Determinants of volume in dark pools

Mark J. Ready*

University of Wisconsin-Madison, Madison, WI, 53706, USA,

October 8, 2010

Abstract

I investigate determinants of trading volume for NASDAQ stocks in three of the "dark pools" that cater to institutional traders: Liquidnet, POSIT and Pipeline. I use both panel data regressions and a simulated method of moments approach to investigate institutional traders' propensity to route orders to these dark pools. The results suggest that dark pool usage is lower for stocks with the lowest spreads per share, which is consistent with trader routing of these stocks to other venues in order to satisfy soft-dollar agreements.

JEL Classifications: G12, G14

Keywords: Microstructure; Liquidity; Trading Volume, Simulated Method of Moments

---

**Contact information: Tel.: 608-262-5226; E-mail address: mjready@facstaff.wisc.edu*
I would like to thank Charles Jones, Mao Ye, and participants at the 2010 AFA meetings and in research seminars at Syracuse University and the University of Pittsburgh.

1. Introduction

Institutions are always looking for ways to reduce trading costs. In the past, institutional traders largely relied on one or more intermediaries, including market makers and sell-side brokers. Of course, intermediaries must be compensated for providing their services, and this compensation is reflected in some combination of commissions and market impact cost. Along with these direct costs, institutions worry that the intermediaries may not carefully guard the information associated with their orders. If this information leaks out, then opportunistic traders will front-run the institutional order, which means that they trade in advance of the institutional trade and in the same direction.[1] Given these potential costs, it is not surprising that institutions are interested in finding ways to bypass market intermediaries and trade directly with one another.[2] Several market centers have evolved to facilitate this direct trading between institutions, and in this paper I investigate three: Liquidnet, POSIT and Pipeline.

These three market centers are sometimes called "Dark Pools," because traders do not publicly reveal their orders in advance. Liquidnet is connected directly to the order management systems of participating institutions, and the Liquidnet software continually scans them looking for potential match. When two institutions happen to have orders for the same security but in the opposite direction, the Liquidnet system sends messages to each institution and they begin a negotiation to trade. POSIT is owned by ITG, and it began as a periodic crossing network.

---

[1] Trading in dark pools is just one way that institutions attempt to hide their trading interest. They also split their orders (see Barclay and Warner (1993)) or post hidden orders on ECNs (see Hasbrouck and Saar (2002)).

[2] Schwarz and Steil (2002) discuss the potential information leakage when using an agency broker, and they also discuss the idea that soft-dollar agreements can induce institutions to use brokers in spite of this problem. I consider soft-dollar agreements as one possible determinant of which orders are sent to dark pools.

Institutions place orders into the POSIT Match[SM] system in advance of each scheduled crossing session, and only trade if there happens to be another institution that submits an order for the same stock but in the opposite direction. In August of 2005, POSIT introduced POSIT Alert[SM], which works with buy-side order management systems in a way that is very similar to Liquidnet. Pipeline accepts firm orders to trade blocks of 10,000 shares or more. These orders are hidden, except that the system uses color coding to indicate stocks for which there is "serious liquidity is available."[3] If an order is entered into Pipeline and there is already an order in the system on the other side, then the system crosses the two at a price inside of the current quotes.

In addition to having no explicit advertisement of trading interest, all three dark pools try to control information leakage by excluding institutions who might try to exploit the information contained in the order flow and by eliminating or restricting smaller orders that might be used to "ping" the system to discover larger trading interest. All three systems rely on quotes from the rest of the market to determine execution prices. For the most part, these systems exclude sell-side firms, although Pipeline is connected to Lava Trading, which is owned by Citigroup and is open to sell-side firms. Like the rest of the trading landscape, dark pools continue to evolve. In the next section, I discuss dark pools in more detail and I also briefly describe the proposed SEC rules regarding pre- and post-trade transparency.

Given the apparent benefits from using Liquidnet, POSIT or Pipeline, it is reasonable to ask why they are not used exclusively for all institutional trading. The first obvious answer is that for trading to occur, the counterparties must enter their orders in the system at the same time, and when both buyers and sellers are present, the maximum volume is the smaller of the total buying and the total selling interest. The second answer is the focus of this paper: sometimes it

---

[3] See http://www.pipelinetrading.com/PLSolutions/PipelineTrading.aspx.

may not be optimal to use the system. Use of these systems generally entails waiting, at least if the trader wants to get a substantial probability of an execution. This waiting can be costly if the price moves unfavorably.[4] Thus, depending on the characteristics of the stock or market conditions, traders may sometimes prefer other strategies to get faster executions. Another potential problem with using these systems is gaming by other traders. In spite of the fact that all three of the systems employ safeguards against this type of behavior, there is still some feeling among traders that it can be a problem.[5] A final potential problem with these venues is that although any executions (if they occur) tend to be at favorable prices and fees are low, they do not provide any soft-dollar benefit. Accordingly, traders who have volume quotas with soft-dollar brokers may route orders elsewhere.

To investigate the factors driving the traders' choice to use Liquidnet, POSIT or Pipeline, I use a sample of quarterly volumes by stock for each of these venues. Obviously, the most important determinant of volume in any one of these venues will be the level of institutional trading during that period. To measure institutional trading, I use the changes in quarterly institutional holdings. I develop a model that includes random arrivals of trading interest within each quarter, and assumes that trader routing decisions are based on characteristics of the stock,

---

[4] If informed traders use the dark pools, then an uninformed trader faces an adverse selection problem. If her order happens to be in the same direction as the order from the informed trader, then her order will compete with (be "crowded out" by) the informed order, and it will be more likely that she will have to subsequently submit her order to a dealer. At the time she trades with the dealer, the information may be public. Even if the information has not been announced, the dealer will charge a wider spread knowing that her order may indirectly reflect the presence of information. See Hendershott and Mendelson (2000) and Ye (2009) for formal models of competition between a crossing network and a dealer market.

[5] See "Trading in a Dark Pool? Watch for Sharks" Wall Street Journal, August 18, 2008.

the order size, and the organization making the trade. I use the model to motivate panel data regressions that relate dark pool shares of volume to stock characteristics and institutional trading activity, and I also estimate the underlying model parameters using simulated method of moments.

I show that these venues appear to attract the lowest share of institutional trades in stocks with the highest consolidated volumes. I test three hypotheses that could explain this pattern. The results suggest that dark pool usage is lower for stocks with lower spreads per share, which is consistent with trader routing of these stocks to other venues in order to satisfy soft dollar agreements. Unrelated to the hypotheses, the results suggest that institutions with higher turnover are less likely to route their orders to the dark pools.

The next section describes dark pools in more detail and briefly describes the current and pending regulations. Section 3 describes the data and presents some summary statistics. In particular, Section 3 shows that the dark pools' share of consolidated volume is lowest in the stocks with highest trading volume. Section 4 introduces hypotheses to explain the negative relation between dark pool shares and consolidated volume, and develops additional predictions based on the hypotheses. Section 5 presents the structural model, and the empirical results are contained in Sections 6 and 7. Section 8 considers ECN usage as a potential alternative to dark pools, and Section 9 concludes.

2. Dark pools and other institutional trading venues

Liquidnet, POSIT and Pipeline are the three oldest dark pools that cater primarily to the buy-side, but there are many other trading mechanisms that cater to these same institutions and do not publicly display trading interest. Of course, to get a trade done it is necessary to reveal something to someone, so in this context "public display" means market-wide display of the order. Using a standard limit order is an example of market-wide display. If the limit order is aggressively priced, it will be included in the National Best Bid and Offer (NBBO).[6] Even if it is not aggressively priced, many limit order markets publicly reveal all of their limit orders.

Buy-side traders look for ways to reveal their orders only to potential counterparties, and keep them hidden from other market participants. Exchange floor brokers are perhaps the oldest example of this type of trading mechanism—they represent large institutional orders and look for other floor brokers who may be representing an order in the same stock but the opposite direction. It appears that the traditional floor brokers have lost considerable market share in recent years, so much so that during 2006 and 2007 the New York Stock Exchange closed three of its five trading rooms.[7]

Large sell-side firms also help to facilitate institutional trading without publicly disclosing trading interest. When a buy-side fund gives a sell-side broker an order to "work" the sell-side broker contacts other buy-side firms looking for counterparties. These contacts can be electronic "indications of interest" sent to buy-side trading desks, and they can be telephone calls. Institutions also submit "iceberg orders" to ECN's. Iceberg orders are orders where most of the shares available are hidden.

Dark pools exhibit many of the features of exchange floor brokers and sell-side brokers. Their current structures were made possible by dramatic improvements in speed of computer

---

[6] The SEC's proposed "Regulation of Non-Public Trading Interest," includes the following discussion: "The term 'dark pool' is not used in the Exchange Act or Commission rules. For purposes of this release, the term refers to ATSs that do not publicly display quotations in the consolidated quotation data."

[7] See "NYSE CEO sees 15-20 pct decline in floor brokers," which is available at http://www.reuters.com/article/idUSN0529031620080505

processing and electronic communications in the late 1990's. Dark pools are registered with the SEC as Alternative Trading Systems (ATSs), and they are governed by SEC's Regulation ATS. This regulation requires all ATS trades to be included in the consolidated trade data (trades from all of the ATSs are coded as "OTC"), and it requires all ATSs that execute more than 5% of the volume in a particular stock to include their quotes in the consolidated quote stream.

As it turns out, the volume of each individual dark pool rarely exceeds the 5% threshold, and the order representation inside of most ATS systems would not meet the SEC's current definition of a quote, so the NBBO generally does not include any information from dark pools. On November 13, 2009 the SEC proposed new rules that would lower the threshold for quote reporting from 5% to 0.25% of consolidated volume and would change the definition of what constitutes a quote to include many of the indications of interests currently used by dark pools. As proposed, the rules would exempt indications of interest where the order value exceeds $200,000. The proposed rules would also require disclosure of the identity of the executing ATS for trades smaller than $200,000.

Because of the dissemination of indications of interest, some market observers have remarked that rather than being "dark," these markets are actually "partially lit." The SEC's concern is that the indications of interest in the dark pools represent selective disclosure of quotes, which could disadvantage investors who do not have access to the dark pools. Many of the ATSs have objected to the proposed requirement to disclose their indications of interest in the public quote. Some have suggested that the requirement would push them to go completely dark.[8]

[8] When Regulation ATS was passed in September 2002, the Island ECN had more than 5% of the volume in three Exchange Traded Funds. Rather than including its quotes in these securities in the NBBO, Island chose to "go dark"

According to Advanced Trading[9], as of February 2010 there were 29 separate dark pools either operating or in the planning stages. Of these, NYFIX Millennium and Instinet Crossing were operating during my entire 2005-2007 sample period. Both of these allow sell-side brokers to participate. NYFIX Millennium began with trading of NYSE-listed securities and they still comprise the majority of the NYFIX Millennium volume. The volumes for Instinet Crossing are included in the total volumes for Instinet, and they are a small fraction of the total. Most of the other dark pools in Advanced Trading's list are either sponsored by sell-side firms or invite their participation.

The SEC's November 2009 proposing release reports that in dark pools executed approximately 7.2% of the total share volume in NMS stocks during the second quarter of 2009, with no single dark pool executing more than 1.3%. The SEC's data are compiled from quarterly filings of form ATS-R, which are not available to the public. Total consolidated NMS daily volume is approximately 8 billion shares, so the SEC's statistics imply total daily dark pool volume of about 570 million shares, and they imply that no single dark pool had daily volumes much above 100 million shares. For 2009, Liquidnet reported[10] average daily volume of 57.2

---

and not allow any displayed orders in these securities. Island's motivation was not to avoid public display--in fact, the full, real-time limit book was already available on their web site. Rather, including their quotes in the NBBO would have required Island to link to the Intermarket Trading System (ITS), and Island felt that the latency in ITS would have cause an unacceptable degradation in the performance of their trading system. See Hendershott and Jones (2005).

[9] See http://advancedtrading.thewallstreetwiki.com/directories/directory-dark-pools.php

[10] See http://www.liquidnet.com/docs/liquidnetBuzz/pressReleases/2010/20100110_liquidnet_announces_december_volumes.pdf.

million shares, so it is still one of the more important dark pools. POSIT and Pipeline do not publicly report their volumes.

3. Data

There are four different sources of data: market center volumes were downloaded from the NASDAQtrader.com web site, institutional holdings are from the Thompson 13F database, stock characteristics are from the CRSP Stock Files, and measures of transaction cost are calculated from TAQ data.

The sample covers the nine calendar quarters beginning with July 2005 and ending with September 2007. Trading is inferred from changes in quarterly institutional holdings, so it is important to avoid stocks with ticker symbol changes or stock splits over the quarter. Also, a change in symbol near the start or end of the quarter has the potential to yield inconsistencies in the report. Accordingly, using the CRSP Stock Files, for each quarter I identify all NASDAQ common stocks that meet the following criteria for the full five-month period extending from one month prior to the start of the quarter through one month after the end of the quarter.

- No changes in ticker symbol, CUSIP, or CRSP share code
- No stock splits
- Non-zero consolidated volume each month
- Minimum month-end share price of $2
- Minimum month-end market capitalization of $100 million
- No other class of common stock issued by the firm

In addition to the above, I also eliminate about 0.5% of the potential observations because the ticker symbol from CRSP does not match the ticker symbol in the volume data, which are described in subsection 2.2. These screens produce a sample of approximately 1,650 stocks each quarter. I provide summary statistics for this full sample in the next two subsections, but the panel data regressions in Section 6 and the Simulated Method of Moments tests in Section 7 restrict the sample to the 500 stocks with the largest consolidated trading volume each quarter. The restricted sample captures about 89% of the consolidated volume from the full sample.

3.1. Market Center Volumes

NASDAQ-listed stocks are traded in many venues and by many participants inside of the NASDAQ market, and also by several other stock exchanges. NASDAQ tracks trading volumes for each of its internal participants using a four-letter Market Participant ID code (MPID). Total monthly consolidated volumes by stock come from the CRSP monthly stock files. Monthly volumes by stock and by MPID were obtained from two different sources on the NASDAQtrader.com web site. At the time the data were gathered, the FTP site ftp://ftp.NASDAQtrader.com/monthlysharevolume/ had 13 monthly files spanning the period from June 2005 through June 2006 that include volumes for all MPIDs. The remaining data were downloaded during late 2007 using the monthly volume reports that were available at that time at www.NASDAQtrader .com, but have recently been replaced by a new dataset. These data were downloaded by month, MPID, and ticker symbol for the 18 months from April 2006 through September 2007.

It is not uncommon for there to be zero dark pool volume for a particular stock in a given month. To confirm that this is not a result of a symbol mismatch between CRSP and the volume data source, I look for volume in that symbol from other MPIDs. These other MPIDs are all available in the early part of the sample. For the latter part of the sample, in addition to Liquidnet, POSIT and Pipeline, volumes were downloaded for several sell-side brokers and a few of the largest Electronic Crossing Networks (ECNs). If no volumes are found in any of the

MPIDs, then I assume it is a symbol mismatch and I discard the stock for that quarter. As discussed at the beginning of this section, that step eliminates a little under 0.5% of the observations.

Table 1 reports the volumes by MPIDs for the second quarter of 2006, for Microsoft (the highest volume NASDAQ common stock that quarter), American Commercial Lines (an inland waterway shipper and barge manufacturer), and for the total across all NASDAQ stocks in the sample that quarter. American Commercial Lines is an example of a less-well known firm, but the majority of the stocks in the sample have even lower volumes. The second quarter 2006 consolidated volume for American Commercial Lines ranks 467^{th} among the 1,688 NASDAQ stocks in the sample for that quarter, so it is included in the restricted sample that is used for the tests in Sections 6 and 7.

As shown in Table 1, Liquidnet (MPID='LQNT'), POSIT (MPID='ITGI') and Pipeline (MPID='BLOK') accounted for 0.3% of Microsoft's consolidated volume, 5.5% of American Commercial Lines' consolidated volume, and 1.0% of consolidated volume for the full sample in the second quarter of 2006. As shown in figure 1, the fraction of consolidated volume captured by the dark pools remains reasonably constant across the early part of the sample, but increases somewhat starting in the latter part of 2006, primarily due to an increase in the share of ITGI.

American Commercial Lines was chosen for Table 1 in part due to the relatively high fraction of total volume executed in the dark pools, but also to bring out two consistent features of the sample: first, the dark pool volume shares are higher in the NASDAQ stocks with consolidated volumes in the middle deciles of the sample, and second, that dark pool shares are higher when institutional volumes are higher. Figure 2 illustrates the first of these features of the sample, by plotting the average dark pool shares by decile of consolidated volume for the third quarter of 2005, the second quarter of 2006, and the third quarter of 2007. Note that roughly speaking, the restricted sample includes the 8th, 9th and 10th deciles shown in the figure.

The pattern in figure 2 is only part of the story for American Commercial Lines, because its dark pool share of 5.5% (shown in Table 1) is even higher than the average for the middle deciles shown in figure 2. It turns out that institutional trading for American Commercial Lines was relatively high in the second quarter of 2006. The relation between dark pool shares and institutional volumes will be discussed in more detail in the next subsection.

Along with the three dark pools, Table 1 lists the 17 MPIDs with highest total share volume across the 1,688 stocks in the second quarter 2006 sample. The two largest of these, Island/Instinet (INET) and BRUT were ECNs. As it turned out, both were acquired by NASDAQ in 2007, and their hardware and software were used to form the basis for some of NASDAQ's new trading systems. The next 4 market centers shown in Table 1 (GSCO, UBSS, MSCO, and SBSH) are sell-side brokers that have large trading desks that cater primarily to institutional clients. Knight (NITE) is the largest of the wholesalers, which are market centers that cater to orders from retail brokerage firms that don't have their own trading desks. Although it is generally categorized as a wholesaler, Knight also serves institutional clients. Of the remaining market centers, five (BOFA, LEHM, FBCO, MLCO, and DBAB) are also sell side brokers. Citadel (CDRG) may be better known for its hedge funds, but it is fast becoming a full-service sell-side firm. Automated Trading Desk (AUTO) and Bloomberg Tradebook (BTRD) are ECN's, although Automated Trading Desk also has a market making arm that provides liquidity (posts limit orders) in their ECN. The other two MPID's in the list (NFSC and ETRD) are the trading operations for two large retail brokerages. NFSC is owned by Fidelity and ETRD is owned by E*Trade. These brokerages "internalize" (trade against for their own account) the orders from their retail brokerage customers.

There is a three-month overlap between the two volume data sources, allowing a check of consistency between the two. For stocks that remain on NASDAQ, the two sources have identical data. There is a minor difference in breadth of coverage, because the newer data source excluded stocks that later transferred from NASDAQ. For example, American Eagle Outfitters transferred from NASDAQ to the NYSE in March of 2007, and its ticker symbol changed from AEOS to AEO. Volumes for AEOS are included in the June 2006 file from the ftp site, but were not included in the downloaded data and are not included in the volume reports that are currently available from NASDAQtrader.com.

Both of the volume sources described above reflect “single-counted” volume. That is, when different MPID’s represent the buyer and the seller, NASDAQ assigns the volume to the MPID that has the responsibility for reporting the trade to the tape. For example, when a trade is between two market makers, the selling market maker has reporting responsibility. This is unambiguous for Liquidnet, POSIT and Pipeline because their structure leads to their representing both sides of each trade. When examining volumes of other market makers, it should be noted that changes in total participation may not translate into the same changes in reported trades.

New monthly volume reports are available from NASDAQtrader.com that start with December 2005. These reports reflect “double-counted” volumes, which means that credit is given to the MPID’s on each side of the trade. If a single MPID represents both sides, then it gets credit for twice the trade volume. Although the prior reports included all MPID’s, participation is now optional and many of the MPID’s have opted not to participate in the new reports. Most of the participants in the new reports are sell-side brokers, and there are no NASDAQ volumes reported for Liquidnet, POSIT or Pipeline.

3.2. Institutional Trading Activity

If we observe trading volume in one of the three trading venues, then it must be the case that two or more of the participating institutions satisfied the following three conditions:

1) Chose to use the trading venue

2) Had orders that were in the same stock on the same day

3) Had orders that were in the opposite direction.

The primary focus of this paper is the determinants of the choice to use the trading venues (condition 1), but in order to make inferences using the observed volume in the venue, it is necessary to control for the trading demands of the institutions (conditions 2 and 3). To do this, I use changes in quarterly holdings for each institution the Thompson 13F database to construct a proxy for the trading demands in each stock for each of the institutions that might use the dark pools.

Institutions like Fidelity and Capital Research and Management (the manger of the American funds) file quarterly 13F reports with the SEC that disclose aggregate holdings of each security across all of the accounts where they have "investment discretion," which would include securities in the mutual funds that they manage, securities in portfolios that they manage for pension funds and high-net-worth clients, and securities in their own proprietary accounts. In addition, during my sample period the mutual funds managed by these firms individually reported their holdings of each security at the end of each fiscal year on Form N-CSR, at the middle of each fiscal year on Form N-CSRS, and at the end of the first and third quarters on Form N-Q.

Large 13F institutions often include many separate portfolios that are separately managed, and one portfolio may be buying a particular stock at the same time that another portfolio is selling the same stock, so the change in total 13F holdings for a stock will often understate the trading within the institution. In principle, one could address this issue by disaggregating the mutual funds that are contained in each 13F institution's reports. I do not take this approach for two reasons. The first issue is that the mapping between mutual funds and 13F institutions in the Thompson database contains several errors, and many of the funds are not identified with a 13F institution. The second issue is that many of the mutual funds have reporting cycles that do not end on the calendar quarters, so the mutual fund holdings cannot be matched to the 13F report, even if the mutual fund is correctly identified with its parent institution.

To get a sense of the relative sizes of 13F institutions and mutual funds and the connection between the two, consider Microsoft holdings as of June 30, 2006, which is the middle of my sample period. As of that date, there were 10.0 billion shares of Microsoft outstanding. The Thompson database includes 1,418 13F institutions that reported at least some shares of Microsoft as of that date, and the total across these institutions was 5.4 billion shares. In total, the Thompson database includes 4,834 mutual funds that issued quarterly holdings as of fiscal quarters ending between May 15, 2006 and August 15, 2006, and the total across these mutual funds was 2.3 billion shares of Microsoft. Of these, 3,826 mutual funds (total of 1.8 billion shares of Microsoft) had quarters that ended exactly on June 30, 2006. The table in the Thompson database that maps between mutual funds and 13F institutions, but this table excludes many of the smaller funds. Out of the 3,826 mutual funds with Microsoft holdings reported on June 30, 2006, 554 mutual funds (1.2 billion shares of Microsoft) are identified with their 13F institution (some incorrectly matched).

For each quarter I collect all 13F institutions that filed 13F reports at both the beginning and end of the quarter. In total there are 2,962 different institutions captured in the sample, and 60% of these are in the sample for all nine quarters. I use changes in holdings (adjusted for stock splits)[11] over the quarter for each institution as proxies for institutional orders for each stock. In contrast, the change in aggregate institutional holdings over a quarter is a measure of net institutional trading, which could also be called the institutional imbalance. Griffin, Harris and Topalogu (2003) examine institutional trading, but their focus is on imbalances. They use proprietary data to measure institutional and retail imbalances, classifying each trade based on the typical clientele of the brokers involved in the trade. They also use quarterly changes in aggregate 13F holdings, but only as a check of the reliability of their institutional imbalance proxy.

The simulations described in Section 5 assume that each change in holdings for each stock represents a single order send from a portfolio manager to a trading desk, and they also assume that the trading desk splits large orders across trading days. The simulations randomly assign the orders from each institution to trading days in the quarter and then randomly rout these orders to the dark pools based on a probit function that uses characteristic of the stocks, the orders, and the institutions as inputs. The goal is to understand the determinants of the routing

[11] The Thompson holdings data are adjusted for stock splits between the "rdate" (which is actual date when the holdings are measured) and the "fdate." If the 13F filing is available for a quarter, then the database includes records where the rdate and the fdate values match. For mutual funds, the earliest fdate is often after the rdate. Accordingly, it is generally necessary to adjust both for splits between the rdate and the fdate, and to adjust for splits that occur over the quarter (between two rdates). My main sample excludes stocks that split over the quarter, but some of the conditioning variables used in the routing decisions compare orders in the sample stocks to the universe of orders in all CRSP stocks.

decisions, so the simulated method of moments estimates of the parameters of this probit function are the quantities of interest.

One issue with my proxy for institutional trading is that not all investment firms are allowed to trade in the dark pools. During my sample period, all three dark pools took steps to exclude certain hedge funds, and they also excluded the brokers and proprietary traders from sell-side firms.

Another issue with the proxies for institutional trading is that each change in holdings represents the minimum number of shares traded by that 13F institution over the quarter. If institution A's holdings of XYZ increased by 100,000 shares over the quarter, it could be the case that 200,000 shares were purchased and 100,000 shares were sold over the quarter. It is probably relatively rare for a single mutual fund to buy and sell the same stock in the same quarter, but this becomes more likely when one considers different mutual funds within a 13F institution.

In some cases, the change in an institution's reported holdings of a particular stock may not represent trading activity. For example, in their October 2006 13F filing, U.S. Trust (a subsidiary of Charles Schwab) reported a new holding of 2.8 million shares of Genlyte Corporation (which is one of the stocks in my sample). This represented about 10% of the outstanding shares. Because the position exceeded 5% of the outstanding shares, U.S. Trust also filed form 13G, which indicated that the shares were part of a trust they had set up for Glenn Bailey, the founder of Genlyte. Evidently these shares were previously held by him, perhaps in a Schwab brokerage account, but the creation of the trust triggered the 13F reporting requirement. Other large changes in shares might result when large holders transfer their trusts from one institution to another, or when new shares are issued by the company. Whenever a 13F filer reports a change in holdings that is very large relative to the share outstanding, one might reasonably suspect that the change could have resulted from something other than normal trading. I address this issue with separate treatment of the changes in holdings that are more than five percent of shares outstanding: I construct separate volume measures for the panel data regressions in Section 6, and I include a separate probability that the large orders are not available to the dark pools in the simulations in Section 7.

Finally, there are two more issues, which are probably less important than the issues discussed above. First, the 13F reports exclude some of the institutional volume, because institutions with holdings less than $100 million are not required to file form 13F and the institutions that do file are allowed to exclude positions that are less than 10,000 shares with market value below $200,000. In most stocks the bulk of the trading is done by institutions that are large enough to file regularly, and these institutions seldom bother with very small positions in stocks. Also, the calculation of changes in holdings implicitly assumes unreported positions are zero, so if the fund goes from a large position to a small, unreported position, the calculated change in holdings will be approximately correct. The second remaining issue is that the 13F filings only show long positions. This may not be very important for the purpose of assessing potential dark pool volume, because institutions with large short positions are likely to be hedge funds, so they may be excluded from participating in the dark pools.

Recall that figure 2 shows dark pool shares of total consolidated volume are lower for the stocks with highest consolidated volumes. Figure 3 shows that this is for two reasons: 1) institutional volumes are a smaller proportion of consolidated volume for the high-volume stocks, and 2) dark pool shares of institutional volume are lower for the high-volume stocks. Trading in dark pools requires both institutional buyers and institutional sellers, so one potential measure of institutional volume is the minimum of institutional buying and institutional selling. Panel A of figure 3 shows this volume measure as a fraction of consolidated volume by decile of consolidated volume. As with figure 2, figure 3 shows the results for the full sample, and the restricted sample is roughly the 8th, 9th and 10th deciles of the full sample. The shape of these plots is not driven by using the minimum of institutional buying and selling; the plots look similar using the maximum of institutional buying and institutional selling. Panel A of figure 3 is the first part of the explanation for the hump-shaped pattern in figure 2; it shows that institutional trading activity makes up a lower fraction of consolidated volume for the highest and lowest volume stocks. As shown in Panel B of figure 3, this is not the whole story. Panel B shows dark pool volume as a fraction of the institutional volume proxy (minimum of institutional buying and selling) by decile of consolidated volume. In panel B, the highest volume stocks still stand out because they have low dark pool volume (again, the plots look similar suing the maximum in place of the minimum when calculating potential volume). This is particularly surprising because the order flow in the highest volume stocks is likely to be much steadier, leading to a greater fraction of orders arriving simultaneously. Thus, if institutional traders routed the same fraction of all orders to dark pools, we would expect to see a higher fraction of institutional orders in high volume stocks executed in the dark pools.

Table 2 shows the estimated institutional volumes for the second quarter of 2006 for both Microsoft and American Commercial Lines (the two stocks whose volumes by market center are reported in Table 1). The 12 institutions listed separately in the table are the three largest net buyers of Microsoft, followed by the three largest net sellers of Microsoft, followed by the three largest net buyers of American Commercial Lines that were not already listed as Microsoft buyers or sellers, followed by the three largest net sellers of American Commercial Lines that were not already listed as Microsoft buyers or sellers.

As shown in the last line of Table 2, the matched institutional volume measure (minimum of institutional buying and selling) for American Commercial Lines is 30% of consolidated volume for the second quarter of 2006. This is much higher than the 15%-20% level for the middle deciles of consolidated volumes shown in Panel A of figure 3. This higher level of institutional activity may explain why American Commercial Lines' dark pool share of consolidated volume in the second quarter of 2006 (5.5% as shown in Table 1) is substantially higher than the 1.5%-2% level for middle volume deciles shown in figure 2.

The institutional trading measures are matched to the NASDAQ volume sample using ticker symbol, and this appears to work in all cases. There are at least some institutional holdings for every stock in the sample in each quarter.

4. Hypotheses and Empirical Predictions

The summary statistics discussed in Section 2 suggest two related questions:

1. Why aren't dark pool volumes higher in general, given the potential benefits associated with avoiding intermediaries?
2. Why aren't dark pool volumes higher in the highest volume stocks, where the likelihood of finding a counterparty should be highest?

4.1. Hypotheses

To attempt to answer the above questions, I begin by assuming that if a counterparty is found, then executions in the dark pools are always the lowest cost, but these cost savings are proportional to the costs from trading in alternative venues. This assumption is motivated by the observation that the execution price in dark pools is often at the midpoint of the current quoted spread. The important point is that although highly liquid stocks have high probability of execution in the dark pools, they will also have low execution costs in other market centers, so the savings in execution cost from using the dark pool is likely to be small. Thus, if institutional traders face constraints or have other objectives besides execution cost, then they may be willing to accept a small execution cost penalty to trade elsewhere. All of the following hypotheses assume that institutional traders attempt to minimize trading costs, but that they face other objectives or constraints. Traders may face multiple constraints, so these hypotheses are not mutually exclusive.

Hypothesis 1: Institutional traders face a constraint that some fixed number of shares must be routed to certain brokers to satisfy soft-dollar agreements. They choose to send orders in the lowest cost stocks to these brokers (so these orders are not sent to a dark pool), because the transaction cost penalty is smallest for these stocks.

Hypothesis 2: Institutional traders worry about prices moving between the time that they get an order from the portfolio manager and the time they execute the trade. They tend to send orders in lower cost/higher volatility stocks to other market centers besides the dark pools in order to get certain, immediate execution.

Hypothesis 3: Institutional traders have limited time to monitor orders. The dark pools require some extra work, so they tend to send the orders with smaller potential savings to other market centers (for example, directly to ECN’s or to multiple market centers through computerized trading algorithms).

4.2. Empirical predictions and definitions of the spread rank and dollar value rank measures

In order to develop the empirical predictions for the above hypotheses, I assume that the potential cost savings per share from using the dark pool is a fixed proportion of the spread. Under Hypotheses 1, I further assume that the trader has a good sense of the total number of orders that will arrive during the quarter. Under these assumptions, if the trader were only seeking to satisfy a quota for total shares sent to soft dollar venues (Hypothesis 1), then orders with the lowest spread per share should be sent to those venues (and not sent to the dark pools).

To test Hypothesis 1, for each institution in each quarter, I combine the buy and sell orders, sort them by spread per share (lowest to highest), and then rank each stock according to where its shares fall in the sorted list. I define a *spread rank measure*, which is the average position of the shares for each stock in the list. For example, if the institution purchased or sold 4 million shares of Microsoft and purchased or sold 1 million shares of American Commercial Lines during the second quarter of 2006, and those were the only trades for the quarter, then that would be a total of 5 million shares traded by that institution in the quarter. The Microsoft average spread per share in the second quarter of 2006 was $0.01, which was lower than the American Commercial Lines average spread per share of $0.09, so the 4 million shares of Microsoft would be ranked ahead of the 1 million shares of American Commercial Lines. The spread rank measure for the Microsoft order from this institution in this quarter would be 0.4, which is calculated by taking the average position of the Microsoft shares in the sorted list (2 million) and dividing by the total number of shares in the list (5 million). Similarly, the spread rank measure for American Commercial Lines would be 4.5 million divided by 5.0 million, which equals 0.9. Hypothesis 1 predicts that the lowest spread stocks will be routed elsewhere, so the probability of routing each order to the dark pool will be an increasing function of the spread rank measure.

Hypothesis 1 predicts that the orders routed to satisfy soft-dollar agreements are selected by moving up the list of orders sorted by spread until the requirement is satisfied, so the impact of the decision rule is inherently non-linear. That is, one would expect especially low frequencies of dark pool usage for orders with the lowest spread rank measures and perhaps similar dark pool usage for orders with medium or high spread rank measures. The difficulty is determining where the cutoff should be. The SEC (1998) reports that institutions they surveyed paid soft dollar commissions on about 8% of their trades. Goldstein et. al. (2009) suggest that this figure understates the total paid for “premium services”; they estimate that58% of institutional orders in 2003 had commissions above discount levels. Of course, some of these higher commissions were likely related to features of the individual order, and the broker may be committing capital or working to find couterparties. In addition, the prediction of Hypothesis 1 is based on the assumption that the soft-dollar broker has higher expected execution cost than the dark pool (so the penalty[12] is minimized by sending stocks with the lowest spread per share). This assumption may be more reasonable for the smaller “boutique” brokers who may have valuable research but lack large trading operations. On the other hand, if a commission agreement is with a broker whose execution costs are competitive with the dark pools, then there is no incentive to select particular orders to satisfy that agreement.

---

[12] Note that even if there is an execution cost penalty, this does not necessarily imply that the soft-dollar arrangement is bad for the fund’s shareholders, because the research that is purchased can be valuable. Indeed, Edelen, Evans, and Kadlec (2008) find that mutual funds’ abnormal returns are positively related to soft dollar payments for research (although negatively related to soft dollar payments for other purposes).

The illustrative example above describes an institution that trades only two stocks, and those two stocks happen to be stocks included in the volume sample. The calculation of the spread rank measure includes all CRSP stocks traded by the institution in the quarter, including NYSE and AMEX stocks and other NASDAQ stocks not in the volume sample. For the second quarter of 2006, the aggregate spread rank measures for Microsoft and American Commercial Lines were 0.06 and 0.98, respectively. The fact that the number for American Commercial Lines is so close to 1.0 suggests that its spread was among the highest of stocks traded by institutions in the quarter. Note that although there were certainly higher spread stocks traded in the quarter, the spread rank measure is based on the position in the list in terms of shares traded, and the higher spread stocks were not traded in high volume.

Hypothesis 2 suggests that the trader is concerned about potential price movements while waiting to see if the order executes, so there will be a tradeoff between price volatility and expected cost savings. Importantly, if the trader's time is not a constraint, then this tradeoff does not depend on the size of the order. Rather, the trader will compare the proportional price volatility (standard deviation of return) to the expected proportional cost savings, which are assumed to be proportional to the spread. Hypothesis 2 predicts that dark pool usage will be an increasing function of the relative spread (the spread as a fraction of share price) and a decreasing function of price volatility. Both are used as inputs to the probit function that governs routing decisions in the simulations in Section 4.

Finally, Hypothesis 3 suggests that the trader has a fixed capacity for the number of order that could be "worked" through the dark pools, which means that the orders with the highest potential dollar cost savings would be routed to the dark pools. To test this hypothesis, I construct a second rank measure, which I call the *dollar value rank measure*, for each institution in each quarter by ranking the product of the spread per share times the shares outstanding (lowest to highest), and then dividing by the total number of orders. Returning to the example above, when an institution trades 4 million shares of Microsoft (spread is $.01 per share) and 1 million shares of American Commercial Lines (spread is $.09 per share), the total dollar spread for the Microsoft order is $.01 per share times 4 million shares = $.04 million. The total dollar spread for the American Commercial Lines order is $.09 per share times 1 million shares = $.09 million. By assumption, the potential dollar savings from using a dark pool is proportional to the total dollar spread. In this example, the total dollar spread for American Commercial Lines is higher than the total dollar spread for Microsoft, so the dollar value rank measure for the Microsoft order would be (1-0.5)/2=0.25 and the dollar value rank measure for the American Commercial lines order would be (2-0.5)/2=0.75. The subtraction of 0.5 from each individual rank causes the measure to have an average value of 0.5, regardless of the number of stocks traded by the institution. Hypothesis 3 predicts that the probability of routing each order to the dark pool will be an increasing function of the dollar value rank measure (the highest ranks are the highest potential dollar savings).

As with the spread rank measure, the dollar value rank measure calculation includes all CRSP stocks traded by the institution in the quarter. I construct a quarterly aggregate dollar value rank measure for each stock by calculating the average of the dollar value rank measures for that stock across the institutions that traded that stock in the quarter, weighting each institutions measure by the number of shares traded. For the second quarter of 2006, the share-weighted average dollar value rank measures for Microsoft and American Commercial Lines are 0.90 and 0.77, respectively. The high value for Microsoft indicates that large trade sizes more than offset the low spreads per share.

Before proceeding to the empirical tests, it is important to note that commissions and execution fees will also impact routing decisions. The typical execution fee for the dark pools is 2 cents per share, and the typical commission from a sell-side broker (excluding any soft dollar charges) is generally in excess of 3 cents per share. The execution fees at ECNs are much smaller, and will often include a small rebate for posting liquidity. Thus, if the natural alternative to the dark pool is a non-soft-dollar sell-side broker, then the benefit from successfully executing in the dark pool includes commission/fee savings as well as the likely execution price improvement from transacting at the midpoint. In contrast, if the natural alternative is an ECN, then the fee difference makes the dark pool less attractive. For example, suppose the quoted bid-ask spread is $.01 and a trader believes it is possible to sell all of the shares in an ECN at the current bid price. In this case, the ECN is preferred to the dark pool, because the dark pool potential price improvement is only $.005 per share, whereas the fee savings is nearly $.02 per share. Thus, in order to complete the interpretation of the dark pool volumes in light of the three hypotheses, it is necessary to test whether ECN's appear to be the primary execution alternative to the dark pools for particular types of stocks and institutional orders.

5. A Structural model for estimation

Recall that if we observe trading volume in one of the three trading venues, then it must be the case that two or more institutions satisfied the following three conditions:

1) Chose to use the trading venue

2) Had orders that were in the same stock on the same day

3) Had orders that were in the opposite direction.

The goal of the model presented in this section is to provide a structure for the empirical tests that incorporates random arrivals of trading interest. The structural model developed in this section is primarily descriptive; it does not explicitly model the optimization problem faced by the trader at a buy-side institution. Rather the model assumes that the result of this optimization can be described by a function that gives the probability of using the venue to attempt to trade, where the function arguments are measures of the characteristics of the stock and the characteristics of the institution.

Although there is no explicit model of the trader's optimal order strategy, it is probably best to think of the framework as capturing the decisions of a trader who has been directed by the portfolio manager to trade and to complete the order in a reasonable time at a reasonable price. The framework is probably less appropriate for capturing strategic informed traders as in Kyle (1985), but this may be reasonable in this case, because the three dark pools that I consider take positive steps to exclude opportunistic traders. Hendershott and Mendelson (2000) and Ye 2009 consider competition between dealer markets and crossing networks, where strategic informed traders are able to use either venue.

5.1. Order arrival and matching assumptions

The model assumes that the orders for each institution arrive randomly over the quarter. Large institutional orders are generally split and executed over multiple days, so the model assumes the number of days for each order is governed by the ratio of the order size to the average daily consolidated volume in that stock (measured over the previous quarter). Discussions with institutional traders suggest that they generally try to avoid being more than ¼ of the daily volume. Thus, it seems that on average an order equal to the average daily volume would take about 4 days to complete. In some cases, institutional portfolio managers direct their traders to buy or sell quickly, and in other cases, they allow the trader to be more patient. In order to capture the splitting of large orders across days, and also capture the fact that some orders may be handled with more or less patience than the average, I assume the number of days for each order is equal to one plus a random variable that has a poisson distribution with parameter 3 times the ratio of the order size to average daily volume. If the number of days implied by this random outcome is greater than the number of days in the quarter, then I assume the order is spaced evenly across all days in the quarter.

The last possible starting day for the order is equal to one plus the number of days in the quarter, minus the number of days in the order. The actual starting day is calculated by multiplying this value by a uniform (0,1) random variable and rounding up to the nearest integer. Thus, the simulation begins with each institution's change in holdings over the quarter, determines the number of sequential trading days, and then randomly inserts this string of trading days into the calendar quarter.

I recognize that if orders randomly arrive throughout the quarter, and they can bridge across the boundary of the quarter, then the distribution of the position of the order is more complex than what I have assumed in my model. To see why, suppose all order sizes are the same and order start dates are uniformly distributed throughout the quarter. If this were the case, then if we observe smaller changes in holdings in a quarter, then these must be the result of orders that bridge across one of the quarter boundaries. Clearly not all orders are the same size, but the example suggests that the measured changes in holdings could give some information about the likely position of the trades in the quarter. Attempting to capture this issue in the simulations would clearly add another layer of complexity. My goal in the simulations is to allow for the fact that finding a counterparty requires simultaneous arrivals of buy and sell orders, and to capture the idea that the probability of a match increases when there more different traders active on both sides of the market. Thus, I hope that the simulations produce reasonable inferences about the parameters governing the routing decisions.

The model treats the three dark pools as a combined venue, effectively assuming that when traders choose to route to any of the three dark pools, they continue to try all three of the dark pools with enough patience and persistence that any potential matching orders for that day ultimately connect with each other. Admittedly this is an extreme assumption, but the results of the simulations suggest that this kind of assumption is necessary to explain the dark pool volumes. Based on the estimated values for the model parameters, a high fraction of orders in these stocks must be routed to the dark pools to explain the observed volumes, even under the assumption that orders will match across the three dark pools. All dark pool volume is assumed to come from the institutional orders, so it is based on the minimum of buy and sell orders sent to the dark pools each day.

The next subsection introduces the notation for data items and calculated quantities. Subsection 5.3 describes the probit function governing the routing decisions.

5.2. Variable Definitions

I use the following notation for the remainder of the paper:

Indexing

q =1,2,…,9 denotes calendar quarter (3rd quarter 2005 through 3rd quarter 2007)

s = 1,2,…,862 denotes the stocks in the sample

i = 1,2,…,2962 denotes the institutions in the sample

n_s = the number of quarters that stock *s* is in the sample (ranges from 1 to 9)

Institutional order data

$O_{i,s,q}$= change in holdings of stock *s* by institution *i* in quarter *q* (the institution's "order")

$P_{s,q}$= total institutional purchases (sum of positive values of $O_{i,s,q}$)

$S_{s,q}$= total institutional sales (absolute value of the sum of negative values of $O_{i,s,q}$)

$L_{i,s,q}$= 1 if the order is 5% or more of the shares outstanding (=0 otherwise)

$LP_{s,q}$= total large purchases (sum of positive values of $L_{i,s,q}O_{i,s,q}$)

$LS_{s,q}$= total large sales (absolute value of the sum of negative values of $L_{i,s,q}O_{i,s,q}$)

Potential determinants of routing decisions

$\sigma_{s,q}$ = standard deviation of daily close-to-close returns over the quarter measured in percentage points. The measure is standardized using the mean and standard deviation across all stocks in the sample for the quarter.

$R_{s,q}$ = (relative) quoted spread (spread per share/price per share) measured in percentage points. The measure is standardized using the mean and standard deviation across all stocks in the sample for the quarter.

$srm_{s,i,q}$ = spread rank measure (for stock *s*, calculated using all of the trades for institution *i* in quarter *q*)

$i20srm_{s,i,q}$ = 1 if $srm_{s,i,q} < 0.20$ (the bottom 20% of spread per share), 0 otherwise

$dvrm_{s,i,q}$ = dollar value rank measure

$dp_{i,q}$ = the natural logarithm of the average of the dollar values of position sizes (across all stocks in the CRSP database) held by institution *i* at the beginning and end of quarter *q*. The measure is standardized using the mean and standard deviation across all institutions in the sample for the quarter.

$to_{i,q}$ = the turnover for institution *i* in quarter *q*, which is equal to the ratio of the total dollar value of the trades by institution *i* in CRSP stocks over quarter *q* to the average of the total holdings at the beginning and end of the quarter. The measure is standardized using the mean and standard deviation across all institutions in the sample for the quarter.

<u>Averages across institutions for a stock</u> (weighted by shares in each order)

$$I20SRM_{s,q} = \sum_{i=1}^{N_I} abs(O_{i,s,q}) i20srm_{s,i,q} \Big/ \sum_{i=1}^{N_I} abs(O_{i,s,q})$$

$$DVRM_{s,q} = \sum_{i=1}^{N_I} abs(O_{i,s,q}) dvrm_{s,i,q} \Big/ \sum_{i=1}^{N_I} abs(O_{i,s,q})$$

$$DP_{s,q} = \sum_{i=1}^{N_I} abs(O_{i,s,q}) dp_{i,q} \Big/ \sum_{i=1}^{N_I} abs(O_{i,s,q})$$

$$TO_{s,q} = \sum_{i=1}^{N_I} abs(O_{i,s,q}) to_{i,q} \Big/ \sum_{i=1}^{N_I} abs(O_{i,s,q})$$

<u>Volume data</u>

$D_{s,q}$ = total share volume in the dark pools for stock *s* in quarter *q*

$C_{s,q}$ = total consolidated share volume for stock *s* in quarter *q*

$F_{s,q} = 100^* D_{s,q}/C_{s,q}$ = percentage of consolidated share volume executed in the dark pools for stock *s* in quarter *q*

5.3. Probit routing function

The probit calculation produces a "value" for routing a particular order to the dark pool. For each stock *s* traded by institution *i* in quarter *q*, the value from routing the order to the dark pools is:

$$V_{i,s,q} = \alpha_q + \beta X_{i,s,q} + \tilde{z}_{i,s,q} + \tilde{e}_s \qquad (1)$$

where the column vector $X_{i,s,q}$ contains the characteristics of the stock and the order and the row vector β contains the parameters of interest. The random outcomes $\tilde{z}_{i,s,q}$ and $\tilde{e}_s$ are independent and normally distributed with mean zero and standard deviations σ_z and σ_e. Once the values of the random outcomes $\tilde{z}_{i,s,q}$ and $\tilde{e}_s$ are determined, the order is routed to the dark pools if $V_{i,s,q} > 0$. If the order is large (if $L_{i,s,q}$= 1), then I use a separate probit with a single constant parameter α_L to determine if the order is available to trade in the dark pool, regardless of $V_{i,s,q}$. The $\tilde{e}_s$ are stock random effects that represent tendency for orders in a particular stock to be sent to (or away from) dark pools. These effects are intended to capture features of the stock that may impact routing decisions but are not captured by the control variables $X_{i,s,q}$. The random effects do not have institution or quarter subscripts, so these effects are assumed to be constant across orders and over time. As an example, they may capture the fact that traders have learned dark pools are a particularly good (or bad) place to find liquidity in stock *s*.

Note that the routing decisions in equation 1 would not be changed if we multiplied the value of each $V_{i,s,q}$ by some constant positive value, so the distribution of routing outcomes would not be changed if we multiplied each of the parameters (including the standard deviations of $\tilde{z}_{i,s,q}$ and $\tilde{e}_s$) by some constant positive value. Accordingly, without loss of generality I fix σ_z =1.

The vector of explanatory variables, $X_{i,s,q}$, includes six components (see section 5.2 for complete definitions): $\sigma_{s,q}$ (return volatility), $R_{s,q}$ (relative spread), $i20srm_{s,i,q}$ (indicator that the spread rank measure is <0.20), $dvrm_{s,i,q}$ (dollar value rank measure), $dp_{i,q}$ (log of institution's average position size), and $to_{i,q}$ (institution's turnover). I use the indicator variable for the spread rank measure because of the mechanism that underlies Hypothesis 1. The idea is that traders know how many shares they need to direct to soft dollar brokers over the quarter, so they keep an eye out for orders in low spread stocks and use those to satisfy the requirements throughout the quarter. Accordingly, if Hypothesis 1 is correct we might expect routing decisions to differ dramatically between very low spread stocks and all other stocks, and expect little difference between somewhat high spread stocks and very high spread stocks. In contrast, Hypothesis 3 implies a comparison each day across the orders that arrive that day. If time permits, traders may use a Dark Pool for a low total spread order on one day, but on a busier day they may send a similar order elsewhere. Thus if Hypothesis 3 is correct, we should expect average routing behavior to change relatively smoothly across the different values of the dollar value rank measure.

Note that the return volatility and relative spread measures are standardized using the distributions across all sample stocks in the quarter, and the log of average position size and the turnover are standardized using the distributions across all institutions in the quarter. The elements of β in the probit function are constant across quarters, so the standardizations reflect an assumption that it is the relative values of the explanatory variables that determine the routing decisions each quarter. The two rank measures are already relative measures, so no additional standardization is required.

6. Panel Data Regressions

The key difficulty with estimating the model described in the previous section is that the routing decisions that are the outcomes of the probit function are not directly observed. Instead, we observe the orders and we observe the ultimate dark pool volumes, the latter of which result not only from routing decisions, but also from the results of random matching with other orders. One simple approach to this problem is to approximate the resulting dark pool volumes as a linear function of the aggregate orders, the characteristics of the stocks, and the average characteristics of the institutions and orders trading the stocks.

One motivation for the model and the associated hypotheses is to explain the pattern in dark pool volumes for different levels of consolidated volumes. The first row of Table 3 shows the average values of $F_{s,q}$ (the percentage of consolidated share volume executed in the dark pools), for three categories of consolidated volume and for the full sample. Recall that figure 2 shows the pattern for all NASDAQ stocks, whereas the sample to be used in the regressions and SMM tests is comprised of the 500 highest volume NASDAQ stocks each quarter. Accordingly, the three categories of the regression sample approximately correspond to the 8^{th}, 9^{th} and 10^{th} deciles of the full sample, which are the three rightmost points plotted in figure 2. Each quarter I compute the average within each volume category, and Table 3 reports the averages of these values across the nine sample quarters. By construction, the averages of the actual dark pool percentages shown in the first row of Table 3 roughly match the average shape and magnitude of the three rightmost values plotted in figure 2. The three curves plotted in figure 2 are based on three separate quarters, but the pattern is similar for the other quarters, so the numbers in the first line of Table 3 roughly correspond to averages across the three curves in figure 2.

The remainder of Table 3 summarizes the characteristics of the stocks and their institutional trading demands, split by volume category and for the full sample. The variables in the “Independent Variables” section of the table are used in regressions later in this subsection, where the independent variable is the percentage of consolidated share volume executed in the dark pools ($F_{s,q}$). The first four variables in this section of the table are defined as follows:

$SMatch_{s,q} = 100*Min[(P_{s,q}-LP_{s,q}),(S_{s,q}-LS_{s,q})]/C_{s,q}$

$SExtra_{s,q} = 100*Max[(P_{s,q}-LP_{s,q}),(S_{s,q}-LS_{s,q})]/C_{s,q}-SMatch_{s,q}$

$LMatchDiff_{s,q} = 100*Min[P_{s,q},S_{s,q}]/C_{s,q} -SMatch_{s,q}$

$LExtraDiff_{s,q} = 100*\{Max[P_{s,q},S_{s,q}]/C_{s,q}-Min[P_{s,q},S_{s,q}]/C_{s,q}\} - SMatch_{s,q}$

To understand these variables, it is helpful to focus on the first two, which are based on small orders. It turns out that there are no large orders for 76% of the stock/quarter observations in my sample, so for these cases the "Diff" variables are equal to zero. The "matched" (minimum of purchases and sales) and the "extra" volume (imbalance between purchases and sales) should both contribute to the realized volume in the dark pool, but at different rates. For example suppose for a particular stock and quarter there were 1 million shares purchased and 1 million shares sold by 13F institutions, and the total consolidated volume for the quarter was 10 million shares. In this example, the matched shares are 10 percent of consolidated volume, so the value of $SMatch_{s,q}$ in this case would be 10. Note that the 1 million shares (10 percent of consolidated volume) is also the maximum possible volume in the dark pools. To achieve this volume, every order would have to be routed to the dark pool and every purchase would have to match with a sale. Thus, when $SMatch_{s,q}$ is included in a regression to explain observed dark pool volumes, the estimated coefficient should be less than 1.0, because not all orders will be sent to the dark pools and not all orders sent to dark pools will happen to find a matching counterparty.

Next suppose that 2 million shares were purchased by 13F institutions and 1 million shares were sold. Compared to the previous example, we should see somewhat more dark pool volume, because some fraction of the extra sell orders should be routed to the dark pools and these extra sell orders should increase the probability that the purchases routed to the dark pools find a matching counterparty. Having additional orders on just one side of the market should not result in as much dark pool volume as having additional orders on both sides of the market, so one would expect that the regression coefficient on $SExtra_{s,q}$ would be less than the coefficient on $SMatch_{s,q}$.

To understand the coefficients on $LMatchDiff_{s,q}$ and $LExtraDiff_{s,q}$ note that if the dark pool routing probability and the matching probability were the same for large and small orders, then the regression coefficient on $LMatchDiff_{s,q}$ should equal the coefficient on $SMatch_{s,q}$, and the coefficient on $LExtraDiff_{s,q}$ should equal the coefficient on $SExtra_{s,q}$. However, given these orders are very large, one would expect them to be spread across many days in the quarter, so they should result in better matching, which would mean that the coefficients on $LMatchDiff_{s,q}$ and $LExtraDiff_{s,q}$ should be larger than the corresponding coefficients on $SMatch_{s,q}$ and $SExtra_{s,q}$. On the other hand, if some of these large orders are not really orders at all (recall the Genlyte example from section 3.2), then the coefficients on the large-order measures should be lower than the coefficients on the smaller-order measures.

For there to be volume in the dark pools, there must be institutional trading (captured by the variables described above) and the institutions must choose to use the dark pools. The remaining variables shown in Table 3 are related to the factors that drive routing probabilities in the model described in section 5.3. Accordingly, these variables are interacted with $SMatch_{s,q}$, and $SMatch_{s,q}$ is also included in the regression. The regression coefficient on $SMatch_{s,q}$ gives the sensitivity of dark pool volume to matched institutional order flow, and the coefficients on the interacting variables can be interpreted as the change in the sensitivity of dark pool to changes in the interacting variable. In principle, these variables could also be interacted with the other three institutional order measures, but it turns out that $SMatch_{s,q}$ is substantially more important than the other three in explaining dark pool volumes, so I limit the interactions to that variable for brevity and to ease interpretation of the results. The volatility and relative spread measures are already standardized across stocks, so a value of zero means that the observation equals the stock-wide average. For the low spread rank variable, a value of zero means that none of the orders for that stock are from the lowest spread orders for the institutions submitting the orders. For the remaining three variables, I subtract the full sample mean from the variable before multiplying by $SMatch_{s,q}$. This allows easier interpretation of the coefficient(s) on $SMatch_{s,q}$, because it makes sense to think of the values of all of the interacting variables being set to zero.

The regression coefficients for the interacting variables can be used to evaluate the three hypotheses. Hypothesis 1 predicts that when the low share rank indicator variable is equal to one, a smaller fraction of the shares will be routed to dark pools, so dark pool volumes will be less sensitive to institutional order flow for stocks with high values of $I20SRM_{s,q}$. This means that the estimated coefficient (on $I20SRM_{s,q}$ multiplied by $SMatch_{s,q}$) should be negative. Using similar logic, Hypothesis 2 predicts that the regression coefficient on the interaction between order flow and return volatility will be negative and the regression coefficient on the interaction between order flow and relative spread will be positive. Finally, hypothesis 3 predicts that the regression coefficient on the interaction between order flow and dollar value rank measure will be positive. The interactions between matched order flow and the remaining two explanatory variables in Table 3 are included to control for the possibility that dark pool usage may be driven by characteristics of the institution as opposed to characteristics of the stock.

Although the share weighted characteristics of the institutions do not differ significantly across the consolidated volume categories, the second to last line of Table 3 shows the value for the average position size is high for all three categories. Recall that this variable is standardized across institutions, so values near one indicate institutions with (log) average positions sizes that are one standard deviation above the mean across institutions. It is not surprising that the majority of the shares in the institutional orders come from the larger institutions. The average (normalized) values for the turnover measure are near zero, because the higher turnover institutions tend to have somewhat smaller average position sizes.

Table 4 shows the results of three different panel data regressions using my sample of 4,500 stock/quarters. Each regression allows for random effects by stock and the t-statistics are based on robust estimates of the parameter variance/covariance matrix. The first two columns of Table 4 show the results of:

$$F_{s,q} = a + b_1 SMatch_{s,q} + b_2 SExtra_{s,q} + b_3 LMatchDiff_{s,q} + b_4 LExtraDiff_{s,q} + \phi_{s,q} \qquad (2)$$

where:

$\phi_{s,q} = \delta_s + \varepsilon_{s,q}$ (stock-specific random effects)

Table 4 shows that regression (2) produces the predicted positive signs for the coefficients all four of the order measures, and as expected they are all less than one and the coefficients on the measures of imbalanced order flow are lower than the coefficients on the measures of balanced order flow. The estimate for b_1 of 0.143 has a natural interpretation: a 1,000-share increase in institutional orders on both sides of the market would on average result in a 143-share increase in dark pool volume. The estimate for b_2 is positive and less than the estimate for b_1. It suggests that a 1,000-share increase in the imbalance of orders would on average result in a 33-share increase in dark pool volume.

The coefficients b_3 and b_4 on the extra volumes associated with large orders are smaller than the corresponding coefficients b_1 and b_2 associated with smaller orders. This provides empirical support for the assumption that the large changes in reported positions may be less likely to reflect actual trading activity. Of course, it is also possible that these very large changes are accomplished through trading, but are not as likely to be routed to dark pools. Recall that if all these changes reflected traded orders, and these orders were as likely as small orders to use dark pools, then one would expect b_3 to be greater than b_1 and expect b_4 to be greater than b_2, because such large orders would be split across several days, thereby increasing their chance to interact with orders on the other side of the market.

If the relationship between dark pool volumes and the four institutional order measures were linear, then the expected value for the intercept in regression (2) would be zero, because if all of the order measures were zero then there should be no dark pool volume. Table 4 shows that the estimate of -0.05 for the intercept in regression (2) is statistically insignificant, and even adding or subtracting two times the standard error yields a fairly small magnitude compared to average dark pool volumes of 2.12 percent.

The middle columns of Table 4 report the results for the following regression:

$$F_{s,q} = a + b_{1,q}SMatch_{s,q} + b_2SExtra_{s,q} + b_3LMatchDiff_{s,q} + b_4LExtraDiff_{s,q} + \phi_{s,q} \qquad (3)$$

In regression (3), the coefficients on $SMatch_{s,q}$ vary by quarter over the sample, which allows for the changes in the routing probabilities over time, possibly because the dark pool firms were acquiring new clients and adding functionality. The results from this regression suggest steadily increasing routing probabilities over time. The intercept is now significantly negative, but the magnitude of the estimate is still relatively small.

The last two columns of Table 4 report the results for the regression used to test the hypotheses:

$$F_{s,q} = a_q + b_{1,q}SMatch_{s,q} + b_2SExtra_{s,q} + b_3LMatchDiff_{s,q} + b_4LExtraDiff_{s,q}$$
$$+ b_5\sigma_{s,q}{*}SMatch_{s,q} + b_6R_{s,q}{*}SMatch_{s,q} + b_7I20SRM_{s,q}{*}SMatch_{s,q}$$

$$+ b_8(DVRM_{s,q} - \overline{DVRM})*SMatch_{s,q} + b_9(DP_{s,q} - \overline{DP})*SMatch_{s,q}$$

$$+ b_{10}(TO_{s,q} - \overline{TO})*SMatch_{s,q} + \phi_{s,q} \qquad (4)$$

A variable name with a bar above it and no subscripts denotes the full sample average. The estimates for b_5 through b_8 from regression (4) provide support for Hypothesis 1 (soft dollar agreements), but they are inconsistent with Hypothesis 2 and Hypothesis 3. The significant negative estimate for b_7 indicates the orders with the lowest dollar spread per share are less likely to be routed to dark pools, which is consistent with Hypothesis 1. The coefficient estimate of -0.054 indicates how much lower the volume sensitivity is for stocks with low dollar spreads per share. Using the 4^{th} sample quarter as an example (the second quarter of 2006), a stock note that $b_{1,4}$ is 0.144, so for a stock that does not have a low spread per share ($I20SRM_{s,4}$=0) an additional 1,000 shares of orders on both sides of the market would result in 144 shares of expected additional dark pool volume. In contrast, for a stock with low spread per share ($I20SRM_{s,4}$=1), the sensitivity falls to 0.144-0.044=0.090, so an additional 1,000 shares of orders on both sides of the market would results in 90 shares of expected additional dark pool volume.

The significant positive estimate for b_5 indicates that orders in more volatile stocks are more likely to be routed to dark pools, which is inconsistent with Hypothesis 2. The significant negative estimate for b_8 indicates that orders with higher total dollar spread are less likely to be routed to dark pools, which is inconsistent with Hypothesis 3.

The estimates for b_9 and b_{10} provide evidence as to the types of institution that are likely to use dark pools. The estimate for b_9 is not statistically significant, but the estimate for b_{10} is significantly negative, indicating that institutions with higher turnover are less likely to use dark pools. This may reflect the fact that the dark pools take steps to exclude institutions that follow order anticipation strategies, and these institutions will tend to have high turnover.

Recall that one motivation for the regressions was to explain the pattern in dark pool volumes for different levels of consolidated volumes. The category of dark pool volume does not produce a significant coefficient when added to regression (4), which suggests that the other features of these stocks that are captured in Table 3 are sufficient to explain the pattern. Note that the first line of Table 3 shows that there is a 1.41 percentage point difference between the dark pool volume shares in the top and bottom categories of consolidated volume. Using regression (3), the average predicted dark pool volumes differ by 1.01 percentage points between the top and bottom consolidated volume categories, so 1.01 percentage points of the observed 1.41 percentage point difference in observed volumes is explained by the differences in the measures of institutional order volume. Using regression (4), the average predicted dark pool volumes differ by 1.25 percentage points between the top and bottom consolidated volume categories, so the added variables in regression (4) explain much of the remaining difference. Although many of the coefficients on the added variables in regression (4) are statistically significant, the values of most of these added variables do not vary much across the consolidated volume categories. The obvious exception is the low spread rank indicator variable. When the value for b_7 is set to zero, the difference in average predicted dark pool volumes between the top and bottom consolidated volume categories falls to 1.07 percentage points, so the low spread rank indicator variable explains 1.25-1.07=0.18 percentage points of the difference.

7. Simulated Method of Moments Results

The Simulated Method of Moments procedure produces explicit estimates for β, which is the vector of slope coefficients in the probit function that governs routing decisions. The estimates of the first four elements of β are used to evaluate the three hypotheses. Hypothesis 1 predicts that the slope coefficient on the low share rank indicator variable will be positive.

Hypothesis 2 predicts that the slope coefficient on the return volatility will be negative and the slope coefficient on the relative spread will be positive. Finally, hypothesis 3 predicts that the slope coefficient on the dollar value rank measure will be positive. The remaining two explanatory variables are included to control for the possibility that dark pool usage may be driven by characteristics of the institution as opposed to characteristics of the stock.

7.1. Simulation details

For a given set of parameter values, the simulation steps are as follows:

Step 1: Convert each institution's order (change in holdings) for each stock into daily orders.

1A. Draw a poisson variable with a mean of 3 times the ratio of the order size to the average daily volume, and truncate to 63 if the outcome is greater than 63. This is the number of trading days for the order. The order size divided by the number of days gives the shares to be traded each day.

1B. Subtract the number of days for the order from the number of trading days in the quarter. This gives the list of potential start days in the quarter. Choose a start day for the order using a random draw that is uniformly distributed over the list of start days.

Step 2: For each stock randomly generate the error terms $\tilde{e}_s$

Step 3: For each institution's order in each stock, calculate $V_{i,s,q}$ by randomly generating the error term $\tilde{z}_{i,s,q}$, adding the error terms for the stock/market from step 2, and adding $\alpha_q + \beta X_{i,s,q}$. If the resulting $V_{i,s,q}$ >0 then route the order to the dark pool using the share amounts and days determined in step 1.

Step 4: If $L_{i,s,q}$= 1 (a large order) then draw a standard normal random variable. If this draw is less than α_L then the order is not available to be routed to dark pools, regardless of the outcome of step 3.

Step 5: For each stock, each day, determine the minimum of the total shares in buy orders sent to the dark pools and the total shares in sell orders sent to the dark pools. This is the simulated dark pool volume for that stock that day. Add the simulated daily volumes for each stock over all of the trading days in the quarter to yield the simulated dark pool volume for that stock over the quarter.

The vector of parameters to be estimated, θ, has 16 elements, including nine values of the quarterly constants α_q, the six elements of β, and the constant in the separate probit calculation for large orders α_L. The remaining parameter, σ_e is chosen in a calibration exercise. For a particular value of σ_e I use SMM to estimate the probit parameters. I then use these parameters to generate simulated datasets and I run the panel data regression from equation 2 in section 6 on these simulated data sets. In addition to producing coefficient estimates, the panel data regression output also contains an estimate of the variance of the stock-specific random effects, which is 0.27 using the actual data. I use trial and error to find a value for σ_e (a value of 0.2 accomplishes this goal) that causes the variance of the stock-specific random effects based on the simulated data sets to have a value of approximately 0.27. The estimated parameters are similar to those reported when values of 0.15 or 0.25 are used for σ_e.

The SMM procedure adjusts the parameters in the probit function so that the averages of the moments from the simulated dataset match the average of the moments from the actual data.

The moments that I use are Generalized Least Squares estimates[13] from a regression of dark pool shares of consolidated volume for each stock/quarter ($F_{s,q}$), on all of the independent variables in regression (2) in section 6. The appendix provides explicit definitions of the moment conditions and provides the details of the SMM procedure.

7.2. SMM Parameter Estimates

Table 5 shows the SMM estimates for the six parameters included in β, which are the parameters that determine the routing decisions in equation 1. The estimate for α_L is -0.21, and a standard normal random draw must be larger than this value for the order to considered available for routing, so there is a 0.58 probability that a larger order is available for routing to the dark pool, before considering the value of $V_{i,s,q}$.

Overall, the results shown in Table 5 are consistent with the results in Section 6 that are based on the panel data regressions. Consistent with hypothesis 1, the coefficient on the low spread rank indicator variable is significantly negative, which indicates that orders with the lowest spreads per share are less likely to be sent to dark pools. The other two hypotheses are not supported. The coefficient on the dollar value rank measure is negative, which indicates that orders with higher total dollar spread are *less* likely to be sent to a dark pool. Finally, consistent with hypothesis 2 the coefficient on the relative spread is significantly positive, but the positive sign of the coefficient on return standard deviation is not consistent with hypothesis 2.

The results in the last two rows of table 5 indicate which type of institution is more likely to route orders to the dark pools. As was the case with the results from the panel data

---

[13] Hennessy and Whited (2005) use regression coefficients in a simulated method of moments estimation of a dynamic model of firm capital structure.

regressions, the SMM results suggest that institutions with higher turnover are less likely to route their orders to the dark pools, and they produce a statistically insignificant result for average position size.

7.3. Features of the simulated data

Most of the orders in the sample are smaller than the average daily volume. Using the estimated values of the parameters and using an equally-weighted average of the stock/quarters, 10.4% of the orders routed to the dark pools in the simulations are executed over multiple days, and the average number of days per order is 1.17. This is reasonably consistent with results from Jones and Lipson (1999), who examine institutional orders in a sample of stocks that switched to the NYSE from either the NASDAQ or the AMEX during 1993-1995. They report that approximately 15% of the orders in their sample were “worked” by the trading desk, in the sense that they resulted either in executions with multiple brokers or executions across multiple days. Thus, the 15% value reported by Jones and Lipson (1999) represents an upper limit on the fraction of orders executed across multiple days.

Table 6 summarizes the simulation results in total and by the same three categories of consolidated volume that were used in Table 4. On average across 150 simulations, about 46% of the shares in the institutional orders are routed to the dark pools, and approximately 24% of these routed shares are executed in the dark pools. Recall that the simulations implicitly assume that orders routed to any of the dark pools on a particular day will ultimately connect with each other, and in spite of this assumption it is necessary to have a fairly high fraction of orders routed to the dark pools in order to generate the levels of total volumes that we observe. When evaluating the high estimated routing rates and low estimated execution rates produced by the SMM procedure, an important caveat is that the structural model implicitly assumes that traders are somewhat impatient. When the order arrives at the trading desk, the model assumes that the trader may split it across days if it is large, but otherwise she at least begins execution on the day she receives the order. The model assumes she is patient enough to find all possible counterparties in the dark pool that day, but she does not wait multiple days to begin execution. If the trader were more patient, then she might try the order in the dark pools over successive days, and such behavior by multiple traders would substantially increase the matching rate. If the matching rate were higher, then one could explain the observed volumes with a lower fraction of the orders routed to the dark pools. If one were to think of the problem in the framework used by Goettler, Parlour, and Rajan (2005), then in my model the traders act as if their private valuation of the security is quite different from the current prices available in either the dark pools or other market centers. Several discussions with actual traders suggest that this assumption may be fairly reasonable. In nearly all buy-side firms the trading function is separate from the portfolio management function, and traders generally report that most portfolio managers expect most orders to be completed by the end of the day.

Table 6 shows that the simulations do a reasonable job of capturing the pattern in dark pool volumes across the three volume categories. One of the goals of this paper is to investigate the cross-sectional differences in routing probabilities, and one way to evaluate the impact of the variables included in the probit is to look at the change in routing probability that results from changing one of the control variables while holding the other variables fixed. The one variable from the probit function that varies substantially across the different volume categories is the low-spread-rank indicator variable. The last three lines of Table 6 show the results from simulations where the probit function coefficient on this variable is set to zero. The main simulations produce a 1.37 percentage point difference between the dark pool shares in the first and third consolidated volume categories. This difference falls to 1.25 percentage points when the coefficient on the low-spread-rank indicator variable is set to zero. Thus the low-spread-rank indicator variable explains 0.12 percentage points of the difference between the first and third categories, which is two thirds of the impact estimated in the panel data regressions in Section 6.

Returning to the examples of Microsoft and American Commercial Lines in the second quarter of 2006, note that Microsoft has a very narrow spread and American Commercial Lines has a relatively wide spread. In addition, because the orders in Microsoft tend to be very large, they tend to have relatively high values of the dollar value rank measure. Averaging across the 150 simulations for that quarter, 28% of the Microsoft orders are routed to the dark pools and 40% of the American Commercial Lines orders are routed to dark pools.

8. ECNs as an alternative to dark pools

As discussed in Section 4, while the low usage of dark pools for low-spread stocks is consistent with the use of these orders to satisfy soft-dollar routing agreements, it is also consistent with use of ECN's as an alternative to dark pools, because ECN's have lower fees. My data set includes execution volumes for two large ECN's over the full sample period: Automated Trading Desk and Bloomberg Tradebook. Together, these two ECN account for an average of 2.72 percent of consolidated volume for my sample stocks, and this percentage is reasonably stable across the nine sample quarters.

If the reduced usage of dark pools for institutional orders in low spread stocks is due to routing of these orders to ECN's, then we should see increased sensitivity of ECN volume to institutional order flow in these stocks. Table 7 repeats regression (4), except that in place of dark pool volume as a percent of consolidated volume, the independent variable is the total volume for the two ECN's as a percent of consolidated volume. The coefficient on the interaction between the low spread rank indicator variable and the matched volume variable is significantly negative, indicating that institutional orders in low-spread stocks are less likely to be routed to these ECN. This is the opposite of what would be expected if the results in the previous sections were driven by a search for lower fees, and it is consistent with Hypothesis 1. Under Hypothesis 1, these orders are the natural ones to rout to soft-dollar brokers so they would also be less likely to be sent to ECNs.

Table 7 also shows a significant positive estimate for b_8, which indicates institutional orders with the highest dollar value ranks (highest value for the product of spread per share and shares in the order) are more likely to be sent to these two ECNs. The results in Sections 6 and 7 suggested these orders were less likely to be routed to dark pools. Taken together, the results suggest institutional traders tend to favor ECN's over dark pools for these orders.

9. Conclusion

I investigate determinants of trading volume for NASDAQ stocks in three of the "dark pools" that cater to institutional traders: Liquidnet, POSIT and Pipeline. I use a sample of volumes by stock for each of these venues that covers the nine calendar quarters from the third quarter of 2005 through the third quarter of 2007. To measure institutional trading, I use the changes in quarterly institutional holdings. I develop a structural model that allows for random arrivals of trading interest within each quarter, and enables me to disentangle the effect of trader choice from the effects of random matching of trading interest. I use panel data regressions that are motivated by the model, and I directly estimate the model's parameters using simulated method of moments.

I show that these venues appear to attract the lowest share of institutional trades in stocks with the highest consolidated volumes. I test three hypotheses that could explain this pattern.

The results suggest that dark pool usage is lower for stocks with lower spreads per share, which is consistent with trader routing of these stocks to other venues in order to satisfy soft dollar agreements. I show that this result cannot be explained by use of lower-fee ECNs to execute these orders, because ECN’s receive a lower fraction of institutional orders in these stocks.

Finally, unrelated to the tested hypotheses, the results suggest that institutions with higher turnover are less likely to route their orders to the dark pools, possibly because dark pools take steps to exclude certain types of institutions that would tend to have higher turnover.

References

Barclay, M., Warner, J., 1993. Stealth trading and volatility: which trades move prices? Journal of Financial Economics 34, 281-305.

Edelen, R., Evans, R., Kadlec, G., 2008. What do soft-dollars buy? Performance, expense shifting, agency costs., working paper

Goettler, R., Parlour, C., Rajan, U., 2005. Equilibrium in a dynamic limit order market. Journal of Finance 60, 2149-2192.

Goldstein, M., Irvine, P., Kandel, E. Wiener, Z., 2009. Brokerage commissions and institutional trading patterns. Review of Financial Studies, forthcoming.

Gourieroux, C., Monfort, A., 1993. Simulation-based inference. Journal of Econometrics 59, 5-33.

Griffin, J., Harris, J., Topaloglu, S., 2003. The dynamics of institutional and individual trading. Journal of Finance 58, 2285-2320.

Hasbrouck, J., Saar, G., 2002. Limit orders and volatility in a hybrid market: the Island ECN. NYU working paper

Hendershott, T., Mendelson, H., 2000. Crossing networks and dealer markets: competition and performance. Journal of Finance 55, 2071-2116.

Hendershott, T., Jones, C., 2005. Island goes dark: transparency, fragmentation, and regulation. Review of Financial Studies 18, 743-793.

Hennesy, C., Whited, T., 2005. Debt dynamics. Journal of Finance 60, 1129-1166.

Hsiao, C., 1986. Analysis of panel data (Cambridge University Press, Melbourne, Australia).

Jones, C., Lipson, M., 1999. Execution costs of institutional equity orders. Journal of Financial Intermediation 8, 123-140.

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

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

Kyle, A., 1985. Continuous auctions and insider trading. Econometrica 53, 1315-1336.

Schwartz, R., Steil, B., 2002. Controlling institutional Trading costs. Journal of Portfolio Management 28, 39–49.

Securities and Exchange Commission. 1998. Inspection report on the soft dollar practices of broker-dealers, investment advisors and mutual funds. http://www.sec.gov/news/studies/softdolr.htm.

Ye, M., 2009. A glimpse in to the dark: price formation, transaction cost and market share of the crossing network. Working paper.

Appendix

SMM Estimation of Parameters and Standard Errors

The estimation procedure closely follows Section 3 in Gourieroux and Monfort (1993) and the notation in this appendix will follow Gourieroux and Monfort, except that Gourieroux and Monfort use *i* to index the independent individuals, and in this paper the independent individuals are the stocks, which are indexed by *s* (in this paper *i* is used to index the institutions). Recall from Section 5 that the subscript *s* identifies each stock, and there are *n*=862 unique stocks in the sample. The column vector y_s contains up to nine quarterly values of the dark pool volume scaled by consolidated volume ($100 {*} D_{s,q}/ C_{s,q}$) for stock *s* in quarter *q*. The procedure uses moment conditions that are a function of y_s and a collection of observable variables, denoted x_s. The moment conditions for stock *s* are given by the vector-valued function $K(y_s, x_s)$, and the expression $K(y, x)$ (without the "*s*" subscripts) refers to the vector-valued average across the stocks, that is,

$$K(y,x) = \frac{1}{n}\sum_{s=1}^{n} K(y_s, x_s)\ .$$

In my application, the sample moment conditions are the Generalize Least Squares coefficients on the 19 independent variables (9 quarterly intercepts and 10 other variables) that are used in the panel data regressions in section 6. The GLS coefficients are based on the assumption that the ratio of the variance of the stock-specific random effects to the variance of the observation-specific effects is τ=0.28, which is the estimate produced by the restricted maximum likelihood procedure used to generate the first two columns of Table 3. Re-running the procedure using OLS moment conditions (equal to the GLS estimates if $\tau = 0$) yielded very similar coefficient estimates. The GLS coefficients are very close to the restricted maximum likelihood coefficients reported in Table 3. The benefit of using GLS coefficients as the moment conditions is that it allows explicit definition of the function $K(y_s, x_s)$.

To construct the moment conditions, I use the fact that in this application with random coefficients, the GLS coefficient estimates are equal to the estimates from an OLS regression on transformed values of the independent and dependent variables (See Frees (2004), page xx and Hsiao (1986), page 37). Let t_s denote the number of time-series (quarterly) observations for stock *s*, so in my sample t_s ranges from one to nine and $\sum_{s=1}^{n} t_s = 4,500$. The transformation subtracts the following fraction of the average across the quarterly values for stock *s* from each quarterly dependent variable values and from each of the quarterly independent variables:

$$\varsigma_s = \left(1 - \frac{1}{\sqrt{\tau t_s + 1}}\right)$$

Let *g* denote the matrix of transformed independent variables (4,500 rows and 19 columns), let g_s denote the portion of this matrix containing the transformed variables for stock *s* (one to nine rows and 19 columns), and let $g_{s,t}$ (t=1,2,…,t_s) denote one row of g_s. Note that the subscript *t* is different from the subscript *q* introduced in section 5 because *t* indexes only the observations where stock *s* is in the sample. Define the set of observable variables for each stock to be:

$$x_s = \{t_s, g_s\}$$

Then the moment conditions are given by:

$$K(x_s, y_s) = \sum_{t=1}^{t_s} n(g'g) g_{s,t}' \left(y_{s,t} - \varsigma_s \sum_{j=1}^{t_s} \frac{y_{s,j}}{t_s} \right)$$

The average of the above across the 862 stocks in the sample gives the 19 GLS moment conditions.

The SMM procedure chooses the value of the parameter vector θ so that the simulated moments are "close" to the sample moments. Let h=1,2,…,H index the simulated data sets and let $\tilde{k}(x_s, u_{h,s}, \theta^*)$ denote the vector of moments for stock s from simulation h using parameters θ^*. Following Gourieroux and Monfort (1993), $u_{h,s}$ represents the full set of simulated latent variables used to calculate the simulated value for y_s. In my application, $u_{h,s}$ includes the random draws used to generate the random outcomes $\tilde{z}_{i,s,q}$ and $\tilde{e}_s$ that are used in each institution's value function for each order (equation 1), as well as the random draws used to determine the number of days and starting position for each order. Once the simulated values of each y_s are determined, the calculation of $\tilde{k}(x_s, u_{h,s}, \theta^*)$ is the same as the calculation of $K(y_s, x_s)$. Thus, if θ_0 denotes the true value of the parameters, then $E\left[K(y,x)\right] = E\left[\tilde{k}(x, u_h, \theta_0)\right]$, where, as before, when the "s" subscript is omitted, the expressions refer to the averages across the stocks.

The estimated value of the parameters is obtained from the following optimization:

$$\hat{\theta} = \arg\min_{\theta} \left\{ \left[K'(y,x) - \frac{1}{H}\sum_{h=1}^{H} \tilde{k}'(x, u_h, \theta) \right] \Omega \left[K(y,x) - \frac{1}{H}\sum_{h=1}^{H} \tilde{k}(x, u_h, \theta) \right] \right\} \quad (2),$$

where Ω is a positive-definite square "weighting" matrix. Note that $K(y,x)$ and $\tilde{k}(x, u_h, \theta)$ are column vectors. The apostrophe denotes the transpose operator. In performing the optimization in equation 2, the values of the simulated latent variables u_h are held fixed as θ is varied (see Gourieroux and Monfort (1993), page 10).

The optimization given in equation 2 is done twice. In the first step, Ω is the identity matrix. The estimated parameter vector from this first step, $\hat{\theta}^1$, is then used to estimate the optimal weighting matrix:

$$\hat{\Omega}^* = \left\{ \frac{1}{n} \sum_{s=1}^{n} \left[K(y_s, x_s) - \frac{1}{H} \sum_{h=1}^{H} \tilde{k}(x_s, u_{h,s}, \hat{\theta}^1) \right] \left[K'(y_s, x_s) - \frac{1}{H} \sum_{h=1}^{H} \tilde{k}'(x_s, u_{h,s}, \hat{\theta}^1) \right] \right\}^{-1} \quad (3)$$

The final parameter estimates are then calculated using the weighting matrix from equation 3 in the minimization shown in equation 2. Gourieroux and Monfort (1993) suggest that the calculation of the optimal weighting matrix in equation 3, which is done once, should be done with a large value of *H*. I use *H*=150 for the optimizations in equation 2 and *H*=500 for the calculation in equation 3.

The estimated asymptotic variance/covariance matrix for $\sqrt{n}(\hat{\theta} - \theta_0)$ is given by

$$\left[D'\hat{\Omega}^* D \right]^{-1} + \frac{1}{H} \left[D'\hat{\Omega}^* D \right]^{-1} D'\hat{\Omega}^* \Phi \hat{\Omega}^* D \left[D'\hat{\Omega}^* D \right]^{-1},$$

where

$$\Phi = \frac{1}{H} \sum_{h=1}^{H} \tilde{k}'(x, u_h, \hat{\theta}) \tilde{k}(x, u_h, \hat{\theta}) - \left[\frac{1}{H} \sum_{h=1}^{H} \tilde{k}'(x, u_h, \hat{\theta}) \right] \left[\frac{1}{H} \sum_{h=1}^{H} \tilde{k}(x, u_h, \hat{\theta}) \right]$$

is the estimated covariance matrix for the simulated moment conditions and D is the matrix of partial derivatives of the vector moment conditions with respect to θ evaluated at $\hat{\theta}$. I estimate the elements of D by re-running the simulations a total of 2*19=38 times. In each pair of simulations, I vary a single parameter in $\hat{\theta}$ by a small amount in each direction and calculate the slope of the change in the average of each moment across the simulations.

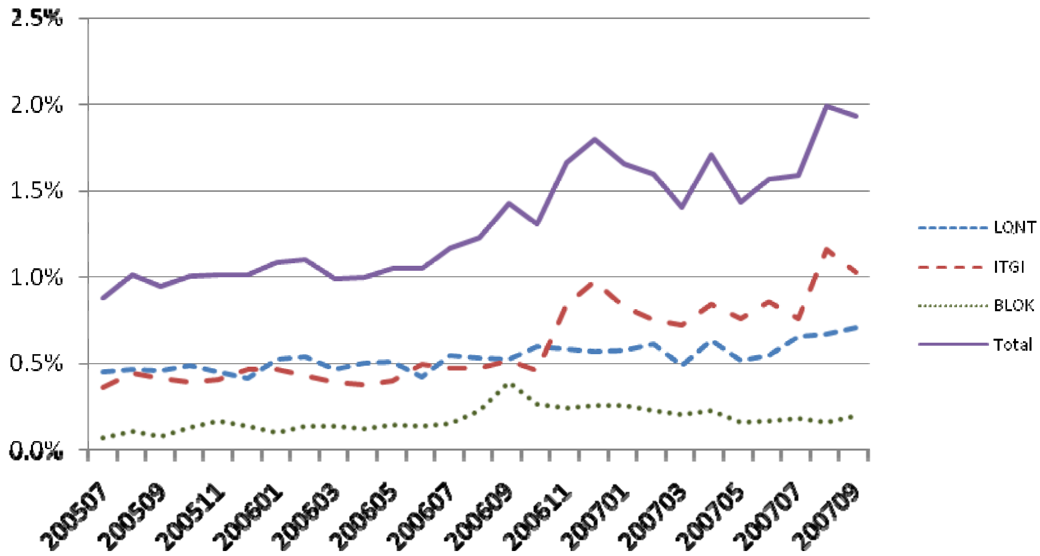

Fig. 1. Dark pool shares of total consolidated volumes across all stocks by month.
Each month, the sample consists of all NASDAQ stocks that pass the data screens for the quarter. Each line in the figure shows total dark share volume (summed across all sample stocks for the month) divided by total consolidated volume (summed across all sample stocks for the month).

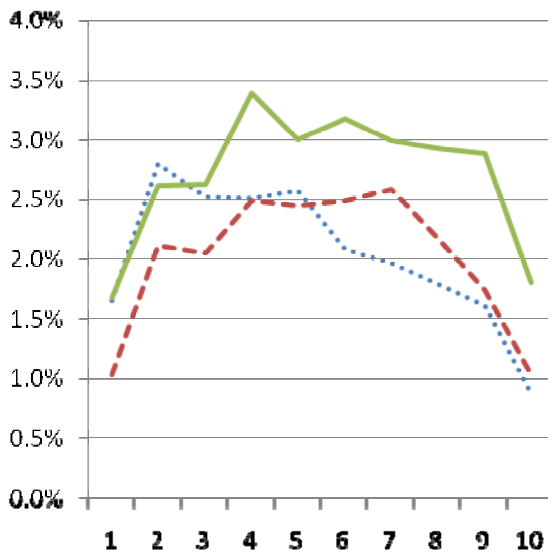

Fig. 2. Dark pool shares of consolidated volume

The figure shows average dark pool share of consolidated volume across the stocks in each decile of consolidated volume. The dotted lines show results for the third quarter of 2005, the dashed lines show results for the second quarter of 2006 and the solid lines shows result for the third quarter of 2007.

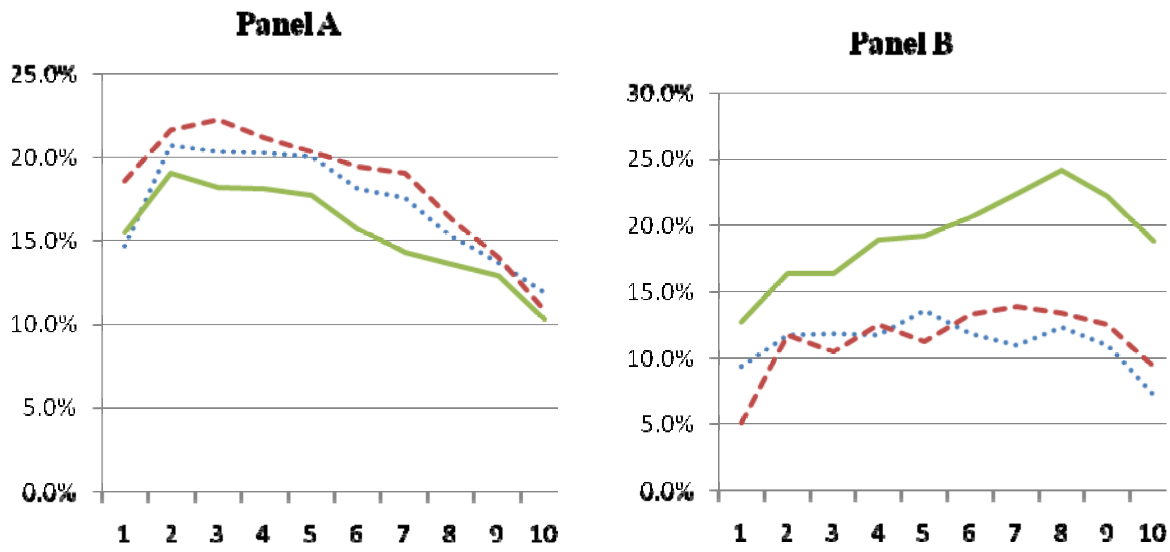

Fig. 3. Institutional volume and dark pool share of potential institutional volume
Both panels reflect averages across the Nasdaq stocks within each decile of quarterly consolidated volume. Panel A is shows institutional volume (equal to the minimum of institutional buying and institutional selling) as a fraction of consolidated volume. Panel B shows dark pool volume as a fraction of this institutional volume measure. The dotted lines show results for the third quarter of 2005, the dashed lines show results for the second quarter of 2006, and the solid lines shows result for the third quarter of 2007.

Table 1
NASDAQ MPID's and volumes for the 2nd quarter of 2006
The full sample period is June 2005 through September 2007. June 2006 is the last month for which volumes are available for all NASDAQ MPIDs. The table shows identities and volumes for the three dark pools (LQNT, ITGI, and BLOK) and for an additional 17 NASDAQ MPID's, which are those with the highest volumes across all of the NASDAQ MPID's in the 2nd quarter of 2006. During the quarter, there were a total of 540 MPID's with at least some volume in the NASDAQ stocks in the sample, and there are a total of 20 MPID's shown in the table, so the "All Other " category in the tables is comprised of 540-20=520 MPID's. The "Other Exchanges" category is the difference between the total NASDAQ reported volume and the consolidated volume obtained from CRSP. The NASDAQ data excludes odd lots, so the "Other Exchanges" category also includes odd lots. All volume amounts are in thousands of shares.

		Microsoft		Am. Comm. Lines		Full Sample	
MPID	Market Participant	Volume	Pct	Volume	Pct	Volume	Pct
LQNT	Liquidnet	3,753	0.1%	948	3.2%	480,162	0.5%
ITGI	ITG	7,968	0.1	596	2.0	424,300	0.4
BLOK	Pipeline Trading Systems	3,700	0.1	72	0.2	136,600	0.1
	Subtotal Dark Pools	15,421	0.3%	1,616	5.5%	1,041,062	1.0%
INET	Island/Instinet	1,850,821	31.7	4,828	16.5	27,626,605	27.3
BRUT	BRUT ECN	635,936	10.9	2,342	8.0	10,065,329	10.0
GSCO	Goldman, Sachs	316,895	5.4	2,794	9.6	4,670,020	4.6
UBSS	UBS Securities	144,051	2.5	1,407	4.8	4,544,339	4.5
SBSH	Citigroup Global Markets	186,191	3.2	260	0.9	2,602,786	2.6
MSCO	Morgan Stanley	131,246	2.3	443	1.5	2,569,009	2.5
NITE	Knight Equity Markets	59,451	1.0	556	1.9	2,439,568	2.4
BOFA	Banc of America Securities	86,583	1.5	245	0.8	1,499,433	1.5
CDRG	Citadel Derivatives Group	31,874	0.5	97	0.3	1,320,935	1.3
BTRD	Bloomberg Tradebook	42,270	0.7	899	3.1	1,319,173	1.3
ETRD	E*Trade Capital Markets	23,905	0.4	2	0.0	1,294,558	1.3
LEHM	Lehman Brothers.	35,996	0.6	237	0.8	1,278,851	1.3
AUTO	Automated Trading Desk	69,975	1.2	191	0.7	1,243,413	1.2
NFSC	National Financial Services	82,535	1.4	0	0.0	1,237,222	1.2
FBCO	Credit Suisse Securities	70,110	1.2	334	1.1	1,195,578	1.2
MLCO	Merrill Lynch	112,735	1.9	837	2.9	1,169,574	1.2
DBAB	Deutsche Bank Securities	75,969	1.3	539	1.8	983,876	1.0
	All Other NASDAQ MPIDs	649,347	11.1	4,344	14.9	10,586,148	10.5
	Other Exchanges	1,210,882	20.8	7,274	24.9	22,394,952	22.2
	Consolidated Volume	5,832,192	100.0%	29,246	100.0%	101,082,433	100.0%

Table 2
Second quarter 2006 changes in institutional holdings
Quarterly holdings data are from the Thompson database, which is based on 13F filings. For all entities that filed reports for both March 31, 2006 and June 30, 2006, changes in holdings are calculated by comparing the number of shares held on March 31 with the number of shares held on June 30, 2006. All volume amounts are in thousands of shares.

	Microsoft		American Commercial Lines	
Investment Firm (13F filer)	Increase	Decrease	Increase	Decrease
Capital Research & Mgmt Co.	48,905			
Franklin Resources Inc.	39,805		245	
Brandes Invt. Partners, LP	34,713			
Fidelity Management & Research		123,182	917	
Wellington Management Co, LLP		60,670	2,691	
Lord, Abbett & Co. LLC		34,650		
American Cent Investment Mgmt.		13,485	1,401	
Tremblant Capital, L.P.			1,117	
Barclays Bank PLC	2,013		853	
Blackrock Inc.		6,685		1,408
Lazard Freres & Company LLC	1,055			498
Trafelet & Company, LLC				425
All others	347,179	448,679	3,843	4,971
Total for all investment firms	473,670	687,351	11,069	7,302
Total number of investment firms with non-zero change in holdings over the quarter	623	796	81	73
Total as a percent of consolidated volume for the quarter	8%	12%	38%	25%
Minimum of institutional buying and selling (as a percent of consolidated volume)	473,670 (8%)		7,302 (25%)	

Table 3

Sample characteristics by category of consolidated volume

The sample covers the nine quarters from the third quarter of 2005 through the third quarter of 2007. Each quarter, the sample includes the 500 NASDAQ stocks with the highest consolidated volumes taken from the set of stocks with a single class of common, no stock splits, share price above $2, market capitalization above $100 million, and no changes in cusip or ticker symbol. In total there are 9x500=4,500 stock/quarter observations. Each quarter, the sample stocks are sorted by total consolidated volume and placed in one of three categories: the lowest third, middle third, and top third. The values in the table reflect the averages across the nine sample quarters within each category of consolidated volume.

		Consolidated volume category			
		Lowest third	Middle third	Top third	Full Sample
Actual Dark Pool volume (percentage of cons. vol.)		2.77	2.23	1.36	2.12
Quarterly consolidated volume (millions of shares)		22.9	54.6	392.2	156.8
<u>Independent Variables</u>					
Institutional order measures (percentage of cons. vol.)					
$SMatch_{s,q}$	minimum of total purchase and total sales, excluding larger orders, as pct. of consolidated volume	16.11	13.61	10.63	13.44
$SExtra_{s,q}$	imbalance between total purchases and sales, excluding larger orders, as pct. of consolidated volume	6.12	5.20	3.59	4.97
$LMatchDiff_{s,q}$	change in Match if large orders are included	0.98	0.89	0.71	0.86
$LExtraDiff_{s,q}$	change in Extra if large orders are included	0.27	0.24	0.14	0.22
Stock Characteristics					
$\sigma_{s,q}$	daily return volatility	-0.19	0.08	0.11	0.00
$R_{s,q}$	relative spread	0.04	-0.05	0.01	0.00
Share-weighted measures of order characteristics					
$I20SRM_{s,q}$	indicator: spread rank measure <0.2	0.00	0.03	0.35	0.13
$SRM_{s,q}$	spread rank measure	0.86	0.68	0.40	0.65
$DVRM_{s,q}$	dollar value rank measure	0.79	0.79	0.83	0.80
$DP_{s,q}$	institutional position value	0.94	0.97	1.13	1.01
$TO_{s,q}$	institutional turnover	0.00	0.04	0.04	0.02

Table 4
Panel data regressions – dark pool volumes
The sample covers the nine quarters from the third quarter of 2005 through the third quarter of 2007. Each quarter, the sample includes the 500 NASDAQ stocks with the highest consolidated volumes taken from the set of stocks with a single class of common, no stock splits, share price above $2, market capitalization above $100 million, and no changes in cusip or ticker symbol. In total there are 9x500=4,500 stock/quarter observations. The dependent variable is dark pool volume (total of Liquidnet, POSIT and Pipeline) as a percent of total consolidated volume for the stock/quarter. The regressions assume stock-specific random effects and the t-statistics are based on robust estimates of the parameter variance/covariance matrix. Models 1, 2, and 3 are based on equations 2, 3, and 4 in section 6.

Independent variables			Model 1		Model 2		Model 3	
Symbol	Description	Coeff.	Est.	t-stat	Est.	t-stat	Est.	t-stat
	intercept	a	-0.051	-0.7	-0.228	-3.4	-0.268	-4.3
$SMatch_{s,q}$	matched orders	b_1	0.143	26.2				
$SMatch_{s,1}$	matched orders, 3rd quarter 2005	$b_{1,1}$			0.107	18.7	0.112	21.2
$SMatch_{s,2}$	matched orders, 4th quarter 2005	$b_{1,2}$			0.115	19.3	0.127	22.6
$SMatch_{s,3}$	matched orders, 1st quarter 2006	$b_{1,3}$			0.138	22.1	0.142	24.4
$SMatch_{s,4}$	matched orders, 2nd quarter 2006	$b_{1,4}$			0.141	21.9	0.144	24.1
$SMatch_{s,5}$	matched orders, 3rd quarter 2006	$b_{1,5}$			0.156	22.9	0.163	25.2
$SMatch_{s,6}$	matched orders, 4th quarter 2006	$b_{1,6}$			0.183	29.8	0.196	31.8
$SMatch_{s,7}$	matched orders, 1st quarter 2007	$b_{1,7}$			0.183	26.5	0.190	30.0
$SMatch_{s,8}$	matched orders, 2nd quarter 2007	$b_{1,8}$			0.184	29.1	0.189	31.6
$SMatch_{s,9}$	matched orders, 3rd quarter 2007	$b_{1,9}$			0.198	31.1	0.211	36.2
$SExtra_{s,q}$	order imbalance	b_2	0.033	8.4	0.039	10.2	0.038	10.0
$LMatchDiff_{s,q}$	change in Match incl. large orders	b_3	0.095	7.1	0.090	7.0	0.099	8.0
$LExtraDiff_{s,q}$	change in Extra incl. large orders	b_4	0.009	1.2	0.005	0.7	0.013	1.6
$\sigma_{s,q}*SMatch_{s,q}$	daily return volatility	b_5					0.006	2.5
$R_{s,q}*SMatch_{s,q}$	relative spread	b_6					0.009	3.0
$I20SRM_{s,q}*SMatch_{s,q}$	indicator: spread rank <0.2	b_7					-0.054	-7.9
$(DVRM_{s,q}-\overline{DVRM})*SMatch_{s,q}$	average dollar value rank	b_8					-0.107	-3.2
$(DP_{s,q}-\overline{DP})*SMatch_{s,q}$	average inst. position value	b_9					0.011	1.4
$(TO_{s,q}-\overline{TO})*SMatch_{s,q}$	average inst. turnover	b_{10}					-0.084	-10.4

Table 5
SMM estimates of the parameters that determine routing decisions
The sample covers the nine quarters from the third quarter of 2005 through the third quarter of 2007. Each quarter, the sample includes the 500 NASDAQ stocks with the highest consolidated volumes taken from the set of stocks with a single class of common, no stock splits, share price above \$2, market capitalization above \$100 million, and no changes in cusip or ticker symbol. In total there are 9x500=4,500 stock/quarter observations. Orders are assumed to be routed to dark pools based on the following value function:

$$V_{i,s,q} = \alpha_q + \beta X_{i,s,q} + \tilde{z}_{i,s,q} + \tilde{e}_s$$

Where $X_{i,s,q}$ is a vector containing characteristics of institution *i*, stock *s*, and the institution's order in that stock in quarter *q*. The order is routed to the dark pool if $V_{i,s,q} > 0$. The table shows SMM estimates for the six parameters contained in β.

Coefficient on:	Estimate	Standard Error	t-statistic
$\sigma_{s,q}$ (return volatility)	0.084	0.021	4.03
$R_{s,q}$ (relative spread)	0.063	0.032	1.96
$i20srm_{s,i,q}$ (low spread rank measure indicator)	-0.215	0.092	-2.34
$dvrm_{s,i,q}$ (dollar value rank measure)	-1.291	0.283	-4.56
$dp_{i,q}$ (institution's average position size)	-0.104	0.080	-1.30
$to_{i,q}$ (institution's turnover)	-0.724	0.118	-6.13

Table 6
Breakdown of simulated dark pool volumes by category of consolidated volume
The sample covers the nine quarters from the third quarter of 2005 through the third quarter of 2007. Each quarter, the sample includes the 500 NASDAQ stocks with the highest consolidated volumes taken from the set of stocks with a single class of common, no stock splits, share price above $2, market capitalization above $100 million, and no changes in cusip or ticker symbol. In total there are 9x500=4,500 stock/quarter observations. Each quarter, the sample stocks are sorted by total consolidated volume and placed in one of three categories: the lowest third, middle third, and top third. The values in the table reflect the averages across the nine sample quarters within each of these three categories of consolidated volume. The fraction of routed shares executed is the simulated dark pool volume divided by the average of the simulated total shares in the purchase orders and the simulated total shares in the sell orders routed to the dark pool. Except for the zero coefficient on the spread rank indicator in the last three lines of the table, the simulated values use the SMM parameter estimates as inputs to the probit function. In all cases, the simulated values represent averages for 150 simulations.

	Consolidated volume category			
	Lowest third	Middle third	Top third	Full Sample
Actual Dark Pool volume (percent of cons. vol.)	2.77	2.23	1.36	2.12
Main Simulation Results (using estimated parameters)				
Fraction of shares routed to Dark Pools	0.48	0.47	0.43	0.46
Fraction of routed shares executed	0.26	0.25	0.22	0.24
Simulated Dark Pool volumes (percent of cons. vol.)	2.73	2.17	1.36	2.09
Simulations setting the low spread rank indicator coefficient to 0				
Fraction of shares routed to Dark Pools	0.48	0.47	0.45	0.47
Fraction of routed shares executed	0.26	0.25	0.23	0.25
Simulated Dark Pool volumes (fraction of cons. vol.)	2.73	2.19	1.48	2.13

Table 7
Panel data regressions – ECN volumes
The sample covers the nine quarters from the third quarter of 2005 through the third quarter of 2007. Each quarter, the sample includes the 500 NASDAQ stocks with the highest consolidated volumes taken from the set of stocks with a single class of common, no stock splits, share price above $2, market capitalization above $100 million, and no changes in cusip or ticker symbol. In total there are 9x500=4,500 stock/quarter observations. The dependent variable is ECN volume (total of Automated Trading Desk and Bloomberg Tradebook) as a percent of total consolidated volume for the stock/quarter. The regressions assume stock-specific random effects and the t-statistics are based on robust estimates of the parameter variance/covariance matrix.

Independent variables				
Symbol	Description	Coeff.	Est.	t-stat
	intercept	a	2.836	38.31
$SMatch_{s,1}$	matched orders, 3rd quarter 2005	$b_{1,1}$	-0.006	-1.29
$SMatch_{s,2}$	matched orders, 4th quarter 2005	$b_{1,2}$	-0.006	-1.32
$SMatch_{s,3}$	matched orders, 1st quarter 2006	$b_{1,3}$	0.031	6.66
$SMatch_{s,4}$	matched orders, 2nd quarter 2006	$b_{1,4}$	0.024	5.02
$SMatch_{s,5}$	matched orders, 3rd quarter 2006	$b_{1,5}$	0.004	0.90
$SMatch_{s,6}$	matched orders, 4th quarter 2006	$b_{1,6}$	0.012	2.59
$SMatch_{s,7}$	matched orders, 1st quarter 2007	$b_{1,7}$	-0.025	-5.81
$SMatch_{s,8}$	matched orders, 2nd quarter 2007	$b_{1,8}$	-0.022	-4.47
$SMatch_{s,9}$	matched orders, 3rd quarter 2007	$b_{1,9}$	-0.005	-0.98
$SExtra_{s,q}$	order imbalance	b_2	0.001	0.29
$LMatchDiff_{s,q}$	change in Match incl. large orders	b_3	0.000	0.07
$LExtraDiff_{s,q}$	change in Extra incl. large orders	b_4	0.005	1.09
$\sigma_{s,q}*SMatch_{s,q}$	daily return volatility	b_5	0.009	5.10
$R_{s,q}*SMatch_{s,q}$	relative spread	b_6	0.002	0.70
$I20SRM_{s,q}*SMatch_{s,q}$	indicator: spread rank <0.2	b_7	-0.059	-9.97
$(DVRM_{s,q}-\overline{DVRM})*SMatch_{s,q}$	average dollar value rank	b_8	0.079	3.28
$(DP_{s,q}-\overline{DP})*SMatch_{s,q}$	average inst. position value	b_9	-0.012	-2.26
$(TO_{s,q}-\overline{TO})*SMatch_{s,q}$	average inst. turnover	b_{10}	0.012	1.84