A Match in the Dark: Understanding Crossing Network Liquidity

Sugata Ray[1]
University of Florida

Abstract

I model the decision of whether or not to use a crossing network (CN) or a traditional quoting exchange (QE) and derive hypotheses regarding the factors that affect this decision. I test these hypotheses on realized CN volumes and find that the likelihood of using CNs increases and then decreases with increasing relative bid ask spread and other measures of market liquidity. These findings are consistent with the model and reflect two countervailing effects: (1) increased savings on spread related transaction costs on CNs and (2) concerns regarding gaming when QE prices are more easily manipulated. Gaming concerns also decrease the consistency of volume on CNs. Additionally, I find that CN use increases with information asymmetry and the difficulty of disguising informed trading on QEs. In addition to providing empirical support for the model, these findings inform the debate on CN regulation and suggest that gaming is likely to inhibit CN growth.

JEL Classifications: D53, G12, G28

Keywords: Crossing networks; exchanges; gaming; liquidity

March 5, 2010

Comments Welcome

---

[1] Contact: Warrington College of Business Administration, University of Florida, P.O. Box 117168, University of Florida, Gainesville FL 32611-7168. sugata.ray@ufl.edu, (352)392-8022. The author is grateful to Alfred Berkeley, David Brown, Ian Domowitz, Andy Ellner, Itay Goldstein, David Musto, Andy Naranjo, Mahendrarajah (Nimal) Nimalendran, Krishna Ramaswamy, Jay Ritter, Parag Shah and Wayne Wagner. Dominique Badoer provided excellent research assistance. Any remaining errors are the responsibility of the author.

1 Introduction

Crossing networks (CNs), or "dark pools of liquidity," have gained prominence in recent years, attracting increased trading volumes and drawing increased financial press and regulatory scrutiny. CNs work by crossing buy and sell orders for the same stock using prices obtained from quoting exchanges (QEs), such as the NASDAQ or NYSE. CNs currently account for approximately 8.5% of equity trading volumes in the US. This volume is the culmination of a number of years of aggressive growth which shows no signs of abating. The number of crossing networks in the US equity markets has also been steadily growing from a solitary venue (POSIT) in the 1990s to 40 today.[1] Across the Atlantic, European CNs have garnered 4% of equity volume share and are on a similar growth trajectory. This growth begins with the microeconomic decision to switch from traditional exchanges to CNs. What induces traders to route a particular order to a CN rather than a traditional exchange? Which stocks are garnering the greatest volume on CNs, and why? Surprisingly little academic research looks at these empirical questions.

This study focuses on the decision of whether or not to use CNs. I find that CNs are more likely to be used when transaction cost savings outweigh the execution uncertainty, and gaming concerns are minimal. Gaming concerns not only depress CN market shares, but they also decrease the consistency of CN volumes as gaming and non-gaming traders engage in a coordination game where the non-gamers try to avoid gamers while still availing of the benefits of CNs. I also find that higher informational asymmetry and difficulty of disguising informed trading on regular exchanges drives more volume to CNs. Understanding these decision criteria is critical to managing the growth and regulation of CNs.

Currently, trades are typically routed to electronic order books (e.g. NASDAQ), hybrid markets (e.g. NYSE), crossing networks, or other trading venues by either the principal initiating the trade, an agent who might be handling the trade, or by an algorithm employed by either the principal or the agent. For smaller orders by retail traders, this most often takes the form of a limit or market order sent to a limit order book where the order is either executed immediately or is added to the book. If such a small order goes through an agent,

[1] Rosenblatt Securities reports that 8.38% of volume in September 2009 was on 16 crossing networks that report volumes to them. The TABB Group, a consultancy, estimates the compound annual growth rate (CAGR) of CN volumes at 42.5% over the 2007-2010 horizon. Ye (2009) estimates that there are 40 CNs in the US currently and 20 more globally.

it may ultimately still reach a crossing network as the agent tries to find the best execution venue for the trade. Larger, block trades are often parsed out over time and over venues by human traders or by algorithms. Crossing networks are among the venues used to fulfil these large orders. This is done to minimize the impact of the large order on prices. Ultimately, either a human or a computer decides whether to send all or part of an order to a CN in order to minimize transactions costs (including price impact costs) and maximize the probability of execution.

In the model described in this study, a liquidity trader considers the advantages and disadvantages of CNs relative to QEs and decides which to use. Using empirical predictions from the model solution, I formally develop hypotheses regarding situations where traders would be more or less likely to use CNs. Appealing to the intuition that CNs and traditional QEs are substitutes, I use the fraction of volume traded on CNs compared to QEs as a proxy for the probability of sending an order to a CN. I compute the fraction of volume on CNs for a sample of stocks over 13 months and use the cross-sectional and time-series variations in this fraction to empirically test my hypotheses.

The study largely focuses on how the liquidity of QEs, specifically the relative bid ask spread, affects CN use. Economic intuition suggests potential countervailing effects: the higher the bid ask spread, the greater the savings from using a CN and the more volume a CN should garner. On the other hand, stocks with large bid ask spreads are more easily gamed, leading non-gamer traders to avoid CNs, reducing volume. This intuition is formalized in the model, which generates a prediction that CN use increases and then decreases with relative bid ask spread. This tradeoff has been documented from an experimental, practitioner's point of view in Altunata, Rakhlin, and Waelbroeck (2010), which finds that almost all potential savings from CNs are lost to adverse selection (or gaming). [2]

Examining volume on CNs, I find evidence the supports the empirical predication from

[2] The simplest form of gaming is trading on QEs to manipulate the best bid and offer before subsequently trading at manipulated prices on a CN. Conversations with ITG and Pipeline suggest that while overtly manipulating prices in thinly traded stocks on QEs was a popular method of gaming CNs in the past, gaming today is more commonly done by trading at favorable prices generated by QEs with poor price discovery, where future mean reversion is apparent and large. While direct manipulation of quotes is easily detectable and banned, high frequency algorithmic trading that takes advantage of short periods of inaccurate pricing on QEs to trade on CNs is much more difficult to detect. This is in line with Naes and Odegaard (2006), which looks at trades initiated by a large institutional trader on CNs and finds that trades executed on CNs perform significantly worse than trades that fail to execute.

the model. The fraction of volume transacted on CNs initially increases with the bid ask spread, rising from 1.48% of total volume for the lowest decile of bid ask spreads to 2.33% of volume when the relative bid ask spread is 0.63%. The fraction of volume subsequently falls as the relative bid ask spread increases further, reaching a low of 0.26% for the highest decile of relative bid ask spreads. While it is difficult to linearly map the fraction of volume on CNs on to the probability of using CNs, these statistically and economically significant differences suggest corresponding differences in the propensity of traders to use CNs across stocks with different relative bid ask spreads (see section 6.1 for further details). This comports with the economic intuition outlined above, and the effect is robust to adding in a variety of controls. I find similar results when considering the effect of price impact measures on the fraction of volume in CNs, but in formal regression tests, this effect is subsumed into the coefficient on the bid ask spread. [3]

I also find the fraction of volume on CNs decreases with absolute monthly returns, total monthly volume, turnover, (defined as monthly volume divided by shares outstanding) and order flow spread contribution (as measured by Roll (1984)). All of these findings are consistent with orders being routed in accordance with the costs and benefits of CNs. Traders weigh the lowered transaction costs and added anonymity afforded by CNs against the longer execution times, lower probability of execution, and the potential for gaming.

Existing empirical academic research on CNs is nascent. Ready (2009) is the empirical study most closely related to this one. Ready (2009) studies the fraction of institutional trades routed to CNs, finds that CN volumes are negatively related to dollar spreads per share and interprets this as evidence of institutional trader routing to satisfy soft dollar commission requirements. Gresse (2006) examines POSIT data and "shows that DM (dealer market) spreads are negatively related to CN executions." This study refines these results by using relative bid ask spreads and by documenting a clear non-monotonic relationship between relative spreads and the fraction of CN volume. This study also incorporates measures of transaction costs and interprets the findings as support for decision criteria for traders/algorithms choosing between CNs and traditional exchanges on the basis of transaction costs, execution probability and gaming concerns. Naes and Odegaard (2006) shows that adverse selection in crossing networks leads to substantial costs from the failure to trade,

[3]The price impact measure is constructed in a manner similar to Amihud (2002), but uses smaller time intervals to measure mean reversion.

which may outweigh the savings in transactions costs.

Theoretical academic studies include Blume (2007), which notes that CNs are fragmenting liquidity in equity markets. Hendershott and Mendelson (2000) looks at the impact of CNs on existing dealer markets and traders and finds that the cost advantages and liquidity of CNs are the key determinants of their competitiveness and their impact on dealer markets. Hendershott and Mendelson (2000) also considers the tradeoff of adverse selection against lowered transaction costs, concluding that informed traders are more likely to use CNs. This is consistent with empirical results reported in my study. Ye (2009) models a transaction system with a crossing network and stock market and theoretically shows that the crossing network will hurt the price discovery process. Conrad, Johnson, and Wahal (2003) shows that alternative trading systems such as crossing networks tend to have favorable costs compared to traditional institutional trading systems, although Naes and Odegaard (2006) and Mittal (2008) suggest that the selection bias of trades executed in CNs may lead to an overestimation of transaction cost savings. I find that savings on transaction costs play a role when deciding whether to use a CN; however, any potential savings are weighed against gaming concerns, which often lead to the selection bias in execution.

In addition to academic studies, there are also a number of practitioner studies regarding crossing networks: Domowitz, Finkelshteyn, and Yegerman (2009) and Sofianos and Jeria (2008) consider the transactions costs from using CNs versus traditional venues and Mittal (2008) provides a comprehensive survey of "dark" trading venues and details various mechanisms which lead to "toxicity" of volume and unfavorable executions on these pools.

The remainder of this paper is divided into three sections. In section 2, I elaborate on the economic intuition motivating the study. In sections 3 and 4, I set up the model, solve it and formalize testable hypotheses. In section 5, I describe the empirical data used and present empirical results from the study. In section 6, I discuss the empirical results and detail some robustness tests. In section 7, I conclude.

2 Motivation

2.1 What are CNs and why (or why not) use them?

Crossing networks function through traders placing a buy or sell order on the CN for a quantity of shares and a predetermined length of time. These orders are not displayed to any other market participants (also called “dark”) and often do not specify a limit price or require immediate execution. If, during the time the order is active, another trader enters an order taking the other side of the trade on the CN, shares are transferred from the selling trader to the buying trader. The prices are typically obtained as the midpoint of the best bid and best offer prices from the QEs.[4]

Many CNs such as POSIT Now and LCX transact orders immediately as the crosses are available. Other CNs (e.g. POSIT Match, Instinet) instead process batch orders. They take orders throughout the course of a time period, and at the end (often a random instant in a prespecified interval), they match buy and sell orders at the then-prevailing midpoint of the national best bid and offer prices (NBBO). In both cases, the presumption is that any transaction will be executed at the midpoint of the NBBO market, but traders have the option of specifying how much more (or less) favorable they need the price to be vis-à-vis the mid. For example, a trader looking to go long and willing to pay a slight premium may specify his willingness to pay up to mid plus a penny. Similarly, a trader willing to sacrifice speed of execution for a slightly better price might specify his willingness to pay up to mid minus a penny.

Traders also have the option to specify a limit price, or the least favorable price at which they would be willing to transact. Regardless of the exact mechanics, the key point is that traders rely on the NBBO as a price discovery venue. There are a variety of CNs, but they all share this one feature: they use other markets for price discovery in a practice known as “parasitic pricing.”

The obvious advantage of using CNs is that participants no longer pay the bid ask spread to transact. Additionally, the orders are confidential, and only if a cross is successfully executed are the price and transacted volume communicated to the transacting parties. Regardless of the crossing outcome, the parties are never publicly identified. Many participants prize this anonymity. In particular, informed participants trading large positions are less likely to leak information when trading on CNs than on traditional exchanges. In addition to the immediate savings on the transaction costs in terms of the bid ask spread, CNs also minimize the price impact of a trade, leading to further savings.[5]

[4] A more detailed treatment of the institutional details regarding CNs can be found in Ye (2009).

[5] Domowitz, Finkelshteyn, and Yegerman (2009) estimate that traders using the ITG POSIT CN save between 12 - 16 basis points in transaction costs compared to alternative non-dark venues. More generally, Domowitz, Finkelshteyn, and Yegerman (2009) estimate ITG’s ‘Dark Algorithm’ (an algorithm that trades

The two main disadvantages of using crossing networks are (1) a potential lack of immediate executions and (2) the potential for gaming. The lack of immediate execution stems from the fact that crossing networks require both buy and sell orders in order to facilitate a cross. Participants may have to leave an order on a crossing network for a period of time before another party takes the other side of the trade. This problem is exacerbated for batching CNs as described above. Not only must participants hope that a counterparty exists, but they also have no way of knowing if a counterparty exists until after the batching occurs. Lack of immediate execution could pose problems for traders with short-lived information (see Hendershott and Mendelson (2000) for a discussion of this effect) as well as for uninformed traders bound by regulation to complete trades within a given timeframe (e.g. a mutual fund tracking a Russell index over the Russell rebalance).

As the volume on CNs increases relative to QEs, gaming poses a greater threat to further CN liquidity growth. The *Wall Street Journal (WSJ)* article, "Gaming in Dark Pools" reports "Orders may not be as well hidden in dark pools as mutual and pension funds believe. Traders known as 'gamers' are thought to be fishing for orders in ways that may be illegal. [The gamers exploit dark pools by buying] small increments of small cap stocks on conventional exchanges, then [selling] a block in a dark pool at a better price. ... Gaming is more viable with small caps because [they are] thinly traded, making it easier to push the price around. Some dark pools specialize in small caps. One firm that manages dark-pool technology for firms, OnePipe, has tracked individual instances of gaming by looking at gyrations in otherwise sleepy small stocks." A portfolio manager interviewed in the article says about this strategy, "It's free money... It's a neat inefficiency in the marketplace, and dark pools are growing tremendously in size."

Gaming affects thinly traded stocks more than liquid ones. Given that bid ask spreads and price impact are much higher for thinly traded stocks, the CN user faces a tradeoff between the cost savings of avoiding a large bid ask spread and potential losses due to gaming. If non-gaming participants on CNs felt they were trading at manipulated prices, they would abandon the CN and return to QEs, strengthening the price discovery process. At the extreme, if nearly all volume were transacted on a CN, price discovery on the open exchange would lack robustness, and thus trading on CNs would cease, leading to an untenable equilibrium.

across a host of dark venues) saves 8 basis points in transactions costs compared to non-dark venues.

I model a trader's decision of whether to use a CN in the following section and highlight the role QE liquidity plays in CN use.

3 Model

The model is a single interval, two period model. At $t = 0$, an uninformed liquidity trader discovers the need to transact x shares of a stock (ticker ABC) within the time interval $t \in (0, 1)$. The trader may use one of two venues for the trade:

1. Send the block order for x shares immediately to the QE. The QE is populated exclusively by a market maker who lists his bid and offer book for the given ticker, ABC, openly and is willing to transact more the further the price is from the initial base price (we can imagine this increasing spread in size as compensation for trading against informed traders, as per Kyle (1985)). Figure 1 shows the depth graphically for such a market. During the interval, the market maker will not replenish orders. At $t = 1$, the market maker will replenish the order book after checking with his sources to see if there has been any fundamental changes in ABC share pricing. This is in line with the concept of slow moving capital outlined in Duffie (2010).

2. Initially send the order to the CN and hope that the order gets crossed. If the order does not get crossed in the time interval, the order will be re-routed to the QE right before the interval ends and will immediately transact there.

The CN has liquidity characteristics such that if the liquidity trader places the order to the CN, it will be filled by another liquidity trader with probability π during the interval. Routing the order to the CN will also result in uncertain and potentially delayed execution. This imposes a cost on the liquidity trader, represented by δ. Additionally, there is an arbitrageur who can pay c to observe the order placed on the CN. The arbitrageur can then trade freely in the QE and CN to maximize profits. If he trades in the CN and the original liquidity trader's order is going to be successfully filled by another liquidity trader, the arbitrageur will only be able to fill a fraction, α, of the order.[6] All traders are informed

[6]This parameter can be thought of as a measure of how CNs route orders between true liquidity traders and arbitrageurs. Although this simply resolves into an additional cost for the arbitrageur in this model, Ye (2009) shows that this parameter is critical to determining CN viability.

in as much as they know that there has been no fundamental change in ABC and the true price is p_0. The game theoretic equilibrium will tie the arbitrageur's optimal actions back to the liquidity trader's initial decision.

3.1 Interpretation of model parameters

In addition to the functional form of the QE depth chart (the function $f(p)$ from Figure 1), there are six parameters in this model. Three of them have liquidity based interpretations, one is a measure of the delay/uncertainty in using CNs, one is a stylistic representation of the costs of observing orders on the CN, and the last is a measure of order assignment on the CN.

- x is the size of the block the liquidity trader needs to transact in the time interval.
- δ is the delay penalty paid by the liquidity trader for routing through the CN. This can be thought of as the cost of waiting for the batching time to occur or waiting for a complementary order to be entered into the CN. This is paid regardless of whether the order executes on the CN.
- k is the liquidity of the QE. Referring to Figure 1, we can see that for a stock where the mid of the bid and ask is p_0, at any given price, p, the market maker on the QE is willing to transact up to $k \times f\left(\left|\frac{p}{p_0} - 1\right|\right)$, where $f(.)$ is a function of the price, p. If k were a very large number, the large block trader would not need to worry about moving the market too much when trading on the QE.
- π is the probability that there is another large liquidity trader on the CN taking the opposite position as our original liquidity trader during the interval. As CN liquidity increases, this probability is likely to increase. In this simplified model, we simply resolve all orders on the CN into a binary probability of filling the order. Of course, it is much more likely that in a less liquid CN, a small part of the x share block order might be filled. Modeling partial fulfilment on the CN adds complexity while the model results remain qualitatively unchanged.
- c can be thought of as the difficulty that an arbitrageur faces in observing the order put on the exchange. If, for example, the arbitrageur has to place a series of small

trades and use that to infer the presence of a large order, the cost would be the risk of those small trades and the cost to unwind. Alternatively, if there is some chance the CN will detect the gaming attempt and ban the arbitrageur, c can be thought of as the expected cost of "being caught."

- α is the fraction of the x shares the arbitrageur will be assigned if there is a large liquidity trader opposite the initial liquidity trader. If α is 0, that would mean the arbitrageur would not be able to trade on the CN at all if there is another liquidity trader taking the other side of the original trade. If α is 1, the arbitrageur will be able to trade the entire x_0 share lot, despite the presence of another liquidity trader.

Section 4 solves the model and presents the equilibrium probabilities for CN use by the liquidity trader and gaming attempts by the arbitrageur as a function of relative bid ask spreads on the QE.

4 Model Solution and Empirical Predictions

I solve the model generally and then highlight the empirical predictions by using a simple functional form of $f(.)$ and a fixed set of parameters. I will assume that the liquidity trader is interested in buying x shares although the argument is exactly symmetrical if it is a sale instead of a purchase. The liquidity trader values immediate execution and wants to minimize transaction costs. Execution on the QE is immediate and incurs no delay penalty. Execution on the CN incurs a delay penalty of δ, representing both the uncertainty and the delays arising from waiting for batching times common in real life CNs.

4.1 Liquidity trader's options

The liquidity trader has the option to trade exclusively on the QE. In this case, the trader will buy until the marginal offer is p' such that

$$x = \int_{p_0}^{p'} k \times f(\frac{p}{p_0} - 1)dp \tag{1}$$

The liquidity trader values the shares he has purchased at p_0. The total loss to the liquidity trader is going to be l_{QE}. The liquidity trader has to make this trade at a loss for exogenous reasons discussed above in section 2.1.

$$l_{QE} = \int_{p_0}^{p'} p \times k \times f(\frac{p}{p_0} - 1)dp \tag{2}$$

Now we consider how this would change if the liquidity trader could use the CN and was certain the arbitrageur would not be trading in the CN. The liquidity trader would send the order to the CN and only if the order did not execute there (probability of $1 - \pi(x_0)$) would the order be sent to the QE. If the order was executed on the CN, the order would be crossed at p_0 for no loss to the liquidity trader. Sending the order to the CN would incur delay penalty δ.

$$l_{CN,NoArb} = (1 - \pi) \times \int_{p_0}^{p'} p \times k \times f(\frac{p}{p_0} - 1)dp + \delta \tag{3}$$

This is a simplification, as the probability of execution on the CN will be related to execution depth. However, the delay penalty, δ, can still be interpreted as the costs of having to wait for a batching period to occur. The decision to use the CN hinges on the tradeoff between the delay penalty and the transaction costs paid for using the QE. In the absence of an arbitrageur, the liquidity trader will choose to use the CN when the following condition holds.

$$\delta < \pi \times \int_{p_0}^{p'} p \times k \times f(\frac{p}{p_0} - 1)dp \tag{4}$$

This leads to a binary relationship between the probability of the liquidity trader using the CN, q, and the depth of the CN, k. This is displayed in Figure 3. In the absence of the arbitrageur, the liquidity trader will never use the CN when the QE has sufficient depth (relative bid ask spread of 0.50% or less under the parameters stated in the figure) and will always try the CN if the QE liquidity is below this threshold.

4.2 Arbitrageur's decision

We model the arbitrageur as having two decisions to make: (1) Whether to pay c to observe the orders on the CN and (2) What to do once they have observed the order. Naturally, the first decision will be made contingent on the optimal actions from the second decision. The arbitrage we focus on once the order has been observed is that the arbitrageur can trade first on the QE and change the mid of the best bid and offer from p_0 to some other price and then trade against the liquidity trader on the CN.[7] For example, if the liquidity trader is buying x on the CN, the arbitrageur can first buy on the QE, raising prices, before selling to the liquidity trader on the CN at the artificially high price. Like the liquidity trader, the arbitrageur values all positions in ABC at p_0 per share.

The cost of increasing the bid-ask on the QE from p_0 to p_{arb} will be $\lambda(p_{arb})$

$$\lambda(p_{arb}) = \int_{p_0}^{p_{arb}} p \times k \times f(\frac{p}{p_0} - 1)dp \tag{5}$$

where p_{arb} is a choice variable for the arbitrageur. Once this buy order has been placed and executed on the QE, the arbitrageur can place a bid at p_{arb} - ϵ (where ϵ is a small number) and the arbitrageur can submit an order to the CN to sell x to the liquidity trader at the higher mid price on the QE (approximately p_{arb}). The gain from this trade is γ.

$$\gamma(p_{arb}) = ((1 - \pi) + \alpha\pi) \times x \times (p_{arb} - p_0) \tag{6}$$

The arbitrageur will be able to sell all x shares if there is no competing liquidity trader (probability of $1-\pi$) and only α fraction of the shares if there is a liquidity trader (probability of π). Thus the arbitrageur will only pay cost c if the gain from arbitrage, g_{arb}, is greater than the cost to observe the order.

$$g_{arbs} = \max_{p_{arb}} [\gamma(p_{arb}) - \lambda(p_{arb})] - c \geq 0 \tag{7}$$

[7] In reality, CNs can easily detect price manipulation in this manner. Gamers generally wait for natural, temporary imbalances in the order book to shift the QE mid price too low or too high before placing the appropriate trade on the CN. Although not explicitly modeled, this would still lead to a higher incidence of gaming for stocks with thinner QE liquidity, as the mid of the bid and ask quotes are easily moved through small liquidity trades and often result in the temporary imbalances necessary for gaming to occur. (see Mittal (2008) for details)

We also note that if the arbitrager pursues the arbitrage trade, the amount lost for the liquidity trader is actually greater than that gained by the arbitrageur. This is because if there is a second liquidity trader in the market, that trader also gets some of the gains from the arbitrageur moving the QE market. Thus the loss to the liquidity trader if there is an arbitrage trade is l_{Arb}

$$l_{Arb} = x \times (p_{arb} - p_0) \tag{8}$$

In order to completely characterize the game theoretic equilibrium, I assume a simplified depth chart for the QE where the market maker simply supplies a uniform amount of liquidity, k, at each price. Figure 2 shows the QE under this assumption. I characterize a mixed strategy Nash equilibrium where the liquidity trader and arbitrageur each choose the probability with which to trade on the CN (q) and pay c to look for arbitrage on the CN (r), respectively . Under the mixed strategy I define the following:

$$\begin{aligned} q &= prob(\text{ Liquidity trader uses CN }) \\ r &= prob(\text{ Arbitrageur pays c }) \end{aligned} \tag{9}$$

The arbitrageur's payout in this mixed strategy (akin to the left hand side of inequality 7) is g_{Arb} and loss to the liquidity trader is l_{LT}

$$\begin{aligned} g_{arb} &= r\left[-c + (1-q)0 + q[(1-\pi)(p_{arb} - p_0)x + \pi(\alpha)(p_{arb} - p_0)x - \frac{(p_{arb} - p_0)^2 k}{2}]\right] \\ l_{LT} &= (1-q)\frac{x^2}{2k} + q\left[\delta + (1-r)(\pi.0 + (1-\pi)\frac{x^2}{2k}) + r(\frac{[1-(1-\alpha)\pi]x^2}{k})\right] \end{aligned} \tag{10}$$

Conditional on the CN being used, the optimal strategy for each party is to select q and r such that the other party is indifferent to their choice of the probability, and the mixed strategy equilibrium yields optimal probabilities, q^* and r^*.

$$q^* = \frac{2kc}{(1-(1-\alpha)\pi)x^2} \tag{11}$$

$$r^* = \frac{\frac{x^2}{2k} - \delta}{\frac{x^2}{2k}(\pi - 1) + \frac{(1-(1-\alpha)\pi)x^2}{k}} \tag{12}$$

Note that q^* is bounded at 1, thus, if the expression in equation 11 ever exceeds 1, it simply means that the arbitrageur can never execute a profitable trade and $q^* = 1$ and $r^* = 0$. The liquidity trader will always trade on the CN first and the arbitrageur will never attempt to arbitrage. Similarly, r^* is bounded below at 0; if equation 12 were to yield a negative expression, that would mean that in order for the liquidity trader to be indifferent, the arbitrageur would have to be improving the liquidity trader's welfare when the liquidity trader uses the CN, which is not possible. This outcome also corresponds to situations where the QE is liquid enough that the cost of delay from using the CN outweighs any potential transaction cost savings (or inequality 4 holds).

These probabilities (q^* and p^*) are plotted in Figure 3. The QE still dominates the CN when relative bid ask spreads are below 0.50% as the QE's immediacy outweighs the minimal transaction cost savings from the CN. However, once the CN becomes a viable option (relative bid ask spread $\geq$ 0.50%), instead of always choosing the CN, the liquidity trader and arbitrageur pick q^* and p^* as per equations 11 and 12, respectively. When the bid ask spread is just above the 0.50% threshold, the liquidity trader will be the most likely to use the CN, as the risk of being gamed by the arbitrageur is minimal. As the bid as spread increases, the prevalence of gaming (r^*) increases and the attractiveness of the CN to the liquidity trader decreases. In the model, all savings from using the CN are lost to gaming. This is in line with empirical findings from Altunata, Rakhlin, and Waelbroeck (2010), who suggest that "opportunistic savings in dark aggregators [which route orders to multiple CNs] are almost entirely lost to adverse selection [trading on CNs during periods of temporary mispricing on QEs]."

4.3 Testable hypotheses

In this section, I formalize testable hypotheses, relying on model solutions and economic intuition from the advantages and disadvantages of CNs outlined above. I propose various factors that affect the probability that a trader (or algorithm) chooses to route a trade to a CN. In testing these hypotheses, I rely on the intuition that CNs and QEs are likely to serve as alternative trading venues, rather than assuming that CNs generate new trades.[8] Thus, I can use the fraction of volume traded on CNs for a given stock in a given month as a proxy for the probability of trader/alogrithm routing orders for that stock in that month to a CN. This allows me to test my hypotheses on empirical data by looking at how factors that I claim influence the decision to use a CN actually affect the fraction of volume traded on CNs.

4.3.1 Bid ask spreads affect the probability of using CNs

As seen in Figure 3, the model predicts that CN use (q^*) is zero for low relative bid ask spreads ($\leq 0.50\%$), increases sharply once transaction costs savings outweigh costs associated with CN delays and uncertainties (relative bid ask spread = 0.50%), and then slowly decreases with relative bid ask spread as gaming concerns become a factor in the liquidity trader's routing decision. This leads to two predictions about the relationship between CN use and the relative bid ask spread:

The probability of using CNs increases with the relative bid ask spread while overall levels of the relative bid ask spread are low As long as gaming concerns are not a factor, the decision of whether to use a CN is driven by whether transaction cost savings from the CN outweigh costs associated with delay and uncertainty from CN use. Thus CN use increases with the relative bid ask spreads.

The probability of using CNs decreases with the relative bid ask spread when overall levels of relative bid ask spreads are higher and gaming concerns are a factor in the liquidity trader's routing decision Pricing for stocks with higher bid ask spreads is more easily manipulated and CN use for these stocks is thus more vulnerable to gaming. This would decrease the likelihood of non-gamers using CNs for such stocks for

[8]In other words, I assume that CNs and QEs are substitutes, rather than complements or separate goods. Economic intuition suggests that they are unlikely to be complements, except in the mildest sense for gamers who manipulate prices on QEs before trading on CNs. A fixed effect regression of CN volume fraction on total monthly volume shows a statistically weak, slightly negative relationship, suggesting that they are more likely to be substitutes rather than separate goods.

fear that any execution on CNs will occur at a manipulated price. This would, in turn, decrease the fraction of volume transacted on CNs.

4.3.2 CNs are more likely to be used by informed traders, particular when it is difficult to disguise informed trading on open exchanges

This hypothesis relates to the advantages afforded by anonymity on CNs. Economic intuition and empirical predictions from Hendershott and Mendelson (2000) suggest that informed traders are more likely to use CNs.[9] This is particularly so when it is difficult for an informed trader to use an open exchange without revealing the information. This hypothesis suggests that CN use increases in information asymmetry (measured as the component of the bid ask spread arising from information asymmetry), and this increase is greater when it is difficult to disguise informed trading on the QE. I use total volume and turnover as proxies for the ease of disguising trades on open exchanges.

4.3.3 Liquidity begets liquidity: The more likely an order is to execute on a CN, the more likely a trader is to use it

This hypothesis relates to the uncertain execution on CNs. Trader concerns regarding uncertain execution would be moderated by historical evidence of successful transactions using the CNs. If this were the case, we would expect persistence in the fraction of volume on CNs.

4.3.4 Stocks that are easier to game, such as those with high bid ask spreads, will have less steady volumes on CNs, and may exhibit negative autocorrelation where high trading volume in one period or more periods is reversed in subsequent periods by gamers discovering such volume

This hypothesis follows from a combination of the higher cost savings and the ease of gaming of less liquid stocks. If there is a period where a large fraction of volume for such a stock is transacted on a CN, we may expect that gamers will start to take notice. Subsequent gaming will lead to a fall in volume in subsequent periods. This is in contrast with stocks

[9] This is an explicit empirical prediction from Hendershott and Mendelson (2000). It does make an exception for traders with very short lived information, who would prefer the immediacy the QEs provide, despite the price impact of their trades.

that are less easily gamed, where we expect a more steady stream of volume to the CN, as outlined above in Hypothesis 4.3.3.

5 Empirical Results

5.1 Data

This study uses transaction data from TAQ, price and volume data from CRSP and CN volume data obtained from the Nasdaqtracker website.[10] I construct a panel dataset of 2,869 stocks over 13 months from June 2005 to June 2006. CN volume is defined in the same manner as Ready (2009), where the market participant ID (MPID) is listed as 'ITGI' (Posit), 'LQNT' (Liquidnet) or 'BLOK' (Pipeline Trading).[11] The summary statistics for the data are presented in Table 1. The average volume traded on the CNs over all stocks during this period is 1.4%, in line with Ready (2009). Average bid ask spreads for each stock are calculated at the daily level from the TAQ database. The average relative spread, defined as the difference between the bid and the ask divided by the price of the stock is 2.3%. The average log percentage spread is the average of the log of the percentage bid ask spreads defined above. Absolute monthly return is simply the absolute with-dividend monthly return as reported by CRSP. Order flow spread contribution is the fraction of the bid ask spread accounted for by the Roll (1984) estimate of the order flow component of bid ask spread.[12] The smaller this number, the larger the contribution of asymmetric information to the bid ask spread (see Kyle (1985) for details). Monthly turnover is monthly volume divided by total shares outstanding. Totaly monthly volume is the number of shares traded in a given month for a given stock.

Table 2 presents the correlations between these variables. CN volume fraction is nega-

[10] This is the same data used by Ready (2009), but due to site changes, the data currently available only spans 13 months from June 2005 through June 2006.

[11] Conversations with ITG suggest that only about a quarter of the volume with 'ITGI' listed as MPID goes through the POSIT crossing network with the rest consisting of algorithmic trades and other brokerage activity. Liquidnet and Pipeline do not suffer from this problem. Robustness tests confirm my results stand if only Liquidnet and Pipeline trades are used for computing CN volume in the analysis.

[12] The order flow spread contribution is calculated as $\frac{2\times\sqrt{cov(R_t,R_{t-1})}}{A-B}$. The numerator is the implicit bid ask spread in the absence of informational asymmetry as approximated by Roll (1984). The denominator is the actual bid ask spread. The fraction represents the component of the bid ask spread that does not arise from information asymmetry concerns.

tively correlated with the average relative bid ask spread, suggesting traders have gaming concerns. The average relative bid ask spread is also negatively correlated with market capitalization, monthly turnover, monthly volume and share price.

5.2 Relative bid ask spreads and the fraction of volume traded on CNs

In Figure 3, I plot the average fraction of volume traded on CNs against the relative bid ask spread in percentage. We see that the fraction of the volume increases with relative bid ask spread until the relative bid ask spread is about 50 bps. Above this point, the fraction of CN volume decreases with relative bid ask spread. The increase and subsequent decrease of CN volume fraction with relative bid ask spread is consistent with the tension between increasing cost savings with the spread (which leads to initial increase in fraction of volume on CNs) and the increased potential for gaming (which leads to the subsequent decline). The low fraction of volume when bid ask spreads are low is also consistent with the low cost savings from CNs for such stocks combined with the uncertain execution on CNs. In fact, the probability of the liquidity trader using the CN in the mixed strategy Nash equilibrium closely mirrors the observed CN volume fraction.

5.3 Univariate analysis

As a more formal test of hypothesis 4.3.1, the hypothesis that bid ask spreads affect the probability of using a CN, I sort the stock months into deciles by relative bid ask spread and compute the average fraction of volume on CNs for each decile. The results are presented in Table 3. In addition to the average fraction of CN volume, I also present standard errors of these estimates. We can see that the fraction of volume on CNs increases with relative bid ask spread until the 4th decile and then decreases. The difference in average fractions of CN volume between the lowest and the 4th decile and the 4th and the highest decile are statistically significant.

Hypothesis 4.3.1 is consistent with this pattern. For stocks where the relative spread is low, the cost savings from using the CN are low and are outweighed by the immediacy of execution on the open exchanges. As the spread increases, the larger cost savings make the CNs more attractive, and thus, the fraction of volume on the CN increases.

However, as the relative spread increases beyond a point, the robustness of price discovery on the open exchanges decreases, and the parasitic pricing of the CNs is open to gaming. This leads to a decrease in the fraction of the volume on CNs.

These findings provide initial support for my hypothesis. In addition to testing the other hypotheses in the following sections, I also see if this effect is robust after controlling for other factors.

5.4 Bivariate analysis

In bivariate analyses, in addition to sorting by the relative bid ask spread, I also sort by the historical CN volume fraction and total volume. These results are presented in Tables 4 and 5, respectively. The stock months are sorted by relative bid ask spread (rows) as before, and by the other variables (columns).

Table 4 presents a bivariate analysis of CN volume depending on relative bid ask spread and the lagged CN volume fraction. Note that the bottom four deciles of observations all have no trading on the CNs. Thus, the first distinct decile above the first is designated the fifth decile. As a robustness check, we see that the univariate pattern of CN volume fraction increasing and then decreasing with the relative spread is largely maintained after controlling for historical CN volume fraction. Specifically, the pattern is maintained for observations in the all but the highest decile of historical CN volume. This is in line with the economic intuition outlined in Section 2.1: if execution on CNs is more likely (i.e. when historical CN volume fraction is high), the advantage of open exchange immediacy for low relative bid ask spreads is somewhat negated.

Table 4 also allows us to directly test Hypothesis 4.3.3. We see a clear trend of increasing CN volume as historical CN volume fraction increases. This supports the hypothesis that liquidity begets liquidity. The more likely execution is on a CN, the less uncertainty faced by a trader and the more likely the trader is to place the order on the CN, further increasing liquidity.

Table 5 presents a bivariate analysis of CN volume depending on relative bid ask spread and total transacted volume. Again, as a robustness check, we can see that even after controlling for total monthly volume, CN volume fraction increases and then decreases with relative bid ask spread. This is the case for all except the 1st and 3rd deciles, where the estimates of CN volume are noisy and do not demonstrate the pattern as clearly.

5.5 Regression analysis

As another test of my hypotheses, I run a regression analysis with the fraction of volume on CNs as the dependent variable. Table 6 presents the results of this regression. We see that the logged relative bid ask spread and the same variable squared have economically and statistically significant effects, both as a standalone regressors in the base specification (first column), as well as in conjunction with other explanatory variables (second column).[13]

I also run a split-sample linear regression to address the non-monotonic nature of the relationship. The third and fourth columns drop the squared log relative bid ask spread and split the sample into observations with relative bid ask spread less than 0.63% and more than 0.63%, respectively. The central point of 0.63% is chosen based on the univariate analysis in Table 3, which suggests that CN volume fraction hits an interior maximum at that relative bid ask spread. From the coefficients on the log relative bid ask spread in these regressions, we can see that the spread has an economically and statistically significant effect on the fraction of volume on the CN. In particular, increasing(decreasing) the log relative bid ask spread by 1 standard deviation from a relative bid ask spread of 0.63% decreases(decreases) the fraction traded on the CN by 0.51% (0.58%), or about one third of the average CN volume percentage. This evidence supports Hypotheses 4.3.1. [14]

The second column also examines the effect of the variables that affect informational aspects of trading. The absolute return is a proxy for whether new information has been released in a given month. We see that in months where new information is released (absolute return is high), CNs account for a lower fraction of the volume. This is most easily explained as a mechanical result of the price discovery function of QEs at work. In months where absolute return is high, price discovery is required, and this increases trading volume on the QE. An alternative explanation is the perception that CNs provide poorer execution during periods of high volatility. [15]

Order flow spread contribution (OFSC) is a measure of the non information-asymmetry

[13]Log spreads are used to mitigate the effect of outliers. log_{10} is used rather than log_e to facilitate interpretations of economic magnitudes of the effect of the relative bid ask spread.

[14] Ready (2009) finds that CN volume decreases with spread rank measure, an institution specific rank ordering of its trades by absolute dollar spreads. This is interpreted as institutions choosing to trade high bid ask spread on quoting venues to satisfy soft dollar agreements.

[15]See Domowitz, Finkelshteyn, and Yegerman (2009) for details on this perception and a comparison of CN performance across different volatility regimes.

component of the bid ask spread. The lower this number, the more of the bid ask spread can be attributed to information asymmetry and the more trading is conducted on CNs. This is in line with Hypothesis 4.3.2 as well as the empirical predictions from Hendershott and Mendelson (2000). When a stock has high information asymmetry, it is harder to disguise trades on QEs and the anonymity of CNs is more appealing. Table 7 presents results of two interaction regression specifications where interaction terms of the OFSC with the total monthly volume and monthly turnover are added as regressors. Although statistically insignificant, these interaction terms are directionally consistent with information asymmetry having a larger effect when informed trading is harder to disguise on QEs due to low trading volume.

The second column also shows that the effect of the relative bid ask spread is robust in the presence of time dummies, various explanatory variables and controls. I run this specification with stock fixed effects to see if the results are cross-sectional or whether they hold in time series as well. The results of this specification are presented in the fifth column. Coefficients for the logged relative bid ask spread squared and absolute return remain statistically significant, and have the same sign, which indicates that there is a time series component to these results as well. For example, the same stock may have time varying bid ask spreads, and we see that empirical tests of the effect of changes in the bid ask spread for that stock support Hypothesis 4.3.1. The lower magnitude of these coefficients indicates that there is a cross-sectional aspect to these results as well.

The lagged CN volume fraction is significant and positive in all specifications except the fixed effects specification. This supports Hypothesis 4.3.3: liquidity begets liquidity. The coefficient on the lagged fraction of CN volume in the specification with controls is 0.260, indicating that a 1% increase in CN volume fraction in the previous month increases the CN volume fraction in the current month by an economically and statically significant 0.26%.

As a test of Hypothesis 4.3.4, I calculate the autocorrelation of CN volume fractions after sorting by relative bid ask spread. Table 8 presents the results from these regressions. I regress the fraction of volume on CNs against the lagged fraction of volume on CNs. As before, there are statistically significant positive relationships. However, the magnitude and statistical significance of the positive relationship decreases as the relative bid ask spread increases. This provides support for the hypotheses that gaming concerns reduce the steadiness of volume on CNs when the relative bid ask spread is high. Higher CN volumes attract attention from gamers, which in turn make CNs less attractive for non-gamers, reducing the level of autocorrelation.

6 Robustness tests and discussion

The empirical results provide broad support for the empirical predictions from the model and hypotheses developed based on the economic intuition outlined in section 2. The results also shed light on to how traders decide to route trades between CNs and QEs. These results suggest that traders consider the strengths and weaknesses of CNs when deciding whether to trade there. The results are robust to a variety of specifications and have both a cross-sectional and a time series component to them. A large component of these results pertain to the tradeoff between transaction cost savings and gaming concerns, which I measure using the relative bid ask spread. In robustness tests, I test how the fraction of CN volume changes with other measure of QE liquidity, such as price impact measures and the effective spread. My results are directionally robust to using these measures in place of the relative bid ask spread. However, the noisiness of these alternative measures eliminates the statistical significance.

As CN volumes grow, there are endogeneity concerns regarding price discovery (e.g. bid ask spreads and price impact measures) and CN volumes. Gresse (2006) argues that CNs lead to tighter dealer market spreads. However, that study shows a correlation, rather than causation, between the spreads and CN volumes. Such a correlation could be the result of the bid ask spreads affecting the decision to use CNs, rather than the other way around. More generally, does the quality of price discovery for a particular stock in a particular month shift volume to CNs or does the shifting of volumes affect the price discovery? Given that CN volume fractions were about 1.5% of total volume in my sample, I adopt the view that the former is more likely than the latter. As CN volumes increase, further analysis into their impact on price discovery and bid ask spreads in dealer markets is warranted.

6.1 Using the fraction of volume on POSIT, Liquidnet and Pipeline as a proxy for the probability of using CNs

In the empirical tests of my hypotheses, I use the fraction of volume traded on POSIT, Liquidnet and Pipeline relative to total volume as a proxy for the probability of using CNs.

There are a number concerns with this analysis:

1. The fraction of volume on CNs may not truly represent the probability of using CNs. This is most easily seen in a hypothetical situation where a number of pairs of buyers and sellers are trying to transact with each other while choosing between a CN and a market exchange. They share a common probability of using the CN, p. The transaction is successful if they both choose the same location. If not, the trade fails. This fragmentation reduces the number of successful trades and fraction of *successful* trades on the CN is $\frac{p^2}{2p^2-2p+1} \leq p \ \forall \ p \in [0, 0.5]$. Thus, for low levels of p, the fraction of volume on CNs underestimates the probability of using them. This is exacerbated when there are multiple CN venues as the fragmentation further reduces success of CN trades. Thus, the fraction of trades on CNs is a lower bound on the probability of choosing CNs at the levels of execution observed in our sample.

2. According to Rosenblatt Securities' monthly dark liquidity tracker, *"Let There be Light,"*, POSIT, Liquidnet and Pipeline accounted for a total of 8% of CN volumes in September 2009, with the remaining volume accounted for by bank and exchange owned CNs. I use these three venues because it is difficult to distinguish agency trades from CN trades in the case of bank owned CNs, since they share the same market participant ID. It is possible that different factors may drive volume for bank owned CNs. For example, soft dollar payment considerations may play a role in routing stock to a bank's internal CN. However, at some level, the economic intuition outlined in this paper will come into play as the decision to choose between the CN or the QE is made. In the case of the previous example, the trader at an agency desk receiving a large client order may have to decide whether or not to try and match the order on the internal CN or on QEs. Thus, many findings from this study will still be applicable.

3. POSIT, Liquidnet and Pipeline historically had minimum volume thresholds, which might have affected the traders considering CNs as a trading venue. To mitigate this, Ready (2009) uses institutional investor holding data from 13-F filings to determine institutional volume change, uses that as the base of potential trades that might have gone to CNs and compares CN volumes to this number. I choose to use total volumes as a denominator. Aside from added simplicity, this provides a more accurate estimate

in situations where, for extremely liquid stocks, institutional investors may buy and sell the same stock in a give quarter.

Given the dearth of micro data on CN usage and confidentiality around algorithm and trader routing behavior, using the fraction of trades executed on CNs as a proxy of the probability of CN use is the best option available. As long as CNs and QEs are mainly substitutes for each other, rather than completely separate goods, the fraction of trades on CNs will be monotonically related to the probability of using CNs. As new data is collected on CN usage, and as CN growth and regulation affect the attractiveness of using CNs, the questions asked in this study may merit re-examination.

7 Conclusion

The rapid growth and proliferation of CNs suggest that they are here to stay. This study shows that CNs are garnering market share in situations where they have advantages relative to QEs. While the volume is currently small (1.5% in external crossing networks in my sample (2005-2006) and between 5 to 10% overall in 2009), it has been growing steadily. The increasing size of CNs is a double-edged sword for further growth, as it increases the probability of crosses, but may reduce the robustness of price discovery on QEs, leading to potential gaming. In addition to anecdotal evidence from the financial media and self regulation by CNs to deter such gaming, the decrease of CN volume on stocks with high relative bid ask spreads documented in this paper is consistent with the presence of gaming concerns. At the extreme, if all volume went through CNs, prices from QEs would be easily manipulated, and CN orders would be vulnerable to gamers; thus, such an equilibrium would be unstable, as orders would return to traditional QEs.

Crossing networks have also been facing greater regulatory scrutiny as regulators are trying to determine if the selective access to CNs has any detrimental effects on market participants who are unable to use them. In Release No. 34-60997, the SEC is "proposing to amend the regulatory requirements of the Securities Exchange Act of 1934 (Exchange Act) that apply to non-public trading interest in National Market System (NMS) stocks, including so-called dark pools of liquidity." One of the proposed measures is for "publicly disseminating consolidated trade data to require real-time disclosure of the identity of dark pools and other ATSs [alternative trading systems] on the reports of their executed trades."

While such measures might "promote the Exchange Act goals of transparency, fairness, and efficiency," they may also exacerbate extant gaming concerns, further inhibiting CN growth.

Finally, there is the increased fragmentation that Blume (2007) addresses. The space is growing rapidly and there are multiple players, both external and internal. The lack of centralized liquidity hampers the growth of CNs and consolidation is a natural next step. While this particular concern is not addressed in this paper, it will certainly have some bearing on the future growth trajectory of CNs. Between potential gaming, increased regulatory scrutiny and rapid, fragmented growth, crossing networks face an interesting future as they find their niche among the variety of trading venues available to market participants today.

References

Altunata, Serhan, Dmitry Rakhlin, and Hanri Waelbroeck, 2010, Adverse selection vs. opportunistic savings in Dark Aggregators, *Journal of Trading* 5.

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

Blume, Marshall E., 2007, Competition and Fragmentation in the Equity Markets: The Effect of Regulation NMS, *Univeristy of Pennsylvania, Working Paper*.

Conrad, Jennifer, Kevin M. Johnson, and Sunil Wahal, 2003, Institutional trading and alternative trading systems, *Journal of Financial Economics* 70, 99–134.

Domowitz, Ian, Ilya Finkelshteyn, and Henry Yegerman, 2009, Cul de Sacs and Highways: An Optical Tour of Dark Pool Trading Performance, *ITG Working Paper*.

Duffie, Darrell, 2010, Asset Price Dynamics with Slow-Moving Capital, *AFA Presidential Address*.

Gresse, Carole, 2006, The Effect of Crossing-Network Trading on Dealer Market's Bid-Ask Spreads, *European Financial Management* 12, 143–160.

Hendershott, Terrence, and Haim Mendelson, 2000, Crossing Networks and Dealer Markets: Competition and Performance, *Journal of Finance* 55, 2071–2116.

Kyle, Albert S, 1985, Continuous Auctions and Insider Trading, *Econometrica* 53, 1315–35.

Mittal, Hitesh, 2008, Are You Playing in a Toxic Dark Pool? A Guide to Preventing Information Leakage, *Journal of Trading* 3, 20–33.

Naes, Randi, and Bernt Arne Odegaard, 2006, Equity trading by institutional investors: To cross or not to cross?, *Journal of Financial Markets* 9, 79–99.

Ready, Mark J., 2009, Determinants of volume in dark pools, *Univeristy of Wisconsin, Working Paper*.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Financial Economics* 39.

Sofianos, George, and David Jeria, 2008, Quantifying the SIGMA X Crossing Benefit, *Journal of Trading* 3, 84–91.

Ye, Mao, 2009, A Glimpse into the Dark: Price Formation, Transaction Cost and Market Share of the Crossing Network, *Cornell University, Working Paper*.

A Model solution details

Conditional on the arbitrageur paying c and determining that there is an order to buy x shares on the CN, the arbitrageur will trade on the QE to manipulate the CN price before trading against the liquidity trader. The arbitrageur will buy until the best ask on the QE is p_1 and will put a smaller bid ϵ below p_1, effectively increasing the mid of the bid ask to p_1.

The optimal p_1 maximizes the following equation:

$$(1-(1-\alpha)\pi)x(p_1-p_0)-\frac{k(p_1-p_0)^2}{2} \tag{13}$$

Differentiating with respect to (p_1-p_0), the first order condition yields

$$p_{arb}-p_0=\frac{(1-(1-\alpha)\pi)x}{k} \tag{14}$$

where p_{arb} is the new best offer price after the arbitrageur has manipulated the QE to maximize profits from trading against the liquidity trader on the CN. The gain to the arbitrageur expressed as in terms of the r, the frequency with which the arbitrageur tries to game the system is

$$g_{arb}=r\big[-c+(1-q)0+q[(1-\pi)(p_{arb}-p_0)x+\pi(\alpha)(p_{arb}-p_0)x-\frac{(p_{arb}-p_0)^2k}{2}]\big] \tag{15}$$

where c is the cost paid to try to game, $(1-q)0$ reflects the probability that the liquidity trader does not use the CN, $(1-\pi)(p_{arb}-p_0)x$ reflects the gains from situations when the liquidity trader uses the CN and there are no other competing traders in the CN, $\pi(\alpha)(p_{arb}-p_0)x$ reflects the situations where the liquidity trader uses the CN but there is competition for trading in the CN and $\frac{(p_{arb}-p_0)^2k}{2}$ reflects the cost to move the QE mid price.

In the mixed strategy Nash equilibrium, the liquidity trader will choose q, the probability of using the CN, such that the arbitrageur will have no incentive to change r, or such that $-c+(1-q)0+q[(1-\pi)(p_{arb}-p_0)x+\pi(\alpha)(p_{arb}-p_0)x-\frac{(p_{arb}-p_0)^2k}{2}]=0$. Thus, the optimal q^* is:

$$q^*=\frac{2kc}{[1-(1-\alpha)\pi]^2x^2} \tag{16}$$

The loss to the liquidity trader, expressed as an expression in q and r, is as follows:

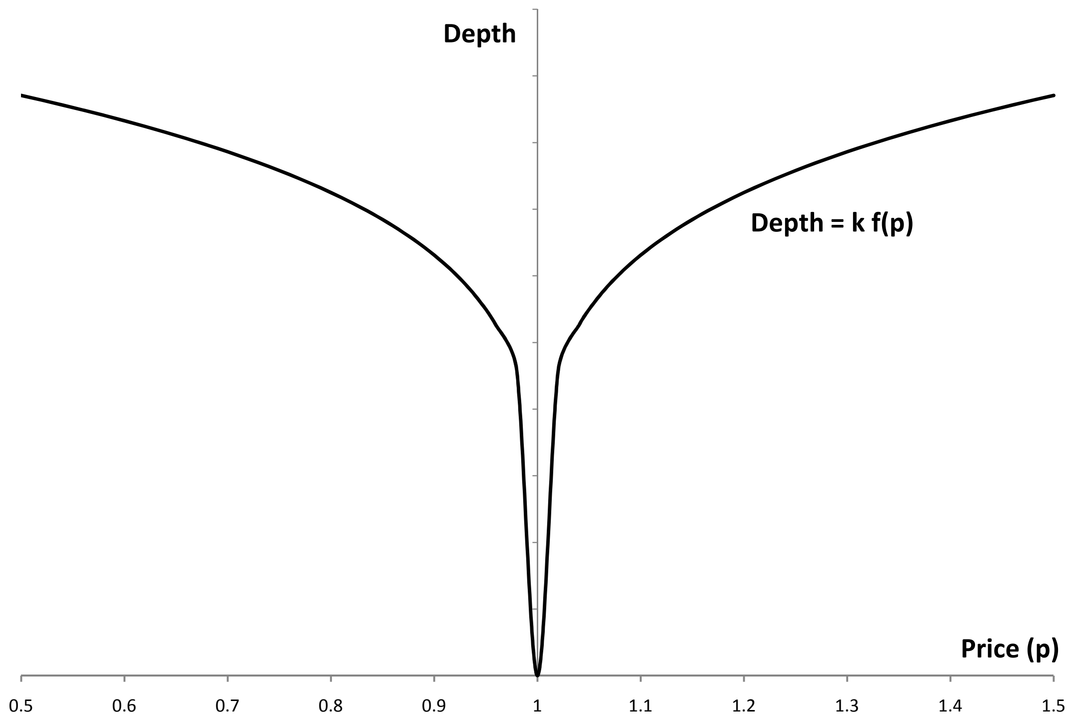
$$l_{LT} = (1-q)\frac{x^2}{2k} + q\big[\delta + (1-r)(\pi.0 + (1-\pi)\frac{x^2}{2k}) + r(\frac{[1-(1-\alpha)\pi]x^2}{k})\big] \tag{17}$$

where $(1-q)\frac{x^2}{2k}$ represents the loss from using the QE if the CN is not used and $q\big[\delta+(1-r)(\pi.0+(1-\pi)\frac{x^2}{2k})+r(\frac{[1-(1-\alpha)\pi]x^2}{k})\big]$ represents expected losses from the gamed CN price and potential eventual QE use if the liquidity does route the trade initially to the CN. The specific terms in the second equation are δ, the cost of trying to use the CN, $(1-r)(\pi.0+(1-\pi)\frac{x^2}{2k})$, the expected loss from reverting to the QE when the arbitrageur is not gaming given that there is only a π probability of execution on the CN, and $r(\frac{[1-(1-\alpha)\pi]x^2}{k})$, the losses to the trading on gamed prices on the CN when the arbitrageur is gaming.

In the mixed strategy Nash equilibrium, the arbitrageur will choose r, the probability of gaming, such that the liquidity trader will have no incentive to change q, or such that $\big[\delta+(1-r)(\pi.0+(1-\pi)\frac{x^2}{2k})+r(\frac{[1-(1-\alpha)\pi]x^2}{k})-\frac{x^2}{2k}\big] = 0$. Thus, the optimal r^* can be represented as:

$$r^* = \frac{\frac{x^2}{2k} - \delta}{\frac{x^2}{2k}(\pi - 1) + \frac{(1-(1-\alpha)\pi)x^2}{k}} \tag{18}$$

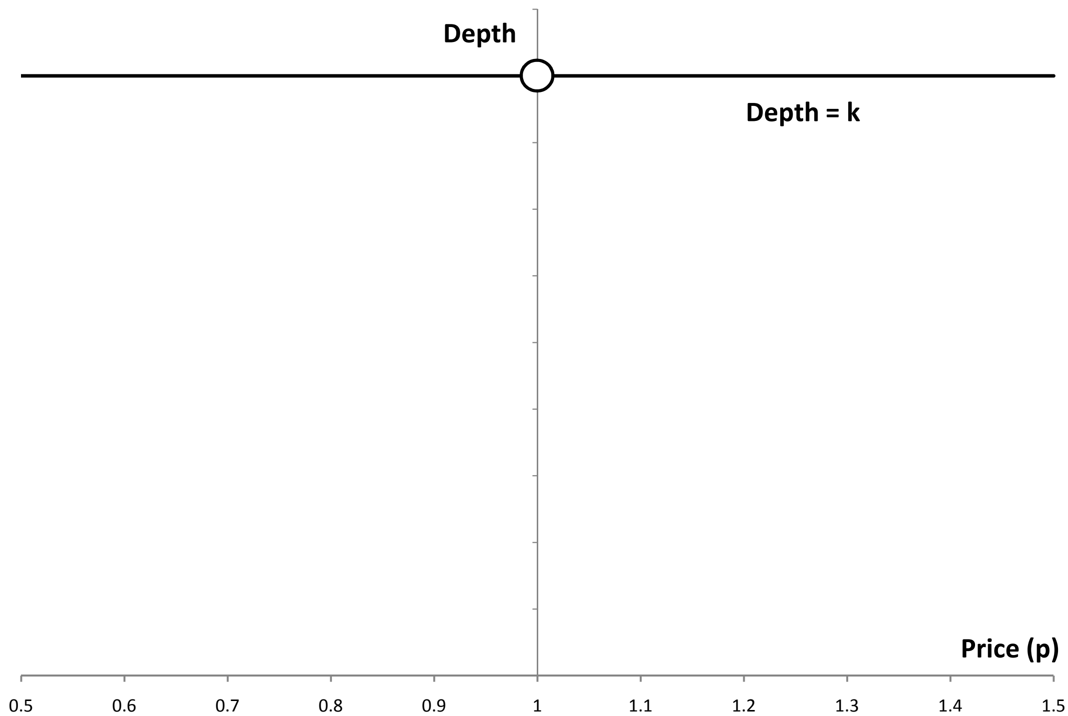
Figure 1: Representative depth chart for a quoting exchange



$p_0 = 1$; Prices below 1 correspond to bids and prices above 1 correspond to asks

This graph presents a representative depth chart. The market maker is willing to transact a increasing amount as the price is bid away from $p_0 = 1$.

Figure 2: Simplified depth chart for a quoting exchange

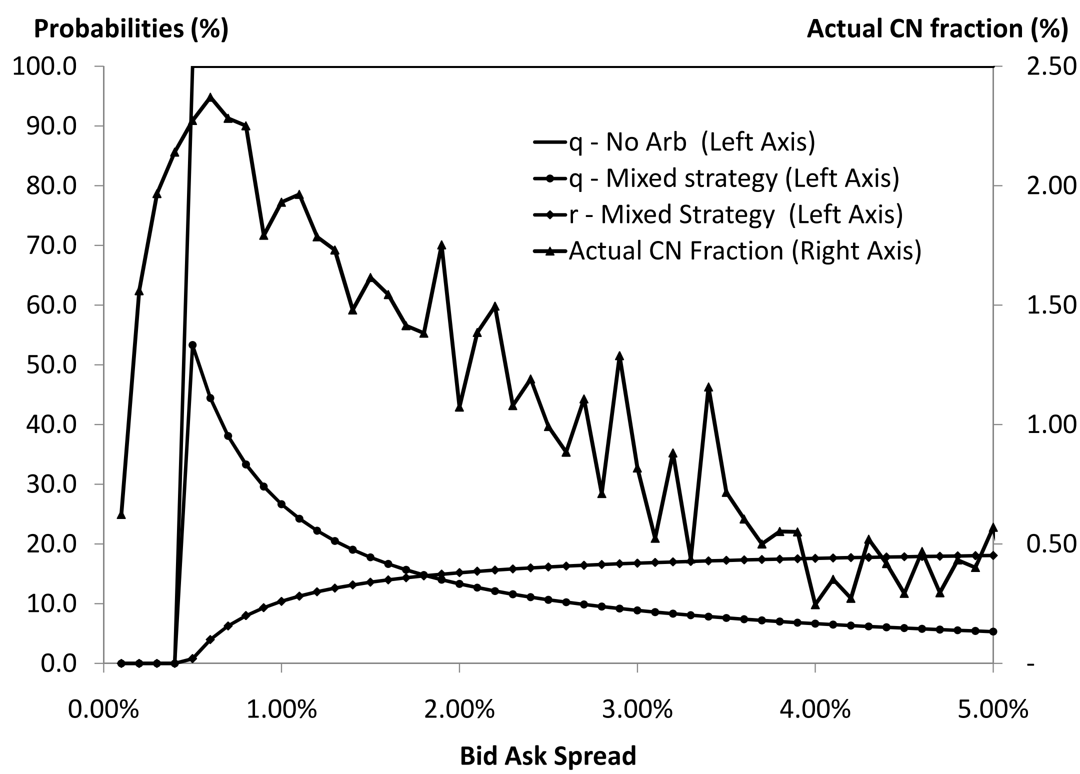


$p_0 = 1$; Prices below 1 correspond to bids and prices above 1 correspond to asks

This graph presents the simplified depth chart used to solve the model in section 4. The market is willing to transact a flat amount, k, at each price.

Figure 3: Optimal probabilities for the liquidity trader using the CN (q^*), the arbitrageur attempting to game the CN (r^*) and the actual fraction of volume on the CN

This graph presents solutions to the liquidity trader and arbitrageur's problems. Specifically it presents (1) the liquidity trader's probability of using the CN if no arbitrageur is present (solution to equation 4) and (2) the liquidity trader and arbitrageur's probabilities of using the CN and trying to game the CN under the simplified quoting exchange depth (equations 11 and 12). The graph also presents the average actual fraction of trades on CNs, sorted by relative bid ask spread.



All the model probabilities are generated using the simplified bid ask depth chart from Figure 2 and the following parameters: $\delta = 3.00$, $x = 1000$ and relative bid ask spreads are calculated using prices to trade 100 shares $p_0 = 1$ to generate parameter k. Additionally, mixed strategy probabilities use the following parameters: $c = 5$, $\pi = 0.25$, $\alpha = 0.5$.

Table 1: Summary Statistics

This table presents the summary statistics over the June 2005 to June 2006 period for the variables used. Average volume traded on the crossing networks (CNs) is defined as volume on the three external crossing networks, ITGI Posit, Liquidnet and Pipeline divided by total volume. The average bid ask spreads for each stock are calculated at the daily level from the TAQ database. The relative spread is defined as the difference between the bid and the ask divided by the price of the stock. The average log percentage spread is average of the of log of the percentage bid ask spreads defined above. Absolute monthly return is the absolute with-dividend monthly return as reported by CRSP. Order flow spread contribution is the fraction of the bid ask spread accounted for by the Roll (1984) estimate of the order flow component of bid ask spread. Monthly turnover is the number of shares divided by total shares outstanding in percentage terms. Totaly monthly volume is the number of shares traded in a given month for a given stock in millions of shares. The market capitalization is the number of shares outstanding multiplied by the average share price over the month expressed in millions of dollars.

Variable	**Mean**	**Std. Dev.**	**N**
Fraction of volume on CNs (%)	1.405	3.337	33631
ITGI volume fraction	0.727	2.257	33631
LQNT volume fraction	0.623	2.082	33631
BLOK volume fraction	0.055	0.546	33631
Average relative BA spread (%)	2.325	2.725	33631
Log average relative BA spread	0.068	0.544	33631
Log relative spread squared	0.301	0.31	33631
Absolute monthly return (%)	8.912	10.647	33206
Order flow spread contrib.	0.003	0.016	33631
Monthly turnover (%)	18.375	50.406	33631
Total monthly volume (MM)	12.734	75.357	33631
Share price	18.246	16.913	33630

Table 2: Cross-correlation table

This table presents the cross correlation between the variables used in my study. Average volume traded on the CNs is defined as volume on the three external crossing networks, ITGI Posit, Liquidnet and Pipeline divided by total volume. The average bid ask spreads for each stock are calculated at the daily level from the TAQ database. The relative spread is defined as the difference between the bid and the ask divided by the price of the stock. Market capitalization is calculated as the share price multiplied by the number of shares outstanding. The market impact is the price impact measure as defined in Amihud (2002) but calculated using 5 minute intervals. Absolute monthly return is the absolute with-dividend monthly return as reported by CRSP. Order flow spread contribution is the fraction of the bid ask spread accounted for by the Roll (1984) estimate of the order flow component of bid ask spread. Monthly turnover is number of shares divided by total shares outstanding in percentage terms. Totaly monthly volume is the number of shares traded in a given month for a given stock in millions of shares. The market capitalization is the number of shares outstanding multiplied by the average share price over the month expressed in millions of dollars.

Variables	CN Volume Fraction	Avg. rel. spread	Market capitalization	Market impact	Abs. return	Order flow fraction	Monthly turnover	Total monthly volume	Share price
CN Volume Fraction	1.000								
Avg. rel. spread	-0.182	1.000							
Market capitalization	-0.013	-0.110	1.000						
Market impact	-0.048	0.225	-0.027	1.000					
Abs. return	-0.047	0.044	-0.031	-0.015	1.000				
Order flow fraction	-0.008	-0.024	0.007	0.026	-0.021	1.000			
Monthly turnover	-0.026	-0.144	0.028	-0.042	0.299	0.004	1.000		
Total monthly volume	-0.026	-0.120	0.697	-0.031	0.030	0.002	0.182	1.000	
Share price	0.088	-0.369	0.214	-0.045	-0.146	0.083	0.073	0.071	1.000

Table 3: Univariate analysis of the impact of the relative bid ask spread on fraction of volume on the CN

This table sorts the observations by relative bid ask spread (bid ask spread divided by the stock price) into deciles. The first column presents the average relative bid ask spread for each decile, in percentage terms. The second column presents the average fraction of volume trading through crossing networks, in percentage terms. The third column presents the standard error around the estimate of the fraction of volume trading on the crossing networks.

Relative BA spread decile	Mean		
	Average Relative BA Spread %	Average CN Volume %	Std Error %
Lowest	0.15	1.48	0.03
2nd	0.28	2.07	0.04
3rd	0.43	2.23	0.06
4th	0.63	2.33	0.06
5th	0.97	1.88	0.06
6th	1.49	1.55	0.07
7th	2.27	1.17	0.07
8th	3.37	0.70	0.06
9th	4.95	0.39	0.05
Highest	8.71	0.26	0.05
Total	2.33	1.41	0.05

Table 4: Impact of percentage spread and last month's fraction of volume traded on the CN on fraction of volume on CN this month

This table sorts the observations into deciles by relative bid ask spread (bid ask spread divided by the stock price) and the previous month's crossing network volume fraction. The rows are in increasing order of relative bid ask spread and the columns are in increasing order of last month's crossing network volume fraction. The interior cells present the average crossing network volume fraction (in percentage terms) for the current month.

	Deciles of Lagged_CN_fraction							
Relative BA spread Decile	Lowest	5th	6th	7th	8th	9th	Highest	Total
Lowest	0.2	0.5	0.9	1.3	1.8	2.2	3.0	1.5
2nd	0.3	0.8	1.3	1.8	2.0	2.7	3.3	2.1
3rd	0.3	0.8	1.4	1.9	2.3	2.8	4.1	2.2
4th	0.3	1.0	1.4	2.0	2.7	2.9	4.4	2.3
5th	0.7	1.1	1.5	1.8	2.0	3.1	3.8	1.9
6th	0.6	1.3	1.2	2.2	1.8	2.8	4.2	1.6
7th	0.5	1.0	1.4	1.7	3.2	3.1	4.7	1.2
8th	0.4	0.3	1.6	1.4	1.4	2.1	5.0	0.7
9th	0.2	0.3	0.6	1.0	0.4	2.4	3.7	0.4
Highest	0.2	0.6	0.2	0.3	0.6	5.8	4.3	0.2
Total	0.3	0.9	1.3	1.8	2.1	2.7	4.0	1.4

Table 5: Impact of percentage spread and total monthly volume on fraction of volume on CN

This table sorts the observations into deciles by relative bid ask spread (bid ask spread divided by the stock price) and the total monthly volume. The rows are in increasing order of relative bid ask spread and the columns are in increasing order of total monthly volume. The interior cells present the average crossing network volume fraction (in percentage terms).

	Deciles of Total_Monthly_Volume										
Relative BA spread Decile	Lowest	2nd	3rd	4th	5th	6th	7th	8th	9th	Highest	Total
Lowest	9.3	0.0	1.4	0.0	0.3	1.1	2.0	2.1	1.9	1.1	1.5
2nd	0.0	0.0	0.0	0.8	1.0	2.3	2.6	2.4	2.2	1.2	2.1
3rd	0.0	0.0	6.0	3.5	2.6	2.9	2.7	2.1	1.7	0.8	2.2
4th	0.0	0.0	1.3	2.4	3.0	2.8	2.5	2.0	1.3	0.5	2.3
5th	0.0	0.6	1.8	2.4	2.4	2.1	1.6	1.2	0.9	0.4	1.9
6th	1.9	0.9	1.6	2.5	1.8	1.5	1.2	0.6	0.5	0.2	1.6
7th	0.5	1.1	2.0	1.8	0.9	0.7	0.4	0.2	0.2	0.1	1.2
8th	0.4	1.0	0.9	0.9	0.8	0.3	0.1	0.6	0.0	0.1	0.7
9th	0.4	0.4	0.5	0.3	0.5	0.1	0.0	0.1	0.0	0.0	0.4
Highest	0.3	0.4	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.3
Total	0.4	0.7	1.1	1.6	1.7	1.9	2.0	1.8	1.7	1.0	1.4

Table 6: Determinants of fraction of volume on CNs

This table presents regression coefficients for the following regression specification: $f_t = \alpha + \sum_i \beta_i g_{i,t} + \epsilon$ where f_t is the fraction of volume on CNs in a given month and g_i are the explanatory variables. The first column presents the log relative bid ask spread and log spread squared as standalone regressors. The second column presents all explanatory variables and controls. The third and fourth columns present this same specification on split samples and without the squared log relative spread. The fifth column presents a stock fixed effects specification of the regression in the second column. The regressors are log average relative bid ask spread, the log relative spread squared, one month lagged crossing network volume fraction, the absolute monthly return, the order flow spread contribution measure (defined as the order flow spread calculated as per Roll (1984)) divided by the total spread, monthly turnover in percentage terms, the total monthly volume in millions of shares and the share price. Time dummies are included but unreported. The t-statistics are computed using standard errors that adjust for heteroskedasticity. *, ** and *** denote statistical significance at the 5%, 1% and 0.1% levels, respectively.

	Base	Controls	Rel BA < 0.63	Rel BA >= 0.63	Fixed Effect
Log average relative BA spread	-0.930***	-0.771***	0.932***	-1.070***	0.489**
	(-36.327)	(-15.956)	(8.923)	(-8.623)	(3.090)
Log relative spread squared	-1.255***	-0.869***			-0.379***
	(-27.306)	(-15.891)			(-3.301)
Lagged CN volume fraction (%)		0.260***	0.349***	0.228***	-0.004
		(11.372)	(13.377)	(8.986)	(-0.172)
Absolute monthly return (%)		-0.006***	-0.007*	-0.005**	-0.003**
		(-3.865)	(-2.223)	(-3.273)	(-2.754)
Order flow spread contrib.		-3.126***	-1.558***	-4.668**	-1.475
		(-3.859)	(-3.519)	(-3.118)	(-1.933)
Monthly turnover (%)		-0.003***	-0.005**	-0.001**	0.000
		(-4.028)	(-2.865)	(-2.637)	(0.295)
Market capitalization (MM)		-0.000**	-0.000*	0.001***	0.000
		(-2.880)	(-2.208)	(4.217)	(0.636)
Total monthly volume (MM)		-0.001*	-0.000	-0.010***	0.000
		(-2.493)	(-0.486)	(-5.771)	(1.331)
Share price		0.003*	0.000	0.002	-0.012**
		(2.379)	(0.391)	(0.729)	(-2.932)
R-squared	0.042	0.131	0.171	0.110	0.002
N	33631	30581	10823	19758	30581

Table 7: Order flow spread contribution (OFSC) and interactions

This table presents regression coefficients for the following regression specification: $f_t = \alpha + \sum_i \beta_i g_{i,t} + \epsilon$ where f_t is the fraction of volume on CNs in a given month and g_i are the explanatory variables. The first column presents the control specification from column two of Table 6. The second and third columns presents interactions of the order flows spread contribution (OFSC) with monthly volume and turnover respectively. The regressors are log average relative bid ask spread, the log relative spread squared, one month lagged crossing network volume fraction, the absolute monthly return, the order flow spread contribution (OFSC) measure (defined as the order flow spread calculated as per Roll (1984)) divided by the total spread, monthly turnover in percentage terms, the total monthly volume in millions of shares and the share price. Time dummies are included but unreported. The t-statistics are computed using standard errors that adjust for heteroskedasticity. *, ** and *** denote statistical significance at the 5%, 1% and 0.1% levels, respectively.

	Base	Interaction 1	Interaction 2
Order flow spread contrib.	-3.126***	-3.209***	-3.754**
	(-3.859)	(-3.847)	(-2.822)
OFSC × Volume		0.009	
		(1.647)	
OFSC × Turnover			0.024
			(1.056)
Total monthly volume (MM)	-0.001*	-0.001**	-0.001*
	(-2.493)	(-2.638)	(-2.515)
Monthly turnover (%)	-0.003***	-0.003***	-0.003***
	(-4.028)	(-4.027)	(-4.018)
Log average relative BA spread	-0.771***	-0.771***	-0.771***
	(-15.956)	(-15.955)	(-15.945)
R-squared	0.131	0.131	0.131
N	30581	30581	30581

Table 8: Autocorrelation of CN volume sorted by relative bid ask spread decile

This table presents regression coefficients for the following specification: $f_t = \alpha + \beta f_{t-1} + \epsilon$ where f_t is the fraction of volume on the CN. The sample is sorted into deciles by the relative bid ask spread and the regression is performed separately for each decile. The t-statistics are computed using standard errors that adjust for heteroskedasticity. *, ** and *** denote statistical significance at the 5%, 1% and 0.1% levels, respectively.

	Dec. 1	Dec. 2	Dec. 3	Dec. 4	Dec. 5	Dec. 6	Dec. 7	Dec. 8	Dec. 9	Dec. 10
Lagged CN volume fraction (%)	0.342***	0.330***	0.371***	0.284***	0.233***	0.276***	0.274***	0.185***	0.197**	0.209**
	(12.662)	(14.307)	(7.863)	(4.788)	(6.102)	(6.635)	(3.913)	(3.735)	(2.770)	(3.017)
Constant Supressed										
R-squared	0.130	0.107	0.158	0.120	0.072	0.066	0.096	0.033	0.036	0.116
N	3105	3116	3097	3100	3026	3060	3053	3068	3098	3038