, also used in Degryse et al. (2011).

The DepthAsk(X) aggregates all shares offered at prices between the midpoint and the midpoint plus $X$ basis points. Similarly, the DepthBid(X) sums the shares offered within the midpoint minus $X$ basis points and the midpoint. The midpoint is the average of the best bid and ask price available in the market, and I choose $X = 10$ basis points relative to the midpoint. The price constraint $X$ garantuees that I sum liquidity at prices close to the midpoint, i.e. only at good price levels. This is important, as liquidity offered deeper in the order book is less likely to be executed, and therefore less relevant to investors. The number of shares in the interval are then converted to the value in GBPs.

Formally, define price level $j = 1, 2, ..., J$ on the pricing grid and the midpoint $M$, the average of the best ask and bid price available in the market, then for venue $v$,

$$DepthAsk(X)_v = \sum_{j=1}^{J} P_{j,v}^{Ask} Q_{j,v}^{Ask} \mathbf{1}\left(P_{j,v}^{Ask} < M(1+X)\right), \tag{9a}$$

$$DepthBid(X)_v = \sum_{j=1}^{J} P_{j,v}^{Bid} Q_{j,v}^{Bid} \mathbf{1}\left(P_{j,v}^{Bid} > M(1-X)\right). \tag{9b}$$

The measures are calculated at the end of every 100 millisecond interval and represent liquidity offered at the bid and ask side, per trading venue. When taking higher values for $X$, liquidity deeper in the order book is also incorporated. Then, comparing different price levels $X$ reveals the shape of the order book. For example, if the depth measure increases rapidly in $X$, the order book is deep while if it increases only slowly, the order book is relatively thin. The order book is asymetric when the absolute difference between

DepthAsk(X) and DepthBid(X) is high.

The measure has several features that make it highly suitable for the empirical approach. First, the measure is calculated per venue, for the bid and ask side, which allows us to analyze correlations between the provision and consumption of liquidity across venues, sides and over time. Second, the measure can directly be related to trading volumes, as both are denoted in pounds. Third, the Depth measure incorporates limit orders beyond the best price levels, making it robust to small, price improving limit orders. Such orders are often placed by high-frequency traders, who mostly drive the dynamics in the model. Fourth, by choosing a fixed interval the measure is independent of the tick size, which varies across stocks. For a detailed discussion and comparison of this measure with related liquidity measures such as the Cost of Roundtrip ($CRT(D)$) and Exchange Liquidity Measure ($XLM(V)$), I refer the interested reader to Degryse et al. (2011).

Table 4 contains summary statistics on the Depth(10) and Depth(50) measures for the bid and ask side, reported in GBPs and calculated per exchange. The statistics are based on single observations per tenth of a second per stock, equal weighted over all stocks. Depending on the tick sizes, the Depth(10) aggregates liquidity of two to five price levels on the bid and ask side. First and most strikingly, we observe that the liquidity offered on Chi-X is 86% of the LSE, while they execute only a third of the LSE volume. As predicted by the model, this implies that indeed a substantial fraction of investors only have access to the LSE. The liquidity available at Bats is roughly 40% of the liquidity at the LSE, while Turquoise and Bats have approximately 20% each. The ask side contains on average 3% more liquidity than the bid side, meaning that the order books are very symmetrical.

The regression analysis works with *changes* in DepthAsk(X) and DepthBid(X), i.e., the value of the current minus the previous observation. As Eq. (9) shows, these changes depend on the activity in the limit order book and on the level of the midpoint. The model relates changes in the depth measures due to limit order book activity only, i.e., the placement, cancellation, modification and execution of limit orders. Therefore, I define $Chg\_DepthAsk(X)$ as the difference in DepthAsk between each period, holding the midpoint constant

$$Chg\_DepthAsk(X)_{i,t} = DepthAsk(X, M_{t-1})_{i,t} - DepthAsk(X, M_{t-1})_{i,t-1}. \tag{10}$$

The measure simply shows how much liquidity in GBPs is added or removed from one period to the next.

## 3.4 Methodology

The model predicts that high-frequency traders operating as market makers place duplicate limit orders on several exchanges. Therefore, we expect that a trade is followed by cancellations of limit orders on the same side of the limit order books of competing venues. This prediction holds equally for buy and sell trades. The second prediction of the model is that the cancellations should occur to a greater extent on venues with a high share of non-SORT traders, i.e., the traditional exchange. The reason is that the market makers place more duplicate limit orders on the LSE than on Chi-X, which in turn will be cancelled after a Chi-X trade. The third prediction, consistent with theories of asymmetric information, is that new limit orders will be placed on the bid side after a buy trade (and on the ask side after a sell), including some duplicate limit orders.

Cancellations of limit orders occur when the $Chg\_DepthAsk(10)$ is negative (after controlling for trading volume). Thus, I regress the $Chg\_DepthAsk(10)$ of venue $v$ on contemporeneous and lagged buy and sell volumes of all venues. I add lags of ten seconds, which are 100 periods as the data are sampled at a 100 millisecond frequency, because the model's predictions apply to a high-frequency trading environment and should be incorporated very quickly. Instead of estimating 100 individual lagged coefficients, I add five variables that average trading volume of 1, 2-4, 5-10, 11-20 and 21-100 periods away, per venue for buy and sell volumes. Section A.2 in the appendix explains in more detail how I obtain the cumulative impact of a transaction over time.

A trade is classified as $Buy$ or $Sell$, and define trading venue $v = 1, ..., 5$, for the current and five lagged groups $l$, stock $i$ and time $t$. I test the model's predictions with the following regressions:

$$\begin{aligned} Chg\_DepthAsk(10)^V_{it} \quad = \quad & c_i + \sum_{v=1}^{5}\sum_{l=0}^{5}\left(\beta^{Buy}_{l,v} \times Buy^v_{it-l} + \beta^{Sell}_{l,v} \times Sell^v_{it-l}\right) + \\ & \sum_{v=1}^{5}\left(\beta^{Buy}_{v} \times BuyHid^v_{it} + \beta^{Sell}_{v} \times SellHid^v_{it}\right) + \varepsilon_{it}. \qquad (11) \end{aligned}$$

The term after the firm fixed effects represents the buy and sell volumes (in GBPs) for the five venues covering the six lagged groups. The term on the second line controls for contemporeneous hidden buy and sell liquidity (observed when executed), which is added for the following reason. The effect of a buy trade on $Chg\_DepthAsk(10)$ of that venue should mechanically be $-1$ : a one pound trade reduces the depth with exactly one pound. However, since a trade executed against a hidden limit order does not reduce DepthAsk(10), I control for executed hidden liquidity.

This regression is executed ten times: for the bid and ask sides of five trading venues. It shows how many pounds close to the midpoint are submitted or cancelled after a one pound buy or sell trade on some venue. Note that the effects do not die out over time, as for example a buy trade might contain positive price information, such that some limit orders will permanently be cancelled on the ask side.

## 3.5 Results

The regression results are reported in Table (5), with the change in DepthAsk and DepthBid of all venues as dependent variables. Each column represents one regression, showing separate coefficients for buy and sell volumes, per trading venue. The dependent and independent variables are all measured in GBPs. Within each venue, the displayed coefficients represent the cumulative effect over time (the running sum). I only show the effect within one tenth of a second, after 1 second and after 10 seconds. Intermediate lagged values are estimated to improve the model fit, but for brevity not reported. Next I discuss the findings for the DepthAsk only, as the results for the DepthBid are symmetric.

In line with the Proposition (4), the first column shows that a one pound buy trade at Chi-X is immediately followed by cancellations on the LSE of -0.21 pounds (-21%). After ten seconds, the effect is -0.61, meaning that more than half of the Chi-X trade size is cancelled on the LSE. The coefficients for Bats are similar, -0.27 immediately and -0.54 after ten seconds (all significant at the 1% level). This effect is economically very large, and implies that trades on entrant venues are immediately followed by cancellations on the traditional market. Note that the effect cannot be explained by investors who simultaneously place trades on several venues, since the regression controls for trades on other venues. Also, the regression controls for the execution of hidden limit orders. The effect of Nasdaq and Turquoise trades on LSE liquidity are negative, but surprisingly small.[21]

The immediate effect of a one pound LSE buy trade on LSE DepthAsk is -0.83 pounds. This implies that while the trade removed 1 pound, either 17 cents is immediately replen-

[21] I need to investigate the possibility that Nasdaq and Turquoise mainly attract trades when only they offer the best price in the market. In this case, we would not expect cancellations.

ished, or, first a new limit order is placed which immediately provokes the trade. The latter explanation is consistent with the findings of Hasbrouck and Saar (2009). A Chi-X buy trade reduces Chi-X DepthAsk with -1.31 (column (2)), implying that beyond the reduction of 1 pound, an additional and significant 31 cents is cancelled. Coefficients of the other venues lie between -0.70 and -1.26.

Consistent with the second part of Proposition (4), the LSE indeed responds more strongly to Chi-X and Bats trades than vice versa. In column (2) and (3), LSE trades reduce liquidity on Chi-X and Bats with -0.18 and -0.05 after ten seconds (compared to -0.61 and -0.54 above).

The cancellation effect also holds particularly strong between Bats and Chi-X, in both directions. Bats and Chi-X have immediate cross-coefficients of -0.58 and -0.18 (column (2-3)), and are the most succesfull entrants in terms of market share. These findings suggest that a fraction of high-frequency market makers operate on the LSE, Chi-X and Bats simultaneously, generating the strong correlations of liquidity between these venues. In contrast, Turquoise and Nasdaq seem more independent, as their liquidity does not respond much to trades on the LSE, Chi-X and Bats and vice versa.[22]

The immediate effect of any venues sell trades on LSE DepthAsk(10) is economically large and positive, with coeffients ranging from 0.10 to 0.30 (column (1), bottom panel). This result is consistent with an information effect: the sell trade conveys negative information about the stock, such that market makers improve prices (and quantities) of their ask limit orders. In addition, the finding is consistent with selling investors who use algorithms that optimally place market orders on the bid sides and new limit orders on the ask sides (as predicted by theory of e.g., Parlour (1998)).

Concluding, the results confirm that consolidated liquidity is overstated in a fragmented market, because a single transaction is followed by substantial cancellations of limit orders on the other venues. In the model, these cancellations would not occur in a single venue setting or when all traders use smart order routing technology.

[22] A possible explanation is that at times, Nasdaq and Turquoise offer zero DepthAsk(10). Obviously, in such periods their liquidity does not respond to competitors trades, pushing the estimated coefficients toward zero.

### 3.6 Drivers of the cancellation effect

In this section I analyze the driving forces behind the cancellation effect. In particular, I am interested in the variation in cancellation rates of the venues across stocks and over time, and how they are influenced by observable market characteristics.

I estimate the previous model for hour $t$ of stock $i$, but use as dependent variable the change in ask (or bid) liquidity summed over all venues, i.e., the change in consolidated Depth(10). These regressions represent the impact of a trade at venue $v$ on market wide depth, consistent with Proposition (3). In a new dataset I store the estimated coefficients $Coef_{v,it}$ accumulated over one second after the trade, for buy trades on the ask side liquidity and sell trades on the bid side liquidity. The sample size is extended from 10 to all 21 trading days in November 2009 for the same 10 stocks used in the previous analysis, which results in 1890 observations. Next, I filter the data by dropping observations where the trading venue $v$ of $Coef_{v,it}$ has executed less than £10.000 or 50 transactions. Without this restriction, the estimated coefficients on the cancellation rates have too high standard errors and are not economically meaningful. As a consequence, we disregard the regressions of Turquoise and Nasdaq OMX altogether because these venues are insufficiently active.[23] In addition, I winsorize $Coef_{v,it}$ at the 1% and 99% level to reduce the impact of outliers.

Based on the new dataset, I run the following two regressions for firm $i$, venue $V = \{LSE, Chi\text{-}X, Bats\}$, and hour $h$

$$
\begin{aligned}
Coef_{V,ih} &= c_i + \delta_h + Frag_{ih} + LS\_LSE_{ih} + \varepsilon_{ih}, &(12)\\
Coef_{V,ih} &= c_i + \delta_{(h)} + Frag_{ih} + LS\_LSE_{ih} + Ln(entrant\ trades)_{ih} + Order\ Imb_{ih} + \\
&\quad Volat_{ih} + Ln(turnover)_{ih} + Ln(Depth(10)\ cons)_{ih} + \varepsilon_{ih}. &(13)
\end{aligned}
$$

Using estimated coefficients as dependent variable in a second step regression does not bias the coefficients of the second step.[24] $Frag$ is the degree of fragmentation of stock-hour $ih$, defined as (1-HHI). HHI is the sum of squared market shares of the five venues based on trading volume (also used in Degryse et al. (2011)). $LS\_LSE$ is the liquidity share of the LSE, defined as the ratio of LSE Depth(10) over consolidated Depth(10). $Ln(entranttrades)$ is the logarithm of the number of transactions at entrant venues. Order

[23] We do incorporate these venues when calculating consolidated liquidity, although the results are qualitatively unaffected when leaving them out.

[24] The measurement error from the first step only increases the standard errors of the coefficients in the second step.

imbalance ($Order\ Imb$) is the difference between the logarithm of buy volume and sell volume. $Volat$ is the realized volatility based on 5-minute underlying stock returns, calculated as the sum of 12 squared 5-minute returns, hour by hour. $Ln(turnover)$ and $Ln(Depth(10)\ cons)$ are the logarithms of trading volume (in GBP) and consolidated Depth(10). We add firm fixed effects and daily dummy variables $\delta_{(h)}$ to absorb fixed stock and day characteristics.

Summary statistics of the regression variables are presented in Table (6). The top panel shows that the degree of fragmentation is fairly constant across stocks and over time, with a mean of 0.59 and standard deviation of 0.09. This also holds for the LSE liquidity share, with a standard deviation of 0.11. In contrast, the number of trades on the entrants vary substantially, with a mean of 435 and a standard deviation of 485. The order imbalance is highly volatile too, since a one standard deviation increase implies that buy volume is 43% larger than sell volume. The bottom panel shows the cancellation rates of trades at the individual venues on consolidated liquidity. The impact on consolidated liquidity is approximately equal to the sum of the impact on the individual venues, as reported in Table (5). The standard deviations of Turquoise and Nasdaq are very large, which justifies excluding them from further analysis.

The regression results for the three trading venues are presented in Table (7). $Frag$ should be a good proxy for the number of traders with SORT technology, which corresponds to Proposition (3). In the model, market makers face higher adverse selection costs when trading with SORT traders, because there is a risk that the SORT trader already bought shares on the competing venue, which in turn implies a larger and more informed trade on average. Thus, we expect that $Frag$ correlates negatively with the amount of cancellations.

The coefficient on $Frag$ is indeed negative for Chi-X and Bats (columns (2-3)), but not for the LSE (column (1)). While this result seems contrasting, a likely explanation is that $Frag$ also proxies for the activity of market makers who operate on several venues. Then, a higher level of fragmentation implies more cross-market activity and *more* cancellations. This explanation mainly holds for trades on the LSE, which is always active, whereas Chi-X and Bats are only active when the market is fragmented.[25] When I control for other variables that proxy for the activity of cross-market market makers (columns (4-6)), $Frag$ becomes strongly negative (and significant). Then, a one standard deviation increase in

[25] An alternative explanation is that some liquidity providers on the LSE do not operate on the entrant venues, i.e., do not cancel their limit orders after they trade on the LSE.

$Frag$ (0.09) reduces cancellations of LSE trades by 5% and Chi-X and Bats trades by 30% and 43%.

The liquidity share of the LSE is negatively correlated with the amount of duplicate limit orders, and significant in all regressions. A large LSE liquidity share means that market makers are less active on the entrants which results in fewer cross-market activity and cancellations. These coefficients are economically large, as a one standard deviation change reduces the cancellation rates with 20-40% (depending on the trading venue).

The remaining control variables in columns (4-6) have the expected signs. An increasing number of entrant trades increases the cancellation rates, which is consistent with market makers operating more actively on entrant venues. Similarly, more overall liquidity (consolidated depth) is consistent with a higher amount of duplicate limit orders. Turnover has a negative sign, which means that in periods of high trading activity the market makers place less duplicate limit orders.

The bottom panel of Table (7) shows the results using the cancellation effects of sell trades on bid liquidity as dependent variable, and are highly symmetric.

# 4 Conclusion

Equity markets have evolved rapidly in recent years due to the increasing number of trading venues and heavy investments in high-frequency trading technology. I show that a specific type of high-frequency traders, those who operate like modern day market makers, might in fact cause a strong overestimation of liquidity aggregated across trading venues. The reason is that these market makers place duplicate limit orders on several venues, and after execution of one limit order they quickly cancel their outstanding limit orders on competing venues. As a result, a single trade on one venue is followed by reductions in liquidity on all other venues.

The empirical analysis confirms that trades are followed by substantial cancellations on competitors. That is, within 100 milliseconds after trades on some venues $39-85\%$ of the order size is cancelled on competitors. After one second this number increases to $98-125\%$, which shows that the impact of a trade on liquidity is in fact twice the trade size. Note that the reduction in liquidity is due to cancellations of limit orders, since the analysis controls for transactions on all individual trading venues.

The analysis is executed on a sample of FTSE100 stocks, which are fairly fragmented in the sample period. The main advantage of using these stocks is the rich data: I observe the entire limit order books of the five trading venues that compete for these stocks. The data represent all publicly displayed liquidity, which is the same information set available to the general public. By sampling the data at one observation per tenth of a second, I study high-frequency trading behavior. The paper contributes to the literature by analyzing the impact of high-frequency trading strategies on the demand and supply of liquidity across exchanges.

The results relate to the benefits and drawbacks of equity market fragmentation. While previous research typically shows a beneficial effect of fragmentation on liquidity, the analysis argues that the benefits, while still positive, are mitigated because of duplicate limit orders. Note that duplicate limit orders will not arise when all investors use smart order routing technology (SORT).

An additional result of the model is that liquidity shocks are highly correlated across trading venues. Therefore, static measures of consolidated liquidity (and quoted depth) overstate the liquidity available to investors, since a single trade reduces liquidity on all venues. However, liquidity measures such as the quoted spread and effective spread are unaffected by the duplicate limit orders. My results are particularly important to algorithms designed to minimize execution costs by splitting up trades over time. Indeed, the strong cancellations mitigate the benefits of order splitting—a result that cannot be observed by static liquidity measures.

My model focusses on two exchanges only, but can already predict a substantial fraction of duplicate limit orders. Therefore, the relevance of the model is only strenghtened by the fact that most European stocks trade on more than four exchanges (with publicly displayed limit order books) and some US stocks on so much as twelve exchanges. A larger number of trading venues encourages market makers to duplicate their limit orders.

The predictions of the model are very relevant to US markets too. Indeed, it seems that all US traders use SORT because the regulator prohibits the execution of trades at prices inferior to the best available price. However, duplicate limit orders will still arise when some traders are unable to split up trades *simultaneously* across venues. When a trader leaves a millisecond delay between the split, high-frequency traders can observe the first part of the trade and quickly cancel duplicate limit orders on other venues before the second part of the trade arrives. Therefore, the results of the model also hold when the trading speeds varies across investors.

The main policy implication of the model is that fair markets require traders to have simultaneous access to all trading venues. In any other case non-SORT or slow traders pay a higher price for liquidity, which is transfered to the SORT traders. This result is consistent with McInish and Upson (2011), who find that high-frequency traders are able to exploit slower traders in other ways too (see also Biais et al. (2011) and Hoffmann (2011)). A second implication is that the decision of a trader to acquire SORT increases the liquidity she has access to, but also imposes a negative externality on other traders. The overall welfare effects depend on the cost of acquiring smart routing technology, and are therefore ambiguous.

# A Appendix

## A.1 Theory

This section summarizes the proofs of the propositions in the model.

- The consolidated liquidity in a fragmented market is strictly larger than liquidity in a single exchange setting (Proposition (1)).

**Proof.** Denote the liquidity in a single exchange setting as $Q_1$, which is the solution from Sandås (2001). I repeat equation (7), and rewrite as

$$Q_{A1} = \frac{p_1 - c - X_t - \lambda\phi}{\lambda} - \frac{\gamma(1-\pi)Q_{B1}}{\gamma(1-\pi) + (\alpha+\gamma\pi)\exp(\frac{Q_{B1}}{\phi})} \equiv Q_1 - c_{B1}Q_{B1} \qquad \text{A1(a)}$$

$$Q_{B1} = \frac{p_1 - c - X_t - \lambda\phi}{\lambda} - \frac{\gamma\pi Q_{A1}}{\gamma\pi + (1-(\alpha+\gamma\pi))\exp(\frac{Q_{A1}}{\phi})} \equiv Q_1 - c_{A1}Q_{A1}. \qquad \text{A1(b)}$$

I need to show that $Q_{A1} + Q_{B1} > Q_1$, which is equivalent to $c_{B1}Q_{B1} + c_{A1}Q_{A1} < Q_1$.

$$\begin{aligned} c_{B1}Q_{B1} + c_{A1}Q_{A1} &= c_{B1}Q_{B1} + c_{A1}Q_1 - c_{A1}c_{B1}Q_{B1} \\ &= c_{A1}Q_1 + (1-c_{A1})c_{B1}Q_{B1} \\ &< c_{A1}Q_1 + (1-c_{A1})Q_1 = Q_1. \end{aligned}$$

In the first equality I simply replace $Q_{A1}$ with the first line of equation A1(a); which I then rewrite in the second line. The inequality holds because $c_{B1}Q_{B1} < Q_1$, since $c_{B1} < 1$ and $Q_{B1} \leq Q_1$. ■

- Both venues co-exist when $\alpha < 1$ and $\beta < 1$. Co-existence means that both venues attract a positive amount of liquidity and market share.

**Proof.** It is sufficient to show that liquidity on both venues is strictly positive, because SORT traders will always use liquidity of both venues when the incoming order size is sufficiently large (i.e., when the size goes to infinity). This is proved in the last equality above, that $Q_{A1} = Q_1 - c_{B1}Q_{B1} > 0$, and symmetrically for venue B. ■

- The first derivates of the solution of $Q_{A1}$ with respect to $\pi, \gamma, \alpha, \beta$ and $Q_{B1}$ are straightforward and not reported, but available upon request (Proposition (2)).

- The market makers' profit of limit orders on price level 1 is given by

$$\int_{0}^{Q_{A1}} (\alpha + \gamma\pi) \int_{q}^{\infty} \Pi_i(x)\mathrm{d}x + \gamma(1-\pi) \int_{q+Q_{B1}}^{\infty} \Pi_i(x)\mathrm{d}x\mathrm{d}q. \tag{14}$$

Calculating the profit of market makers involves integrating over the profitability of all marginal shares at each price level. I calculate the profit of each marginal share by integrating over the incoming trade $x$, which are traders who either go immediately to venue A, $(\alpha + \gamma\pi)$, or traders that go to venue B first and then to A, $\gamma(1 - \pi)$. For brevity, denote $\Pi_i(x) = (p_i - X_t - c - \lambda x)f(x)$, then the market makers profit of all limit orders on price level 1 of venue A is

## A.2 Empirical

This section shows that the methodology in section 3.4 measures cumulative effects over time. That is, in regression (11) I add contemporeneous terms and lagged values of trading volumes 100 periods ago (i.e., ten seconds). Instead of estimating 100 coefficients, I create six variables representing the averaged lagged volumes of the current, 1, 2-4, 5-10, 11-20 and 21-100 periods away, per venue for buy and sell volumes. Define $t$ as the current period,

$i$ as the start and $j$ as the end of the intervals (e.g., $i = 2$ and $j = 4$). Then

$$Vol_{i-j} = \frac{1}{j - i + 1} \sum_{n=i}^{j} Vol_{t-n}.$$

The periods of lagged values are chosen such that they maximize the model fit. An example of how the data looks like is shown in the table below, where a £1.00 trade occurs on some venue at time $t = 1$. The first four columns show the values of the contemporeneous and three lagged groups. The fifth column shows the cumulative effect of the regression coefficients over time, calculated as a running sum of the individually estimated coefficients. By constructing the variables as averages, the long-term effect of a trade is simply the sum of the estimated coefficients. The standard errors are also calculated based on this sum.

**Data example of a trade at time $t = 1$.**
Each $\beta_t$ represents the estimated coefficient of lagged average volumes $t$ periods away, as described in regression (11).

| Time | $\text{Vol}_0$ | $\text{Vol}_1$ | $\text{Vol}_{2-4}$ | $\text{Vol}_{5-10}$ | Cumulative effect |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1.00 | 0 | 0 | 0 | $\beta_0$ |
| 2 | 0 | 1.00 | 0 | 0 | $\beta_0 + \beta_1$ |
| 3 | 0 | 0 | 0.33 | 0 | $\beta_0 + \beta_1 + 0.33\beta_{2-4}$ |
| 4 | 0 | 0 | 0.33 | 0 | $\beta_0 + \beta_1 + 0.66\beta_{2-4}$ |
| 5 | 0 | 0 | 0.33 | 0 | $\beta_0 + \beta_1 + \beta_{2-4}$ |
| 6 | 0 | 0 | 0 | 0.20 | $\beta_0 + \beta_1 + \beta_{2-4} + 0.2\beta_{5-10}$ |
| 7 | 0 | 0 | 0 | 0.20 | $\beta_0 + \beta_1 + \beta_{2-4} + 0.4\beta_{5-10}$ |

# References


B. Biais, C. Bisière, and C. Spatt. Imperfect competition in financial markets: An empirical study of island and nasdaq. *Management Science*, 56(12):2237–2250, 2010.

B. Biais, T. Foucault, and S. Moinas. Equilibrium high frequency trading. *Working Paper Toulouse School of Economics*, 2011.

S. Buti, B. Rindi, Y. Wen, and I. Werner. Tick size regulation, intermarket competition and sub-penny trading. *Working Paper Fisher College of Business, Ohio State University*, 2011.

Á. Cartea and J. Penalva. Where is the value in high frequency trading? *Working Paper Universidad Carlos III de Madrid*, 2011.

H. Degryse, F. de Jong, and V. van Kervel. The impact of dark trading and visible fragmentation on market quality. *Working paper Tilburg University*, 2011.

T. Foucault and A. Menkveld. Competition for order flow and smart order routing systems. *Journal of Finance*, 63(1):119–158, feb 2008. URL http://ideas.repec.org/p/ebg/heccah/0831.html.

L. Glosten. Is the electronic open limit order book inevitable? *Journal of Finance*, 49(4): 1127–1161, Sep. 1994.

L. Glosten. Competition, design of exchanges and welfare. *Working paper Columbia University*, 1998.

L. Glosten and L. Harris. Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21(1):123–142, may 1988.

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

F. Guilbaud and H. Pham. Optimal high frequency trading with limit and market orders. *Working paper Université Paris*, 2011.

J. Hasbrouck and G. Saar. Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets*, 12:143–172, 2009.

J. Hasbrouck and G. Saar. Low-latency trading. *Working paper Stern school of business*, 2011.

T. Ho and H. R. Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1):47–73, March 1981. URL http://ideas.repec.org/a/eee/jfinec/v9y1981i1p47-73.html.

P. Hoffmann. A dynamic limit order market with fast and slow traders. *Working Paper European Central Bank*, 2011.

B. Jovanovic and A. Menkveld. Middlemen in limit order markets. *Working paper New York University*, 2011.

A. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

T. McInish and J. Upson. Strategic liquidity supply in a market with fast and slow traders. *Working Paper University of Texas, El Paso*, 2011.

A. Menkveld. High frequency trading and the new-market makers. *Working paper Free University Amsterdam*, 2011.

M. O'Hara and M. Ye. Is market fragmentation harming market quality? *Journal of Financial Economics*, 100(3):459–474, June 2011.

M. Pagano. Trading volume and asset liquidity. *Quarterly Journal of Economics*, 104: 25–74, 1989.

C. Parlour. Price dynamics in limit order markets. *Review of Financial Studies*, 11(4): 789–816, 1998. doi: doi:10.1093/rfs/11.4.789.

P. Sandås. Adverse selection and competitive market making: Empirical evidence from a limit order market. *The Review of Financial Studies*, 14(3):705–734, 2001.

**Table (1) Numerical Example**
We solve the model for the quantities offered on the four best ask price levels of venue A (upper panel) and B (second panel). We vary the parameter values of the fraction of investors with access to venue A only ($\alpha$) and venue B only ($\beta$). The fraction of investors with smart order routing technology ($\gamma$) varies accordingly, defined as $1-\alpha-\beta$. The lower panels show the market shares and the per trade total expected profits to market makers of each venue. The expected profits are expressed in basis points relative to the midpoint. The remaining model parameters are held fixed. The average trade size is 1, with a per unit price impact of 20 basis points. The best ask price is £10.00, the tick size is 0.5 pennies and the fixed trading cost is one tenth of a cent. I specifically set the fundamental value to £9.993943, such that offered liquidity is constant at each price level when all investors use SORT (column 2). When both venues offer the best price, SORT traders are equally likely to go to venue A or B first ($\pi = 0.5$).

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| $\alpha$ | 1 | 0 | 0.1 | 0.1 | 0.4 | 0.5 | 0.5 |
| $\beta$ | 0 | 0 | 0 | 0.1 | 0.4 | 0.5 | 0 |
| $\gamma$ | 0 | 1 | 0.9 | 0.8 | 0.2 | 0 | 0.5 |
| Venue A | | | | | | | |
| $Q_{A1}$ | 1.53 | 1.25 | 1.29 | 1.30 | 1.46 | 1.53 | 1.42 |
| $Q_{A2}$ | 2.50 | 1.25 | 1.78 | 1.87 | 2.47 | 2.50 | 2.37 |
| $Q_{A3}$ | 2.50 | 1.25 | 2.49 | 2.86 | 2.58 | 2.50 | 2.55 |
| $Q_{A4}$ | 2.50 | 1.25 | 2.64 | 2.94 | 2.52 | 2.50 | 2.50 |
| Venue B | | | | | | | |
| $Q_{B1}$ | 0 | 1.25 | 1.25 | 1.30 | 1.46 | 1.53 | 1.25 |
| $Q_{B2}$ | 0 | 1.25 | 1.23 | 1.87 | 2.47 | 2.50 | 1.15 |
| $Q_{B3}$ | 0 | 1.25 | 0.78 | 2.86 | 2.58 | 2.50 | 0.15 |
| $Q_{B4}$ | 0 | 1.25 | 0.03 | 2.94 | 2.52 | 2.50 | 0.00 |
| Market shares | | | | | | | |
| A | 1.00 | 0.50 | 0.58 | 0.50 | 0.50 | 0.50 | 0.78 |
| B | 0.00 | 0.50 | 0.42 | 0.50 | 0.50 | 0.50 | 0.22 |
| Market maker profits in basis points | | | | | | | |
| A | 2.24 | 0.75 | 0.93 | 0.84 | 1.06 | 1.12 | 1.55 |
| B | 0.00 | 0.75 | 0.66 | 0.84 | 1.06 | 1.12 | 0.35 |

**Table (2) Summary statistics of sample firms.**
Summary statistics are presented for the 10 FTSE100 sample stocks, calculated using data of November 2009. Panel A shows market cap (in millions), price, average daily traded volume (millions of shares), turnover (millions of pounds), realized volatility and the market shares of the five trading venues. The trading venues are the London Stock Exchange (LSE), Chi-X, Bats Trading, Turquoise and Nasdaq OMX Europe. Panel B shows limit order book data, where statistics are equally weighted based on one observation per tenth of a second, per stock. For each venue, statistics on the average number of transactions and limit order updates per minute are shown, per venue. The stocks in our sample are Aviva, Hsbc Holdings Plc, Itv Plc, Kingfisher, Lonmin, National Grid, Pearson, Sage group, Vedanta Resources and Xstrata.

| | Mean | Stdev | Max |
|---|---|---|---|
| | Panel A: Stock characteristics | | |
| Market Cap | 21,062 | 37,827 | 125,930 |
| Price | 8.06 | 7.13 | 23.3 |
| Cumvolume | 527 | 554 | 1,883 |
| Turnover | 3,104 | 4,188 | 13,650 |
| Volatility | 3.92 | 1.38 | 6.10 |
| Share LSE | 66.2 | 5.24 | 73.8 |
| Share Chi-X | 20.5 | 3.52 | 25.8 |
| Share Bats | 6.44 | 1.68 | 9.49 |
| Share Turq. | 5.17 | 1.07 | 6.29 |
| Share Nasdaq | 1.77 | 0.65 | 2.73 |
| | Panel B: Limit order book | | |
| Trades LSE | 5.16 | 9.70 | 347 |
| Trades Chi-X | 4.32 | 7.46 | 283 |
| Trades Bats | 1.80 | 3.59 | 84 |
| Trades Turq. | 1.19 | 2.41 | 51 |
| Trades Nasdaq | 0.61 | 1.54 | 56 |
| Limits LSE | 159.84 | 216.31 | 6,999 |
| Limits Chi-X | 218.46 | 373.08 | 11,934 |
| Limits Bats | 123.92 | 231.60 | 7,308 |
| Limits Turq. | 98.25 | 146.95 | 3,296 |
| Limits Nasdaq | 75.63 | 154.66 | 7,081 |

**Table (3) Summary statistics of trading volume per minute.**
Buy and sell volumes in GBPs are reported, showing averages of volume per minute, equal weighted over 10 stocks and 10 trading days. The left panel shows all volume and the right panel volume executed against hidden limit orders.

| | Volume | | | Hidden | | |
|---|---|---|---|---|---|---|
| | Mean | Stdev | Max | Mean | Stdev | Max |
| Sell LSE | 39,960 | 131,673 | 12,910,331 | 2,692 | 35,785 | 4,001,096 |
| Sell Chi-X | 17,492 | 43,216 | 990,636 | 578 | 3,501 | 155,871 |
| Sell Bats | 5,909 | 18,996 | 1,149,016 | 798 | 4,344 | 150,729 |
| Sell Turq | 4,082 | 12,535 | 349,188 | 425 | 2,719 | 86,744 |
| Sell Nasdaq | 1,429 | 6,667 | 482,222 | 8 | 358 | 30,040 |
| Buy LSE | 41,989 | 149,532 | 12,280,297 | 3,122 | 37,028 | 1,644,776 |
| Buy Chi-X | 17,674 | 44,352 | 1,186,893 | 625 | 3,944 | 318,701 |
| Buy Bats | 5,404 | 18,715 | 1,536,919 | 760 | 4,217 | 209,244 |
| Buy Turq | 4,088 | 13,377 | 396,196 | 532 | 3,317 | 126,089 |
| Buy Nasdaq | 1,553 | 7,034 | 342,203 | 26 | 845 | 65,950 |

**Table (4) Summary statistics Depth(X) measure.**
Summary statistics are presented using limit order book data, containing one observation per tenth of second per stock, for 10 trading days in November 2009. The statistics are equally weighted over observations. The mean, standard deviation and maximum of Depth(10) and Depth(50) on the ask and bid side are shown. The Depth(10) on the ask side reflects the available liquidity, in pounds, offered with prices in the interval of Midpoint and Midpoint + 10bps. Similarly, the Depth(10) on the bid side reflects the liquidity offered with prices between Midpoint - 10bps and Midpoint. Depth(50) sums liquidity within 50 basis points from the midpoint. The trading venues are LSE, Chi-X, Bats Trading, Turquoise and Nasdaq OMX Europe.

| *Ask side* | Mean | Stdev | Max |
|---|---|---|---|
| Depth(10) LSE | 66,620 | 117,961 | 8,940,000 |
| Depth(10) Chi-X | 57,693 | 78,948 | 1,100,000 |
| Depth(10) Bats | 26,311 | 39,358 | 527,947 |
| Depth(10) Turq. | 14,915 | 18,817 | 411,974 |
| Depth(10) Nasdaq | 12,824 | 21,345 | 213,733 |
| Depth(50) LSE | 446,687 | 371,316 | 9,040,000 |
| Depth(50) Chi-X | 268,467 | 210,066 | 1,650,000 |
| Depth(50) Bats | 88,363 | 94,112 | 800,103 |
| Depth(50) Turq. | 71,495 | 47,170 | 570,524 |
| Depth(50) Nasdaq | 58,096 | 44,992 | 332,529 |
| *Bid side* | | | |
| Depth(10) LSE | 63,244 | 90,327 | 4,420,000 |
| Depth(10) Chi-X | 55,907 | 77,345 | 1,020,000 |
| Depth(10) Bats | 25,404 | 38,187 | 553,743 |
| Depth(10) Turq. | 14,273 | 19,277 | 525,517 |
| Depth(10) Nasdaq | 13,241 | 22,891 | 538,609 |
| Depth(50) LSE | 431,658 | 339,853 | 5,100,000 |
| Depth(50) Chi-X | 270,657 | 213,624 | 1,790,000 |
| Depth(50) Bats | 85,748 | 92,792 | 1,170,000 |
| Depth(50) Turq. | 74,360 | 51,850 | 638,623 |
| Depth(50) Nasdaq | 61,156 | 51,443 | 664,181 |

**Table (5)** **The cumulative impact of turnover on Depth(10).**

Each column represents one regression, showing the cumulative effect over time of buy and sell turnover on changes in DepthAsk(10) and DepthBid(10) of one venue. Changes in the DepthAsk(10) reflect changes in liquidity offered with prices in the interval of Midpoint and Midpoint + 10bps. These changes stem from limit order book activity (placement, cancellations and execution of limit orders). The data consist of one observation per tenth of a second, for each stock. The independent variables are contemporeneous and lagged buy and sell trading volumes for each of the five venues, in GBP. We show the cumulative effect over time (i.e., the running sum) of current trades, and trades one and ten seconds ago. Accordingly, each panel shows the immediate, short and long-term effects of one venues transactions on another venues liquidity. The regressions also contain executed hidden volume as control variables (not reported for brevity). Standard errors are clustered per firm - halfhour, a single asterix indicates significance at the 1% level.

| | | Ask Side | | | | | Bid Side | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| £ *Buys* | Sec | LSE | Chi-X | Bats | Turq. | Nasdaq | LSE | Chi-X | Bats | Turq. | Nasdaq |
| LSE | 0 | -0.83* | -0.25* | -0.09* | -0.02* | -0.02* | 0.28* | 0.24* | 0.09* | 0.02* | 0.01* |
| LSE | 1 | -0.80* | -0.30* | -0.14* | -0.04* | -0.05* | 0.35* | 0.31* | 0.15* | 0.05* | 0.05* |
| LSE | 10 | -0.67* | -0.18* | -0.05* | -0.03* | -0.04* | 0.33* | 0.23* | 0.09* | 0.04* | 0.04* |
| Chi-X | 0 | -0.21* | -1.31* | -0.18* | -0.02* | -0.03* | 0.26* | 0.68* | 0.18* | 0.03* | 0.03* |
| Chi-X | 1 | -0.52* | -1.47* | -0.46* | -0.09* | -0.13* | 0.50* | 1.00* | 0.45* | 0.13* | 0.12* |
| Chi-X | 10 | -0.61* | -1.29* | -0.37* | -0.09* | -0.13* | 0.67* | 1.11* | 0.46* | 0.15* | 0.14* |
| Bats | 0 | -0.27* | -0.58* | -1.26* | -0.04* | -0.10* | 0.37* | 0.60* | 0.51* | 0.05* | 0.08* |
| Bats | 1 | -0.46* | -0.79* | -1.21* | -0.07* | -0.17* | 0.52* | 0.88* | 0.69* | 0.09* | 0.16* |
| Bats | 10 | -0.54* | -0.83* | -1.01* | -0.08* | -0.15* | 0.87* | 1.16* | 0.81* | 0.14* | 0.16* |
| Turq | 0 | -0.04 | -0.04* | -0.05* | -0.70* | -0.03* | 0.13* | 0.11* | 0.08* | 0.14* | 0.04* |
| Turq | 1 | -0.11* | -0.08* | -0.02 | -0.69* | -0.04* | 0.22* | 0.15* | 0.06* | 0.19* | 0.08* |
| Turq | 10 | -0.13 | -0.06 | 0.04 | -0.68* | -0.02 | 0.17 | 0.13 | 0.01 | 0.18* | 0.07* |
| Nasdaq | 0 | -0.03 | 0.03 | -0.01 | 0.01 | -0.75* | -0.05 | -0.08 | 0.04 | 0.07* | 0.14* |
| Nasdaq | 1 | -0.08 | 0.08 | 0.04 | 0.04 | -0.63* | 0.00 | -0.18 | -0.08 | 0.02 | 0.12* |
| Nasdaq | 10 | -0.24 | 0.02 | 0.07 | 0.06 | -0.62* | 0.43 | -0.19 | -0.11 | 0.06 | 0.19* |
| £ *Sells* | | | | | | | | | | | |
| LSE | 0 | 0.30* | 0.27* | 0.10* | 0.02* | 0.02* | -0.78* | -0.29* | -0.10* | -0.02* | -0.02* |
| LSE | 1 | 0.38* | 0.35* | 0.17* | 0.06* | 0.06* | -0.75* | -0.35* | -0.16* | -0.05* | -0.06* |
| LSE | 10 | 0.39* | 0.29* | 0.12* | 0.05* | 0.05* | -0.65* | -0.23* | -0.06* | -0.04* | -0.04* |
| Chi-X | 0 | 0.27* | 0.70* | 0.19* | 0.02* | 0.02* | -0.24* | -1.28* | -0.20* | -0.02* | -0.03* |
| Chi-X | 1 | 0.46* | 0.99* | 0.43* | 0.12* | 0.11* | -0.51* | -1.40* | -0.43* | -0.08* | -0.13* |
| Chi-X | 10 | 0.57* | 1.08* | 0.44* | 0.13* | 0.12* | -0.53* | -1.15* | -0.31* | -0.08* | -0.13* |
| Bats | 0 | 0.29* | 0.56* | 0.43* | 0.04* | 0.08* | -0.21* | -0.58* | -1.15* | -0.05* | -0.10* |
| Bats | 1 | 0.37* | 0.81* | 0.62* | 0.07* | 0.15* | -0.38* | -0.77* | -1.11* | -0.09* | -0.16* |
| Bats | 10 | 0.53* | 1.00* | 0.70* | 0.09* | 0.16* | -0.41* | -0.69* | -0.88* | -0.08* | -0.15* |
| Turq | 0 | 0.10* | 0.08* | 0.08* | 0.13* | 0.04* | -0.06* | -0.03 | -0.07* | -0.63* | -0.05* |
| Turq | 1 | 0.17* | 0.12* | 0.06* | 0.18* | 0.08* | -0.08 | -0.03 | -0.06* | -0.63* | -0.03* |
| Turq | 10 | 0.33* | 0.11 | -0.02 | 0.20* | 0.09* | -0.17 | -0.02 | 0.00 | -0.59* | -0.04 |
| Nasdaq | 0 | 0.21* | 0.04 | 0.12* | 0.11* | 0.16* | -0.24* | -0.04 | -0.05 | -0.03 | -0.82* |
| Nasdaq | 1 | 0.22 | -0.01 | 0.07 | 0.11* | 0.20* | -0.43* | -0.04 | 0.00 | -0.03 | -0.76* |
| Nasdaq | 10 | 0.35 | -0.14 | -0.05 | 0.09* | 0.23* | -0.84* | -0.18 | -0.00 | -0.03 | -0.68* |
| R2 | | .123 | .115 | .069 | .025 | .026 | .104 | .113 | .071 | .022 | .027 |

**Table (6)** **Descriptive statistics.**

Descriptive statistics of the independent variables of regression (11) are shown. The variables are constructed per stock-hour. The top panel shows Frag, defined as 1 - HHI based on trading volume, and the LSE liquidity share (the LSE Depth(10) divided by the consolidated Depth(10)). Next is the logarithm of the number of trades on entrants, and the Order Imbalance, defined as the logarithm of buy volume minus the logarithm of sell volume. Ln(Turnover) the logarithm of turnover summed over all venues, and Ln(Depth(10) Cons) is the logarithm of the consolidated Depth(10). Last is the realized volatility, defined as the sum of squared 5 minute returns measured hour-by-hour. The bottom panel shows the cancellation effect of a buy (sell) trade of one venue on the consolidated ask (bid) liquidity. The cancellation effect is the impact (summed over one second) of a 1 share buy-trade at one venue on market wide liquidity offered on the ask sides. These variables are the coefficients of regression (11), estimed per stock-hour using as dependent variable the consolidated Depth(10) on the ask or bid side. We only use estimates when the venue in question executes >50 trades and >£100.000, and winsorize at the 99% level. The reported variables are winsorized at the 1% level, and .

| | Mean | Stdev | Max |
|---|---|---|---|
| Frag | 0.59 | 0.09 | 0.78 |
| Liq Share LSE | 0.40 | 0.11 | 0.95 |
| Ln(entrant trades) | 5.63 | 0.98 | 8.38 |
| Order Imbalance | -0.01 | 0.43 | 4.28 |
| Ln(Turnover) | 14.99 | 1.38 | 19.04 |
| Ln(Depth(10) Cons) | 12.43 | 1.01 | 16.09 |
| Realized Vol | 0.02 | 0.12 | 0.96 |
| Cancellation effect | | | |
| Buy-Ask LSE | 1.44 | 0.71 | 3.70 |
| Buy-Ask Chi-X | 2.07 | 1.11 | 5.73 |
| Buy-Ask Bats | 1.96 | 1.74 | 7.19 |
| Buy-Ask Turq | 1.14 | 1.65 | 7.57 |
| Buy-Ask Nasdaq | 1.44 | 3.31 | 14.69 |
| Sell-Bid LSE | 1.53 | 0.77 | 4.18 |
| Sell-Bid Chi-X | 1.99 | 1.06 | 5.48 |
| Sell-Bid Bats | 1.91 | 1.68 | 7.20 |
| Sell-Bid Turq | 0.95 | 1.50 | 5.41 |
| Sell-Bid Nasdaq | 1.51 | 2.99 | 14.62 |
| Observations | 1890 | | |

**Table (7)** **Drivers of the cancellation effect.**

We show the variables that determine the cancellation effect of trades on consolidated liquidity, per trading venue. The cancellation effect is the impact (summed over one second) of a venues buy-trades on market wide liquidity offered on the ask side (top panel). The bottom panel shows the impact of sell-trades on liquidity on the bid side. This effect is estimated per stock-hour in regression (11), using as dependent variable the consolidated Depth(10) on the ask (bid) side. These coefficients are the dependent variables in the regressions below. We add the following independent variables. Frag is defined as 1 - HHI based on trading volume, and Liq Share LSE is the LSE liquidity share (the LSE Depth(10) divided by the consolidated Depth(10)). Ln(entrant trades) is the logarithm of the number of trades on entrant venues and Ln(Turnover) the logarithm of market wide turnover. Ln(Depth(10) Cons) is the logarithm of the consolidated Depth(10), and Realized Vol is the realized volatility defined as the sum of squared 5 minute returns measured per hour-by-hour. We add firm and day fixed effects. We apply robust Newey-West standard errors (HAC) with 9 lags, t-statistics are shown in parantheses.

| | LSE | Chi-X | Bats | LSE | Chi-X | Bats |
|---|---|---|---|---|---|---|
| | Buy trades on consolidated Ask liquidity | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Frag | 1.147*** | -1.087** | -0.788 | -0.520* | -3.358*** | -4.742*** |
| | (4.09) | (-2.39) | (-0.71) | (-1.74) | (-4.31) | (-3.59) |
| Liq Share LSE | -2.261*** | -3.659*** | -3.130*** | -1.775*** | -3.110*** | -2.275** |
| | (-9.05) | (-8.84) | (-3.36) | (-7.79) | (-7.21) | (-2.49) |
| Ln(entrant trades) | | | | 0.585*** | 0.927*** | 1.750*** |
| | | | | (7.10) | (4.71) | (4.97) |
| Order Imbalance | | | | -0.291*** | -0.0314 | -0.341* |
| | | | | (-7.00) | (-0.43) | (-1.85) |
| Ln(Turnover) | | | | -0.640*** | -0.774*** | -1.535*** |
| | | | | (-9.10) | (-4.21) | (-5.05) |
| Ln(Depth(10) cons) | | | | 0.423*** | 0.506*** | 0.408** |
| | | | | (10.93) | (6.62) | (2.20) |
| Realized Vol | | | | -0.288** | 0.0833 | -1.107 |
| | | | | (-2.33) | (0.42) | (-1.10) |
| Observations | 1,718 | 1,689 | 1,058 | 1,718 | 1,689 | 1,058 |
| R2 | 0.143 | 0.078 | 0.018 | 0.261 | 0.143 | 0.047 |
| | Sell trades on consolidated Bid liquidity | | | | | |
| | (7) | (8) | (9) | (10) | (11) | (12) |
| Frag | 1.175*** | -1.296*** | -0.981 | -0.891** | -3.028*** | -4.790*** |
| | (3.55) | (-2.86) | (-0.81) | (-2.37) | (-4.42) | (-3.03) |
| Liq Share LSE | -1.413*** | -3.911*** | -3.001*** | -0.602* | -3.437*** | -2.336** |
| | (-3.98) | (-10.33) | (-3.32) | (-1.77) | (-8.93) | (-2.46) |
| Ln(entrant trades) | | | | 0.572*** | 0.625*** | 1.654*** |
| | | | | (6.02) | (3.78) | (4.47) |
| Order Imbalance | | | | 0.256*** | 0.0803 | -0.170 |
| | | | | (4.04) | (1.01) | (-0.91) |
| Ln(Turnover) | | | | -0.651*** | -0.597*** | -1.461*** |
| | | | | (-7.73) | (-3.98) | (-4.72) |
| Ln(Depth(10) cons) | | | | 0.607*** | 0.467*** | 0.277 |
| | | | | (11.12) | (6.89) | (1.54) |
| Realized Vol | | | | -0.185 | -0.374** | 0.287 |
| | | | | (-0.92) | (-2.12) | (0.17) |
| Observations | 1,718 | 1,689 | 1,057 | 1,718 | 1,689 | 1,057 |
| R2 | 0.067 | 0.090 | 0.017 | 0.211 | 0.139 | 0.043 |

# CDS Auctions *

Mikhail Chernov,[†] Alexander S. Gorbenko,[‡] and Igor Makarov[§]

November 15, 2011

**Abstract** We analyze credit default swap settlement auctions theoretically and evaluate them empirically. In our theoretical analysis we show that the current auction design may not give rise to the fair bond price, and we suggest modifications to minimize this mispricing. In our empirical study, we find that auctions undervalue bonds by an average of 10% on the auction day, and link the undervaluation to the number of bonds exchanged. We also find that underlying bond prices follow a V pattern around the day of the auction: in the preceding 10 days prices decrease by 30% on average, and in the subsequent 10 days they revert to their pre-auction levels.

**JEL Classification Codes:** G10, G13, D44

**Keywords:** credit default swaps, auctions, settlement, open interest

---

*We thank Bruno Biais, Christopher Hennessy, Dwight Jaffee, Francis Longstaff, Oleg Rytchkov, Dimitri Vayanos, Nancy Wallace, Bilge Yilmaz and seminar participants at LBS, LSE, Nottingham, Ohio State, Temple, Toulouse, and UC Berkeley. We also thank Anton Lines for copy editing.

[†]London School of Economics, and CEPR; m.chernov@lse.ac.uk.

[‡]London Business School; agorbenko@london.edu.

[§]London Business School; imakarov@london.edu.

# Introduction

Credit Default Swaps (CDS) have been one of the most significant financial innovations in the last 20 years. They have become very popular among investment and commercial banks, insurance companies, pension fund managers and many other economic agents. As a result, the market has experienced enormous growth. According to the Bank of International Settlements (BIS), the notional amount of single-name CDS contracts grew from \$5.1 trillion in December 2004 to \$33.4 trillion in June 2008, and was still \$18.4 trillion in June 2010 following a decline in the aftermath of the credit crisis.

The recent crisis put CDS in the spotlight, with policymakers now assigning them a central role in many reforms. The success of these reforms depends on the efficient functioning of the CDS market and on a thorough understanding of how it operates. Recognizing this, much research has been dedicated to the valuation of CDS contracts, econometric analysis of CDS premia, violations of the law of one price in the context of basis trades, search frictions, counterparty risk, private information, and moral hazard problems associated with holding both bonds issued by a particular entity and CDS protection on this entity.[1]

In this paper we focus on another aspect of CDS. We study how the payoff of a CDS contract is determined when a credit event occurs. Our theoretical analysis of the unusual auction-based procedure reveals that this mechanism may lead to deviations from fundamental value. The mispricing is attributable, in large part, to strategic bidding on the part of investors holding CDS. Empirically, we find that CDS auctions undervalue the underlying securities, by 10% on average. Because the magnitude of this mispricing is economically large, our findings may have implications for how CDS are valued, used and analyzed.

In a nutshell, a CDS is a contract that protects a buyer against the loss of a bond's principal in the case of a credit event (e.g., default, liquidation, debt restructuring, etc.). Initially, CDS were settled physically with the cheapest-to-deliver option. Under such settlement, the protection buyer was required to deliver any bond issued

[1] This work includes, but is not limited to, Acharya and Johnson (2007), Arora, Gandhi, and Longstaff (2009), Bolton and Oehmke (2011), Duffie (1999), Duffie and Zhu (2011), Garleanu and Pedersen (2011), Longstaff, Mithal, and Neis (2005), Pan and Singleton (2008), and Parlour and Winton (2010).

by the reference entity to the protection seller in exchange for the bond's par value. But as a result of the rapid development of the CDS market, the notional amount of outstanding CDS contracts came to exceed the notional amount of deliverable bonds many times over. This made physical settlement impractical and led the industry to develop a cash settlement mechanism. This mechanism is the object of our study.

While many derivatives are settled in cash, the settlement of CDS in this way is challenging for two reasons. First, the underlying bond market is opaque and illiquid, which makes establishing a benchmark bond price difficult. Second, parties with both CDS and bond positions face recovery basis risk if their positions are not closed simultaneously.[2] The presence of this risk renders it necessary that the settlement procedure include an option to replicate an outcome of the physical settlement.

In response to these challenges, the industry has developed a novel two-stage auction. In the first stage of the auction, parties that wish to replicate the outcome of the physical settlement submit their requests for physical delivery via dealers. These requests for physical delivery are aggregated into the net open interest ($NOI$). Dealers also submit bid and offer prices with a commitment to transact in a predetermined minimal amount at the quoted prices. These quotations are used to construct the initial market midpoint price ($IMM$). The $IMM$ is used to derive a limit on the final auction price, which is imposed to avoid potential price manipulation. The limit is referred to as the price cap. The $NOI$ and the $IMM$ are announced to all participants.

In the second stage a uniform divisible good auction is implemented, in which the net open interest is cleared. Each participant may submit limit bids that are combined with the bids of the dealers from the first stage. The bid that clears the net open interest is declared to be the final auction price, which is then used to settle the CDS contracts in cash.

We analyze the auction outcomes from both theoretical and empirical perspectives.

[2]Recovery basis risk can be illustrated as follows. Imagine a party that wishes to hedge a long position in a bond by buying a CDS with the same notional amount. The final physically-settled position is known in advance: the protection buyer delivers a bond in exchange for a predetermined cash payment equal to par value. However, the cash-settled position is uncertain before the auction: the protection buyer keeps the bond, pays the auction-determined bond value (unknown at the outset) to the protection seller, and receives par value in exchange. The difference between the market value of the bond held by the protection buyer and the auction-determined value is the recovery basis.

To study price formation, we follow Wilson (1979) and Back and Zender (1993). We formalize the auction using an idealized setup in which all auction participants are risk-neutral and have identical expected valuations of the bond, $v$. This case is not only tractable, but also provides a useful benchmark against which to test whether CDS auctions lead to the fair-value price. While Wilson (1979) shows that a standard uniform divisible good auction can result in underpricing specifically, we demonstrate that the current auction design can yield a final price either above or below $v$.

Our conclusion differs because participants of CDS auctions can have prior positions in derivatives on the asset being auctioned. If a participant chooses to settle her entire CDS position physically, her final payoff is not affected by the auction outcome. However, in the case of cash settlement, buyers of protection benefit if the auction price is set below fair value, while sellers benefit if it is set above. Therefore, an auction outcome depends on the size of the net CDS positions; that is, positions that remain after participants submit their physical settlement requests.

To be specific, consider the case of positive $NOI$: a second-stage auction in which the agents buy bonds. When the net CDS positions of protection sellers are less than the $NOI$, the Wilson (1979) argument still holds. Underpricing occurs if the auction participants choose not to bid aggressively. The current auction rule is such that bids above the final price are guaranteed to be fully filled, so participants are not sufficiently rewarded for raising their bids. On the other hand, when the net CDS positions of sellers are larger than the $NOI$, bidding above the fair value and realizing a loss from buying $NOI$ units of bonds is counterbalanced by a reduction in the net payoff of the existing CDS contracts. In the absence of a cap, the auction price would be at least $v$.

Our theory delivers a rich set of testable predictions. Full implementation of such tests requires data on individual CDS positions and bids, which are not available. Nonetheless, we are able to analyse some aspects of the auction data and find evidence that is consistent with our theoretical predictions. We use TRACE bond data to construct the reference bond price. Using the reference bond price on the day before the auction as a proxy for $v$, we find that the auction price is set at the price cap whenever there is overpricing. Furthermore, the extent of overpricing does not exceed the spread between the price cap and $IMM$. When the final auction price is uncapped and the $NOI$ is positive (a typical situation), the bonds are undervalued and the degree of undervaluation increases with the $NOI$. In addition, underlying bond prices follow a V pattern around the auction day. In the 10 days before the auction, prices decrease by 30% on average. They reach their lowest level on the day of the auction (average underpricing of 10%), before reverting to their pre-auction levels over the next 10 days. This evidence suggests that our conclusions are robust to the choice of the reference bond price.

Our findings prompt us to consider ways to mitigate the observed mispricing. In a standard setting, in which agents have no prior positions in derivative contracts written on the asset being auctioned, Kremer and Nyborg (2004) suggest a likely source of underpricing equilibria. They show that a simple change of allocation rule from pro-rata on the margin to pro-rata destroys all underpricing equilibria. We show that the same change of allocation rule would be beneficial in our setting. In addition, we suggest that imposing an auction price cap conditional on the outcome of the first stage could further reduce mispricing in equilibrium outcomes.

To our knowledge there are four other papers that examine CDS auctions, two of which were carried out independently and contemporaneously with our work. Three of the papers analyse the auctions empirically. Helwege, Maurer, Sarkar, and Wang (2009) find no evidence of mispricing in an early sample of 10 auctions, of which only four used the current auction format. Coudert and Gex (2010) study a somewhat different sample of auctions, using Bloomberg data for reference bond prices. They document a large gap between a bond's price on the auction date and the final auction price. However, they do not link the gap to the net open interest, nor do they provide any theoretical explanations for their findings. Gupta and Sundaram (2011) also document a V pattern in bond prices around the auction day. Under a simplifying assumption that bidders in the second stage of the auction have zero CDS positions, they find that a discriminatory auction format could reduce the mispricing. Finally, Du and Zhu (2011) examine the outcome types that are possible in CDS auctions. Their paper considers a special case of our model, in which they implicitly assume that all market participants can buy and short-sell bonds of distressed companies at the fair value $v$ without any restrictions. This setup implies that only overpricing equilibria can exist. Further, they treat physical settlement requests as given. We show that this setup results in fair pricing if agents choose physical settlement optimally. We allow for a more realistic setup, where there are constraints on short selling, and where some participants cannot hold distressed debt. We show that there can be substantial underpricing in this case.

The remainder of the paper is organized as follows. Section 1 describes the CDS auction methodology as it is currently employed. Section 2 describes the auction model. Section 3 provides the main theoretical analysis. Section 4 relates the predictions of the theoretical model to empirical data from CDS auctions. Section 5 discusses modifications that could potentially improve the efficiency of the auction. Section 6 concludes. The appendix contains proofs that are not provided in the main text.

# 1 The Auction Format

This discussion is based on a reading of the auction protocols available from the ISDA website. Initially, CDS auctions were designed for cash settlement of contracts on credit indexes. The first auction that allowed single-name CDS to be settled in cash was the Dura auction, conducted on November 28, 2006. The auction design used in this case, and for all subsequent credit events, consists of two stages.

In the first stage, participants in the auction submit their requests for physical settlement. Each request for physical settlement is an order to buy or sell bonds at the auction price. To the best of the relevant party's knowledge, the order must be in the same direction as – and not in excess of – the party's market position, which allows the participants to replicate traditional physical settlement of the contracts. For example, if a party is long one unit of protection and submits a request to physically deliver one bond, the resulting cash flow is 100 and is identical to that of physical settlement.

In addition, a designated group of agents (dealers) makes a two-way market in the defaulted assets by submitting bids and offers with a predefined maximum spread and associated quotation size. The spread and quotation sizes are subject to specification prior to each auction and may vary depending on the liquidity of the defaulted assets.[3]

The first stage inputs are then used to calculate the net open interest ($NOI$) and an 'initial market midpoint' ($IMM$), which are carried through to the second part

[3] The most common value of the spread is 2% of par. Quotation sizes range from \$2 to \$10 million; \$2 million is the most common amount.

of the auction. The $NOI$ is computed as the difference of the physical-settlement buy and sell requests. The $IMM$ is set by discarding crossing/touching bids and offers, taking the 'best half' of each, and calculating the average. The best halves would be, respectively, the highest bids and the lowest offers. If a dealer's quotation is crossed and is on the wrong side of the $IMM$, she must make a payment, called an adjustment amount, to the ISDA. That is, she pays the adjustment amount if her bid is higher than the $IMM$ and the $NOI$ is to sell, or if an offer is lower than the $IMM$ and the $NOI$ is to buy. The adjustment amount itself is a product of the quotation amount and the difference between the quotation and the $IMM$.

As an example, consider the Nortel Limited auction of February 10, 2009. Table 1 lists the market quotes submitted by participating dealers. Once these quotes have been received, the bids are sorted in descending order and the offers in ascending order. The highest bid is then matched with the lowest offer, the second highest bid with the second lowest offer, and so on. Figure 1 displays the quotes from Table 1 which are organized in this way. For example, the Citibank bid of 10.5 and the Barclays offer of 6.0 create a tradeable market.

The $IMM$ is computed from the non-tradeable quotes. First, the 'best half' of the non-tradeable quotes is selected (i.e., the first five pairs). Second, the $IMM$ is computed as an average of bid and offer quotes in the best half, rounded to the nearest one-eighth of a percentage point. In our example there are nine pairs of such quotes. The relevant bids are: three times 7.0 and two times 6.5. The relevant offers are: two times 8.0; two times 8.5; and 9. The average is 7.6 and the rounded average is 7.625.

Given the established $IMM$ and the direction of open interest, dealers whose quotes have resulted in tradeable markets pay the adjustment amount to the ISDA. In the case of Nortel, the open interest was to sell. Thus, dealers whose bids crossed the markets were required to pay an amount equal to (Bid-$IMM$) times the quotation amount, which was \$2 MM. Citigroup had to pay $(10.5 - 7.625)/100 \times \$2MM =$ \$57500 and Banc of America Securities had to pay $(9.5 - 7.625)/100 \times \$2MM =$ \$37500.

Finally, the direction of open interest determines the cap on the final price, where the price itself is set in the second part of the auction. In the Nortel example the open interest was to sell, which meant the final price could exceed the $IMM$ by a maximum of 1.0. Thus the price cap was 8.625, as depicted in Figure 1.

After the publication of the $IMM$, the $NOI$, and the adjustment amounts, the second stage of the auction begins. If the $NOI$ is zero, the final price is set equal to the $IMM$. If the $NOI$ is non-zero, dealers may submit corresponding limit orders on behalf of their customers (including those without CDS positions) – and for their own account – to offset the $NOI$. Agents submit 'buy' limit orders if the $NOI$ is greater than zero and 'sell' limit orders if it is less than zero. In practice, it is unlikely that all agents involved in the first stage will participate in the second stage as well. Participants in the CDS market are diverse in terms of their investment objectives and institutional constraints. For example, many mutual and pension funds may not be allowed to hold any of the defaulted bonds.

Upon submission of the limit orders, if the $NOI$ is to buy, the auction administrators match the open interest against the market bids from the first stage of the auction, and against the limit bids from the second stage of the auction. They start with the highest bid, proceeding through the second highest bid, third highest bid, and so on, until either the entire net open interest or all of the bids have been matched. If the $NOI$ is cleared, the final price is set equal to the lowest bid corresponding to the last matched limit order. However, if this bid exceeds the $IMM$ by more than the cap amount (typically half of the bid-offer spread), the final price is simply set equal to the $IMM$ plus the cap amount. If all bids are matched before the $NOI$ clears, the final price will be zero and all bids will be filled on a pro-rata basis. The procedure is similar if the $NOI$ is to sell. If there are not enough offers to match the net open interest, the final price is set to par.

# 2 The Auction Model

The main question we wish to address in this paper is whether the current auction format may result in mispricing. Our approach is motivated by the classic work of Wilson (1979) and Back and Zender (1993) who show how this can happen in a standard divisible-good auction. As in Wilson (1979), we assume that all agents are risk-neutral and have identical expectations about the value of the bonds. This case is not only tractable, but also provides a useful benchmark from which to judge whether the auction leads to the fair-value price. This approach is popular in the auction literature because if equilibria that result in mispricing can be found in this admittedly basic setup, it is likely they will also be possible in more realistic scenarios.

The goal of this section is to formalize the auction process described in Section 1. There are two dates: $t = 0$ and $t = 1$. There is a set $\mathcal{N}$ of strategic players and the total number of agents is $|\mathcal{N}| = N$. A set of dealers $\mathcal{N}_d$ constitutes a subset of all players, $\mathcal{N}_d \subseteq \mathcal{N}$, $|\mathcal{N}_d| = N_d$. Each agent $i \in \mathcal{N}$ is endowed with $n_i \in \mathbb{R}$ units of CDS contracts and $b_i \in \mathbb{R}$ units of bonds. Agents with positive (negative) $n_i$ are called protection buyers (sellers). Because a CDS is a derivative contract, it is in zero net supply, $\sum_i n_i = 0$. One unit of bond pays $\tilde{v} \in [0, 100]$ at time $t = 1$. The auction takes place at time $t = 0$ and consists of two stages.

## 2.1 First Stage

In the first stage, the auction initial market midpoint ($IMM$) and the net open interest ($NOI$) are determined. Agent $i$ may submit a request to sell $y_i$ (or buy if $y_i < 0$) units of bonds at par (100). Each protection buyer, $n_i > 0$, is only allowed to submit a request to sell $y_i \in [0, n_i]$ units of bonds, while each protection seller, $n_i < 0$, may only submit a request to buy $y_i \in [n_i, 0]$ units of bonds. Given these requests, the $NOI$ is determined as follows:

$$NOI = \sum_{i=1}^{N} y_i. \tag{1}$$

In addition, all dealers from the set $\mathcal{N}_d$ are asked to provide a price quote $\pi_i$. Given $\pi_i$, dealer $i$ must stand ready to sell or buy $L$ units of bonds at bid and offer prices $\pi_i + s$ and $\pi_i - s$, $s > 0$. Quotes from dealers whose bids and offers cross are discarded. The $IMM$, denoted by $p^M$, is then set equal to the average of the remaining mid-quotations.

## 2.2 Second Stage

At this stage, a uniform divisible good auction is held. If $NOI = 0$ then $p^A = p^M$. If $NOI > 0$, participants bid to buy $NOI$ units of bonds. In this case, each agent $i$ may submit a left-continuous non-increasing demand schedule $x_i(p) : [0, p^M + s] \to \mathbb{R}_+ \cup 0$.

Let $X(p) = \sum_{i \in \mathcal{N}} x_i(p)$ be the total demand. The final auction price $p^A$ is the highest price at which the entire $NOI$ can be matched:

$$p^A = \max\{p | X(p) \geq NOI\}.$$

If $X(0) \leq NOI$, $p^A = 0$. Given $p^A$, the allocations $q_i(p^A)$ are determined according to the 'pro-rata at the margin' rule:

$$q_i(p^A) = x_i^+(p^A) + \frac{x_i(p^A) - x_i^+(p^A)}{X(p^A) - X^+(p^A)} \times (NOI - X^+(p^A)), \tag{2}$$

where $x_i^+(p^A) = \lim_{p \downarrow p^A} x_i(p)$ and $X^+(p) = \lim_{p \downarrow p^A} X(p)$ are the individual and total demands, respectively, above the auction clearing price.

If $NOI < 0$, participants offer to sell $|NOI|$ units of bonds. Each agent $i$ may then submit a right-continuous non-decreasing supply schedule $x_i(p) : [100, p^M - s] \rightarrow \mathbb{R}_- \cup 0$.

As before, the total supply is $X(p) = \sum_{i \in \mathcal{N}} x_i(p)$. And the final auction price $p^A$ is the lowest price at which the entire $NOI$ can be matched:

$$p^A = \min\{p | X(p) \leq NOI\}.$$

If $X(100) \geq NOI$, $p^A = 100$. Given $p^A$, the allocations $q_i(p^A)$ are given by:

$$q_i(p^A) = x_i^-(p^A) + \frac{x_i(p^A) - x_i^-(p^A)}{X(p^A) - X^-(p^A)} \times (NOI - X^-(p^A)),$$

where $x_i^-(p^A) = \lim_{p \uparrow p^A} x_i(p)$ and $X^-(p) = \lim_{p \uparrow p^A} X(p)$ are the individual and total supplies, respectively, below the auction clearing price.

## 2.3 Preferences

Two types of agents participate in the auction: dealers and common participants. In our setup, all agents are risk-neutral and have identical expected valuations of the bond payoff, $v$. The agents' objective is to maximize their wealth, $\Pi_i$, at date 1, where

$$\begin{aligned}\Pi_i &= \underbrace{(v - p^A)q_i}_{\text{auction-allocated bonds}} + \underbrace{(n_i - y_i)(100 - p^A)}_{\text{net CDS position}} \\ &+ \underbrace{100y_i}_{\text{physical settlement}} + \underbrace{v(b_i - y_i)}_{\text{remaining bonds}} \quad . \end{aligned} \tag{3}$$

and $q_i$ is the number of auction-allocated bonds.

Dealers differ from common participants in that they submit quotes ($\pi_i$) in the first stage, which are made public after the auction. Thus, due to regulatory and reputational concerns, dealers may be reluctant to quote prices that are very different from $v$ unless the auction results in a large gain. To model these concerns we assume that dealers' utility has an extra term $-\frac{\gamma}{2}(\pi_i - v)^2$, $\gamma \geq 0$.

## 2.4 Trading Constraints

So far we have assumed a frictionless world in which every agent can buy and sell bonds freely. This is a very strong assumption which is violated in practice. Therefore, we extend our setup to allow market imperfections. Specifically, we place importance on the following two frictions.

First, some auction participants, such as pension funds or insurance companies, may not be allowed to hold bonds of defaulted companies. To model this, we introduce Assumption 1.

**Assumption 1** *Only a subset $\mathcal{N}_+ \subseteq \mathcal{N}$, $\mathcal{N}_+ \neq \emptyset$ of the set of agents can hold a positive amount of bonds after the auction.*

Second, because bonds are traded in OTC markets, short-selling a bond is generally difficult. To model this, we introduce Assumption 2.

**Assumption 2** *Each agent $i$ can sell only $b_i$ units of bonds.*

In what follows, we solve for the auction outcomes both in the frictionless world and under Assumptions 1 and 2.

# 3 Analysis

We now turn to a formal analysis of the auction described in the preceding section. We solve for the auction outcomes using backward induction. We start by solving for the equilibrium outcome in the second stage of the auction, for a given $IMM$ and $NOI$. We then find optimal dealer quotations $\pi_i$ and optimal physical settlement requests in the first stage given the equilibrium outcomes of the second stage.

## 3.1 Second Stage

As previously noted, stage two consists of a uniform divisible good auction with the goal of clearing the net open interest generated in the first stage. A novel feature of our analysis is that we study auctions where participants have prior positions in derivative contracts written on the asset being auctioned. We show that equilibrium outcomes in this case can be very different from those realized in 'standard' auctions (that is, auctions in which $n_i = 0$ for all $i$).

We first consider the case in which all CDS positions are common knowledge. (This assumption is relaxed later.) If this is the case, each agent $i$ takes the following as given: the $NOI$, a set of all CDS positions $n_i$, a set of physical settlement requests $y_i$, $i \in \mathcal{N}$, and the demand of other agents $x_{-i}(p)$. Therefore, from equation (3), each agent's demand schedule $x_i(p)$ solves the following optimization problem:

$$\max_{x_i(p)} \; (v - p(x_i(p), x_{-i}(p)))\, q_i(x_i(p), x_{-i}(p)) + (n_i - y_i)\,(100 - p(x_i(p), x_{-i}(p)))\,. \quad (4)$$

The first term in this expression represents the payoff realized by participating in the auction, while the second term accounts for the payoff from the remaining CDS positions, $n_i - y_i$, which are settled in cash on the basis of the auction results.

To develop intuition about the forthcoming theoretical results, consider the bidding incentives of the auction participants. The objective function (4) implies that, holding the payoff from the auction constant, an agent who has a short (long) remaining CDS position wishes the final price to be as high (low) as possible. However, agents with opposing CDS positions do not have the same capacity to affect the auction price. The auction design restricts participants to submit one-sided limit orders depending on the sign of the $NOI$. If the $NOI > 0$, only buy limit orders are allowed, and therefore agents with short CDS positions are capable of bidding up the price. By contrast, all that an agent with a long CDS positions can do to promote her desired outcome is not to bid at all. The situation is reversed when the $NOI < 0$.

Continuing with the case of the $NOI > 0$, consider an example of one agent with a short CDS position. She has an incentive to bid the price as high as possible if the $NOI$ is lower than the notional amount of her CDS contracts (provided she is allowed to hold defaulted bonds). This is because the cost of purchasing the bonds at a high auction price is offset by the benefit of cash-settling her CDSs at the same high price. In contrast, if the $NOI$ is larger than the notional amount of her CDS position, she would not want to bid more than the fair value of the bond, $v$. This is because the cost of purchasing bonds at a price above $v$ is not offset by the benefit of cash-settling CDSs. In what follows, we show that this intuition can be generalized to multiple agents, as a long as we consider the size of their aggregate net CDS positions relative to the $NOI$.

**Proposition 1** *Suppose that $NOI > 0$ and Assumption 1 holds.*

1. *If*

$$\sum_{i \in \mathcal{N}_+ : n_i < 0} |n_i - y_i| \geq NOI, \tag{5}$$

*and $p^M + s > v$, then in any equilibrium the final auction price $p^A \in [v, p^M + s]$. Furthermore, there always exists an equilibrium in which the final price is equal to the cap: $p^A = p^M + s$. If $p^M + s < v$ then the final price is always equal to the cap: $p^A = p^M + s$.*

2. *If*

$$\sum_{i \in \mathcal{N}_+ : n_i < 0} |n_i - y_i| < NOI, \tag{6}$$

*then only equilibria with $p^A \leq \min\{p^M + s, v\}$ exist.*

**Proof.** *Part 1.* Intuitively, if condition (5) holds, there is a subset of agents for whom a joint loss incurred by acquiring a number of bonds equal to the $NOI$, at a price above $v$, is dominated by a joint gain from paying less on a larger number of short CDS contracts that remain after the physical settlement. As a result, these agents bid aggressively and can push the auction price above $v$ unless it is constrained by the $IMM$. In the latter case, $p^A = p^M + s$.

Formally, suppose that $p^M + s > v$, $p^A < v$ and condition (5) holds. We show that this cannot be true in equilibrium. Let the equilibrium allocation of bonds to agent $i$ be $q_i$. Consider a change in the demand schedule of player $i$ from $x_i$ to $x_i'$ that leads to the auction price $p \in [p^A, v]$. Denote the new bond allocation of agent $i$ by $q_i'$. Since demand schedules are non-decreasing, $q_i' \geq q_i$. Agent $i$'s change in profit is thus

$$\delta_i = \left[(v - p^A)q_i - p^A(n_i - y_i)\right] - \left[(v - p)q_i' - p(n_i - y_i)\right] =$$

$$= (p - p^A)(n_i - y_i + q_i) - (v - p)(q_i' - q_i) \leq (p - p^A)(n_i - y_i + q_i). \tag{7}$$

Equilibrium conditions require that $\delta_i \geq 0$ for all $i$. Summing over all $i$ such that $n_i < 0$, it must be that

$$0 \leq \sum_{i:n_i<0} \delta_i \leq (v - p^A) \sum_{i:n_i<0} (n_i - y_i + q_i)\,.$$

Because all $q_i \geq 0$,

$$\sum_{i:n_i<0} (n_i - y_i + q_i) \leq \sum_{i:n_i<0} (n_i - y_i) + NOI \leq 0, \tag{8}$$

where we use (5). Thus in any equilibrium with $p^A < v$, it must be that $\delta_i = 0$ for all $i$ with $n_i < 0$. (7) and (8) then imply that for any deviation $x_i'$ that leads to $p \in [p^A, v]$, it must be that $q_i' = q_i$. Since this is true for any $p \in [p^A, v]$ the initial total demand $X(p)$ must be constant over $[p^A, v]$, and therefore $p^A = v$. Thus we arrive at a contradiction.

Next, consider the following set of equilibrium strategies:

$$x_i(p): \quad \begin{cases} x_i = NOI \times (n_i - y_i)/(\sum_{j:n_j<0}(n_j - y_j)) & \text{if} \quad v < p \leq p^M + s, \\ x_i = NOI & \text{if} \quad p \leq v, \end{cases}$$

for agents with net negative CDS positions after physical settlement request submission, and $x_i(p) \equiv 0$ for agents with positive CDS positions. It is not difficult to see that it supports $p^A = p^M + s$.

*Part 2.* Finally, suppose that condition (5) does not hold and there exists an equilibrium with $p^A > v$. Then there also exists an $i$ such that agent $i$'s equilibrium second stage allocation $q_i > |n_i - y_i|$. Consider a variation of this agent's demand schedule, in which she submits zero demand at $p^A > v$ and demand equal to the $NOI$ at $p^A = v$. Given this variation, the new auction price will be higher than or equal to $v$. Thus her profit increases by at least $(p^A - v)(q_i + n_i - y_i) > 0$, so $p^A > v$ cannot be an equilibrium outcome. *QED.*

The next lemma shows that when all agents are allowed to hold bonds after the auction (that is, Assumption 1 does not hold), condition (5) always holds. As a result, the final price is always at least $v$ unless it is capped.

**Lemma 1** *If $\mathcal{N}_+ = \mathcal{N}$ then condition (5) holds.*

**Proof.**

$$\sum_{i:n_i<0}(n_i - y_i) + NOI = \sum_{i:n_i<0}(n_i - y_i) + \sum_i y_i = \sum_{i:n_i<0} n_i + \sum_{i:n_i>0} y_i \leq \sum_{i:n_i<0} n_i + \sum_{i:n_i>0} n_i = 0.$$

*QED.*

Proceeding to the case where $NOI < 0$, we obtain the following result.

**Proposition 2** *Suppose that $NOI < 0$ and there are no short-selling constraints. If $p^M - s < v$, then in any equilibrium, $p^A \in [p^M - s, v]$. If $p^M - s > v$ then $p^A = p^M - s$.*

This result is a natural counterpart of Part 1 of Proposition 1, and the proof follows the same logic. Without Assumption 2, we do not have a counterpart to Part 2 because all agents can participate in the second stage. With short-selling constraints, equilibria in which the bond is overpriced and the price is not capped can also exist. The conditions allowing for these equilibria are more stringent than those in Part 2 of Proposition 1 because the short-selling constraints are assumed to hold at the individual level. Proposition 3 characterizes these conditions.

**Proposition 3** *Suppose that $NOI < 0$ and Assumption 2 is imposed.*

1. *If for all $i$ such that $n_i > 0$,*

$$b_i \geq -NOI \times \frac{n_i - y_i}{\sum_{j:n_j>0}(n_j - y_j)} \tag{9}$$

*then there exists an equilibrium in which* $p^A = p^M - s$.

2. *If*

$$\sum_{i:n_i>0} b_i < -NOI, \tag{10}$$

*then only equilibria with* $p^A \geq \max\{p^M - s, v\}$ *exist.*

**Proof.** Part 2 is straightforward: under the assumption of short-sale constraints and (10), $NOI$ units of bonds cannot be sold solely by agents with long CDS positions. Agents with non-positive CDS positions, however, will not sell bonds at a price below $v$. Thus we only need to prove Part 1. To do this, consider the following set of strategies (assuming that $p^M - s < v$):

$$x_i(p): \quad \begin{cases} x_i = NOI \times (n_i - y_i)/(\sum_{j:n_j<0}(n_j - y_j)) & \text{if} \quad p^M + s \leq p < v, \\ x_i = -b_i & \text{if} \quad p \geq v, \end{cases}$$

for agents with net positive CDS positions after physical request submission, and

$$x_i(p): \quad \begin{cases} x_i = 0 & \text{if} \quad p^M + s \leq p < v, \\ x_i = -b_i & \text{if} \quad p \geq v. \end{cases}$$

for agents with positive CDS positions. It is not difficult to see that this set of strategies constitutes an equilibrium and supports $p^A = p^M - s$. *QED*.

## 3.2 First Stage

To solve for a full game equilibrium, the last step is to determine physical settlement requests $y_i$, the $NOI$ and the $IMM$, given the outcomes in the second stage of the auction. The $IMM$ does not contain any information in our setup, which precludes uncertainty. Nevertheless, it can still play an important role because it provides a cap on the final price. We start our analysis by assuming that the second-stage auction does not have a cap. After we solve for (and develop intuition about) the optimal physical settlement requests and the $NOI$, we discuss the effect of the cap.

#### 3.2.1 Second-Stage Auction Without a Cap

First, we show that in a frictionless world, only equilibria with the auction price different from $v$ exist. Furthermore, in all of these equilibria agents obtain the same utility.

**Proposition 4** *Suppose that there are no trading frictions, i.e. Assumptions 1 and 2 are not imposed. Then any equilibrium will be one of three types: (i) $p^A \in (v, 100]$ and $NOI \geq 0$, where agents with initial long CDS positions choose physical delivery and receive zero bond allocation in the auction; (ii) $p^A \in [0, v)$ and $NOI \leq 0$, where agents with initial short CDS positions choose physical delivery and do not sell bonds in the auction; and (iii) $p = v$. In each of the three cases, all agents attain the same utility.*

**Proof.** Suppose that $p^A \in (v, 100]$. Lemma 1, Part 2 of Proposition 1, and Proposition 2 imply that this can be the case only if $NOI \geq 0$. Clearly, only agents with negative remaining CDS positions after the first stage of the auction will be willing to buy bonds at a price above $v$. Agents with initial long CDS positions receive zero bond allocation. From (3) each of their utility functions will be

$$\Pi_i = n_i(100 - v) + (y_i - n_i)(p^A - v) + b_i v. \tag{11}$$

If $p^A > v$, utility (11) is maximized if $y_i$ is as large as possible. Therefore, $y_i = n_i$ and $\Pi_i = n_i(100 - v)$ for $n_i > 0$. Thus in any such equilibrium agents with initial long CDS positions choose physical delivery, receive zero bond allocation, and attain the same utility. The $NOI$ is

$$NOI = \sum_i y_i = \sum_{i:n_i>0} n_i + \sum_{i:n_i<0} y_i = -\sum_{i:n_i<0} (n_i - y_i) \geq 0. \tag{12}$$

In other words, the $NOI$ is equal to the sum of outstanding CDS positions (after the first stage) held by agents with initial short CDS positions. As a result, any gain from buying at a price above $v$ (due to the existing CDS positions) is exactly offset by the loss incurred by buying bonds at this price. From (3), the utility of agents with initial short CDS positions is given by

$$\Pi_i = n_i(100 - v) + (y_i - n_i - q_i)(p^A - v) + b_i v. \tag{13}$$

Because every agent can always guarantee utility $\Pi_i = n_i(100-v)$ by choosing physical delivery, $q_i$ cannot be higher than $-(n_i - y_i)$. In addition, (12) implies that $q_i$ cannot be lower than $-(n_i - y_i)$. Therefore, $q_i = -(n_i - y_i)$ and $\Pi_i = n_i(100 - v)$ for each $i : n_i < 0$. The proof when $p^A \in [0, v)$ is similar. $QED$.

Proposition 4 shows that in a frictionless world, all mispricing equilibria are unidirectional – that is, there is no under- (over-) pricing if the $NOI$ is positive (negative). Furthermore, agents can undo any loss of utility resulting from auction mispricing by optimally choosing between cash and physical settlement of their positions.

We now turn to more realistic setups that include trading frictions. Our analysis in section 3.1 shows that there can be a continuum of equilibria in the second stage, which makes solving for every equilibrium in a two-stage auction a daunting problem. Instead of characterizing all of the equilibria, we show that in the presence of trading frictions, as outlined in Section 2.4, there exists a subset of equilibria of the two-stage game that results in bond mispricing. This result answers, in the affirmative, our main question as to whether mispricing is possible in the auction. Proposition 5 characterizes sufficient conditions for underpricing to occur.

**Proposition 5** *Suppose that Assumption 1 holds,*

$$(i) \qquad \sum_{i:n_i>0} n_i + \sum_{i\in\mathcal{N}_+:n_i<0} n_i > 0, \tag{14}$$

*and for any* $n_i > 0$*,*

$$(ii) \quad n_i > \frac{\sum_{j:n_j>0} n_j + \sum_{j\in\mathcal{N}_+:n_j<0} n_j}{K+1}, \tag{15}$$

*where $K$ is a total number of agents with initial long CDS positions. Then there exist a multitude of subgame perfect underpricing equilibria for the two-stage auction, in which (i) $NOI > 0$,*

$$(ii) \quad \frac{\partial p^A(NOI)}{\partial NOI} < 0, \quad \textit{and (iii)} \quad 0 \le v - p^A(NOI) \le NOI \times \left|\frac{\partial p^A(NOI)}{\partial NOI}\right|. \tag{16}$$

*In particular, there exists a subset of full-game equilibria where for any $NOI$ that can be realized in the first stage, the second stage leads to a final price $p^A$ which is a linear function of the $NOI$:*

$$p^A = v - \delta \times NOI \geq 0, \quad \delta > 0. \tag{17}$$

**Proof.** See Appendix.

We give a formal proof by construction in the Appendix and merely describe the intuition here. In the proof, we show that if optimal physical settlement requests satisfy condition (6) instead then there exist second-stage equilibria with $p^A \leq v$, where agents play the following strategies:

$$x_i(p) = \max\{c(v-p)^\lambda - n_i + y_i, 0\}, \tag{18}$$

$c$ and $\lambda$ are specified in the Appendix. A similar set of strategies is used in Back and Zender (1993) to construct equilibria in a standard auction without CDS positions. There could also be other classes of equilibrium second-stage strategies. We use strategies (18) mainly because they lead to a closed form solution. The main challenge in the rest of the proof is to solve jointly for equilibrium physical settlement requests and the second-stage equilibrium price.

A closer inspection of (3) reveals that if the final auction price is lower than $v$ and is not affected by agents' physical requests (i.e., participants always choose to play same-price equilibria as long as the $NOI$ is high enough to ensure second-stage underpricing), agents with long (short) CDS positions only have an incentive to choose full cash (physical) settlement in the first stage. This first-stage play implies that the $NOI$ must be negative. As a result, second-stage underpricing equilibria in which $\partial p^A / \partial NOI = 0$ cannot be equilibria of the full game. However, if the strategies played in the second stage are such that the final auction price is a negative function of the $NOI$, then the incentives of agents with long CDS positions become non-trivial. Submission by such agents of a partial physical settlement request could lead to a larger $NOI$ and in turn to a lower final auction price, increasing the payoff they receive from their partial cash settlement. The larger the initial positions of agents with long CDS positions, the stronger the incentives to lower the price via partial physical settlement. Condition (15) guarantees that the long positions of agents are sufficiently large to ensure that they choose physical settlement of enough positions to render the resulting $NOI$ positive.

The subset of equilibria characterized in Proposition 5 is the simplest and serves as an example of underpricing. There may be other equilibria resulting in underpricing that we have not found. While Lemma 1 implies that condition (14) is necessary for an underpricing equilibrium to exist, condition (15) can be relaxed at the expense of a more complicated proof.

Finally, notice that if there are short-sale constraints (that is, Assumption 2 is imposed), the logic of Proposition 4 may also break down. In this case, agents with initial long CDS positions are able to choose only $b_i$ units of bonds for physical settlement. If for at least one such agent $n_i > b_i$, and sufficiently many agents with remaining short CDS positions participate, the agents as a group could become strictly better off by pushing the price above $v$. Proposition 6 characterizes the effect of short-sale constraints on auction outcomes.

**Proposition 6** *Suppose that only Assumption 2 is imposed and there exists an $i$ such that*

$$n_i > b_i > 0, \tag{19}$$

*and*

$$\sum_{j:n_j<0} |n_j| > \sum_{j:n_j>0} \max\{b_j, 0\}. \tag{20}$$

*Then for the two-stage auction there exists a subgame perfect overpricing equilibrium in which $NOI = \sum_{j:n_j>0} \max\{b_j, 0\} > 0$, $p^A = 100$, and agents with initial short CDS positions attain strictly greater utility than when $p^A = v$.*

**Proof.** The proof is by construction. As in Proposition 4, if $p^A = 100$, agents who are initially long CDS contracts will choose physical delivery, and only agents with negative remaining CDS positions after the first stage will be willing to buy bonds in the auction. Proposition 1 Part 1 shows that for any $NOI > 0$, if condition (5) holds (which turns out to be the case in the constructed equilibrium), then $p^A = 100$ is an equilibrium of the second stage if agents play the following strategies:

$$x_i(p): \quad \begin{cases} x_i = NOI \times (n_i - y_i)/(\sum_{j:n_j<0}(n_j - y_j)) & \text{if} \quad v < p \leq 100, \\ x_i = NOI & \text{if} \quad p \leq v. \end{cases}$$

for agents with net negative CDS positions after physical request submission, and $x_i(p) \equiv 0$ for other agents. The profit earned by agent $i$ with $n_i < 0$ is therefore

$$\Pi_i = \left( y_i - NOI \frac{n_i - y_i}{\sum_{j:n_j<0}(n_j - y_j)} \right)(100 - v) + b_i v. \tag{21}$$

Taking the F.O.C. at $y_i = 0$, one can verify that it is optimal for agents with initial short CDS positions to choose cash settlement. Thus $NOI = \sum_{j:n_j>0} \max\{b_j, 0\}$ and the profit accruing to any agent $i$ with an initial short CDS position $n_i < 0$ is

$$\Pi_i = -(100 - v)n_i \times \frac{\sum_{j:n_j>0} \max\{b_j, 0\}}{\sum_{j:n_j<0} n_j} + b_i v > (100 - v)n_i + b_i v,$$

where the expression on the right hand side is the agent's utility if $p^A$ is equal to $v$. *QED*.

Propositions 5 and 6 show that there can be either underpricing or overpricing equilibria in the two-stage game with $NOI > 0$, if there are trading frictions. A similar set of results can be obtained for $NOI < 0$.

### 3.2.2 Second-Stage Auction with a Cap

We now discuss the implications of the second-stage price cap, which imposes an upper bound of $p^M + s$ on the final price. In the presence of the cap, mispricing in the auction depends on the bidding behavior of dealers in the first stage. The next proposition shows that the $IMM$ is equal to $v$ when there are no trading frictions.

**Proposition 7** *Suppose that there are no trading frictions (Assumptions 1 and 2 are not imposed) and $\gamma > 0$. Then $IMM = v$. Therefore, of the overpricing equilibria described in Proposition 4 there can exist only equilibria with $|p^A - v| \leq s$.*

**Proof.** Proposition 4 shows that in all possible equilibria, common participants attain the same utility. Because dealers have regulatory and reputational concerns, captured by the extra term $-\gamma(\pi_i - v)^2$, their optimal quotes, $\pi_i$, are equal to $v$. Thus, $IMM = v$. *QED*.

This result further restricts the set of possible full-game equilibria. When there are no frictions, the final auction price cannot differ from the fair value of the bond by more than the size of the spread, $s$.

In the presence of frictions, there can be either underpricing or overpricing equilibria in the auction without a cap (Propositions 5 and 6). The cap cannot eliminate underpricing equilibria.[4] Additionally, if the cap is set too low it rules out equilibria with $p^A = v$.

The cap can, however, be effective at eliminating overpricing equilibria. As an illustration, consider a simple example in which all dealers have zero CDS positions.[5] Proposition 6 shows that when there are short-sale constraints, the final auction price can be as high as 100 in the absence of a cap. Following the same logic as in Proposition 6, one can show that if the cap is greater than $v$, then there exists an equilibrium with the final price equal to the cap. Since in any such equilibrium dealers do not realize any profit but have regulatory and reputational concerns, their optimal quotes are equal to $v$. Thus, $IMM = v$ and $p^A = v + s$.

# 4 Empirical Evidence

Our theoretical analysis shows that CDS auctions may result in both overpricing and underpricing of the underlying bonds. In this section, we seek to provide empirical evidence indicating which outcomes occur in practice. Unfortunately, the true value of deliverable bonds is never observed. Because of this, we use available bond prices from the day before the auction to construct a proxy for the bond value $v$. Admittedly this measure is not a perfect substitute for the true value of the bond, and so we also consider a number of alternatives to show the robustness of our results. We first describe our data before presenting the empirical analysis.

[4] For example, consider an extreme case in which all dealers have large positive CDS positions and the conditions of Proposition 5 hold. Following the logic of Proposition 5, one can show that there exists a subgame perfect equilibrium in which $p^A = 0$ and $IMM = v$.

[5] While this is a simplification it is arguably also realistic, as dealers try to maintain zero CDS positions in their capacity as market makers.

## 4.1 Data

Our data come from two primary sources. The details of the auction settlement process are publicly available from the Creditfixings website (www.creditfixings.com). As of December 2010, there have been 86 CDS and Loan CDS auctions, settling contracts on both US and international legal entities. To study the relationship between auction outcomes and the underlying bond values, we merge these data with bond price data from the TRACE database. TRACE reports corporate bond trades for US companies only. Thus, our merged dataset contains 23 auctions.

Table 2 summarizes the results of the auctions for these firms. It reports the settlement date, the type of credit event and the auction outcomes. Most of the auctions took place in 2009 and were triggered by the Chapter 11 event. In only two of the 23 auctions (Six Flags and General Motors) was the net open interest to buy ($NOI < 0$). The full universe of CDS auctions contains 61 auctions in which the net open interest was to sell, 19 auctions where the net open interest was to buy, and 6 auctions with zero net open interest.

Table 3 provides summary statistics of the deliverable bonds for each auction for which we have bond data.[6] Deliverable bonds are specified in the auction protocols, available from the Creditfixings website. The table also reports the ratio of net open interest to the notional amount of deliverable bonds ($NOI/NAB$). This shows how many units of bonds changed hands during an auction, as a percentage of the total amount of bonds. There is strong heterogeneity in $NOI/NAB$ across different auctions, with absolute values ranging from 0.38% to 56.81%. In practice, $NOI$ has never exceeded $NAB$.

We construct daily bond prices by weighing the price for each trade against the trade size reported in TRACE, as in Bessembinder, Kahle, Maxwell, and Xuet (2009). These authors advocate eliminating all trades under $100,000 as they are likely to be non-institutional. The larger trades have lower execution costs; hence they should reflect the underlying bond value with greater precision. For each company, we build

[6] A clarification regarding the auctions of Abitibi and Bowater is in order. AbitibiBowater is a corporation, formed by Abitibi and Bowater for the sole purpose of effecting their combination. Upon completion of the combination, Abitibi and Bowater became subsidiaries of AbitibiBowater and the businesses that were formerly conducted by Abitibi and Bowater became the single business of AbitibiBowater. The CDS contracts were linked to the entities separately, and, as a result, there were two separate auctions.

a time-series of bond prices in the auction event window of -30 to +30 trading days. Because all credit events occur no more than one calendar month before the CDS auction, our choice of the event window ensures that our sample contains all relevant data for the post-credit-event prices. The last column of Table 3 reports a weighted average bond price on the day before the auction, $p_{-1}$. We use this as our proxy for the bond value $v$.

## 4.2 The Impact of the First Stage

The theoretical results of Section 3 imply that the first and the second stages of the auction are not independent. The first stage yields the mid-point price, $p^M$, which determines a cap on the final settlement price. Our model shows that when the final price, $p^A$, is capped, it can be either above or below the true value of the bond, $v$, depending on the initial CDS and bond positions of different agents.

Our analysis suggests a way of differentiating between the two cases. To be more specific, consider outcomes in which $NOI > 0$ (outcomes in which $NOI < 0$ follow similar logic). According to Proposition 1 Part 1, the price can be higher than $v$ if, after the first stage, the aggregate short net CDS position of agents participating in the second stage is larger than the net open interest. In this case, protection sellers have an incentive to bid above the true value of the bond to minimize the amount paid to their CDS counterparties. Notice that while bidding at a price above $v$, they would like to minimize the amount of bonds acquired at the auction for a given final auction price. Thus, if the price is above $v$ they will never bid to buy more than $NOI$ units of bonds.

The case in which $p^A$ is capped and lies below the true value of the bond is brought about when dealers set $p^M$ so that $p^M + s$ is below $v$. This prevents the agents from playing second-stage equilibrium strategies with the final price above the cap. In this case, submitting a large demand at the cap price leads to greater profit. Thus, in the presence of competition and sharing rules, agents have an incentive to buy as many bonds as possible and would bid for substantially more than $NOI$ units.

The final price is capped in 19 of the 86 credit-event auctions.[7] Figure 2 shows the entities and the individual bids at the cap price. The individual bids are represented

[7]Of these 19 auctions, only one (Ecuador) has a negative $NOI$. So the above discussion for the case of positive $NOI$ should be adjusted appropriately for Ecuador.

by different colors, and bid sizes are scaled by $NOI$ to streamline their interpretation. For example, there are seven bids at the cap price in the case of General Growth Properties. Six of these are equal to $NOI$ and the seventh one is approximately one-fourth of $NOI$.

We can see that in all but two auctions (Kaupthing Bank and Glitnir), the bids at the price cap do not exceed $NOI$. The results suggest that in these cases the final auction price is above the true bond value. Of the 19 auctions with a capped price, we have bond data for only five companies: Smurfit-Stone, Rouse, Charter Communications, Capmark and Bowater. Comparing the final auction price from Table 2 with the bond price from Table 3, we can see that that the bond price (our proxy for the true bond value) is below the final auction price for these five companies, as expected.

We can compare the bond and auction prices for the rest of the companies for which TRACE data are available. Figure 3 shows the ratio of final auction prices to bond prices, $p^A/p_{-1}$. We see that in all but seven auctions, the final auction price, $p^A$, is below the bond price, $p_{-1}$. It seems likely that underpricing equilibria were played out in these auctions. The exceptions include the aforementioned five companies with capped auction prices, as well as the General Motors and Six Flags auctions, where the price was not capped but $NOI < 0$. In these last two cases, the auction prices are expected to exhibit a reverse pattern.

## 4.3 Price Impact at the Second Stage

In the preceding section, our evidence showed that in the absence of a cap, the auction yields a price below the bond value. According to Proposition 1, if $NOI > 0$ such an outcome can occur only if the aggregate net short CDS position of the agents who participate in the second stage, $\sum_{i\in\mathcal{N}_+:n_i<0}|n_i - y_i|$, is smaller than or equal to the net open interest. But as we do not have data on individual bids and positions we cannot test this proposition directly. Instead, we provide empirical evidence that complements our theoretical analysis. Specifically, we study the effect of the $NOI$ on the degree of price discrepancy resulting from the auction. We scale the net open interest by the notional amount of deliverable bonds, giving the quantity $NOI/NAB$, to allow for a meaningful cross-sectional examination.

Tables 2 and 3 reveal that $NOI/NAB$ is greatest in the auctions with the largest discrepancy in prices. At the same time, $NOI/NAB$ is lowest in the auctions where the final price is capped, which is again consistent with Propositions 1 and 5. We quantify this relationship using a simple cross-sectional regression of $p_{-1}/p^A$ on $NOI/NAB$ :

$$p^A/p^{-1} = \alpha + \beta \times NOI/NAB + \varepsilon. \tag{22}$$

Figure 4 shows the results. The normalized $NOI$ explains 55% of the variation in the ratio of the lag of the market price of bonds to the final price. The $\beta$ is significantly negative. For every one-percentage-point increase in the normalized $NOI$, the underpricing increases by 1.2%.

This evidence is consistent with Proposition 5, which shows that there exist second-stage equilibria in which the final price, $p^A$, depends linearly on the $NOI$ (equation (17)). Since the only theoretical restriction on the slope ($\delta$) is its sign, the linear relationship (17) can be written as

$$p^A/v = 1 + \beta \times NOI/NAB, \quad \beta < 0.$$

If agents play equilibrium strategies with the same $\beta$ across auctions, the estimated cross-sectional regression $\beta$ will also be an estimate of the within-auction relationship. While the assumption of the same linear dependence across auctions is admittedly strong, it can be accommodated by the following argument. If all agents in an auction take historical information about previous types of equilibria into account when forming their perceptions, then $\beta$ is unlikely to vary much across auctions. Finally, the estimated $\alpha$ is insignificantly different from one, which is again consistent with the theory.

## 4.4 Robustness Checks

### 4.4.1 Fair Value Proxy

Our conclusions so far rest on the assumption that $p_{-1}$ is a good proxy for the actual fair value $v$. One could argue that auctions exist precisely because it is difficult to establish a bond's fair value by observing bond markets. Moreover, even if $p_{-1}$ were to reflect the bond value accurately, it would still be the value on the day before the auction. It is conceivable that the auction process establishes a $v$ that differs from $p_{-1}$ simply due to the arrival of new information between time $-1$ and 0, and/or as a result of the centralized clearing mechanism of the auction.

We expand the auction event window to check the robustness of our results to these caveats. In our sample, the shortest time between a credit event and an auction is 8 days. This prompts us to select an event window of -8 to +12 days. The choice of boundary is dictated by liquidity considerations: liquidity generally declines after the auction. Figure 5 (a) displays daily bond prices normalized by the auction final price, $p_t/p^A$, equally weighted across the 22 auctions for which we have reliable bond data.[8] We see that the price generally declines, reaches its minimum on the auction day, then reverts to its initial level. The figure shows that no matter which day we look at, the auction final price is, on average, at least 10% lower.

We have a sample of 22 auctions with reliable data. This small sample size may raise concerns that our results are sensitive to outliers. In what follows, we discuss the effects of two types of such outliers.

First, the Tribune auction stands out because of the large magnitude of the underpricing it generated. This can be seen in Figure 4: the point in the lower right corner of the plot. The normalized $NOI$ was also the largest, so the magnitude of the underpricing on its own is consistent with our theory. Nonetheless, to be sure that the pattern of average prices is not driven by this one company, we remove Tribune from our sample and recompute the pattern. Figure 5 (b) shows the results. We see that the magnitude of the average smallest underpricing declines to 5%, but all qualitative features remain intact.

Second, there are six auctions that resulted in overpricing. Four auctions (Smurfit-Stone, Rouse, Capmark and Bowater) had positive $NOI$ and final price equal to the cap (see Section 4.2 for details). The two remaining auctions (GM and Six Flags) had negative $NOI$. Therefore, the presence of these names in our average may only bias our results against finding underpricing.

The documented V shape of the discrepancy alleviates the concern that the correct value $v$ differs from $p_{-1}$ simply because the latter does not reflect the bond value

[8] We exclude the auction for Charter, which has only 10 trades in the [-10,0] window during which our proxy for $v$ is constructed. Of these 10 trades, only 6 are in sizes greater than \$1M. The second-worst company in terms of data reliability, Chemtura, has 35 trades and all of them are above \$1M.

correctly. If this were the case, one would expect bond prices to remain in the region of the auction price after the auction, whereas in practice they increase.

### 4.4.2 The Cheapest-to-Deliver Option

Another potential concern about using weighted daily bond prices as a proxy for the underlying value of auctioned bonds is that agents will likely use only cheapest-to-deliver bonds for physical delivery. As a result, our methodology may overestimate the fair value. This argument is not applicable when the credit event is Chapter 11, and all the deliverable bonds are issued by the holding company and cross-guaranteed by all subsidiaries. In Chapter 11, bonds with no legal subordination are treated as identical, see for example Guha (2002).[9] The reasons for this are that all the bonds stop paying coupons and mature (cease to exist) at the same time, with identical terminal payouts to all bondholders. Hence there is no concern that some bonds are cheaper to deliver due to the difference in their fundamental value.

As an example, Figure 6 shows weighted daily prices of each individual WaMu bond issue, identified by its CUSIP. We see that there are large difference between the prices of different bonds in the period leading to the credit event (trading day -19). After this day the prices of all bonds are very similar. The prices cannot be literally identical because trades may occur at different times of the day, and because trades may be either buyer- or seller-initiated which means prices will be closer to bid or ask prices, respectively.

In our sample, 13 out of 23 credit events are triggered by Chapter 11 bankruptcy and have one issuer. These companies should not have bonds that diverge in value. Nonetheless, we manually confirm that this is indeed the case. There are three companies that filed for Chapter 11 and have multiple subsidiaries issuing bonds, but for which TRACE contains trade data for only one subsidiary in the event window (CIT, Lyondell, and Quebecor). We treat these three names the same way as the 13 firms without subsidiaries.

There are four companies that filed for Chapter 11 and have multiple subsidiaries, and where we have data for the bonds of these subsidiaries (Bowater, Charter, Nortel and Smurfit-Stone). In all of these cases the bonds of the different subsidiaries are

[9] CDS contracts on bonds with different seniorities are settled in different auctions. Examples of this in our data are the Dura/Dura Sub auctions.

legally pari-passu with each other, but some of them may be structurally subordinated to others and, therefore, could be cheaper. For this reason, we select the cheapest bonds in the case of these four companies (however, the differences are not large in practice). There are three companies with a credit event other than Chapter 11 (Abitibi, Capmark and Rouse) in which we also select the cheapest bonds.

Finally, to account for other potential deliverables selection issues that could work against our findings, we treat the aforementioned differences in bond prices (due to bid-ask spread and timing differences) as real differences, and select the lowest-priced bonds. Specifically, we take representative daily prices of a company's deliverable bonds to be equal to the weighted daily prices of their bond issues with the lowest pre-auction price, provided that these bond issues are relatively actively traded.[10] The results are displayed in Figure 5. It can be seen that even with these conservative bond selection criteria, the average underpricing on the day of the auction is still 10%, and follows a V pattern as before.[11]

# 5 Extensions

Section 4 documents our finding that when $NOI/NAB$ is large, the auction generally results in a price considerably below fair value. We now suggest several modifications to the auction design that can reduce mispricing, and discuss some of the assumptions of the model.

## 5.1 Allocation Rule at the Second Stage

As usual, we focus on the case of $NOI > 0$. Proposition 1 shows that if condition (6) holds, the CDS auction is similar to a 'standard' auction, so the price can be below $v$. Kremer and Nyborg (2004) show that in a setting without CDS positions, a simple change of the allocation rule from pro-rata on the margin (2) to 'pro-rata' destroys all underpricing equilibria, so that only $p^A = v$ remains. Under the pro-rata rule, the

[10]The requirement is that the trading volume over the five trading days before the auction constitutes at least 5% of total trading volume for the company.

[11]Gupta and Sundaram (2011) address the cheapest-to-deliver issue using an alternative procedure based on econometric modelling of issue-specific pricing biases, and arrive at similar conclusions.

equilibrium allocations $q_i$ are given by

$$q_i(p^A) = \frac{x_i(p^A)}{X(p^A)} \times NOI.$$

That is, the total rather than marginal demand at $p^A$ is rationed among agents. The next proposition extends the result of Kremer and Nyborg (2004) to our setting. We demonstrate that if $p^M + s \geq v$, then the second-stage equilibrium price $p^A$ cannot be less than $v$. This is true even if the agents are allowed to hold non-zero quantities of CDS contracts.

**Proposition 8** *Suppose that the auction sharing rule is pro-rata. In this case, if* $NOI > 0$ *then* $p^A \geq \min\{p^M + s, v\}$. *If* $NOI < 0$ *then* $p^A \leq \max\{p^M - s, v\}$.

**Proof.** See Appendix.

To develop intuition for this result, consider the case of positive $NOI$. According to Proposition 1 Part 2, if condition (6) holds, the pro-rata on the margin allocation rule may inhibit competition and lead to underpricing equilibria. The presence of agents who are short CDS contracts does not help in this case. The pro-rata allocation rule (i) does not guarantee the agents their inframarginal demand above the clearing price, and (ii) closely ties the proportion of allocated bonds to the ratio of individual to total demand at the clearing price. Therefore, a switch to such a rule would increase competition for bonds among agents. As a result, even agents with long positions would bid aggressively. If $p^A < v$, demanding the $NOI$ at a price only slightly higher than $p^A$ allows an agent to capture at least half of the surplus. As a result, only fair-price equilibria survive.

## 5.2 The Price Cap

Our theoretical analysis in Section 4.2 shows that the presence of a price cap can result in auction outcomes with either lower or higher mispricing. The cap is likely to help when $|NOI|$ is small and the temptation to manipulate the auction results is highest. At the same time, the cap allows dealers to limit the final price to below $v$ in the second stage.

These results suggest that making the cap conditional on the outcome of the first stage of a CDS auction can lead to better outcomes. In our base model without uncertainty, the optimal conditional cap is trivial. Again, we consider the case of $NOI > 0$. If $p^M < v$, setting $s^* = v - p^M$ ensures that the set of second-stage equilibria includes $v$. If $p^M \geq v$, it is best to set $s^* = 0$. While the conditional cap cannot eliminate the worst underpricing equilibria, it can ensure that agents who want to bid aggressively will be able to do so.

In practice, $v$ and $n_i$ are unobservable. Thus, making the cap conditional on $NOI$ and on the ratio $p^M/p_{-1}$ could lead to the final auction price being closer to the fair bond value. For example, if $p^M/p_{-1} \leq \alpha$, $NOI$ is large and $\alpha < 1$ is reasonably small, the auctioneer can set a higher cap; if $p^M/p_{-1} > \alpha$ and $NOI$ is small, a lower cap can be set.

## 5.3 Risk-averse agents

So far we have considered only risk-neutral agents. This allowed us to abstract from risk considerations. If agents are risk-averse, the reference entity's risk is generally priced. Even though a CDS is in zero net supply, its settlement leads to a reallocation of risk among the participants in the auction; hence it can lead to a different equilibrium bond price. In particular, when $NOI/NAB$ is large and positive, and there are only a few risk-averse agents willing to hold defaulted bonds, the auction results in highly-concentrated ownership of the company's risk and can thus lead to a lower equilibrium bond price.

Notice, however, that risk-aversion does not automatically imply a lower auction price. For example, if marginal buyers of bonds in the auction are agents who previously had large negative CDS positions (as in Proposition 5), their risk exposure after the auction may actually decrease. As a result they could require a lower risk premium.

Due to the fact that we do not have data on individual agents' bids and positions, we cannot determine whether the observed price discrepancy is due to mispricing equilibria or risk-aversion. It is likely that both factors work together in the same direction. Data on individual agents' bids and positions could help to quantify the effect of the two factors on the observed relationship between the auction price and the size of net open interest.

### 5.4 Private information

Up to this point we have restricted our attention to the simplest case in which agents' CDS positions are common knowledge. This may seem like a very strong assumption given that CDS contracts are traded in the OTC market. Notice, however, that in the type of equilibria constructed in Propositions 5 (linear case) and 6, conditions (14), (15) and (19), (20) completely define the two equilibria. Therefore, Propositions 5 and 6 continue to hold with private CDS positions as long as (14), (15) and (19), (20) are public knowledge.[12] One can argue that this is likely to be the case. For example, (20) assumes that total short CDS positions are larger than total bond holdings of agents with long CDS positions. The aggregate net CDS positions are known to market participants.[13] Therefore, whether condition (20) holds can be easily verified in every auction. Similarly, (19) assumes that there is an agent whose long position in CDSs is larger than her bond holdings. Given the much larger size of CDS contracts compared to the value of bonds outstanding, (19) holds as long aggregate long CDS positions are larger than the value of the outstanding bonds. The latter is true for most (if not all) of the auctions.

We also assume that agents value bonds identically, and that this value is common knowledge. This assumption provides a stark benchmark: we are able to show that the auction results in mispricing even in such a basic case. We conjecture that it would be even harder for the current auction mechanism to arrive at the fair value when agents have private or heterogeneous valuations.

## 6 Conclusion

We present a theoretical and empirical analysis of the settlement of CDS contracts when a credit event takes place. A two-stage, auction-based procedure aims to establish a reference bond price for cash settlement and to provide market participants

[12] The formal proofs follow closely the original proofs for the full information case and are available upon request.

[13] For example, they are available from Markit reports.

with the option to replicate a physical settlement outcome. The first stage determines the net open interest ($NOI$) in the physical settlement and the auction price cap (minimum or maximum price, depending on whether the $NOI$ is to sell or to buy). The second stage is a uniform divisible good auction with a marginal pro-rata allocation rule that establishes the final price by clearing the $NOI$.

In our theoretical analysis, we show that the auction may result in either overpricing or underpricing of the underlying bonds. Our empirical analysis establishes that the former case is more prevalent in practice. Bonds are underpriced by 10% on average, and the amount of underpricing increases with the $NOI$ (normalized by the notional amount of deliverable bonds). We propose introducing a pro-rata allocation rule and a conditional price cap to mitigate this mispricing.

# References


[1] Viral Acharya and Timothy Johnson, 2007. "Insider Trading in Credit Derivatives," Journal of Financial Economics, vol. 84(1), pages 110-141.

[2] Navneet Arora, Priyank Gandhi, and Francis A. Longstaff, 2009. "Counterparty Credit Risk and the Credit Default Swap Market," working paper, UCLA.

[3] Kerry Back and Jaime F. Zender, 1993. "Auctions of Divisible Goods: On the Rationale for the Treasury Experiment," Review of Financial Studies, vol. 6(4), pages 733-764.

[4] Hendrik Bessembinder, Kathleen M. Kahle, William F. Maxwell, and Danielle Xu, 2009. "Measuring Abnormal Bond Performance," Review of Financial Studies, vol. 22(10), pages 4219-4258.

[5] Patrick Bolton and Martin Oehmke, 2011. "Credit Default Swaps and the Empty Creditor Problem," Review of Financial Studies, forthcoming.

[6] Virginie Coudert and Mathieu Gex, 2010. "The Credit Default Swap Market and the Settlement of Large Defaults," working paper no. 2010-17, CEPII.

[7] Songzi Du and Haoxiang Zhu, 2011. "Are CDS Auctions Biased?" working paper, Stanford GSB.

[8] Durrell Duffie, 1999. "Credit Swap Valuation," Financial Analysts Journal, January-February, pages 73-87.

[9] Durrell Duffie and Haoxiang Zhu, 2011. "Does a Central Clearing Counterparty Reduce Counterparty Risk?" Review of Asset Pricing Studies, forthcoming.

[10] Nicolae Garleanu and Lasse Pedersen, 2011. "Margin-Based Asset Pricing and Deviations from the Law of One Price," Review of Financial Studies, forthcoming.

[11] Rajiv Guha, 2002. "Recovery of Face Value at Default: Theory and Empirical Evidence," working paper, LBS.

[12] Sudip Gupta and Rangarajan K. Sundaram, 2011. "CDS Credit-Event Auctions," working paper, New York University.

[13] Jean Helwege, Samuel Maurer, Asani Sarkar, and Yuan Wang, 2009. "Credit Default Swap Auctions and Price Discovery," Journal of Fixed Income, Fall 2009, pages 34-42.

[14] Ilan Kremer and Kjell G. Nyborg, 2004. "Divisible-Good Auctions: The Role of Allocation Rules," RAND Journal of Economics, vol. 35(1), pages 147-159.

[15] Francis A. Longstaff, Sanjay Mithal and Eric Neis, 2005. "Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit-Default Swap Market," Journal of Finance, vol. 60, pages 2213-2253.

[16] Jun Pan and Kenneth Singleton, 2008. "Default and Recovery Implicit in the Term Structure of Sovereign CDS Spreads," Journal of Finance, vol. 63, pages 2345-2384.

[17] Christine Parlour and Andrew Winton, 2010. "Laying Off Credit Risk: Loan Sales and Credit Default Swaps," working paper, UC Berkeley.

[18] Robert Wilson, 1979. "Auctions of Shares," Quarterly Journal of Economics, vol. 93(4), pages 675-689.

# Appendix

## Proof of Proposition 5

The proof is by construction. We construct a subgame perfect two-stage equilibrium in which the final auction price is a decreasing function of the $NOI$. In a similar fashion to Kremer and Nyborg (2004), it can be shown that one's attention can be restricted w.l.o.g. to equilibria in differentiable strategies. For simplicity, we provide the proof for the case in which agents have large long CDS positions. Specifically, we assume that for all $i : n_i > 0$ :

$$n_i \geq NOI. \tag{A1}$$

Under this additional assumption, we can solve for the equilibrium in closed form. The general case follows similar logic, except that the number of the agents who submit nonzero demand for bonds at the second stage depends on the configuration of CDS positions. When A1 holds, only agents with non-positive CDS positions receive nonzero allocations in the equilibrium.

The proof consists of several steps. In step 1, we derive the F.O.C. for the optimal strategies at the second stage, given the remaining CDS positions of the agents after the first stage. In step 2, we derive the F.O.C. for the optimal physical settlement requests. In step 3, we show that the second-stage equilibrium with price $p^A$ can be supported if agents play the following second-stage strategies:

$$x_i(p) = \max\{c(v-p)^{\lambda} - n_i + y_i, 0\}$$

($c$ and $\lambda$ are specified later). In step 4, we solve for optimal physical requests of agents, given the above second-stage strategies. Finally, we solve for the $NOI$.

**Step 1.** Recall that at the second stage, player $i$ solves problem (4):

$$\max_{x_i(p)} \; (v - p(x_i(p), x_{-i}(p)))\, q_i(x_i(p), x_{-i}(p)) + (n_i - y_i) \times (100 - p(x_i(p), x_{-i}(p)))\,.$$

In any equilibrium of the second stage, the sum of the demand of agent $i$, $x_i(p^A)$, and the residual demand of the other players, $x_{-i}(p^A)$, must equal the $NOI$. Therefore, solving for the optimal $x_i(p)$ is equivalent to solving for the optimal price, $p^A$, given the residual demand of the other players. Thus, the F.O.C. for agent $i$ at the equilibrium price, $p^A$, can be written as

$$(v - p^A)\frac{\partial x_{-i}(p^A)}{\partial p} + x_i(p^A) + n_i - y_i = 0 \qquad \text{if} \qquad x_i(p^A) > 0, \tag{A2}$$

$$(v - p^A)\frac{\partial x_{-i}(p^A)}{\partial p} + x_i(p^A) + n_i - y_i \geq 0 \qquad \text{if} \qquad x_i(p^A) = 0. \tag{A3}$$

**Step 2.** Recall that agent $i$'s profit is given by equation (3):

$$\begin{aligned} \Pi_i \;=\; & \underbrace{(v - p^A)q_i}_{\text{auction-allocated bonds}} + \underbrace{(n_i - y_i) \times (100 - p^A)}_{\text{remaining CDS}} \\ + \; & \underbrace{100y_i}_{\text{physical settlement}} + \underbrace{v(b_i - y_i)}_{\text{remaining bonds}}. \end{aligned}$$

Using the fact that $\partial NOI/\partial y_i = 1$, we have that the F.O.C. for the optimal settlement amount, $y_i$, for agent $i$, satisfies

$$\frac{\partial \Pi_i}{\partial y_i} = 0 \quad \text{if} \quad y_i \neq 0 \quad \text{and} \quad y_i \neq n_i, \tag{A4}$$

$$\frac{\partial \Pi_i}{\partial y_i} \leq 0 \quad \text{if} \quad y_i = 0 \quad \text{and} \quad n_i > 0, \quad \text{or} \quad y_i = n_i \quad \text{if} \quad n_i < 0, \tag{A5}$$

$$\frac{\partial \Pi_i}{\partial y_i} \geq 0 \quad \text{if} \quad y_i = 0 \quad \text{and} \quad n_i < 0, \quad \text{or} \quad y_i = n_i \quad \text{if} \quad n_i > 0, \tag{A6}$$

where

$$\frac{\partial \Pi_i}{\partial y_i} = -\frac{\partial p^A(NOI)}{\partial NOI}(n_i - y_i + q_i) - (v - p^A(NOI))\left(1 - \frac{\partial q_i}{\partial y_i}\right). \tag{A7}$$

**Step 3.** Let $M$ be the number of agents with nonpositive CDS positions who are allowed to hold bonds, and let $\lambda = 1/(M-1)$. Then consider the following set of strategies at the second stage:

$$x_i(p) = \max\left\{\frac{NOI + \sum_{j \in \mathcal{N}_+ : n_j < 0}(n_j - y_j)}{M}\frac{(v-p)^\lambda}{(v - p^A(NOI))^\lambda} - n_i + y_i, 0\right\}. \tag{A8}$$

Demand schedules (A8) imply that agents with non-positive CDS positions who are allowed to hold bonds receive, at $p = p^A$, the following bond allocations:

$$q_i = \frac{NOI + \sum_{j \in \mathcal{N}_+ : n_j < 0} (n_j - y_j)}{M} - (n_i - y_i). \tag{A9}$$

Equation (A3) implies that agents with initial long CDS positions receive zero equilibrium bond allocations at the second stage, as long as

$$n_i - y_i \geq \frac{NOI + \sum_{j \in \mathcal{N}_+ : n_j < 0} (n_j - y_j)}{M - 1}. \tag{A10}$$

If this is the case, equation (A2) implies that strategies (A8) form an equilibrium at the second stage, with the equilibrium price equal to $p^A$.[14]

**Step 4.** Consider now the optimal physical settlement requests of agents with initial short CDS positions. We need consider only those agents who are allowed to hold bonds after the auction. As part of the equilibrium constructed in step 3, these agents receive $q_i$ units of bonds, as given in (A9). So we can write condition (A7) as

$$\frac{\partial \Pi_i}{\partial y_i} = -\frac{\partial p^A(NOI)}{\partial NOI} \frac{NOI + \sum_{j \in \mathcal{N}_+ : n_j < 0} (n_j - y_j)}{M} - \frac{v - p^A(NOI)}{M}. \tag{A11}$$

For simplicity, we solve for the interior solution so that $\frac{\partial \Pi_i}{\partial y_i} = 0$. Direct computations show that in such an equilibrium it must be the case that

$$NOI + \sum_{j \in \mathcal{N}_+ : n_j < 0} (n_j - y_j) = \left(v - p^A(NOI)\right) / \left| \frac{\partial p^A(NOI)}{\partial NOI} \right|. \tag{A12}$$

Now consider the optimal physical settlement requests of agents with initial long CDS positions. If these agents receive a zero equilibrium bond allocation, conditions

[14] Technically, one extra condition is needed to ensure the existence of the constructed equilibrium. Inequality (A10) must continue to hold for every possible deviation $\hat{y}_j > y_j$ by each participant $j \in \mathcal{N}$. If for some such $\hat{y}_j$ this condition breaks down for agent $i$ with a long CDS position, then this agent will participate in the second stage of the auction, which could increase the profit earned by agent $j$. (Of course, agent $i$ could increase $y_i$ itself, in which case $j = i$). This extra condition does not hold for $\hat{y}_i > y_i$ when $M = 2$, which leads to existence of $p^A = 0$ underpricing equilibria only. When $M > 2$ there exist underpricing equilibria with $p^A > 0$, in which out-of-equilibrium submission of physical settlement requests does not lead agents with long CDS positions to participate in the second stage of the auction. The details are available upon request.

(A4) and (A7) imply that their optimal physical requests satisfy

$$y_i = \max \left\{ n_i - \left(v - p^A(NOI)\right) / \left| \frac{\partial p^A(NOI)}{\partial NOI} \right|, 0 \right\}. \tag{A13}$$

Using equilibrium condition (A12) together with condition (A10), we can see that agents with initial long CDS positions will receive a zero equilibrium bond allocation at the second stage if

$$n_i \geq \frac{\left(v - p^A(NOI)\right) / |\frac{\partial p^A(NOI)}{\partial NOI}|}{M-1}. \tag{A14}$$

Assumption (A1), along with condition (16), guarantee an interior solution for the optimal physical requests of agents with initial long CDS positions.

**Step 5.** Finally, the optimal physical requests of the agents must sum to the $NOI$:

$$\sum_{i:n_i>0} \left( n_i - \frac{v - p^A(NOI)}{\left| \frac{\partial p^A(NOI)}{\partial NOI} \right|} \right) + \sum_{i \in \mathcal{N}_+ : n_i < 0} y_i = NOI. \tag{A15}$$

Using (A12), we can write (A15) as

$$\sum_{i:n_i>0} n_i + \sum_{i \in \mathcal{N}_+ : n_i < 0} n_i - \frac{v - p^A(NOI)}{\left| \frac{\partial p^A(NOI)}{\partial NOI} \right|}(K+1) = 0, \tag{A16}$$

where $K$ is the number of agents with initial long CDS positions. Consider the case where $p^A(NOI) = v - \delta \times NOI$. Under this specification,

$$\frac{v - p^A(NOI)}{\left| \frac{\partial p^A(NOI)}{\partial NOI} \right|} = NOI.$$

Condition (A15) gives a simple formula for the $NOI$:

$$NOI = \frac{\sum_{i:n_i>0} n_i + \sum_{i \in \mathcal{N}_+ : n_i < 0} n_i}{K+1} > 0. \tag{A17}$$

$QED$.

## Proof of Proposition 8

As usual we focus on the case where $NOI > 0$. Note that the pro-rata allocation rule satisfies the *majority property* (Kremer and Nyborg, 2004): an agent whose demand at the clearing price is above 50% of the total demand is guaranteed to be allocated at least $(50\% + \eta) \times NOI$, where $\eta > 0$.

First, suppose that $v \leq p^M + s$. The proof that $p^A$ cannot be above $v$ is the same as in Proposition 1. We now prove that $p^A$ cannot be below $v$. Suppose instead that $p^A < v$. The part of agent $i$'s utility that depends on her equilibrium allocation and the final price is:

$$(v - p^A) \times q_i - p^A \times (n_i - y_i).$$

Suppose first that there is at least one agent for which $q_i < 0.5$. Suppose that this agent changes her demand schedule to:

$$x'_i(p) = \begin{cases} NOI, & p \leq p^A + \varepsilon \\ 0, & \text{otherwise,} \end{cases} \tag{A18}$$

where $0 < \varepsilon < v - p^A$. After this deviation, the new clearing price is $p^A + \varepsilon$. Since $X_{-i}(p^A + \varepsilon) < NOI$ (otherwise $p^A + \varepsilon$ would have been the clearing price), agent $i$ demands more than 50% at $p^A + \varepsilon$, and under the pro-rata allocation rule receives $q'_i > 0.5 \times NOI$. The lower bound on the relevant part of agent $i$'s utility is now:

$$(v - p^A - \varepsilon) \times 0.5 \times NOI - (p^A + \varepsilon) \times (n_i - y_i).$$

We can write the difference between agent $i$'s utility under deviation and her utility under the assumed equilibrium as follows:

$$(0.5 \times NOI - q_i) \times (v - p^A) - \varepsilon(n_i - y_i + 0.5 \times NOI). \tag{A19}$$

For small enough $\varepsilon$ and under the assumption that $p^A < v$, (A19) is greater than zero and hence equilibria with $p^A < v$ cannot exist.

If there are no agents with $q_i < 0.5 \times NOI$ we are in an auction with two bidders only. In this case, each of them gets exactly $0.5 \times NOI$. At price $p^A + \varepsilon$ ($0 < \varepsilon < p^M + s - v$), there is at least one player (player $i$), for which $x_i(p^A + \varepsilon) < 0.5 \times NOI$.

Then, if the opposite agent uses demand schedule (A18), the new clearing price will be $p^A + \varepsilon$ and this agent will receive at least $(0.5 + \eta) \times NOI$. For small enough $\varepsilon$ the difference between agent $i$'s utility under the deviation and her utility under the assumed equilibrium is:

$$\eta \times (v - p^A) - \varepsilon(n_i - y_i + (0.5 + \eta) \times NOI) > 0. \tag{A20}$$

Therefore, equilibria with $p^A < v$ cannot exist. We conclude that if $v \leq p^M + s$, then $p^A = v$ is the only clearing price in any equilibrium under the pro-rata allocation rule.

Finally, suppose that $p^M + s < v$. The proof for this case is the same, except that there is no feasible deviation to a higher price if $p^A = p^M + s$. Hence, $p^A = p^M + s < v$ is the only clearing price in any equilibrium under the pro-rata allocation rule. QED.

# Tables and Figures

Table 1: *Nortel Limited Market Quotes*

| Dealer | Bid | Offer |
|---|---|---|
| Banc of America Securities LLC | 9.5 | 11.5 |
| Barclays Bank PLC | 4.0 | 6.0 |
| BNP Paribas | 7.0 | 9.0 |
| Citigroup Global Markets Inc. | 10.5 | 12.5 |
| Credit Suisse International | 6.5 | 8.5 |
| Deutsche Bank AG | 6.0 | 8.0 |
| Goldman Sachs & Co. | 6.0 | 8.0 |
| J.P. Morgan Securities Inc. | 7.0 | 9.0 |
| Morgan Stanley & Co. Incorporated | 5.0 | 7.0 |
| The Royal Bank of Scotland PLC | 6.5 | 8.5 |
| UBS Securities LLC | 7.0 | 9.0 |

Table 1 shows the two-way quotes submitted by dealers at the first stage of the Nortel Ltd. auction.

Table 2: *Auction Summaries*

| Name | Date | Credit Event | Inside Market Quote | Net Open Interest | Final Price |
|---|---|---|---|---|---|
| Dura | 28 Nov 2006 | Chapter 11 | 24.875 | 20.000 | 24.125 |
| Dura Subordinated | 28 Nov 2006 | Chapter 11 | 4.250 | 77.000 | 3.500 |
| Quebecor | 19 Feb 2008 | Chapter 11 | 42.125 | 66.000 | 41.250 |
| Lehman Brothers | 10 Oct 2008 | Chapter 11 | 9.750 | 4920.000 | 8.625 |
| Washington Mutual | 23 Oct 2008 | Chapter 11 | 63.625 | 988.000 | 57.000 |
| Tribune | 6 Jan 2009 | Chapter 11 | 3.500 | 765.000 | 1.500 |
| Lyondell | 3 Feb 2009 | Chapter 11 | 23.250 | 143.238 | 15.500 |
| Nortel Corp. | 10 Feb 2009 | Chapter 11 | 12.125 | 290.470 | 12.000 |
| Smurfit-Stone | 19 Feb 2009 | Chapter 11 | 7.875 | 128.675 | 8.875 |
| Chemtura | 14 Apr 2009 | Chapter 11 | 20.875 | 98.738 | 15.000 |
| Great Lakes | 14 Apr 2009 | Ch 11 of Chemtura | 22.875 | 130.672 | 18.250 |
| Rouse | 15 Apr 2009 | Failure to pay | 28.250 | 8.585 | 29.250 |
| Abitibi | 17 Apr 2009 | Failure to pay | 3.750 | 234.247 | 3.250 |
| Charter Communications | 21 Apr 2009 | Chapter 11 | 1.375 | 49.2 | 2.375 |
| Capmark | 22 Apr 2009 | Failure to pay | 22.375 | 115.050 | 23.375 |
| Idearc | 23 Apr 2009 | Chapter 11 | 1.375 | 889.557 | 1.750 |
| Bowater | 12 May 2009 | Chapter 11 | 14.000 | 117.583 | 15.000 |
| R.H.Donnelly Corp. | 11 Jun 2009 | Chapter 11 | 4.875 | 143.900 | 4.875 |
| General Motors | 12 Jun 2009 | Chapter 11 | 11.000 | -529.098 | 12.500 |
| Visteon | 23 Jun 2009 | Chapter 11 | 4.750 | 179.677 | 3.000 |
| Six Flags | 9 Jul 2009 | Chapter 11 | 13.000 | -62.000 | 14.000 |
| Lear | 21 Jul 2009 | Chapter 11 | 40.125 | 172.528 | 38.500 |
| CIT | 1 Nov 2009 | Chapter 11 | 70.250 | 728.980 | 68.125 |

Table 2 summarizes the auction results for 23 US firms for which TRACE data are available. It reports the settlement date, type of credit event, inside market quote (per 100 of par), net open interest (in millions of USD), and final auction settlement price (per 100 of par).

Table 3: *Tradable Deliverable Bond Summary Statistics*

| Name | Number of deliverable bonds | Notional amount of bonds outstanding (NAB) | NOI/NAB (%) | Average price on the day before the auction |
|---|---|---|---|---|
| Dura | 1 | 350,000 | 5.71 | 25.16 |
| Dura Subordinated | 1 | 458,500 | 16.79 | 5.34 |
| Quebecor | 2 | 600,000 | 11.00 | 42.00 |
| Lehman Brothers | 157 | 42,873,290 | 11.47 | 12.98 |
| Washington Mutual | 9 | 4,750,000 | 20.80 | 64.79 |
| Tribune | 6 | 1,346,515 | 56.81 | 4.31 |
| Lyondell | 3 | 475,000 | 30.15 | 26.57 |
| Nortel Corp. | 5 | 3,149,800 | 9.22 | 14.19 |
| Smurfit-Stone | 5 | 2,275,000 | 5.65 | 7.77 |
| Chemtura | 3 | 1,050,000 | 9.40 | 26.5 |
| Great Lakes | 1 | 400,000 | 32.65 | 26.71 |
| Rouse | 4 | 1,350,000 | 0.63 | 29.00 |
| Abitibi | 10 | 3,000,000 | 7.81 | 4.61 |
| Charter Communications | 17 | 12,769,495 | 0.38 | 2.00 |
| Capmark | 2 | 1,700,000 | 6.79 | 22.75 |
| Idearc | 1 | 2,849,875 | 31.21 | 2.15 |
| Bowater | 6 | 1,875,000 | 6.27 | 14.12 |
| R.H.Donnelly Corp. | 7 | 3,770,255 | 3.81 | 5.12 |
| General Motors | 16 | 18,180,552 | -2.91 | 11.17 |
| Visteon | 2 | 1,150,000 | 15.62 | 74.87 |
| Six Flags | 4 | 1,495,000 | -4.14 | 13.26 |
| Lear | 3 | 1,298,750 | 13.28 | 39.27 |
| CIT | 281 | 22,584,893 | 3.29 | 69.35 |

Table 3 provides summary statistics of deliverable bonds for 23 US firms for which TRACE data are available. Column three reports the ratio of Table 2's net open interest (NOI) to the notional amount outstanding of deliverable bonds. The last column shows a weighted average bond price on the day before the auction, constructed as described in Section 4.1.

Figure 1: *IMM Determination: The Case of Nortel*

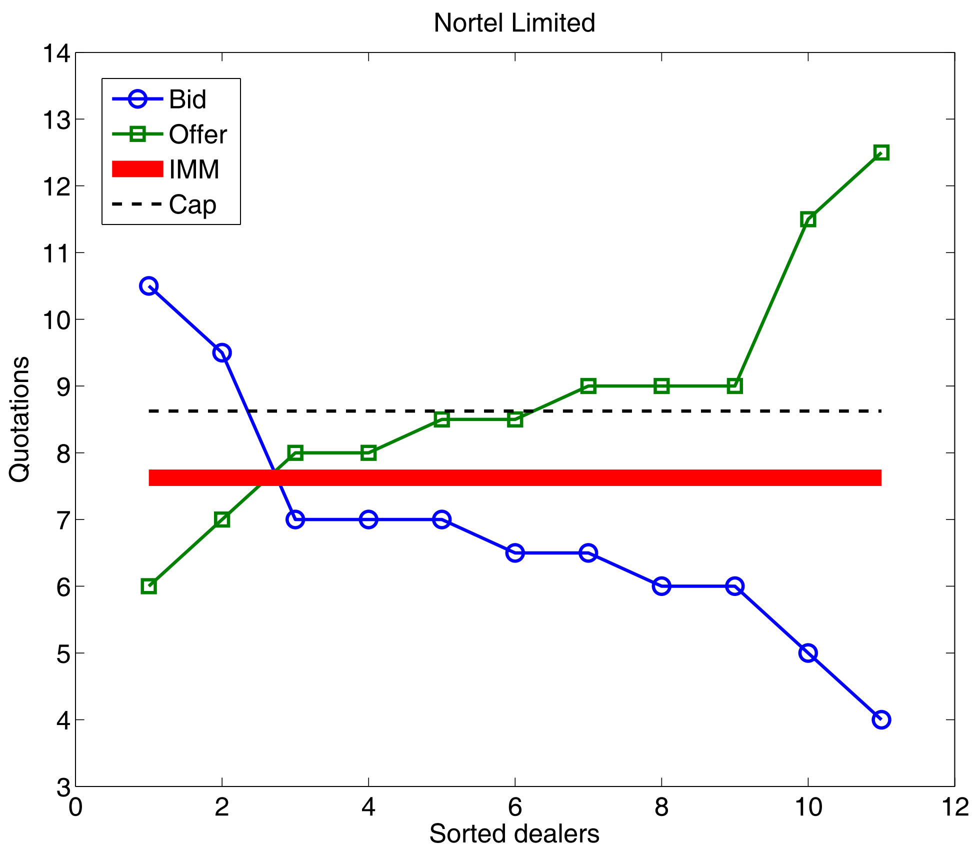



Figure 1 displays all bids (sorted in descending order) and all offers (sorted in ascending order). Tradeable quotes (bid greater than offer) are discarded for the purposes of computing IMM. Dealers quoting tradeable markets must pay a penalty (adjustment amount) to ISDA. The cap price is higher than the IMM by 1% of par and is used in determining the final price. (If the open interest is to buy, the cap price is set below the IMM.)

Figure 2: *Bids at the Cap Price*

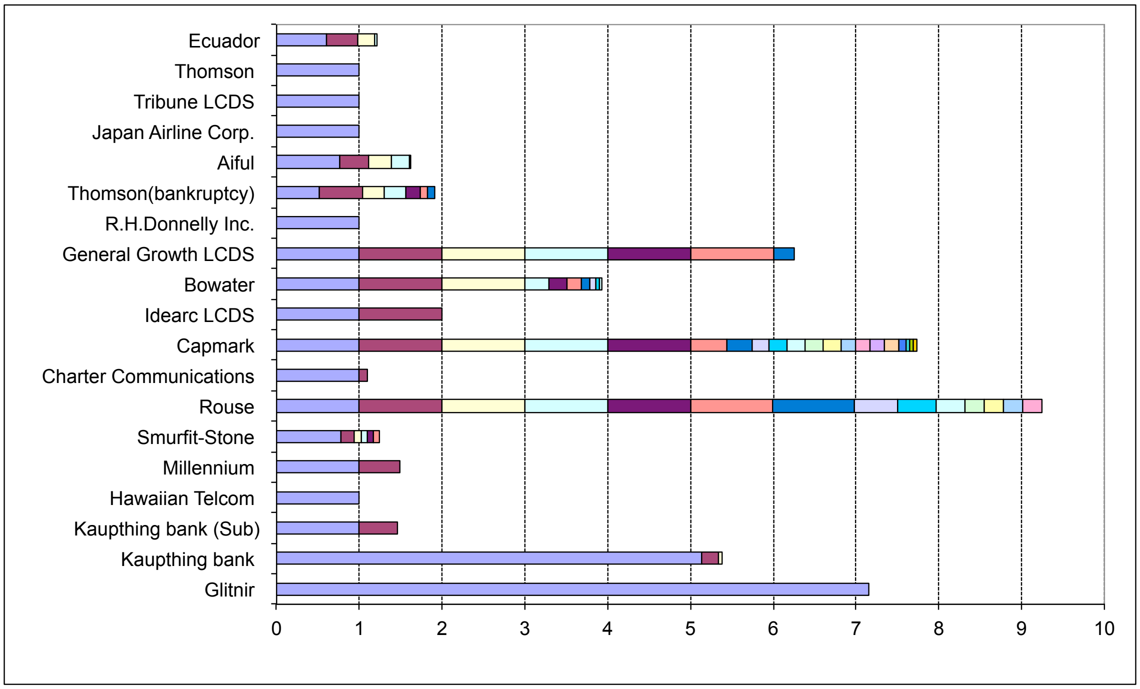



Figure 2 shows individual bids scaled by the $NOI$ at the cap price (in auctions where the price is capped). Each bid within an auction is represented by a different color.

Figure 3: *Bond and Auction Prices*

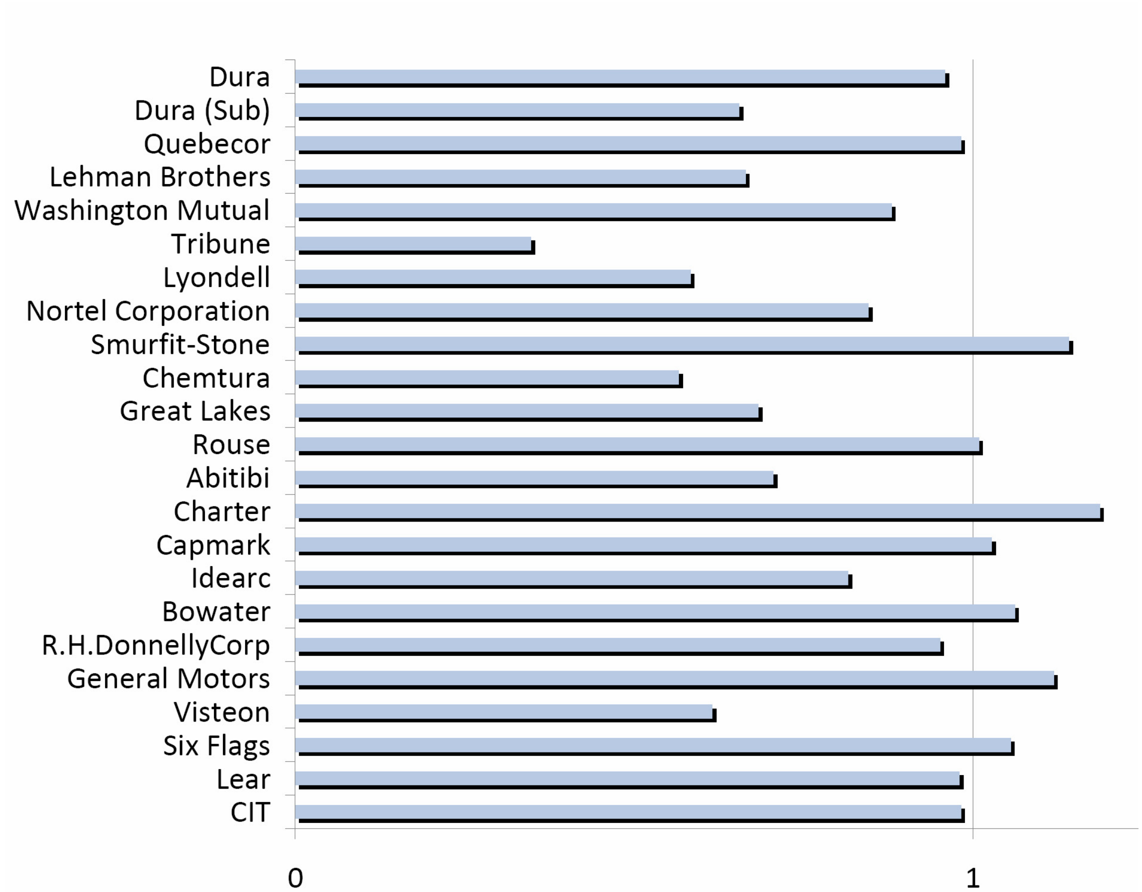



Figure 3 shows the final auction price, scaled by the weighted-average market price of the bonds a day before the auction.

Figure 4: *Price Discount*

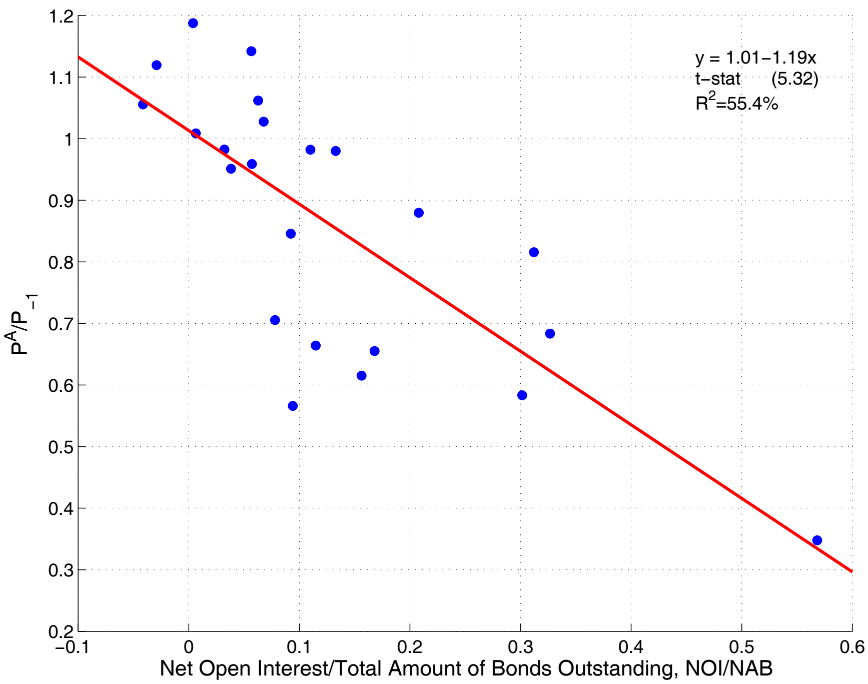
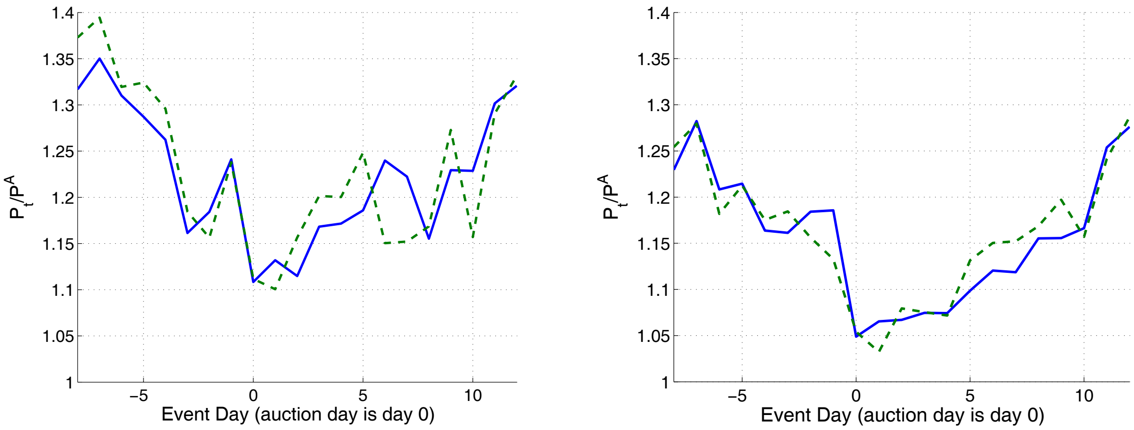



Figure 4 shows the result of an OLS regression where the dependent variable is the ratio of the final auction price to the weighted-average market price of bonds a day before the auction, and the explanatory variable is the scaled $NOI$:

$$y_i = \alpha + \beta \times NOI_i/NAB_i + \varepsilon_i.$$

Figure 5: *Price Impact*




$(a)$ $(b)$

Figure 5 Panel (a) displays daily bond prices, normalized by the auction final price, $p_t/p^A$, and equally weighted across the 22 auctions reported in Table 2 (the Charter auction is excluded due to a lack of reliable bond data). Panel (b) shows the same prices but excluding the Tribune auction, which has the largest degree of underpricing. The blue line shows the prices based all available bond issues. The green line shows prices based only on bond issues with the lowest price.

Figure 6: *Washington Mutual Bond Prices*

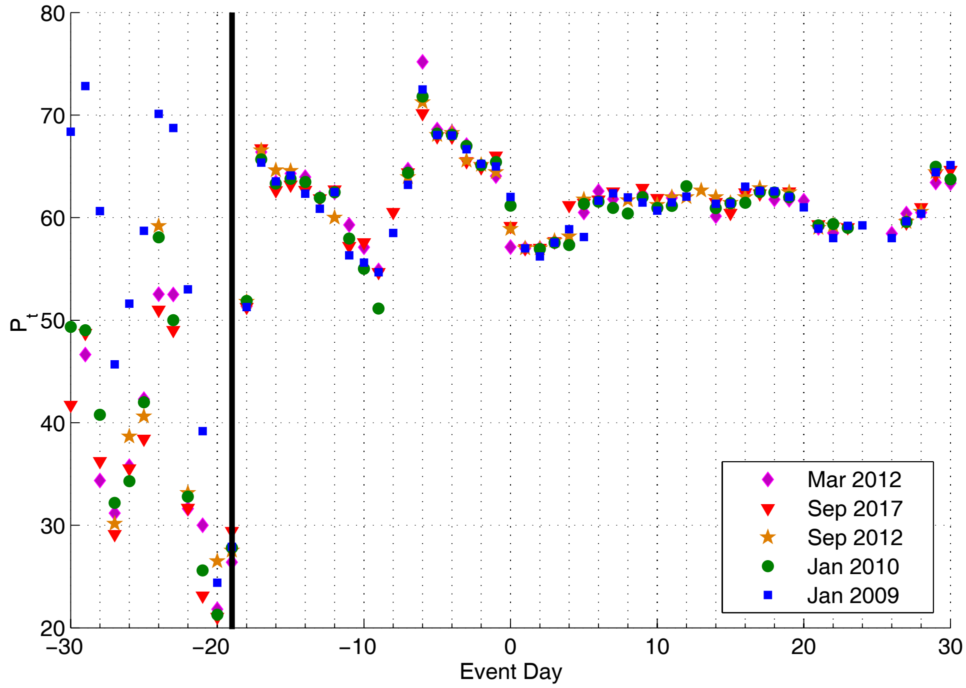



Figure 6 shows daily prices of Washington Mutual outstanding bond issues around the day of bankruptcy (indicated by a vertical black line). The legend shows the maturity date of each issue. The daily price at a given date is a volume-weighted average for all trades at this date. Further details on the construction of this graph are given in Section 4.1.

# The Effect of Algorithmic Trading on Liquidity in the Options Market

**Suchi Mishra**
Knight Ridder Associate Research Professor of Finance
Florida International University

**Robert T. Daigler**
Knight Ridder Research Professor of Finance
Florida International University

**Richard Holowczak**
Baruch College
City University of New York

Keywords: microstructure, algorithmic trading, liquidity, options

We thank Sasanka Vadlamudi for his computer assistance, without which this paper would not have been possible.

Suchi Mishra is Knight Ridder Research Associate Professor of Finance, Florida International University, Department of Finance RB208, Chapman Graduate School of Business, Miami FL 33199. Email: Mishras@fiu.edu. Phone: 305-348-4282.

Robert T. Daigler is Knight Ridder Research Professor of Finance, Florida International University, Department of Finance RB206, Chapman Graduate School of Business, Miami FL 33199. Email: Daiglerr@fiu.edu. Phone: 305-348-3325.

Richard Holowczak is Associate Professor of Finance and Director of the Subotnick Financial Services Center, Baruch College, CUNY, 1 Bernard Baruch Way, New York, NY. Email: Richard.Holowczak@baruch.cuny.edu. Phone: 646-312-1544

May 21, 2012

# The Effect of Algorithmic Trading on Liquidity in the Options Market


## Abstract

Algorithmic trading consistently reduces the bid-ask spread in options markets, regardless of firm size, option strike price, call or put option, or volatility in the markets. However, the effect on depth depends on the categorization of the data. The examination of the introduction of penny quotes provides a successful robustness test for the importance of algorithmic trading on liquidity. Overall, this study provides a controlled analysis of options with different levels of activity and different types of market participants across strikes/calls/puts/underlying stocks. Our findings also contribute to the extant literature on the characteristics of the liquidity of options markets during the growth period of algorithmic trading.


During the past several years the widespread development of automated order execution systems (algorithmic or algo trading) has transformed the financial markets. In particular, the promulgation of Order Protection Rule 611 under Regulation NMS in 2005 promoted the use of electronic trading and subsequently computerized algorithms. According to Rule 611, limit orders that are "immediately and automatically accessible" via an "Immediate or Cancel" (IOC) order have their prices protected from another trade execution at an inferior price. Consequently, Regulation NMS leveled the playing field across all U.S. exchanges regarding order executions.[1] These rule changes caused exchanges to compete based on trading fees, the speed of order handling, and the quality of execution in order to obtain a greater share of trading volume (Palmer, 2009). Because of the proliferation of electronic trading across all exchanges, the use of algorithms became indispensable for the trading process of institutions, market makers, and

[1] On April 6, 2005, the Securities and Exchange Commission adopted Regulation NMS, a series of initiatives designed to modernize and strengthen the national market system for equities. Regulation NMS was published in Securities Exchange Act Release No. 51808 (Jun. 9, 2005), 70 FR 37496 (Jun. 29, 2005) ("NMS Release"). These initiatives include: (1) Rule 610, which addresses the access to markets; (2) Rule 611, which provides inter-market price priority for displayed and accessible quotations; (3) Rule 612, which establishes minimum pricing increments; and (4) amendments to the joint-industry plans and rules governing the dissemination of market data. Rule 611, among other things, requires a trading center to establish, maintain, and enforce written policies and procedures reasonably designed to prevent "trade-throughs" – the execution of trades at prices inferior to protected quotations displayed by other trading centers. In order to be protected a quotation must be immediately and automatically accessible. (See Palmer (2009)).

professional traders. This resulted in algorithmic trading taking over the market making function for smaller size trades in the stock market due to its speed and cost advantages (see Hendershott and Moulton (2007)). More generally, Hendershott, Jones and Menkveld (2011) explain the use of algorithmic trading as follows:

> Algorithms are used to supply as well as to demand liquidity. For example, liquidity demanders use smart order routers (SORs) to decide the placement of a liquidity order, whereas liquidity suppliers such as hedge funds and broker-dealers use algorithms to supply liquidity. Overall, as trading became more electronic, it became easier and cheaper to replicate a floor trader's activity with a computer program doing algorithmic trading.

The growth of algorithmic trading has spurred interest in its potential effects on market dynamics (Hendershott and Riordan, 2009). In particular, such mechanized trading systems potentially could both reduce liquidity and increase volatility, particularly in times of market stress.[2] Two sides to the argument exist concerning the use of algorithmic trading. On the one hand, algos can increase competition and result in an increase in liquidity, thereby lowering the cost of immediacy. On the other hand, liquidity could decrease if algorithmic trades are used mainly to demand liquidity. For example, whereas limit order submitters supply liquidity by granting a trading option to others, liquidity demanders attempt to identify and pick-off beneficial trading opportunities by increasing the cost of submitting limit orders by causing spreads to widen. An example of liquidity demanders are algo traders executing large institutional blocks in short periods of time. Empirically, Hendershott et al. (2011) and Hendershott and Riordan (2009) find that the net effect of algo trading is to reduce bid-ask spreads and aid in the pricing efficiency in the stock market.

[2] The Flash Crash of May 6th, 2010 is an example of how algorithmic trading *may* have led to extreme volatility and the disappearance of liquidity. This potential liability of algorithmic trading has caused critics to support curbs to be placed on such trading. More recently, algorithmic trading also was criticized because of its "unfair" advantage over non-computerized traders, since algos possess a sub-second timing advantage in placing quotes and the related potential of front running of larger block orders. Here we concentrate on the effect of algorithmic trading on *options market* pricing for market scenarios other than the Flash Crash.

We extend the pioneering work of Hendershott et al. (2011) on the effects of algorithmic trading in the stock market to options. The importance of algorithmic trading for options on the demand side is found in the “Smart Routing” of options and the algorithmic execution of price improving multi-leg orders, as well as spread enhancing market-making activities across strikes, expirations, calls/puts, and on as many as seven options exchanges at once. Alternatively, the multitude of options challenges the ability of this market mechanism to generate liquidity for supply side activities. Supply side traders require execution of positions at current bid/ask prices such that the bid-ask spread widens and depth declines. Large supply side option orders challenge the ability of a potentially think market (such as options with many strikes, expirations, and exchanges) to consistently provide liquidity.

Preliminary evidence on the extent of algorithmic trading in the options markets is found in Figure 1, which shows the growth of OPRA message traffic from 2006 to 2008. Such activity is clearly visible in 2007 and increases in 2008. We examine the relation between algorithmic trading and liquidity by analyzing the bid-ask spread and the best bid-ask depth values for the Options Price Reporting Authority (OPRA) data feed for the flow of option messages as a proxy of algo trading. We differentiate between “call” and “put” options, and among “in-”, “near-” and “out-of-the-money” options, as well as providing separate results by market capitalization, volume, and volatility quintiles. Given the liquidity differences among the various options groupings, we have the advantage of analyzing the effect of algo trading on liquidity for a wide range of instrumental characteristics. These results provide more definitive conclusions than stocks concerning the ability of algo trading to supply liquidity effectively across a wide range of different characteristics (option strikes/expirations/calls-puts), thereby determining to what extent bid-ask spreads and depth responds to non-human intervention. Such results and conclusions are critical to regulators who make decisions concerning the benefit of algorithmic trading relative to the risk of liquidity disappearing during flash crashes.

We find broad evidence to support the benefits of algorithmic trading to reduce the bid-ask spread measure of liquidity, as well as providing an analysis of conflicting results for the depth of the market. We support our analysis with a robustness check by using the introduction of penny quotes as an exogenous event to support the liquidity impact of message traffic. Our findings also support the Cao and Wei (2009) results of the existence of a material liquidity factor in the options market. Moreover, our spread and depth analysis of the different strike categories ("in-", “near-” and “out-of-the-money”), as well as both calls and puts, supports the breadth of liquidity in options. We also find a differential impact of the underlying stock market capitalization and volatility, and the option characteristic of volume, on option bid-ask spreads and depth. Thus, we provide evidence on liquidity commonalities in the options market.

In conclusion, our results add to the developing literature on the liquidity of options, as well as more specifically substantiate the beneficial liquidity impacts of algo trading.[3] Consequently, potential regulatory restrictions on algorithmic trading should consider the benefits of such strategies on complex markets such as options, as well as the disadvantages of much slower human traders who enter the market for fundamental reasons separate from algo liquidity supply effects from market making and related strategies.

## I. Algorithmic Trading and Options

Our study contributes to two related strands of academic literature: The impact of algorithmic

[3] Microstructure research in options is complicated by the multitude of strike prices and expirations dates, the number of revisions in the bid-ask quotes, and the difficulty in obtaining data. Our findings add to the relatively thin literature on this direction as well as the even smaller subset of literature on option market liquidity (Vijh, 1990; Cao and Wei, 2009).

trading on the market environment and its impact on option market liquidity. The literature on the impact of algo trading in general is still at its infancy (Hasbrouck and Saar, 2010). In addition, the area of option market liquidity is a relatively nascent area compared to liquidity research on the equity and debt markets (Cao and Wei, 2009). The benefit of examining exchange-traded options is that it provides a natural laboratory for studying how trading mechanisms and the competitive structure of the industry affect market quality, given the large number of strike prices per underlying stocks and the relatively large number of exchanges trading options (Mayhew, 2002). Our paper ties a knot between these two fields by studying the impact of algo trading on option market liquidity.

The first area of algo research is the examination of the characteristics of algorithmic trading and algo trading strategies (especially the effect of the speed of transmission on trading strategies). Riordan and Storkenmaier (2008), Easley, Hendershott, and Ramadorai (2009), and Hasbrouck and Saar (2010) examine the effect of the speed of order transmission and algo strategies. For example, Riordan and Storkenmaier state that algo traders increase liquidity by reducing latency in order transmission from 50 ms to 10 ms, thereby reducing trading costs by 1 to 4 basis points.

The second area of research is the impact of algo trading on the market environment, such as information dissemination and the liquidity variables of bid-ask spread and depth. Hendershott and Riordan (2009), Brogaard (2010), Karagozoglu (2011), and Hendershott, Jones, and Menkveld (2011) are the primarily sources dealing with the impact of algo trading on market quality factors such as price discovery and liquidity. More specifically, Hendershott and Riordan examine the 30 DAX stocks, finding that algorithmic trades create a larger price impact than non-algorithmic trades and therefore tend to contribute more to price discovery. Brogaard investigates the impact of algo trading on equity market quality by using a dataset of 26 high-frequency traders in 120 stocks. He reports that high-frequency traders contribute to the liquidity provision in the market, that their trades improve price discovery more than trades of other market participants, and that their activity appears to lower volatility. Karagozoglu examines algorithmic trading in relation to futures markets, finding that spreads are decreased and market depth is increased in five different futures contracts. The only related liquidity study using options to examine market quality is Cao and Wei (2009), who show the existence of a material common liquidity factor in the options market, although they do not relate this common factor to algo trading; thus, option liquidity does have a factor that flows across the strike prices and calls and puts of an option series.[4]

Hendershott, Jones, and Menkveld (2011) is the most related research to this paper and forms the basis of the experimental design for our study. Hendershott, et al. uses a measure of NYSE message traffic as a proxy for algo trading to study its impact on the liquidity of stocks, without differentiating among the various strategies used by algo traders. They also include an event study approach around the introduction of autoquoting as an exogenous instrument to examine the effect of algorithmic trading. The authors document that an increase in the number of algorithmic trading messages affect the liquidity of only the largest stocks. For these stocks, liquidity improved in terms of a decline in the quoted and effective spreads, although quoted depth decreased. The use of the autoquoting period confirms the key results of their paper.

[4] Regarding the general research on options not directly related to algo trading, Biais and Hillion (1991) and John, Koticha and Subrahmanyam (1991) develop models that examine the *equilibrium* bid and ask prices for individual equity options markets. Ho and Macris (1984) analyze the transaction price and bid-ask spread relation for AMEX individual equity options; George and Longstaff (1993) determine the cross-sectional differences among individual equity options for different strikes; Mayhew (2002) examines the effects of competition and market structure using individual equity option bid-ask spreads; and Cai, Hudson, and Keasey (2004) examine equities on the London Stock Exchange (LSE) and find a L-shape in the bid-ask spread, a two-humped shape for volume, and a U-shape for volatility.

## II. Data

Options microstructure research provides several challenges related to data structure. First, the number of strike prices and expiration dates multiplies the number of data series, with the different strikes/expirations possessing differing price response characteristics. Second, the number of quote revisions (algo messages) has geometrically increased over the past few years, creating data analysis and storage issues. Finally, data availability for all quotations for all stock options for all exchanges is limited. Thus, unlike organized microstructure data for the equity markets, there is dearth of comprehensive microstructure research for exchange-traded options.

The data for this study employs the Options Price Reporting Authority (OPRA) data feed. The OPRA feed consists of trade execution and the best bid and offer quotes and size from *each* of the seven U.S. equity options exchanges. OPRA flags each quote with an indicator stating the quote's bid-ask relative to the national best bid and offer (NBBO). We employ the Baruch Options Data Warehouse database of options, which processes the full OPRA feed and generates data extracts and statistics on trade and quote messages.

This paper uses data for calendar years 2007 and 2008, representing 2,328,185 unique options series on 5,100 underlying equities, ETFs and indexes. The two years of data contain 311,567,675 trades and approximately 1.3 trillion quotes, requiring 65 terabytes of disk storage. We focus on 2007 and 2008 because algorithmic trading in options markets increased starting in 2007 (as shown in Figure 1) and because 2008 provides a unique opportunity to examine how volatility affects both the spread and depth of options markets, especially in terms of the relation between algorithmic trading and the financial crisis. In addition, our research design and time interval includes the introduction of penny quotes for options markets, which was initiated in 2007.

We compute the quoted spread for each option series for each stock employed in this study by determining the average National Best Bid and Offer (NBBO) bid-ask spread over the entire trading day for each day in both years, as well as the total dollar value for each options series traded. In this process we employ the traditional filters for spreads and depth. For example, we ignore negative spreads and stub quotes (a quote with a zero bid and a very large ask, such as 199,999).[5] The data on market capitalization, and equity returns for the calculation of the daily volatility, are obtained from COMPUSTAT and CRSP.

## III. Liquidity Measures and Methodology

*A. Liquidity, Algo Trading, and Control Measures*

Our goal is to examine the relation between algorithmic trading and the liquidity of the associated options market by using the number of messages as the measure of algorithmic trading in the market.[6] Algorithmic trading is variously reported to account for 50% to 70% of the total volume in today's equity market, implying that both the amount and changes in algo trading messages dominate the number of messages in a market.

We examine the relation between message traffic and both the bid-ask spread and depth measures of liquidity in cross sectional panel regressions, where controls are established for the underlying firm size, volatility of the underlying stock, and the dollar volume of option trading. We examine panel regressions employing every intraday bid-ask quote and depth observation

[5] Only "eligible" quotes are employed. An eligible quote is a NBBO quote representing a firm (i.e. "executable") quote that is neither a stub quote nor not a zero price bid quote; quotes with zero size bids or offers are also ignored. All stub quotes are removed from the database, which includes initial opening and closing stub quotes, as well as "non-firm" quotes at the start of the day. The messages include both quotes and trades; however, more than 99.95% of the option messages are quotes. Therefore, for options, messages and bid-ask quotes are effectively equivalent.

[6] Hendershott et al. suggest either a measure of message traffic normalized by volume, or the use of raw message traffic to represent algorithmic trading. We employ raw message traffic; however, we do control for the volume of trading in the regression analysis. The results are unchanged when message traffic normalized by volume of trading is employed.

and accumulate this data into daily algo messages and daily average bid-ask spread and depth data. The volume and volatility control variables are total values for the day. Separate values for the spread and depth are calculated for each option strike, expiration, and call/put for each underlying stock. The percentage spread is calculated as follows:

$$Percentage\ Spread = \left( \frac{(Ask - Bid)}{0.5(Ask + Bid)} \right)(100) \tag{1}$$

where bid and ask prices are the NBBO values.

Depth is calculated as:

$$Depth = \left( \frac{(BestAskSize + BestBidSize)}{2} \right) \tag{2}$$

We sort the options based on three different criteria: (1) by market capitalization of the underlying instruments (stocks and ETFs, generally referred to generically as "stocks"); (2) by dollar volume traded for the options over the entire year; and (3) by volatility of the underlying stocks. We sort the options based on the market capitalization of the underlying stocks into quintiles in descending order, choosing the largest forty stocks from each group. Therefore, we examine the option series data for 200 underlying equities of stratified capitalizations. As noted, we also sort the option series by the respective option trading volume generated by all of the exchanges for the entire year, as well as sorting independently by the volatility of the underlying stock, again in descending order.

We employ the Garman-Klass (1980) measure to calculate the daily stock volatility, as defined by:

$$Var(GK) = \frac{1}{2}\left[\ln(High) - \ln(Low)\right]^2 - \left[2\ln(2) - 1\right]\left[\ln(Open) - \ln(Close)\right]^2 \tag{3}$$

The Garman-Klass measure allows for an examination of volatility within an interval as opposed to the traditional volatility measures that examine volatility between or across intervals. As noted by Garman and Klass, their measure is eight times more efficient than using a close-to-close measure of volatility.[7]

For each sort the first quintile represents stocks with the highest values for that variable, whereas quintile five represents stocks with the lowest quintile values for that variable. For each sort we classify the option series into "call" and "put" options, and further into "in-", "near-" and "out-of-the-money" options. The "in-", "near-" and "out-of the money" option groups are created by employing the following procedure: First, we calculate the difference between the stock price of the last trade and the strike price, labeled the "stock-strike difference." The option is grouped as a near-the-money option if the stock-strike difference is within 2.5 (5) points for stocks below (above) $20. It is grouped as an out-of-the-money call option if the stock-strike difference is -2.5 to -10 (-5 to -20) for stocks below (above) $20, and an in-the-money call option if the difference is 2.5 to 10 (5 to 20) for stocks below (above) $20. Signs are reversed for put options. Options outside these ranges possess little trading interest and therefore are removed from the analysis.

We call the above sample the general sample (or non-penny quote sample), since we remove the stocks with penny quote options from the sample in order to provide inferences on the impact of message traffic (algorithmic trading) independent of the effects of the penny pilot on option market activity.[8]

*B. Panel Regressions*

For the general sample we estimate the following OLS regressions for each category as follows:

---

[7] Efficiency in this context refers to the reduction in the error of the estimate.

[8] The penny pilot option project and its importance are described in the next sub-section.

$$l_{it} = \alpha_i + \beta_i A_{it} + \delta_{it} X_{it} + \epsilon_{it} \quad (4)$$

where $l_{it}$ is the liquidity variable (either the bid-ask spread or the depth), $A_{it}$ is the message traffic representing algorithmic trading, and $X_{it}$ is the set of control variables, i.e. market capitalization, the Garman-Klass volatility of the underlying stock, and the dollar trading volume of the stock's options.

We conduct our tests of option algorithmic trading in two phases. In the first phase we examine the relation between algorithmic trading and liquidity by examining the bid-ask spread and the depth of the market for the non-penny quote (general sample) options. For this step we filter the non-penny quote options so as to provide inferences of message traffic (algorithmic trading) and market quality on option market activity, independent of the consequences of moving from the five/ten cent quotes to penny quotes. In the second phase we design a model for a robustness check (and to establish causality) by picking the introduction of penny quotes in 2007 and 2008 to option series affected by the penny quotes as an exogenous factor that could potentially increase the incidence of algorithmic trading. In fact, the reason to change to penny quotes for stock options was not to benefit algorithmic trading. However, a smaller tick size theoretically should create more quote changes using the penny quote procedure, especially for the more active stock options (American Stock Exchange, 2007; Louton, Saraoglu, and Holowczak, 2009). Moreover, more frequent quotes provide critical new information concerning the fair price of an option to algorithms. Thus, the immediate feedback traders receive from penny quotes should increase algorithmic trading activity, which is especially crucial to options given their extensive number of strikes and expiration dates.

*C. The Penny Pilot as a Robustness Check*

Our approach to *verifying* the relevance of algorithmic trading is to explore the relation between message traffic and option market liquidity by using stocks with option penny quotes. The penny quote sample period starts one month before the penny quote initiation date and ends one month after the penny quote initiation. Note that the transition to option penny quotes occurred in three phases during this time period; we examine each phase independently.[7]

We estimate the following regressions for the sample with penny quotes:

$$l_{it} = \alpha_i + \gamma_t + \beta_i A_{it} + \delta_{it} X_{it} + \epsilon_{it} \tag{5}$$

where $l_{it}$ is the liquidity variable (either bid-ask spread or depth), $A_{it}$ is the message traffic representing algorithmic trading, and $X_{it}$ is the set of control variables such as market capitalization, Garman-Klass volatility of the underlying stock, and the trading volume of the option. Equation (5) includes the additional variable $\gamma_t$ to represent the time dummy for before and after the penny quotes were introduced. Since our principal goal in this analysis is to understand the effects of algorithmic liquidity supply on market quality, we employ the penny-quote dummy ($\gamma_t$) as an instrument for algorithmic trading in the panel regression framework. By including time dummies in the panel specification, we can employ non-penny quoted stocks as controls, comparing the penny-quoted stocks to the not-yet-penny-quoted stocks. The percentage spread and depth used in the penny quote analysis is measured in the same manner as with the total sample. The penny quote regression model is calculated using the GMM (Generalized Method of Moments) procedure.

## III. Results

*A. Basic Statistics*

Tables 1 and 2 show the basic call and put statistics, respectively, by option category for each quintile for the spread, depth, and algorithmic messages, as well as for the control variables of market capitalization, Garman-Klass volatility, and dollar option volume. The average quoted

[7] We separate the general sample and the penny quote sample. This separation provides the opportunity to interpret the results and present inferences for each sample independently. We also examine an integrated sample (not shown here), finding that the results were not significantly different than the general and penny quote samples.

spread as a percentage of the option price is smallest for the in-the-money options, next largest for the near-the-money options, and largest for the out-of-the-money options. This is logical given the size of the prices for the in-, near-, and out-of-the-money option categories. An important characteristic of the option series is that the spreads are almost always higher for the 2008 relative to 2007, with larger differences and spreads occurring for the smaller stocks (lager numbered quintiles). Moreover, the increase in the spread is larger for the "in-" and "out-of-the-money" groups than for the "near-the-moneys."

The depth in Tables 1 and 2 is substantially higher for the first quintile of stocks, which is associated with institutional interest in these options. The depth is much smaller for the other quintiles. Moreover, the near-the-money options possess the largest depth for quintiles one to three. The most striking depth results are for 2008, where the depth for quintile one is typically less than half of the depth existing in 2007; however, the depth for the other quintile is often *larger* in 2008 than in 2007. This result indicates the extent of evaporating liquidity in the options market for the largest stocks due to the financial crisis and increase in algorithmic trading.

The number of algorithmic messages is substantially higher for the first quintile, which is consistent with the underlying stocks for this quintile being the largest and potentially most active stocks. Moreover, the number of algo messages increase significantly from 2007 to 2008, especially for puts and quintile 1,with quintile 5 being the lone exception. In terms of the control variables, the Garman-Klass volatility for 2008 increases by a factor of six for the first quintile and by a factor of 2.3 for quintile five. The market capitalization and dollar volume variables remained relatively stable over the two year period for most categories.

B. *Spread Results*

This section examines the bid-ask spread results for the general sample for 2007 and 2008. Our goal is to understand the effects of algorithmic trading on the liquidity of options. Tables 3 and 4 provide the quintile spread results sorted in terms of each of the control variables in quintile descending order for 2007 and 2008, respectively. Table 3 shows that the standardized spread decreases with increasing message traffic for all categories and both years for the market capitalization sort.[9] The number of messages is larger for the larger capitalization firms (e.g. quintile 1), therefore the coefficients are smaller in these cases. More importantly, the statistical significance of the message traffic variable almost always is larger for the larger firms such as quintile one, showing that the consistency and reliability of the results is stronger for quintile one. Moreover, the decrease in the spread is significant for all quintiles and option categories. The volume ranking by quintiles shows the same decrease in spreads and decline in significance on the market capitalization results, although quintiles 4 and 5 often are not significant. The volume quintile results are consistent since the largest capitalized companies often possess the largest options dollar volume.

Tables 3 and 4 also show that the spread declines with message traffic for the volatility sorted groups. However, the significance level of the spread decrease for these results is consistently the greatest for the *lowest* volatility group (i.e. quintile five). This result is intuitive since the *highest* volatility group (quintile 1) should include active options for the more volatile smaller cap stocks in this group which would be more diverse in their response to algo trading as well as be less liquid, whereas the lowest volatility group (quintile 5) would include larger capitalized firms; thus, the largest significance for the spread decrease for the volatility grouping

---

[9] We also examine the spread and depth results after removing the data for the financial crisis time period in 2008. We follow Anand, Puckett, Irvine, and Venkataraman (2011) to determine the crisis time period. The results for the crisis period in 2008 are essentially equivalent to the entire 2008 year, and are available upon request.

is in quintile five whereas the most significant results for the market capitalization and volume sorted results discussed above are in quintile one.

*C. Depth Results*

Depth as a measure of liquidity has received minimal attention in the literature. In particular, in relation to algo trading only Hendershott et al. (2011) examines the roll of depth, finding that depth actually *declines* as algo trading increases. Thus, this measure of liquidity may actually be reduced due to the frequent quote revisions associated with algo trading. The reasoning by Hendershott et al. is based on the smaller trade size created by certain strategies for algo trading, although the evidence is anecdotal.

Tables 5 and 6 show that our analysis of depth for options typically *increases* as algo trading increases, especially for the market capitalization and volatility groupings, contrary to the market capitalization results of Hendershott et al.. Unlike the spread results, there is no pattern in the size of the significance values across quintiles or option categories. For the volume grouping the results are mixed, both in terms of the sign and whether the quintiles are significant, although the quintile one results often are most significant. Overall, there is no conclusive pattern for the depth variable using the volume sorted quintiles. These results can be due to algorithmic trading orders being sliced into smaller orders and executed in batches rather than being executed as large volume orders.

*D. The Penny Pilot as an Exogenous Event*

We next examine the penny pilot quotation for options as an exogenous factor that could potentially increase the incidence of algorithmic trading. The penny pilot program for options was a Securities and Exchange Commission (SEC) initiative to quote stock options with the most activity in terms of pennies rather than nickels/dimes in order to decrease price spreads, provide better prices to retail customers, and reduce the payment for order flow. For our purposes, the introduction of penny quotes for the options market during 2007 and 2008 provides an opportunity to examine the effects of an exogenous factor. In fact, although algorithmic trading is not the intentioned beneficiary of the penny pilot program, by design it promotes the practice of algorithmic trading. Thus, a smaller tick size caused by penny quotes should create more quote changes, especially for the more active stock options (American Stock Exchange, 2007). Moreover, more frequent quotes provide critical new information concerning the fair price of an option. Thus, the immediate feedback that traders receive from penny quotes is consistent with an increase in algorithmic trading activity, which is especially crucial for option trading because of the complexity of their strike/expiration/multiple exchange structure.

Table 7 presents the basic statistics for the penny quote sample. The penny quote sample we employ possesses basic characteristics that are almost equivalent to the first volume sorted quintile in the general sample; the explanation for this similarity is that the stocks used for the penny pilot in 2007/2008 are large capitalization stocks that possess very actively traded options. As with the general sample, the average spread as a percentage of the price for the penny stocks is smallest for the in-the-money options, next largest for the near-the-money options, and largest for the out-of-the-money options. The depth is typically largest for the near-the-money options for calls and the out-of-the-money options for puts. The depth is consistent with the results for the first quintile of the market capitalization ranking for the general option sample. This is consistent with both the penny quote sample and the first quintile from the general sample being dominated by underlying stocks that are of interest to institutions. Also, the number of algo messages is substantially higher for the near-the-money group, since near-the-moneys are the most active category.

Phase I of the penny pilot program (PPP) was adopted by six options exchanges on January 26th of 2007 and included 13 securities; Phase II of the PPP began on September 28, 2007 and included 22 securities; and Phase III began on March 28, 2008 and covered 28 securities. Our underlying general sample in 2007 and 2008 excludes these penny pilot securities. With the introduction of the penny pilot there were two major changes that could confound our results. First, penny quotes should increase the slicing and dicing of orders, since smaller-sized orders can be placed at better prices. Second, there could be more effective market making due to the existence of algo traders and their speed and number advantages. However, there could be less depth in the market due to less clustering of orders around the NBBO because of such slicing and dicing of algorithmic orders.

For each phase we examine one month before and one month after the penny quote is introduced. Thus, we generate daily panel regressions according to the specification in equation (2). Tables 8 presents the spread results for all three phases. In the penny quote sample the results for the “in-”, “near-” and “out-of-the-money” categories show interesting differences. the bid-ask spread declines with message traffic, consistent with the general sample, for all 12 regressions (three penny pilot phases) for the near- and out-of-the-money option groups, but is not significant for the in-the-money regressions.

Table 9 presents the depth results for the penny quote sample. Except for one case the depth significantly decreases. These results contradict the depth results for the market capitalization and volatility groupings for the general sample.[10] However, unlike the general sample, the decline in depth for the penny pilot sample is a natural consequence of the introduction of smaller penny quotes with more frequent quote revisions, as well as due to an

[10] Of course, the in-the-money results can be related to thin trading.

increase in the slicing of orders into smaller sizes.[11]

These penny quote results are consistent with the results for the volume ranking of the first quintile of the general sample. In addition, the penny quote sample results for both the spread and depth for the options are similar to those of Hendershott et al. (2011) for equities. In general, the penny quote stocks are the most active and/or liquid stocks in the market, and therefore the increase in message traffic means a smaller spread due to a greater liquidity supply because of a larger number of (algorithmic) market makers, and a decreased depth due to the slicing and dicing of orders. Therefore, with increased message traffic, both the trading cost and the depth is reduced.

*E. Discussion of Results*

Overall, our results differ from Hendershott et al. (2011) in obvious ways, especially in terms of the signs on the depth variable. These differences stem from the fact that our study examines stock options, whereas Hendershott et al. analyzes stocks. For example, Hendershott et al.'s discussion focuses o the most liquid stocks (quintile 1) of the market capitalization group, whereas our larger market capitalization group does not necessarily employ the most active options. In fact, the volatility of the underlying stocks is a predominant motivation for trading options, with the volatility grouping showing a positive increase in depth for both 2007 and 2008. The volume of option trading is the most transparent method of determining the most active options. In fact, the volume sorting sample results closely mirror the Hendershott et al. results for both spreads and depth, i.e. the spread and depth these variables typically decline with higher message traffic, with this relation existing with less significance as the comparison changes from quintile one to quintile five. Finally, note that the decline in depth is consistent with the slicing and dicing of orders from buy side algorithms, as well as by the competition on

[11] See the "Penny Quoting Pilot Program Report" by the American Stock Exchange (2007).

the algorithmic liquidity supply side, which potentially can lead to a smaller size offered by each market maker at the best bid and offer. Moreover, for the penny quote sample the results mirror quintile one of Hendershott, et al.

## IV. Conclusions

Empirical research on the market impact of algorithmic trading is important for both policy makers and market participants because of the potential impact of algo trading on the bid-ask spread and the depth of the market. Previous research examines the impact of algo trading on the stock and futures market. We extend this research on a market with various levels of trading activity due to different stocks, a range of strike prices, different expiration dates, and a multitude of exchanges. These factors make the application of algorithmic trading more difficult, as well as more useful. We employ the Options Price Reporting Authority (OPRA) data feed, using the flow of messages as a proxy of algo trading. Thus, our results offer evidence on the liquidity impacts of algorithmic trading in the options market. In addition, we employ the introduction of penny quotes in option markets as an exogenous event to test the liquidity impact of message traffic.

Given the liquidity differences among the various groups of options, we have the advantage of examining the effect of algorithmic trading on liquidity in a more in-depth context. Our analysis of the general sample for 2007 and 2008, and sorting them by the characteristics of the underlying stock (by market capitalization and volatility) as well as by dollar option volume, provides evidence that supports Hendershott et al. (2011). Moreover, we provide an explanation as to why a reduction in depth with algorithmic trading can exist, as with our penny quote sample and the results found in Hendershott et al.

The issue of liquidity in financial markets is a timely and crucial factor. Additional analysis of more complicated and integrated markets such as options would provide crucial information to aid appropriate regulatory interests in making the markets "fair and efficient." Moreover, further investigation of the impacts of algorithmic traders on the markets is essential in determining the tradeoffs between the additional liquidity algo traders provide in normal markets versus the potential for market crashes when algo traders remove their liquidity, as happened for the Flash Crash.

**REFERENCES:**


Anand A., A. Puckett, P. J. Irvine, K. Venkataraman, 2011. “Performance of Institutional Trading Desks: An Analysis of Persistence in Trading Cost.” *Review of Financial Studies,* forthcoming.

Brogaard, J. A., 2010. “High Frequency Trading and its Impact on Market Quality.” Working paper, Northwestern University.

Biais, B., & P. Hillion, 1991. “Option Prices, Insider Trading, and Interdealer Competition.” CEPR financial markets working paper, INSEAD.

Cai, C. X., R. Hudson, and K. Keasy, 2004. “Intraday Bid-Ask Spreads, Trading Volume and Volatility: Recent Empirical Evidence from the London Stock Exchange.” *Journal of Business Finance and Accounting*, 31, 647-676.

Easley, D., T. Hendershott, and T. Ramadorai, 2009. “Leveling the Trading Field.” Working paper, Cornell University.

Garman, M. B., M. J. Klass, 1980. “On the Estimation of Security Price Volatility from Historical Data.” *Journal of Business*, 53, 67-78.

George, T., and F. Longstaff, 1993. “Bid-Ask Spreads and Trading Activity in the S&P 100 Index Options Market.” *Journal of Financial and Quantitative Analysis*, 28, 381-397.

Hasbrouck, J., and G. Saar, 2010. “Low-Latency Trading.” NYU Working paper.

Hendershott, T., C. M. Jones, A. J. Menkveld, 2011. “Does Algorithmic Trading Improve Liquidity?” *Journal of Finance*, 66, 1-33.

Hendershott, T., and P. Moulton, 2007. “The Shrinking New York Stock Exchange Floor and the Hybrid Market.” Manuscript, UC Berkeley.

Hendershott, T., and R. Riordan, 2009. “Algorithmic Trading and Information.” Manuscript, UC Berkeley.

Ho, T., and R. G. Macris, 1984. “Dealer Bid-Ask Quotes and Transaction Prices: An Empirical Study of some AMEX Options.” *Journal of Finance*, 39, 23-45.

John, K., A. Koticha, and M. Subrahmanyam, 1994. “The Microstructure of Options Markets: Informed Trading, Liquidity, Volatility and Efficiency.” Working paper, New York University.

Karagozoglu, A. (2011). “Direct Market Access in Exchange-Traded Derivatives: Effects of Algorithmic Trading on Liquidity in Futures Markets.” *Review of Futures Markets*, 19 Special Issue, 95-142.

Kohen, A., 2008. “Options Algorithms and Algorithmic Trading of Options.” *Journal of Trading*, Fall: 36-38.

Louton, D., Saraoglu, H. and Holowczak, R., 2009. "Quote Behavior and Liquidity Implications of the Options Penny Pilot project: A Comparative Study”. *Proceedings of the Eastern Finance Association Annual Meetings.* Washington, D.C.

Mayhew, S., 2002. “Competition, Market Structure and Bid-Ask Spreads in Stock Options Markets.” *Journal of Finance*, 57(2), 931-958.

Melanie, C., and J. Wei, 2010. “Option Market Liquidity: Commonality and Other Characteristics.” *Journal of Financial Markets*, 13, 20-48.

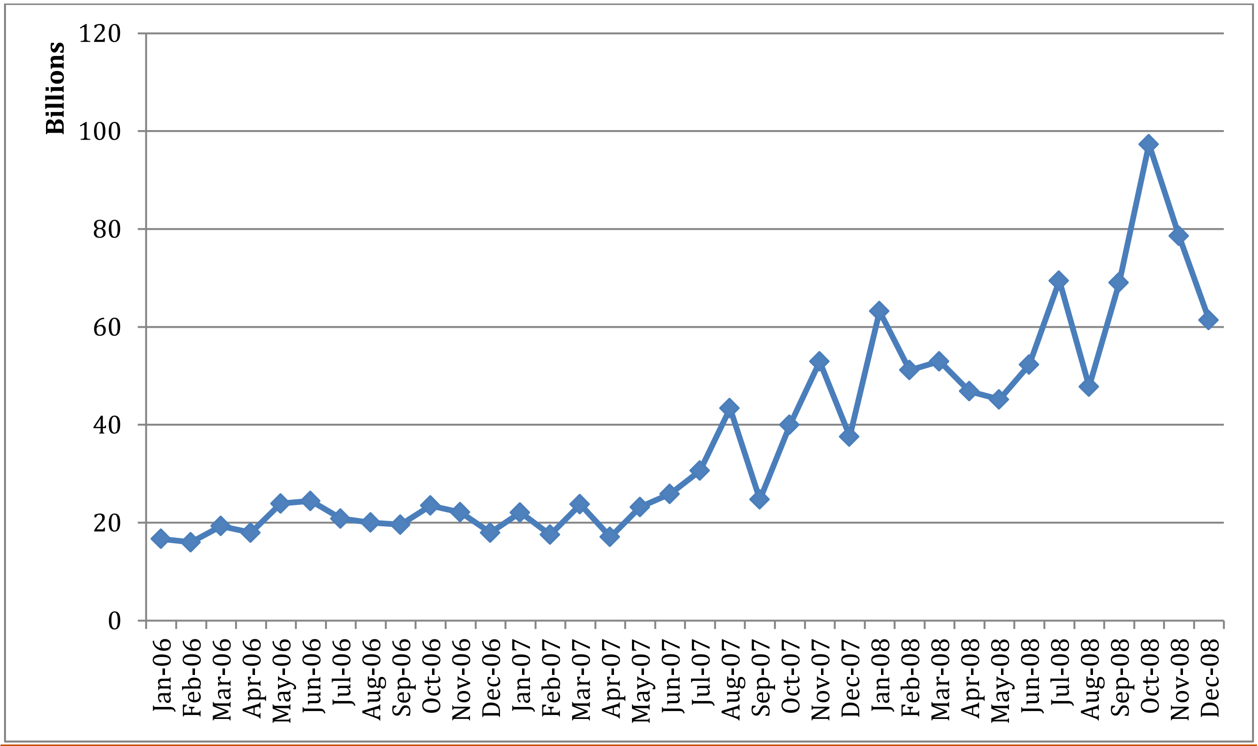
Palmer, M., 2009. “Algorithmic Trading: A Primer.” *Journal of Trading*, Summer (2009): 30-39.

Riordan, R., and A. Storkenmaier, 2008. “Optical Illusions: The Effects of Exchange System Latency on Liquidity.” Working paper, University of Karlsruhe, Germany.

Vijh, M. A., 1990. “Liquidity of the CBOE Equity Options.” *Journal of Finance*, 45, 1157-59.

Figure 1

OPRA Message Traffic per Month in Billions of Messages




This figure examines the growth in option messages for before and during the study period.

## Table 1: Summary Statistics for Calls

| | | CALLS 2007 | | | | | CALLS 2008 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Q1 | Q2 | Q3 | Q4 | Q5 | Q1 | Q2 | Q3 | Q4 | Q5 |
| Quoted Spread | | | | | | | | | | | |
| | in | 0.0240 | 0.0520 | 0.0730 | 0.1040 | 0.1720 | 0.0390 | 0.0930 | 0.1240 | 0.2180 | 0.2200 |
| | near | 0.0860 | 0.2060 | 0.2840 | 0.5090 | 0.5740 | 0.0760 | 0.2240 | 0.3480 | 0.4990 | 0.6990 |
| | out | 0.2760 | 0.4060 | 0.6150 | 0.8140 | 0.7920 | 0.3180 | 0.4880 | 0.7480 | 1.0010 | 1.2810 |
| Quoted Depth | | | | | | | | | | | |
| | in | 665 | 77 | 60 | 46 | 19 | 387 | 74 | 55 | 41 | 50 |
| | near | 1,768 | 145 | 74 | 44 | 21 | 764 | 182 | 75 | 46 | 43 |
| | out | 1,160 | 125 | 57 | 34 | 22 | 701 | 103 | 58 | 39 | 30 |
| Messages | | | | | | | | | | | |
| | in | 32,713 | 8,025 | 4,064 | 1,653 | 1,494 | 42,865 | 14,107 | 4,866 | 1,906 | 909 |
| | near | 41,861 | 7,331 | 3,434 | 1,573 | 1,206 | 74,134 | 11,998 | 4,542 | 1,992 | 1,098 |
| | out | 26,916 | 4,140 | 1,776 | 1,018 | 1,550 | 46,271 | 7,811 | 2,973 | 1,386 | 873 |
| GK Volatility | | | | | | | | | | | |
| | in | 7.01947 | 11.0200 | 11.9430 | 19.8429 | 33.9894 | 42.1542 | 37.9620 | 43.7990 | 51.9070 | 73.5807 |
| | near | 7.01947 | 11.0200 | 11.9430 | 19.8429 | 33.9894 | 42.1542 | 37.9620 | 43.7990 | 51.9070 | 73.5807 |
| | out | 7.01947 | 11.0200 | 11.9430 | 19.8429 | 33.9894 | 42.1542 | 37.9620 | 43.7990 | 51.9070 | 73.5807 |
| Market Cap | | | | | | | | | | | |
| | in | 17.0658 | 14.9900 | 14.3680 | 13.5043 | 13.2806 | 16.8769 | 14.9660 | 13.9270 | 13.1640 | 12.6521 |
| | near | 17.0658 | 14.9900 | 14.3680 | 13.5043 | 13.2806 | 16.8769 | 14.9660 | 13.9270 | 13.1640 | 12.6521 |
| | out | 17.0658 | 14.9900 | 14.3680 | 13.5043 | 13.2806 | 16.8769 | 14.9660 | 13.9270 | 13.1640 | 12.6521 |
| Volume | | | | | | | | | | | |
| | in | 1238.0480 | 299.2944 | 234.4949 | 282.4840 | 61.1345 | 632.8810 | 399.7656 | 235.3451 | 81.7114 | 181.8925 |
| | near | 2831.8604 | 240.7275 | 109.7671 | 63.6481 | 45.7641 | 2257.5096 | 230.8528 | 72.1126 | 42.8126 | 26.0978 |
| | out | 963.9726 | 78.8107 | 47.1090 | 31.2650 | 217.3550 | 708.951 | 95.1252 | 29.6912 | 16.3591 | 13.7386 |

Based on the option code we divide the data into call and put options and then into in-, near-, and out-of-the-money strikes. The table provides daily averages for each variable for the call options for the general sample for 2007 and 2008. We group/rank the options by the underlying's (equity's) market capitalization. For each quintile we then provide averages for the quoted spread, quoted depth, number of messages, Garman-Klass volatility, market capitalization and dollar option volume by each equity subgroup and for calls and puts and "in-", "near-" and "out-of-the-money" options. The values for the market capitalization and volatility variables are equivalent for the in-, near-, and out-of-the-money categories since they are based on the underlying stocks. Dollar option volume is the average per strike price for each stock in the category and then divided by 100 (the strikes include those without a trade but with a quote).

## Table 2: Summary Statistics for Puts

| | | PUTS 2007 | | | | | PUTS 2008 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Q1 | Q2 | Q3 | Q4 | Q5 | Q1 | Q2 | Q3 | Q4 | Q5 |
| Quoted Spread | | | | | | | | | | | |
| | in | 0.0313 | 0.0677 | 0.1295 | 0.2514 | 0.2773 | 0.0371 | 0.0916 | 0.1226 | 0.1816 | 0.2768 |
| | near | 0.0814 | 0.2022 | 0.2712 | 0.4329 | 0.5295 | 0.0632 | 0.1928 | 0.2921 | 0.4543 | 0.5843 |
| | out | 0.2445 | 0.5072 | 0.6026 | 1.0273 | 0.7164 | 0.2089 | 0.3871 | 0.5737 | 0.8737 | 1.3039 |
| Quoted Depth | | | | | | | | | | | |
| | in | 619 | 91 | 59 | 39 | 27 | 426 | 92 | 56 | 45 | 41 |
| | near | 1,975 | 151 | 72 | 42 | 26 | 843 | 188 | 77 | 47 | 40 |
| | out | 1,706 | 121 | 51 | 35 | 17 | 711 | 102 | 56 | 38 | 32 |
| Messages | | | | | | | | | | | |
| | in | 36,700 | 8,187 | 3,973 | 1,411 | 905 | 47,843 | 13,609 | 4,771 | 2,444 | 1,335 |
| | near | 40,163 | 7,027 | 3,549 | 1,499 | 1,285 | 69,149 | 11,585 | 4,620 | 2,094 | 1,256 |
| | out | 25,143 | 3,537 | 1,718 | 1,872 | 954 | 41,296 | 7,107 | 3,009 | 1,551 | 1,169 |
| GK Volatility | | | | | | | | | | | |
| | in | 7.01947 | 11.024 | 11.943 | 19.843 | 33.989 | 42.1542 | 37.96192 | 43.799 | 51.9068 | 73.5807 |
| | near | 7.01947 | 11.024 | 11.943 | 19.843 | 33.989 | 42.1542 | 37.96192 | 43.799 | 51.9068 | 73.5807 |
| | out | 7.01947 | 11.024 | 11.943 | 19.843 | 33.989 | 42.1542 | 37.96192 | 43.799 | 51.9068 | 73.5807 |
| Market Cap | | | | | | | | | | | |
| | in | 17.0658 | 14.99 | 14.368 | 13.504 | 13.281 | 16.8769 | 14.96557 | 13.927 | 13.1639 | 12.6521 |
| | near | 17.0658 | 14.99 | 14.368 | 13.504 | 13.281 | 16.8769 | 14.96557 | 13.927 | 13.1639 | 12.6521 |
| | out | 17.0658 | 14.99 | 14.368 | 13.504 | 13.281 | 16.8769 | 14.96557 | 13.927 | 13.1639 | 12.6521 |
| Volume | | | | | | | | | | | |
| | in | 2101.8805 | 368.0625 | 373.0354 | 91.5764 | 108.5181 | 897.4038 | 462.8728 | 254.1447 | 122.1596 | 72.5502 |
| | near | 3529.1479 | 194.3081 | 99.5504 | 42.5228 | 87.6818 | 2831.0195 | 227.7973 | 85.5088 | 53.0268 | 30.8512 |
| | out | 1011.6183 | 88.3165 | 39.2363 | 19.9890 | 16.2646 | 944.0373 | 97.2319 | 38.8053 | 26.3321 | 23.4810 |

Based on the option code we divide the data into call and put options and then into in-, near-, and out-of-the-money strikes. The table provides daily averages for each variable for the put options for the general sample for 2007 and 2008. We group/rank the options by the underlying's (equity's) market capitalization. For each quintile we then provide averages for the quoted spread, quoted depth, number of messages, Garman-Klass volatility, market capitalization and dollar option volume by each equity subgroup and for calls and puts and "in-", "near-" and "out-of-the-money" options. The values for the market capitalization and volatility variables are equivalent for the in-, near-, and out-of-the-money categories since they are based on the underlying stocks. Dollar option volume is the average per strike price for each stock in the category and then divided by 100 (the strikes include those without a trade but with a quote).

## Table 3: The Effect of Algorithmic Trading on Bid-Ask Spreads (2007, Calls)

| Group/Sorting Criteria | Q1 | Q2 | Q3 | Q4 | Q5 | Volume | Market Cap | GK Volatility |
|---|---|---|---|---|---|---|---|---|
| CALLS (IN) 2007 | | | | | | | | |
| Qspread for Volume | -0.0001 | -0.0011 | -0.0057 | -0.0032 | -0.0051 | -0.5043 | -3.7400 | 0.1462 |
| | (-55.21) | (-32.94) | (-17.71) | (-1.95) | (-0.62) | (-1.31) | (-77.93) | (65.97) |
| Qspread for Market Cap | -0.0001 | -0.0004 | -0.0022 | -0.0017 | -0.0025 | -0.1478 | -3.1800 | 0.0000 |
| | (-66.69) | (-28.04) | (-22.80) | (-23.31) | (-2.28) | (-4.66) | (-56.65) | (77.42) |
| Qspread for GK Volatility | -0.0004 | -0.0001 | -0.0001 | -0.0002 | -0.0002 | -14.400 | -13.4300 | 6.9487 |
| | (-22.76) | (-10.37) | (-15.74) | (-19.99) | (-43.34) | (-7.45) | (-45.20) | (22.92) |
| CALLS (NEAR) 2007 | | | | | | | | |
| Qspread for Volume | -0.0008 | -0.0096 | -0.0231 | -0.0121 | -0.0104 | -29.4100 | 3.8200 | 0.3635 |
| | (-48.14) | (-49.88) | (-31.60) | (-8.02) | (-2.06) | (-101.58) | (8.30) | (13.57) |
| Qspread for Market Cap | -0.0007 | -0.0007 | -0.0132 | -0.0078 | -0.0141 | -34.1100 | 13.6300 | 0.0000 |
| | (-40.75) | (-13.67) | (-39.04) | (-18.92) | (-7.38) | (-107.77) | (18.80) | (14.29) |
| Qspread for GK Volatility | 0.0003 | -0.0002 | -0.0007 | -0.0008 | -0.0009 | -62.7400 | -39.5400 | 2.1600 |
| | (2.46) | (-0.33) | (-0.96) | (-12.73) | (-40.47) | (-68.82) | (-27.54) | (14.94) |
| CALLS (OUT) 2007 | | | | | | | | |
| Qspread for Volume | -0.0020 | -0.0271 | -0.0478 | 0.0014 | 0.0060 | -102.5100 | 8.6900 | 1.0900 |
| | (-40.68) | (-38.15) | (-16.34) | (0.30) | (1.09) | (-197.35) | (9.92) | (27.63) |
| Qspread for Market Cap | -0.0016 | -0.0022 | -0.0476 | -0.0120 | -0.0075 | -102.5100 | 22.4400 | 0.0002 |
| | (-28.93) | (-18.90) | (-27.06) | (-10.88) | (-1.96) | (-179.58) | (15.48) | (30.40) |
| Qspread for GK Volatility | -0.0019 | -0.0007 | -0.0007 | -0.0023 | -0.0031 | -122.2100 | -16.8200 | 1.5600 |
| | (-6.88) | (-8.02) | (-6.23) | (-16.28) | (-35.78) | (-85.85) | (-7.00) | (10.15) |

## Table 3: The Effect of Algorithmic Trading on Bid-Ask Spreads (2007, Puts) Continued…

**PUTS (IN) 2007**

| Group/Sorting Criteria | Q1 | Q2 | Q3 | Q4 | Q5 | Volume | Market Cap | GK Volatility |
|---|---|---|---|---|---|---|---|---|
| Qspread for Volume | -0.0002 | -0.0022 | -0.0136 | -0.0052 | -0.0219 | -4.6400 | -2.8500 | 0.1606 |
| | (-19.86) | (-15.85) | (-8.66) | (-0.80) | (-1.43) | (-28.72) | (-13.14) | (20.39) |
| Qspread for Market Cap | -0.0002 | -0.0004 | -0.0038 | -0.0019 | -0.0111 | -4.39 | -5.0900 | 0.0000 |
| | (-19.03) | (-18.61) | (-10.90) | (-9.27) | (-3.16) | (-23.98) | (-13.50) | (11.55) |
| Qspread for GK Volatility | -0.0004 | -0.0006 | -0.0001 | -0.0004 | -0.0002 | -18.51 | -22.5400 | -0.1015 |
| | (-5.23) | (-10.40) | (-8.22) | (-13.88) | (-26.05) | (-25.45) | (-21.06) | (-1.37) |

**PUTS (NEAR) 2007**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Qspread for Volume | -0.0007 | -0.0079 | 0.0224 | -0.0069 | -0.0057 | -21.0900 | 1.5700 | 0.3239 |
| | (-40.82) | (-37.29) | (-25.32) | (-3.00) | (-0.91) | (-81.53) | (3.71) | (13.38) |
| Qspread for Market Cap | -0.0007 | -0.0007 | -0.0127 | -0.0071 | -0.0194 | -26.4800 | 4.5200 | 0.0000 |
| | (-36.93) | (-14.77) | (-29.50) | (-15.28) | (-8.97) | (-89.64) | (6.68) | (11.92) |
| Qspread for GK Volatility | -0.0005 | -0.0002 | 0.0004 | -0.0009 | -0.0008 | -51.7400 | -21.3000 | 0.8620 |
| | (-3.68) | (-3.37) | (0.61) | (-13.78) | (-38.29) | (-56.49) | (-15.29) | (7.50) |

**PUTS (OUT) 2007**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Qspread for Volume | -0.0019 | -0.0242 | -1.0000 | -0.0141 | -0.0092 | 77.1800 | 8.9500 | 1.0800 |
| | (-40.14) | (-24.05) | (-14.33) | (-1.79) | (-0.23) | (-160.94) | (9.97) | (25.98) |
| Qspread for Market Cap | -0.0017 | -0.0025 | -0.0381 | -0.0158 | -0.1005 | -81.4500 | 11.7900 | 0.0001 |
| | (-30.92) | (-19.97) | (-15.02) | (-8.59) | (-3.97) | (-152.16) | (8.60) | (24.27) |
| Qspread forGK Volatility | -0.0011 | -0.0011 | -0.0008 | -0.0030 | -0.0026 | -109.2100 | -31.9200 | 7.7900 |
| | (-3.49) | (-10.37) | (-6.96) | (-16.15) | (-35.20) | (-65.25) | (-10.42) | (20.07) |

The Table regresses the quoted Spread (QSpread) on a proxy for algorithmic trading (message traffic) and the three control variables of market capitalization, Garman-Klass volatility of the underlying stock, and the dollar volume of the stock's options for the general sample for 2007. The control variable values given here are for quintile 1. The specification is: $l_{it} = \alpha_i + \beta_i A_{it} + \delta_{it} X_{it} + \epsilon_{it}$ where $l_{it}$ is the liquidity variable (quoted spread in this case), $A_{it}$ is the

message traffic representing algorithmic trading, and $X_{it}$ is the set of control variables such as market capital, Garman-Klass volatility of the underlying stock, and the dollar volume of the option. Volume is the logarithm of the average volume per strike and per stock after dividing by 100.

## Table 4: The Effect of Algorithmic Trading on Bid-Ask Spreads (2008, Calls)

**CALL(IN)2008**

| Group/Sorting Criteria | Q1 | Q2 | Q3 | Q4 | Q5 | Volume | Market Cap | GK Volatility |
|---|---|---|---|---|---|---|---|---|
| Qspread for Volume | -0.0003 | -0.0015 | -0.0027 | -0.0128 | 0.0056 | -3.3500 | 1.7700 | 0.0069 |
| | (-27.05) | (-19.07) | (-4.72) | (-3.62) | (0.36) | (-16.11) | (7.52) | (8.59) |
| Qspread for Market Cap | -0.0004 | -0.0005 | -0.0013 | -0.0010 | -0.0077 | -5.0100 | 2.0100 | 0.1516 |
| | (-30.47) | (-31.04) | (-12.30) | (-22.64) | (-4.90) | (-19.65) | (4.34) | (44.10) |
| Qspread for GK Volatility | -0.0091 | -0.0007 | -0.0017 | -0.0014 | -0.0004 | -2.7500 | -17.2900 | 0.0535 |
| | (-5.99) | (-13.34) | (-7.25) | (-10.12) | (-18.49) | (-0.91) | (-2.04) | (1.40) |

**CALL(NEAR)2008**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Qspread for Volume | -0.0005 | -0.0056 | -0.0138 | -0.0267 | -0.0243 | -14.3600 | 4.5300 | 0.0186 |
| | (-54.71) | (-44.77) | (-28.56) | (-14.22) | (-5.18) | (-59.11) | (12.92) | (16.56) |
| Qspread for Market Cap | -0.0007 | -0.0007 | -0.0078 | -0.0016 | -0.0352 | -22.6500 | 1.2900 | 0.1008 |
| | (-58.71) | (-39.11) | (-33.69) | (-22.80) | (-18.37) | (-76.17) | (1.97) | (40.89) |
| Qspread for GK Volatility | -0.0111 | -0.0012 | -0.0011 | -0.0015 | -0.0008 | -61.4200 | -282.5700 | -0.0889 |
| | (-3.38) | (-11.94) | (-4.97) | (-15.94) | (-9.72) | (-11.86) | (-28.69) | (-1.55) |

**CALL(OUT)2008**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Qspread for Volume | -0.0022 | -0.0158 | -0.0308 | -0.0640 | 0.0018 | -97.3200 | 18.4100 | 0.0279 |
| | (-83.72) | (-48.95) | (-20.89) | (-10.85) | (0.14) | (-210.46) | (23.40) | (14.57) |
| Qspread for Market Cap | -0.0020 | -0.0023 | -0.0222 | -0.0049 | -0.0047 | -108.2300 | 0.1238 | 0.1357 |
| | (-59.49) | (-50.95) | (-29.43) | (-29.57) | (-10.87) | (-197.49) | (0.08) | (40.87) |
| Qspread for GK Volatility | -0.0170 | -0.0020 | -0.0080 | -0.0067 | -0.0040 | -123.3700 | -232.1600 | -0.1090 |
| | (-3.07) | (-12.08) | (-10.08) | (-19.20) | (-13.40) | (-19.09) | (-18.84) | (-1.58) |

## Table 4: The Effect of Algorithmic Trading on Bid-Ask Spreads (2008, Puts) Continued…

**PUT(IN)2008**

| | Q1 | Q2 | Q3 | Q4 | Q5 | Volume | Market Cap | GK Volatility |
|---|---|---|---|---|---|---|---|---|
| Qspread for Volume | -0.0003 | -0.0013 | 0.0007 | 0.0016 | 0.0171 | -0.2.6700 | -1.6200 | 0.0050 |
| | (-34.50) | (-16.0) | (1.80) | (0.94) | (1.27) | (-16.80) | (-7.84) | (13.95) |
| Qspread for Market Cap | -0.0004 | -0.0005 | -0.0011 | -0.0008 | -0.0048 | -4.2300 | -5.5400 | 0.0139 |
| | (-37.68) | (-36.23) | (-7.72) | (-22.41) | (-4.10) | (-21.67) | (-13.66) | (20.19) |
| Qspread for GK Volatility | -0.0025 | -0.0008 | 0.0005 | -0.0003 | -0.0004 | -26.5300 | -21.3100 | -0.0017 |
| | (-1.92) | (-11.21) | (0.46) | (-6.27) | (-6.59) | (-10.23) | (-4.26) | (-0.42) |

**PUT(NEAR)2008**

| | Q1 | Q2 | Q3 | Q4 | Q5 | Volume | Market Cap | GK Volatility |
|---|---|---|---|---|---|---|---|---|
| Qspread for Volume | -0.0003 | -0.0045 | -0.0095 | -0.0210 | -0.0263 | -9.1900 | 3.3200 | 0.0083 |
| | (-46.88) | (-38.59) | (-19.28) | (-10.58) | (-3.83) | (-47.72) | (11.61) | (9.40) |
| Qspread for Market Cap | -0.0005 | -0.0005 | -0.0073 | -0.0016 | -0.0120 | -15.1500 | -1.4600 | 0.0316 |
| | (-50.82) | (-32.5) | (-28.36) | (-23.16) | (-8.94) | (-64.18) | (-2.72) | (15.43) |
| Qspread for GK Volatility | -0.0030 | -0.0014 | -0.0004 | -0.0010 | -0.0007 | -54.4300 | 4.5400 | 0.2978 |
| | (-1.37) | (-15.86) | (-2.56) | (-11.76) | (-9.36) | (-19.07) | (0.79) | (6.69) |

**PUT(OUT)2008**

| | Q1 | Q2 | Q3 | Q4 | Q5 | Volume | Market Cap | GK Volatility |
|---|---|---|---|---|---|---|---|---|
| Qspread for Volume | -0.0015 | -0.0077 | -0.0336 | -0.0185 | -0.0319 | -60.7000 | 2.7200 | 0.0096 |
| | (-59.43) | (-25.84) | (-12.89) | (-2.81) | (-1.62) | (-139.72) | (3.89) | (4.57) |
| Qspread for Market Cap | -0.0014 | -0.0018 | -0.0203 | -0.0047 | -0.0373 | -63.8900 | -9.2000 | 0.3208 |
| | (-41.43) | (-39.69) | (-17.14) | (-24.16) | (-5.37) | (-41.43) | (-6.55) | (28.61) |
| Qspread for GK Volatility | -0.0771 | -0.0026 | -0.0071 | -0.0070 | -0.0042 | -90.0600 | 62.7100 | 0.3808 |
| | (-3.36) | (-12.65) | (-11.89) | (-17.74) | (-10.41) | (-6.28) | (1.59) | (2.20) |

The Table regresses the quoted Spread (QSpread) on a proxy for algorithmic trading (message traffic) and the three control variables of market capitalization, Garman-Klass volatility of the underlying stock, and the dollar volume of the stock's options for the general sample for 2008. The control variable values given here are for quintile 1. The specification is: $l_{it} = \alpha_i + \beta_i A_{it} + \delta_{it} X_{it} + \epsilon_{it}$ where $l_{it}$ is the liquidity variable (quoted spread in this case), $A_{it}$ is the

message traffic representing algorithmic trading, and $X_{it}$ is the set of control variables such as market capital, Garman-Klass volatility of the underlying stock, and the dollar volume of the option. Volume is the logarithm of the average volume per strike and per stock after dividing by 100.

## Table 5: The Effect of Algorithmic Trading on Depth (2007, Calls)

| Group/Sorting Criteria | Q1 | Q2 | Q3 | Q4 | Q5 | Volume | Market Cap | GK Volatility |
|---|---|---|---|---|---|---|---|---|
| **CALL(IN)2007** | | | | | | | | |
| Qdepth for Volume | 0.0036 | -0.0006 | -0.0023 | -0.0005 | 0.0012 | -15.5500 | 27.0500 | -3.0200 |
| | (27.30) | (-8.30) | (-8.92) | (-0.88) | (1.81) | (-8.79) | (12.30) | (-29.75) |
| Qdepth for Market Cap | 0.0041 | 0.0094 | 0.0012 | 0.0012 | 0.0079 | -4.2945 | 68.4400 | -6.1800 |
| | (29.38) | (43.00) | (9.12) | (11.87) | (10.06) | (-0.24) | (21.17) | (-31.94) |
| Qdepth for GK Volatility | 0.0008 | 0.0008 | 0.0030 | 0.0008 | 0.0002 | 6.2269 | 3.6400 | 2.0100 |
| | (19.94) | (36.52) | (51.23) | (10.076) | (13.75) | (1.48) | (5.61) | (30.030) |
| **CALL(NEAR)2007** | | | | | | | | |
| Qdepth for Volume | -0.0010 | -0.0006 | -0.0010 | -0.0001 | 0.0001 | -90.1200 | 33.5800 | -11.5300 |
| | (-4.39) | (-6.28) | (-7.30) | (-0.88) | (0.07) | (-23.19) | (5.43) | (-32.06) |
| Qdepth for Market Cap | 0.0031 | 0.0140 | 0.0025 | -0.0005 | 0.0048 | 6.2000 | -101.7600 | -28.4900 |
| | (10.64) | (57.55) | (20.36) | (-3.76) | (15.74) | (1.47) | (-10.50) | (-41.14) |
| Qdepth for GK Volatility | 0.0004 | 0.0005 | 0.0033 | -0.0006 | -0.0048 | -3.8693 | 35.5300 | 2.1400 |
| | (5.83) | (19.48) | (46.17) | (-5.81) | (-15.83) | (-0.85) | (49.64) | (29.80) |
| **CALL(OUT)2007** | | | | | | | | |
| Qdepth for Volume | 0.0010 | 0.0001 | -0.0001 | -0.0002 | 0.0001 | -104.0500 | 173.4600 | -4.4200 |
| | (4.09) | (0.60) | (-3.75) | (-1.34) | (0.68) | (-40.41) | (39.93) | (-22.54) |
| Qdepth for Market Cap | 0.0034 | 0.0048 | 0.0073 | 0.0007 | 0.0026 | -85.0300 | 224.0500 | -12.1300 |
| | (11.67) | (16.97) | (17.07) | (2.35) | (5.18) | (-28.75) | (29.84) | (-30.084) |
| Qdepth for GK Volatility | 0.0003 | 0.0006 | 0.0012 | 0.0018 | -0.0046 | -3.6600 | 39.5800 | 0.87070 |
| | (0.034) | (18.91) | (10.69) | (10.36) | (-10.09) | (-7.04) | (45.12) | (15.49) |

## Table 5: The Effect of Algorithmic Trading on Depth (2007, Puts) Continued…

| Group/Sorting Criteria | Q1 | Q2 | Q3 | Q4 | Q5 | Volume | Market Cap | GK Volatility |
|---|---|---|---|---|---|---|---|---|
| **PUT(IN)2007** | | | | | | | | |
| Qdepth for Volume | 0.0032 | -0.0015 | -0.0016 | 0.0009 | 0.0045 | -30.8100 | 24.9500 | -2.0400 |
| | (27.62) | (-11.58) | (-5.60) | (1.76) | (5.50) | (-18.11) | (10.91) | (-24.58) |
| Qdepth for Market Cap | 0.0040 | 0.0059 | 0.0011 | 0.0015 | 0.0063 | -1.9800 | 119.8600 | -3.7700 |
| | (32.34) | (24.40) | (5.72) | (8.79) | (7.92) | (-1.15) | (33.84) | (-24.53) |
| Qdepth for GK Volatility | 0.0004 | 0.0007 | 0.0039 | 0.0011 | -0.0008 | 7.6400 | 8.1900 | 0.7930 |
| | (7.37) | (24.69) | (44.68) | (11.73) | (-0.54) | (14.96) | (10.90) | (15.28) |
| **PUT(NEAR)2007** | | | | | | | | |
| Qdepth for Volume | -0.0065 | -0.0006 | -0.0006 | 0.0001 | 0.0003 | 55.0500 | 3.7600 | -14.2200 |
| | (-23.57) | (-5.32) | (-3.41) | (0.83) | (0.61) | (13.26) | (0.055) | (-36.58) |
| Qdepth for Market Cap | -0.0039 | 0.0087 | 0.0027 | -0.0002 | 0.0051 | 129.8600 | -250.6500 | -36.1600 |
| | (-11.37) | (31.04) | (16.08) | (-1.38) | (11.27) | (27.11) | (-22.85) | (-46.40) |
| Qdepth for GK Volatility | 0.0004 | 0.0003 | 0.0028 | -0.0010 | -0.0133 | -0.99600 | 38.1900 | 1.2800 |
| | (5.12) | (12.88) | (30.93) | (-7.25) | (-38.00) | (-1.87) | (47.02) | (19.07) |
| **PUT(OUT)2007** | | | | | | | | |
| Qdepth for Volume | -0.0033 | 0.0002 | 0.0064 | 0.0007 | 0.0022 | -77.3100 | 176.6200 | -10.0700 |
| | (-9.63) | (0.84) | (10.43) | (0.02) | (2.61) | (-22.72) | (27.73) | (-34.14) |
| Qdepth for Market Cap | 0.0003 | 0.0049 | 0.0047 | -0.0006 | 0.0271 | -55.6000 | 115.4400 | -23.7500 |
| | (0.89) | (10.16) | (8.06) | (-1.62) | (6.82) | (-14.11) | (11.43) | (-43.58) |
| Qdepth for GK Volatility | 0.0005 | 0.0003 | 0.0008 | 0.0012 | -0.0136 | -6.5000 | 46.3800 | 25.4000 |
| | (4.72) | (9.50) | (5.89) | (4.82) | (-24.09) | (-10.41) | (40.59) | (17.55) |

The Table regresses the quoted depth (Qdepth) on a proxy for algorithmic trading (message traffic) and the three control variables of market capitalization, Garman-Klass volatility of the underlying stock, and the dollar volume of the stock's options for the general sample for 2007. The control variable values given here are for quintile 1. The specification is: $l_{it} = \alpha_i + \beta_i A_{it} + \delta_{it} X_{it} + \epsilon_{it}$ where $l_{it}$ is the liquidity variable (quoted spread in this case), $A_{it}$ is the

message traffic representing algorithmic trading, and $X_{it}$ is the set of control variables such as market capital, Garman-Klass volatility of the underlying stock, and the dollar volume of the option. Volume is the logarithm of the average volume per strike and per stock after dividing by 100.

## Table 6: The Effect of Algorithmic Trading on Depth (2008, Calls)

| CALL(IN)2008 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group/Sorting Criteria | Q1 | Q2 | Q3 | Q4 | Q5 | Volume | Market Cap | GK Volatility |
| Qdepth for Volume | 0.0004 | -0.0000 | 0.0000 | -0.0000 | -0.0001 | -23.1000 | 6.7500 | -0.0332 |
| | (71.29) | (-0.35) | (0.87) | (-0.63) | (-0.44) | (-20.26) | (5.24) | (-7.30) |
| Qdepth for Market Cap | 0.0003 | 0.0000 | 0.0000 | 0.0002 | 0.0004 | -16.9600 | 4.9400 | -0.3830 |
| | (60.58) | (50.68) | (10.01) | (74.84) | (9.79) | (-15.99) | (2.56) | (-26.76) |
| Qdepth for GK Volatility | 0.0001 | 0.0000 | 0.0001 | 0.0001 | 0.0002 | 3.8400 | 13.9200 | -0.0144 |
| | (10.50) | (12.25) | (8.42) | (19.18) | (22.16) | (1.68) | (2.18) | (-0.50) |
| CALL(NEAR)2008 | | | | | | | | |
| Qdepth for Volume | 0.0002 | -0.0002 | -0.0000 | 0.0000 | -0.0000 | -67.1600 | -15.1800 | -0.0661 |
| | (58.08) | (-21.18) | (-8.72) | (2.31) | (-0.76) | (-63.08) | (-9.89) | (-13.43) |
| Qdepth for Market Cap | 0.0003 | 0.0001 | 0.0002 | 0.0005 | 0.0001 | -44.4800 | -61.8900 | -0.1959 |
| | (86.76) | (36.32) | (24.85) | (148.02) | (12.01) | (-41.84) | (-26.47) | (-22.23) |
| Qdepth for GK Volatility | -0.0002 | 0.0001 | -0.0000 | 0.0001 | 0.0002 | 46.8200 | -73.8700 | -0.1554 |
| | (-7.08) | (23.51) | (-0.00) | (12.91) | (23.77) | (7.65) | (-6.35) | (-2.29) |
| CALL(OUT)2008 | | | | | | | | |
| Qdepth for Volume | 0.0002 | 0.0000 | -0.0000 | 0.0001 | -0.0000 | -65.9800 | 27.0300 | -0.0537 |
| | (49.13) | (0.99) | (-2.29) | (6.40) | (-1.16) | (-67.27) | (16.20) | (-13.20) |
| Qdepth for Market Cap | 0.0006 | 0.0001 | 0.0001 | 0.0006 | 0.0001 | -69.6600 | -76.7800 | -0.1389 |
| | (85.02) | (21.15) | (9.50) | (71.07) | (5.12) | (-61.18) | (-24.70) | (-20.14) |
| Qdepth for GK Volatility | -0.0001 | 0.0001 | -0.0001 | 0.0002 | 0.0002 | 24.7400 | -45.2200 | -0.0657 |
| | (-3.87) | (15.28) | (-4.28) | (14.22) | (13.20) | (6.63) | (-6.36) | (-1.64) |

## Table 6: The Effect of Algorithmic Trading on Depth (2008, Puts) Continued…

| PUT(IN)2008 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group/Sorting Criteria | Q1 | Q2 | Q3 | Q4 | Q5 | Volume | Market Cap | GK Volatility |
| Qdepth for Volume | 0.0003 | -0.0000 | -0.0000 | -0.0000 | -0.0001 | -17.9800 | 9.2044 | -0.0120 |
| | (62.76) | (-6.13) | (-3.71) | (-1.33) | (-1.07) | (-18.65) | (0.73) | (-5.45) |
| Qdepth for Market Cap | 0.0002 | 0.0005 | 0.0001 | 0.0002 | 0.0001 | -14.7700 | 26.08 | 0.0739 |
| | (50.12) | (66.79) | (12.76) | (53.57) | (5.25) | (-15.14) | (12.86) | (21.45) |
| Qdepth for GK Volatility | 0.0005 | 0.0001 | 0.0000 | 0.0001 | 0.0001 | 18.2700 | -28.02 | -0.00411 |
| | (5.68) | (12.03) | (0.28) | (16.02) | (21.19) | (10.06) | (-8.00) | (-1.38) |
| PUT(NEAR)2008 | | | | | | | | |
| Qdepth for Volume | 0.0002 | -0.0002 | -0.0000 | 0.0000 | -0.0003 | -60.5100 | -28.7300 | -0.0595 |
| | (56.51) | (-17.02) | (-10.31) | (3.02) | (-0.73) | (-52.97) | (-16.95) | (-11.33) |
| Qdepth for Market Cap | 0.0004 | 0.0001 | 0.0002 | 0.0005 | 0.0017 | -43.7000 | -98.9400 | -0.1910 |
| | (84.84) | (24.12) | (19.19) | (109.73) | (11.54) | (-37.81) | (-37.56) | (-19.00) |
| Qdepth for GK Volatility | -0.0002 | 0.0001 | -0.0000 | 0.0000 | 0.0020 | -25.2000 | -13.5700 | -0.0803 |
| | (-1.83) | (18.80) | (-4.62) | (6.71) | (16.41) | (-1.83) | (-4.20) | (-3.21) |
| PUT(OUT)2008 | | | | | | | | |
| Qdepth for Volume | 0.0002 | 0.0000 | 0.0013 | -0.0000 | 0.0002 | -71.0400 | -15.2700 | -0.0479 |
| | (39.54) | (3.39) | (5.85) | (-0.66) | (0.16) | (-58.48) | (-7.82) | (-8.13) |
| Qdepth for Market Cap | 0.0005 | 0.0000 | 0.0014 | 0.0006 | 0.0013 | -67.1700 | -229.1100 | -1.0800 |
| | (60.85) | (11.91) | (4.37) | (50.42) | (2.43) | (-45.04) | (-58.38) | (-34.29) |
| Qdepth for GK Volatility | 0.0002 | 0.0000 | -0.0011 | 0.0002 | 0.0052 | -3.2500 | 29.9000 | -0.0203 |
| | (6.52) | (3.53) | (-3.99) | (11.18) | (16.04) | (-1.69) | (5.64) | (-0.88) |

The Table regresses the quoted depth (Qdepth) on a proxy for algorithmic trading (message traffic) and the three control variables of market capitalization, Garman-Klass volatility of the underlying stock, and the dollar volume of the stock's options for the general sample for 2008. The control variable values given here are for quintile 1. The specification is: $l_{it} = \alpha_i + \beta_i A_{it} + \delta_{it} X_{it} + \epsilon_{it}$ where $l_{it}$ is the liquidity variable (quoted spread in this case), $A_{it}$ is the

message traffic representing algorithmic trading, and $X_{it}$ is the set of control variables such as market capital, Garman-Klass volatility of the underlying stock, and the dollar volume of the option. Volume is the logarithm of the average volume per strike and per stock after dividing by 100.

## Table 7: Summary Statistics for the Penny Quote Sample

| | | CALLS | PUTS |
|---|---|---|---|
| Quoted Spread | | | |
| | in | 0.0213 | 0.0348 |
| | near | 0.0989 | 0.1050 |
| | out | 0.6063 | 0.5353 |
| Quoted Depth | | | |
| | in | 692 | 584 |
| | near | 1,322 | 1,447 |
| | out | 1,126 | 1,593 |
| Messages | | | |
| | in | 37,958 | 37,689 |
| | near | 42,756 | 39,396 |
| | out | 22,561 | 18,696 |
| GK Volatility | | | |
| | in | 5.4287 | 5.4287 |
| | near | 5.4287 | 5.4287 |
| | out | 5.4287 | 5.4287 |
| Market Cap | | | |
| | in | 16.9911 | 16.9911 |
| | near | 16.9911 | 16.9911 |
| | out | 16.9911 | 16.9911 |
| Volume | | | |
| | in | 556.5453 | 708.1939 |
| | near | 1767.2345 | 1960.4648 |
| | out | 367.3980 | 358.2757 |

Based on the option code we divide the data into call and put options and then into in-, near-, and out-of-the-money options. The above table provides the averages for the call and put options for the penny quote sample for 2007 and 2008 for the variables of interest. Dollar option volume is the average per strike price for each stock in the category and then divided by 100 (the strikes include those without a trade but with a quote).

## Table 8: The Effect of Algorithmic Trading on Bid-Ask Spreads for the Penny Quote Sample (Calls)

**CALL(IN)2007**

| | Messages | GK Volatility | Market Cap | Volume |
|---|---|---|---|---|
| Qspread for phase1 | -0.0000 | 0.0006 | -0.0024 | 0.0002 |
| | (-0.06) | (0.04) | (-0.09) | (0.01) |
| Qspread for phase2 | -0.0001 | 0.0006 | -0.0036 | 0.0006 |
| | (-0.09) | (0.12) | (-0.18) | (0.03) |
| Qspread for phase3 | 0.0000 | -0.0004 | -0.0088 | -0.0013 |
| | (0.07) | (-0.05) | (-0.07) | (-0.04) |

**CALL(NEAR)2007**

| | | | | |
|---|---|---|---|---|
| Qspread for phase1 | -0.0002 | 0.0029 | 0.0115 | -0.0230 |
| | (-5.45) | (2.36) | (2.21) | (-2.58) |
| Qspread for phase2 | -0.0002 | 0.0030 | 0.0013 | -0.0218 |
| | (-10.47) | (8.65) | (0.82) | (-11.16) |
| Qspread for phase3 | -0.0000 | 0.0034 | 0.0350 | -0.0206 |
| | (-10.54) | (10.95) | (10.47) | (-6.18) |

**CALL(OUT)2007**

| | | | | |
|---|---|---|---|---|
| Qspread for phase1 | -0.0001 | 0.0567 | -0.0061 | 0.0258 |
| | (-6.40) | (7.63) | (-0.68) | (0.92) |
| Qspread for phase2 | -0.0000 | 0.0079 | 0.0107 | -0.0720 |
| | (-9.67) | (11.16) | (2.75) | (-16.65) |
| Qspread for phase3 | -0.0000 | 0.0070 | 0.0954 | -0.0206 |
| | (-16.60) | (15.15) | (20.18) | (-2.98) |

## Table 8: The Effect of Algorithmic Trading on Bid-Ask Spreads for the Penny Quote Sample (Puts)

**PUT(IN)2007**

| | Messages | GK Volatility | Market Cap | Volume |
|---|---|---|---|---|
| Qspread for phase1 | 0.0000 | -0.0070 | 0.0001 | 0.0005 |
| | (0.03) | (-0.03) | (0.00) | (0.01) |
| Qspread for phase2 | -0.0000 | 0.0006 | -0.0006 | -0.0048 |
| | (-0.58) | (0.49) | (-0.16) | (-1.54) |
| Qspread for phase3 | 0.0000 | -0.0031 | -0.0465 | -0.0019 |
| | (0.47) | (-0.46) | (-0.47) | (-0.36) |

**PUT(NEAR)2007**

| | Messages | GK Volatility | Market Cap | Volume |
|---|---|---|---|---|
| Qspread for phase1 | -0.0000 | 0.0034 | 0.0066 | 0.0065 |
| | (-5.55) | (1.18) | (1.77) | (0.64) |
| Qspread for phase2 | -0.0000 | 0.0021 | 0.0022 | -0.0209 |
| | (-9.32) | (6.89) | (1.68) | (-10.62) |
| Qspread for phase3 | -0.0000 | 0.0025 | 0.0293 | -0.0119 |
| | (-8.13) | (7.03) | (6.88) | (-3.63) |

**PUT(OUT)2007**

| | Messages | GK Volatility | Market Cap | Volume |
|---|---|---|---|---|
| Qspread for phase1 | -0.0002 | 0.0961 | -0.0236 | 0.1049 |
| | (-4.39) | (4.43) | (-1.80) | (2.18) |
| Qspread for phase2 | -0.0000 | 0.0050 | 0.0042 | -0.0482 |
| | (-12.07) | (9.78) | (1.32) | (-13.41) |
| Qspread for phase3 | -0.0000 | 0.0032 | 0.0548 | 0.0266 |
| | (-13.88) | (8.85) | (13.99) | (3.24) |

The Table regresses the quoted spread on a proxy for algorithmic trading (message traffic) and various controls such as market capitalization, the Garman-Klass volatility of the underlying stock, dollar trading volume of the stock's options and a dummy variable which takes the value of 1 if it is after the penny quote introduction. The specification is: $l_{it} = \alpha_i + \gamma_t + \beta_i A_{it} + \delta_{it} X_{it} + \epsilon_{it}$ where $l_{it}$ is the liquidity variable (either spread or depth), $A_{it}$ is the message traffic representing algorithmic trading, and $X_{it}$ is the set of control variables such as market capital, Garman-Klass volatility of the underlying stock, and the trading volume of the option. The equation includes an additional variable $\gamma_t$ to represent the time dummy for before and after the penny quotes were introduced. Volume is the logarithm of the average volume per strike and per stock after dividing by 100.

## Table 9: The Effect of Algorithmic Trading on Depth for the Penny Quote Sample (Calls)

**CALL(IN)2007**

| | Messages | GK Volatility | Market Cap | Volume |
|---|---|---|---|---|
| Qdepth for phase1 | -0.0002 | -0.0440 | -0.5286 | 0.2054 |
| | (-12.92) | (-3.19) | (-16.86) | (8.69) |
| Qdepth for phase2 | -0.0000 | -0.0302 | -0.0089 | 0.0581 |
| | (-4.62) | (-6.23) | (-0.54) | (2.67) |
| Qdepth for phase3 | 0.0000 | -0.0111 | -0.1160 | -0.0250 |
| | (1.36) | (-1.27) | (-1.04) | (-0.78) |

**CALL(NEAR)2007**

| | | | | |
|---|---|---|---|---|
| Qdepth for phase1 | -0.0007 | -0.4652 | -1.5322 | 2.2927 |
| | (-13.14) | (-21.78) | (-16.98) | (12.40) |
| Qdepth for phase2 | -0.0000 | -0.0659 | 0.0565 | 0.2299 |
| | (-31.37) | (-27.81) | (6.67) | (20.29) |
| Qdepth for phase3 | -0.0000 | 0.0254 | 0.4499 | 0.3652 |
| | (-38.63) | (25.52) | (33.76) | (36.99) |

**CALL(OUT)2007**

| | | | | |
|---|---|---|---|---|
| Qdepth for phase1 | -0.0009 | 0.1309 | -0.2900 | 0.8277 |
| | (-2.01) | (1.03) | (-1.78) | (1.65) |
| Qdepth for phase2 | -0.0001 | -0.0189 | 0.0960 | 0.4000 |
| | (-26.22) | (-2.86) | (7.08) | (22.88) |
| Qdepth for phase3 | -0.0001 | 0.0038 | 0.1571 | 0.4254 |
| | (-17.52) | (1.92) | (9.32) | (16.16) |

## Table 9: The Effect of Algorithmic Trading on Depth for the Penny Quote Sample (Puts) Continued…

**PUT(IN)2007**

| | Messages | GK Volatility | Market Cap | Volume |
|---|---|---|---|---|
| Qdepth for phase1 | 0.0007 | -0.3993 | 0.0963 | 0.0276 |
| | (0.43) | (-0.46) | (0.31) | (0.15) |
| Qdepth for phase2 | -0.0000 | -0.0272 | 0.0153 | 0.0355 |
| | (-3.24) | (-5.61) | (0.58) | (1.54) |
| Qdepth for phase3 | 0.0002 | -0.0924 | -1.3156 | -0.0490 |
| | (3.09) | (-3.19) | (-3.12) | (-1.78) |

**PUT(NEAR)2007**

| | Messages | GK Volatility | Market Cap | Volume |
|---|---|---|---|---|
| Qdepth for phase1 | -0.0007 | -0.3434 | -0.9793 | 2.7788 |
| | (-9.45) | (-5.57) | (-13.37) | (10.60) |
| Qdepth for phase2 | -0.0000 | -0.0581 | 0.1101 | 0.3246 |
| | (-40.61) | (-35.32) | (15.71) | (31.10) |
| Qdepth for phase3 | -0.0000 | 0.0210 | 0.4320 | 0.2901 |
| | (-6.65) | (3.05) | (4.89) | (5.53) |

**PUT(OUT)2007**

| | Messages | GK Volatility | Market Cap | Volume |
|---|---|---|---|---|
| Qdepth for phase1 | -0.0053 | 1.6353 | -1.3258 | 3.8893 |
| | (-4.75) | (4.23) | (-5.88) | (4.41) |
| Qdepth for phase2 | -0.0000 | -0.0333 | 0.1247 | 0.6238 |
| | (-34.71) | (-6.75) | (13.64) | (34.68) |
| Qdepth for phase3 | -0.0000 | 0.0037 | 0.1642 | 0.5181 |
| | (-13.09) | (1.70) | (8.45) | (12.33) |

The Table regresses the quoted depth on a proxy for algorithmic trading (message traffic) and various controls such as market capitalization, the Garman-Klass volatility of the underlying stock, dollar trading volume of the stock's options and a dummy variable which takes the value of 1 if it is after the penny quote introduction. The specification is: $l_{it} = \alpha_i + \gamma_t + \beta_i A_{it} + \delta_{it} X_{it} + \epsilon_{it}$ where $l_{it}$ is the liquidity variable (either spread or depth), $A_{it}$ is the message traffic representing algorithmic trading, and $X_{it}$is the set of control variables such as market capital, Garman-Klass volatility of the underlying stock, and the trading volume of the option. The equation includes an additional variable $\gamma_t$ to represent the time dummy for before and after the penny quotes were introduced. Volume is the logarithm of the average volume per strike and per stock after dividing by 100.

# News Trading and Speed*

Thierry Foucault† Johan Hombert‡ Ioanid Roşu§

May 24, 2012

**Abstract**

Adverse selection occurs in financial markets because certain investors have either (a) more precise information, or (b) superior speed in accessing or exploiting information. To disentangle the effects of precision and speed on market performance, we compare two models in which a dealer and a more precisely informed trader continuously receive news about the value of an asset. In the first model the trader and the dealer are equally fast, while in the second model the trader receives the news one instant before the dealer. Compared with the first model, in the second model: (1) the fraction of trading volume due to the informed investor increases from near zero to a large value; (2) liquidity decreases; (3) short-term price changes are more correlated with asset value changes; (4) informed order flow autocorrelation decreases to zero. Our results suggest that the speed component of adverse selection is necessary to explain certain empirical regularities from the world of high frequency trading.

KEYWORDS: Insider trading, Kyle model, noise trading, trading volume, algorithmic trading, informed volatility, price impact.

---

*We thank Pete Kyle, Stefano Lovo, Victor Martinez, Dimitri Vayanos; and seminar participants at Paris Dauphine, Copenhagen Business School, Univ. Carlos III in Madrid, and ESSEC for valuable comments.

†HEC Paris, foucault@hec.fr.

‡HEC Paris, hombert@hec.fr.

§HEC Paris, rosu@hec.fr.

# 1 Introduction

The recent advent of high frequency trading (HFT) in financial markets has raised numerous questions about the role of high frequency traders and their strategies.[1] Because of the proprietary nature of HFT and its extraordinary speed, it is difficult to characterize HFT strategies in general.[2] Nevertheless, there is increasing evidence that at least one category of high frequency traders exploits very quick access to public information in an attempt to analyze the news and trade before everyone else. For example, in their online advertisement for real-time data processing tools, Dow Jones states: "Timing is everything and to make lucrative, well-timed trades, institutional and electronic traders need accurate real-time news available, including company financials, earnings, economic indicators, taxation and regulation shifts. Dow Jones is the leader in providing high-frequency trading professionals with elementized news and ultra low-latency news feeds for algorithmic trading."[3] This category of HFT can also use public market data to infer information from related securities. We call this category high frequency news trading (HFNT) or, in short, *news trading*.

Clearly, news trading generates adverse selection.[4] In general, adverse selection occurs because some investors have either (a) more precise information, or (b) superior speed in accessing or exploiting information. Traditionally, the market microstructure literature, e.g., Kyle (1985), has mainly focused on the first type of adverse selection. In contrast, the speed component of adverse selection has received little attention. Our paper focuses on this second type of adverse selection in the context of news trading.

To separate the role of precision and speed, we consider two models of trading under

[1]In many markets around the world, high frequency trading currently accounts for a majority of trading volume. Hendershott, Jones, and Menkveld (2011) report that in 2009 as much as 73% of trading volume in the United States was due to high frequency trading. A similar result is obtained by Brogaard (2011) for NASDAQ, and Chaboud, Chiquoine, Hjalmarsson, and Vega (2009) for various Foreign Exchange markets. High frequency trading has been questioned espectially after the U.S. "Flash Crash" on May 6, 2010. See, e.g., Kirilenko, Kyle, Samadi, and Tuzun (2011).

[2]SEC (2010) attributes the following characteristics to HFT: (1) the use of extraordinarily high-speed and sophisticated computer programs for generating, routing, and executing orders; (2) use of co-location services and individual data feeds offered by exchanges and others to minimize network and other types of latencies.

[3]See http://www.dowjones.com/info/HighFrequencyTrading.asp.

[4]Hendershott and Riordan (2011) find that on NASDAQ the marketable orders of high frequency traders have a significant information advantage and are correlated with future price changes.

asymmetric information. In both models, a risk-neutral informed trader and a competitive dealer (or market maker) continuously learn about the value of an asset. In both models, the informed trader receives a more precise stream of news than that received by the dealer. The only difference lies in the timing of access to the stream of news. In the first model, the *benchmark model*, the informed trader and the dealer are equally fast.[5] In the second model, the *fast model*, the informed trader receives the news one instant before the dealer. We show that even an infinitesimal speed advantage leads to large differences in the predictions of the two models.

We further argue that the fast model is better suited to describe the world of high frequency trading. For example, consider the recent increase in trading volume observed in various exchanges around that world, which in part has been attributed to the rise of HFT. At high frequencies, traditional models such as Kyle (1985), or extensions such as our benchmark model have difficulty in generating a large trading volume of investors with superior information. To see, this, consider Figure 1. As it is apparent from the plot, the fast model can account for a significant participation rate of informed trading at higher frequencies, while the informed trader in the benchmark model is essentially invisible at high frequencies.[6] Thus, accounting for adverse selection due to speed is important if we want to explain the large observed trading volume due to HFT.

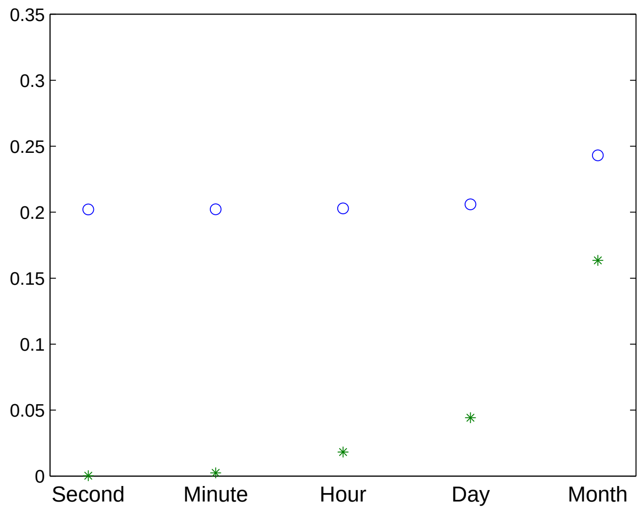
Why would a small speed advantage for the informed trader translate into such a large different in outcomes? For this, we need to understand the difference in optimal strategies of the informed trader in the two models. In principle, when the asset value changes over time, there are two components of the optimal strategy:

(1) *Level Trading* (or the low-frequency, drift, or deterministic component). This is a multiple of the difference between the asset value and the price, and changes slowly over time. Also, it is proportional to the time interval between two trades, thus it is small relative to the other component.

[5]The benchmark model is similar to that of Back and Pedersen (1998), except that in our model the dealer also receives news about the asset value.

[6]In our benchmark model, as in Kyle (1985), there is a single informed trader. We have checked that the pattern shown in the figure can be obtained in models with multiple informed traders, such as Back, Cao, and Willard (2000).

**Figure 1: Informed participation rate at various trading frequencies.** The figure plots the fraction of the trading volume due to the informed trader in a discrete time model for various lengths of time between trading periods (second, minute, hour, day, month) in (a) the benchmark model, marked with "$*$"; and (b) the fast model, marked with "$\circ$". The parameters used are $\sigma_u = \sigma_v = \sigma_e = \Sigma_0 = 1$ (see Theorem 1). The liquidation date $t = 1$ corresponds to 10 calendar years.




(2) *Flow Trading* (or the high-frequency, volatility, or stochastic component). This is a multiple of the new signal, i.e., the innovation in asset value, and changes every instant. This component is relatively much larger than the level trading component.

With no asymmetry in speed, the informed trader in the benchmark model does not have any incentive to trade on the asset value innovation, and trades only on the level of the asset value: the price impact of flow trading would otherwise be too high. By contrast, in the fast model the informed trader also engages in flow trading, in anticipation of a price move in the next instant due to the incorporation of news by the public.

These two components of the optimal strategy of the informed trader drive all the comparisons between the benchmark model and the fast model. To begin with, trading volume is higher in the fast model: in addition to the noise trading which is assumed the same in the two models, there is the large flow trading component from the informed trader (the level component is too small to matter at high frequencies). As observed in Figure 1, the fraction of trading volume due to the informed trader is much larger at high frequencies, due to the large flow trading component.

Liquidity is smaller in the fast model: besides the usual adverse selection coming from the superior precision of the informed trader, anticipatory trading generates additional adverse selection.

The comparison of price informativeness in the two models is more subtle. In the fast model, trades are more correlated with current innovations in asset value because of the flow trading component. Therefore, price changes are also more correlated with innovations in asset value. However, the variance of the pricing error is the same in both models. The reason is that there is a substitution between level trading and flow trading: there is flow trading in the fast model, but level trading is less intense than in the benchmark model. Therefore, in the fast model, trades are more correlated with current innovations in asset value, but also less correlated with past innovations. These two effects exactly offset and leave the variance of the pricing error identical in both models.

The effect of fast trading on price volatility is similarly complex. Price volatility arises from both trading and quote revisions, since the dealer also learns about the asset value and updates quotes. In the fast model, the contribution of trades to price volatility is larger, because of the volatile flow trading component of informed trading. The flip side is that when the market maker receives information, part of it has already been revealed through trading. Therefore, quote revisions are of a smaller magnitude, and price volatility unrelated to trading is lower in the fast model. These two effect on volatility exactly offset each other so that total price volatility is the same in both models and equal to the volatility of the asset value.

In the benchmark model, the informed order flow is autocorrelated: there is only the level trading component, which changes direction only very slowly over time. In the fast model, the informed order flow has zero autocorrelation: at high frequencies, flow trading dominates level trading, and the innovations in asset value are uncorrelated. Our results suggest that the fast model is better suited than the benchmark model to describe the strategies of high frequency traders: Brogaard (2011) observes that their order flow is indeed volatile, and there is little evidence of autocorrelation.

To discuss empirical implications of our paper, we start by arguing that the informed trader of the fast model fits certain stylized facts about high frequency traders: (i) large trading volume: the informed investor in the fast model trades in large quantities, while in the benchmark model informed trading volume is essentially zero at high frequencies; (ii) low order flow autocorrelation: the fast informed investor's trades have low serial correlation, compared to a large autocorrelation in the benchmark model; (iii) anticipatory trading: the order flow of the fast investor has a significant correlation with current price changes, compared to a low correlation in the benchmark model.

We stress that our model applies to the specific category of high frequency traders who engage in flow trading, but not necessarily to other types of high-frequency trading strategies such as high-frequency market making.[7] Recognizing this distinction is important for testing the predictions of our model.

We have two types of empirical predictions: (i) the effect of HFNT on various market outcomes; and (ii) the effect of various market characteristics on HFNT activity. For (i), we analyze the causal effect of HFNT by comparing the equilibrium outcomes when one moves from the benchmark model to the fast model. In the fast model, the informed trader is able to access information before the public does. This can occur, for example, by purchasing access to various high frequency news feeds, by co-location services offered by the exchange, by increasing automation, etc. The converse move from the fast model to the benchmark model is also of interest: it can represent, e.g., the effect of regulation aimed at dampening high frequency trading. From the discussion above, we see that eliminating the speed advantage of the informed trader (a) reduces trading volume; (b) reduces overall adverse selection, and thus increases market liquidity.

The second type of empirical prediction can be obtained in the context of the fast model, by studying the effect of various parameters on informed trading activity. For example, we find that an increase in the precision of public news increases the amount of flow trading, yet improves liquidity. To understand why, recall that flow trading arises because the informed trader is willing to trade based on his signal just before the market maker updates the quotes based on a correlated signal. The more precise

[7]See Jovanovic and Menkveld (2011) for a theoretical and empirical analysis of liquidity provision by fast traders.

the public news, the higher the correlation between the informed trader's signal and the market maker's signal. This increases the benefits of trading in anticipation of the quotes updates. Therefore, flow trading increases.[8] At the same time, more public news also improves liquidity. The reason is simple: having more precise public news reduces adverse selection. Interestingly, it implies that if the amount of public news changes (over time or across securities) then flow trading and liquidity move in the same direction. This is not because flow trading improves liquidity; indeed, we saw that the opposite is true when the informed trader acquires a speed advantage. Instead, this is because more public news increases *both* flow trading and liquidity.

Another example is the effect of price volatility. Holding constant the relative precision of public news, an increase in price volatility can be modeled as an increase in the volatility of the innovation of the asset value. Then, an increase in price volatility causes both an increase in flow trading activity, and a reduction in liquidity. The intuition is straightforward. When the asset is more volatile, the anticipation effect is stronger, and thus the flow trading increases. Because flow trading is more intense, there is more adverse selection due to speed, and liquidity is negatively affected.

Our paper is part of a growing theoretical literature on trading and speed. Biais, Foucault, and Moinas (2011) analyze the welfare implications of the speed advantage of HFTs in a 3-period model: HFTs raise trading volume and gains from trade, but increase adverse selection. In a search model with symmetric information, Pagnotta and Philippon (2011) show that trading platforms seeking to attract order flow have an incentive to relax price competition by differentiating along the speed dimension. Previously, the market microstructure literature has focused on the precision component of adverse selection, e.g., Kyle (1985), Back, Cao, and Willard (2000). In all these models, the behavior of the informed traders is similar to that of the informed trader in our benchmark model. In fact, we can describe our benchmark model as a mixture of Back and Pedersen (1998) and Chau and Vayanos (2008). From Back and Pedersen (1998) our benchmark model borrows the moving asset value; and from Chau and Vayanos

[8] This prediction can be tested in the cross-section of securities, if one has a proxy for the amount of public news that is released over time. It can also be tested in the time-series of a specific security, if there is time-variation in the amount of public news.

(2008) it borrows the periodic release of public information. In neither of these models the informed trader has a speed advantage. Our fast model contributes to the literature by showing that even an infinitesimal speed advantage for the informed trader results in a large difference in outcomes, e.g., speed causes a large participation rate of the informed trader, and an uncorrelated informed order flow.

The paper is organized as follows. Section 2 describes our two models: the benchmark model, and the fast model. The models are set in continuous time, but in Appendix A we present the corresponding discrete versions. Section 3 describes the resulting equilibrium price process and trading strategies, and compares the various coefficients involved. Section 4 discusses empirical implications of the model. Section 5 concludes.

# 2 Model

Trading occurs over the time interval $[0, 1]$. The risk-free rate is taken to be zero. During $[0, 1]$, a single informed trader ("he") and uninformed noise traders submit market orders to a competitive market maker ("she"), who sets the price at which the trading takes place. There is a risky asset with liquidation value $v_1$ at time 1. The informed trader learns about $v_1$ over time, and the expectation of $v_1$ conditional on his information available until time $t$ follows a Gaussian process given by

$$v_t \;=\; v_0 + \int_0^t \mathrm{d}v_\tau, \quad \text{with} \quad \mathrm{d}v_t \;=\; \sigma_v \, \mathrm{d}B_t^v, \tag{1}$$

where $v_0$ is normally distributed with mean 0 and variance $\Sigma_0$, and $B_t^v$ is a Brownian motion.[9] We refer to $v_t$ as the asset value or the fundamental value, and to $\mathrm{d}v_t$ as the innovation in asset value. Thus, the informed trader observes $v_0$ at time 0 and, at each

[9]This assumption can be justified economically as follows. First, define the asset value $v_t$ as the full information price of the asset, i.e., the price that would prevail at $t$ if all information until $t$ were to become public. Then, assume that (i) $v_t$ is a martingale (true, if the market is efficient), and (ii) $v_t$ is continuous (technically, it has continuous sample paths). Then, $v_t$ can be represented as an Itô integral with respect to a Brownian motion, by the Martingale Representation Theorem (see, e.g., Karatzas and Shreve (1991, Theorem 3.4.2)); our representation (1) is a simple particular case, with zero drift and constant volatility. But, even if $v_t$ has jumps (e.g., at Poisson-distributed random times), we conjecture that our key result of a non-zero $\mathrm{d}v_t$ component in the optimal trading strategy of the informed trader stays the same.

time $t + \mathrm{d}t \in [0,1]$ observes $\mathrm{d}v_t$.

The aggregate position of the informed trader at $t$ is denoted by $x_t$. The informed trader is risk-neutral and chooses $x_t$ to maximize expected utility at $t = 0$ given by

$$U_0 \;=\; \mathsf{E}\left[\int_0^1 (v_1 - p_{t+\mathrm{d}t})\,\mathrm{d}x_t\right] \;=\; \mathsf{E}\left[\int_0^1 (v_1 - p_t - \,\mathrm{d}p_t)\,\mathrm{d}x_t\right], \tag{2}$$

where $p_{t+\mathrm{d}t} = p_t + \mathrm{d}p_t$ is the price at which the order $\mathrm{d}x_t$ is executed.[10]

The aggregate position of the noise traders at $t$ is denoted by $u_t$, which is an exogenous Gaussian process given by

$$u_t \;=\; u_0 + \int_0^t \mathrm{d}u_\tau, \quad \text{with} \quad \mathrm{d}u_t \;=\; \sigma_v\,\mathrm{d}B_t^u, \tag{3}$$

where $B_t^u$ is a Brownian motion independent from $B_t^v$.

The market maker also learns about the asset value. At $t + \mathrm{d}t$, she receives a noisy signal of the innovation in asset value:

$$\mathrm{d}z_t \;=\; \mathrm{d}v_t + \mathrm{d}e_t, \quad \text{with} \quad \mathrm{d}e_t \;=\; \sigma_e\,\mathrm{d}B_t^e, \tag{4}$$

where $B_t^e$ is a Brownian motion independent from all the others. She does not observe the individual orders, but only the aggregate order flow

$$\mathrm{d}y_t \;=\; \mathrm{d}u_t + \mathrm{d}x_t. \tag{5}$$

Because the market maker is competitive and risk-neutral, at any time the price equals the conditional expectation of $v_1$ given the information available to her until that point. In the following, we will refer to the conditional expectation of $v_1$ just before trading takes place at time $t{+}\mathrm{d}t$ as the quote, and we denote it by $q_t$. One possible interpretation for the quote $q_t$ is that it is the bid-ask midpoint in a limit order book with zero tick size and zero bid-ask spread.[11] The conditional expectation of $v_1$ just after trading takes

[10]Because the optimal trading strategy of the informed trader might have a stochastic component, we cannot set $\mathsf{E}(\mathrm{d}p_t \mathrm{d}x_t) = 0$ as, e.g., in the Kyle (1985) model.

[11]This interpretation is correct if the price impact is increasing in the signed order flow and a zero order flow has zero price impact. These conditions are satisfied in the linear equilibrium we consider in

place at time $t + \mathrm{d}t$ is the execution price and is denoted by $p_{t+\mathrm{d}t}$.

We consider two different models: the *benchmark model* and the *fast model*, which differ according to the timing of information arrival and trading. A simplified timing of each model is presented in Table 1. Figure 2 shows the exact sequence of quotes and prices in each model.

**Table 1: Timing of events during $[t, t + \mathrm{d}t]$ in the benchmark model and in the fast model**

| Benchmark Model | Fast Model |
|---|---|
| 1. Informed trader observes $\mathrm{d}v_t$ | 1. Informed trader observes $\mathrm{d}v_t$ |
| 2. Market maker observes $\mathrm{d}z_t = \mathrm{d}v_t + \mathrm{d}e_t$ | 2. Trading |
| 3. Trading | 3. Market maker observes $\mathrm{d}z_t = \mathrm{d}v_t + \mathrm{d}e_t$ |

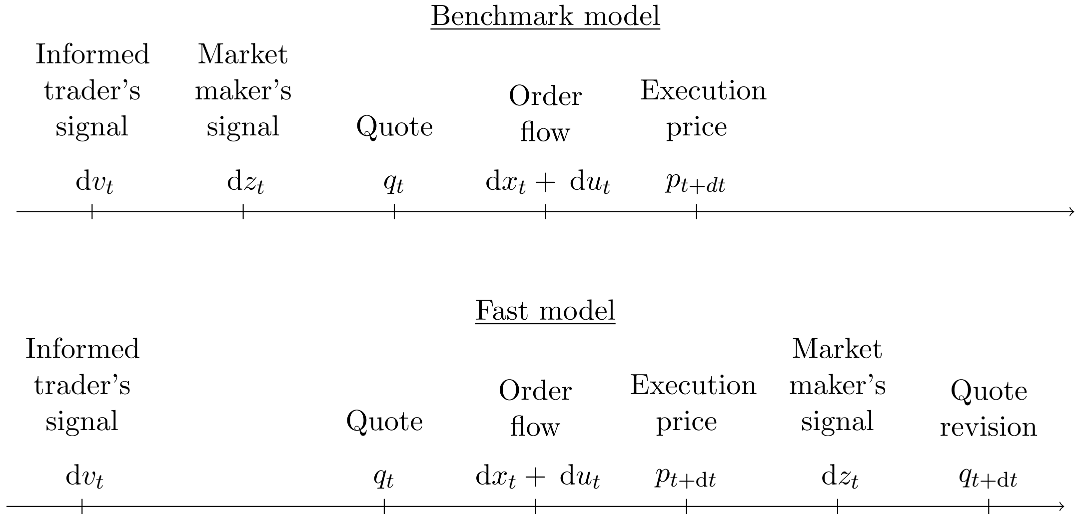
In the benchmark model, the order of events during the time interval $[t, t + \mathrm{d}t]$ is as follows. First, the informed trader observes $\mathrm{d}v_t$ and the market maker receives the signal $\mathrm{d}z_t$. The market maker sets the quote $q_t$ based on the information set $\mathcal{I}_t \cup \mathrm{d}z_t$, where $\mathcal{I}_t \equiv \{z_\tau\}_{\tau \leq t} \cup \{y_\tau\}_{\tau \leq t}$. The information set includes the order flow and the market maker's signal until time $t$, as well as the new signal $\mathrm{d}z_t$. Then, the informed trader submits a market order $\mathrm{d}x_t$ and noise traders also submit their order $\mathrm{d}u_t$. The information set of the market maker when she sets the execution price $p_{t+\mathrm{d}t}$ is $\mathcal{I}_t \cup \mathrm{d}z_t \cup \mathrm{d}y_t$ as it includes the new order flow.

In the fast model, the informed trader can move faster than the market maker. First, the informed trader observes $\mathrm{d}v_t$. Then, the market maker posts quotes before she observes her own signal. Therefore, the quote $q_t$ is based on the information set $\mathcal{I}_t$. The informed trader submits the market order $\mathrm{d}x_t$ along with the noise traders' orders $\mathrm{d}u_t$. The execution price $p_{t+\mathrm{d}t}$ is conditional on the information set $\mathcal{I}_t \cup \mathrm{d}y_t$. After trading has taken place, the market maker receives the signal $\mathrm{d}z_t$ and updates the quotes based on the information set $\mathcal{I}_t \cup \mathrm{d}z_t \cup \mathrm{d}y_t$. The new quote $q_{t+\mathrm{d}t}$ will be the prevalent quote in the next trading round.

The benchmark model is similar to models of the Kyle (1985) type. Formally, the

Section 3.

**Figure 2: Timing of events**




benchmark model is an extension of Back and Pedersen (1998) with the additional assumption that the market maker also learns about $\mathrm{d}v_t$. In all these versions of the Kyle model, the informed trader has more precise information than the market maker, but no speed advantage. By contrast, in the fast model, the informed trader has a speed advantage in addition to more precise information.

# 3 Equilibrium

The equilibrium concept is similar to that of Kyle (1985) or Back and Pedersen (1998). We look for linear equilibria defined as follows.

In the benchmark model, we look for an equilibrium in which the quote revision is linear in the market maker's signal

$$q_t \;=\; p_t + \mu_t \, \mathrm{d}z_t, \tag{6}$$

and the price impact is linear in the order flow

$$p_{t+\mathrm{d}t} \;=\; q_t + \lambda_t \, \mathrm{d}y_t. \tag{7}$$

In the fast model, we look for an equilibrium in which the price impact is linear in the order flow as in equation (7), and the subsequent quote revision is linear in the unexpected part of the market maker's signal[12]

$$q_{t+\mathrm{d}t} \;=\; p_{t+\mathrm{d}t} + \mu_t(\,\mathrm{d}z_t - \rho_t\,\mathrm{d}y_t). \tag{8}$$

In both models, we look for a strategy of the informed trader of the form

$$\mathrm{d}x_t \;=\; \beta_t(v_t - p_t)\,\mathrm{d}t + \gamma_t\,\mathrm{d}v_t, \tag{9}$$

i.e., we solve for $\beta_t$ and $\gamma_t$ so that the strategy defined in equation (9) maximizes the informed trader's expected profit (2). In Appendix A, we use the discrete time version of both models to show that, as long as the equilibrium has a linear pricing rule, the optimal strategy of the informed trader has the same form as in (9).[13]

In what follows, we refer to the first term of trading strategy, $\beta_t(v_t - p_t)\,\mathrm{d}t$, as *level trading*, as it consists in trading on the difference between the level of the asset value and the price level. This term appears in essentially all models of trading of the Kyle (1985) type, such as Back and Pedersen (1998), Back, Cao, and Willard (2000), etc. The second term of the trading strategy, $\gamma_t\,\mathrm{d}v_t$, consists in trading on the innovation of the asset value, and we call it *flow trading*. The next result shows that flow trading is zero in the benchmark model, but nonzero in the fast model.

**Theorem 1.** *In the benchmark model there is a unique linear equilibrium:*

$$\mathrm{d}x_t \;=\; \beta_t^B(v_t - p_t)\,\mathrm{d}t + \gamma_t^B\,\mathrm{d}v_t, \tag{10}$$

$$\mathrm{d}p_t \;=\; \mu_t^B\,\mathrm{d}z_t + \lambda_t^B\,\mathrm{d}y_t, \tag{11}$$

[12]In the fast model, the market maker's signal $\mathrm{d}z_t$ is predictable from the order flow $\mathrm{d}y_t$, thus the quote update is done only using the unexpected part of $\mathrm{d}z_t$.

[13]In fact, in discrete time the optimal strategy has $q_t$ instead of $p_t$. But because the difference between $p_t$ and $q_t$ is infinitesimal, the difference vanishes in continuous time when multiplying by $\mathrm{d}t$. In the proof of Theorem 1, we use $p_t$ for the benchmark model, and $q_t$ for the fast model, since these are well defined Itô processes with the same type of increment, $\lambda_t\,\mathrm{d}y_t + \mu_t(\,\mathrm{d}z_t - \rho_t\,\mathrm{d}y_t)$.

*with coefficients given by*

$$\beta_t^B = \frac{1}{1-t}\,\frac{\sigma_u}{\Sigma_0^{1/2}}\left(1+\frac{\sigma_v^2\sigma_e^2}{\Sigma_0(\sigma_v^2+\sigma_e^2)}\right)^{1/2}, \tag{12}$$

$$\gamma_t^B = 0, \tag{13}$$

$$\lambda_t^B = \frac{\Sigma_0^{1/2}}{\sigma_u}\left(1+\frac{\sigma_v^2\sigma_e^2}{\Sigma_0(\sigma_v^2+\sigma_e^2)}\right)^{1/2}, \tag{14}$$

$$\mu_t^B = \frac{\sigma_v^2}{\sigma_v^2+\sigma_e^2}. \tag{15}$$

*In the fast model there is a unique linear equilibrium:*[14]

$$\mathrm{d}x_t = \beta_t^F(v_t-q_t)\,\mathrm{d}t+\gamma_t^F\,\mathrm{d}v_t, \tag{16}$$

$$\mathrm{d}q_t = \lambda_t^F\,\mathrm{d}y_t+\mu_t^F(\,\mathrm{d}z_t-\rho_t^F\,\mathrm{d}y_t), \tag{17}$$

*with coefficients given by*

$$\beta_t^F = \frac{1}{1-t}\,\frac{\sigma_u}{(\Sigma_0+\sigma_v^2)^{1/2}}\,\frac{1}{\left(1+\frac{\sigma_e^2}{\sigma_v^2}f\right)^{1/2}}\left(1+\frac{(1-f)\sigma_v^2}{\Sigma_0}\,\frac{1+\frac{\sigma_e^2}{\sigma_v^2}+\frac{\sigma_e^2}{\sigma_v^2}f}{2+\frac{\sigma_e^2}{\sigma_v^2}+\frac{\sigma_e^2}{\sigma_v^2}f}\right), \tag{18}$$

$$\gamma_t^F = \frac{\sigma_u}{\sigma_v}\,f^{1/2} = \frac{\sigma_u}{(\Sigma_0+\sigma_v^2)^{1/2}}\,\frac{\left(1+\frac{\sigma_e^2}{\sigma_v^2}f\right)^{1/2}(1+f)}{2+\frac{\sigma_e^2}{\sigma_v^2}+\frac{\sigma_e^2}{\sigma_v^2}f}, \tag{19}$$

$$\lambda_t^F = \frac{(\Sigma_0+\sigma_v^2)^{1/2}}{\sigma_u}\,\frac{1}{\left(1+\frac{\sigma_e^2}{\sigma_v^2}f\right)^{1/2}(1+f)}, \tag{20}$$

$$\mu_t^F = \frac{1+f}{2+\frac{\sigma_e^2}{\sigma_v^2}+\frac{\sigma_e^2}{\sigma_v^2}f}, \tag{21}$$

$$\rho_t^F = \frac{\sigma_v^2}{\sigma_u(\Sigma_0+\sigma_v^2)^{1/2}}\,\frac{(1+\frac{\sigma_e^2}{\sigma_v^2}f)^{1/2}}{2+\frac{\sigma_e^2}{\sigma_v^2}+\frac{\sigma_e^2}{\sigma_v^2}f}, \tag{22}$$

*and $f$ is the unique root in $(0,1)$ of the cubic equation*

$$f = \frac{\left(1+\frac{\sigma_e^2}{\sigma_v^2}f\right)(1+f)^2}{\left(2+\frac{\sigma_e^2}{\sigma_v^2}+\frac{\sigma_e^2}{\sigma_v^2}f\right)^2}\,\frac{\sigma_v^2}{\sigma_v^2+\Sigma_0}. \tag{23}$$

[14]Note that the level trading component in (16) has $q_t$ instead of $p_t$. This is the same formula, since (8) implies $(p_t-q_t)\,\mathrm{d}t=0$. We use $q_t$ as a state variable, because it is a well defined Itô process.

*In both models, when $\sigma_v \to 0$, the equilibrium converges to the unique linear equilibrium in the continuous time version of Kyle (1985).*

*Proof.* See Appendix B. □

To give some intuition for the theorem, note that in both models the optimal strategy of the informed trader has a non-zero level trading component. This is because in both models the informed trader receives more precise signals than the market maker: the informed trader knows $v_t$ exactly, while the market maker's best forecast is $p_t$. Therefore, it is optimal for the informed trader to trade on the forecast error of the market maker $v_t - p_t$. This forecast error is slowly moving, because its change is of the order of $(\mathrm{d}v_t - \mathrm{d}p_t)\,\mathrm{d}t$, which at high frequencies is negligible. The informed trader trades smoothly on the forecast error, in the sense that the level trading component is of the order $\mathrm{d}t$.

The key difference between the two models is that only in the fast model the informed trader's optimal strategy has a flow trading component. The reason is that in the benchmark model, when the trader submits the order $\mathrm{d}x_t$, all the signals $\{\,\mathrm{d}v_\tau\}_{\tau \le t}$ that he has received are given the same weight in the optimal strategy. By contrast, in the fast model the marker maker has not incorporated the signal $\mathrm{d}z_t = \mathrm{d}v_t + \mathrm{d}e_t$ in the price yet. Therefore, it is optimal for the informed trader to trade aggressively on $\mathrm{d}v_t$ before the market maker receives information $\mathrm{d}z_t$.

The flow trading component is volatile since it is an innovation in a random walk process. It also generates a much larger order flow than the level trading component, because it is of order $\mathrm{d}t^{1/2}$.

We give some comparative statics for the coefficients from Theorem 1.

**Proposition 1.** *In the context of Theorem 1, for all values of the parameters we have the following inequalities:*

$$\beta_0^F < \beta_0^B \tag{24}$$

$$\lambda^F > \lambda^B \tag{25}$$

$$\mu^F < \mu^B. \tag{26}$$

*In both the benchmark equilibrium and the fast equilibrium,*

$\beta_0$ *increases in* $\sigma_v$, $\sigma_u$, $\sigma_e$*; and decreases in* $\Sigma_0$*;*

$\lambda$ *increases in* $\sigma_v$, $\sigma_e$, $\Sigma_0$*; and decreases in* $\sigma_u$*;*

$\mu$ *increases in* $\sigma_v$*; decreases in* $\sigma_e$*; and is constant in* $\sigma_u$*;*

$\mu$ *is constant in* $\Sigma_0$ *in the benchmark, but decreases in* $\Sigma_0$ *in the fast equilibrium.*

*Additionally, in the fast equilibrium,*

$\gamma$ *increases in* $\sigma_v$, $\sigma_u$*; and decreases in* $\sigma_e$, $\Sigma_0$*;*

$\rho$ *increases in* $\sigma_v$*; and decreases in* $\sigma_u$, $\sigma_e$, $\Sigma_0$.

*Proof.* See Appendix B. □

The intuition for some of these comparative statics is discussed in the next section.

# 4 Empirical Implications

## 4.1 High Frequency News Trading

In this section we argue that the informed trader of the fast model shares some of the characteristics that are attributed to the broad category of High Frequency Traders (HFTs). Specifically, we show that the informed trader (i) is responsible for a large fraction of the order flow; (ii) his order flow exhibits low serial correlation; and (iii) he engages in anticipatory trading. This is not to say that our model can be applied to study all types of HFTs. Indeed, the spectrum of strategies which can be classified under the umbrella of high frequency trading is quite large.[15] Our paper focuses on one of these strategies, namely, high frequency news trading (HFNT). Therefore, the empirical predictions and policy implications of our model apply to HFNT, but not necessarily to other categories of HFT.

[15]For instance, SEC (2010) places high frequency trading under four categories: (a) Passive Market Making, which generates large volumes by submitting and canceling many limit orders; (b) Arbitrage, which is based on correlation strategies (statistical arbitrage, pairs trading, index arbitrage, etc.); (c) Structural, which involves identifying and exploiting other market participants that are slow; and (d) Directional, which implies taking significant, unhedged positions based on anticipation of intraday price movements.

First, we define the *Informed Participation Rate (IPR)* as the contribution of the informed trader to total order flow

$$IPR_t \;=\; \frac{\mathsf{Var}(\,\mathrm{d}x_t)}{\mathsf{Var}(\,\mathrm{d}y_t)} \;=\; \frac{\mathsf{Var}(\,\mathrm{d}x_t)}{\mathsf{Var}(\,\mathrm{d}u_t) + \mathsf{Var}(\,\mathrm{d}x_t)}. \tag{27}$$

**Proposition 2.** *The informed participation rate is zero in the benchmark while it is positive in the fast model,*

$$IPR_t^B \;=\; 0, \qquad IPR_t^F \;=\; \frac{f}{1+f}, \tag{28}$$

*where $f$ is defined in Theorem 1.*

*Proof.* See Appendix B. □

In the benchmark model, the informed trader's optimal strategy has only a level trading component. The level trading component consists in a drift in the asset holding $x_t$. This generates a trading volume that is an order of magnitude smaller than the trading volume generated by the noise traders. Formally, informed trading volume is of the order $\mathrm{d}t$ while noise trading volume is of the order $\mathrm{d}t^{1/2}$. By contrast, in the fast model, the informed trader's optimal strategy includes a flow trading component. The flow trading component is volatile (i.e., stochastic), which generates a trading volume that is of the same order of magnitude as the noise trading volume.

In the discrete time version of the model, informed trading volume is non zero but it converges quickly to zero as the trading frequency increases. In Figure 1 in the Introduction, we have already seen that in the benchmark model the trading volume is already small when trading takes place at the daily frequency. In the fast model, informed trading volume converges to a limit between zero and one when the trading frequency becomes very large.

Next, we consider the serial correlation of the informed order flow.

**Proposition 3.** *The autocorrelation of the informed order flow is positive in the bench-*

*mark while it is zero in the fast model: for $\tau > 0$,*

$$\mathsf{Corr}(\,\mathrm{d}x_t^B,\,\mathrm{d}x_{t+\tau}^B) \;=\; \left(\frac{1-t-\tau}{1-t}\right)^{\frac{1}{2}+\lambda^B\beta_0^B} \;>\; 0, \tag{29}$$

$$\mathsf{Corr}(\,\mathrm{d}x_t^F,\,\mathrm{d}x_{t+\tau}^F) \;=\; 0. \tag{30}$$

*Proof.* See Appendix B. □

The intuition behind Proposition 3 is that the level trading component is slowly moving, i.e., it is serially correlated. This explains why the informed order flow is positively autocorrelated in the benchmark model. By contrast, the flow trading component is not serial correlated as it only depends on the current innovation in the asset value. Since the flow trading component generates a much larger order flow than the level component, the autocorrelation of the informed order flow is zero in the fast model. Note that the fact that the autocorrelation is exactly zero may depend on the specific assumptions of the model, e.g., the informed trader has no inventory cost. The more robust result related to Proposition 3 is that the serial correlation of the informed order flow is lower in the fast model than in the benchmark.

The empirical evidence about HFT order flow autocorrelation is mixed. For instance, using US stock trading data aggregated across all HFTs, Brogaard (2011) and Hendershott and Riordan (2011) find a positive autocorrelation of the aggregate HFT order flow. By contrast, Menkveld (2011) using data on a single HFT on the European stock market, and Kirilenko, Kyle, Samadi, and Tuzun (2011) using data on the Flash Crash of May 2010, find clear evidence of mean reverting inventories. These opposite results reflect the fact that HFT strategies are diverse and may come in a wide variety of autocorrelation patterns.[16] Empirical studies which consider HFTs as a whole measures the average autocorrelation across all types of HFT strategies, and HFNT is only one of those.

Finally, we measure *Anticipatory Trading (AT)* by the correlation between the in-

[16]In addition, the definition of HFTs can introduce a bias. For instance, in Brogaard (2011), Hendershott and Riordan (2011), and Kirilenko et al. (2011), one of the criteria to classify a trader as HFT is that it keeps its inventories close to zero.

formed order flow and the next instant return,

$$AT_t \;=\; \mathsf{Corr}(\,\mathrm{d}x_t, q_{t+\,\mathrm{d}t} - p_{t+\,\mathrm{d}t}), \tag{31}$$

where we recall that $p_{t+\,\mathrm{d}t}$ is the price at which the order flow $\mathrm{d}x_t$ is executed, and $q_{t+\,\mathrm{d}t}$ is the next quote revision that takes place when the market maker receives her next signal.

**Proposition 4.** *Anticipatory trading is zero in the benchmark while it is positive in the fast model*

$$AT_t^B \;=\; 0, \qquad AT_t^F \;=\; \frac{(1-\rho^F\gamma^F)\sigma_v}{\sqrt{(1-\rho^F\gamma^F)^2\sigma_v^2 + \sigma_e^2 + (\rho^F)^2\sigma_u^2}} > 0. \tag{32}$$

There is anticipatory trading in the fast model because the flow trading component of the strategy anticipates the very next quote revision. This is consistent with Kirilenko et al. (2011) and Hendershott and Riordan (2011) who find that, on average over all categories of high frequency trading strategies, HFTs' aggressive orders are correlated with future price changes at a short horizon.

## 4.2 The Effect of High Frequency News Trading

In this section we study the effect of HFNT on several market outcomes: liquidity, price discovery, price volatility, and price impact. To do that, we compare the equilibrium of the market when one moves from the benchmark to the fast model. Indeed, in the fast model, the informed trader is able to access information and trade based on it quickly, that is, before the information is incorporated into quotes. In practice, this can occur because the exchange increases automation, offers co-location services, or implements any other change that lowers the execution time for market orders. Alternatively, one can view a shift from the fast model to the benchmark model as the result of a move by the regulator or the trading platform to dampen HFNT.

We already proved the following result in Proposition 1:

**Corollary 1.** *Liquidity is lower in the fast model than in the benchmark:* $\lambda^F > \lambda^B$.

The market is less liquid in the fast model since there is more adverse selection than in the benchmark model. Indeed, the informed trader has more precise information in both models, and, in the fast model only, the informed is also faster. This generates a second source of adverse selection, coming from speed.

Previous empirical work has investigated the effect of high frequency trading in general on liquidity. Some find evidence of a positive (e.g., Hendershott, Jones, and Menkveld (2011), Hasbrouck and Saar (2010)) while others find the opposite (e.g., Hendershott and Moulton (2011)). These papers have considered high frequency traders as a group, and have therefore measured their average impact across the entire spectrum of HFT strategies. We predict that HFNT reduces liquidity, but it may be the case that high frequency market making improves the liquidity, and that the overall effect on liquidity is positive.

Another measure of the price impact of trades is the *Cumulative Price Impact (CPI)* defined as the covariance between the informed order flow trade per unit of time at $t$ and the subsequent price change over the time interval $[t, t+\tau]$ for $\tau > 0$:[17]

$$CPI_t(\tau) \;=\; \mathsf{Cov}\left(\frac{\mathrm{d}x_t}{\mathrm{d}t}, p_{t+\tau} - p_t\right). \tag{33}$$

Because the optimal strategy of the informed trader is of the type $\mathrm{d}x_t = \beta_t(v_t - p_t)\,\mathrm{d}t + \gamma_t\,\mathrm{d}v_t$, the cumulative price impact can be decomposed into two terms:

$$CPI_t(\tau) \;=\; \beta_t\,\mathsf{Cov}(v_t - p_t, p_{t+\tau} - p_t) + \frac{1}{\mathrm{d}t}\gamma_t\,\mathsf{Cov}(\,\mathrm{d}v_t, p_{t+\tau} - p_t), \tag{34}$$

and note that the second term is well defined, because $\mathsf{Cov}(\,\mathrm{d}v_t, p_{t+\tau} - p_t)$ is of the order of $\mathrm{d}t$, since the asset value, $v_t$, is a Gaussian process.

**Proposition 5.** *In the benchmark model, the cumulative price impact is*

$$CPI_t^B(\tau) \;=\; k_1^B\left[1 - \left(1 - \frac{\tau}{1-t}\right)^{\lambda^B\beta_0^B}\right], \tag{35}$$

[17] Using $p_t$ or $q_t$ in the definition of $CPI_t(\tau)$ is equivalent because the difference between the two is smaller than $p_{t+\tau} - p_t$ by an order of magnitude.

*while in the fast model it is*

$$CPI_t^F(\tau) \;=\; k_0^F + k_1^F \left[1 - \left(1 - \frac{\tau}{1-t}\right)^{(\lambda^F - \mu^F \rho^F)\beta_0^F}\right], \tag{36}$$

*where*

$$\begin{aligned} k_1^B &= \beta_0^B \Sigma_0, & (37)\\ k_0^F &= \gamma^F((\lambda^F - \mu^F \rho^F)\gamma^F + \mu^F)\sigma_v^2, & (38)\\ k_1^F &= \beta_0^F \Sigma_0 + \gamma^F(1 - (\lambda^F - \mu^F \rho^F)\gamma^F - \mu^F)\sigma_v^2. & (39) \end{aligned}$$

*Proof.* See Appendix B. □

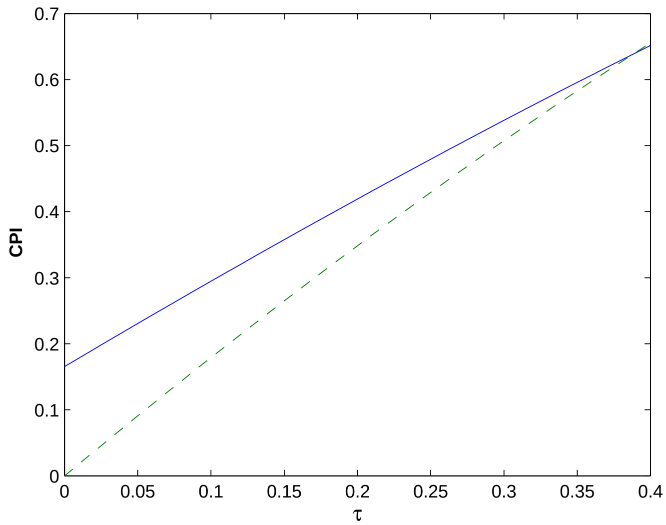
One can see from the formulas, or from Figure 3, that in the benchmark model the cumulative price impact starts from near-zero values when $\tau$ is very small, while in the fast model it starts from a positive value, $k_0^F$. Then, the cumulative price builds up over time in both models, because the level trading component is correlated with all prices changes in the future. To sum up, the intercept in Figure 3 is evidence of flow trading, while the positive slope is evidence of level trading. Note that the cumulative price impact is a univariate covariance. If we want to obtain a causal impact of trades, we need to control for the future order flow. This can be done using a VAR model, as will be shown in Section 4.4.

Next, we consider the effect of HFNT on the price discovery process. We define price informativeness at any given point in time $t$ as the (squared) pricing error

$$\Sigma_t \;=\; \mathsf{E}\big((v_t - p_t)^2\big). \tag{40}$$

More insight can be gained by decomposing this pricing error into errors about the last change in asset value and errors about the level of the asset value. First, we note that (40) can rewritten as follows:

**Figure 3: Cumulative Price Impact at Different Horizons.** The figure plots the cumulative price impact at $t = 0$, $\mathsf{Cov}\left(\frac{\mathrm{d}x_0}{\mathrm{d}t}, p_\tau - p_0\right)$ against the horizon $\tau \in (0, 1]$ in (a) the benchmark model, with a dotted line; and (b) the fast model, with a solid line. The parameters used are $\sigma_u = \sigma_v = \sigma_e = \Sigma_0 = 1$ (see Theorem 1). The liquidation date $t = 1$ corresponds to 10 calendar years.




**Lemma 1.** *In both the benchmark and the fast models,*

$$\Sigma_t \;=\; (1+t)\Sigma_0 + 2t\sigma_v^2 - 2\int_0^t \mathsf{Cov}(\,\mathrm{d}p_\tau, v_{\tau+\,\mathrm{d}\tau}). \tag{41}$$

*Proof.* See Appendix B. □

Intuitively, if price changes are more correlated with the asset value ($\mathsf{Cov}(\mathrm{d}p_\tau, v_{\tau+\,\mathrm{d}\tau})$ is larger), the price ends up being closer on average to the asset value ($\Sigma_t$ is smaller). Moreover, we have the following decomposition:[18]

$$\mathsf{Cov}(\,\mathrm{d}p_t, v_{t+\mathrm{d}t}) \;=\; \mathsf{Cov}(\,\mathrm{d}p_t, v_t) + \mathsf{Cov}(\,\mathrm{d}p_t,\,\mathrm{d}v_t). \tag{42}$$

**Proposition 6.** $\mathsf{Cov}(\mathrm{d}p_t, \mathrm{d}v_t)$ *is higher in the fast model than in the benchmark; while* $\mathsf{Cov}(\mathrm{d}p_t, v_t)$ *is higher in the benchmark than in the fast model.* $\Sigma_t$ *is the same in both the benchmark and the fast models.*

Returns are more informative about the level of the asset value in the benchmark

[18]In this equation, $\mathrm{d}p_t$ denotes $p_{t+\mathrm{d}t} - p_t$ in the benchmark model, and $q_{t+\mathrm{d}t} - q_t$ in the fast model.

model, while they are more informative about changes in the asset value in the fast model. The reason for the latter comes from the flow trading component. In the benchmark, the contemporaneous correlation between changes in the price and in the asset value comes from quote revisions only:

$$\mathsf{Cov}(\,\mathrm{d}p_t^B,\,\mathrm{d}v_t) \;=\; \mathsf{Cov}(\mu^B\,\mathrm{d}z_t,\,\mathrm{d}v_t) \;=\; \mu^B\sigma_v^2\,\mathrm{d}t. \tag{43}$$

In the fast model, flow trading adds to this covariance:

$$\mathsf{Cov}(\mathrm{d}p_t^F,\mathrm{d}v_t) \;=\; \mathsf{Cov}(\lambda^F\mathrm{d}x_t^F+\mu^F(\mathrm{d}z_t-\rho^F\,\mathrm{d}x_t),\mathrm{d}v_t) \;=\; \big(\mu^B+(\lambda^F-\mu^F\rho^F)\big)\sigma_v^2\mathrm{d}t. \tag{44}$$

It implies that returns are more correlated with the innovations of the asset value in the fast model.

By contrast, the covariance of returns with the level of the asset value is higher in the benchmark model. The reason is that the level component of informed trading is less intense in the fast model than in the benchmark model. Indeed, there is a substitution between flow trading and level trading. The intuition for this *substitution effect* is that the informed trader competes with himself when using his information advantage. Trading more on news now consumes the profit from trading on the level in the future. Therefore, when flow trading increases in the fast model, level trading has to decrease.

In terms of total pricing error, these two effects—higher correlation of returns with changes and lower correlation with levels—exactly cancel out, and the pricing error is the same in both models. In the fast model, new information is incorporated more quickly into the price while older information is incorporated less quickly, leaving the total pricing error equal in both models. The more formal reason why these two effects exactly offset each other is that, in both the benchmark and the fast models, the informed trader finds it optimal to release information at a constant rate to minimize price impact. Therefore, $\Sigma_t$ decreases linearly over time in both models. Moreover, the transversality condition for optimization requires that no money is left on the table at $t = 1$, i.e., $\Sigma_1 = 0$. Since the initial value $\Sigma_0$ is exogenously given, the evolution of $\Sigma_t$ is the same in both models.

We now consider the effect of HFNT on price volatility. Following Hasbrouck (1991a, 1991b) we decompose price volatility into the volatility coming from trades and the volatility coming from quotes:

$$\mathsf{Var}(\mathrm{d}p_t) \;=\; \mathsf{Var}(p_{t+\,\mathrm{d}t} - q_t) + \mathsf{Var}(q_t - p_t). \tag{45}$$

The first term of the decomposition if the variance of the price impact of trades $(p_{t+\mathrm{d}t} - q_t)$. The second term of the decomposition is the variance of quote revisions unrelated to trading $(q_t - p_t)$.

**Proposition 7.** $\mathsf{Var}(p_{t+\mathrm{d}t} - q_t)$ *is higher in the fast model than in the benchmark; while* $\mathsf{Var}(q_t - p_t)$ *is higher in the benchmark than in the fast model.* $Var(\mathrm{d}p_t)$ *is the same in benchmark and in the fast models and it equals*

$$\mathsf{Var}(\,\mathrm{d}p_t) \;=\; \sigma_v^2 + \Sigma_0. \tag{46}$$

More information is incorporated through trading in the fast model. This is because the informed trader acts on the news before the market marker revises the quotes. Therefore, trading is more intense and price volatility coming from trades is higher in the fast model. The flip side is that the quote revision is less intense, and the price volatility coming from quotes is lower in the fast model.

In terms of total price volatility, these two effects cancel each other and price volatility is the same in both models. The reason why the two effects exactly offset each other is that in an efficient market price changes are not autocorrelated. Therefore, the short-term price variance per unit of time is always equal to the long-term price variance per unit of time, which is itself equal to the variance per unit of time of the (exoegenous) asset value.

## 4.3 The Determinants of High Frequency News Trading

Because we identify HFNT with the activity of the informed trader in the fast model, in this section we study the determinants of HFNT by doing comparative statics on various parameters in the fast model. We measure HFNT activity by the informed participation rate defined in Equation (27).

Consider first the effect of the precision of public news. Holding constant the variance of the innovation of the asset value $\sigma_v^2$, more precise public news about the changes in asset value amounts to a lower $\sigma_z^2 = \sigma_v^2 + \sigma_e^2$, or, equivalently, a lower $\sigma_e^2$.

**Proposition 8.** *An increase in the precision of public news, i.e., a decrease in $\sigma_e$, increases HFNT activity (increases $IPR_t^F$) and improves liquidity (decreases $\lambda^F$).*

*Proof.* By Propositon 2, $IPR^F = \frac{f}{1+f}$, thus the informed participation rate in the fast model has the same dependence on $\sigma_e$ as $f$. From (19), $\gamma^F = \frac{\sigma_u}{\sigma_v} f^{1/2}$, thus $f$ has the same dependence on $\sigma_e$ as $\gamma^F$. Therefore, $IPR^F$ has the same depdendence on $\sigma_e$ as $\gamma^F$. But Proposition 1 shows that $\gamma^F$ is decreasing in $\sigma_e$. Finally, we use again Proposition 1 to show that $\lambda^F$ is increasing in $\sigma_e$. □

The fast trader needs a precise news environment in order to take advantage of anticipatory trading. Otherwise, if the public signal is imprecise, i.e. $\sigma_e$ is large, the market maker does not adjust quotes by much ($\mu^F$ is small), the informed trader cannot benefit much from his speed advantage and does not trade intensely on the news component. This prediction can be tested in the cross-section of securities, if one has a proxy for the amount of public news that is released over time. It can also be tested in the time-series of a specific security, if there is time-variation in the amount of public news.

As stated in Proposition 8, more public news also improves liquidity because it reduces adverse selection. Interestingly, it implies that if the amount of public news changes (over time or across securities) then HFNT and liquidity move in the same direction. This is not because HFNT improves liquidity; instead, this is because more public news increases both HFNT and liquidity.

Next, we consider the effect of price volatility. From Equation (46), $\mathsf{Var}(\mathrm{d}p_t) = \sigma_v^2 + \Sigma_0$, thus we model an increase in price volatility as an increase in the variance of the innovation of the asset value, $\sigma_v^2$, while holding costant the relative precision of public news, i.e., the ratio $\sigma_e^2/\sigma_v^2$. We can prove the following result.

**Proposition 9.** *An increase in price volatility (higher $\sigma_v$ while holding $\sigma_e/\sigma_v$ constant) increases HFNT activity (increases $IPR_t^F$) and reduces liquidity (increases $\lambda^F$).*

Because the informed trader acts in anticipation of price changes, more volatility increases the intensity of flow trading, and therefore the informed participation rate ($IPR$), or HFNT activity. As a result, there is more adverse selection, and liquidity is thus negatively affected.

## 4.4 Methodological Issues in Empirical Analysis of HFNT

Our framework can be used to shed light on some methodological issues in the empirical analysis of HFNT. In order to make our model more comparable to econometric models, we consider the discrete time version of our continuous time model, as in Appendix A. It works very similarly to the continuous time model, the main difference being that the infinitesimal time interval $\mathrm{d}t$ is replaced by a real number $\Delta t > 0$. We also consider that $\Delta t$ is small and we approximate the equilibrium variables ($\beta_t$, $\gamma_t$, $\lambda_t$, $\mu_t$, $\rho_t$) in the discrete time model by their continuous time counterpart. Letting $T = \frac{1}{\Delta t}$ be the number of trading periods, time is indexed by $t = 0, 1, \ldots, T-1$. The informed order flow at time $t$ is equal to

$$\Delta x_t \;=\; \beta_t(v_t - q_t)\Delta t + \gamma_t \Delta v_t, \tag{47}$$

where $q_t$ is the quote just before the order flow arrives, and $p_{t+1}$ is the execution price.

### 4.4.1 Timing Issues in Defining Returns

There are several issues when one measures returns empirically. For instance, when returns are computed from trade to trade, the econometrician can either use the transaction price, or the mid-quote just after the trade, or the bid or the ask depending on the direction of the order flow, or the mid-quote after the next quote revision, etc. Lags in trade reporting and time aggregation of data can also impose constraints on how trade-to-trade returns are defined. To emphasize the consequence of these timing assumptions, we contrast two different definitions of returns in the context of our model.

A first option is to compute returns using the quotes just after the order is filled ("post-trade quotes"). With this assumption, the return contemporaneous to the order flow $\Delta x_t + \Delta u_t$ is $r_t = p_{t+1} - p_t$. A second possibility is to compute returns using the quotes just before the next trade takes place ("pre-trade quotes"). In this case, the return contemporaneous to the order flow $\Delta x_t + \Delta u_t$ is $r_t = q_{t+1} - q_t$.

To illustrate the implications of these two assumptions for the empirical analysis, we consider the following VAR model with $K \geq 1$ lags in the spirit of Hasbrouck (1996):[19]

$$
\begin{aligned}
r_t &= \sum_{k=1}^{K} a_k r_{t-k} + \sum_{k=0}^{K} b_k \, \mathrm{d}x_{t-k} + \sum_{k=0}^{K} c_k \, \mathrm{d}u_{t-k} + \varepsilon_t, && (48)\\
dx_t &= \sum_{k=1}^{K} d_k r_{t-k} + \sum_{k=1}^{K} e_k \, \mathrm{d}x_{t-k} + \sum_{k=1}^{K} f_k \, \mathrm{d}u_{t-k} + \eta_t, && (49)\\
du_t &= \sum_{k=1}^{K} g_k r_{t-k} + \sum_{k=1}^{K} h_k \, \mathrm{d}x_{t-k} + \sum_{k=1}^{K} i_k \, \mathrm{d}u_{t-k} + \zeta_t. && (50)
\end{aligned}
$$

We compute the coefficients of the VAR model under the two timing assumptions. To allege the notations, we now omit the superscript $^F$ when we refer to the equilibrium variables in the fast model.

**Proposition 10.** *When post-trade quotes are used:* $b_0 = c_0 = \lambda$, $b_1 = \mu(1 - \rho\gamma)/\gamma$, $c_1 = -\mu\rho$, *and all other coefficients are zero. When pre-trade quotes are used:* $b_0 = \lambda - \mu\rho + \mu/\gamma$, $c_0 = \lambda - \mu\rho$, *and all other coefficients are zero.*

*Proof.* See Appendix B. □

Depending on how returns are measured, the estimated $b_1$ may be positive or equal to zero. When returns are computed using post-trade quotes, the informed order flow is positively related to the next period return ($b_1 > 0$). The economic interpretation is that the informed trader engages in anticipatory trading. By contrast, when returns are measured using pre-trade quotes, $b_1 = 0$ because the time $t$ order flow is incorrectly considered as being contemporaneous to the subsequent quote revision $q_{t+1} - p_{t+1}$. In this case, we fail to reject the incorrect null hypothesis of no anticipatory trading. This

[19] This specification is used, e.g., by Brogaard (2010).

suggests that using the quotes immediately after trading takes place, or the price at which the last unit of the order flow is executed, may be necessary to detect anticipatory trading in the data.

### 4.4.2 Sampling Issues

It is customary to aggregate data over time. This can be due to limited data availability, or it may be a deliberate choice of the econometrician to make data analysis more manageable. In this section we look at the consequence of time aggregation and we show that the time interval at which data are aggregated affects the results of the empirical analysis. In particular, when the sampling frequency is low relative to the trading frequency, the empirical moments are biased in the sense that they differ from the theoretical moments of the model.

Assume that each observation in the data spans $n \geq 1$ trading rounds. In this case, the data are a time-series of length $T/n$. For $j = 1, \ldots, \frac{T}{n}$, the $j$th observation corresponds to trading during the $n$ trading rounds starting at time $t = (j-1)n$. The order flow of the informed trader is $\Delta x_j(n) \equiv \Delta x_t + \cdots + \Delta x_{t+n-1}$, and, assuming that prices are defined as post-trade quotes, the return is $r_j(n) \equiv p_{t+n} - p_t$.

First, we consider the measure of anticipatory trading defined in equation (31). Its empirical counterpart when data are sampled every $n$ trading rounds is

$$AT_j(n) \; = \; \mathsf{Corr}(\Delta x_j(n), r_{j+1}(n)). \tag{51}$$

**Proposition 11.** *The empirical measure of anticipatory trading $AT_j(n)$ decreases with $n$ and converges to zero when $n \to +\infty$.*

*Proof.* See Appendix B. □

The aggregated order flow spans $n$ trading periods. Moreover, each trade anticipates news that is incorporated in the quotes in the next trading round. Therefore, only the last trade of the aggregated order flow $\Delta x_j(n)$ is correlated with the next aggregated return $r_{j+1}(n)$. As a result, when $n$ increases, the correlation between $\Delta x_j(n)$ and $r_{j+1}(n)$ decreases. When $n$ becomes too large, the correlation becomes almost zero.

This result suggests that sampling data at a sufficiently high frequency is important for detecting anticipatory trading.

We now turn to the informed participation rate (27). Its empirical counterpart when data are sampled each $n$ trading periods can be defined as

$$IPR_j(n) \;=\; \frac{\mathsf{Var}(\Delta x_j(n))}{\mathsf{Var}(\Delta x_j(n)) + \mathsf{Var}(\Delta u_j(n))}. \tag{52}$$

In order to obtain closed-form formula solutions, we consider the limit case where the trading frequency is large, holding fixed the time interval $\tau = n\Delta t$ at which data are aggregated. In this case, the informed participation can therefore be written as a function of $\tau$: $IPR_j(\tau) = \lim_{\Delta t \to 0} IPR_j(\tau/\Delta t)$.

**Proposition 12.** *The empirical informed participation rate $IPR_j(\tau)$ increases with the sampling interval $\tau$.*

*Proof.* See Appendix B. □

The level trading component is positively autocorrelated over time. Therefore, the variance of the informed order flow increases faster than the time horizon $\tau$ at which the variance is computed. Since the noise trading order flow is serially uncorrelated, the fraction of the order flow variance due to the informed trader increases with the sampling interval $\tau$.

Finally, consider $\mathsf{Corr}(\Delta x_j(n), \Delta x_{j+1}(n))$, the empirical autocorrelation of the informed order flow. Again, in order to obtain closed-form formulas, we hold constant the sampling interval $\tau = n\Delta t$ and we let $\Delta t \to 0$. As a result, the autocorrelation of the informed order flow is now a function of $\tau$.

**Proposition 13.** *The informed order flow autocorrelation $\mathsf{Corr}(\Delta x_j(\tau), \Delta x_{j+1}(\tau))$ increases with $\tau$.*

*Proof.* See Appendix B. □

The level trading component of the informed order flow is positively correlated over time, while the flow trading component is not. When data are sampled at a very high frequency, the flow trading component represents a large fraction of the informed order flow variance. In this case, the autocorrelation of the informed order is therefore close to zero. By contrast, at a lower frequency, the level trading component becomes a larger part of the the variance of the informed order flow, and the autocorrelation of the informed order flow increases.

# 5 Conclusions

We have argued that adverse selection has two components: a precision component, and a speed component. To analyze the effect of speed on market quality, we have proposed two models of trading with an informed trader who continuously observes a stream of news. In the benchmark model, the informed trader learns about the asset value at the same time as the market maker. In the fast model, the informed trader has an infinitesimal speed advantage. We have shown that the difference in equilibrium outcomes between the two models is large. In particular, we have shown that in the fast model the optimal strategy of the informed trader has a flow trading component, which is an order of magnitude larger and more volatile than the level trading component.

As a consequence, in the fast model the fraction of trading volume due to the informed investor is large, while in the benchmark model this fraction is essentially zero at high frequencies. As a result of an extra component of adverse selection, liquidity is lower in the fast model, compared to the benchmark. Nevertheless, price volatility and price informativeness are the same, due to a substitution effect. In the fast model, there is more flow trading, but less level trading.

Our results are consistent with stylized facts about high frequency trading, and we generate additional predictions about (i) the causal effect of high frequency trading on various market performance measures; (ii) the effect of various determinants of high frequency trading, both in the cross section and in the time series. For example, we find that an increase in the precision of public news increases the amount of high frequency trading, yet, surprisingly, liquidity is improved.

# A Models in Discrete Time

## A.1 Discrete Time Fast Model

We divide the interval $[0,1]$ into $T$ equally spaced intervals of length $\Delta t = \frac{1}{T}$. Trading takes place at equally spaced times, $t = 1, 2, \dots, T-1$. The sequence of events is as follows. At $t = 0$, the informed trader observes $v_0$. At each $t = 1, \dots, T-1$, the informed trader observes $\Delta v_t = v_t - v_{t-1}$; and the market maker observes $\Delta z_{t-1} = \Delta v_{t-1} + \Delta e_{t-1}$, except at $t = 1$. The error in the market maker's signal is normally distributed, $\Delta e_{t-1} \sim \mathcal{N}(0, \sigma_e^2 \Delta t)$. The market maker quotes the bid price = the ask price = $q_t$. The informed trader then submits $\Delta x_t$, and the liquidity traders submit in aggregate $\Delta u_t \sim \mathcal{N}(0, \sigma_u^2 \Delta t)$. The market maker observes only the aggregate order flow, $\Delta y_t = \Delta x_t + \Delta u_t$, and sets the price at which the trading takes place, $p_t$. The market maker is competitive, i.e., makes zero profit. This translates into the following formulas:

$$
\begin{aligned}
p_t &= \mathsf{E}(v_t \mid \mathcal{I}_t^p), \qquad & \mathcal{I}_t^p &= \{\Delta y_1, \dots, \Delta y_t, \Delta z_1, \dots, \Delta z_{t-1}\}, && (53)\\
q_{t+1} &= \mathsf{E}(v_t \mid \mathcal{I}_t^q), \qquad & \mathcal{I}_t^q &= \{\Delta y_1, \dots, \Delta y_t, \Delta z_1, \dots, \Delta z_t\}. && (54)
\end{aligned}
$$

We also denote

$$
\begin{aligned}
\Omega_t &= \mathsf{Var}(v_t \mid \mathcal{I}_t^p), && (55)\\
\Sigma_t &= \mathsf{Var}(v_t \mid \mathcal{I}_t^q). && (56)
\end{aligned}
$$

**Definition 1.** *A pricing rule $p_t$ is called linear if it is of the form $p_t = q_t + \lambda_t \Delta y_t$, for all $t = 1, \dots, T-1$.*[20] *An equilibrium is called linear if the pricing rule is linear, and the informed trader's strategy $\Delta x_t$ is linear in $\{v_\tau\}_{\tau \le t}$ and $\{q_\tau\}_{\tau \le t}$.*

The next result shows that if the pricing rule is linear, the informed trader's strategy is also linear, and furthermore it can be decomposed into a level trading component, $\beta_t(v_{t-1} - q_t)$, and a flow trading component, $\gamma_t \Delta v_t$.

[20] We could defined more generally, a pricing rule to be linear in the whole history $\{\Delta y_\tau\}_{\tau \le t}$, but as Kyle (1985) shows, this is equivalent to the pricing rule being linear only in $\Delta y_t$.

**Theorem 2.** *Any linear equilibrium must be of the form*

$$\begin{aligned}
\Delta x_t &= \beta_t(v_{t-1} - q_t)\Delta t + \gamma_t \Delta v_t, && (57)\\
p_t &= q_t + \lambda_t \Delta y_t, && (58)\\
q_{t+1} &= p_t + \mu_t(\Delta z_t - \rho_t \Delta y_t), && (59)
\end{aligned}$$

*for $t = 1, \ldots, T-1$, where $\beta_t$, $\gamma_t$, $\lambda_t$, $\mu_t$, $\rho_t$, $\Omega_t$, and $\Sigma_t$ are constants that satisfy*

$$\lambda_t = \frac{\beta_t \Sigma_{t-1} + \gamma_t \sigma_v^2}{\beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2}, \tag{60}$$

$$\mu_t = \frac{\left(\sigma_u^2 + \beta_t^2 \Sigma_{t-1}\Delta t - \beta_t \gamma_t \Sigma_{t-1}\right)\sigma_v^2}{\left(\beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2\right)\sigma_e^2 + \left(\beta_t^2 \Sigma_{t-1}\Delta t + \sigma_u^2\right)\sigma_v^2}, \tag{61}$$

$$m_t = \lambda_t - \rho_t \mu_t = \frac{\beta_t \Sigma_{t-1}(\sigma_v^2 + \sigma_e^2) + \gamma_t \sigma_v^2 \sigma_e^2}{\left(\beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2\right)\sigma_e^2 + \left(\beta_t^2 \Sigma_{t-1}\Delta t + \sigma_u^2\right)\sigma_v^2}, \tag{62}$$

$$\rho_t = \frac{\gamma_t \sigma_v^2}{\beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2}, \tag{63}$$

$$\Omega_t = \Sigma_{t-1} + \sigma_v^2 \Delta t - \frac{\beta_t^2 \Sigma_{t-1}^2 + 2\beta_t \gamma_t \Sigma_{t-1}\sigma_v^2 + \gamma_t^2 \sigma_v^4}{\beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2}\,\Delta t, \tag{64}$$

$$\Sigma_t = \Sigma_{t-1} + \sigma_v^2 \Delta t - \frac{\beta_t^2 \Sigma_{t-1}^2(\sigma_v^2 + \sigma_e^2) + \beta_t^2 \Sigma_{t-1}\Delta t \sigma_v^4 + \sigma_v^4 \sigma_u^2 + \gamma_t^2 \sigma_v^4 \sigma_e^2 + 2\beta_t \gamma_t \Sigma_{t-1}\sigma_v^2 \sigma_e^2}{\left(\beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2\right)\sigma_e^2 + \left(\beta_t^2 \Sigma_{t-1}\Delta t + \sigma_u^2\right)\sigma_v^2}\,\Delta t. \tag{65}$$

*The value function of the informed trader is quadratic for all $t = 1, \ldots, T-1$:*

$$\pi_t = \alpha_{t-1}(v_{t-1} - q_t)^2 + \alpha'_{t-1}(\Delta v_t)^2 + \alpha''_{t-1}(v_{t-1} - q_t)\Delta v_t + \delta_{t-1}. \tag{66}$$

*The coefficients of the optimal trading strategy and the value function satisfy*

$$\beta_t\Delta t = \frac{1-2\alpha_t m_t}{2(\lambda_t-\alpha_t m_t^2)}, \tag{67}$$

$$\gamma_t = \frac{1-2\alpha_t m_t(1-\mu_t)}{2(\lambda_t-\alpha_t m_t^2)}, \tag{68}$$

$$\alpha_{t-1} = \beta_t\Delta t(1-\lambda_t\beta_t\Delta t)+\alpha_t(1-m_t\beta_t\Delta t)^2, \tag{69}$$

$$\alpha'_{t-1} = \alpha_t(1-\mu_t-m_t\gamma_t)^2+\gamma_t(1-\lambda_t\gamma_t), \tag{70}$$

$$\alpha''_{t-1} = \beta_t\Delta t+\gamma_t(1-2\lambda_t\beta_t\Delta t)+2\alpha_t(1-m_t\beta_t\Delta t)(1-\mu_t-m_t\gamma_t), \tag{71}$$

$$\delta_{t-1} = \alpha_t\big(m_t^2\sigma_u^2+\mu_t^2\sigma_e^2\big)\Delta t+\alpha'_t\sigma_v^2\Delta t+\delta_t. \tag{72}$$

*The terminal conditions are*

$$\alpha_T = \alpha'_T = \alpha''_T = \delta_T = 0. \tag{73}$$

*The second order condition is*

$$\lambda_t-\alpha_t m_t^2 > 0. \tag{74}$$

*Given $\Sigma_0$, conditions (60)–(74) are necessary and sufficient for the existence of a linear equilibrium.*

*Proof.* First, we show that Equations (60)–(65) are equivalent to the zero profit conditions of the market maker. Second, we show that Equations (67)–(74) are equivalent to the informed trader's strategy (57) being optimal.

**Zero Profit of Market Maker:** Let us start with with the market maker's update due to the order flow at $t=1,\dots,T-1$. Conditional on $\mathcal{I}_{t-1}^q$, the variables $v_{t-1}-q_t$ and $\Delta v_t$ have a bivariate normal distribution:

$$\begin{bmatrix} v_{t-1}-q_t \\ \Delta v_t \end{bmatrix} \mid \mathcal{I}_{t-1}^q \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{t-1} & 0 \\ 0 & \sigma_v^2 \end{bmatrix}\right). \tag{75}$$

The aggregate order flow at $t$ is of the form

$$\Delta y_t \;=\; \beta_t(v_{t-1} - q_t)\Delta t + \gamma_t \Delta v_t + \Delta u_t. \tag{76}$$

Denote by

$$\Phi_t \;=\; \mathsf{Cov}\left(\begin{bmatrix} v_{t-1} - q_t \\ \Delta v_t \end{bmatrix}, \Delta y_t\right) \;=\; \begin{bmatrix} \beta_t \Sigma_{t-1} \\ \gamma_t \sigma_v^2 \end{bmatrix} \Delta t. \tag{77}$$

Then, conditional on $\mathcal{I}_t = \mathcal{I}_{t-1}^q \cup \{\Delta y_t\}$, the distribution of $v_{t-1} - q_t$ and $\Delta v_t$ is bivariate normal:

$$\begin{bmatrix} v_{t-1} - q_t \\ \Delta v_t \end{bmatrix} \mid \mathcal{I}_t \;\sim\; \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right), \tag{78}$$

where

$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \;=\; \Phi_t \,\mathsf{Var}(\Delta y_t)^{-1} \Delta y_t \;=\; \begin{bmatrix} \beta_t \Sigma_{t-1} \\ \gamma_t \sigma_v^2 \end{bmatrix} \frac{1}{\beta_t^2 \Sigma_{t-1} \Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2}\, \Delta y_t, \tag{79}$$

and

$$\begin{aligned} \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \;&=\; \mathsf{Var}\left(\begin{bmatrix} v_{t-1} - q_t \\ \Delta v_t \end{bmatrix}\right) - \Phi_t \,\mathsf{Var}(\Delta y_t)^{-1} \Phi_t' \\ = \begin{bmatrix} \Sigma_{t-1} & 0 \\ 0 & \sigma_v^2 \Delta t \end{bmatrix} &- \frac{1}{\beta_t^2 \Sigma_{t-1} \Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2} \begin{bmatrix} \beta_t^2 \Sigma_{t-1}^2 & \beta_t \gamma_t \Sigma_{t-1} \sigma_v^2 \\ \beta_t \gamma_t \Sigma_{t-1} \sigma_v^2 & \gamma_t^2 \sigma_v^4 \end{bmatrix} \Delta t. \end{aligned} \tag{80}$$

We compute

$$p_t - q_t \;=\; \mathsf{E}(v_t - q_t \mid \mathcal{I}_t) \;=\; \mu_1 + \mu_2 \;=\; \frac{\beta_t \Sigma_{t-1} + \gamma_t \sigma_v^2}{\beta_t^2 \Sigma_{t-1} \Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2}\, \Delta y_t, \tag{81}$$

which proves Equation (60) for $\lambda_t$. Also,

$$\begin{aligned} \Omega_t \;&=\; \mathsf{Var}(v_t - q_t \mid \mathcal{I}_t) \;=\; \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2 \\ &=\; \Sigma_{t-1} + \sigma_v^2 \Delta t - \frac{\beta_t^2 \Sigma_{t-1}^2 + 2\beta_t\gamma_t \Sigma_{t-1}\sigma_v^2 + \gamma_t^2 \sigma_v^4}{\beta_t^2 \Sigma_{t-1} \Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2}\, \Delta t, \end{aligned} \tag{82}$$

which proves (64).

Next, to compute $q_{t+1} = \mathsf{E}(v_t \mid \mathcal{I}_t^q)$, we start from the same prior as in (75), but we consider the impact of both the order flow at $t$ and the market maker's signal at $t+1$:

$$\begin{aligned} \Delta y_t &= \beta_t(v_{t-1} - q_t)\Delta t + \gamma_t \Delta v_t + \Delta u_t, &(83)\\ \Delta z_t &= \Delta v_t + \Delta e_t. &(84)\end{aligned}$$

Denote by

$$\begin{aligned} \Psi_t &= \mathsf{Cov}\left(\begin{bmatrix} v_{t-1} - q_t \\ \Delta v_t \end{bmatrix}, \begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix}\right) = \begin{bmatrix} \beta_t \Sigma_{t-1} & 0 \\ \gamma_t \sigma_v^2 & \sigma_v^2 \end{bmatrix} \Delta t, &(85)\\ V_t^{yz} &= \mathsf{Var}\left(\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix}\right) = \begin{bmatrix} \beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2 & \gamma_t \sigma_v^2 \\ \gamma_t \sigma_v^2 & \sigma_v^2 + \sigma_e^2 \end{bmatrix} \Delta t. &(86)\end{aligned}$$

Conditional on $\mathcal{I}_t^q = \mathcal{I}_{t-1}^q \cup \{\Delta y_t, \Delta z_t\}$, the distribution of $v_{t-1} - q_t$ and $\Delta v_t$ is bivariate normal:

$$\begin{bmatrix} v_{t-1} - q_t \\ \Delta v_t \end{bmatrix} \mid \mathcal{I}_t^q \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right), \tag{87}$$

where

$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \Psi_t \,(V_t^{yz})^{-1} \begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \frac{\begin{bmatrix} \beta_t \Sigma_{t-1}(\sigma_v^2 + \sigma_e^2)\Delta y_t \; - \; \beta_t \gamma_t \Sigma_{t-1} \sigma_v^2 \Delta z_t \\ \gamma_t \sigma_v^2 \sigma_e^2 \Delta y_t \; + \; (\beta_t^2 \Sigma_{t-1}\Delta t + \sigma_u^2)\sigma_v^2 \Delta z_t \end{bmatrix}}{(\beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2)\sigma_e^2 + (\beta_t^2 \Sigma_{t-1}\Delta t + \sigma_u^2)\sigma_v^2}, \tag{88}$$

and

$$\begin{aligned} \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} &= \mathsf{Var}\left(\begin{bmatrix} v_{t-1} - q_t \\ \Delta v_t \end{bmatrix}\right) - \Psi_t \,(V_t^{yz})^{-1}\, \Psi_t' \\ &= \begin{bmatrix} \Sigma_{t-1} & 0 \\ 0 & \sigma_v^2 \Delta t \end{bmatrix} - \frac{\begin{bmatrix} \beta_t^2 \Sigma_{t-1}^2 (\sigma_v^2 + \sigma_e^2) & \beta_t \gamma_t \Sigma_{t-1} \sigma_v^2 \sigma_e^2 \\ \beta_t \gamma_t \Sigma_{t-1} \sigma_v^2 \sigma_e^2 & (\beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_e^2 + \sigma_u^2)\sigma_v^4 \end{bmatrix}}{(\beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2)\sigma_e^2 + (\beta_t^2 \Sigma_{t-1}\Delta t + \sigma_u^2)\sigma_v^2} \Delta t. \end{aligned} \tag{89}$$

Therefore,

$$\begin{aligned} q_{t+1} - q_t &= \mu_1 + \mu_2 \\ &= \frac{\big(\beta_t \Sigma_{t-1}(\sigma_v^2 + \sigma_e^2) + \gamma_t \sigma_v^2 \sigma_e^2\big)\Delta y_t \;+\; \big(\sigma_u^2 + \beta_t^2 \Sigma_{t-1}\Delta t - \beta_t \gamma_t \Sigma_{t-1}\big)\sigma_v^2 \Delta z_t}{(\beta_t^2 \Sigma_{t-1}\Delta t + \gamma_t^2 \sigma_v^2 + \sigma_u^2)\,\sigma_e^2 + (\beta_t^2 \Sigma_{t-1}\Delta t + \sigma_u^2)\,\sigma_v^2} \qquad (90) \\ &= m_t \Delta y_t \;+\; \mu_t \Delta z_t \;=\; (\lambda_t - \rho_t \mu_t)\Delta y_t \;+\; \mu_t \Delta z_t, \qquad (91) \end{aligned}$$

which proves Equations (61), (62), and (63). Also,

$$\begin{aligned} \Sigma_t &= \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2 \\ &= \Sigma_{t-1} + \sigma_v^2 \Delta t - \frac{\beta_t^2 \Sigma_{t-1}^2(\sigma_v^2 + \sigma_e^2) + \beta_t^2 \Sigma_{t-1}\Delta t \sigma_v^4 + \sigma_v^4 \sigma_u^2 + \gamma_t^2 \sigma_v^4 \sigma_e^2 + 2\beta_t \gamma_t \Sigma_{t-1}\sigma_v^2 \sigma_e^2}{(\beta_t^2 \Sigma_{t-1} + (\beta_t + \gamma_t)^2 \sigma_v^2 + \sigma_u^2)\,\sigma_e^2 + (\beta_t^2 \Sigma_{t-1} + \sigma_u^2)\,\sigma_v^2}\,\Delta t, \qquad (92) \end{aligned}$$

which proves (65).

**Optimal Strategy of Informed Trader:** At each $t = 1, \ldots, T-1$, the informed trader maximizes the expected profit: $\pi_t = \max \sum_{\tau=t}^{T-1} \mathsf{E}\big((v_T - p_\tau)\Delta x_\tau\big)$. We prove by backward induction that the value function is quadratic and of the form given in (66): $\pi_t = \alpha_{t-1}(v_{t-1} - q_t)^2 + \alpha'_{t-1}(\Delta v_t)^2 + \alpha''_{t-1}(v_{t-1} - q_t)\Delta v_t + \delta_{t-1}$. At the last decision point $(t = T-1)$ the next value function is zero, i.e., $\alpha_T = \alpha'_T = \alpha''_T = \delta_T = 0$, which are the terminal conditions (73). This is the transversality condition: no money is left on the table. In the induction step, if $t = 1, \ldots, T-1$, we assume that $\pi_{t+1}$ is of the desired form. The Bellman principle of intertemporal optimization implies

$$\pi_t \;=\; \max_{\Delta x}\; \mathsf{E}\Big((v_t - p_t)\Delta x + \pi_{t+1} \mid \mathcal{I}_t^q, v_t, \Delta v_t\Big). \qquad (93)$$

Equations (58) and (59) show that the quote $q_t$ evolves by $q_{t+1} = q_t + m_t \Delta y_t + \mu_t \Delta z_t$, where $m_t = \lambda_t - \rho_t \mu_t$. This implies that the informed trader's choice of $\Delta x$ affects the trading price and the next quote by

$$
\begin{aligned}
p_t &= q_t + \lambda_t(\Delta x + \Delta u_t), && (94)\\
q_{t+1} &= q_t + m_t(\Delta x + \Delta u_t) + \mu_t \Delta z_t. && (95)
\end{aligned}
$$

Substituting these into the Bellman equation, we get

$$
\begin{aligned}
\pi_t \;=\; & \max_{\Delta x} \; \mathsf{E}\Big(\Delta x(v_{t-1} + \Delta v_t - q_t - \lambda_t \Delta x - \lambda_t \Delta u_t) \\
& \;+\; \alpha_t(v_{t-1} + \Delta v_t - q_t - m_t\Delta x - m_t \Delta u_t - \mu_t \Delta v_t - \mu_t \Delta e_t)^2 \;+\; \alpha'_t \Delta v_{t+1}^2 \qquad (96)\\
& \;+\; \alpha''_t(v_{t-1} + \Delta v_t - q_t - m_t\Delta x - m_t \Delta u_t - \mu_t \Delta v_t - \mu_t \Delta e_t)\Delta v_{t+1} \;+\; \delta_t\Big) \\
=\; & \max_{\Delta x} \; \Delta x(v_{t-1} - q_t + \Delta v_t - \lambda_t \Delta x) \\
& \;+\; \alpha_t\Big((v_{t-1} - q_t - m_t \Delta x + (1-\mu_t)\Delta v_t)^2 + (m_t^2 \sigma_u^2 + \mu_t^2 \sigma_e^2)\Delta t\Big) \;+\; \alpha'_t \sigma_v^2 \Delta t \qquad (97)\\
& \;+\; 0 \;+\; \delta_t.
\end{aligned}
$$

The first order condition with respect to $\Delta x$ is

$$
\Delta x \;=\; \frac{1 - 2\alpha_t m_t}{2(\lambda_t - \alpha_t m_t^2)}(v_{t-1} - q_t) \;+\; \frac{1 - 2\alpha_t m_t (1-\mu_t)}{2(\lambda_t - \alpha_t m_t^2)} \Delta v_t, \qquad (98)
$$

and the second order condition for a maximum is $\lambda_t - \alpha_t m_t^2 > 0$, which is (74). Thus, the optimal $\Delta x$ is indeed of the form $\Delta x_t = \beta_t(v_{t-1} - q_t)\Delta t + \gamma_t \Delta v_t$, where $\beta_t \Delta t$ and $\gamma_t$ are as in Equations (67) and (68). We substitute $\Delta x_t$ in the formula for $\pi_t$ to obtain

$$
\begin{aligned}
\pi_t \;=\; & \Big(\beta_t \Delta t(1 - \lambda_t \beta_t \Delta t) + \alpha_t (1 - m_t \beta_t \Delta t)^2\Big)(v_{t-1} - q_t)^2 \\
& + \Big(\alpha_t(1 - \mu_t - m_t \gamma_t)^2 + \gamma_t(1 - \lambda_t \gamma_t)\Big)\Delta v_t^2 \qquad (99)\\
& + \Big(\beta_t \Delta t + \gamma_t(1 - 2\lambda_t \beta_t \Delta t) + 2\alpha_t(1 - m_t \beta_t \Delta t)(1 - \mu_t - m_t \gamma_t)\Big)(v_{t-1} - q_t)\Delta v_t \\
& + \alpha_t\big(m_t^2 \sigma_u^2 + \mu_t^2 \sigma_e^2\big)\Delta t \;+\; \alpha'_t \sigma_v^2 \Delta t \;+\; \delta_t.
\end{aligned}
$$

This proves that indeed $\pi_t$ is of the form $\pi_t = \alpha_{t-1}(v_{t-1} - q_t)^2 + \alpha'_{t-1}(\Delta v_t)^2 + \alpha''_{t-1}(v_{t-1} - q_t)\Delta v_t + \delta_{t-1}$, with $\alpha_{t-1}$, $\alpha'_{t-1}$, $\alpha''_{t-1}$ and $\delta_{t-1}$ as in Equations (69)–(72). □

We now briefly discuss the existence of a solution for the recursive system given in

Theorem 2. The system of equations (60)–(72) can be numerically solved backwards, starting from the boundary conditions (73). We also start with an arbitrary value of $\Sigma_T > 0$.[21] By backward induction, suppose $\alpha_t$ and $\Sigma_t$ are given. One verifies that Equation (65) implies

$$\Sigma_{t-1} = \frac{\Sigma_t\big(\sigma_v^2\sigma_u^2 + \sigma_v^2(\sigma_u^2 + \gamma_t^2\sigma_e^2)\big) - \sigma_v^2\sigma_u^2\sigma_e^2\Delta t}{\big(\sigma_u^2\sigma_e^2 + \sigma_v^2(\sigma_u^2 + \gamma_t^2\sigma_e^2) + \beta_t^2\Delta t^2\sigma_v^2\sigma_e^2 - 2\gamma_t\beta_t\Delta t\sigma_v^2\sigma_e^2\big) - \Sigma_t\beta_t^2\Delta t\big(\sigma_v^2 + \sigma_e^2\big)}. \tag{100}$$

Then, Equations (60)–(62) can be rewritten to express $\lambda_t, \mu_t, m_t$ as functions of $(\Sigma_t, \beta_t, \gamma_t)$ instead of $(\Sigma_{t-1}, \beta_t, \gamma_t)$. Next, we use (67) and (68) to express $\lambda_t, \mu_t, m_t$ as functions of $(\lambda_t, \mu_t, m_t, \alpha_t, \Sigma_t)$. This gives a system of polynomial equations, whose solution $\lambda_t, \mu_t, m_t$ depends only on $(\alpha_t, \Sigma_t)$. Numerical simulations show that the solution is unique under the second order condition (74), but the authors have not been able to prove theoretically that this is true in all cases. Once the recursive system is computed for all $t = 1, \ldots, T-1$, the only condition left to do is to verify that the value obtained for $\Sigma_0$ is the correct one. However, unlike in Kyle (1985), the recursive equation for $\Sigma_t$ is not linear, and therefore the parameters cannot be simply rescaled. Instead, one must numerically modify the initial choice of $\Sigma_T$ until the correct value of $\Sigma_0$ is reached.

## A.2 Discrete Time Benchmark Model

The setup is the same as for the fast model, except that the market maker gets the signal $\Delta z$ at the same time as the informed trader observes $\Delta v$. The sequence of events is as follows. At $t = 0$, the informed trader observes $v_0$. At each $t = 1, \ldots, T-1$, the informed trader observes $\Delta v_t = v_t - v_{t-1}$; and the market maker observes $\Delta z_t = \Delta v_t + \Delta e_t$, with $\Delta e_t \sim \mathcal{N}(0, \sigma_e^2\Delta t)$. The market maker quotes the bid price = the ask price = $q_t$. The informed trader then submits $\Delta x_t$, and the liquidity traders submit in aggregate $\Delta u_t \sim \mathcal{N}(0, \sigma_u^2\Delta t)$. The market maker observes only the aggregate order flow, $\Delta y_t = \Delta x_t + \Delta u_t$, and sets the price at which the trading takes place, $p_t$. The

[21] Numerically, it should be of the order of $\Delta t$.

market maker is competitive, i.e., makes zero profit. This implies

$$p_t = \mathsf{E}(v_t \mid \mathcal{I}_t^p), \qquad \mathcal{I}_t^p = \{\Delta y_1, \ldots, \Delta y_t, \Delta z_1, \ldots, \Delta z_t\}, \tag{101}$$

$$q_t = \mathsf{E}(v_t \mid \mathcal{I}_t^q), \qquad \mathcal{I}_t^q = \{\Delta y_1, \ldots, \Delta y_{t-1}, \Delta z_1, \ldots, \Delta z_t\}. \tag{102}$$

We also denote

$$\Sigma_t = \mathsf{Var}(v_t \mid \mathcal{I}_t^p), \tag{103}$$

$$\Omega_t = \mathsf{Var}(v_t \mid \mathcal{I}_t^q). \tag{104}$$

The next result shows that if the pricing rule is linear, the informed trader's strategy is also linear, and furthermore it only has a level trading component, $\beta_t(v_t - q_t)$.

**Theorem 3.** *Any linear equilibrium must be of the form*

$$\Delta x_t = \beta_t(v_t - q_t)\Delta t, \tag{105}$$

$$p_t = q_t + \lambda_t \Delta y_t, \tag{106}$$

$$q_t = p_{t-1} + \mu_{t-1}\Delta z_t, \tag{107}$$

*for $t = 1, \ldots, T-1$, where by convention $p_0 = 0$, and $\beta_t$, $\gamma_t$, $\lambda_t$, $\mu_t$, $\Omega_t$, and $\Sigma_t$ are constants that satisfy*

$$\lambda_t = \frac{\beta_t \Sigma_t}{\sigma_u^2}, \tag{108}$$

$$\mu_t = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2}, \tag{109}$$

$$\Omega_t = \frac{\Sigma_t \sigma_u^2}{\sigma_u^2 - \beta_t^2 \Sigma_t \Delta t}, \tag{110}$$

$$\Sigma_{t-1} = \Sigma_t + \frac{\beta_t^2 \Sigma_t^2}{\sigma_u^2 - \beta_t^2 \Sigma_t \Delta t}\,\Delta t - \frac{\sigma_v^2 \sigma_e^2}{\sigma_v^2 + \sigma_e^2}\,\Delta t. \tag{111}$$

*The value function of the informed trader is quadratic for all $t = 1, \ldots, T-1$:*

$$\pi_t = \alpha_{t-1}(v_t - q_t)^2 + \delta_{t-1}. \tag{112}$$

*The coefficients of the optimal trading strategy and the value function satisfy*

$$\beta_t \Delta t \;=\; \frac{1-2\alpha_t\lambda_t}{2\lambda_t(1-\alpha_t\lambda_t)}, \tag{113}$$

$$\alpha_{t-1} \;=\; \beta_t\Delta t(1-\lambda_t\beta_t\Delta t)+\alpha_t(1-\lambda_t\beta_t\Delta t)^2, \tag{114}$$

$$\delta_{t-1} \;=\; \alpha_t\big(\lambda_t^2\sigma_u^2+\mu_t^2(\sigma_v^2+\sigma_e^2)\big)\Delta t+\delta_t. \tag{115}$$

*The terminal conditions are*

$$\alpha_T \;=\; \delta_T \;=\; 0. \tag{116}$$

*The second order condition is*

$$\lambda_t(1-\alpha_t\lambda_t) \;>\; 0. \tag{117}$$

*Given $\Sigma_0$, conditions (108)–(117) are necessary and sufficient for the existence of a linear equilibrium.*

*Proof.* First, we show that Equations (108)–(111) are equivalent to the zero profit conditions of the market maker. Second, we show that Equations (113)–(117) are equivalent to the informed trader's strategy (105) being optimal.

**Zero Profit of Market Maker:** Let us start with with the market maker's update due to the order flow at $t=1,\dots,T-1$. Conditional on $\mathcal{I}_t^q$, $v_t$ has a normal distribution, $v_t|\mathcal{I}_t^q \sim \mathcal{N}(q_t,\Omega_t)$. The aggregate order flow at $t$ is of the form $\Delta y_t = \beta_t(v_t-q_t)\Delta t+\Delta u_t$. Denote by

$$\Phi_t \;=\; \mathsf{Cov}(v_t-q_t,\Delta y_t) \;=\; \beta_t\Omega_t\Delta t. \tag{118}$$

Then, conditional on $\mathcal{I}_t^p = \mathcal{I}_t^q \cup \{\Delta y_t\}$, $v_t \sim \mathcal{N}(p_t,\Sigma_t)$, with

$$p_t \;=\; q_t+\lambda_t\Delta y_t, \tag{119}$$

$$\lambda_t \;=\; \Phi_t\,\mathsf{Var}(\Delta y_t)^{-1} \;=\; \frac{\beta_t\Omega_t}{\beta_t^2\Omega_t\Delta t+\sigma_u^2}, \tag{120}$$

$$\begin{aligned}\Sigma_t \;&=\; \mathsf{Var}(v_t-q_t)-\Phi_t\,\mathsf{Var}(\Delta y_t)^{-1}\Phi_t' \\ &=\; \Omega_t-\frac{\beta_t^2\Omega_t^2}{\beta_t^2\Omega_t\Delta t+\sigma_u^2}\,\Delta t \;=\; \frac{\Omega_t\sigma_u^2}{\beta_t^2\Omega_t\Delta t+\sigma_u^2}.\end{aligned} \tag{121}$$

To obtain Equation (108) for $\lambda_t$, note that the above equations for $\lambda_t$ and $\Sigma_t$ imply $\frac{\lambda_t}{\Sigma_t} = \frac{\beta_t}{\sigma_u^2}$. Equation (110) is obtained by solving for $\Sigma_t$ in Equation (121).

Next, consider the market maker's update at $t = 1, \ldots, T-1$ due to the signal $\Delta z_t = \Delta v_t + \Delta e_t$. From $v_{t-1}|\mathcal{I}^p_{t-1} \sim \mathcal{N}(p_{t-1}, \Sigma_{t-1})$, we have $v_t|\mathcal{I}^p_{t-1} \sim \mathcal{N}(p_{t-1}, \Sigma_{t-1} + \sigma_v^2 \Delta t)$ . Denote by

$$\Psi_t \; = \; \mathsf{Cov}(v_t - p_{t-1}, \Delta z_t) \; = \; \sigma_v^2 \Delta t. \tag{122}$$

Then, conditional on $\mathcal{I}^q_t = \mathcal{I}^p_{t-1} \cup \{\Delta z_t\}$, $v_t|\mathcal{I}^q_t \sim \mathcal{N}(q_t, \Omega_t)$, with

$$\begin{aligned} q_t &= p_{t-1} + \mu_t \Delta z_t, && (123) \\ \mu_t &= \Psi_t \, \mathsf{Var}(\Delta z_t)^{-1} \; = \; \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2}, && (124) \\ \Omega_t &= \mathsf{Var}(v_t - p_{t-1}) - \Psi_t \, \mathsf{Var}(\Delta z_t)^{-1} \Psi_t' \\ &= \Sigma_{t-1} + \sigma_v^2 \Delta t - \frac{\sigma_v^4}{\sigma_v^2 + \sigma_e^2} \, \Delta t \; = \; \Sigma_{t-1} + \frac{\sigma_v^2 \sigma_e^2}{\sigma_v^2 + \sigma_e^2} \, \Delta t. && (125) \end{aligned}$$

which proves Equation (109) for $\mu_t$. Note that Equation (125) gives a formula for $\Sigma_{t-1}$ as a function of $\Omega_t$, and we already proved (110), which expresses $\Omega_t$ as a function of $\Sigma_t$. We therefore get $\Sigma_{t-1}$ as a function of $\Sigma_t$, which is (111).

**Optimal Strategy of Informed Trader:** At each $t = 1, \ldots, T-1$, the informed trader maximizes the expected profit: $\pi_t = \max \sum_{\tau=t}^{T-1} \mathsf{E}\big((v_T - p_\tau)\Delta x_\tau\big)$. We prove by backward induction that the value function is quadratic and of the form given in (112): $\pi_t = \alpha_{t-1}(v_t - q_t)^2 + \delta_{t-1}$. At the last decision point ($t = T-1$) the next value function is zero, i.e., $\alpha_T = \delta_T = 0$, which are the terminal conditions (116). In the induction step, if $t = 1, \ldots, T-1$, we assume that $\pi_{t+1}$ is of the desired form. The Bellman principle of intertemporal optimization implies

$$\pi_t \; = \; \max_{\Delta x} \; \mathsf{E}\Big((v_t - p_t)\Delta x + \pi_{t+1} \mid \mathcal{I}^q_t, v_t, \Delta v_t\Big). \tag{126}$$

Equations (106) and (107) show that the quote $q_t$ evolves by $q_{t+1} = q_t + m_t \Delta y_t + \mu_t \Delta z_{t+1}$. This implies that the informed trader's choice of $\Delta x$ affects the trading price and the next quote by

$$p_t = q_t + \lambda_t(\Delta x + \Delta u_t), \tag{127}$$

$$q_{t+1} = q_t + \lambda_t(\Delta x + \Delta u_t) + \mu_t \Delta z_{t+1}. \tag{128}$$

Substituting these into the Bellman equation, we get

$$\begin{aligned}\pi_t \;=\; & \max_{\Delta x} \; \mathsf{E}\Big(\Delta x(v_t - q_t - \lambda_t \Delta x - \lambda_t \Delta u_t) \\ & + \; \alpha_t(v_t + \Delta v_{t+1} - q_t - \lambda_t \Delta x - \lambda_t \Delta u_t - \mu_t \Delta z_{t+1})^2 \; + \; \delta_t\Big)\end{aligned} \tag{129}$$

$$\begin{aligned} = \; & \max_{\Delta x} \; \Delta x(v_t - q_t - \lambda_t \Delta x) \\ & + \; \alpha_t\Big((v_t - q_t - \lambda_t \Delta x)^2 + (\lambda_t^2 \sigma_u^2 + \mu_t^2(\sigma_v^2 + \sigma_e^2))\Delta t\Big) \; + \; \delta_t.\end{aligned} \tag{130}$$

The first order condition with respect to $\Delta x$ is

$$\Delta x \;=\; \frac{1 - 2\alpha_t \lambda_t}{2\lambda_t(1 - \alpha_t \lambda_t)}(v_t - q_t), \tag{131}$$

and the second order condition for a maximum is $\lambda_t(1 - \alpha_t \lambda_t) > 0$, which is (117). Thus, the optimal $\Delta x$ is indeed of the form $\Delta x_t = \beta_t(v_t - q_t)\Delta t$, where $\beta_t \Delta t$ satisfies Equation (113). We substitute $\Delta x_t$ in the formula for $\pi_t$ to obtain

$$\pi_t \;=\; \Big(\beta_t \Delta t(1 - \lambda_t \beta_t \Delta t) + \alpha_t(1 - \lambda_t \beta_t \Delta t)^2\Big)(v_t - q_t)^2 \;+\; \alpha_t\big(\lambda_t^2 \sigma_u^2 + \mu_t^2(\sigma_v^2 + \sigma_e^2)\big)\Delta t \;+\; \delta_t. \tag{132}$$

This proves that indeed $\pi_t$ is of the form $\pi_t = \alpha_{t-1}(v_t - q_t)^2 + \delta_{t-1}$, with $\alpha_{t-1}$ and $\delta_{t-1}$ as in Equations (114) and (115). □

Equations (108)–(111) and (113)–(115) form a system of equations. As before, it is solved backwards, starting from the boundary conditions (116), and so that $\Sigma_t = \Sigma_0$ at $t = 0$.

# B Proofs

## B.1 Proof of Theorem 1

**Benchmark model:** We compute the optimal strategy of the informed trader at $t + \mathrm{d}t$. As we have seen in the discrete version of the model, in Appendix A, we need to consider only strategies $\mathrm{d}x_\tau$ of the type $\mathrm{d}x_\tau = \beta_\tau(v_\tau - p_\tau)\,\mathrm{d}\tau + \gamma_\tau\,\mathrm{d}v_\tau$. Recall that $\mathcal{I}_t^p$ is the market maker's information set immediately after trading at $t$. If we denote by $\mathcal{J}_t^p = \mathcal{I}_t^p \cup \{v_\tau\}_{\tau \le t+\mathrm{d}t}$ the trader's information set before trading at $t + \mathrm{d}t$, the expected profit from trading after $t$ is

$$\pi_t \;=\; \mathsf{E}\left(\int_t^1 (v_1 - p_{\tau+\mathrm{d}\tau})\,\mathrm{d}x_\tau \mid \mathcal{J}_t^p\right). \tag{133}$$

From (11), $p_{\tau+\mathrm{d}\tau} = p_\tau + \mu_\tau(\mathrm{d}v_\tau + \mathrm{d}e_\tau) + \lambda_\tau(\mathrm{d}x_\tau + \mathrm{d}u_\tau)$. For $\tau \ge t$, denote by

$$V_\tau \;=\; \mathsf{E}\big((v_\tau - p_\tau)^2 \mid \mathcal{J}_t^p\big). \tag{134}$$

Then the expected profit is

$$\begin{aligned} \pi_t \;&=\; \mathsf{E}\left(\int_t^1 (v_\tau + \,\mathrm{d}v_\tau - p_\tau - \mu_\tau\,\mathrm{d}v_\tau - \lambda_\tau\,\mathrm{d}x_\tau)\,\mathrm{d}x_\tau \mid \mathcal{J}_t^p\right) & (135)\\ &=\; \int_t^1 \Big(\beta_\tau V_\tau + (1 - \mu_\tau - \lambda_\tau\gamma_\tau)\gamma_\tau\sigma_v^2\Big)\,\mathrm{d}\tau. & (136)\end{aligned}$$

$V_\tau$ can be computed recursively:

$$\begin{aligned} V_{\tau+\mathrm{d}\tau} \;&=\; \mathsf{E}\big((v_{\tau+\mathrm{d}\tau} - p_{\tau+\mathrm{d}\tau})^2 \mid \mathcal{J}_t^p\big)\\ &=\; \mathsf{E}\big((v_\tau + \,\mathrm{d}v_\tau - p_\tau - \mu_\tau\,\mathrm{d}v_\tau - \mu_\tau\,\mathrm{d}e_\tau - \lambda_\tau\,\mathrm{d}x_\tau - \lambda_\tau\,\mathrm{d}u_\tau)^2 \mid \mathcal{J}_t^p\big) \qquad (137)\\ &=\; V_\tau + (1 - \mu_\tau - \lambda_\tau\gamma_\tau)^2\sigma_v^2\,\mathrm{d}\tau + \mu_\tau^2\sigma_e^2\,\mathrm{d}\tau + \lambda_\tau^2\sigma_u^2\,\mathrm{d}\tau - 2\lambda_t\beta_t V_\tau\,\mathrm{d}\tau.\end{aligned}$$

therefore the law of motion of $V_\tau$ is a first order differential equation

$$V_\tau' \;=\; -2\lambda_t\beta_t V_\tau + (1 - \mu_\tau - \lambda_\tau\gamma_\tau)^2\sigma_v^2 + \mu_\tau^2\sigma_e^2 + \lambda_\tau^2\sigma_u^2, \tag{138}$$

or equivalently $\beta_\tau V_\tau = \frac{-V'_\tau + (1-\mu_\tau-\lambda_\tau\gamma_\tau)^2\sigma_v^2 + \mu_\tau^2\sigma_e^2 + \lambda_\tau^2\sigma_u^2}{2\lambda_\tau}$. Substitute this into (133) and integrate by parts. Since $V_t = 0$, we get

$$\begin{aligned}\pi_t \;=\; & -\frac{V_1}{2\lambda_1} + \int_t^1 V_\tau \Big(\frac{1}{2\lambda_\tau}\Big)' \,\mathrm{d}\tau \\ & + \int_t^1 \bigg(\frac{(1-\mu_\tau-\lambda_\tau\gamma_\tau)^2\sigma_v^2 + \mu_\tau^2\sigma_e^2 + \lambda_\tau^2\sigma_u^2}{2\lambda_\tau} + (1-\mu_\tau-\lambda_\tau\gamma_\tau)\gamma_\tau\sigma_v^2\bigg) \,\mathrm{d}\tau.\end{aligned} \tag{139}$$

This is essentially the argument of Kyle (1985): we have eliminated the choice variable $\beta_\tau$ and replaced it by $V_\tau$. Since $V_\tau > 0$ can be arbitrarily chosen, in order to get an optimum we must have $\big(\frac{1}{2\lambda_\tau}\big)' = 0$, which is equivalent to

$$\lambda_\tau \;=\; \text{constant}. \tag{140}$$

For a maximum, the transversality condition $V_1 = 0$ must be also satisfied.

We next turn to the choice of $\gamma_\tau$. The first order condition is

$$-(1-\mu_\tau-\lambda_\tau\gamma_\tau) + (1-\mu_\tau-\lambda_\tau\gamma_\tau) - \lambda_\tau\gamma_\tau \;=\; 0 \quad\Longrightarrow\quad \gamma_\tau \;=\; 0. \tag{141}$$

Thus, there is no flow trading in the benchmark model. Note also that the second order condition is $\lambda_\tau > 0$.[22]

Next, we derive the pricing rules from the market maker's zero profit conditions. The equations $p_t = \mathsf{E}(v_1|\mathcal{I}_t^p)$ and $q_t = \mathsf{E}(v_1|\mathcal{I}_t^p, \mathrm{d}z_t)$ imply that $q_t = p_t + \mu_t\,\mathrm{d}z_t$, where

$$\mu_t \;=\; \frac{\mathsf{Cov}(v_1,\, \mathrm{d}z_t \mid \mathcal{I}_t^p)}{\mathsf{Var}(\,\mathrm{d}z_t \mid \mathcal{I}_t^p)} \;=\; \frac{\mathsf{Cov}(v_0 + \int_0^1 \mathrm{d}v_\tau,\, \mathrm{d}v_t + \,\mathrm{d}e_t \mid \mathcal{I}_t^p)}{\mathsf{Var}(\,\mathrm{d}v_t + \,\mathrm{d}e_t \mid \mathcal{I}_t^p)} \;=\; \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2}. \tag{142}$$

The equations $q_t = \mathsf{E}(v_1|\mathcal{I}_{t+\mathrm{d}t}^q)$ and $p_{t+\mathrm{d}t} = \mathsf{E}(v_1|\mathcal{I}_{t+\mathrm{d}t}^q, \mathrm{d}y_t)$ imply that $p_{t+\mathrm{d}t} = q_t + \lambda_t \mathrm{d}y_t$, where

$$\lambda_t \;=\; \frac{\mathsf{Cov}(v_1,\, \mathrm{d}y_t \mid \mathcal{I}_{t+\mathrm{d}t}^q)}{\mathsf{Var}(\,\mathrm{d}y_t \mid \mathcal{I}_{t+\mathrm{d}t}^q)} \;=\; \frac{\mathsf{Cov}(v_1, \beta_t(v_t - p_t)\,\mathrm{d}t + \,\mathrm{d}u_t \mid \mathcal{I}_{t+\mathrm{d}t}^q)}{\mathsf{Var}(\beta_t(v_t - p_t)\,\mathrm{d}t + \,\mathrm{d}u_t \mid \mathcal{I}_{t+\mathrm{d}t}^q)} \;=\; \frac{\beta_t\Sigma_t}{\sigma_u^2}, \tag{143}$$

[22]The condition $\lambda_\tau > 0$ is also a second order condition with respect to the choice of $\beta_\tau$. To see this, suppose $\lambda_\tau < 0$. Then if $\beta_\tau > 0$ is chosen very large, Equation (138) shows that $V_\tau$ is very large as well, and thus $\beta_\tau V_\tau$ can be made arbitrarily large. Thus, there would be no maximum.

where $\Sigma_t = \mathsf{E}\big((v_t - p_t)^2 | \mathcal{I}_t^p\big)$.[23] The information set of the informed trader, $\mathcal{J}_t^p$, is a refinement of the market maker's information set, $\mathcal{I}_t^p$. Therefore, by the law of iterated expectations, $\Sigma_t$ satisfies the same equation as $V_t$:

$$\Sigma_t' \;=\; -2\lambda_t\beta_t\Sigma_t + (1-\mu_t-\lambda_t\gamma_t)^2\sigma_v^2 + \mu_t^2\sigma_e^2 + \lambda_t^2\sigma_u^2, \tag{144}$$

except that it has a different initial condition. One can solve this differential equation explicitly and show that the transversality condition $V_1 = 0$ is equivalent to $\int_0^1 \beta_t \,\mathrm{d}t = +\infty$, and in turn this is equivalent to $\Sigma_1 = 0$. Since $\lambda_t$, $\mu_t$ and $\gamma_t = 0$ are constant, by (143) $\beta_t\Sigma_t$ is also constant. Equation (144) then implies that $\Sigma_t'$ is constant. Since $\Sigma_1 = 0$, $\Sigma_t = (1-t)\Sigma_0$, and $\beta_t = \frac{\beta_0}{1-t}$. Finally, we integrate (144) between 0 and 1, and substituting for $\lambda = \frac{\beta_0\Sigma_0}{\sigma_u^2}$, $\mu = \frac{\sigma_v^2}{\sigma_v^2+\sigma_e^2}$, and $\gamma = 0$, we obtain $\beta_0$ and $\lambda$ as stated in the Theorem.

**Fast model:** The informed trader has the same objective function as in (133):

$$\pi_t \;=\; \mathsf{E}\left(\int_t^1 (v_1 - p_{\tau+\,\mathrm{d}\tau})\,\mathrm{d}x_\tau \mid \mathcal{J}_t^p\right). \tag{145}$$

but here we use $q_t$ instead of $p_t$ as a state variable. From (7), $p_{t+\mathrm{d}t} = q_t + \lambda_t \mathrm{d}y_t$. Also, from (17), $q_{\tau+\,\mathrm{d}\tau} = q_\tau + \mu_\tau(\mathrm{d}z_\tau - \rho_\tau\mathrm{d}y_\tau) + \lambda_\tau(\mathrm{d}y_\tau)$, and we obtain

$$\begin{aligned} q_{\tau+\,\mathrm{d}\tau} \;&=\; \mu_\tau\,\mathrm{d}z_\tau + m_\tau\,\mathrm{d}y_\tau, \quad \text{with} \qquad (146)\\ m_\tau \;&=\; \lambda_\tau - \mu_\tau\rho_\tau. \qquad (147)\end{aligned}$$

As we have seen in the discrete version of the model, in Appendix A, we need to consider only strategies $\mathrm{d}x_\tau$ of the type (16), $\mathrm{d}x_\tau = \beta_\tau(v_\tau - q_\tau)\,\mathrm{d}\tau + \gamma_\tau\,\mathrm{d}v_\tau$. For $\tau \geq t$, denote by

$$V_\tau \;=\; \mathsf{E}\big((v_\tau - q_\tau)^2 \mid \mathcal{J}_t^p\big). \tag{148}$$

[23]Because $\mathcal{I}_{t+\mathrm{d}t}^q = \mathcal{I}_t^p \cup \{\mathrm{d}z_t\}$, the two information sets differ only by the infinitesimal quantity $\mathrm{d}z_t$, and thus we can also write $\Sigma_t = \mathsf{E}\big((v_t-p_t)^2|\mathcal{I}_{t+\mathrm{d}t}^q\big) = \mathsf{E}\big((v_t-p_t)^2|\mathcal{I}_t^p\big)$.

The expected profit is

$$\pi_t = \mathsf{E}\left(\int_t^1 (v_\tau + \,\mathrm{d}v_\tau - q_\tau - \lambda_\tau \,\mathrm{d}x_\tau)\,\mathrm{d}x_\tau \mid \mathcal{J}_t^p\right) \tag{149}$$

$$= \int_t^1 \Big(\beta_\tau V_\tau + (1-\lambda_\tau\gamma_\tau)\gamma_\tau\sigma_v^2\Big)\,\mathrm{d}\tau. \tag{150}$$

$V_\tau$ is computed as in the benchmark model, except that $\lambda_\tau$ is replaced by $m_\tau$:

$$\begin{aligned} V_{\tau+\mathrm{d}\tau} &= \mathsf{E}\big((v_{\tau+\mathrm{d}\tau} - q_{\tau+\mathrm{d}\tau})^2 \mid \mathcal{J}_t^p\big) \\ &= V_\tau + (1-\mu_\tau - m_\tau\gamma_\tau)^2\sigma_v^2\,\mathrm{d}\tau + \mu_\tau^2\sigma_e^2\,\mathrm{d}\tau + m_\tau^2\sigma_u^2\,\mathrm{d}\tau - 2m_t\beta_t V_\tau\,\mathrm{d}\tau. \end{aligned} \tag{151}$$

therefore the law of motion of $V_\tau$ is a first order differential equation

$$V_\tau' = -2m_t\beta_t V_\tau + (1-\mu_\tau - m_\tau\gamma_\tau)^2\sigma_v^2 + \mu_\tau^2\sigma_e^2 + m_\tau^2\sigma_u^2, \tag{152}$$

or equivalently $\beta_\tau V_\tau = \frac{-V_\tau' + (1-\mu_\tau - m_\tau\gamma_\tau)^2\sigma_v^2 + \mu_\tau^2\sigma_e^2 + m_\tau^2\sigma_u^2}{2m_\tau}$. Substitute this into (133) and integrate by parts. Since $V_t = 0$, we get

$$\begin{aligned} \pi_t &= -\frac{V_1}{2m_1} + \int_t^1 V_\tau\Big(\frac{1}{2m_\tau}\Big)'\,\mathrm{d}\tau \\ &\quad + \int_t^1 \left(\frac{(1-\mu_\tau - m_\tau\gamma_\tau)^2\sigma_v^2 + \mu_\tau^2\sigma_e^2 + m_\tau^2\sigma_u^2}{2m_\tau} + (1-\lambda_\tau\gamma_\tau)\gamma_\tau\sigma_v^2\right)\,\mathrm{d}\tau. \end{aligned} \tag{153}$$

Since $V_\tau > 0$ can be arbitrarily chosen, in order to get an optimum we must have $\big(\frac{1}{2m_\tau}\big)' = 0$, which is equivalent to $m_\tau = $ constant. For a maximum, the transversality condition $V_1 = 0$ must be also satisfied.

We next turn to the choice of $\gamma_\tau$. The first order condition is

$$-(1-\mu_\tau - m_\tau\gamma_\tau) + (1-\lambda_\tau\gamma_\tau) - \lambda_\tau\gamma_\tau = 0 \quad\Longrightarrow\quad \gamma_\tau = \frac{\mu_\tau}{2\lambda_\tau - m_\tau} = \frac{\mu_\tau}{\lambda_\tau + \mu_\tau\rho_\tau}. \tag{154}$$

Thus, we obtain a nonzero flow trading component. The second order condition is $\lambda_\tau + \mu_\tau\rho_\tau > 0$. There is also a second order condition with respect to $\beta$: $m_\tau > 0$: see Footnote 22.

Next, we derive the pricing rules from the market maker's zero profit conditions. As in the benchmark model, we compute

$$\lambda_t \;=\; \frac{\mathsf{Cov}_t(v_1,\, \mathrm{d}y_t)}{\mathsf{Var}_t(\,\mathrm{d}y_t)} \;=\; \frac{\mathsf{Cov}_t(v_1, \beta_t(v_t - p_t)\,\mathrm{d}t + \gamma_t\,\mathrm{d}v_t +\, \mathrm{d}u_t)}{\mathsf{Var}(\beta_t(v_t - p_t)\,\mathrm{d}t + \gamma_t\,\mathrm{d}v_t +\, \mathrm{d}u_t)} \;=\; \frac{\beta_t\Sigma_t + \gamma_t\sigma_v^2}{\gamma_t^2\sigma_v^2 + \sigma_u^2}, \tag{155}$$

$$\rho_t \;=\; \frac{\mathsf{Cov}_t(\,\mathrm{d}z_t,\, \mathrm{d}y_t)}{\mathsf{Var}_t(\,\mathrm{d}y_t)} \;=\; \frac{\gamma_t\sigma_v^2}{\gamma_t^2\sigma_v^2 + \sigma_u^2}, \tag{156}$$

$$\mu_t \;=\; \frac{\mathsf{Cov}_t(v_1,\, \mathrm{d}z_t - \rho_t\,\mathrm{d}y_t)}{\mathsf{Var}_t(\,\mathrm{d}z_t - \rho_t\,\mathrm{d}y_t)} \;=\; \frac{-\rho_t\beta_t\Sigma_t + (1-\rho_t\gamma_t)\sigma_v^2}{(1-\rho_t\gamma_t)^2\sigma_v^2 + \rho_t^2\sigma_u^2 + \sigma_e^2}. \tag{157}$$

By the same arguments as for the benchmark model, $\Sigma_t = (1-t)\Sigma_0$, $\beta_t = \frac{\beta_0}{1-t}$, and $\beta_t\Sigma_t$, $\lambda_t$, $\rho_t$, $\mu_t$ are constant. Since $\Sigma_t$ satisfies the same Equation (152) as $V_t$, and $\Sigma_t' = -\Sigma_0$, we obtain

$$-\Sigma_0 \;=\; -2m_t\beta_t\Sigma_t + (1-\mu_\tau - m_\tau\gamma_\tau)^2\sigma_v^2 + \mu_\tau^2\sigma_e^2 + m_\tau^2\sigma_u^2. \tag{158}$$

We now define the following constants:

$$a \;=\; \frac{\sigma_u^2}{\sigma_v^2}, \quad b \;=\; \frac{\sigma_e^2}{\sigma_v^2}, \quad c \;=\; \frac{\Sigma_0}{\sigma_v^2}, \tag{159}$$

$$f \;=\; \frac{\gamma^2}{a}, \quad \tilde{\lambda} \;=\; \lambda\gamma, \quad \tilde{\rho} \;=\; \rho\gamma, \quad \nu = \frac{\beta_0\Sigma_0}{\sigma_u^2}\gamma, \quad \tilde{m} = m\gamma. \tag{160}$$

With these notations, Equations (154)–(158) become

$$\begin{aligned} \tilde{\lambda} \;&=\; \mu(1-\tilde{\rho}), \quad \tilde{\lambda} \;=\; \frac{\nu + f}{1+f}, \quad \tilde{\rho} \;=\; \frac{f}{1+f}, \quad \mu \;=\; \frac{1-\nu}{1+b(1+f)} \\ c \;&=\; \frac{2\nu}{f} - (1-\mu-\tilde{m})^2 - \mu^2 b - \frac{\tilde{m}^2}{f}. \end{aligned} \tag{161}$$

Substitute $\tilde{\lambda}$, $\tilde{\rho}$, $\mu$ in $\tilde{\lambda} = \mu(1-\tilde{\rho})$ and solve for $\nu$:

$$\nu \;=\; \frac{1-(1+b)f - bf^2}{2+b+bf} \;=\; \frac{1+f}{2+b+bf} - f. \tag{162}$$

The other equations, together with $\tilde{m} = \tilde{\lambda} - \mu\tilde{\rho}$, imply

$$\begin{aligned} \tilde{\lambda} \;&=\; \frac{1}{2+b+bf}, \quad \tilde{\rho} \;=\; \frac{f}{1+f}, \quad \mu \;=\; \frac{1+f}{2+b+bf}, \quad \tilde{m} \;=\; \frac{1-f}{2+b+bf}, \\ 1+c \;&=\; \frac{(1+bf)(1+f)^2}{f(2+b+bf)^2}. \end{aligned} \tag{163}$$

Putting together (162) and the last equation in (163), we compute

$$\beta_0 \;=\; \frac{a}{c\gamma}\,\nu \;=\; \frac{a^{1/2}}{cf^{1/2}}\,\nu \;=\; \frac{a^{1/2}}{cf^{1/2}(1+c)}\,\frac{1}{(1+f)^{1/2}}\left(c+(1-f)\frac{1+b+bf}{2+b+bf}\right). \quad (164)$$

Now substitute $a$, $b$, $c$ from (159) in Equations (163)–(164) and use $\gamma = a^{1/2}f^{1/2}$ to obtain Equations (18)–(23). One can also check that the second order conditions $\lambda+\mu\rho > 0$ and $m > 0$ are equivalent to $f \in (-1,1)$. Next, we show that the equation $1+c = \frac{(1+bf)(1+f)^2}{f(2+b+bf)^2}$ has a unique solution $f \in (-1,1)$, which in fact lies in $(0,1)$. This can be shown by noting that

$$F_b(f) \;=\; 1+c, \quad \text{with} \quad F_b(x) \;=\; \frac{(1+bx)(1+x)^2}{x(2+b+bx)^2}. \quad (165)$$

One verifies $F_b'(x) = \frac{(x+1)(x-1)(2+b+3bx)}{x^2(2+b+bx)^2}$, so $F_b(x)$ decreases on $(0,1)$. Since $F_b(0) = +\infty$ and $F_b(1) = \frac{1}{1+b} < 1$, there is a unique $f \in (0,1)$ so that $F_b(f) = 1+c$.[24]

## B.2 Proof of Proposition 1

We use the notations from the proof of Theorem 1. We start by showing that $\mu^F < \mu^B$; by computation, $\frac{1+f}{2+b+bf} < \frac{1}{1+b}$ is equivalent to $f < 1$, which is true since $f \in (0,1)$.

We show that $\lambda^F > \lambda^B$, i.e., $\frac{(c+1)^{1/2}}{a^{1/2}}\frac{1}{(1+bf)^{1/2}(1+f)} > \frac{c^{1/2}}{a^{1/2}}\left(1+\frac{b}{c(1+b)}\right)^{1/2}$. After squaring the two sides, and using $1+c = \frac{(1+bf)(1+f)^2}{f(2+b+bf)^2}$, we need to prove that $\frac{1}{f(2+b+bf)^2} > c + 1 - \frac{1}{1+b}$, or equivalently $\frac{1}{1+b} > \frac{(1+bf)(1+f)^2}{f(2+b+bf)^2} - \frac{1}{f(2+b+bf)^2}$. This can be reduced to proving $1+b+(1-f)(1+bf) > 0$, which is true.

The same type of calculations can be used to show that $\beta_0^F < \beta_0^B$, or to prove the other comparative statics.

## B.3 Proof of Proposition 2

In the benchmark model, $Var(dx_t) = (\beta_t^B)^2\Sigma_t \mathrm{d}t^2$ and $Var(du_t) = \sigma_u^2\mathrm{d}t$. Therefore, $IPR_t^B = 0$.

In the fast model, $Var(dx_t) = (\gamma_t^B)^2\sigma_v^2\mathrm{d}t$. Therefore, $IPR_t^F = (\gamma_t^B)^2\sigma_v^2/((\gamma_t^B)^2\sigma_v^2 +$

[24] One can check that $F_b(x) = 1+c$ has no solution on $(-1,0)$: When $b \le 1$, $F_b(x) < 0$ on $(-1,0)$. When $b > 1$, $F_b(x)$ attains its maximum on $(-1,0)$ at $x^* = -\frac{2+b}{3b}$, for which $F_b(x^*) = \frac{(b-1)^3}{b(b+2)^3} < 1$.

$\sigma_u^2) = f/(f+1)$, using the equation for $\gamma_t^F$ in Theorem 1.

## B.4 Proof of Proposition 3

We start with a useful preliminary result:

**Lemma 2.** *In the benchmark model and in the fast model, for all $s < u$, we have*

$$Cov(v_s - p_s, v_u - p_u) = \Sigma_s \left(\frac{1-u}{1-s}\right)^{m\beta_0}, \tag{166}$$

$$Cov(dv_s, v_u - p_u) = (1 - m\gamma - \mu)\sigma_v^2 \left(\frac{1-u}{1-s}\right)^{m\beta_0} ds, \tag{167}$$

*where $m \equiv \lambda - \mu\rho$.*

*Proof.* We start from

$$\begin{aligned} Cov(v_s - p_s, v_u - p_u) &= Cov(v_s - p_s, v_s - p_s) - \int_s^u Cov(v_s - p_s, dp_h)\, dh \\ &= \Sigma_s - \int_s^u Cov(v_s - p_s, m\beta_h(v_h - p_h))\, dh \end{aligned}$$

Differentiating with respect to $u$ we obtain

$$\frac{\partial}{\partial u} Cov(v_s - p_s, v_u - p_u) = -m\beta_u Cov(v_s - p_s, v_u - p_u),$$

which rewrites as

$$\frac{\partial}{\partial \tau} \log Cov(v_s - p_s, v_u - p_u) = -m\beta_u = -m\beta_0 \frac{1}{1-u} = m\beta_0 \frac{\partial}{\partial u} \log(1-u).$$

Integrating between $s$ and $u$ and using $Cov(v_s - p_s, v_s - p_s) = \Sigma_s$, we obtain equation (166).

Similarly, we have

$$\begin{aligned} Cov(dv_s, v_u - p_u) &= Cov(dv_s, dv_s - dp_s) - \int_s^u Cov(dv_s, dp_h)\, dh \\ &= (1 - m\gamma - \mu)\sigma_v^2 ds - \int_s^u m\beta_h Cov(dv_s, v_h - p_h)\, dh. \end{aligned}$$

Proceeding as above we obtain (167). □

We can now prove Proposition 3. The formula in the benchmark model follows immediately from Equation (166). In the fast model, the auto-covariance of the order flow is of order $\mathrm{d}t^2$ while the variance is of order $\mathrm{d}t$, therefore the autocorrelation is of order $\mathrm{d}t$, which is zero in continuous time.

## B.5 Proof of Proposition 5

Follows immediately from Lemma 2.

## B.6 Proof of Lemma 1

We start from

$$\Sigma_t = Var(v_t - p_t) = Var(v_t) + Var(p_t) - 2Cov(v_t, \int_0^t dp_\tau).$$

We have $Var(v_t) = \Sigma_0 + t\sigma_v^2$. Since the price is a martingale and given that we prove in the proof of Proposition 7 that the volatility of the price is equal to the volatility of the asset value, we have $Var(p_t) = t(\Sigma_0 + \sigma_v^2)$. Finally, using that price change cannot be correlated with future innovation in asset value, we obtain equation (41).

## B.7 Proof of Proposition 7

In the benchmark model, $Var(p_{t+\mathrm{d}t} - q_t) = (\lambda^B)^2\sigma_u^2\mathrm{d}t$ and $Var(q_{t+\mathrm{d}t} - p_{t+\mathrm{d}t}) = (\mu^B)^2\sigma_v^2\mathrm{d}t$. Using the equilibrium parameter values of Theorem 1 we obtain $Var(dp_t) = \Sigma_0 + \sigma_v^2 t$.

Similarly, in the fast model, $Var(p_{t+\mathrm{d}t} - q_t) = (\lambda^F)^2((\gamma^F)^2\sigma_v^2 + \sigma_u^2)\mathrm{d}t$ and $Var(q_{t+\mathrm{d}t} - p_{t+\mathrm{d}t}) = (\mu^B)^2((1 - \rho^F\gamma^F)^2\sigma_v^2 + \sigma_e^2 + (\rho^F)^2\sigma_u^2)\mathrm{d}t$. Using the equilibrium parameter values of Theorem 1, we obtain that $Var(p_{t+\mathrm{d}t} - q_t)$ is higher than in the benchmark, $Var(q_{t+\mathrm{d}t} - p_{t+\mathrm{d}t})$ is lower than in the benchmark, and $Var(dp_t) = \Sigma_0 + \sigma_v^2 t$ is the same as in the benchmark.

## B.8 Proof of Proposition 10

When returns are computed using post-trade quotes, we have

$$r_t = \mu_{t-\Delta t}(\Delta z_{t-\Delta t} - \rho_{t-\Delta t}\Delta y_{t-\Delta t}) + \lambda_t \Delta y_t.$$

Using that $\Delta x_{t-\Delta t} \approx \gamma_{t-\Delta t}\Delta v_{t-\Delta t}$ when $\Delta t \to 0$, and that $\gamma_t$, $\lambda_t$, $\mu_t$, and $\rho_t$ are constant over time, the above equation rewrites as

$$r_t \approx \lambda\Delta x_t + \lambda\Delta u_t + \mu(1/\gamma - \rho)\Delta x_{t-\Delta t} - \mu\rho\Delta u_{t-\Delta t} + \mu\Delta e_{t-\Delta t}.$$

Similarly, when pre-trade quotes are used, we have

$$\begin{aligned} r_t &= \lambda_t\Delta y_t + \mu_t(\Delta z_t - \rho_t\Delta y_t) \\ &= (\lambda - \mu\rho + \mu/\gamma)\Delta x_t + (\lambda - \mu\rho)\Delta u_t + \mu\Delta e_t. \end{aligned}$$

## B.9 Proof of Proposition 11

In the limit $\Delta t \to 0$, we have

$$\begin{aligned} &Cov(\Delta x_j(n), r_{j+1}(n)) = \mu\gamma(1-\rho\gamma)\sigma_v^2\Delta t, \\ &Var(\Delta x_j) = n\gamma^2\sigma_v^2\Delta t, \\ &Var(r_{j+1}) = (\sigma_v^2 + \Sigma_0)n\Delta t. \end{aligned}$$

Therefore

$$Corr(\Delta x_j, r_{j+1}) = \frac{\mu(1-\rho\gamma)\sigma_v}{n\sqrt{\sigma_v^2 + \Sigma_0}}$$

is decreasing in $n$ and goes to 0 when $n$ goes to infinity.

## B.10 Proof of Proposition 12

We consider the limit $\Delta t \to 0$ and $n \to +\infty$ such that $n\Delta t = \tau$ is fixed. In this case, we have $Var(\Delta x_j) = Var(x_{t+\tau} - x_t)$, where $t = (j-1)\tau$. Then, we can write

$$
\begin{aligned}
Var(x_j) &= Var(\int_{s=t}^{t+\tau} \beta_s(v_s - p_s)ds + \gamma_s dv_s) \\
&= \int_{s=t}^{t+\tau} \gamma_s^2 Var(dv_s) \\
&\quad +2\int_{s=t}^{t+\tau}\int_{u=s}^{t+\tau} \beta_s\beta_u Cov\left(v_s - p_s, v_u - p_u\right) ds\, du \\
&\quad +2\int_{s=t}^{t+\tau}\int_{u=s}^{t+\tau} \gamma_s\beta_u Cov\left(dv_s, v_u - p_u\right) du.
\end{aligned}
$$

It then follows from Lemma 2 that

$$Var(x_j) = \gamma^2\sigma_v^2\tau + \beta_t(\beta_0\Sigma_0 + \gamma(1 - m\gamma - \mu)\sigma_v^2)\tau^2 + o(\tau^2)$$

when $\tau$ is small and where $m \equiv \lambda - \mu\rho$. Using that $Var(u_j) = \sigma_u^2\tau$, we obtain

$$IPR_j = \frac{\gamma^2\sigma_v^2}{\gamma^2\sigma_v^2 + \sigma_u^2} + \frac{\sigma_u^2}{(\gamma^2\sigma_v^2 + \sigma_u^2)^2}\beta_t(\beta_0\Sigma_0 + \gamma(1 - m\gamma - \mu)\sigma_v^2)\tau + o(\tau).$$

## B.11 Proof of Proposition 13

We consider the limit $\Delta t \to 0$ and $n \to +\infty$ such that $n\Delta t = \tau$ is fixed. In this case, we have

$$
\begin{aligned}
&Cov(\Delta x_j, \Delta x_{j+1}) = \beta_t(\beta_0\Sigma_0 + \gamma(1 - m\gamma - \mu)\sigma_v^2)\tau^2 + o(\tau^2), \\
&Var(\Delta x_j) = Var(\Delta x_{j+1}) = \gamma^2\sigma_v^2\tau + \beta_t(\beta_0\Sigma_0 + \gamma(1 - m\gamma - \mu)\sigma_v^2)\tau^2 + o(\tau^2).
\end{aligned}
$$

Therefore

$$Corr(\Delta x_j, \Delta x_{j+1}) = \frac{\beta_t^2\Sigma_t + \beta_t\gamma_t(1 - m\gamma - \mu)\sigma_v^2}{\gamma^2\sigma_v^2}\tau + o(\tau).$$

# References


[1] Back, Kerry, Henry Cao, and Gregory Willard (2000): "Imperfect Competition

among Informed Traders," *Journal of Finance*, 55, 2117–2155.

[2] Back, Kerry, and Hal Pedersen (1998): "Long-Lived Information and Intraday Patterns," *Journal of Financial Markets*, 1, 385–402.

[3] Biais, Bruno, Thierry Foucault, and Sophie Moinas (2011): "Equilibrium High Frequency Trading," Working Paper.

[4] Brogaard, Jonathan (2010): "High Frequency Trading and Its Impact on Market Quality," Working Paper.

[5] Brogaard, Jonathan (2011): "The Activity of High Frequency Traders," Working Paper.

[6] Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega (2009): "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market," Working Paper, Board of Governors of the Federal Reserve System.

[7] Chau, Minh, and Dimitri Vayanos (2008): "Strong-Form Efficiency with Monopolistic Insiders," *Review of Financial Studies*, 21, 2275–2306.

[8] Hasbrouck, Joel (1991a): "Measuring the Information Content of Stock Trades," *Journal of Finance*, 46, 179–207.

[9] Hasbrouck, Joel (1991b): "The Summary Informativeness of Stock Trades: An Econometric Analysis," *Review of Financial Studies*, 4, 571–595.

[10] Hasbrouck, Joel (1996): "Order Characteristics and Stock Price Evolution: An Application to Program Trading," *Journal of Financial Economics*, 41, 129–149.

[11] Hasbrouck, Joel, and Gideon Saar (2011): "Low-Latency Trading," Working Paper.

[12] Hendershott, Terrence, Charles Jones, and Albert Menkveld (2011): "Does Algorithmic Trading Improve Liquidity?," *Journal of Finance*, 66, 1–33.

[13] Hendershott, Terrence, and Pamela Moulton (2011): "Automation, Speed, and Stock Market Quality: The NYSE's Hybrid," *Journal of Financial Markets*, 14, 568–604.

[14] Hendershott, Terrence, and Ryan Riordan (2010): "Algorithmic Trading and Information," Working Paper.

[15] Hendershott, Terrence, and Ryan Riordan (2011): "High Frequency Trading and Price Discovery," Working Paper.

[16] Jovanovic, Boyan, and Albert Menkveld (2011): "Middlemen in Limit-Order Markets," Working Paper.

[17] Karatzas, Ioannis, and Steven Shreve (1991): *Brownian Motion and Stochastic Calculus*, Springer Verlag.

[18] Kirilenko, Andrei, Albert Kyle, Mehrdad Samadi, and Tugkan Tuzun (2011): "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market," Working Paper.

[19] Kyle, Albert (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1335.

[20] Menkveld, Albert (2012): "High Frequency Trading and the New-Market Makers," Working Paper.

[21] Pagnotta, Emiliano, and Thomas Philippon (2011): "Competing on Speed," Working Paper.

[22] SEC (2010): "Concept Release on Equity Market Structure," Release No. 34-61358; File No. S7-02-10.

# How (Un)Informed Is Trading?

Eric K. Kelley and Paul C. Tetlock*

May 2012

**Abstract**

We estimate a structural model of strategic trader behavior that sheds light on the determinants of trading volume and stock returns. Our novel identification approach exploits enormous empirical variation in trading and volatility associated with the time of day and public news arrival. Over 95% of trading occurs during regular market hours (9:30am to 4pm), even though prices exhibit considerable volatility during extended hours, especially when news arrives. For the model to explain the data, discretionary liquidity trading must constitute the bulk of trading volume and must increase significantly after news arrives. However, from 2001 to 2010, informed trading increasingly contributes to volume and stock price discovery because our estimate of the cost of acquiring private information falls by a factor of 12 in this decade.

* University of Arizona, ekelley@eller.arizona.edu and Columbia University, paul.tetlock@columbia.edu. An earlier version of this paper circulated under the title “Round-the-Clock News, Trading, and Returns.” We thank Lauren Cohen, Rick Sias, and seminar participants at Arizona, Columbia, Georgia, and Tennessee for helpful discussions and Dow Jones for providing the news data used in this study.

DeBondt and Thaler (1995) argue that the high trading volume observed in financial markets "is perhaps the single most embarrassing fact to the standard finance paradigm." Since then, Chordia, Roll, and Subrahmanyam (2011) show that turnover in US markets has actually *increased* fivefold, implying annual volume is now *tens of trillions* of dollars. Although they argue that informed trading drives the recent trend, a complete characterization of traders' motives remains elusive. This paper attempts to fill this void by estimating a structural model with strategic informed and uninformed traders. The model sheds light on the determinants of trading volume and the relative importance of private and public information in price discovery.

In our model, as in Admati and Pfleiderer (1988), both informed and (uninformed) "liquidity" traders optimally choose to trade when others are trading because such clustering behavior minimizes traders' impact on prices. As in standard models, informed traders choose to enter by weighing the cost of acquiring and acting on information ($c$ parameter) against their expected trading profits. The novel feature is that we allow liquidity traders' net benefits of learning and acting on their trading needs ($h$ parameter) to vary with the time of day and the occurrence of news. This parameterization captures variation in traders' awareness of data, portfolio monitoring costs, rebalancing needs, and opportunity costs. Such variation causes trader attention to the market, and thus trading volume, volatility, and liquidity, to fluctuate as well.

To identify the model parameters, we exploit enormous empirical variation in trading and volatility associated with the time of day and news arrival in the electronic trading era. This study is the first to adopt this promising approach. Since 2000, revolutions in information and trading technology have enabled virtually round-the-clock market-related activity in US stocks. Newswires now arrive continuously throughout the day. Electronic communication networks give any institution or retail trader with a brokerage account the ability to trade outside regular market hours (from 9:30am to 4pm). Despite having this ability, few choose to trade in extended hours. Less than 5% of total trading in our sample takes place in the pre-market (7am to 9:30am in our study) and the after-market (4pm to 6:30pm). In contrast, stock return volatility during extended hours periods is more than half of that during the regular market. In fact, in periods when news arrives, extended hours volatility is on par with regular hours volatility.

We use these strong contemporaneous relationships between volume, volatility, and news arrival in each intraday period to estimate our structural model's parameters. Our estimates of the parameter ($h$) designed to measure time-varying trader attention to the market are highest during regular trading hours and trading periods soon after news arrives, which accords with intuition. In the model, these high $h$ values motivate liquidity traders to enter the market and indirectly spur informed traders to enter by increasing their expected profits. Our estimates of the parameter ($c$) designed to measure the cost of acquiring private information are low enough so that dozens of informed traders choose to acquire information and trade during the regular market, but sufficiently high so that few opt to participate during extended hours sessions. When news occurs, more informed traders choose to enter because more liquidity traders enter the market and increase liquidity.

For the model to explain the data, liquidity trading must account for the bulk of trading activity, particularly during regular market hours. The central reason is that compared to extended hours periods, the level of trading volume is quite high in relation to volatility, suggesting non-informational trading is important. Our parameter estimates imply that non-informational trading accounts for about 95% of regular market volume in 2001-2005, though it declines to 85% in 2006-2010. This modest decline is matched by an increase in informed trading: from 3.6% to 11.4% of volume in the regular period. In large cap stocks (with size over

$10B), informed trading accounts for much more of regular, pre-, and after-market trading (18%, 41%, and 37%, respectively). These patterns are consistent with the qualitative findings in Chordia, Roll, and Subrahmanyam (2011) who analyze how empirical proxies for informed trading vary over time and across firms.

Our estimates indicate that dramatic changes in the information cost parameter ($c$) explain these patterns in informed trading. Whereas the attention parameter ($h$) is quite stable over time, information costs fall by a factor of 12 in the past decade. The decline in $c$ is most pronounced for large cap stocks, where it falls by a factor of 30. These estimates are consistent with the theory that the widespread adoption of new information gathering and trading technologies has transformed trading, especially in large stocks.

The stark reduction in the cost of acquiring information has important implications for price discovery, too. In 2001 to 2005, private information revealed through trading explains just 8% of the variance in returns during the regular market, whereas it explains 76% of variance from 2006 to 2010. This change is again most stark for large cap stocks, though it pervades small caps (under $1B) and mid caps (between $1B and $10B) as well. These results arise because it is far less costly to gather and trade on value-relevant information in the later time period. By the time this information is publicly observable, most of it is already incorporated in prices. As a result, we find that the role of public information in price discovery has diminished over time.

Next we test our model's predictions for how prices and volume will respond in periods *after* news arrival. In the model, to mimic the 24-hour news cycle, we assume news affects business activities and awareness of data ($h$) for one full trading day. We find that this simple model correctly predicts the price response to news is immediate whereas the volume response can be delayed. Intuitively, because prices respond to information while volume arises primarily from non-informational motives, the two need not coincide.[1] For example, when public news arrives during the after-market period, price responds immediately, but the bulk of abnormal trading occurs during the regular market on the following trading day. The delayed and prolonged trading after news contrasts with the timing in models that generate trading mainly through differences in beliefs, such as those based on pre-event or event-period information (*e.g.*, Kim and Verrecchia (1991, 1994, and 1997)) or those based purely on differences in opinion (*e.g.*, Harris and Raviv (1993), Kandel and Pearson (1995), and Banerjee and Kremer (2010)). Such models make the counterfactual prediction that the impacts of news on volume and volatility coincide because both are caused by changes in investors' beliefs.

We further distinguish our model of liquidity trading from belief-based models of trading by investigating market activity around news events sorted on the basis of changes in analysts' beliefs about quarterly earnings. We separately consider news events in which analyst forecast dispersion decreases, remains similar, and increases within one week after the news. Although news events in which dispersion either increases or decreases are associated with the highest return volatility, such events are actually associated with slightly *lower* trading activity. This lack of trading when beliefs change appears inconsistent with the disagreement models of trading volume, but it can be reconciled with models of discretionary liquidity traders. Such traders may expect their trades to have more impact on prices when news has greater impact on the belief distribution, thereby deterring them from entering the market and trading.

We organize the paper as follows. Section 1 reviews the relevant literature. Section 2 presents the structural model of market activity and explains which empirical moments identify

[1] These predictions are not mechanically generated by the fact that we allow trader attention to vary with news. We also estimate a version of our model in which news does not affect attention but it does affect the extent of acquirable private information. This version makes nearly identical equilibrium predictions: volume after news again arises primarily from discretionary liquidity traders, who in this case choose to enter because news reduces information asymmetry and lowers their trading costs.

the key model parameters. Section 3 describes how we apply the model to the empirical data. It also describes the features of our databases on trading activity and news and provides summary statistics of the key moments. Section 4 presents our estimates of the model parameters, along with estimates of two alternative parameterizations. Section 5 analyzes decompositions of return variance and trading volume under the main parameterization. Section 6 tests the distinct predictions of the discretionary liquidity trader model and belief-based models of trading. Section 7 concludes and suggests directions for future research.

## 1. Literature Review

In this section, we briefly review three strands of literature: the determinants of trading volume, the roles of private and public information in stock price formation, and the extended hours markets. Readers with knowledge of these areas may prefer to skip this section.

### *A. Determinants of Trading*

Classic models such as Milgrom and Stokey (1982) and Kyle (1985) consider two motives for trading: information and liquidity, where liquidity is simply an exogenous demand to transact for some reason other than information. While these models form the basis for most theory and intuition in market microstructure, they do not explain important stylized facts such as the tendency of trading to cluster near the beginning and ending of trading periods (Jain and Joh, 1986) or following information releases such as earnings announcements (Beaver, 1968). In light of this, Admati and Pfleiderer (1988) introduce *discretionary* liquidity traders who endogenously choose when to trade with the goal of minimizing their expected trading losses. To minimize the price impact of trading, each uninformed trader prefers to trade when other uninformed traders choose to enter the market, leading to clustering in trading volume.

In most other theories, changes in traders' relative beliefs cause trading around information releases. Kim and Verrecchia (1991) introduce a model in which a public signal resolves trader disagreement initially caused by noisy private signals about firm value. In contrast, Kim and Verrecchia (1994) model public signals that generate new private information and thus disagreement. Kim and Verrecchia (1997) allow for both types of public signals—that resolve or generate disagreement—and show that both can generate trading. Harris and Raviv (1993) offer an alternative model in which traders have different opinions about the impact of a public signal, which causes trading when the signal is informative. Similarly, Kandel and Pearson (1995) model volume that occurs when traders update their estimates of firm value using different likelihood functions, motivated by different prior beliefs. A recent model by Banerjee and Kremer (2010) presents a simple characterization of volume in differences in opinion models as resulting from changes in the level of investor disagreement.

Empirical work attributes most of the increase in trading volume in US stocks from $3.5 trillion in 1994 to $32.0 trillion in 2010 to institutional trading, which is often perceived as informed (Chordia et al. (2011)). There is also some support for the numerous theories based on information or differences in opinion in conjunction with information releases (see Bamber, Barron, and Stevens, 2011 for a thorough review). However, the notion of discretionary liquidity traders emphasized in our model has received almost no attention in the empirical literature. One notable exception is Kross, Ha, and Heflin (1994) who find that absolute change in beta around earnings announcements is positively related to announcement-period trading and attribute this trading to portfolio rebalancing, a specific type of discretionary liquidity trading. A second is

Chae (2005), who finds that trading volume declines before and then increases after scheduled earnings announcements. He interprets this pattern as discretionary liquidity traders postponing their orders until the information release resolves information asymmetry.

*B. The Roles of Private and Public Information on Price Formation*

Related research studies the roles of public and private information in price formation. French and Roll (1986) compare trading and non-trading periods and provide evidence that 86% of stock return variance is *not* caused by the arrival of public news and is *not* transitory. By process of elimination, one could infer that private information is the chief driver of price movements. Bolstering this view, evidence in Barclay, Litzenberger, and Warner (1990) and Ito, Lyons, and Melvin (1998) shows that hourly stock return variance is several times higher during trading hours, when informed trading can occur, as compared to non-trading hours. Taking a different approach, Roll (1988), Cutler, Poterba, and Summers (1989), Berry and Howe (1994), and Mitchell and Mulherin (1994) consider the volatility-news relationship. These studies complement the message of French and Roll (1986) by showing that public information arrival is only weakly linked to market activity.

In contrast, Jones, Kaul, and Lipson (1994) compare return variance during *endogenously determined* non-trading periods to variance when trading occurs. They find that a large proportion of variance occurs in the absence of trading, providing "evidence that public (versus private) information is the major source of short-term return volatility." Ultimately, the relative importance of private and public information in market activity remains unknown both because these constructs are difficult to measure and because the world has changed dramatically since the initial studies. In the past 15 years, news production and trading volume have increased by an order of magnitude and return variance has risen sharply and fallen precipitously, as shown by Tetlock (2010), Chordia et al. (2011), and Brandt et al. (2009), respectively. Our model provides a simple variance decomposition into private and public information components that speaks directly to these issues.

### *C. The Extended Hours Markets*

Finally, a small literature investigates extended hours markets, which many view as less liquid and important than the regular market. Barclay and Hendershott (2004) find that bid-ask spreads outside normal trading hours are about three or four times those during normal trading and attribute the difference to greater adverse selection. Barclay and Hendershott (2003) show that information asymmetry and price discovery per trade are highest during the pre-market.

Two other papers are more closely related to our work. Zdorovtsov (2004) finds that volatility in both extended and regular hours increases in the presence of public news. Jiang, Likitapiwat, and McInish (2011) report higher trading volume and lower quoted and effective spreads during extended hours periods containing earnings announcements than in periods without announcements. They also find that earnings announcement periods contribute more to 24-hour price changes than do non-announcement periods. We build on this literature by providing a comprehensive analysis of the relationship between volume, volatility, and news arrival during regular and extended trading hours in the context of a structural model.

## 2. Structural Model of Market Activity

### *A. Theory*

Inspired by Admati and Pfleiderer (1988), we model prices and volume in a market for a single risky asset (hereafter "stock"), where liquidity traders and informed traders endogenously choose to participate. Each trading round in the model corresponds to an intraday period, such as the pre-market period. Because all traders choose whether to participate in each period, time variation in market activity can be rich even though there is effectively only one period in the model. The timing in each period is that public information arrives, traders choose whether to enter the market, private signals are realized, and trading occurs at a price set by the market maker.

The stock's value ($F$) at some distant future trading round $T$ is given by:

$$F = \overline{F} + \sum_{t=1}^{T} d_t + \sum_{t=1}^{T} \varepsilon_t, \tag{1}$$

where $\overline{F}$ is the initial value of the stock in period 0, $t$ indexes periods, and $d_t$ and $\varepsilon_t$ are independently distributed with means of zero and variances of $\sigma_{dt}^2$ and $\sigma_{\varepsilon t}^2$. In period $t$, both $d_t$ and $\varepsilon_t$ are revealed publicly. The key difference is that informed traders can acquire information about $d_t$ one period in advance, whereas information about $\varepsilon_t$ cannot be acquired. Thus, $d_t$ and $\varepsilon_t$ represent tractable and intractable information, respectively, in the sense of Mendelson and Tunca (2004). We assume that public news ($news_t \in \{0,1\}$) is the sole mechanism for releasing intractable information, implying that $\varepsilon_t = 0$ in non-news periods when $news_t = 0$.

In each period, each informed trader ($i$) chooses whether to enter the market based on whether her expected trading profits exceeds her information gathering and processing costs ($c$). By paying $c$, informed traders can observe $d_{t+1}$, which is normally distributed. Each informed trader is risk neutral and thus selects demand to maximize expected profits given her signal. The number of informed traders ($m_t$) is determined endogenously by free entry subject to a zero trading profit condition.

Each discretionary liquidity trader ($j$) enters the market if his expected benefit of rebalancing net of entry costs exceeds his expected trading losses. The main entry cost is the opportunity cost of time and attention required to monitor his portfolio exposures and determine his optimal trade ($y_{jt}$). If he enters the market, discretionary liquidity trader $j$ buys $y_{jt}$ shares, where $y_{jt}$ is normally distributed with mean zero and variance $\sigma_{yt}^2$. We allow opportunity costs of time, attention, and rebalancing needs to vary across traders and summarize these considerations using a simple function $h_t/j$. This implies that liquidity traders ($j$) are indexed by their expected rebalancing benefit net of entry costs, which decrease from $h_t$ to 0 as $j$ increases. The number of discretionary liquidity traders ($k_t$) is determined endogenously by the free entry of liquidity traders who choose to participate. There are no non-discretionary liquidity traders, though liquidity traders with low $j$ values exhibit strong desires to trade—especially when $h_t$ is high.

News not only releases intractable information in the current intraday period, but it also may have longer-term impacts on trader attention and rebalancing needs, as captured by $h_t$. Motivated by the modern day 24-hour news cycle, we allow $h_t$ to depend on whether news has arrived in the past 24 hours. Sequences of related news stories about a firm may unfold during a 24-hour period, bringing the news and the stock to the attention of more traders who may realize their current stock holdings differ significantly from their optimal allocation. Formally, we assume $h_t = h(RecentNews_t, t)$, where recent news is defined as:

$$RecentNews_t = 1-(1-news_t)(1-news_{t-1})(1-news_{t-2})(1-news_{t-3}). \tag{2}$$

The risk neutral market maker only observes total order flow and thus cannot distinguish between orders from each trader. Assuming market making is competitive, the stock price is set such that the market maker’s expected profit is zero conditional on total order flow. We suppose further that the market maker's pricing function is linear in total order flow:

$$p_t = F_t + \lambda_t Q_t = F_t + \lambda_t\left(\sum x_{it} + \sum y_{jt}\right), \text{ where } F_t = \overline{F} + \sum_{s=1}^{t} d_s + \sum_{s=1}^{t} \varepsilon_s. \tag{3}$$

In this equation, $F_t$ represents expectation of $F$ conditional on public information, $\lambda_t$ is the sensitivity of price to net order flow ($Q_t$).

We solve the model by conjecturing an equilibrium and evaluating whether agents have an incentive to deviate using backward induction. We suppose that the $m_t$ informed traders select their demands ($x_{it}$) to depend linearly on their signals, implying $x_{it} = \beta_t d_{t+1}$, where $\beta_t$ measures trader aggressiveness. In the Appendix, we characterize the unique (symmetric) equilibrium in which informed traders and market makers follow the (same) linear strategies above and both informed and liquidity traders endogenously choose to participate in the market. The key endogenous parameters are the equilibrium sensitivity of prices to order flow ($\lambda_t$), trader aggressiveness ($\beta_t$), and the number of informed ($m_t$) and liquidity traders ($k_t$):

$$\lambda_t = c^{-1}(h_t / c + 1)^{-2}\sigma^2_{dt+1} \tag{4}$$

$$\beta_t = (h_t + c)\sigma^{-2}_{dt+1} \tag{5}$$

$$m_t = h_t / c \tag{6}$$

$$k_t = ch_t(h_t / c + 1)^2 \sigma^{-2}_{dt+1}\sigma^{-2}_{yt}. \tag{7}$$

*B. Empirically Identifying the Model*

We identify the exogenous parameters in the model, $\sigma^2_{dt}$, $\sigma^2_{\varepsilon t}$, $h_t$, and $c$, using return variance ($Var(r_t)$) and average volume ($E(v_t)$) conditional on news ($news_t \in \{0,1\}$) in the four periods. We allow $\sigma^2_{dt}$ and $\sigma^2_{\varepsilon t}$ to vary only across the four intraday periods. Intuitively, return variances in news and non-news periods provide estimates of $\sigma^2_{dt}$ and $\sigma^2_{\varepsilon t}$, which are measures of tractable and intractable information. Examining volume in news and non-news periods allows us to isolate the changing net benefits of entry ($h_t$) for liquidity traders, as well as the cost ($c$) of informed traders acquiring information.

We compute simple returns ($r_t$) using:

$$r_t = p_t - p_{t-1} = \lambda_t Q_t - \lambda_{t-1}Q_{t-1} + d_t + \varepsilon_t. \tag{8}$$

The variance of returns is then:

$$
\begin{aligned}
Var(r_t) &= \sigma_{\varepsilon t}^2 + \lambda_t^2 Var(Q_t) + Var(d_t - \lambda_{t-1} Q_{t-1}) \\
&= \sigma_{\varepsilon t}^2 + (h_t / c)(h_t / c + 1)^{-1} \sigma_{dt+1}^2 + (h_{t-1} / c + 1)^{-1} \sigma_{dt}^2 .
\end{aligned} \tag{9}
$$

We define volume as in Admati and Pfleiderer (1988) using

$$
v = \max(\sum buys, \sum sells), \tag{10}
$$

where buys and sells are measured using quantities of shares. We can decompose expected volume into one half of the sum of buys and sells plus the absolute value of net buying by the market maker, which can be written as:

$$
\begin{aligned}
E(v_t) &= \frac{1}{2}(m_t \beta_t E[|\, d_{t+1} \,|] + k_t E[|\, y_{jt} \,|]) + \frac{1}{2} E[|\, m_t \beta_t d_{t+1} + \sum y_{jt} \,|] \\
&= \frac{h_t (h_t / c + 1)}{\sqrt{2\pi}\sigma_{dt+1}} \left[ 1 + (h_t + c)\sigma_{dt+1}^{-1} \sigma_y^{-1} + \sqrt{1 + c / h_t} \right].
\end{aligned} \tag{11}
$$

The first two terms in brackets come from trading between liquidity and informed traders and among liquidity traders, while the third term accounts for trades involving the market maker.

To eliminate confounding impacts of repeated news stories, we define each firm's first news event in a 24-hour period as $FirstNews_t = news_t (1 - news_{t-1})(1 - news_{t-2})(1 - news_{t-3})$. This means that $FirstNews_t = 1$ if $news_t = 1$ and there was no news in the prior three intraday periods. We compare firm return variance in situations where there is first news to that in situations where there has been no recent news. We use the variable subscript $nt$ to denote periods $t$ in which $FirstNews_t = RecentNews_t = n$. Thus, $Var(r_{1t})$ is the firm's return variance in periods $t$ when the firm has first news (and thus recent news), whereas $Var(r_{0t})$ is return variance in periods when the firm has no recent news (and thus no first news). By varying the $n$ and $t$ subscripts, we separately measure return variance and trading volume depending on whether news occurs during the regular market, pre-market, overnight, and after-market periods. Hereafter, we refer to periods with $FirstNews_t = RecentNews_t = 1$ as news periods and those with $FirstNews_t = RecentNews_t = 0$ as non-news periods.

The lack of overnight trading activity has implications for the model parameters. Because overnight volume is zero, regardless of whether news occurs, overnight $h_t$ is zero for news ($h_{1t}$) and non-news ($h_{0t}$) periods. Similarly, we assume that no tractable information is available overnight while no trading occurs. This implies that $\sigma_{dt}^2 = 0$ for the pre-market period (following the overnight period) when such tractable information would have become public.

After imposing these restrictions, we estimate the remaining 14 parameters: $\sigma_{dt}^2$ in the three trading periods, $\sigma_{\varepsilon t}^2$ in all four intraday periods, one $c$ parameter, three $h_{1t}$ parameters, and three $h_{0t}$ parameters.[2] Fortunately, the four intraday return variance equations and three intraday expected trading volume equations for news and non-news periods ($Var(r_{1t})$, $Var(r_{0t})$, $E(v_{1t})$, and $E(v_{0t})$) provide 14 empirical moments that can be used to exactly identify the 14 parameters above. The next two sections describe the data and procedures used in this estimation.

## 3. Data and Empirical Moments

Our eligible sample spans 2001 to 2010 and includes NYSE, AMEX, and NASDAQ stocks. We pool intraday observations across similarly-sized firms each year in the sample and separately estimate the 14 moments described above for each intraday period (pre-market, regular hours, after-market, and overnight). Then we average across size groups and years to obtain a set of moments for each intraday period. This procedure mitigates measurement error resulting from firm-level estimates, assigns each size group and year equal weight, and allows us to analyze and control for differences across size groups. Throughout the paper, we compute standard errors based on 1,000 block bootstrap samples. The samples are stratified by year, where each year consists of 50 randomly drawn one-week blocks of return and volume data for

[2] Because the impact of news lasts for 24 hours, the values $h_{1t}$ and $h_{0t}$ not only denote the parameters in periods when $FirstNews_t = RecentNews_t = 1$ and $FirstNews_t = RecentNews_t = 0$, respectively, but they also characterize the parameter values in all periods when $RecentNews_t = 1$ and $RecentNews_t = 0$, regardless of whether first news occurs.

all firms and intraday periods in that week. The standard error is the standard deviation of the estimate of interest—*e.g.*, empirical moment or model parameter—across these 1,000 samples. This procedure assumes independence of returns and volume across weekly blocks of data, but it allows for arbitrary correlation of returns and volume at higher frequencies and across firms. The next subsection introduces the data on stock prices, trading volume, and news releases that forms the basis for the moment estimation.

### *A. Data*

For the pre-market, regular market, and after-market, we obtain trade-by-trade price and volume data from the NYSE TAQ database.[3] We do not include any trades during the overnight period because few trades are reported for most of our sample period. We adjust for non-standard opening and closing times such that the regular trading period only includes the hours in which the market is open, and we consider weekends and weekdays in which the market is closed as part of the overnight period. As discussed in the Appendix, we employ several standard techniques from the microstructure literature to compute an accurate trade-based return for each intraday period. We adjust all firm returns for market returns by subtracting the contemporaneous intraday return of the SPDR S&P 500 ETF (SPY). We compute share turnover from each period as the market value of share volume scaled by market cap at the end of the prior calendar year.

We measure firm-specific news using the Dow Jones archive. These data include all DJ newswire and *Wall Street Journal* stories from 2001 to 2010. For each story, DJ provides stock codes indicating which firms are meaningfully mentioned and a timestamp indicating when the

[3] Barclay and Hendershott (2008) argue that extended hours trades and quotes are adequately represented in the TAQ database. See their section 2.1 for a detailed analysis.

story became publicly accessible. To focus on firm-specific news, we only consider stories that mention at most two publicly trades U.S. stocks. The variable $news_{it}$ equals one when news mentions firm $i$ during intraday period $t$ and zero otherwise. For small and mid cap (large cap) firms, we require at least one (two) story mention(s) to constitute news. This convention does not count isolated stories for large cap stocks as news because large firms frequently receive news coverage even when no new public information exists.

We employ additional sample filters based on size, news coverage, and extended hours trading activity from the prior calendar year. First, we retain only stocks having market capitalizations above \$100 million and share prices greater than \$1 at the end of the prior year. Second, we require a firm to have a news story in a minimum of four pre-market periods and four after-market periods. Third, firms must also have trading in at least 20 pre-market periods and 20 after-market periods. Finally, we divide the firms into three size subsamples based on market capitalization from the prior year-end. We define "large cap," "mid cap," and "small cap" stocks as those with market capitalizations within the intervals [\$10B,$\infty$), [\$1B,\$10B), and [\$100M,\$1B), respectively. These size groups contain an average of 95, 245, and 237 firms per year, respectively.

*B. Empirical Moments*

For each of the four intraday periods, Table 1 and Figure 1 describe the probability of news arrival for the full sample and separately for each size group. Three general patterns are noteworthy. First, while the regular period has the highest probability of news, the majority of news stories occur in the other three periods. The unconditional probability of news in a regular period is about 0.15. The probabilities of news in the pre-market, overnight, and after-market periods are 0.10, 0.12, and 0.08, respectively. Furthermore, measured as an hourly arrival rate (not shown), the probability of news is actually highest during the pre-market and after-market periods. Second, regardless of the period, news occurs more frequently for large cap stocks than for mid or small caps. Third, as shown in Figure 1, the probability of news arrival in every intraday period increases substantially around 2003 and plateaus through 2010. These patterns generally hold for the variable $FirstNews_t$ (the probability of the first news occurrence in a 24-hour period) as well.

[Insert Table 1 here.]

[Insert Figure 1 here.]

Table 2 presents conditional return variance moments used in the estimation along with unconditional variances that serve as a benchmark. We account for spurious reversal due to transitory noise in the period price $t$ by computing $Var(r_t)^* = Var(r_t) + Cov(r_t, r_{t-1}) + Cov(r_t, r_{t+1})$. Interestingly, this adjustment affects return variance by less than 15% in each intraday period, implying that microstructure noise is not too severe. To ease comparison across periods, we report hourly volatilities in percent by dividing each variance by the number of hours in the intraday period and then taking the square root.

[Insert Table 2 here.]

The upshot of Table 2 is that news is consistently associated with higher volatility—volatility conditional on *FirstNews*=1 always exceeds that conditional on *RecentNews=0*, and the difference is economically staggering. In the regular period, volatility on news days is about 130% of that on non-news days. In the pre-market, overnight, and after-market periods, this ratio is even higher at 293%, 200%, and 364%, respectively. Qualitatively similar patterns appear within each size group.

We also note that hourly volatility during extended trading hours is of a similar magnitude as that during regular trading, especially when comparing periods with news. This finding is surprising in light of earlier evidence from French and Roll (1986) that volatility when the regular market is open far exceeds volatility when the regular market is closed. Our results may differ from theirs because new technologies have changed the nature of trading and news dissemination.

Table 3 presents conditional volume moments for the pre-, regular, and after-market periods. Numbers in the table are hourly turnover expressed in basis points. Similar to the conditional variance results, there is far more trading in periods with news than in periods without news. For the regular, pre-, and after-market periods, respectively, turnover with news is 145%, 723%, and 688% of that without news. However, unlike the variance patterns, almost all trading takes place during regular market hours irrespective of the occurrence of news. These stark patterns in variance and trading volume are key to identifying the model in Section 2.

[Insert Table 3 here.]

We summarize the volatility and volume results over time in Figure 2. The four panels represent the intraday periods. The bars depict the difference in turnover during news and non-news periods, while the lines depict differences in volatility. This figure reveals that the positive associations between news arrival and volatility and between news arrival and volume are quite robust over time. In the next section, we estimate our structural model's exogenous parameters.

[Insert Figure 2 here.]

## 4. Estimates of the Model Parameters

### *A. Mapping the Model into the Data*

To facilitate comparisons of empirical moments and parameter estimates across firms and over time, we defined scaled versions (denoted by *) of the moments and parameters in terms of each firm's shares outstanding ($\theta$) and share price ($p$):

$$Var(r_t^*) = Var(r_t / p) \tag{12}$$

$$E(v_t^*) = E(v_t / \theta) \tag{13}$$

$$h_t^* = h_t / (p\theta) \tag{14}$$

$$c^* = c / (p\theta) \tag{15}$$

$$\sigma_{dt+1}^* = \sigma_{dt+1} / p \tag{16}$$

$$\sigma_{\varepsilon t}^* = \sigma_{\varepsilon t} / p \tag{17}$$

$$\sigma_y^* = \sigma_y / \theta \tag{18}$$

$$\lambda_t^* = \lambda_t \theta / p \tag{19}$$

$$\sigma_{rt}^* = \sigma_{rt} / p \tag{20}$$

With these definitions, one can verify that the equations expressing the parameter estimates in terms of the empirical moments remain virtually identical to the original equations in Section 3. The only differences are that simple returns become percentage returns, share volume becomes share turnover, and all parameters in the new equations have asterisks. For practical purposes, we measure shares outstanding and share price at the end of the previous period.[4] In what follows, model parameters are scaled as in Equations (12) to (20), but we suppress the asterisk superscripts to economize on notation.

To obtain a conservative estimate of the importance of liquidity trading, we set the volatility of each liquidity trader's demand ($\sigma_y$) equal to the maximum value subject to the constraint that at least one liquidity trader must trade in each period. At least one liquidity trader is necessary for the model to predict that trading will occur, which holds empirically in each period. This maximum value for liquidity trading size is approximately $\sigma_y$ = 0.2 bps of firm value. In the smallest sample firms with market capitalizations of \$100M, this $\sigma_y$ = 0.2 bps trade size corresponds to \$2,000, which is empirically reasonable for a single retail trade. In the largest firms with market capitalizations of \$100B, this implies a typical liquidity trade size of \$2M. One could interpret this large liquidity trade as either a single large institutional investor or as a

[4] Technically, this timing induces a tiny approximation error in the parameter estimates, but it is typically negligible because the average intraday gross return is very close to 1.0.

group of 100 perfectly correlated retail trades of $20,000 each.[5]

*B. Parameter Estimates*

We estimate the model using the efficient generalized method of moments (GMM). Specifically, we minimize the model's squared prediction errors for each moment using a weighting matrix equal to the inverse of the covariance matrix of the empirical moments. Hansen (1982) shows that this weighting scheme is efficient. As explained in Section 3, we obtain the 14 by 14 covariance matrix of the 14 empirical moments using a block bootstrap technique.

This efficient GMM procedure yields a solution for the 14 model parameters that *exactly* fits the 14 empirical moment equations. This demonstrates that our modeling assumptions do not impose any restrictions on parameters that are infeasible given the empirical moments.[6] Table 4 presents point estimates and standard errors for the key model parameters, $h$, $c$, $\sigma_d$, and $\sigma_{\varepsilon 1}$ ($\sigma_\varepsilon(news)$) for each of the four intraday periods. We compute the parameters' standard errors using the standard GMM formula based on the delta method—*i.e.*, using the moments' bootstrapped covariance matrix and the sensitivity of each moment to each parameter. Panel A shows the estimates for the full sample, while Panels B and C show results for the 2001 to 2005 and 2006 to 2010 periods. Panels D, E, and F present results for firms in the three size groups.

[Insert Table 4 here.]

Focusing on Panel A, one sees the estimated expected benefits of rebalancing net of entry costs ($h$ values) range within an order of magnitude of $1 per $1 million in market capitalization. The estimated net benefit of entry is $1,000 for a typical $1B firm, which is in the range of

[5] If we use smaller $\sigma_y$ values, we obtain results indicating that liquidity trading is more important than in our base case with $\sigma_y$ = 0.2 bps.
[6] Technically, the model is over-identified if we also restrict each of the parameters to be non-negative. These overidentifying restrictions are satisfied even when we do not impose this restriction on the parameters.

plausible values for certainty equivalents of hedging needs and opportunity costs of time for wealthy investors. Comfortingly, this estimated $1,000 benefit from entry is considerably smaller than typical retail order sizes of $10,000 and is far smaller than common institutional order sizes. In addition, the estimated $h$ value is the expected benefit of the first trader who would choose to enter, whereas the benefit for the $j^{th}$ trader is only $h/j$, which is orders of magnitude lower than $h$ if $j$ is high. As discussed below, our parameter estimates imply that far more liquidity traders enter during the regular market period and in periods when news occurs.

The $h$ parameter is dramatically higher during the regular market period, as compared to the pre- or after-market. This suggests either rebalancing is less necessary or market participation is more costly (lower $h$) during extended market hours—both of which are plausible. The values of investors' other asset holdings probably change most during normal business hours. Monitoring portfolio exposures and attending to the market is presumably more costly when the regular market is closed.

Both data accessibility and rebalancing needs can also help explain why $h$ is higher in periods with news. By providing information about the stock and bringing attention to it, news lowers investors' costs of monitoring and evaluating the stock's suitability in their portfolios. If investors have not rebalanced their positions in a long time, their desired holdings may have changed since their last portfolio evaluation. In addition, news about a stock can affect its risk. Changes in a stock's idiosyncratic risk can affect investors' desired holdings if they hold non-market weights on the stock. Changes in systematic risk could affect investors' desired holdings if their other asset holdings are exposed to similar risks.

In contrast to the dramatic differences across intraday periods, the $h$ parameters are remarkably stable over time, as shown in Panels B and C. Mechanically, this is somewhat surprising because higher $h$ values (all else equal) are associated with more trading activity; and such activity has increased significantly over time. However, there is no economic reason to expect that traders' entry costs have increased, which is consistent with the fact that *h* estimates in Panel C are similar or lower than those in Panel B. Panels D, E, and F show that the *h* estimates across the size groups accord with intuition, too. For traders with large stakes, rebalancing needs measured as a fraction of firm value are likely to be greater in small firms. In addition, the difference between *h* in news and non-news periods is larger in small firms, which could happen because relevant data about small firms is less widely available.

The second key model parameter is the cost of acquiring private information (*c*), which is estimated to be approximately $2.13 per $1 million in market capitalization or 0.0213 bps of firm value. This value is plausible in light of the French (2008) estimate of the cost of active management as a percent of firm value, which is stable at roughly 67 bps per year or 0.27 bps per trading day. Comparing our cost of acquiring information about a firm in one intraday period to the French (2008) daily estimate for an entire portfolio, one would infer that the typical active portfolio manager acquires information about 13 firms (0.27 / 0.0213) in at least one intraday period, which is in the realm of plausibility.

A comparison of Panels B and C in Table 4 reveals that the cost of acquiring and trading on private information (*c*) fell by a factor of 12 (from $4.0*10^{-6}$ to $3.3*10^{-7}$) across the two 5-year samples. In the model, lowering *c* induces informed traders to enter and increases volume in all periods, though it affects regular market volume most. The reason is that the number of informed traders is highest in the regular market and such traders do not internalize their impacts of entering the market on existing traders' profits. More subtly, because entry by informed traders affects variance more when there are fewer informed traders, lowering *c* increases regular market variance relatively less than it increases variance in other periods. In the past decade of data, we observe a relative increase in regular market volume and a relative decrease in regular market variance, which corresponds to a reduction of $c$ in the model. The observed overall decrease in variance across all periods in the past decade corresponds to an even larger reduction in $c$.

Panels D, E, and F show that the cost of acquiring information as a percentage of firm value is over 5 times lower for large firms ($9.1*10^{-7}$) than it is for small firms ($4.9*10^{-6}$). This makes sense if there is some fixed cost of acquiring information. In unreported results, we find that the cost of acquiring information declines even more sharply for large firms in the 2006 to 2010 period (from $2.2*10^{-6}$ to $7.4*10^{-8}$, a factor of 30). As a basis for comparison, a recent *Wall Street Journal* story provides a direct estimate of the cost of informed trading in large firms.[7] Several hedge funds paid up to $10,000 each to acquire private information about a December 8th, 2009 health care law that affected four large health care stocks. As in the model, the information was acquired during the regular market, one intraday period in advance of its release during the after-market period. Based on the cumulative market capitalization of the stocks (about $100B), the estimated model cost of $7.4*10^{-8}$ would imply that each hedge fund would need to pay $7,440, which is similar to the reported cost of the meeting.

Lastly, the model estimates the amounts of tractable ($\sigma_d$) and intractable information ($\sigma_{\varepsilon 1}$) to be of the same order of magnitude in the pre-, regular, and after-markets. It is somewhat surprising that similar price discovery occurs in these three intraday periods even though just 1% of trading activity takes place during extended market hours. We analyze this phenomenon further in Section 5 when we decompose return variance and volume. There we also discuss the general implications of all of our structural estimates for the nature of trading and the role of private and public information in price discovery.

[7] *Wall Street Journal*, December 20, 2011, "Inside Capital, Investor Access Yields Rich Tips" by Brody Mullins and Susan Pulliam.

*C. Alternative Parameterization of the Model*

This subsection demonstrates that the model's key parameters are not sensitive to two of our assumptions. First, we consider the impact of our preferred parameterization in which news affects trader attention ($h$) rather than the amount of tractable information ($\sigma_d$) that can be acquired. Here we estimate an alternative model in which news does not affect trader attention ($h$) but it does affect the extent of learnable private information ($\sigma_d$). In this alternative model, only the subscripts on $\sigma_d$ and $h$ in the predicted moment equations (9) and (11) change. As before, we estimate the model using efficient GMM and obtain an exact fit to the 14 empirical moments. Panel A in Table 5 presents the results of this estimation for the full sample.

[Insert Table 5 here.]

The main result in Table 5A is that the parameter estimates look very similar to those in Table 4A, implying both versions of the model make similar equilibrium predictions. Specifically, the estimates of $h$(*no news*) in Table 4A are nearly identical to the estimates of $h$ in Table 5A; and the estimates of $\sigma_d$ in Table 4A are nearly identical to the estimates of $\sigma_d$(*no news*) in Table 5A. The parameter estimates in Table 5A indicate that news reduces information asymmetry between traders by lowering tractable private information ($\sigma_d$(*news*) < $\sigma_d$(*no news*)).[8]

The economic explanations for trading around news are slightly different in the models in Tables 4A and 5A. In the model where $h$ varies with news, discretionary liquidity traders choose to enter because either their direct costs of entry decrease or their rebalancing needs increase ($h$(*news*) > $h$(*no news*)). In the model where $\sigma_d$ varies with news, discretionary liquidity traders choose to enter because news reduces information asymmetry ($\sigma_d$(*news*) < $\sigma_d$(*no news*)) and lowers their trading costs. However, in both models, the ultimate effect is that discretionary

[8] Tetlock (2010) studies return predictability around news and argues that news reduces information asymmetry.

liquidity traders enter in massive quantities around news, so most of the testable predictions of the two models are indistinguishable.[9]

Next, we consider the impact of our assumption that liquidity trading is independent across traders. We do this by increasing the assumed size of each liquidity trader ($\sigma_y$). This effectively increases the correlation among liquidity trades because each trader's demand is perfectly correlated with itself. We can increase the correlation in liquidity trading only in the regular market period because it is already near one in the extended hours periods when few traders choose to enter—*e.g.*, in the pre-market period without news, only one trader enters. Panel B in Table 5 presents estimates from the model in which the regular market trade size is now $\sigma_y$ = 2 bps, while it remains at $\sigma_y$ = 0.2 bps during extended hours.

Table 5B shows that this tenfold increase in liquidity trader size almost doubles implied liquidity needs ($h$), while reducing implied information costs ($c$) and the amount of tractable information ($\sigma_d$) by factors of 4 and 2. The main impact is that liquidity (informed) trading becomes a smaller (larger) component of overall trading. However, informed trading contributes less to price discovery because of the decline in tractable information. Despite these nontrivial quantitative changes, none of the qualitative statements in the forthcoming volume and variance decomposition would be affected by this increase in correlation. Moreover, a single liquidity trade of 2 bps of firm value is quite large, exceeding the typical trade size of an institutional portfolio transition, which Obhizhaeva (2009) estimates to be 1.5 bps. Such transitions represent unusually large liquidity-motivated trades and thus provide a reasonable upper bound. In fact, Obhizhaeva (2009) documents that only 19% of such trades is executed on the first day, suggesting that 0.19 * 1.5 bps = 0.29 bps is a more realistic trade size for a single day.

[9] In the model in Table 4, there is also some entry by informed traders around news, but we will see in Section 5 that this effect is small compared to the influx of non-informational traders around news.

## 5. Implications of the Model for Price Discovery and Trading Activity

*A. Theoretical Volume and Variance Decompositions*

The model in Section 2 allows one to analyze the relative contributions of informed and liquidity traders to market activity. To decompose trading volume, we separately consider the quantities of buy and sell orders that transact between traders and the net order flow ($Q$) in which market makers take the other side of the trade. We assign half the volume arising from a trade to each counterparty participating in the trade. Thus, the volume attributable to market makers ($v_{Mt}$) is half of the expected net order flow aggregated across both trader groups, which is:

$$E(v_{Mt}) = \frac{1}{2}E(|Q|) = \frac{\sqrt{ch_t}(h_t/c+1)^{3/2}}{2\sqrt{2\pi}\sigma_{dt+1}} \tag{21}$$

The trading volume attributable to each group of traders is then half of the sum of their buy orders and sell orders plus half of their proportion of trading with market makers. The proportion of variance in net order flow ($Q$) arising from informed traders is $h_t(h_t+c)^{-1}$. Expected trading volume arising from informed traders ($v_I$) is given by:

$$E(v_{It}) = \frac{h_t(h_t/c+1)}{\sqrt{2\pi}\sigma_{dt+1}}\left[1+\frac{1}{2}\frac{h_t}{h_t+c}\sqrt{1+c/h_t}\right], \tag{22}$$

where the first term reflects trades with liquidity traders and the second reflects trades with market makers. Lastly, the expected volume from liquidity traders ($v_{Lt}$) is:

$$E(v_{Lt}) = \frac{h_t(h_t/c+1)}{\sqrt{2\pi}\sigma_{dt+1}}\left[(h_t+c)\sigma_{dt+1}^{-1}\sigma_y^{-1}+\frac{1}{2}(h_t/c+1)^{-1}\sqrt{1+c/h_t}\right]. \tag{23}$$

The first term comes from trading among liquidity traders as well as between liquidity and informed traders, while the second term accounts for trading with the market maker. Because these three components equal total expected volume, we can express each as a fraction of the total using:

$$\frac{E(v_{It})}{E(v_t)} + \frac{E(v_{Lt})}{E(v_t)} + \frac{E(v_{Mt})}{E(v_t)} = 1. \tag{24}$$

This equation is the basis for the volume decompositions that we report.

We can also decompose return variance in Equation (9) into three components. We report variance decompositions in which each component is expressed as a fraction of the total variance as described below:

$$\frac{\sigma_{\varepsilon t}^2}{Var(r_t)} + \frac{h_t(h_t + c)^{-1}\sigma_{dt+1}^2}{Var(r_t)} + \frac{(h_{t-1}/c + 1)^{-1}\sigma_{dt}^2}{Var(r_t)} = 1. \tag{25}$$

The first term reflects intractable information, which only arrives in periods with public news. The second term represents price discovery arising from trading on private information. The third term measures the revelation of tractable information that the previous period's price did not fully reveal. This term comprises public information revealed by sources other than Dow Jones news, including softer information sources, such as social media, television, radio, and word of mouth.

*B. Estimated Volume Decomposition*

In Figures 3A to 3C, we report the decomposition of expected volume from Equation (24) for the regular market (3A), pre-market (3B), and after-market (3C) periods. Table 6 shows more detailed volume decompositions. The four panels represent the intraday periods, while the six rows in each panel indicate the sample used: full sample, two 5-year subperiods, and the three firm size groups. To construct the tables and figures, we substitute our parameter estimates for news and non-news periods from Table 4 into Equations (21) to (23), weighting the periods by the probability of news occurring in the past 24 hours.[10]

[Insert Figure 3 here.]

[10] Recall that the *h* parameter depends on whether news occurs in the past 24 hours.

[Insert Table 6 here.]

The most striking fact in the figures is that discretionary liquidity trading accounts for the vast majority of volume in each of the three intraday periods, especially in the regular market where it is 92% of volume. One can infer the importance of discretionary liquidity trading from the fact that regular market volatility is comparable to volatility in the other periods, whereas regular market volume is nearly 100 times higher. This implies that liquidity is high in the regular market, which motivates many liquidity traders seeking to minimize their price impact to enter the market. For the same reason, many informed traders enter the market and trade aggressively on their information. However, such intense trading reveals their information almost perfectly, which lowers informed trading profits and deters further entry. Because there is no such counterbalancing force stopping liquidity traders from entering, discretionary liquidity trading constitutes the lion's share of volume.

Subperiod analysis reveals the implications of the decline in the cost of acquiring information documented in Table 4. In the regular market, informed trading accounts for 3.6% of trading volume from 2001 to 2005 and 11.4% from 2006 to 2010. Although the percentage of informed trading volume appears low even in the more recent period, one must remember that total volume is $32 trillion in 2010. This implies that informed volume exceeds $3.5 trillion. While the percentage of informed volume is slightly higher at 25% in the pre- and after-market periods (from 2006 to 2010), the dollar amount of informed trading in these periods pales in comparison to the tens of trillions traded in the regular market.

*C. Estimated Variance Decomposition*

Figures 4A to 4D show the decomposition of return variance in Equation (25) for each of the four intraday periods, again based on the model parameters shown in Table 4. The last three columns in Table 6 report the variance decomposition results. The dramatic difference between the variance and volume decomposition results is that a large fraction of return variance comes from trading on private information, particularly in the second half of the sample. In the 2006 to 2010 period, 77% of return variance during the regular market comes from informed trading. The sharp decrease in information acquisition costs causes this increase in price discovery coming from informed trading.

[Insert Figure 4 here.]

A second notable finding is that the fraction of variance (*e.g.*, in the regular market) attributable to public news (7%) is far lower than the fraction of variance coming from the delayed public release of undiscovered tractable information (60%), shown in the second column of Table 6 labeled "other public." In the regular market, the difference between these two components of variance is largest (6% versus 86%) from 2001 to 2005, when high information acquisition costs deterred the collection of tractable information. In 2006 to 2010, when information acquisition costs fell by an order of magnitude, the difference narrowed to 7% versus 16%. Only in the after-market period is the fraction of variance coming from measurable (DJ) public news similar to the "other public" fraction. This happens both because after-market news is particularly important and because informed traders choose to collect most tractable information that is available during the (prior) regular market period.

## 6. Testing Predictions from Competing Models

### *A. Discretionary Liquidity Trader Model Predictions*

We now use the main model presented in Table 4A for predictive analyses. Given a set of parameter estimates, the model makes testable predictions about return variance and trading volume in the $a$ = 1, 2, 3 periods after news occurs in period $t$:

$$Var(r_{1t+a}) = \sigma^2_{\varepsilon t+a} + h_{1t+a}(h_{1t+a}+c)^{-1}\sigma^2_{dt+a+1} + (h_{1t+a-1}/c+1)^{-1}\sigma^2_{dt+a} \tag{26}$$

$$E(v_{1t+a}) = \frac{h_{1t+a}(h_{1t+a}/c+1)}{\sqrt{2\pi}\sigma_{dt+a+1}}\left[1+(h_{1t+a}+c)\sigma^{-1}_{dt+a+1}\sigma^{-1}_{y}+\sqrt{1+c/h_{1t+a}}\right]. \tag{27}$$

The first term in Equation (26) reflects a simplifying assumption that the variance of intractable information is independent of whether first news occurred in the preceding period.[11] Figures 5A, 5B, 5C, and 5D report the predicted hourly return volatility and trading volume from Equations (26) and (27) in event time after news occurs. The lines in the figures represent return volatility and the bars represent volume. The period in which news occurs ($t$) varies across the figures and the number of periods after news ($a$) varies along the $x$-axis within each figure. Each figure shows the actual return volatility and trading volume observed in the data for comparison purposes. Predicted and actual values are all reported in excess of unconditional expectations.

[Insert Figure 5 here.]

Although the model's predictions are simplistic, they can explain two key stylized facts: 1) prices respond to news primarily in the period when news arrives; and 2) volume responds mainly during the regular period, regardless of when news arrives. The most stark and surprising demonstrations of these facts appear in Figures 5B and 5D, which show market activity after news arrives in the pre-market and after-market periods. In both figures, return volatility is by far the highest in the period when news arrives (event period 0), whereas turnover peaks during the regular market period (event period 1 for pre-market news; and period 3 for after-market news).

Although it makes some quantitative errors, the model correctly predicts these qualitative patterns in the figures. Variance is highest when news arrives because news releases intractable information, which is immediately incorporated in prices, and most periods following news do not release additional intractable information. Volume is highest during the regular market

[11] The variance of intractable information following a non-news period comes solely from first news, so it is the probability of first news arriving multiplied by the variance of intractable information in first news periods—*i.e.*, the first term in Equation (26).

period even though news immediately lowers entry costs for discretionary liquidity trades. Still, because their entry costs remain lowest during the regular market period, more discretionary liquidity traders choose to enter at this time. Quantitatively, the model does not predict sufficiently large trading volume in the regular market following after-market news arrival in Figure 5D. One could reconcile this with the model by allowing after-market news to exert an especially large impact on *h*, which is reasonable because after-market news releases the most intractable information—*i.e.*, the after-market $\sigma_{\varepsilon 1}$ values are the highest in Table 4. The other notable model error is that actual regular market volatility is higher than predicted following pre-market news arrival (in Figure 5B). This error could arise from the simplifying assumption that the variance of intractable information is unaffected by the presence of recent news, which is violated if regular market news is especially informative following pre-market news.

The empirical patterns observed for regular market and overnight news in Figures 5A and 5C are quite similar to those in 5B and 5D, though two differences arise. First, after regular market news in Figure 5A, return volatility is unusually high in the following pre-market period (event period 3). Surprisingly, the model correctly predicts this delayed increase in variance because expected rebalancing needs net of entry costs are especially high in the pre-market period after news occurs, which induces more informed traders to acquire and trade on private information. A simple economic story is that liquidity traders may read the news in the morning, which would lower their cost of trading in the pre-market period. Second, after overnight news in Figure 5C, pre-market variance is actually slightly higher than overnight variance. The same idea—that discretionary liquidity traders may read the news in the morning—could explain this fact, too, which is why the model is again able to qualitatively match the empirical data.

Lastly, we briefly consider the model's predictions for market liquidity, though we do not quantitatively test them because liquidity in the model's one-shot call auction structure is unlikely to be directly comparable to liquidity observed in continuous double auctions. Previous empirical research shows that market liquidity is higher in the regular market period (Barclay and Hendershott, 2004); b) after news occurs (Tetlock, 2010); and c) in large stocks (Hasbrouck, 2009). The model predicts each of these three features in liquidity data mainly because of variation in the model parameters $h$ and $c$, neither of which is estimated using liquidity data. Higher rebalancing needs ($h$) clearly increase liquidity by reducing adverse selection. Lower information acquisition costs ($c$) also promote liquidity by enticing more informed traders to enter the market, which increases competition and leads to aggressive trading, thereby revealing more private information and improving liquidity.

*B. Differences in Opinion Model Predictions*

Here we contrast predictions from our model of liquidity trading with those from belief-based models of trading, such as Kim and Verrecchia (1991), Harris and Raviv (1993), Kandel and Pearson (1995), Hong and Stein (2003), and Scheinkman and Xiong (2003). As discussed in Section 2, these theories predict that changes in traders' relative beliefs cause trading, as those with relatively optimistic interpretations of news buy from those with pessimistic interpretations. To test this idea, we analyze market activity around news events sorted on the basis of changes in analysts' beliefs about quarterly earnings.

To measure changes in relative beliefs about news, we consider only news events in which at least two analysts updated an earnings forecast in the four weeks prior to the news and at least one of this same set of analysts updates within one week after the news as recorded in the I/B/E/S database. We compute relative belief changes using the difference between analyst forecast dispersion before and after the news scaled by the firm's stock price four weeks before the news. We use earnings forecasts and stock prices that are adjusted to account for splits. We group news events by terciles ranked by whether analyst beliefs converge (the bottom tercile of relative belief changes), remain similar (middle tercile), or diverge (top tercile).

For news in each tercile of belief changes, we measure hourly return volatility and turnover in the period when news arrives and the following three periods. Figures 6A, 6B, 6C, and 6D report these measures for news events arriving in each of the four intraday periods. The three bars in each figure depict turnover occurring after each type of news event—convergence, no change, or divergence in analysts' relative beliefs—while the three lines represent volatility after each type of news.

[Insert Figure 6 here.]

The patterns in the four figures do not support belief-based models of trading activity. For example, in Figure 6B showing responses to pre-market news, turnover is very similar in each event period regardless of whether analysts' relative beliefs changed significantly around news. The same observation applies to Figures 6A, 6C, and 6D. This finding casts doubt on the idea that investor disagreement is a major determinant of trading activity after news. An alternative interpretation is that analysts' beliefs are such a poor proxy for investors' beliefs that their relationship with trading activity cannot be observed. However, this latter interpretation would call into question a voluminous empirical literature that uses analyst forecast dispersion as a proxy for investor disagreement—*e.g.*, Diether, Malloy, and Scherbina (2002) and Chordia, Huh, and Subrahmanyam (2007) among many others.[12]

Like the previous figures, Figures 6A through 6D show that return volatility after news typically occurs immediately in event period 0, whereas most trading volume that follows news events occurs in the regular market. This delayed volume response is difficult to reconcile with

[12] A recent paper by Giannini and Irvine (2012) advocates using a novel difference of opinion measure based on disagreement between the tone of media coverage and posts on stocktwits.com.

differences in opinion models. The timing of volatility implies that market prices aggregate most traders' beliefs in event period 0, which suggests that most belief-based trading occurred in event period 0, too. But most trading following news occurs in the regular market, which is not typically event period 0, implying that most trading is not driven by beliefs. In contrast, the timing of volatility and volume can be reconciled with models of discretionary liquidity traders. If these traders expect their trades to have less impact on prices when news has less impact on beliefs, they would choose to enter the market and trade after such news events.

## 7. Concluding Discussion

We estimate a structural model of strategic trader behavior to match the rich relationships between volume, volatility, and news arrival in the electronic trading era. For the model to fully explain the magnitude of the volume and volatility patterns across the intraday periods, discretionary liquidity trading must constitute the vast majority of overall trading volume—*e.g.*, 92% in the regular market. Although the model is a simplification of reality, it suggests that policymakers should carefully consider the welfare of such uninformed traders when evaluating alternative market structures and regulations. In the model, because there is perfect competition among market makers and among informed traders, welfare is solely determined by the surplus received by liquidity traders. The inframarginal traders with the highest rebalancing needs net of entry costs receive aggregate surplus given by:

$$\sum_{j=1}^{k_t} h_t / j - k_t (h_t / k_t) = h_t (\sum_{j=1}^{k_t} 1/ j - 1) \approx h_t \ln(2k_t / 3) \text{ for } k_t >> 1. \tag{28}$$

This surplus increases with the number of liquidity traders who enter the market, suggesting broad participation in trading is a reasonable proxy for welfare. Because the number of liquidity traders is proportional to market liquidity in equilibrium, improvements in liquidity also signify improvements in welfare. Thus, our model provides a theoretical basis for the participation and liquidity objectives that the SEC often cites among its central goals.

Furthermore, our parameter estimates imply the cost of acquiring and acting on information fell sharply in the past decade, causing increases in participation and liquidity and thus trader welfare. This positive account of the impact of advances in trading technology on trader welfare contrasts with more negative populist arguments. Such arguments typically ignore the possibilities that uninformed agents can choose whether to trade and that informed traders aggressively compete against each other. Because it includes these two key features, our model predicts that technological advances can improve liquidity and price discovery, both of which can be socially beneficial.

We show that this model correctly predicts that stock prices respond immediately to news, while trading volume typically responds with a delay. News not only releases intractable information, which is priced immediately, but it also triggers trader attention for an extended period of time. During this window, uninformed traders choose to enter the market when it is cheapest to trade—usually in regular trading hours—regardless of when news arrives.

This emphasis of the role of discretionary liquidity traders and trading frictions contrasts with recent models that highlight the importance of changes in traders' relative beliefs. Although many patterns in volume, volatility, and news can be explained without resorting to trader disagreement, a model that includes both trading frictions *and* disagreement is likely to produce an even richer set of predictions. The welfare implications of a model in which traders' sometimes act on irrational beliefs are also likely to differ from those discussed above. Future researchers could estimate and test such a model using data on individual traders, which could allow for separate estimates of disagreement, attention constraints, and rebalancing needs.

# Appendix

Solving the Model

Here we solve for the equilibrium strategies and endogenous outcomes in the model introduced in Section 2. For an informed trader *i* who acquires a signal $d_{t+1}$ and chooses to trade an amount $x_{it}$, her expected profits ($\pi_{it}$) are given by:

$$E[\pi_{it} \mid d_{t+1}] = E[x_{it}(F_{t+1} - p_t) \mid d_{t+1}] = x_{it}[1-(m_t - 1)\lambda_t \beta_t] d_{t+1} - \lambda_t x_{it}^2. \tag{29}$$

Maximizing this quadratic equation in $x_{it}$ gives:

$$x_{it} = \frac{1}{2}[1/\lambda_t - (m_t - 1)\beta_t] d_{t+1}. \tag{30}$$

To identify a symmetric equilibrium, we equate the $\beta_t$ coefficient in the optimal $x_{it}$ strategy with the conjectured linear strategy above:

$$\beta_t = (1/\lambda_t)(m_t + 1)^{-1}. \tag{31}$$

The equilibrium market depth can be found from the zero profit condition for the market maker:

$$\lambda_t = \frac{Cov(Q_t, d_{t+1})}{Var(Q_t)} = \frac{\sigma_{dt+1}}{(m_t + 1)\sigma_{yt}} \sqrt{\frac{m_t}{k_t}}. \tag{32}$$

The second-order condition for informed traders' maximization problem is $\lambda > 0$, which is always satisfied when there are some liquidity traders—*i.e.*, $k_t \sigma_{yt} > 0$.

Based on their benefits from rebalancing and information acquisition costs, the discretionary liquidity traders and informed traders simultaneously choose whether to enter the market, which endogenously determines $m_t$ and $k_t$ in equilibrium. After substitutions and simplification, the expected trading profit of each informed trader is:

$$E[\pi_{it} \mid d_{t+1}] = \frac{\sigma_{d+1}\sigma_{yt}}{(m+1)} \sqrt{\frac{k_t}{m_t}}. \tag{33}$$

Entry of informed traders occurs until the marginal trader attains zero profits, which occurs when $E[\pi_{it}|d_{t+1}] = c$ or:

$$c^2 m_t (m_t + 1)^2 = k_t \sigma_{yt}^2 \sigma_{dt+1}^2. \tag{34}$$

This implies that equilibrium illiquidity is:

$$\lambda_t = \frac{\sigma_{dt+1}^2}{c(m+1)^2}. \tag{35}$$

The (negative) expected trading profit ($\pi_{jt}$) of each discretionary liquidity traders is:

$$E[\pi_{jt}] = E[y_{jt}(F_{t+1} - p_t)] = -\lambda_t E[y_{jt}^2] = -\frac{\sigma_{dt+1}^2 \sigma_{yt}^2}{c(m_t+1)^2}, \tag{36}$$

where the second equality uses the equilibrium entry condition for informed traders.

Entry of discretionary liquidity traders occurs until the marginal trader's expected trading loss is equal to his expected benefit from rebalancing ($h_t/k_t$), when the equilibrium number of liquidity traders will be

$$k_t = ch_t(m_t+1)^2 \sigma_{dt+1}^{-2} \sigma_{yt}^{-2}. \tag{37}$$

Substituting this condition into the informed trader's zero profit condition, we obtain the equilibrium number of informed traders:

$$m_t = h_t / c. \tag{38}$$

Substituting this value of $m_t$ into the $k_t$ equation gives the equilibrium solution for $k_t$ reported in the text, which then allows one to compute the reported equilibrium values of $\lambda_t$ and $\beta_t$.

<u>Measuring Returns Using TAQ Data</u>

During regular trading, we only keep trades and quotes meeting standard filters used in the microstructure literature. We drop trades with non-positive price or size and those with correction codes not equal to zero or condition code of M, Q, T, or U. For the pre-market and after-market periods, however, the filters for trades necessarily differ. Most importantly, we do not exclude trades with a condition code of T, which explicitly identifies extended hours trades. For extended hours periods, we exclude those that occur at prices probably determined within the trading day (*e.g.*, crosses and block trades), appear out of sequence, or contain non-standard delivery options. This filter eliminates any trades from NYSE, AMEX, or CBOE and trades with "cond" codes B, G, K, M, L, N, O, P, W, U, Z, 4, 5, 6, 8, or 9. We drop trades of at least 10,000 shares or $200,000 regardless of their "cond" codes as these are likely pre-negotiated blocks. Finally, we drop all trades and quotes in the final minute of the pre-market and in the first minute of the after-market period to mitigate effects of bid-ask bounce.

Within each of the pre-market, regular, and after-market periods, we construct a beginning and ending trade price as the volume-weighted average price (VWAP) based on the first and last minute of trades in the dataset and then compute trade-based returns. Using a VWAP instead of a single trade price further mitigates the effects of bid-ask bounce in returns. The overnight period return is the percent change from the last after-market price to the first pre-market price on the subsequent trading day, accounting for dividends and stock splits when the subsequent trading day is the ex date. When there is only one trade observation in an intraday period, its return is computed from the last price from the most recent intraday period. When there is no trading in a period, the return is zero. CRSP computes a stock return on trading day $t$ based on the last regular hours transaction (or quote, in the event of non-trading) prices on day $t$ and $t - 1$. Ignoring our 1-minute buffers at the end of the pre-market and the beginning of the after-market periods, day $t$ CRSP return is approximated by compounding our after-market return from day $t - 1$ and our overnight, pre-market, and regular period returns on day $t$.

<u>Bibliography</u>


Admati, Anat R., and Paul Pfleiderer, 1988, A theory of intraday patterns: Volume and price variability, *Review of Financial Studies* 1, 3-40.

Bamber, Linda Smith, Orie E. Barron, and Douglas E. Stevens, 2011, Trading volume around earnings announcements and other financial reports: Theory, research design, empirical evidence, and directions for future research, *Contemporary Accounting Research* 28, 431-471.

Banerjee, Snehal, and Ilan Kremer, 2010, Disagreement and learning: Dynamic patterns of trade, *Journal of Finance* 65, 1269-1302.

Barclay, Michael J., and Terrence Hendershott, 2003, Price discovery and trading after hours, *Review of Financial Studies* 16, 1041-1073.

Barclay, Michael J., and Terrence Hendershott, 2004, Liquidity externalities and adverse selection: Evidence from trading after hours, *Journal of Finance* 59, 681-710.

Barclay, Michael J., and Terrence Hendershott, 2008, A comparison of trading and non-trading mechanisms for price discovery, *Journal of Empirical Finance* 15, 839-849.

Barclay, Michael J., Robert H. Litzenberger, and Jerold B. Warner, 1990, Private information, trading volume, and stock-return variances, *Review of Financial Studies* 3, 233-253.

Beaver, William H., 1968, The information content of annual earnings announcements, *Journal of Accounting Research* 6, 67-92.

Berry, Thomas D., and Keith M. Howe, 1994, Public information arrival, *Journal of Finance* 49, 1331-1346.

Brandt, Michael W., Alon Brav, John R. Graham, and Alok Kumar, 2010. The idiosyncratic volatility puzzle: Time trend or speculative episodes? *Review of Financial Studies* 23, 863-899.

Chae, Joon, 2005, Trading volume, information asymmetry, and timing information, *Journal of Finance* 60, 413-442.

Chordia, Tarun, Sahn-Wook Huh, and Avanidhar Subrahmanyam, 2007, The cross-section of expected trading activity, *Review of Financial Studies* 20: 709-740.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2011, Recent trends in trading activity and market quality, *Journal of Financial Economics* 101, 243-263.

Cutler, David M., James M. Poterba, and Lawrence H. Summers, 1989, What moves stock prices? *Journal of Portfolio Management* 15, 4-12.

DeBondt, Werner F.M., and Richard H. Thaler, 1995, Financial decision-making in markets and firms: A behavioral perspective, In *Handbooks in Operations Research and Management Science: Finance*. Edited by Robert A. Jarrow, Vojislav Maksimovic, and William T. Ziemba. Amsterdam: Elsevier: 385-410.

Diether, Karl B, Christopher J. Malloy, and Anna Scherbina, 2002, Differences of opinion and the cross section of stock returns, *Journal of Finance* 57, 2113-2141.

Fang, Lily, and Joel Peress, 2009, Media coverage and the cross-section of stock returns, *Journal of Finance* 64, 2023-2052.

French, Kenneth R., 2008, The cost of active investing, *Journal of Finance* 58, 1537-1573.

French, Kenneth R., and Richard Roll, 1986, Stock return variances: The arrival of information and the reaction of traders, *Journal of Financial Economics* 17, 5-26.

Giannini, Robert, and Paul J. Irvine, 2012, The impact of divergence of opinions about earnings within a social network, working paper, University of Georgia.

Harris, Milton, and Artur Raviv, 1993, Differences of opinion make a horse race, *Review of Financial Studies* 6, 473-506.

Hong, Harrison, and Jeremy C. Stein, 2003, Differences of opinion, short-sales constraints, and market crashes, *Review of Financial Studies* 16, 487-525.

Ito, Takatoshi, Richard K. Lyons, and Michael T. Melvin, 1998, Is there private information in the FX market? The Tokyo experiment, *Journal of Finance* 53, 1111-1130.

Jain, Prem C., and Gun-Ho Joh, 1988, The dependence between hourly prices and trading volume, *Journal of Financial and Quantitative Analysis* 23, 269-283.

Jiang, Christine X., Tanakorn Likitapiwat, and Thomas McInish, 2011, Information content of earnings announcements: Evidence from after hours trading, *Journal of Financial and Quantitative Analysis*, forthcoming.

Jones, Charles M., Gautam Kaul, and Marc L. Lipson, 1994, Information, trading, and volatility, *Journal of Financial Economics* 36, 127-154.

Kandel, Eugene, and Neil D. Pearson, 1995, Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy* 103, 831-872.

Kim, Oliver, and Robert E. Verrecchia, 1991, Trading volume and price reactions to public announcements, *Journal of Accounting Research* 29, 302-321.

Kim, Oliver, and Robert E. Verrecchia, 1994, Market liquidity and volume around earnings announcements, *Journal of Accounting and Economics* 17, 41-67.

Kim, Oliver, and Robert E. Verrecchia, 1997, Pre-announcement and event-period private information, *Journal of Accounting and Economics* 24, 395-419.

Kross, William, Gook-Lak Ha, and Frank Heflin, 1994, A test of risk clientele effects via an examination of trading volume response to earnings announcements, *Journal of Accounting and Economics* 18, 67-87.

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

Mendelson, Haim, and Tunay I., Tunca, 2004, Strategic trading, liquidity, and information acquisition, *Review of Financial Studies* 17, 295-337.

Michaely, Roni, Amir Rubin, and Alexander Vedrashko, 2011, Corporate governance and the timing of earnings announcements, Cornell University working paper.

Milgrom, Paul, and Nancy Stokey, 1982, Information, trade, and common knowledge, *Journal of Economic Theory* 26, 17-27.

Mitchell, Mark L., and J. Harold Mulherin, 1994, The impact of public information on the stock market, *Journal of Finance* 49, 923-950.

Roll, Richard, 1988, $R^2$, *Journal of Finance* 43, 541-566.

Scheinkman, José A., and Wei Xiong, 2003, Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183-1219.

Tetlock, Paul C., 2010, Does public financial news resolve asymmetric information? *Review of Financial Studies* 23, 3520-3557.

Zdorovtsov, Vladimir, 2004, Firm-specific news, extended-hours trading and variances over trading and nontrading periods, State Street Global Advisors working paper.

**Table 1: Probability of News Arrival**

This table presents probabilities of news arrival during 2001-2010 for each of four intraday periods: the Regular Market (9:30 AM to 4:00 PM), the Pre-Market (7:00 AM to 9:30 AM), Overnight (6:30 PM to the following 7:00 AM), and the After-Market (4:00 PM to 6:30 PM). The variable *news* is 1 if there is at least one story (two stories) in the Dow Jones Newswires mentioning a particular Small or Mid Cap (Large Cap) firm and 0 otherwise. The variable *FirstNews* is 1 if *news*=1 for the current period and *news*=0 for each of the past three intraday periods, and 0 otherwise. All probabilities are calculated by pooling observations for all firms within each year and size group and then averaging across groups. Panels A-D provide results for All, Large Cap, Mid Cap, and Small Cap firms with market capitalizations in the intervals [\$100M, ∞), [\$10B,¸∞), [\$1B,\$10B), and [\$100M,\$1B), respectively.

| Panel A: All Firms | Regular | Pre-Market | Overnight | After-Market |
|---|---|---|---|---|
| P(*news*=1) | 0.1476 | 0.1013 | 0.1189 | 0.0750 |
| P(*FirstNews*=1) | 0.1015 | 0.0743 | 0.0850 | 0.0532 |
| Firms Per Year/Size Group | 192 | | | |

| Panel B: Large Caps | | | | |
|---|---|---|---|---|
| P(*news*=1) | 0.2565 | 0.1295 | 0.1612 | 0.1008 |
| P(*FirstNews*=1) | 0.1632 | 0.0733 | 0.0855 | 0.0529 |
| Firms Per Year | 95 | | | |

| Panel C: Mid Caps | | | | |
|---|---|---|---|---|
| P(*news*=1) | 0.1225 | 0.1036 | 0.1206 | 0.0747 |
| P(*FirstNews*=1) | 0.0927 | 0.0860 | 0.1021 | 0.0619 |
| Firms Per Year | 245 | | | |

| Panel D: Small Caps | | | | |
|---|---|---|---|---|
| P(*news*=1) | 0.0639 | 0.0707 | 0.0750 | 0.0496 |
| P(*FirstNews*=1) | 0.0485 | 0.0635 | 0.0675 | 0.0449 |
| Firms Per Year | 237 | | | |

**Table 2: Hourly Return Volatility**

This table presents hourly stock return volatilities during 2001-2010 for each of four intraday periods: the Regular Market, the Pre-Market, Overnight, and the After-Market. Variances are calculated by pooling observations for all firms within each year and size group and then averaging across groups. Conditional measures are based only on observations meeting the condition *FirstNews*=1 or *RecentNews*=0. Panels A-D provide results for All, Large Cap, Mid Cap, and Small Cap firms, respectively. Intraday periods, variables, and subsamples are as defined above. Numbers in the table are converted to hourly volatilities by dividing the variance by the number of hours in the intraday period and then taking the square root. Bootstrapped standard errors appear in parentheses.

| Panel A: All Firms | Regular | Pre-Market | Overnight | After-Market |
|---|---|---|---|---|
| Unconditional | 1.275 | 0.698 | 0.344 | 0.811 |
| | (0.019) | (0.013) | (0.007) | (0.016) |
| *FirstNews*=1 | 1.569 | 1.405 | 0.611 | 2.380 |
| | (0.023) | (0.029) | (0.022) | (0.059) |
| *RecentNews*=0 | 1.192 | 0.479 | 0.305 | 0.653 |
| | (0.018) | (0.011) | (0.006) | (0.018) |

| Panel B: Large Caps | | | | |
|---|---|---|---|---|
| Unconditional | 0.954 | 0.625 | 0.271 | 0.599 |
| | (0.025) | (0.025) | (0.012) | (0.028) |
| *FirstNews*=1 | 0.987 | 1.003 | 0.408 | 2.050 |
| | (0.035) | (0.060) | (0.039) | (0.142) |
| *RecentNews*=0 | 0.901 | 0.459 | 0.236 | 0.445 |
| | (0.025) | (0.018) | (0.013) | (0.030) |

| Panel C: Mid Caps | | | | |
|---|---|---|---|---|
| Unconditional | 1.177 | 0.663 | 0.318 | 0.760 |
| | (0.018) | (0.014) | (0.006) | (0.018) |
| *FirstNews*=1 | 1.419 | 1.307 | 0.511 | 2.160 |
| | (0.029) | (0.047) | (0.020) | (0.060) |
| *RecentNews*=0 | 1.082 | 0.450 | 0.281 | 0.593 |
| | (0.018) | (0.015) | (0.006) | (0.020) |

| Panel D: Small Caps | | | | |
|---|---|---|---|---|
| Unconditional | 1.606 | 0.794 | 0.425 | 1.018 |
| | (0.019) | (0.013) | (0.007) | (0.019) |
| *FirstNews*=1 | 2.097 | 1.790 | 0.831 | 2.851 |
| | (0.033) | (0.052) | (0.033) | (0.067) |
| *RecentNews*=0 | 1.511 | 0.526 | 0.380 | 0.854 |
| | (0.019) | (0.012) | (0.007) | (0.020) |

**Table 3: Hourly Turnover**

This table presents hourly turnover during 2001-2010 for each of four intraday periods: the Regular Market, the Pre-Market, Overnight, and the After-Market. Average turnover is calculated by pooling observations for all firms within each year and size group and then averaging across groups. Conditional measures are based only on observations meeting the condition *FirstNews*=1 or *RecentNews*=0. Panels A-D provide results for All, Large Cap, Mid Cap, and Small Cap firms, respectively. Intraday periods, variables, and subsamples are as defined above. Numbers in the table are converted to hourly turnover by dividing the average turnover by the number of hours in the intraday period. Bootstrapped standard errors appear in parentheses.

| Panel A: All Firms | Regular | Pre-Market | After-Market |
|---|---|---|---|
| Unconditional | 25.62 | 0.34 | 0.46 |
| | (0.257) | (0.009) | (0.009) |
| *FirstNews*=1 | 32.50 | 0.94 | 2.27 |
| | (0.430) | (0.044) | (0.085) |
| *RecentNews*=0 | 22.39 | 0.13 | 0.33 |
| | (0.218) | (0.003) | (0.006) |
| **Panel B: Large Caps** | | | |
| Unconditional | 17.60 | 0.17 | 0.24 |
| | (0.181) | (0.004) | (0.008) |
| *FirstNews*=1 | 18.16 | 0.34 | 1.45 |
| | (0.329) | (0.019) | (0.093) |
| *RecentNews*=0 | 16.72 | 0.09 | 0.17 |
| | (0.167) | (0.002) | (0.003) |
| **Panel C: Mid Caps** | | | |
| Unconditional | 29.38 | 0.29 | 0.48 |
| | (0.258) | (0.008) | (0.009) |
| *FirstNews*=1 | 35.10 | 0.74 | 2.25 |
| | (0.384) | (0.037) | (0.106) |
| *RecentNews*=0 | 25.38 | 0.11 | 0.35 |
| | (0.217) | (0.002) | (0.006) |
| **Panel D: Small Caps** | | | |
| Unconditional | 29.88 | 0.55 | 0.66 |
| | (0.437) | (0.024) | (0.017) |
| *FirstNews*=1 | 44.25 | 1.72 | 3.12 |
| | (0.968) | (0.125) | (0.170) |
| *RecentNews*=0 | 25.06 | 0.18 | 0.47 |
| | (0.361) | (0.006) | (0.011) |

**Table 4: Estimates of Model Parameters When *h* Varies with News**

This table presents efficient GMM estimates of exogenous parameters of the model described in Section 2. Where appropriate, the table reports separate estimates of parameters for each of four intraday periods—the Regular Market, the Pre-Market, Overnight, and the After-Market—and conditional on *news*=1 or 0. Panels A-C provide results based on the Full Sample (2001-2010), 2001-2005, and 2006-2010, respectively. Panels D-F provide results based on Large Cap, Mid Cap, and Small Cap firms, respectively. Intraday periods and subsamples are as defined above. GMM standard errors appear in parentheses.

| Panel A: Full Sample | Regular | Pre-Market | Overnight | After-Market |
|---|---|---|---|---|
| *h(news)* x $10^{-6}$ | 8.262 | 1.094 | | 0.741 |
| | (0.248) | (0.055) | | (0.045) |
| *h(no news)* x $10^{-6}$ | 7.116 | 0.206 | | 0.140 |
| | (0.229) | (0.005) | | (0.017) |
| *c* x $10^{-6}$ | 2.132 | | | |
| | (0.288) | | | |
| $\sigma_d$ | 2.552 | | 1.113 | 2.071 |
| | (0.130) | | (0.027) | (0.157) |
| $\sigma_\varepsilon(news)$ | 2.580 | 1.650 | 1.870 | 3.586 |
| | (0.061) | (0.064) | (0.084) | (0.095) |

| Panel B: 2001-2005 | | | | |
|---|---|---|---|---|
| *h(news)* x $10^{-6}$ | 6.376 | 1.320 | | 0.654 |
| | (0.362) | (0.084) | | (0.041) |
| *h(no news)* x $10^{-6}$ | 5.168 | 0.277 | | 0.101 |
| | (0.308) | (0.010) | | (0.008) |
| *c* x $10^{-6}$ | 4.037 | | | |
| | (0.254) | | | |
| $\sigma_d$ | 3.338 | | 1.120 | 1.325 |
| | (0.077) | | (0.034) | (0.089) |
| $\sigma_\varepsilon(news)$ | 2.975 | 1.676 | 2.249 | 4.359 |
| | (0.088) | (0.095) | (0.133) | (0.147) |

| Panel C: 2006-2010 | | | | |
|---|---|---|---|---|
| *h(news)* x $10^{-6}$ | 5.550 | 0.529 | | 0.802 |
| | (0.153) | (0.023) | | (0.033) |
| *h(no news)* x $10^{-6}$ | 4.962 | 0.117 | | 0.285 |
| | (0.136) | (0.004) | | (0.009) |
| *c* x $10^{-6}$ | 0.333 | | | |
| | (0.025) | | | |
| $\sigma_d$ | 1.292 | | 1.427 | 2.492 |
| | (0.045) | | (0.049) | (0.063) |
| $\sigma_\varepsilon(news)$ | 2.135 | 1.812 | 1.392 | 2.560 |
| | (0.081) | (0.071) | (0.070) | (0.082) |

**Table 4 (continued)**

| Panel D: Large Caps | Regular | Pre-Market | Overnight | After-Market |
|---|---|---|---|---|
| *h(news)* x $10^{-6}$ | 4.273 | 0.449 | | 0.481 |
| | (0.274) | (0.047) | | (0.046) |
| *h(no news)* x $10^{-6}$ | 4.136 | 0.145 | | 0.082 |
| | (0.264) | (0.007) | | (0.020) |
| *c* x $10^{-6}$ | 0.908 | | | |
| | (0.312) | | | |
| $\sigma_d$ | 1.957 | | 0.869 | 1.550 |
| | (0.242) | | (0.053) | (0.291) |
| $\sigma_\varepsilon(news)$ | 1.022 | 1.117 | 1.178 | 3.133 |
| | (0.149) | (0.135) | (0.164) | (0.227) |
| **Panel E: Mid Caps** | | | | |
| *h(news)* x $10^{-6}$ | 7.558 | 0.759 | | 0.784 |
| | (0.274) | (0.107) | | (0.065) |
| *h(no news)* x $10^{-6}$ | 6.683 | 0.167 | | 0.203 |
| | (0.225) | (0.008) | | (0.050) |
| *c* x $10^{-6}$ | 1.138 | | | |
| | (0.456) | | | |
| $\sigma_d$ | 1.991 | | 1.078 | 2.203 |
| | (0.295) | | (0.061) | (0.255) |
| $\sigma_\varepsilon(news)$ | 2.328 | 1.638 | 1.511 | 3.238 |
| | (0.085) | (0.103) | (0.082) | (0.097) |
| **Panel F: Small Caps** | | | | |
| *h(news)* x $10^{-6}$ | 12.007 | 2.062 | | 0.903 |
| | (0.320) | (0.103) | | (0.053) |
| *h(no news)* x $10^{-6}$ | 9.395 | 0.298 | | 0.153 |
| | (0.267) | (0.009) | | (0.010) |
| *c* x $10^{-6}$ | 4.947 | | | |
| | (0.275) | | | |
| $\sigma_d$ | 3.489 | | 1.365 | 2.263 |
| | (0.068) | | (0.026) | (0.094) |
| $\sigma_\varepsilon(news)$ | 3.672 | 2.105 | 2.612 | 4.275 |
| | (0.104) | (0.109) | (0.125) | (0.110) |

**Table 5: Alternate Estimates of Model Parameters**

This table presents efficient GMM estimates of exogenous parameters of the model described in Section 2. Where appropriate, the table reports separate estimates of parameters for each of four intraday periods—the Regular Market, the Pre-Market, Overnight, and the After-Market—and conditional on *news*=1 or 0. In Panel A, the $\sigma_d$ parameter varies with news. In Panel B, the $\sigma_y$ parameter is 2 basis points in the Regular period and 0.2 basis points in other periods. Intraday periods and subsamples are as defined above. GMM standard errors appear in parentheses.

Panel A: $\sigma_d$ Varies with News

| | Regular | Pre-Market | Overnight | After-Market |
|---|---|---|---|---|
| $h$ x $10^{-6}$ | 7.125 | 0.206 | | 0.141 |
| | (0.228) | (0.005) | | (0.017) |
| $c$ x $10^{-6}$ | 2.119 | | | |
| | (0.292) | | | |
| $\sigma_d(news)$ | 0.759 | | 0.373 | 1.711 |
| | (0.049) | | (0.012) | (0.131) |
| $\sigma_d(no\ news)$ | 2.547 | | 1.113 | 2.078 |
| | (0.133) | | (0.027) | (0.159) |
| $\sigma_\varepsilon(news)$ | 2.800 | 2.209 | 1.870 | 3.629 |
| | (0.069) | (0.045) | (0.084) | (0.094) |

Panel B: $\sigma_y$ = 2 Basis Points for the Regular Period

| | Regular | Pre-Market | Overnight | After-Market |
|---|---|---|---|---|
| $h(news)$ x $10^{-6}$ | 12.797 | 0.636 | | 0.924 |
| | (0.372) | (0.027) | | (0.024) |
| $h(no\ news)$ x $10^{-6}$ | 10.994 | 0.171 | | 0.311 |
| | (0.326) | (0.005) | | (0.008) |
| $c$ x $10^{-6}$ | 0.502 | | | |
| | (0.040) | | | |
| $\sigma_d$ | 1.505 | | 1.372 | 2.809 |
| | (0.045) | | (0.033) | (0.050) |
| $\sigma_\varepsilon$ | 2.592 | 1.915 | 1.871 | 3.550 |
| | (0.060) | (0.050) | (0.084) | (0.096) |

**Table 6: Trading Volume and Return Variance Decompositions**

This table decomposes trading volume and return variance using the model introduced in Section 2. The values below result from substituting the estimates of the model parameters in Table 4 into Equations (21) to (25). Panels A-D provide sets of decompositions for the Regular Market, Pre-Market, Overnight, and After-Market periods, respectively. Each panel contains results for the Full Sample (2001-2010), 2001-2005, and 2006-1020 as well as Large Cap, Mid Cap, and Small Cap firms. Numbers in the sample are fractions of either trading volume or return variance.

| | Volume Decomposition | | | Variance Decomposition | | |
|---|---|---|---|---|---|---|
| Panel A: Regular | Liquidity | Market Maker | Informed | News | Other Public | Private |
| Full Sample | 0.921 | 0.022 | 0.057 | 0.068 | 0.598 | 0.334 |
| 2001-2005 | 0.947 | 0.017 | 0.036 | 0.060 | 0.857 | 0.083 |
| 2006-2010 | 0.846 | 0.040 | 0.114 | 0.072 | 0.163 | 0.765 |
| Large Caps | 0.892 | 0.030 | 0.079 | 0.031 | 0.607 | 0.362 |
| Mid Caps | 0.902 | 0.026 | 0.072 | 0.062 | 0.426 | 0.513 |
| Small Caps | 0.943 | 0.017 | 0.040 | 0.042 | 0.739 | 0.219 |
| Panel B: Pre-Market | | | | | | |
| Full Sample | 0.738 | 0.123 | 0.138 | 0.174 | | 0.826 |
| 2001-2005 | 0.769 | 0.123 | 0.107 | 0.151 | | 0.849 |
| 2006-2010 | 0.614 | 0.134 | 0.252 | 0.298 | | 0.702 |
| Large Caps | 0.633 | 0.168 | 0.199 | 0.110 | | 0.890 |
| Mid Caps | 0.690 | 0.135 | 0.175 | 0.229 | | 0.771 |
| Small Caps | 0.802 | 0.103 | 0.095 | 0.191 | | 0.809 |

**Table 6 (continued)**

| | Volume Decomposition | | | Variance Decomposition | | |
|---|---|---|---|---|---|---|
| Panel C: Overnight | Liquidity | Market Maker | Informed | News | Other Public | Private |
| Full Sample | | | | 0.204 | 0.796 | |
| 2001-2005 | | | | 0.230 | 0.770 | |
| 2006-2010 | | | | 0.147 | 0.853 | |
| Large Caps | | | | 0.146 | 0.854 | |
| Mid Caps | | | | 0.191 | 0.809 | |
| Small Caps | | | | 0.203 | 0.797 | |

| Panel D: After-Market | | | | | | |
|---|---|---|---|---|---|---|
| Full Sample | 0.838 | 0.084 | 0.077 | 0.379 | 0.547 | 0.074 |
| 2001-2005 | 0.876 | 0.079 | 0.046 | 0.494 | 0.469 | 0.037 |
| 2006-2010 | 0.638 | 0.115 | 0.247 | 0.222 | 0.209 | 0.569 |
| Large Caps | 0.781 | 0.100 | 0.119 | 0.485 | 0.404 | 0.111 |
| Mid Caps | 0.795 | 0.088 | 0.117 | 0.405 | 0.439 | 0.156 |
| Small Caps | 0.877 | 0.077 | 0.046 | 0.306 | 0.658 | 0.036 |

**Figure 1: News Probabilities Over Time**

This figure shows how probabilities of news arrival for each of four intraday periods (the Regular Market, the Pre-Market, Overnight, and the After-Market) vary by year during 2001-2010. The indicator variable *news* is 1 if there is at least one story (two stories) in the Dow Jones Newswires mentioning a particular Small or Mid Cap (Large Cap) firm and 0 otherwise. All probabilities are calculated by pooling observations for all firms within each year and size group and then averaging across size groups. Intraday periods are as defined above.

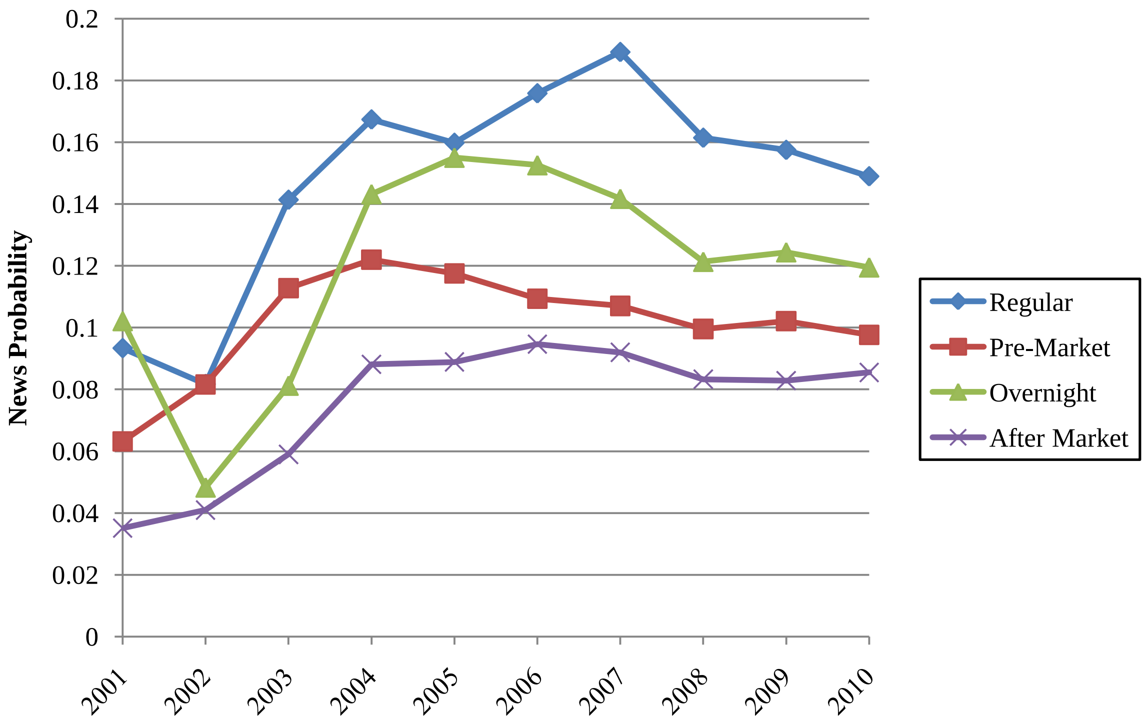
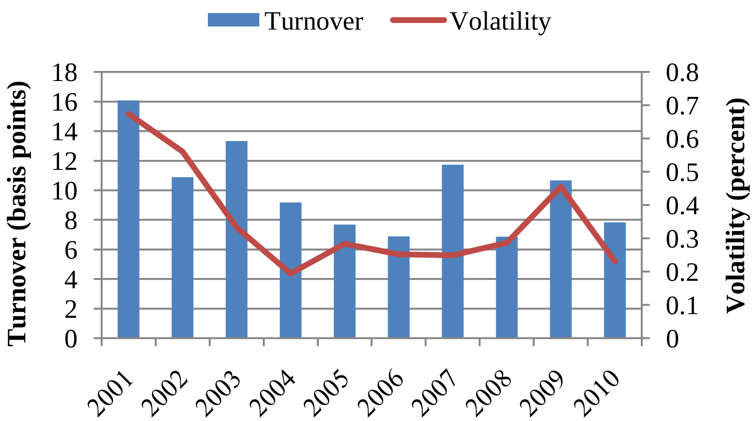
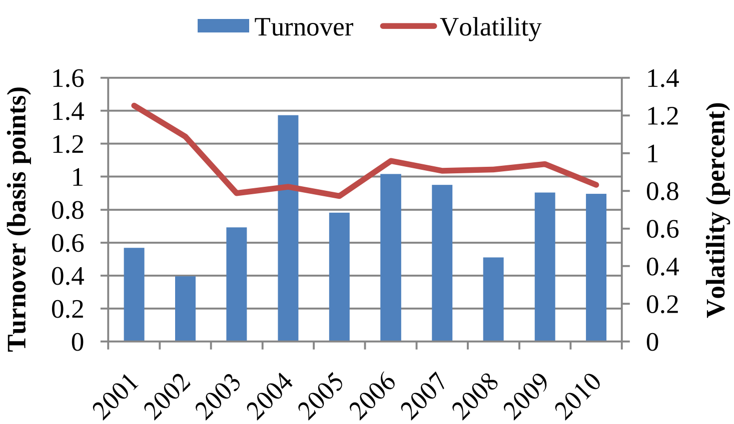


**Figure 2: Differences in News and Non-news Hourly Turnover and Volatility.**
This figure shows how hourly turnover (bars) and volatility (lines) for news and non-news periods vary by year during 2001-2010. News (non-news) periods are intraday periods with *FirstNews* = 1 (*RecentNews*=0). Panels A-D represent the Regular Market, Pre-Market, Overnight, and After-Market periods, respectively. Intraday periods and variables are as defined above.

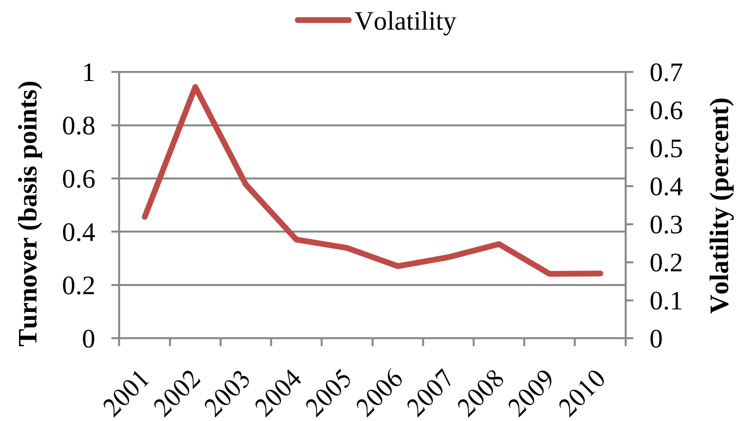
Panel A. Regular Market.




Panel B. Pre-Market.




Panel C. Overnight.




Panel D. After-Market

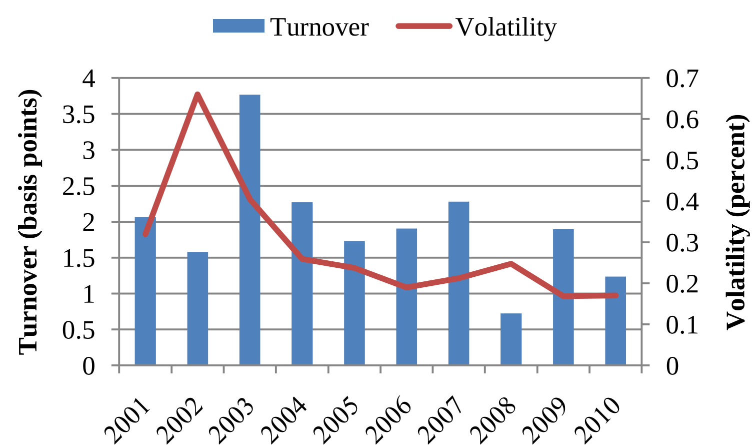
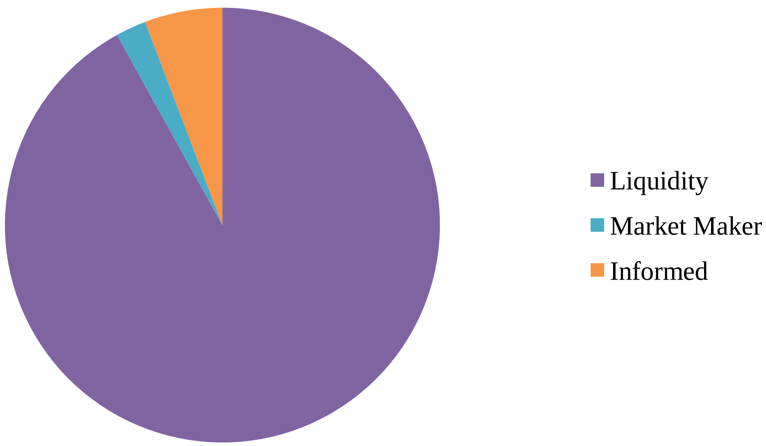


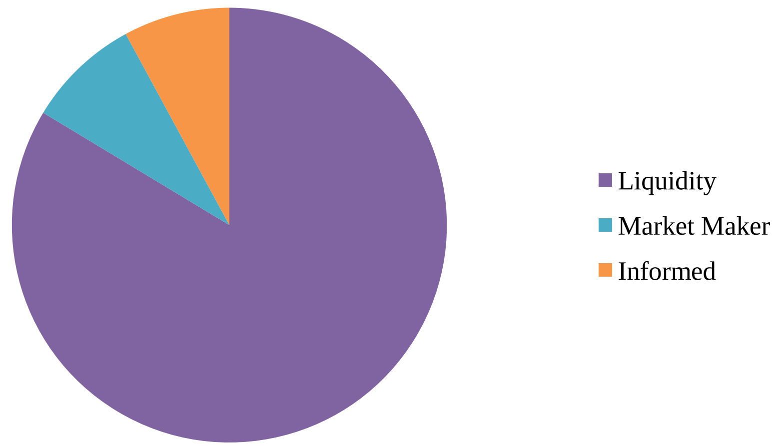
**Figure 3: Trading Volume Decompositions**

This figure decomposes trading volume during 2001-2010 according to Equations (21) to (24). Panels A-C provide decompositions for the Regular Market, Pre-Market, and After-Market periods, respectively.

Panel A. Regular Market Volume




Panel C. After-Market Volume




Panel B. Pre-Market Volume

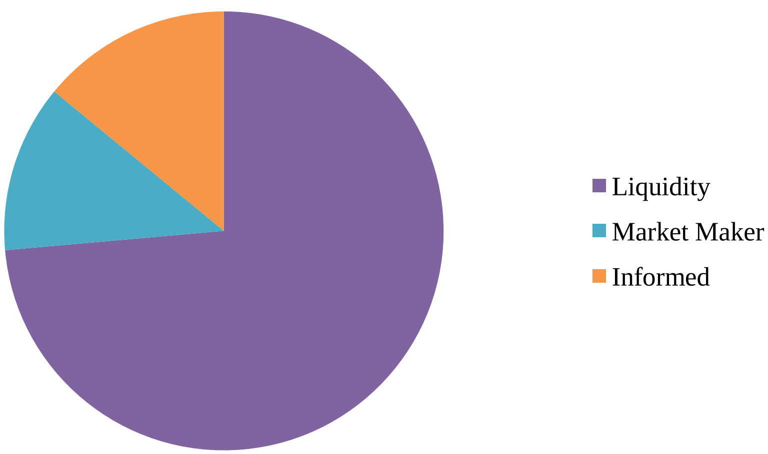
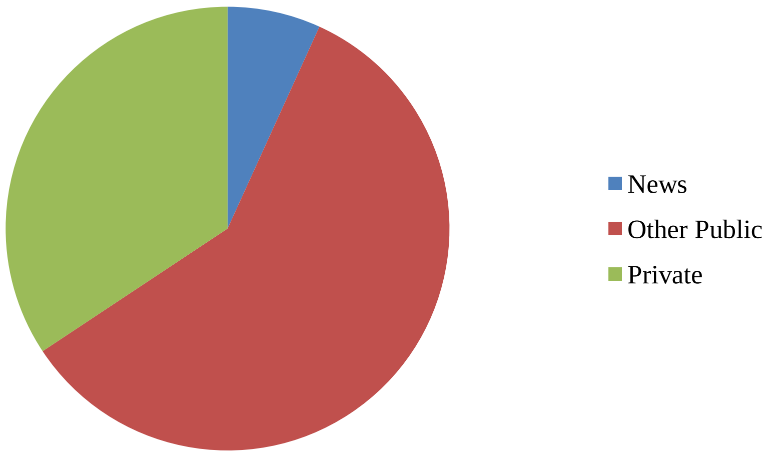


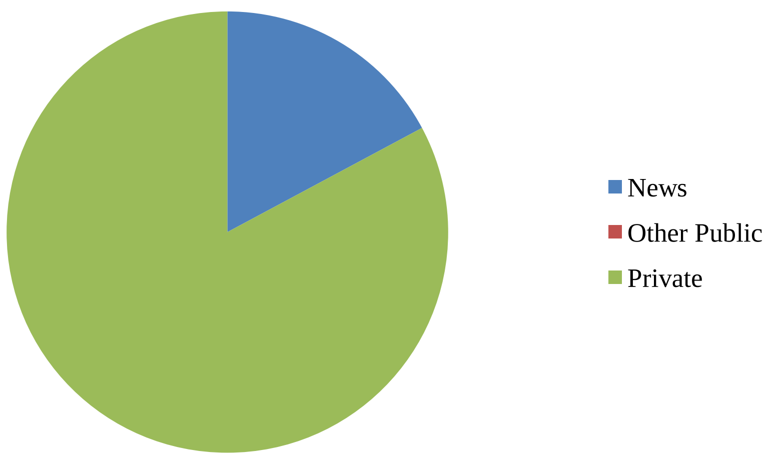
**Figure 4: Return Variance Decompositions**

This figure decomposes return variance during 2001-2010 according to Equation (25). Panels A-D provide decompositions for the Regular Market, Pre-Market, Overnight, and After-Market periods, respectively.

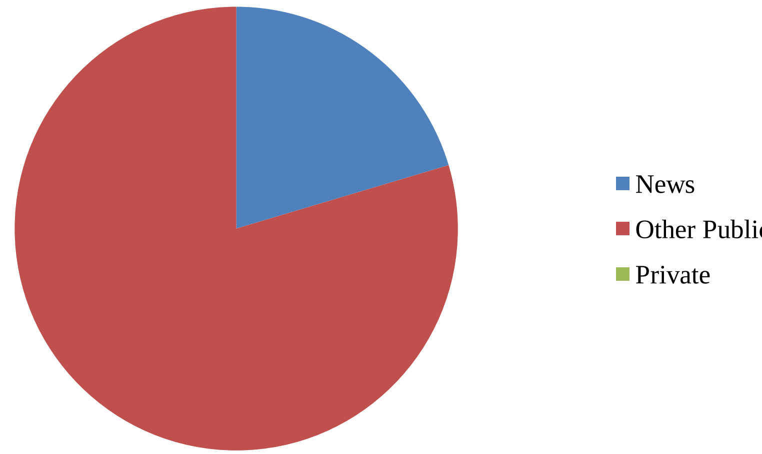
Panel A. Regular Period Variance




Panel B. Pre-Market Variance




Panel C. Overnight Variance




Panel D. After-Market Variance

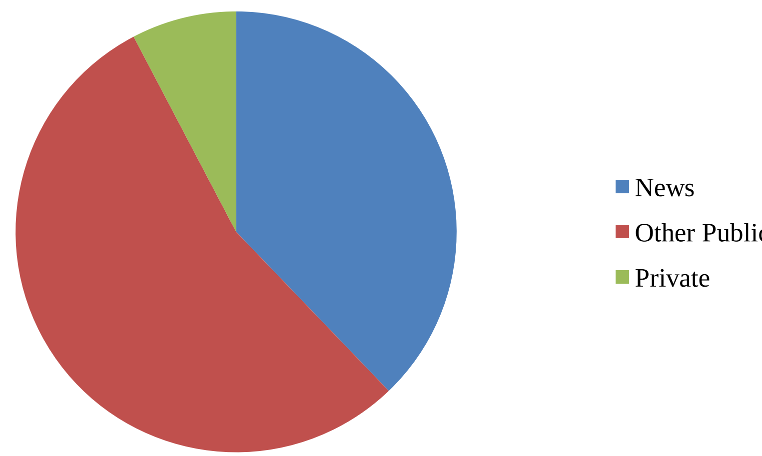
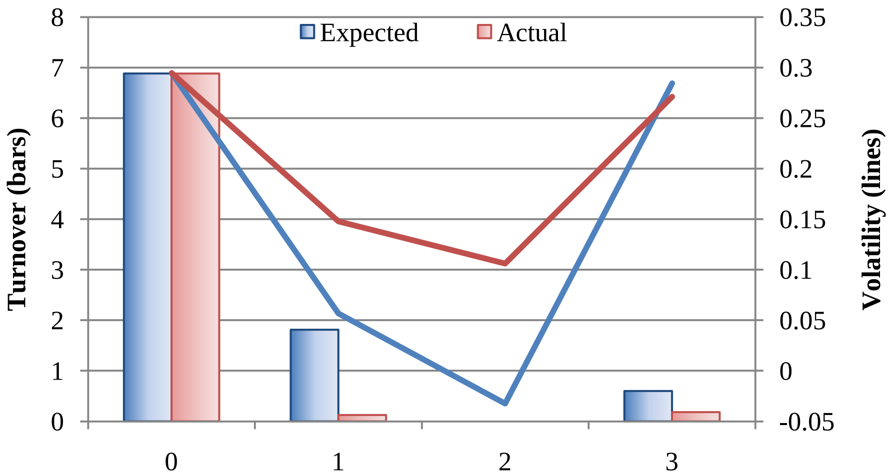


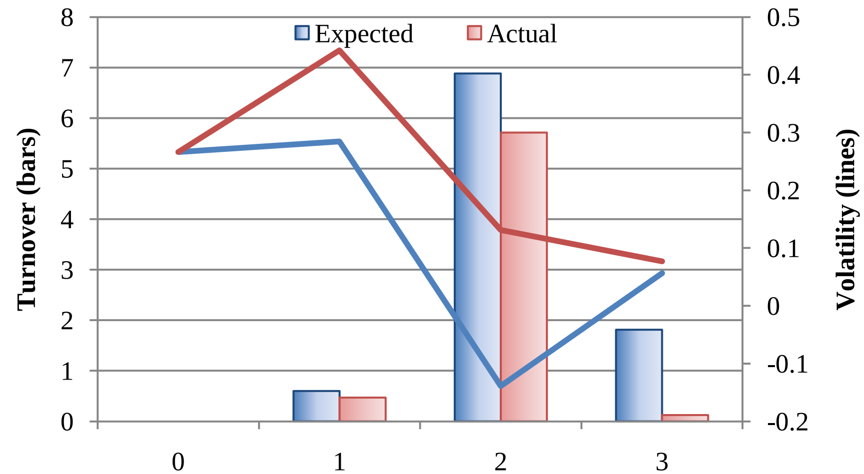
**Figure 5: Event-Time Hourly Turnover and Volatility Following News**

This figure presents predicted and actual hourly turnover and volatility for intraday periods surrounding news arrival (*FirstNews*=1). The predictions come from Equations (26) and (27) in Section 6, using the model parameters from the Full Sample estimates in Table 4. The horizontal access is the period, in event time, relative to the arrival of news. Panels A-D represent news that arrives in the Regular Market, Pre-Market, Overnight, and After-Market periods, respectively. Predicted and actual values are all reported in excess of unconditional expectations. Intraday periods and variables are as defined above.

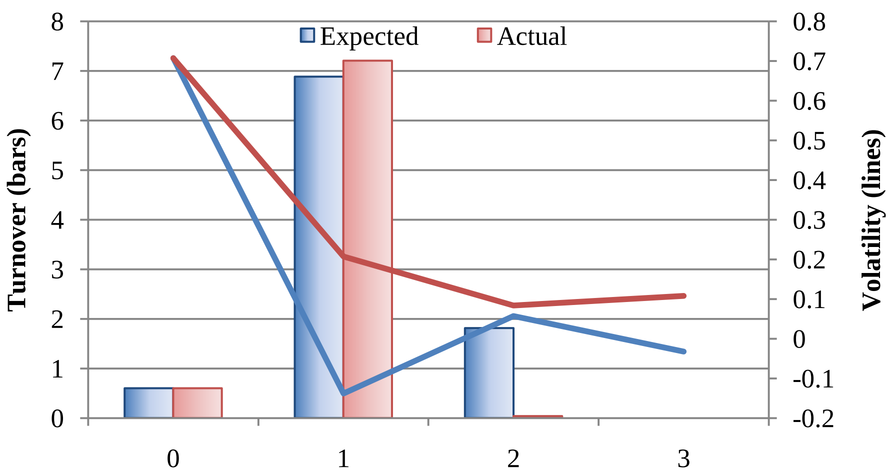
Panel A. Regular Market News




Panel C. Overnight News




Panel B. Pre-Market News




Panel D. After-Market News

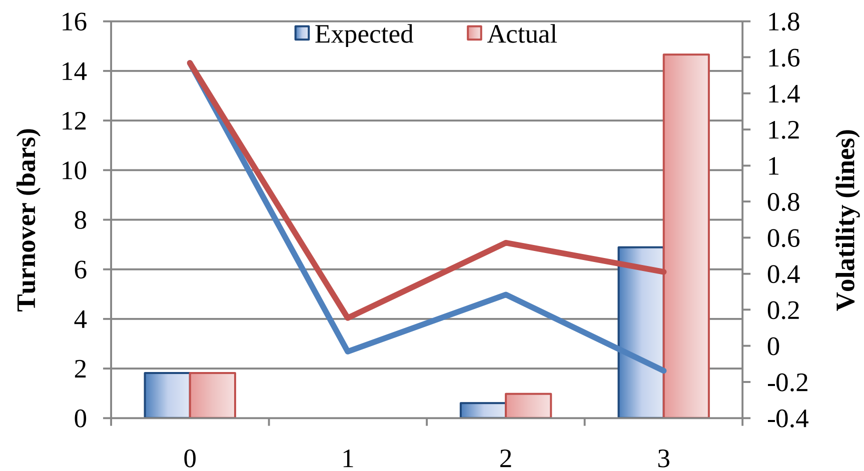
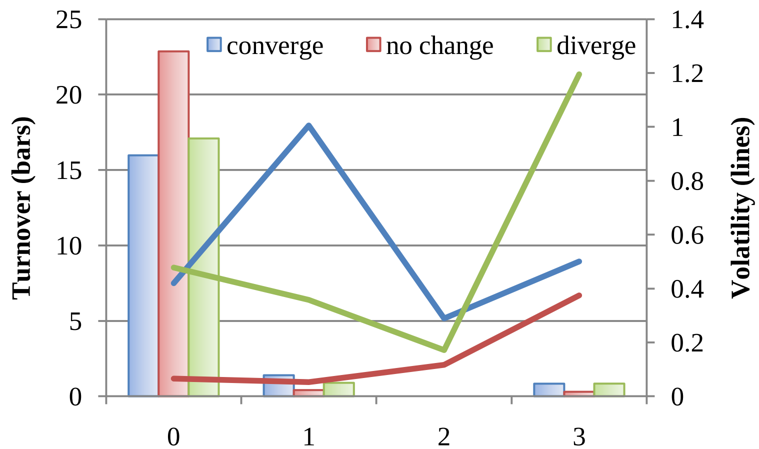
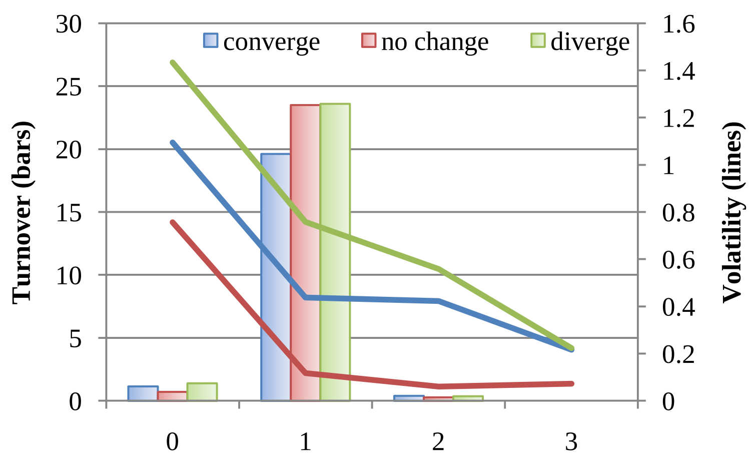


**Figure 6: Event-Time Hourly Turnover and Volatility Following News Sorted by Change in Analyst Forecast Dispersion**
This figure presents predicted and actual hourly turnover and volatility during 2001-2010 for intraday periods surrounding news arrival (*FirstNews*=1). The predictions come from Equations (26) and (27) in Section 6, using the model parameters from the Full Sample estimates in Table 4. The horizontal access is the period, in event time, relative to the arrival of news. Panels A-D represent news that arrives in the Regular Market, Pre-Market, Overnight, and After-Market periods, respectively. Predicted and actual values are all reported in excess of unconditional expectations. Each panel provides separate analysis using news events associated with convergence, no change, or divergence in analysts' quarterly earnings forecasts. These three groups are based on news events in which at least two analysts provide updates in the four weeks prior to the news and at least one updates in the week following the news. Intraday periods and variables are as defined above.

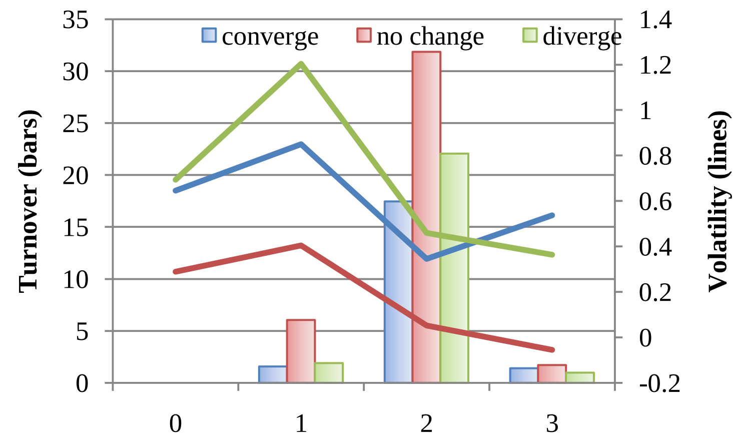
Panel A. Regular Market News




Panel B. Pre-Market News




Panel C. Overnight News




Panel D. After-Market News